Semiparametric Analysis of an Expanded Cox Proportional Hazards Model with Time-Varying Covariates

Wenying Zheng

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Ying Qing Chen, Chair

Kwun Chuen Gary Chan

Ying Huang

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

Semiparametric Analysis of an Expanded Cox Proportional
Hazards Model with Time-Varying Covariates

Wenying Zheng

Chair of the Supervisory Committee:
Affiliate Professor Ying Qing Chen
Department of Biostatistics

Time-varying covariates are often encountered in survival analysis. The Cox proportional hazards model can incorporate time-varying covariates, while the interpretation of regression parameters is less straightforward. We instead propose a complementary log-log survival model. When covariates are time-independent, the proposed model reduces to the Cox proportional hazards model; however, when they are time-varying, the proposed model provides a direct interpretation of regression parameters in the survival function. We develop semi-parametric estimation procedures based on estimating equations, and establish the asymptotic properties of the estimators for the regression parameters and survival functions. In addition, we include weight functions to the estimating equations to improve efficiency. We demonstrate the proposed methods by simulation studies and application to the Mayo Clinic Primary Biliary Cirrhosis data and data from a landmark HIV randomized prevention trial.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my immeasurable appreciation to my committee chair, Dr. Ying Qing Chen, for his guidance, encouragement and patience during the dissertation writing process. I am grateful to my committee members, Dr. Kwun Chuen Gary Chan, Dr. Ying Huang and Dr. N. David Yanez, for their helpful advise on this dissertation. In addition, I thank my family for their support and encouragement.

DEDICATION

to my family

Chapter 1
INTRODUCTION

In many clinical and epidemiological studies, the analysis interest is on the time-to-event outcome, e.g. the infant survival through age 18 months. However, not all the events would occur by the end of study. The rest would be usually considered "censored" with respect to the event. In survival analysis with censored data, the Cox [1972] proportional hazards model is most commonly used to estimate a covariate's effect. The Cox model can include many covariates of various types [Kalbfleisch and Prentice, 2002]. This is particularly appealing for controlled clinical trials, as well-designed well-conducted clinical trials usually collect much more covariates' information, with the intention to perform in-depth primary and secondary analyses. Covariates are not necessarily constant with time. Time-varying covariates are often encountered in practice since survival data are collected over a period of time. Sometimes, derived covariates, such as the interaction between baseline covariates and follow-up time, can be of interest.

In general, for time-independent covariates, the regression coefficients in the Cox model can be interpreted straightforwardly, as "relative risk" per Lehmann's alternative, but need considerable care for the time-varying covariates [Cox and Oakes, 1984]. Specifically, the challenge for the Cox model to deal with time-varying covariates lies in that it does not directly translate their coefficients into interpretable summary statistics of survival probabilities, even for the so-called "external covariates" [Kalbfleisch and Prentice, 2002]. To see this, if a covariate is the function of t, namely $Z(t)$, the Cox model usually assumes that the hazard function of the event time T given $Z(t)$ is

$$\lambda\{t|Z(t)\} = \lambda_0(t)\exp\{\beta Z(t)\},$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function, β is the regression coefficient. Then the associated cumulative hazard function becomes

$$\begin{aligned}\Lambda\{t|Z(u), 0 \leq u \leq t\} &= \int_0^t \lambda_0(u)\exp\{\beta Z(u)\}du \\ &= \int_0^t \frac{f_0(u)}{S_0(u)}\exp\{\beta Z(u)\}du \\ &= \int_0^\infty \frac{\mathrm{I}(u \leq t)}{S_0(u)}\exp\{\beta Z(u)\}f_0(u)du \\ &= \mathrm{E}_{T_0}\left[\frac{\mathrm{I}(T_0 \leq t)\exp\{\beta Z(T_0)\}}{S_0(T_0)}\right],\end{aligned}$$

where T_0 is the event time for $Z(t) \equiv 0, t \geq 0$, f_0 is the density function of T_0, and $S_0(t) = \Pr\{T_0 > t\}$. The corresponding survival function is then of the form

$$S\{t|Z(u), 0 \leq u \leq t\} = \exp\left(-\mathrm{E}_{T_0}\left[\frac{\mathrm{I}(T_0 \leq t)\exp\{\beta Z(T_0)\}}{S_0(T_0)}\right]\right). \tag{1.1}$$

Unless $Z(t)$ is of some particular form, such as constant, the interpretation of β in (1.1) is no longer in relative risk, as it can hardly be expressed by some simple function of survival probabilities.

In this dissertation, we instead propose a natural extension of the Cox proportional hazards model to study the effect of time-varying covariates on the survival function. To be general, now let $\boldsymbol{Z}(t)$ denote the value of a p-vector of possibly time-varying covariates measured at time t, and $\widetilde{\boldsymbol{Z}}(t) = \{\boldsymbol{Z}(s) : 0 \leq s \leq t\}$ be the covariate history up to t. Our proposed model is in a complementary log-log form of the survival function, assuming that

$$\log[-\log\{S(t|\widetilde{\boldsymbol{Z}})\}] = \log[-\log\{S_0(t)\}] + \boldsymbol{\beta}^T\boldsymbol{Z}(t), \tag{1.2}$$

where $S(\cdot|\widetilde{\boldsymbol{Z}})$ is the survival function of T given the covariate history $\widetilde{\boldsymbol{Z}}(\cdot)$, and $\boldsymbol{\beta}$ is the associated p-vector of regression coefficients. Apparently, the proposed expanded model is the Cox model when covariates are all time-independent. When time-varying covariates are present, model (1.2) has an advantage in the relative risk interpretation while the Cox model does not. Model (1.2) can also be considered the complementary log-log survival model, under the general framework of Peng and Huang [2007]. In Peng and Huang's paper, the focus is nevertheless on the development of inference procedures for the varying-coefficients in the complementary log-log survival model. That is, the regression coefficients are time-varying but the covariates remain time-independent. To our knowledge, this dissertation is the first study to incorporate time-varying covariates in the complementary log-log survival model.

The rest of the dissertation is organized as follows. Chapter 2 describes the scientific and theory background of this dissertation. Specifically, we review a class of semiparametric regression models for the analysis of censored survival data and the corresponding inference procedures for the regression coefficients. We discuss the role of time-varying covariates in the survival analysis and present real-life examples of the time-varying covariates. We introduce some basic concepts of the counting process theory that are relevant to this dissertation. In addition, we give a brief review of some classical estimation methods for conducting inference about the survival function.

In Chapter 3, we propose the semiparametric inference procedures for estimation of model (1.2) and study the asymptotic properties of the resulting estimators. Specifically, we formulate a class of martingale-based estimating equations for the unknown parameters in the proposed model, and derive estimators by solving these equations. We also study the weighted estimating equations. The weight functions can be chosen to improve the estimation efficiency. On the other hand, we develop the pointwise confidence intervals for the subject-specific survival curves under the proposed model. We show that the estimators for the regression parameters and survival functions are consistent and asymptotically normal using standard counting process arguments.

Simulation studies that demonstrate our methods are given in Chapter 4. Specifically, we make numerical comparison of the proposed estimator with the Cox maximum partial likelihood estimator [Cox, 1972, 1975] for the regression parameter in the complementary log-log survival model with time-independent covariates. We also make numerical comparison of the proposed unweighted estimator with the proposed optimal weighted estimator for the regression parameter in the complementary log-log survival model with time-varying covariates. Moreover, we carry out extensive simulations that evaluate the performance of the proposed confidence interval procedures for the survival function for practical sample sizes.

In Chapter 5, we apply the proposed methods to the well-known Mayo Clinic Primary Biliary Cirrhosis data and data from a landmark HIV randomized prevention trial. Chapter 6 contains some closing remarks and a discussion of future research.

Chapter 2

BACKGROUNDS

2.1 Semiparametric Survival Models

Time to event data are often collected in clinical and epidemiological studies. Survival models have been developed to analyze such data, and are in widespread use to identify risk factors, to investigate covariate effects and to classify individuals into groups.

In practice, the Cox [1972] proportional hazards model is most commonly used. The Cox proportional hazards model has been extensively studied and its asymptotic properties have been well justified with the counting process martingale theory [Cox, 1972, 1975, Tsiatis, 1981, Andersen and Gill, 1982]. Let T be the failure time and $\boldsymbol{Z}$ be the associated p-dimensional vector of covariates. The Cox model specifies that, the hazard function of T given $\boldsymbol{Z}$ is

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

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function, $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression parameters and the superscript T denotes matrix transpose. The Cox model is viewed as a semiparametric model in that it consists of a non-parametric baseline and a parametric regression function. The baseline distribution is in general unknown in biomedical applications, thus it is desirable to leave the baseline hazard unspecified. The hazard function measures the instantaneous risk of failure [Kalbfleisch and Prentice, 2002], defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}.$$

In the Cox proportional hazards model, $\boldsymbol{\beta}$ estimates the hazard ratio associated with per unit change in $\boldsymbol{Z}$, which is assumed to be constant over time. In contrast to the Cox proportional hazards model, the additive risk model assumes an additive covariate effect on the hazard function [Lin and Ying, 1994], that is

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t) + \boldsymbol{\beta}^T \boldsymbol{Z}.$$

The additive risk model is also semiparametric as the baseline hazard function $\lambda_0(t)$ is unspecified.

Survival function, the probability that the subject survives beyond a given time point $\Pr(T > t)$, is a key measure of the survival experience. The survival function is related to the hazard function via

$$S(t) = \exp\{-\int_0^t \lambda(s) ds\}.$$

Note that

$$\Lambda(t) = \int_0^t \lambda(s) ds$$

is termed as the cumulative hazard function. Modeling survival functions directly provides a different framework for survival analysis in contrast to the hazard-based models, and gives direct interpretations of covariate effects on the survival function. The proportional odds model specifies that the log-odds for survival conditional on covariates is the sum of the unknown baseline log-odds and a regression function of covariates [Pettitt, 1982, Bennett, 1983],

$$\log \frac{S(t|\boldsymbol{Z})}{1 - S(t|\boldsymbol{Z})} = \log \frac{S_0(t)}{1 - S_0(t)} + \boldsymbol{\beta}^T \boldsymbol{Z},$$

where $\boldsymbol{\beta}$ estimates the odds ratio associated with per unit change in $\boldsymbol{Z}$. One important property of the proportional odds model is that the hazard ratio comparing any two covariate groups converges to one as time increases [Murphy et al., 1997]. The proportional odds model provides a useful alternative to the Cox proportional hazards model in the case of converging hazards. Based on the relationship between survival function and hazard function, the Cox proportional hazards model is often used to render inference about the survival function. When covariates are all time-independent, the Cox proportional hazards model can be reformulated as

$$\log[-\log\{S(t|\boldsymbol{Z})\}] = \log[-\log\{S_0(t)\}] + \boldsymbol{\beta}^T \boldsymbol{Z},$$

which is referred to as the complementary log-log survival model [Peng and Huang, 2007].

The linear transformation model can be viewed as a unification and generalization of many classical survival regression models [Chen et al., 2002]. Both the Cox proportional hazards model and the proportional odds model are special cases of the linear transformation model. The linear transformation model assumes that

$$\mathrm{H}(T) = -\boldsymbol{\beta}^T \boldsymbol{Z} + \epsilon,$$

where H is an unspecified monotone function and ϵ is a random variable whose distribution is specified and independent of $\boldsymbol{Z}$. For example, the extreme value and standard logistic distributions for ϵ correspond to the Cox proportional hazards and proportional odds models, respectively [Chen et al., 2002]. The accelerated failure time model generalizes the log linear regression model to the survival data [Buckley and James, 1979, Kalbfleisch and Prentice, 2002], which specifies that

$$\log(T) = \boldsymbol{\beta}^T \boldsymbol{Z} + e,$$

where e is a random variable with an unspecified distribution. Although the accelerated failure time model has a sound interpretation of regression parameters in the failure time, it is seldom used in biomedical applications because the inference procedures are in general complex with censored data [Wei, 1992].

2.2 Time-Varying Covariates

In survival analysis, the value of covariate may change with time. Such covariates are referred to as the time-varying covariates. Consider the following real examples.

Example 2.2.1. The Stanford Heart Transplantation program developed a valuable database that consists of 103 patients admitted to the program between September 1967 and March 1974 [Miller and Halpern, 1982]. One important objective of the study is to investigate the effect of heart transplantation on the survival. Some patients received transplant over follow-up, and other patients did not receive transplant due to the death or the closure of study on April 1, 1974. The waiting time to transplant also varied among heart-transplanted patients, from a few days to a year. Patients who survived longer had a greater chance of receiving transplant. Hence the indicator of heart transplant should be treated as time-varying covariate in the analysis.

Example 2.2.2. The HIVNET 012 randomized trial was conducted to assess the efficacy and safety of short-course nevirapine versus short-course zidovudine for the prevention of mother-to-child transmission of human immunodeficiency virus type-1 (HIV-1) between November 1997 and January 2001 [Jackson et al., 2003]. The study investigators also aimed to compare the 18-month infant HIV-1 free survival between the two treatment groups. The mother-to-child transmission of HIV-1 is likely to occur through postnatal breast-feeding. Maternal HIV-RNA viral loads and maternal CD4+ counts collected over follow-up after delivery are possible confounders and should be adjusted for as time-varying covariates in the analysis.

On the other hand, time-varying covariates could be interactions of covariates and some function of time. This kind of time-varying covariates is useful for checking the adequacy of survival models. For example, the proportional hazards assumption in the Cox proportional hazards model implies that the difference in $\log(-\log)$ survival probability between groups is constant over time. A hypothesis test on the covariate-time interaction in the complementary log-log survival model can serve as a goodness-of-fit tool for the Cox proportional hazards model. Also temporal covariate effect can be assumed by incorporating the interaction of time-independent covariate and specific function of time. For example, consider the following model,

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp\{\beta \times Z \times \log(t)\},$$

where Z takes values 0 or 1. Then the hazard ratio for $Z = 1$ compared to $Z = 0$ is of the form t^β, which increases or decreases with time.

Time-varying covariates can be classified into "external" and "internal" based upon whether or not the covariate depends on the individual's survival process [Kalbfleisch and Prentice, 2002]. The value of external time-varying covariate is independent of the individual's survival status. Mathematically, Kalbfleisch and Prentice [2002] define the external time-varying covariate as follows

$$\Pr\{Z(s)|Z(t), T \geq t\} = \Pr\{Z(s)|Z(t), T = t\},$$

for any $s > t$, where $Z(t)$ denotes the value of covariate Z observed at time t. By contrast, the internal time-varying covariate can only be obtained when the individual is alive and thus the existence of the values of internal time-varying covariates implies that the individual was alive when the values were collected. The time-varying covariates in the survival model are limited to be "external" in order for the interpretation of survival model to be meaningful. In the HIVNET 012 example, the follow-up maternal HIV-RNA viral loads and CD4+ counts are external time-varying covariates when investigating the infant survival, but would be "internal" if the outcome is the maternal survival.

Survival models with time-varying covariates have received considerable attention in recent years. Let $\widetilde{\boldsymbol{Z}}(t) = \{\boldsymbol{Z}(s) : 0 \leq s \leq t\}$ be the covariate history up to time t. The Cox proportional hazards model can incorporate time-varying covariates by assuming

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

where $\lambda(\cdot|\widetilde{\boldsymbol{Z}})$ is the hazard function of T given the covariate history $\widetilde{\boldsymbol{Z}}(\cdot)$ [Kalbfleisch and Prentice, 2002]. However the interpretation of regression parameter is less straightforward under the Cox model with time-varying covariates. When time-varying covariates are present, the Cox model implies that

$$\log\{S(t|\widetilde{\boldsymbol{Z}})\} = -\int_0^t \lambda_0(s) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}(s)\} ds,$$

in which $\boldsymbol{\beta}$ does not provide a direct interpretation of covariate effect on the survival function. It is then apparent that $\boldsymbol{\beta}$ in the Cox model estimates the difference in $\log(-\log)$ survival probability associated with per unit change in Z only when covariates are all time-invariant.

Lin and Ying [1994] developed semiparametric estimation methods for the additive risk model that allows time-varying covariates,

$$\lambda(t|\widetilde{\boldsymbol{Z}}) = \lambda_0(t) + \boldsymbol{\beta}^T \boldsymbol{Z}(t).$$

Zeng and Lin [2006, 2007] extended the class of the linear transformation model to include the time-varying covariates. This extended transformation model specifies that

$$\mathrm{A}(t|\widetilde{\boldsymbol{Z}}) = \mathrm{G}\Big[\int_0^t \mathrm{I}(T \geq s)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}(s)\}d\mathrm{A}(s)\Big],$$

where $\mathrm{A}(.)$ is an unspecified increasing function, $\mathrm{G}(.)$ is a specified strictly increasing and continuously differentiable function and $\mathrm{I}(.)$ is the indicator function, which reduces to the usual linear transformation model when covariates are all time-invariant,

$$\log \mathrm{A}(T) = -\boldsymbol{\beta}^T\boldsymbol{Z} + \log \mathrm{G}^{-1}(-\log \epsilon_0),$$

where ϵ_0 follows a specified uniform distribution. Most recently, Chen et al. [2012] studied the proportional odds model in the presence of time-varying covariates,

$$\log \frac{S(t|\widetilde{\boldsymbol{Z}})}{1 - S(t|\widetilde{\boldsymbol{Z}})} = \log \frac{S_0(t)}{1 - S_0(t)} + \boldsymbol{\beta}^T\boldsymbol{Z}(t),$$

and well established the asymptotic properties of their estimators via the counting process theory. Peng and Huang [2007] developed the varying-coefficient complementary log-log survival model,

$$\log[-\log\{S(t|\boldsymbol{Z})\}] = \log[-\log\{S_0(t)\}] + \boldsymbol{\beta}(t)^T\boldsymbol{Z},$$

where unknown regression coefficients $\boldsymbol{\beta}(t)$ vary with time and covariates $\boldsymbol{Z}$ are time-invariant. In the existing literature, the complementary log-log survival model is confined to time-independent covariates. This dissertation aims to fill this important gap by proposing and studying the complementary log-log survival model that includes the time-varying covariates.

2.3 Introduction of Counting Process Martingale Theory

The counting process martingale theory plays an important role in survival analysis. Censored survival data can be expressed in terms of counting processes and the theoretical justifications of many survival methods involve the use of counting process martingale theory. Fleming and Harrington [1991] gives a complete and comprehensive description regarding the counting process martingale theory. Here, we simply give a brief introduction. All the definitions and theorems mentioned in this section can be found in Fleming and Harrington [1991], Chapter 1, 2 and 5.

The martingale is a special case of stochastic process. Let us begin with the definition of stochastic process. Consider the probability space $(\Omega, \mathcal{F}, P)$, where Ω is a space of outcomes of a random experiment, $\mathcal{F}$ is a σ-algebra of subsets of Ω and P is a probability measure.

Definition 2.3.1. *A (real-valued) stochastic process is a family of random variables* $X = \{X(t) : t \in \Gamma\}$ *indexed by a set* Γ*, all defined on the same probability space* $(\Omega, \mathcal{F}, P)$*.*

Definition 2.3.2. *A family of sub-*σ*-algebras* $\{\mathcal{F}_t : t \geq 0\}$ *of a* σ*-algebra* $\mathcal{F}$ *is called increasing if* $s \leq t$ *implies* $\mathcal{F}_s \subset \mathcal{F}_t$ *(i.e., if for* $s \leq t$*,* $A \in \mathcal{F}_s$ *implies* $A \in \mathcal{F}_t$*). An increasing family of sub-*σ*-algebras is called a filtration.*

Definition 2.3.3. *A stochastic process* $\{X(t) : t \geq 0\}$ *is adapted to a filtration if, for every* $t \geq 0$*,* $X(t)$ *is* $\mathcal{F}_t$*-measurable.*

In other words, a stochastic process is a measurable mapping from $\Gamma \times \Omega$ to the real line $\mathcal{R}$. For example, let $\{M(t) : t \geq 0\}$ denote a stochastic process defined on the probability space; then for each fixed $t \geq 0$, $M(t) : \Omega \to \mathcal{R}$ is a random variable with domain Ω and range $\mathcal{R}$. A commonly used filtration is defined as the history path of stochastic process collected up to t, $\mathcal{F}_t = \sigma\{M(s) : 0 \leq s \leq t\}$. It is then apparent that $M(t)$ is adapted to $\mathcal{F}_t$.

The concept of the martingale is not complex. A martingale is basically a stochastic process with some additional characteristics.

Definition 2.3.4. *Let* $M = \{M(t) : t \geq 0\}$ *be a right-continuous stochastic process with left-hand limits and* $\{\mathcal{F}_t : t \geq 0\}$ *a filtration, defined on a common probability space.* M *is called a martingale with respect to* $\{\mathcal{F}_t : t \geq 0\}$ *if*

1. M *is adpted to* $\{\mathcal{F}_t : t \geq 0\}$*,*

2. $\mathrm{E}|M(t)| < \infty$*,*

3. $\mathrm{E}\{M(t+s)|\mathcal{F}_t\} = M(t)$ *a.s. for all* $s \geq 0, t \geq 0$.

The definition of a martingale implies an important property which is central in the methodological development of survival analysis, that is $\mathrm{E}\{dM(t)|\mathcal{F}_{t}-\} = 0$. Intuitively speaking, given the history path up to but not including t, the increments of a martingale have expectation zero. Let C denote the censoring variable and be independent of the failure time T. Let $X = \min(T, C)$ and $\Delta = \mathrm{I}(T \leq C)$, where $\mathrm{I}(\cdot)$ is the indicator function. Define $N(t) = \mathrm{I}(X \leq t, \Delta = 1)$, $Y(t) = \mathrm{I}(X \geq t)$, $\mathcal{F}_t = \sigma\{N(s), Y(s), \widetilde{\boldsymbol{Z}}(s) : 0 \leq s \leq t\}$ and

$$M(t) = N(t) - \int_0^t Y(s)\lambda(s)ds,$$

where $\lambda(\cdot)$ is the hazard function of T. Then $\{M(t) : t \geq 0\}$ is a martingale with respect to the filtration $\{\mathcal{F}_t : t \geq 0\}$. This is an important example of martingale in the survival analysis, and is discussed in Section 1.3 of Fleming and Harrington [1991]. Specifically, Theorem 1.3.1 of Fleming and Harrington [1991] states the claim of $\{M(t) : t \geq 0\}$ being a martingale and the corresponding proof is rigorous.

There are other useful examples of stochastic processes.

Definition 2.3.5. *A counting process is a stochastic process* $\{N(t) : t \geq 0\}$ *adapted to a filtration* $\{\mathcal{F}_t : t \geq 0\}$ *with* $N(0) = 0$ *and* $N(t) < \infty$ *a.s., and whose paths are with probability one right-continuous, piecewise constant, and have only jump discontinuities, with jumps of size +1.*

Definition 2.3.6. *A process* X *is called predictable with respect to a filtration if, as a mapping from* $[0, \inf) \times \Omega$ *to* $\mathcal{R}$*, it is measurable with respect to the predictable* σ*-algebra generated by that filtration. We call* X *an* $\mathcal{F}_t$*-predictable process.*

Theorem 2.3.1. *Let* N *be a counting process with* $\mathrm{E}N(t) < \infty$ *for any* t*. Let* $\{\mathcal{F}_t : t \geq 0\}$ *be a right-continuous filtration such that*

1. $M = N - A$ *is an* $\mathcal{F}_t$*-martingale, where* $A = \{A(t) : t \geq 0\}$ *is an increasing* $\mathcal{F}_t$*-predictable process with* $A(0) = 0$*;*

2. H is a bounded, $\mathcal{F}_t$-predictable process.

Then the process L given by

$$L(t) = \int_0^t H(u)dM(u)$$

is an $\mathcal{F}_t$-martingale.

Intuitively speaking, the counting process records the number of events observed by time t. Given history up to but not including t, the predictable process is deterministic at t. In the preceding example, $\{N(t) = \mathrm{I}(X \leq t, \Delta = 1) : t \geq 0\}$ is a counting process recording the number of observed failure times, but is not a predictable process; $\{Y(t) = \mathrm{I}(X \geq t) : t \geq 0\}$ is a predictable process indicating whether the subject is at risk at t. Theorem 2.3.1 claims that processes of from $\{\int_0^t H(u)dM(u) : t \geq 0\}$ are also martingales. In fact, many survival data statistics are functions of martingales. It is appealing that the transformation preserves the martingale structure as long as H is a bounded predictable process.

Finally, we present some theorems that are used to calculate the variances and covariances of martingales.

Theorem 2.3.2. *Let M be a right-continuous martingale with respect to a right-continuous filtration $\{\mathcal{F}_t : t \geq 0\}$ and assume $\mathrm{E}M^2(t) < \infty$ for any $t \geq 0$. Then there exists a unique increasing right-continuous predictable process $< M, M >$, called the predictable quadratic variation of M, such that $< M, M > (0) = 0$ a.s., $\mathrm{E} < M, M > (t) < \infty$ for each t, and $\{M^2(t) - < M, M > (t) : t \geq 0\}$ is a right-continuous martingale.*

Theorem 2.3.3. *Let M_1 and M_2 be two right-continuous martingales with respect to a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$, and assume $\mathrm{E}M_i^2(t) < \infty$ for $t \geq 0$ and $i = 1, 2$. Then there exists a right-continuous predictable process $< M_1, M_2 >$, called a predictable covariation process, with $< M_1, M_2 > (0) = 0$, $\mathrm{E} < M_1, M_2 > (t) < \infty$, such that*

1. $< M_1, M_2 >$ is the difference of two increasing right-continuous predictable processes, and

2. $M_1 M_2 - < M_1, M_2 >$ *is a martingale.*

Theorem 2.3.4. *Let* $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$ *be a stochastic basis, and assume:*

1. H_i *is a bounded* $\mathcal{F}_t$*-predictable process,*

2. N_i *is a bounded counting process,*

3. *the* $\mathcal{F}_t$*-martingale* $M_i = N_i - A_i$ *satisfies* $\mathrm{E}M_i^2(t) < \infty$ *for any* t*,*

then

$$< \int H_1 \, dM_1, \int H_2 \, dM_2 >= \int H_1 H_2 \, d < M_1, M_2 > .$$

The predictable quadratic variation process is used to calculate the variance of the martingale. It satisfies

$$d < M, M > (t) = \mathrm{var}\{dM(t) | \mathcal{F}_t -\}.$$

Similary,

$$d < M_1, M_2 > (t) = \mathrm{cov}\{dM_1(t), dM_2(t) | \mathcal{F}_t -\}.$$

The covariance between two martingales can be obtained through the predictable covariation process. In particular, M_1 and M_2 are said to be uncorrelated if $< M_1, M_2 >= 0$ and $M_1(0), M_2(0)$ are uncorrelated. It is also straightforward to calculate the variance and covariance of martingale transformations with Theorem 2.3.4.

Moreover, similar to the well-known central limit theorem for random variables, there is an elegant Martingale Central Limit Theorem that establishes the large sample properties for sequences of martingales, and is frequently used in the theoretical development of statistical methods involving martingales. The Chapter 5 of Fleming and Harrington [1991] states and discusses the Martingale Central Limit Theorem and the corresponding regularity conditions at great length. The main results of the Martingale Central Limit Theorem are summarized in Theorem 5.2.3 of Fleming and Harrington [1991], and a rigorous proof is provided accordingly.

2.4 Estimation Methods for Semiparametric Survival Model

2.4.1 Partial Likelihood Approach

In general, inference about the Cox proportional hazards model is based on the partial likelihood. Cox [1972, 1975] proposed the partial likelihood approach that focused on the estimation of regression coefficients and ignored the baseline hazard. The baseline hazard is often of less interest in practice and thus it is desirable to provide estimates only for the regression coefficients $\boldsymbol{\beta}$.

Suppose that there are no ties in the observed failure times. Let t_i be the ith ordered failure time such that $t_1 < t_2 < \ldots < t_{k-1} < t_k$, where k is the number of distinct observed failure times. Denote $\boldsymbol{Z}_{(i)}$ the vector of covariates for the subject who fails at time t_i, $i = 1, 2, \ldots, k$. The risk set R_i, $i = 1, 2, \ldots, k$ is defined as the subjects who survive beyond time t_i, i.e. the subjects with observed survival times greater than or equal to time t_i. The partial likelihood is given by

$$\mathcal{PL}(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_{(i)}\}}{\sum_{j \in R_i} \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j)}.$$

In fact, the partial likelihood is the product of conditional probabilities that the subject with $\boldsymbol{Z}_{(i)}$ and not the others failed at time t_i given the risk set R_i, for $i = 1, 2, \ldots, k$. Cox [1972] maximized the partial likelihood to estimate the regression coefficients $\boldsymbol{\beta}$; that is, the maximum partial likelihood estimator solves

$$\frac{\partial \log \mathcal{PL}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{k} \left\{ \boldsymbol{Z}_{(i)} - \frac{\sum_{j \in R_i} \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j) \boldsymbol{Z}_j}{\sum_{j \in R_i} \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j)} \right\} = 0.$$

The inverse of the second derivative of the partial log-likelihood with respect to $\boldsymbol{\beta}$ is then used to estimate the covariance matrix of the maximum partial likelihood estimator. Also, partial likelihood ratio tests can be conducted that are similar to the usual likelihood ratio tests.

The partial likelihood approach is efficient, flexible, convenient and has sound theoretical justification. It allows time-varying covariates. The resulting maximum partial likelihood estimator reaches the semiparametric efficiency bound. Andersen and Gill [1982] beautifully established the large-sample properties of the maximum partial likelihood estimator with the counting process martingale theory. In addition, the maximum partial likelihood estimator along with the Breslow [1972] estimator of the cumulative baseline hazard function can be used to estimate the subject-specific survival curves. More importantly, the computations of the partial likelihood approach can be easily carried out using many commercial packages. One drawback of the partial likelihood approach is that the assumption of no ties in the observed failure times is often violated in practice. However, ties would not cause practical problems when the number of failure times is modest. There are also approximation methods to deal with the ties in the partial likelihood [Kalbfleisch and Prentice, 2002].

2.4.2 Maximum Likelihood Method

Murphy et al. [1997] studied the maximum likelihood estimation method for the proportional odds model. Unlike the partial likelihood, the maximum likelihood estimation involves both the baseline and regression parameters. The estimators for the parameters are obtained by maximizing the likelihood in which the cumulative baseline hazard function is regarded as a step function with jumps at the observed failure time.

Zeng and Lin [2007] proposed a general maximum likelihood estimation method for the class of semiparametric regression models with right-censored survival data. The resulting estimators are referred to as the non-parametric maximum likelihood estimators. In fact, the maximum likelihood estimator studied by Murphy et al. [1997] is the non-parametric maximum likelihood estimator for the proportional odds model. Zeng and Lin [2007] maximized the likelihood with respect to the baseline and regression parameters, and treated the cumulative baseline hazard function as a step function with jumps at the observed failure time to ensure the existence of a maximizer. The inference procedures proposed by Zeng and Lin [2007] can be applied to a variety of semiparametric survival models with censored observations. In particular, Chen et al. [2012] studied the non-parametric maximum likelihood estimator for the proportional odds model in the presence of time-varying covariates, following the general non-parametric maximum likelihood estimation procedures.

Let $X = \min(T, C)$ and $\Delta = \mathrm{I}(T \leq C)$, where $\mathrm{I}(\cdot)$ is the indicator function. Define $N(t) = \mathrm{I}(X \leq t, \Delta = 1)$, $Y(t) = \mathrm{I}(X \geq t)$. In general, with the hazard function $\lambda(t|\boldsymbol{Z})$, the likelihood is proportional to

$$\prod_{i=1}^{n} \prod_{t \leq \tau} \big\{ Y_i(t)\lambda(t|\boldsymbol{Z}_i) \big\}^{dN_i(t)} \exp\big\{ - \int_0^{\tau} Y_i(t)\lambda(t|\boldsymbol{Z}_i)dt \big\}.$$

The explicit expressions of $\lambda(t|\boldsymbol{Z})$ vary under different semiparametric regression models but all involve the unknown cumulative baseline hazard Λ_0 and regression coefficients $\boldsymbol{\beta}$. Mathematically, unless the cumulative baseline hazard Λ_0 is discrete, the maximizer of the likelihood does not exist. Thus, for each observed failure time t, $d\Lambda_0(t)$ is replaced by the jump size of Λ_0 at t, $\Delta\Lambda_0(t) = \Lambda_0(t) - \Lambda_0(t-)$. Then, the non-parametric maximum likelihood estimators are obtained by maximizing the likelihood with respect to $\boldsymbol{\beta}$ and $\Delta\Lambda_0(\cdot)$. To estimate the covariance matrix of non-parametric maximum likelihood estimator, one may invert the second derivative of log-likelihood. One may also use the profile likelihood method [Murphy and Van der Vaart, 2000], which is easier to calculate, to estimate the covariance matrix of non-parametric maximum likelihood estimator for $\boldsymbol{\beta}$.

The non-parametric maximum likelihood estimators are shown to be consistent, asymptotic normal and asymptotically efficient [Murphy et al., 1997, Zeng and Lin, 2007], which is of great importance in practical inferences. The asymptotic properties of non-parametric maximum likelihood estimators have been well justified via the modern empirical process theory [Van der Vaart and Wellner, 1996] and semiparametric efficiency theory [Bickel et al., 1993]. One major limitation of the non-parametric maximum likelihood estimation inference procedures is that the method is computationally intensive. Since the cumulative baseline hazard function Λ_0 is regarded as a step function with positive jumps at all the observed failure times, the maximum likelihood estimation involves a large number of parameters, that is, p regression parameters and k baseline parameters, where p is the number of covariates and k is the number of observed failure times. In general, it would require considerable computation to derive estimators for a high-dimensional vector of parameters in the maximum likelihood estimation.

2.4.3 Estimating Equation Method

Consider any function $\boldsymbol{G}$ of the unknown parameter $\boldsymbol{\beta}$ and random variable $\boldsymbol{X}$. If $\boldsymbol{G}$ is unbiased at the true parameter value $\boldsymbol{\beta}_0$, that is,

$$\mathrm{E}\,\boldsymbol{G}(\boldsymbol{X}|\boldsymbol{\beta}_0) = \mathbf{0}.$$

Then the estimating equation estimator $\widehat{\boldsymbol{\beta}}_n$ for $\boldsymbol{\beta}$ is defined as the solution to

$$\sum_{i=1}^{n} \boldsymbol{G}(\boldsymbol{X}_i|\boldsymbol{\beta}) = \mathbf{0},$$

that is, $\widehat{\boldsymbol{\beta}}_n$ satisfies

$$\sum_{i=1}^{n} \boldsymbol{G}(\boldsymbol{X}_i|\widehat{\boldsymbol{\beta}}_n) = \mathbf{0}.$$

In general, the estimating function is the weighted sum of observed minus expected. For example, the well-known least square estimator for the linear regression model $Y = \beta_0 + \boldsymbol{\beta}^T\boldsymbol{Z} + \varepsilon$ belongs to the class of estimating equation estimators, which is obtained by solving the following least square estimating equations

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \beta_0 - \boldsymbol{\beta}^T\boldsymbol{Z}_i) &= 0,\\ \sum_{i=1}^{n}\boldsymbol{Z}_i\,(Y_i - \beta_0 - \boldsymbol{\beta}^T\boldsymbol{Z}_i) &= \mathbf{0}.\end{aligned}$$

In survival analysis with censored data, estimating equation procedures are often used to estimate the parameters in survival models. The estimating equation estimators are generally easy to compute and take much less computation time than the non-parametric maximum likelihood estimators. However, they are not as efficient as the non-parametric maximum likelihood estimators although the efficiency can be improved by including proper weight functions to the estimating functions.

Cheng et al. [1995] proposed estimating equation procedures for the linear transformation model

$$\mathrm{H}(T) = -\boldsymbol{\beta}^T \boldsymbol{Z} + \epsilon.$$

The proposed estimating equations are based on the expectations of the dichotomous variables $\{\mathrm{I}(T_i \geq T_j), i \neq j = 1, 2, \ldots, n\}$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} w(\boldsymbol{Z}_{ij}^T \boldsymbol{\beta}) \boldsymbol{Z}_{ij} \left\{ \frac{\Delta_j \mathrm{I}(X_i \geq X_j)}{\hat{G}^2(X_j)} - \xi(\boldsymbol{Z}_{ij}^T \boldsymbol{\beta}) \right\} = \mathbf{0},$$

where $w(\cdot)$ are positive weights, $X_i = \min(T_i, C_i)$, $\Delta_i = \mathrm{I}(T_i \leq C_i)$, $\boldsymbol{Z}_{ij} = \boldsymbol{Z}_i - \boldsymbol{Z}_j$, $\hat{G}$ is the Kaplan-Meier [1958] estimator for the survival function, $\mathrm{I}(\cdot)$ is the indicator function and

$$\xi(s) = \int_{-\infty}^{\infty} \{1 - F(t + s)\} dF(t)$$

with F the specified distribution function of ϵ. Cheng et al. [1995] also provided rigorous theoretical justification for their estimating equation procedures. However, the validity of the proposed procedures relies on the assumption that the censoring distribution is independent of the covariate $\boldsymbol{Z}$ and this assumption is often violated in biomedical studies.

Yang and Prentice [1999] proposed a class of estimating equation estimators using the weighted empirical odds functions for the estimation of the proportional odds model. Their estimating equation procedures are based on the weighted empirical processes,

$$\begin{aligned}
\widehat{H}_{nj}(t) &= \sum_{i=1}^{n} Z_{ij} \Delta_i \mathrm{I}(X_i \leq t), \\
\widehat{H}_{n\gamma j}(t; \boldsymbol{\beta}) &= \sum_{i=1}^{n} Z_{ij} \gamma(\boldsymbol{\beta}, \boldsymbol{Z}_i) \Delta_i \mathrm{I}(X_i \leq t), \\
\widehat{K}_{nj}(t) &= \sum_{i=1}^{n} Z_{ij} \mathrm{I}(X_i \geq t),
\end{aligned}$$

where for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, p,$ $\gamma(\boldsymbol{\beta}, \boldsymbol{Z}_i) = \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i),$ $X_i = \min(T_i, C_i),$ $\Delta_i = \mathrm{I}(T_i \leq C_i)$ and Z_{ij} is the jth component of covariate $\boldsymbol{Z}_i$. They defined the pseudo-maximum likelihood estimator for the regression parameters in the proportional odds model by solving

$$\sum_{i=1}^{n} \boldsymbol{Z}_i \frac{\Delta_i \gamma(\boldsymbol{\beta}, \boldsymbol{Z}_i) - \widehat{R}_{n0}(X_i; \boldsymbol{\beta})}{\gamma(\boldsymbol{\beta}, \boldsymbol{Z}_i) + \widehat{R}_{n0}(X_i; \boldsymbol{\beta})} = \mathbf{0},$$

where,

$$\begin{aligned}
\widehat{R}_{nj}(t;\boldsymbol{\beta}) &= \frac{1}{\widehat{P}_{nj}(t)}\int_{s\le t}\widehat{P}_{nj-}(s)\widehat{\Lambda}_{n\gamma j}(ds;\boldsymbol{\beta}),\\
\widehat{\Lambda}_{nj}(t) &= \int_0^t \frac{1}{\widehat{K}_{nj}(s)}d\widehat{H}_{nj}(s),\\
\widehat{P}_{nj}(t) &= \prod_{s\le t}\{1-\Delta\widehat{\Lambda}_{nj}(s)\},\\
\widehat{\Lambda}_{n\gamma j}(t;\boldsymbol{\beta}) &= \int_0^t \frac{1}{\widehat{K}_{nj}(s)}\widehat{H}_{n\gamma j}(ds;\boldsymbol{\beta}),
\end{aligned}$$

with $\Delta\widehat{\Lambda}_{nj}(s)=\widehat{\Lambda}_{nj}(s)-\widehat{\Lambda}_{nj-}(s)$. In addition, they defined a class of estimators as the zero of the following functions

$$\int_0^\infty\{\widehat{R}_{nj}(t;\boldsymbol{\beta})-\widehat{R}_{n0}(t;\boldsymbol{\beta})\}\Phi_j(dt;\boldsymbol{\beta}), j=1,2,\ldots,p,$$

where $\Phi(\cdot)$ is a specified function to stabilize the integration. Yang and Prentice [1999] established large-sample properties of the resulting estimators; the corresponding variance estimator takes a closed form.

Motivated by the fact that the increments of a martingale have expectation zero given the history path, various authors have considered martingale-based estimating equation methods for model estimation. Lin and Ying [1994] used martingale-based estimating equations to estimate the regression parameters in the additive risk model,

$$\lambda(t|\widetilde{\boldsymbol{Z}})=\lambda_0(t)+\boldsymbol{\beta}^T\boldsymbol{Z}(t).$$

Chen et al. [2002] developed semiparametric estimation procedures based on martingale functions for the linear transformation model

$$\mathrm{H}(T)=-\boldsymbol{\beta}^T\boldsymbol{Z}+\epsilon.$$

Chen et al. [2012] studied the proportional odds model in the presence of time-varying covariates,

$$\log\frac{S(t|\widetilde{\boldsymbol{Z}})}{1-S(t|\widetilde{\boldsymbol{Z}})}=\log\frac{S_0(t)}{1-S_0(t)}+\boldsymbol{\beta}^T\boldsymbol{Z}(t),$$

and obtained their estimators for the regression parameters from martingale-based estimating equations. Peng and Huang [2007] developed the varying-coefficient complementary log-log survival model in which regression coefficients vary with time and covariates are time-invariant,

$$\log[-\log\{S(t|\boldsymbol{Z})\}] = \log[-\log\{S_0(t)\}] + \boldsymbol{\beta}(t)^T\boldsymbol{Z},$$

and they proposed a martingale-based estimating equation approach for parameter estimation. The method proposed in this paper also relies on the martingale-based estimating equations.

Consider observed data $\{X_i = \min(T_i, C_i), \Delta_i = \mathrm{I}(T_i \leq C_i), \boldsymbol{Z}_i(\cdot) : i = 1, 2, \ldots, n\}$. Let $N(t) = \mathrm{I}(X \leq t, \Delta = 1)$, $Y(t) = \mathrm{I}(X \geq t)$, $\mathcal{F}_t = \sigma\{N(s), Y(s), \widetilde{\boldsymbol{Z}}(s) : 0 \leq s \leq t\}$ and

$$M(t) = N(t) - \int_0^t Y(s)\lambda(s)ds,$$

where $\lambda(\cdot)$ is the hazard function of T. As described in Section 2.3, $\{M(t) : t \geq 0\}$ is a martingale with respect to the filtration $\{\mathcal{F}_t : t \geq 0\}$ and

$$\mathrm{E}\{dM(t)|\mathcal{F}_t-\} = 0.$$

Then the martingale-based estimating equations are defined as

$$\begin{aligned} \sum_{i=1}^n \int_0^\tau dM_i(t) &= 0, \\ \sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i(t)\, dM_i(t) &= \mathbf{0}, \end{aligned}$$

where τ is the duration of the study. The resulting solution is a reasonable estimator for the unknown parameter. For example, under the additive risk model, the martingale-based estimating equations are given by

$$\begin{aligned} \sum_{i=1}^n \int_0^\tau \left[dN_i(t) - Y_i(t)\{d\Lambda_0(t) + \boldsymbol{\beta}^T\boldsymbol{Z}_i(t)dt\}\right] &= 0, \\ \sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i(t)\left[dN_i(t) - Y_i(t)\{d\Lambda_0(t) + \boldsymbol{\beta}^T\boldsymbol{Z}_i(t)dt\}\right] &= \mathbf{0}, \end{aligned}$$

[Lin and Ying, 1994] and it is straightforward to solve these equations to derive the closed-form estimators

$$\begin{aligned}\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}}) &= \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u)\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(u)du\}}{\sum_{i=1}^n Y_i(u)},\\ \widehat{\boldsymbol{\beta}} &= \frac{\sum_{i=1}^n \int_0^\tau \{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t)\}dN_i(t)}{\sum_{i=1}^n \int_0^\tau Y_i(t)\{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t)\}^{\otimes 2}dt},\end{aligned}$$

where $v^{\otimes 2}$ denote vv^T for any vector v and

$$\overline{\boldsymbol{Z}}(t) = \frac{\sum_{i=1}^n Y_i(t)\boldsymbol{Z}_i(t)}{\sum_{i=1}^n Y_i(t)}.$$

Though many martingale-based estimating equations may not be solved explicitly and numerical approaches are used to search for solutions. The martingale-based estimating equation procedures generally render easy computation and straightforward covariance estimation. The large-sample properties of the resulting estimators can be established via the counting process martingale theory.

2.5 Estimation Methods of Survival Function

It is common in biomedical studies to be interested in estimating survival curves. The survival estimates are useful in comparing survival experience between groups, classifying subjects into different risk groups, and predicting survival status for future subjects. In the one-sample case, a common approach for estimating the survival function is to use the Kaplan-Meier [1958] product limit estimator. The Kaplan-Meier estimator is non-parametric and easy to calculate. Let t_i be the ith ordered failure time such that $t_1 < t_2 < \ldots < t_{k-1} < t_k$, where k is the number of distinct observed failure times. Denote d_i the number of subjects failed at time t_i, r_i the number of subjects at risk (including failure) at time t_i, $i = 1, 2, \ldots, k$. For given time t, the Kaplan-Meier product limit estimator for survival function $S(t) = \Pr(T > t)$ is given by

$$\widehat{S}_{km}(t) = \prod_{i:t_i \le t} \frac{r_i - d_i}{r_i}.$$

The Greenwood [1926] formula provides a reasonable estimate for the standard error of the Kaplan-Meier estimator, that is,

$$\widehat{S}_{km}(t)\sqrt{\sum_{i:t_i\leq t}\frac{d_i}{r_i(r_i-d_i)}}.$$

Confidence intervals to accompany the Kaplan-Meier estimator can be constructed using a normal approximation of the Greenwood formula.

In many applications, the survival regression models are used to estimate the survival curves for subjects with specific covariates. Under the survival regression model, survival estimators have closed-form expressions involving parameters in the model. A consistent estimator for the survival function is available when estimators for the parameters in the survival model are consistent and the model fits the data well. Apparently, the variation in the survival estimator is due to the approximation of the model parameters by the corresponding estimates. The asymptotic results for the survival estimator then can be obtained through the large-sample properties of parameter estimators.

Under the Cox proportional hazards model, the estimator for the survival function at time t for subjects with time-independent covariates $\boldsymbol{Z}$ takes the explicit form

$$\widehat{S}_{cox}(t|\boldsymbol{Z}) = \exp\{-\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z})\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})\},$$

where $\widehat{\boldsymbol{\beta}}$ is the Cox maximum partial likelihood estimator and $\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})$ is the Breslow [1972] estimator for the cumulative baseline hazard function,

$$\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}}) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_i)}.$$

Using standard counting process arguments, $\widehat{S}_{cox}(t|\boldsymbol{Z})$ is shown to be consistent and asymptotically normal [Tsiatis, 1981]. The asymptotic variance of $\widehat{S}_{cox}(t|\boldsymbol{Z})$ takes a closed form which is related to the asymptotic distribution of the partial likelihood estimator and Breslow estimator. Let t_i be the ith ordered failure time such that $t_1 < t_2 < \ldots < t_{k-1} < t_k$, where k is the number of distinct observed failure times. The risk set R_i, $i = 1, 2, \ldots, k$ is defined as the subjects who survive beyond time t_i, i.e. the subjects with observed survival times greater than or equal to time t_i. An estimator for the asymptotic variance of $\widehat{S}_{cox}(t|\boldsymbol{Z})$ derived by Tsiatis [1981] is of the form,

$$\widehat{var}\{\widehat{S}_{cox}(t|\boldsymbol{Z})\} = \{\widehat{S}_{cox}(t|\boldsymbol{Z})\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z})\}^2\Big[\sum_{i=1}^{l}\frac{1}{\big\{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_j)\big\}^2} + \boldsymbol{a}^T\widehat{var}(\widehat{\boldsymbol{\beta}})\boldsymbol{a}\Big],$$

where

$$a_j = \sum_{i=1}^{l}\frac{\sum_{j\in R_i} Z_{jj}\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_j)}{\big\{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_j)\big\}^2} - Z_j\frac{1}{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_j)},\ j = 1, 2, \ldots, p,$$

Z_{jj} is the jth component of $\boldsymbol{Z}_j$, Z_j is the jth component of $\boldsymbol{Z}$, $\widehat{var}(\widehat{\boldsymbol{\beta}})$ is the inverse of the second derivative of the partial log-likelihood with respect to $\boldsymbol{\beta}$, and l is such that $t_l < t, t_{l+1} \geq t$.

Lin and Ying [1994] derived an estimator of the survival function for the additive risk model,

$$\widehat{S}_{ar}(t|\widetilde{\boldsymbol{Z}}) = \exp\Big\{-\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}}) - \int_0^t \widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)du\Big\},$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})$ are the parameter estimators derived from the martingale-based estimating equations. Lin and Ying [1994] also established the large-sample properties of the survival estimator by standard martingale analysis. Their estimate for the asymptotic variance of $\widehat{S}_{ar}(t|\widetilde{\boldsymbol{Z}})$ takes the explicit from,

$$\widehat{S}_{ar}(t|\widetilde{\boldsymbol{Z}})^2\Big[\int_0^t\frac{n\sum_{i=1}^n dN_i(u)}{\{\sum_{i=1}^n Y_i(u)\}^2} + G^T(t|\widetilde{\boldsymbol{Z}})A^{-1}BA^{-1}G(t|\widetilde{\boldsymbol{Z}}) + 2G^T(t|\widetilde{\boldsymbol{Z}})A^{-1}D(t)\Big],$$

where

$$
\begin{aligned}
A &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty} Y_i(t)\{\boldsymbol{Z}_i(t)-\overline{\boldsymbol{Z}}(t)\}^{\otimes 2}dt,\\
B &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty} \{\boldsymbol{Z}_i(t)-\overline{\boldsymbol{Z}}(t)\}^{\otimes 2}dN_i(t),\\
D(t) &= \int_0^t \frac{\sum_{i=1}^n\{\boldsymbol{Z}_i(u)-\overline{\boldsymbol{Z}}(u)\}dN_i(u)}{\sum_{i=1}^n Y_i(u)},\\
G(t|\widetilde{\boldsymbol{Z}}) &= \int_0^t \{\boldsymbol{Z}(u)-\overline{\boldsymbol{Z}}(u)\}du,\\
\overline{\boldsymbol{Z}}(t) &= \frac{\sum_{i=1}^n Y_i(t)\boldsymbol{Z}_i(t)}{\sum_{i=1}^n Y_i(t)}.
\end{aligned}
$$

For the linear transformation model, inference procedures for the survival function based on estimating equations have been developed by Cheng et al. [1997]. They also derived a closed-form estimator for the asymptotic variance of survival estimate which is a complex expression.

Ying et al. [1992] proposed a consistent estimator for the survival function $S(t|\boldsymbol{Z})$ under the accelerated failure time model. However, the asymptotic variance of the survival estimator can not be consistently estimated. They instead proposed interval estimation procedures to form the corresponding confidence intervals of survival function without using the normal approximation for the variances of survival estimators. Let $\widehat{S}_{\boldsymbol{\beta}}(\cdot)$ be the Kaplan-Meier estimate for the survival function and ϕ be a continuously differentiable function on (0,1). Define $X = \min(\log T, C)$, $\Delta = \mathrm{I}(\log T \leq C)$. To test the hypothesis $H_0 : S(t|\boldsymbol{Z}) = \tau_0$, for some given τ_0, they consider the minimum dispersion test statistic, that is

$$
Q(\tau_0) = \min_{\boldsymbol{\beta}}\Big[n\Big\{\frac{\widehat{S}_{\boldsymbol{\beta}}(t_{\boldsymbol{\beta}})-\tau_0}{\hat{\sigma}_1(\widehat{\boldsymbol{\beta}})}\Big\}^2 + n\xi_{1n}^T(\boldsymbol{\beta})\{\hat{V}_1(\widehat{\boldsymbol{\beta}})\}^{-1}\xi_{1n}(\boldsymbol{\beta})\Big],
$$

where

$$
\hat{\sigma}_1^2(\widehat{\boldsymbol{\beta}}) = \widehat{S}^2_{\widehat{\boldsymbol{\beta}}}(t_{\widehat{\boldsymbol{\beta}}})\sum_{i=1}^{n}\frac{n\Delta_i \mathrm{I}\{e_i(\widehat{\boldsymbol{\beta}})\leq t_{\widehat{\boldsymbol{\beta}}}\}}{\big[\sum_j \mathrm{I}\{e_j(\widehat{\boldsymbol{\beta}})\geq e_i(\widehat{\boldsymbol{\beta}})\}\big]^2},
$$

$$
\begin{aligned}
\xi_{1n}(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n}\Delta_i\phi[1-\widehat{S}_{\boldsymbol{\beta}}\{e_i(\boldsymbol{\beta})\}]\Big[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n}\boldsymbol{Z}_j\mathrm{I}\{e_j(\boldsymbol{\beta})\geq e_i(\boldsymbol{\beta})\}}{\sum_{j=1}^{n}\mathrm{I}\{e_j(\boldsymbol{\beta})\geq e_i(\boldsymbol{\beta})\}}\Big],\\
\hat{V}_1(\widehat{\boldsymbol{\beta}}) &= \frac{1}{n}\sum_{i=1}^{n}\phi^2[1-\widehat{S}_{\widehat{\boldsymbol{\beta}}}\{e_i(\widehat{\boldsymbol{\beta}})\}]\Delta_i\\
&\quad\times\Big(\frac{\sum_j \mathrm{I}\{e_j(\widehat{\boldsymbol{\beta}})\geq e_i(\widehat{\boldsymbol{\beta}})\}\boldsymbol{Z}_j^{\otimes 2}}{\sum_j \mathrm{I}\{e_j(\widehat{\boldsymbol{\beta}})\geq e_i(\widehat{\boldsymbol{\beta}})\}} - \Big[\frac{\sum_j \mathrm{I}\{e_j(\widehat{\boldsymbol{\beta}})\geq e_i(\widehat{\boldsymbol{\beta}})\}\boldsymbol{Z}_j}{\sum_j \mathrm{I}\{e_j(\widehat{\boldsymbol{\beta}})\geq e_i(\widehat{\boldsymbol{\beta}})\}}\Big]^{\otimes 2}\Big),\\
t_{\boldsymbol{\beta}} &= \log t - \boldsymbol{\beta}^T\boldsymbol{Z}, \qquad e_i(\boldsymbol{\beta}) = X_i - \boldsymbol{\beta}^T\boldsymbol{Z}_i.
\end{aligned}
$$

Based on $Q(\tau_0)$, one may reject H_0 if $Q(\tau_0) > \chi_1^2(\alpha)$, where $\chi_1^2(\alpha)$ is the upper α quantile of the chi-squared distribution with one degree of freedom. Therefore, a valid $100(1-\alpha)\%$ confidence interval of the survival function is given by $\{\tau : Q(\tau) \leq \chi_1^2(\alpha)\}$.

Chapter 3
METHODS

3.1 Proposed Model

We propose a complementary log-log survival model that incorporates time-varying covariates. Let T be the failure time, $\boldsymbol{Z}(t)$ denote the value of the associated p-dimensional vector of possibly time-varying covariates measured at time t, and $\widetilde{\boldsymbol{Z}}(t) = \{\boldsymbol{Z}(s) : 0 \leq s \leq t\}$ be the covariate history collected up until t. The proposed model assumes that

$$\log[-\log\{S(t|\widetilde{\boldsymbol{Z}})\}] = \log[-\log\{S_0(t)\}] + \boldsymbol{\beta}^T \boldsymbol{Z}(t), \tag{3.1}$$

where $S(\cdot|\widetilde{\boldsymbol{Z}})$ is the survival function of T given the covariate history $\widetilde{\boldsymbol{Z}}(\cdot)$, $S_0(\cdot)$ is an unspecified baseline survival function, $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression parameters and the superscript T denotes matrix transpose.

Model (3.1) implicitly requires that $\log[-\log\{S_0(t)\}] + \boldsymbol{\beta}^T \boldsymbol{Z}(t)$ is increasing in t. In view of the fact that the survival function is a decreasing function, $\log[-\log\{S(t|\widetilde{\boldsymbol{Z}})\}]$ is increasing in t. Thus, $\log[-\log\{S_0(t)\}] + \boldsymbol{\beta}^T \boldsymbol{Z}(t)$ is restricted to be increasing in t. There is no constraints for the range of $\boldsymbol{\beta}$ and $\boldsymbol{Z}(t)$. Since

$$S_0(t) \in (0,1) \Longrightarrow \log\{S_0(t)\} \in (-\infty, 0) \Longrightarrow \log[-\log\{S_0(t)\}] \in (-\infty, +\infty),$$

and similarly $\log[-\log\{S(t|\widetilde{\boldsymbol{Z}})\}] \in (-\infty, +\infty)$.

The time-varying covariates in model (3.1) are confined to be external time-varying covariates in order for the interpretation of model (3.1) to be meaningful. The external time-varying covariates are not dependent on an individual's survival process [Kalbfleisch and Prentice, 2002], which satisfy that for any $s > t$,

$$\Pr\{Z(s)|Z(t), T \geq t\} = \Pr\{Z(s)|Z(t), T = t\}.$$

The internal time-varying covariate measures from the individual directly and can only be obtained when the individual is alive. The existence of internal time-varying covariates reveals the survival status of the individual.

Model (3.1) reduces to the usual Cox proportional hazards model when covariates are all time-invariant. Let $\lambda(\cdot|\boldsymbol{Z})$ be the hazard function of T given the time-independent covariates $\boldsymbol{Z}$ and $\lambda_0(\cdot)$ be the unspecified baseline hazard function. It is seen that

$$\begin{aligned}
&\log[-\log\{S(t|\boldsymbol{Z})\}] = \log[-\log\{S_0(t)\}] + \boldsymbol{\beta}^T\boldsymbol{Z} \\
\Rightarrow\ &\log[\int_0^t \lambda(u|\boldsymbol{Z})du] = \log[\int_0^t \lambda_0(u)du] + \boldsymbol{\beta}^T\boldsymbol{Z} \\
\Rightarrow\ &\int_0^t \lambda(u|\boldsymbol{Z})du = \int_0^t \lambda_0(u)du \times \exp(\boldsymbol{\beta}^T\boldsymbol{Z}) \\
\Rightarrow\ &\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}).
\end{aligned}$$

In the presence of time-varying covariates, model (3.1) provides direct interpretations of covariate effects on the survival function while the Cox model does not: $\boldsymbol{\beta}$ estimates the difference in $\log(-\log)$ survival probability associated with per unit change in $\boldsymbol{Z}(\cdot)$.

3.2 Estimation of Regression Parameters

3.2.1 Data

We consider estimation of model (3.1) with right censored data. As usual, we assume that the censoring time, say C, is independent of T conditional on $\widetilde{\boldsymbol{Z}}$. Let $X = \min(T, C)$ and $\Delta = \mathrm{I}(T \le C)$, where $\mathrm{I}(\cdot)$ is the indicator function. Suppose that there are n subjects in the study. For the ith subject, $i = 1, 2, \ldots, n$, the bivariate vector (X_i, Δ_i) is observed instead of the failure time T_i; also, the covariate history $\widetilde{\boldsymbol{Z}}_i(X_i) = \{\boldsymbol{Z}_i(s) : 0 \le s \le X_i\}$ is recorded. The observed data consist of n iid copies of $\{X, \Delta, \widetilde{\boldsymbol{Z}}(X)\}$, that is, $\{X_i, \Delta_i, \widetilde{\boldsymbol{Z}}_i(X_i) : i = 1, 2, \ldots, n\}$.

3.2.2 Estimating Equation Method

Let $\alpha(t) = \log[-\log\{S_0(t)\}]$ and $\Lambda(t|\widetilde{\boldsymbol{Z}})$ is the cumulative hazard function of T given the covariate history $\widetilde{\boldsymbol{Z}}(t)$. Under the complementary log-log survival model, we can write

$$\begin{aligned} \Lambda(t|\widetilde{\boldsymbol{Z}}) &= -\log\{S(t|\widetilde{\boldsymbol{Z}})\} \\ &= \exp\{\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}(t)\}, \\ d\Lambda(t|\widetilde{\boldsymbol{Z}}) &= d\exp\{\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}(t)\}. \end{aligned}$$

Let $\{\alpha_0(t), \boldsymbol{\beta}_0\}$ denote the true value of $\{\alpha(t), \boldsymbol{\beta}\}$. Define $N(t) = \mathrm{I}(X \leq t, \Delta = 1)$, $Y(t) = \mathrm{I}(X \geq t)$, $\mathcal{F}_t = \sigma\{N(s), Y(s), \widetilde{\boldsymbol{Z}}(s) : 0 \leq s \leq t\}$ and

$$M(t) = N(t) - \int_0^t Y(s)\, d\exp\{\alpha_0(s) + \boldsymbol{\beta}_0^T \boldsymbol{Z}(s)\}.$$

As described in Section 2.3 and Fleming and Harrington [1991], $\{M(t) : t \geq 0\}$ is a martingale with respect to the filtration $\{\mathcal{F}_t : t \geq 0\}$. In view of the fact that the increments of a martingale have expectation zero given the history path, $\mathrm{E}\{dM(t)|\mathcal{F}_t-\} = 0$, we can define an estimator for $\{\alpha_0(t), \boldsymbol{\beta}_0\}$ by solving

$$\sum_{i=1}^n \int_0^\tau [dN_i(t) - Y_i(t)\, d\exp\{\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}] = 0, \tag{3.2}$$

$$\sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i(t)[dN_i(t) - Y_i(t)\, d\exp\{\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}] = \boldsymbol{0}, \tag{3.3}$$

where $\{N_i(t), Y_i(t)\}$, $i = 1, 2, \ldots, n$ are sample analogues of $\{N(t), Y(t)\}$ and τ is the duration of the study.

Equation (3.2) implies, for all $t \in [0, \tau]$,

$$\sum_{i=1}^n dN_i(t) - P_n(t; \boldsymbol{\beta})\, d\exp\{\alpha(t)\} - Q_n(t; \boldsymbol{\beta}) \exp\{\alpha(t)\} = 0, \tag{3.4}$$

where

$$\begin{aligned} P_n(t; \boldsymbol{\beta}) &= \sum_{i=1}^n Y_i(t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}, \\ Q_n(t; \boldsymbol{\beta}) &= \sum_{i=1}^n Y_i(t)\, d\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}. \end{aligned}$$

(3.4) satisfies a first-order ordinary differential equation and has a unique closed-form solution. It is seen that

$$\begin{aligned}
(3.4) &\Rightarrow d\exp\{\alpha(t)\} + \frac{Q_n(t;\boldsymbol{\beta})}{P_n(t;\boldsymbol{\beta})}\exp\{\alpha(t)\} = \frac{\sum_{i=1}^n dN_i(t)}{P_n(t;\boldsymbol{\beta})} \\
&\Rightarrow e^{\int_0^t \frac{Q_n(u;\boldsymbol{\beta})}{P_n(u;\boldsymbol{\beta})}}\Big[d\exp\{\alpha(t)\} + \frac{Q_n(t;\boldsymbol{\beta})}{P_n(t;\boldsymbol{\beta})}\exp\{\alpha(t)\}\Big] = e^{\int_0^t \frac{Q_n(u;\boldsymbol{\beta})}{P_n(u;\boldsymbol{\beta})}}\frac{\sum_{i=1}^n dN_i(t)}{P_n(t;\boldsymbol{\beta})} \\
&\Rightarrow d\Big[e^{\int_0^t \frac{Q_n(u;\boldsymbol{\beta})}{P_n(u;\boldsymbol{\beta})}}\exp\{\alpha(t)\}\Big] = e^{\int_0^t \frac{Q_n(u;\boldsymbol{\beta})}{P_n(u;\boldsymbol{\beta})}}\frac{\sum_{i=1}^n dN_i(t)}{P_n(t;\boldsymbol{\beta})} \\
&\Rightarrow e^{\int_0^t \frac{Q_n(u;\boldsymbol{\beta})}{P_n(u;\boldsymbol{\beta})}}\exp\{\alpha(t)\} - e^{\int_0^0 \frac{Q_n(u;\boldsymbol{\beta})}{P_n(u;\boldsymbol{\beta})}}\exp\{\alpha(0)\} = \int_0^t \Big\{e^{\int_0^u \frac{Q_n(s;\boldsymbol{\beta})}{P_n(s;\boldsymbol{\beta})}}\frac{\sum_{i=1}^n dN_i(u)}{P_n(u;\boldsymbol{\beta})}\Big\},
\end{aligned}$$

and as $t \to 0$,

$$\exp\{\alpha(t)\} \to \exp\{\alpha(0)\} = -\log\{S_0(0)\} = 0 \Rightarrow e^{\int_0^0 \frac{Q_n(u;\boldsymbol{\beta})}{P_n(u;\boldsymbol{\beta})}}\exp\{\alpha(0)\} = 1 \times 0 = 0.$$

Therefore, the estimator for $\alpha_0(t)$ is obtained as the solution to (3.4), that is,

$$\widehat{\alpha}_n(t;\boldsymbol{\beta}) = \log\Big[\frac{1}{D_n(t;\boldsymbol{\beta})}\int_0^t \Big\{D_n(u;\boldsymbol{\beta})\frac{\sum_{i=1}^n dN_i(u)}{P_n(u;\boldsymbol{\beta})}\Big\}\Big],$$

where

$$D_n(t;\boldsymbol{\beta}) = \exp\Big\{\int_0^t \frac{Q_n(u;\boldsymbol{\beta})}{P_n(u;\boldsymbol{\beta})}\Big\}.$$

Further, replacing $\alpha(t)$ by $\widehat{\alpha}_n(t;\boldsymbol{\beta})$ in (3.3) yields the estimating equations for $\boldsymbol{\beta}$,

$$\sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i(t)[dN_i(t) - Y_i(t)\, d\exp\{\widehat{\alpha}_n(t;\boldsymbol{\beta}) + \boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}] = \mathbf{0}.$$

The equations are then approximately solved by the use of the Newton-Raphson method [Atkinson, 1989] to derive the estimator of $\boldsymbol{\beta}_0$. Let $\widehat{\boldsymbol{\beta}}_n$ denote the estimator of $\boldsymbol{\beta}_0$. Specifically, we apply the following computation algorithms to get $\widehat{\boldsymbol{\beta}}_n$:

- Start with an initial value $\boldsymbol{\beta}^{(0)}$
- At iteration k, update the current estimate $\boldsymbol{\beta}^{(k-1)}$ to the new estimate $\boldsymbol{\beta}^{(k)}$ by

$$\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^{(k-1)} - \{\nabla\mathbf{f}_n(\boldsymbol{\beta}^{(k-1)})\}^{-1}\mathbf{f}_n(\boldsymbol{\beta}^{(k-1)}),$$

where

$$\mathbf{f}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i(t)[dN_i(t) - Y_i(t)\, d\exp\{\widehat{\alpha}_n(t;\boldsymbol{\beta}) + \boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}]$$

$$\nabla \mathbf{f}_n(\boldsymbol{\beta}) = \partial \mathbf{f}_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$$

- Repeat updating until convergence, e.g. $||\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}^{(k-1)}|| < 0.0001$

In general, the iterative algorithms work well with arbitrary initial values. Nevertheless, one may obtain good initial values by fitting the Cox proportional hazards model.

With all covariates being time-invariant, we have

$$\begin{aligned}
D_n(t;\boldsymbol{\beta}) &= \exp\Big\{ \int_0^t \frac{Q_n(u;\boldsymbol{\beta})}{P_n(u;\boldsymbol{\beta})} \Big\} \\
&= \exp\Big\{ \int_0^t \frac{\sum_{i=1}^n Y_i(u)\, d\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i\}}{\sum_{i=1}^n Y_i(u) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i\}} \Big\} \\
&= \exp\Big\{ \int_0^t \frac{0}{\sum_{i=1}^n Y_i(u) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i\}} \Big\} \\
&= 1;
\end{aligned}$$

then $\exp\{\widehat{\alpha}_n(t;\boldsymbol{\beta})\}$ reduces to

$$\begin{aligned}
\exp\{\widehat{\alpha}_n(t;\boldsymbol{\beta})\} &= \frac{1}{D_n(t;\boldsymbol{\beta})} \int_0^t \Big\{ D_n(u;\boldsymbol{\beta}) \frac{\sum_{i=1}^n dN_i(u)}{P_n(u;\boldsymbol{\beta})} \Big\} \\
&= \frac{1}{1} \int_0^t \Big\{ 1 \times \frac{\sum_{i=1}^n dN_i(u)}{P_n(u;\boldsymbol{\beta})} \Big\} \\
&= \int_0^t \Big\{ \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i\}} \Big\},
\end{aligned}$$

which is precisely the Breslow estimator for the cumulative baseline hazard function. Meanwhile, the estimating function for $\boldsymbol{\beta}$ is identical to the Cox partial likelihood score function,

$$\begin{aligned}
&\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i[dN_i(t) - Y_i(t)\, d\exp\{\widehat{\alpha}_n(t;\boldsymbol{\beta}) + \boldsymbol{\beta}^T \boldsymbol{Z}_i\}] \\
= &\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i[dN_i(t) - Y_i(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)\, d\exp\{\widehat{\alpha}_n(t;\boldsymbol{\beta})\}].
\end{aligned}$$

Thus, in the case of time-independent covariates, our estimator for $\boldsymbol{\beta}$ is in the same form as the Cox maximum partial likelihood estimator.

3.2.3 Large Sample Properties

Recall that $\{\alpha_0(t), \boldsymbol{\beta}_0\}$ denotes the true value of $\{\alpha(t), \boldsymbol{\beta}\}$ and $\{\widehat{\alpha}_n(t; \widehat{\boldsymbol{\beta}}_n), \widehat{\boldsymbol{\beta}}_n\}$ denotes the estimator of $\{\alpha_0(t), \boldsymbol{\beta}_0\}$. Let $\widehat{R}_n(t; \boldsymbol{\beta}) = \exp\{\widehat{\alpha}_n(t; \boldsymbol{\beta})\}$, $R_0(t) = \exp\{\alpha_0(t)\}$, $B_i(t; \boldsymbol{\beta}) = \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}$ and $v^{\otimes 2}$ denote vv^T for any vector v. Define

$$\overline{\boldsymbol{Z}}(t; \boldsymbol{\beta}) = \frac{\sum_{i=1}^n Y_i(t) B_i(t; \boldsymbol{\beta}) \boldsymbol{Z}_i(t)}{\sum_{i=1}^n Y_i(t) B_i(t; \boldsymbol{\beta})},$$
$$\boldsymbol{u}(t; \boldsymbol{\beta}) = \lim_{n \to \infty} \overline{\boldsymbol{Z}}(t; \boldsymbol{\beta}).$$

We assume the following regularity conditions.

- **C1**: The true regression parameter $\boldsymbol{\beta}_0$ lies in the interior of a compact set Θ.

- **C2**: The baseline density function f_0 of T and its derivative f_0' are bounded, satisfying, for some $\epsilon > 0$,

$$\int_0^\infty \left\{\frac{f_0'(t)}{f_0(t)}\right\}^2 f_0(t) dt < \infty, \text{and} \int_0^\infty t^\epsilon f_0(t) dt < \infty.$$

- **C3**: The density function g of C is uniformly bounded, that is, $\sup_t g(t) < \infty$.

- **C4**: There exists finite $\tau > 0$ such that $\Pr(T > \tau) > 0$, $\Pr(C \geq \tau) = \Pr(C = \tau) > 0$.

- **C5**: There exists $\kappa_1 > 0$ and $\kappa_2 > 0$ such that

$$\sup_{|s-t| \leq n^{-\kappa_2}} n^{-1} \sum_i ||\boldsymbol{Z}_i(s) - \boldsymbol{Z}_i(t)|| = O(n^{-1/2-\kappa_1}),$$

  and, for any positive sequence $d_n \to 0$, there exists $\kappa_3 > 0$ such that

$$\sup_{|s-t| \leq d_n} n^{-1} \sum_i ||\boldsymbol{Z}_i(s) - \boldsymbol{Z}_i(t)|| = o\{\max(d_n^{\kappa_3}, n^{-\kappa_3})\}.$$

- **C6**: There exists $\epsilon > 0$ such that $\Pr\{||\boldsymbol{Z}_i(t) - \boldsymbol{u}(t; \boldsymbol{\beta}_0)|| > \epsilon, i = 1, 2, \ldots, n\} > 0$, which essentially means that covariate processes cannot be identical for all the subjects.

- **C7**:

$$\| \int_0^\tau \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{u}(t;\boldsymbol{\beta}_0) \times \mathrm{E}[dN(t) - Y(t)R(t;\boldsymbol{\beta}_0)\, d\exp\{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)\}]$$
$$+ \int_0^\tau \mathrm{E}\Big(Y(t)\{\boldsymbol{Z}(t) - \boldsymbol{u}(t;\boldsymbol{\beta}_0)\}\Big[d\exp\{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)\}\frac{\partial}{\partial \boldsymbol{\beta}} R(t;\boldsymbol{\beta}_0)$$
$$+ R(t;\boldsymbol{\beta}_0)\frac{\partial}{\partial \boldsymbol{\beta}} d\exp\{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)\}\Big]^T\Big) \| > 0.$$

Regularity conditions C1-C5 have been commonly used in the literature [Chen et al., 2002, Peng and Huang, 2007, Chen et al., 2012]. Regularity conditions C1-C3 ensure the central limit theorem for martingales. By regularity condition C4, we assume that some subjects are at risk at the end of the study, a finite time point τ, and subjects alive at time τ are considered censored. Regularity condition C5 is the smoothness condition for time-varying covariates. To show the consistency of $\widehat{\beta}_n$, we assume two additional regularity conditions C6 and C7. These regularity conditions are generally satisfactory in practice.

In the following theorem, we show the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$.

Theorem 3.2.1. *Under regularity conditions C1-C7, as $n \to \infty$, $\widehat{\boldsymbol{\beta}}_n$ is strongly consistent, and $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ converges weakly to a p-variate normal with mean zero and a covariance matrix $\boldsymbol{U}^{-1}\boldsymbol{V}(\boldsymbol{U}^{-1})^T$, where*

$$\begin{aligned}
\boldsymbol{U} &= \lim_{n\to\infty} -\frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big[\frac{\partial \overline{\boldsymbol{Z}}(t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\{dN_i(t) - Y_i(t) \times \widehat{R}_n(t;\boldsymbol{\beta}_0) dB_i(t;\boldsymbol{\beta}_0)\} \\
&\quad + Y_i(t)\{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t;\boldsymbol{\beta}_0)\}\Big\{\frac{\partial \hat{R}_n(t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} dB_i(t;\boldsymbol{\beta}_0) + \widehat{R}_n(t;\boldsymbol{\beta}_0)\frac{\partial dB_i(t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big\}^T\Big], \\
\boldsymbol{V} &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \int_0^\tau \boldsymbol{\xi}_i(t;\boldsymbol{\beta}_0)^{\otimes 2} Y_i(t)\{dR_0(t) \times B_i(t;\boldsymbol{\beta}_0) + R_0(t)\, dB_i(t;\boldsymbol{\beta}_0)\}, \\
\boldsymbol{\xi}_i(t;\boldsymbol{\beta}) &= \boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t;\boldsymbol{\beta}) - \frac{D_n(t;\boldsymbol{\beta})}{P_n(t;\boldsymbol{\beta})}\sum_{k=1}^n \int_t^\tau \Big[\{\boldsymbol{Z}_k(u) - \overline{\boldsymbol{Z}}(u;\boldsymbol{\beta})\}\frac{Y_k(u)\, dB_k(u;\boldsymbol{\beta})}{D_n(u;\boldsymbol{\beta})}\Big].
\end{aligned}$$

Moreover, consistent estimators for $\boldsymbol{U}$ *and* $\boldsymbol{V}$ *are given respectively by*

$$\begin{aligned}
\widehat{\boldsymbol{U}}_n &= -\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Big[\frac{\partial \overline{\boldsymbol{Z}}(t;\widehat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}}\{dN_i(t) - Y_i(t)\times \widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)\, dB_i(t;\widehat{\boldsymbol{\beta}}_n)\} \\
&\quad + Y_i(t)\{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t;\widehat{\boldsymbol{\beta}}_n)\}\Big\{\frac{\partial \widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}} dB_i(t;\widehat{\boldsymbol{\beta}}_n) + \widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)\frac{\partial dB_i(t;\widehat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}}\Big\}^T\Big], \\
\widehat{\boldsymbol{V}}_n &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \boldsymbol{\xi}_i(t;\widehat{\boldsymbol{\beta}}_n)^{\otimes 2} Y_i(t)\{d\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)\times B_i(t;\widehat{\boldsymbol{\beta}}_n) + \widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)\, dB_i(t;\widehat{\boldsymbol{\beta}}_n)\}.
\end{aligned}$$

Proof of Theorem 3.2.1 follows standard counting process theory. For ease of presentation, we state the proof of Theorem 3.2.1 in a univariate setting, which can be generalized easily to the multivariate case.

Proof of Theorem 3.2.1. The proof consists of four steps.

Step 1. Define

$$S_n(\beta_0) = \sum_{i=1}^{n}\int_0^\tau Z_i(t)[dN_i(t) - Y_i(t)\{d\widehat{R}_n(t;\beta_0)B_i(t;\beta_0) + \widehat{R}_n(t;\beta_0)dB_i(t;\beta_0)\}].$$

Here we establish a martingale representation of $S_n(\beta_0)$.

By definition of martingale under the complementary log-log survival model, we have

$$\begin{aligned}
& M_i(t) = N_i(t) - \int_0^t Y_i(s)\, d\exp\{\alpha_0(s) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_i(s)\} \\
\Rightarrow & \sum_{i=1}^{n}\{dN_i(t) - Y_i(t)\, dR_0(t)\, B_i(t;\beta_0) - Y_i(t)\, R_0(t)\, dB_i(t;\beta_0)\} = \sum_{i=1}^{n} dM_i(t). \qquad (3.5)
\end{aligned}$$

By equation (3.4), we have

$$\begin{aligned}
& \sum_{i=1}^{n} dN_i(t) - P_n(t;\beta_0)\, d\exp\{\widehat{\alpha}_n(t;\beta_0\} - Q_n(t;\beta_0)\exp\{\widehat{\alpha}_n(t;\beta_0\} = 0 \\
\Rightarrow & \sum_{i=1}^{n}\{dN_i(t) - Y_i(t)\, d\widehat{R}_n(t;\beta_0)\, B_i(t;\beta_0) - Y_i(t)\, \widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} = 0, \qquad (3.6)
\end{aligned}$$

Then, with some algebra, we can write

$$
\begin{aligned}
S_n(\beta_0) &= \sum_{i=1}^{n} \int_0^{\tau} Z_i(t)[dN_i(t) - Y_i(t)\{d\widehat{R}_n(t;\beta_0)B_i(t;\beta_0) + \widehat{R}_n(t;\beta_0)dB_i(t;\beta_0)\}] \\
&= \sum_{i=1}^{n} \int_0^{\tau} Z_i(t)\{dN_i(t) - Y_i(t)\, d\widehat{R}_n(t;\beta_0)\, B_i(t;\beta_0) - Y_i(t)\, \widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} \\
&\quad - \int_0^{\tau} \bar{Z}(t;\beta_0) \sum_{i=1}^{n}\{dN_i(t) - Y_i(t)\, d\widehat{R}_n(t;\beta_0)\, B_i(t;\beta_0) - Y_i(t)\, \widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} \\
&= \sum_{i=1}^{n} \int_0^{\tau} Z_i(t)\{dN_i(t) - Y_i(t)\, d\widehat{R}_n(t;\beta_0)\, B_i(t;\beta_0) - Y_i(t)\, \widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} \\
&\quad - \sum_{i=1}^{n} \int_0^{\tau} \bar{Z}(t;\beta_0)\{dN_i(t) - Y_i(t)\, d\widehat{R}_n(t;\beta_0)\, B_i(t;\beta_0) - Y_i(t)\, \widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} \\
&= \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t;\beta_0)\}\{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} \\
&\quad - \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t;\beta_0)\}Y_i(t)\, d\widehat{R}_n(t;\beta_0)\, B_i(t;\beta_0) \\
&= \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t;\beta_0)\}\{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} \\
&\quad - \int_0^{\tau} d\widehat{R}_n(t;\beta_0) \sum_{i=1}^{n}\{Z_i(t)\, Y_i(t)\, B_i(t;\beta_0)\} + \int_0^{\tau} d\widehat{R}_n(t;\beta_0) \\
&\quad \times \frac{\sum_{i=1}^{n} Z_i(t)Y_i(t)B_i(t;\beta_0)}{\sum_{i=1}^{n} Y_i(t)B_i(t;\beta_0)} \sum_{i=1}^{n}\{Y_i(t)B_i(t;\beta_0)\} \\
&= \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t;\beta_0)\}\{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} \\
&= \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t;\beta_0)\}\{dM_i(t) + Y_i(t)\, dR_0(t)\, B_i(t;\beta_0) + Y_i(t)\, R_0(t)\, dB_i(t;\beta_0) \\
&\quad - Y_i(t)\widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} \\
&= \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t;\beta_0)\}dM_i(t) + \sum_{i=1}^{n} \int_0^{\tau} \left[\{Z_i(t) - \bar{Z}(t;\beta_0)\}Y_i(t)\, dB_i(t;\beta_0)\right. \\
&\quad \left. \times \{R_0(t) - \widehat{R}_n(t;\beta_0)\}\right] + \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t;\beta_0)\}Y_i(t)\, dR_0(t)\, B_i(t;\beta_0)
\end{aligned}
$$

$$= \sum_{i=1}^{n} \int_0^\tau \{Z_i(t) - \bar{Z}(t;\beta_0)\} dM_i(t) + \sum_{i=1}^{n} \int_0^\tau \big[\{Z_i(t) - \bar{Z}(t;\beta_0)\} Y_i(t)\, dB_i(t;\beta_0)$$
$$\times \{R_0(t) - \widehat{R}_n(t;\beta_0)\}\big] + \int_0^\tau dR_0(t) \sum_{i=1}^{n} \{Z_i(t)\, Y_i(t)\, B_i(t;\beta_0)\} - \int_0^\tau dR_0(t)$$
$$\times \frac{\sum_{i=1}^n Z_i(t) Y_i(t) B_i(t;\beta_0)}{\sum_{i=1}^n Y_i(t) B_i(t;\beta_0)} \sum_{i=1}^{n} \{Y_i(t) B_i(t;\beta_0)\}$$
$$= \sum_{i=1}^{n} \int_0^\tau \{Z_i(t) - \bar{Z}(t;\beta_0)\} dM_i(t) + \sum_{i=1}^{n} \int_0^\tau \big[\{Z_i(t) - \bar{Z}(t;\beta_0)\} Y_i(t)\, dB_i(t;\beta_0)$$
$$\times \{R_0(t) - \widehat{R}_n(t;\beta_0)\}\big].$$

We can construct a martingale representation of $\widehat{R}_n(t;\beta_0) - R_0(t)$. By subtracting (3.5) from (3.6), we get

$$\sum_{i=1}^{n} dM_i(t) - d\{\widehat{R}_n(t;\beta_0) - R_0(t)\}\, P_n(t;\beta_0) - \{\widehat{R}_n(t;\beta_0) - R_0(t)\}\, Q_n(t;\beta_0) = 0. \qquad (3.7)$$

This is similar to the equation from which we obtain $\hat{R}_n(t;\beta)$. It is seen that

$$d\{\widehat{R}_n(t;\beta_0) - R_0(t)\} + \frac{Q_n(t;\beta)}{P_n(t;\beta)} \{\widehat{R}_n(t;\beta_0) - R_0(t)\} = \frac{\sum_{i=1}^n dM_i(t)}{P_n(t;\beta)}$$
$$\Rightarrow e^{\int_0^t \frac{Q_n(u;\beta)}{P_n(u;\beta)}} \Big[d\{\widehat{R}_n(t;\beta_0) - R_0(t)\} + \frac{Q_n(t;\beta)}{P_n(t;\beta)} \{\widehat{R}_n(t;\beta_0) - R_0(t)\}\Big] = e^{\int_0^t \frac{Q_n(u;\beta)}{P_n(u;\beta)}} \frac{\sum_{i=1}^n dM_i(t)}{P_n(t;\beta)}$$
$$\Rightarrow d\Big[e^{\int_0^t \frac{Q_n(u;\beta)}{P_n(u;\beta)}} \{\widehat{R}_n(t;\beta_0) - R_0(t)\}\Big] = e^{\int_0^t \frac{Q_n(u;\beta)}{P_n(u;\beta)}} \frac{\sum_{i=1}^n dM_i(t)}{P_n(t;\beta)}$$
$$\Rightarrow e^{\int_0^t \frac{Q_n(u;\beta)}{P_n(u;\beta)}} \{\widehat{R}_n(t;\beta_0) - R_0(t)\} - 1 \times \{\hat{R}_n(0;\beta_0) - R_0(0)\} = \int_0^t \Big\{e^{\int_0^u \frac{Q_n(s;\beta)}{P_n(s;\beta)}} \frac{\sum_{i=1}^n dM_i(u)}{P_n(u;\beta)}\Big\},$$

and as $t \to 0$,

$$\widehat{R}_n(t;\beta_0) - R_0(t) \to \widehat{R}_n(0;\beta_0) - R_0(0) = \int_0^0 \Big\{\frac{D_n(u;\boldsymbol{\beta}) \sum_{i=1}^n dN_i(u)}{D_n(0;\boldsymbol{\beta})\, P_n(u;\boldsymbol{\beta})}\Big\} + \log\{S_0(0)\} = 1 \times 0 + 0$$

Therefore,

$$\widehat{R}_n(t;\beta_0) - R_0(t) = \frac{1}{D_n(t;\beta_0)} \int_0^t \Big\{D_n(u;\beta_0) \frac{\sum_{i=1}^n dM_i(u)}{P_n(u;\beta_0)}\Big\}$$
$$= \sum_{i=1}^{n} \int_0^t \Big\{\frac{D_n(u;\beta_0)}{D_n(t;\beta_0)\, P_n(u;\beta_0)} dM_i(u)\Big\}.$$

Replacing $\hat{R}_n(t;\beta_0) - R_0(t)$ by its martingale representation, we then have

$$S_n(\beta_0) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t;\beta_0)\} dM_i(t) - \sum_{i=1}^n \int_0^\tau \Big[\{Z_i(t) - \bar{Z}(t;\beta_0)\} Y_i(t)\, dB_i(t;\beta_0) \times \sum_{k=1}^n \int_0^t \Big\{ \frac{D_n(u;\beta_0)}{D_n(t;\beta_0)\, P_n(u;\beta_0)} dM_k(u) \Big\}\Big].$$

As a result of integration by parts,

$$\begin{aligned}
&\sum_{i=1}^n \int_0^\tau \Big[\{Z_i(t) - \bar{Z}(t;\beta_0)\} Y_i(t)\, dB_i(t;\beta_0) \sum_{k=1}^n \int_0^t \Big\{ \frac{D_n(u;\beta_0)}{D_n(t;\beta_0)\, P_n(u;\beta_0)} dM_k(u) \Big\}\Big] \\
&= \int_0^\tau \Big(\Big\{ \sum_{k=1}^n \int_0^t \frac{D_n(u;\beta_0)}{P_n(u;\beta_0)} dM_k(u) \Big\} \times \Big[\sum_{i=1}^n \{Z_i(t) - \bar{Z}(t;\beta_0)\} \frac{Y_i(t)\, dB_i(t;\beta_0)}{D_n(t;\beta_0)} \Big] \Big) \\
&= \Big\{ \sum_{k=1}^n \int_0^\tau \frac{D_n(u;\beta_0)}{P_n(u;\beta_0)} dM_k(u) \Big\} \times \Big[\int_0^\tau \sum_{i=1}^n \{Z_i(t) - \bar{Z}(t;\beta_0)\} \frac{Y_i(t)\, dB_i(t;\beta_0)}{D_n(t;\beta_0)} \Big] \\
&\quad - \Big\{ \sum_{k=1}^n \int_0^0 \frac{D_n(u;\beta_0)}{P_n(u;\beta_0)} dM_k(u) \Big\} \times \Big[\int_0^0 \sum_{i=1}^n \{Z_i(t) - \bar{Z}(t;\beta_0)\} \frac{Y_i(t)\, dB_i(t;\beta_0)}{D_n(t;\beta_0)} \Big] \\
&\quad - \int_0^\tau \Big(\Big\{ \sum_{k=1}^n \frac{D_n(t;\beta_0)}{P_n(t;\beta_0)} dM_k(t) \Big\} \times \Big[\int_0^t \sum_{i=1}^n \{Z_i(u) - \bar{Z}(u;\beta_0)\} \frac{Y_i(u)\, dB_i(u;\beta_0)}{D_n(u;\beta_0)} \Big] \Big) \\
&= \int_0^\tau \Big(\Big\{ \sum_{i=1}^n \frac{D_n(t;\beta_0)}{P_n(t;\beta_0)} dM_i(t) \Big\} \times \Big[\int_t^\tau \sum_{k=1}^n \{Z_k(u) - \bar{Z}(u;\beta_0)\} \frac{Y_k(u)\, dB_k(u;\beta_0)}{D_n(u;\beta_0)} \Big] \Big).
\end{aligned}$$

Finally, we obtain a martingale representation of $S_n(\beta_0)$,

$$\begin{aligned}
S_n(\beta_0) &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t;\beta_0)\} dM_i(t) - \sum_{i=1}^n \int_0^\tau \frac{D_n(t;\beta_0)}{P_n(t;\beta_0)} \\
&\qquad \times \Big[\sum_{k=1}^n \int_t^\tau \{Z_k(u) - \bar{Z}(u;\beta_0)\} \frac{Y_k(u)\, dB_k(u;\beta_0)}{D_n(u;\beta_0)} \Big] dM_i(t) \\
&= \sum_{i=1}^n \int_0^\tau \Big[Z_i(t) - \bar{Z}(t;\beta_0) - \frac{D_n(t;\beta_0)}{P_n(t;\beta_0)} \sum_{k=1}^n \int_t^\tau \{Z_k(u) - \bar{Z}(u;\beta_0)\} \\
&\qquad \times \frac{Y_k(u)\, dB_k(u;\beta_0)}{D_n(u;\beta_0)} \Big] dM_i(t) \\
&\equiv \sum_{i=1}^n \int_0^\tau \xi_i(t;\beta_0) dM_i(t).
\end{aligned}$$

<u>*Step 2.*</u> Here we show the consistency of $\hat{\beta}_n$.

Define

$$
\begin{aligned}
u(t;\beta) &= \lim_{n\to\infty} \bar{Z}(t;\beta) = \lim_{n\to\infty} \frac{n^{-1}\sum_{i=1}^n Z_i(t)Y_i(t)\exp\{\beta Z_i(t)\}}{n^{-1}\sum_{i=1}^n Y_i(t)\exp\{\beta Z_i(t)\}} \\
&= \frac{\mathrm{E}[Z(t)Y(t)\exp\{\beta Z(t)\}]}{\mathrm{E}[Y(t)\exp\{\beta Z(t)\}]}, \\
R(t;\beta) &= \lim_{n\to\infty} \hat{R}_n(t;\beta) = \lim_{n\to\infty} \frac{1}{D_n(t;\beta)} \int_0^t D_n(u;\beta) \frac{n^{-1}\sum_{i=1}^n dN_i(u)}{n^{-1}\sum_{i=1}^n Y_i(u)\exp\{\beta Z_i(u)\}} \\
&= \frac{1}{D(t;\beta)} \int_0^t D(u;\beta) \frac{\mathrm{E}\{dN(u)\}}{\mathrm{E}[Y(u)\exp\{\beta Z(u)\}]}, \\
D(t;\beta) &= \lim_{n\to\infty} D_n(t;\beta) = \lim_{n\to\infty} \exp\Big[\int_0^t \frac{n^{-1}\sum_{i=1}^n Y_i(u)\,d\exp\{\beta Z_i(u)\}}{n^{-1}\sum_{i=1}^n Y_i(u)\exp\{\beta Z_i(u)\}}\Big] \\
&= \exp\Big(\int_0^t \frac{\mathrm{E}[Y(u)\,d\exp\{\beta Z(u)\}]}{\mathrm{E}[Y(u)\exp\{\beta Z(u)\}]}\Big).
\end{aligned}
$$

For arbitrary $\beta \neq \beta_0$, we have

$$
\begin{aligned}
&\frac{1}{n}S_n(\beta) - \frac{1}{n}S_n(\beta_0) \\
&= \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t;\beta)\}\{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta)\,dB_i(t;\beta)\} \\
&\quad - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t;\beta_0)\}\{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta_0)\,dB_i(t;\beta_0)\} \\
&= \int_0^\tau \big\{\frac{1}{n}\sum_{i=1}^n Y_i(t)\,dB_i(t;\beta)\big\}\bar{Z}(t;\beta)\widehat{R}_n(t;\beta) - \int_0^\tau \big\{\frac{1}{n}\sum_{i=1}^n Y_i(t)\,dB_i(t;\beta_0)\big\}\bar{Z}(t;\beta_0)\widehat{R}_n(t;\beta_0) \\
&\quad - \int_0^\tau \big\{\frac{1}{n}\sum_{i=1}^n Z_i(t)\,Y_i(t)\,dB_i(t;\beta)\big\}\widehat{R}_n(t;\beta) + \int_0^\tau \big\{\frac{1}{n}\sum_{i=1}^n Z_i(t)\,Y_i(t)\,dB_i(t;\beta_0)\big\}\widehat{R}_n(t;\beta_0) \\
&\quad - \int_0^\tau \big\{\frac{1}{n}\sum_{i=1}^n dN_i(t)\big\}\{\bar{Z}(t;\beta) - \bar{Z}(t;\beta_0)\} \\
&\longrightarrow_p \int_0^\tau \mathrm{E}[Y(t)\,d\exp\{\beta Z(t)\}]u(t;\beta)R(t;\beta) - \int_0^\tau \mathrm{E}[Y(t)\,d\exp\{\beta_0 Z(t)\}]u(t;\beta_0)R(t;\beta_0) \\
&\quad - \int_0^\tau \mathrm{E}[Z(t)Y(t)\,d\exp\{\beta Z(t)\}]R(t;\beta) + \int_0^\tau \mathrm{E}[Z(t)Y(t)\,d\exp\{\beta_0 Z(t)\}]R(t;\beta_0) \\
&\quad - \int_0^\tau \mathrm{E}\{dN(t)\}\{u(t;\beta) - u(t;\beta_0)\} \equiv \Gamma(\beta).
\end{aligned}
$$

It is seen that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\xi_i(t;\beta_0)\\
&= \frac{1}{n}\sum_{i=1}^{n}\Big[Z_i(t) - \bar{Z}(t;\beta_0) - \frac{D_n(t;\beta_0)}{P_n(t;\beta_0)}\sum_{k=1}^{n}\int_t^\tau \{Z_k(u) - \bar{Z}(u;\beta_0)\}\frac{Y_k(u)\,dB_k(u;\beta_0)}{D_n(u;\beta_0)}\Big]\\
&= \frac{1}{n}\sum_{i=1}^{n}\{Z_i(t)\} - \bar{Z}(t;\beta_0) - \frac{D_n(t;\beta_0)}{P_n(t;\beta_0)}\sum_{k=1}^{n}\int_t^\tau \Big[\{Z_k(u) - \bar{Z}(u;\beta_0)\}\frac{Y_k(u)\,dB_k(u;\beta_0)}{D_n(u;\beta_0)}\Big]\\
&= \frac{1}{n}\sum_{i=1}^{n}\{Z_i(t)\} - \bar{Z}(t;\beta_0) - \frac{D_n(t;\beta_0)}{n^{-1}\sum_{i=1}^{n}Y_i(t)\exp\{\beta_0 Z_i(t)\}}\int_t^\tau \frac{1}{n}\sum_{k=1}^{n}\Big[\{Z_k(u) - \bar{Z}(u;\beta_0)\}\\
&\qquad \times \frac{Y_k(u)\,d\exp\{\beta_0 Z_k(u)\}}{D_n(u;\beta_0)}\Big]\\
&\longrightarrow_p \mathrm{E}\{Z(t)\} - u(t;\beta_0) - \frac{D(t;\beta_0)}{\mathrm{E}[Y(t)\exp\{\beta_0 Z(t)\}]}\int_t^\tau \mathrm{E}\Big[\{Z(s) - u(s;\beta_0)\}\\
&\qquad \times \frac{Y(s)\,d\exp\{\beta_0 Z(s)\}]}{D(s;\beta_0)}\Big].
\end{aligned}
$$

In addition, by the property of martingale, we have $\mathrm{E}\{dM(t)|\mathcal{F}_{t-}\} = 0$. It then follows from the weak law of large numbers that

$$
\frac{1}{n}S_n(\beta_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \xi_i(t;\beta_0)dM_i(t) \rightarrow_p 0.
$$

Therefore, $n^{-1}S_n(\beta) \rightarrow_p \Gamma(\beta)$.

Moreover, we obtain that

$$
\begin{aligned}
&\frac{\partial}{\partial\beta}\{\frac{1}{n}S_n(\beta)\}\\
&= \frac{\partial}{\partial\beta}\{\frac{1}{n}S_n(\beta) - \frac{1}{n}S_n(\beta_0)\}\\
&= \frac{\partial}{\partial\beta}\Big[\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{Z_i(t) - \bar{Z}(t;\beta)\}\{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta)\,dB_i(t;\beta)\}\Big]\\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Big[\frac{\partial}{\partial\beta}\bar{Z}(t;\beta) \times \{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta)\,dB_i(t;\beta)\}\\
&\qquad + Y_i(t)\{Z_i(t) - \bar{Z}(t;\beta)\}\{dB_i(t;\beta)\frac{\partial}{\partial\beta}\widehat{R}_n(t;\beta) + \widehat{R}_n(t;\beta)\frac{\partial}{\partial\beta}dB_i(t;\beta)\}\Big],
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{\partial}{\partial\beta}\Gamma(\beta)\\
&= \frac{\partial}{\partial\beta}\Big(\int_0^\tau \mathrm{E}[Y(t)\,d\exp\{\beta Z(t)\}]u(t;\beta)R(t;\beta) - \int_0^\tau \mathrm{E}\{dN(t)\}\{u(t;\beta)\}\\
&\quad - \int_0^\tau \mathrm{E}[Z(t)Y(t)\,d\exp\{\beta Z(t)\}]R(t;\beta)\Big)\\
&= -\int_0^\tau \frac{\partial}{\partial\beta}u(t;\beta)\times \mathrm{E}[dN(t)-Y(t)R(t;\beta)\,d\exp\{\beta Z(t)\}]\\
&\quad - \int_0^\tau \mathrm{E}\Big(Y(t)\{Z(t)-u(t;\beta)\}\Big[d\exp\{\beta Z(t)\}\frac{\partial}{\partial\beta}R(t;\beta) + R(t;\beta)\frac{\partial}{\partial\beta}d\exp\{\beta Z(t)\}\Big]\Big).
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial}{\partial\beta}\bar{Z}(t;\beta) &= \frac{n^{-1}\sum_{i=1}^n[Z_i^2(t)Y_i(t)\exp\{\beta Z_i(t)\}]}{n^{-1}\sum_{i=1}^n[Y_i(t)\exp\{\beta Z_i(t)\}]} - \bar{Z}^2(t;\beta),\\
\frac{\partial}{\partial\beta}\widehat{R}_n(t;\beta) &= -\frac{\partial D_n(t;\beta)/\partial\beta}{D_n(t;\beta)}\widehat{R}_n(t;\beta) + \frac{1}{D_n(t;\beta)}\int_0^t \frac{n^{-1}\sum_{i=1}^n\{dN_i(u)\}}{n^{-1}\sum_{i=1}^n[Y_i(u)\exp\{\beta Z_i(u)\}]}\\
&\qquad \times\Big(\frac{\partial}{\partial\beta}D_n(u;\beta) - D_n(u;\beta)\frac{n^{-1}\sum_{i=1}^n[Z_i(u)Y_i(u)\exp\{\beta Z_i(u)\}]}{n^{-1}\sum_{i=1}^n[Y_i(u)\exp\{\beta Z_i(u)\}]}\Big),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial\beta}D_n(t;\beta) &= D_n(t;\beta)\int_0^t \frac{n^{-1}\sum_{i=1}^n[Y_i(u)\exp\{\beta Z_i(u)\}dZ_i(u)]}{n^{-1}\sum_{i=1}^n[Y_i(u)\exp\{\beta Z_i(u)\}]}\times\Big(1\\
&\qquad + \frac{\beta n^{-1}\sum_{i=1}^n[Z_i(u)Y_i(u)\exp\{\beta Z_i(u)\}dZ_i(u)]}{n^{-1}\sum_{i=1}^n[Y_i(u)\exp\{\beta Z_i(u)\}dZ_i(u)]}\\
&\qquad - \frac{\beta n^{-1}\sum_{i=1}^n[Z_i(u)Y_i(u)\exp\{\beta Z_i(u)\}]}{n^{-1}\sum_{i=1}^n[Y_i(u)\exp\{\beta Z_i(u)\}]}\Big),\\
\frac{\partial}{\partial\beta}u(t;\beta) &= \frac{\mathrm{E}[Z^2(t)Y(t)\exp\{\beta Z(t)\}]}{\mathrm{E}[Y(t)\exp\{\beta Z(t)\}]} - u^2(t;\beta),\\
\frac{\partial}{\partial\beta}R(t;\beta) &= -\frac{\partial D(t;\beta)/\partial\beta}{D(t;\beta)}R(t;\beta) + \frac{1}{D(t;\beta)}\int_0^t \frac{\mathrm{E}\{dN(u)\}}{\mathrm{E}[Y(u)\exp\{\beta Z(u)\}]}\\
&\qquad \times\Big(\frac{\partial}{\partial\beta}D(u;\beta) - D(u;\beta)\frac{\mathrm{E}[Z(u)Y(u)\exp\{\beta Z(u)\}]}{\mathrm{E}[Y(u)\exp\{\beta Z(u)\}]}\Big),\\
\frac{\partial}{\partial\beta}D(t;\beta) &= D(t;\beta)\int_0^t \frac{\mathrm{E}[Y(u)\exp\{\beta Z(u)\}dZ(u)]}{\mathrm{E}[Y(u)\exp\{\beta Z(u)\}]}\times\Big(1\\
&\qquad + \frac{\beta\mathrm{E}[Z(u)Y(u)\exp\{\beta Z(u)\}dZ(u)]}{\mathrm{E}[Y(u)\exp\{\beta Z(u)\}dZ(u)]} - \frac{\beta\mathrm{E}[Z(u)Y(u)\exp\{\beta Z(u)\}]}{\mathrm{E}[Y(u)\exp\{\beta Z(u)\}]}\Big).
\end{aligned}
$$

Apparently $\Gamma(\beta_0) = 0$. It is seen that $|\frac{\partial}{\partial\beta}\Gamma(\beta_0)| > 0$, by conditions (C6),(C7). It then follows that $\Gamma(\beta)$ is strictly monotone around a neighborhood of β_0, $U(\beta_0)$. By the Glivenko-Cantelli lemma, $\| \frac{\partial}{\partial\beta}\{\frac{1}{n}S_n(\beta)\} - \frac{\partial}{\partial\beta}\Gamma(\beta) \| \to_p 0$ uniformly in $\beta \in U(\beta_0)$. Also we have $n^{-1}S_n(\beta) \to_p \Gamma(\beta)$. As a result, for any $\epsilon > 0$, there exists sufficiently large n and $\nu > 0$, such that

$$S_n(\beta_0 - \epsilon) \times S_n(\beta_0 + \epsilon) < 0, \text{ and } \Pr\big\{\widehat{\beta}_n \in (\beta_0 - \epsilon, \beta_0 + \epsilon)\big\} > 1 - \nu,$$

which implies that $\widehat{\beta}_n$ is consistent.

Step 3. Here we show the asymptotic normality of $\hat{\beta}_n$.

An application of the Taylor expansion gives that

$$\begin{aligned} S_n(\widehat{\beta}_n) - S_n(\beta_0) &= \frac{\partial S_n(\beta_0)}{\partial\beta}(\widehat{\beta}_n - \beta_0) + o(\widehat{\beta}_n - \beta_0) \\ \Longrightarrow \sqrt{n}(\widehat{\beta}_n - \beta_0) &= -\frac{1}{\sqrt{n}}S_n(\beta_0)\Big\{\frac{1}{n}\frac{\partial S_n(\beta_0)}{\partial\beta}\Big\}^{-1} + o(\widehat{\beta}_n - \beta_0). \end{aligned}$$

It follows from the martingale representation of $S_n(\beta_0)$ and the martingale central limit theorem that $\frac{1}{\sqrt{n}}S_n(\beta_0)$ is asymptotically mean-zero normal with variance

$$\begin{aligned} V &= \lim_{n\to\infty} < \frac{1}{\sqrt{n}}S_n(\beta_0), \frac{1}{\sqrt{n}}S_n(\beta_0) > (\tau) \\ &= \lim_{n\to\infty} < \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\cdot} \xi_i(t;\beta_0)dM_i(t), \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\cdot} \xi_i(t;\beta_0)dM_i(t) > (\tau) \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \xi_i^2(t;\beta_0)\,Y_i(t)\,d\Lambda_i(t) \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \xi_i^2(t;\beta_0)\,Y_i(t)\{dR_0(t)\,B_i(t;\beta_0) + R_0(t)\,dB_i(t;\beta_0)\}. \end{aligned}$$

Then by the Delta method and the consistency of $\widehat{\beta}_n$, it is seen that

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \longrightarrow_d \mathrm{N}(0, V/U^2),$$

where

$$
\begin{aligned}
U &= \lim_{n\to\infty} \frac{1}{n}\frac{\partial S_n(\beta_0)}{\partial\beta} \\
&= \lim_{n\to\infty} -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \Big[\frac{\partial}{\partial\beta}\bar{Z}(t;\beta_0) \times \{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta_0)\,dB_i(t;\beta_0)\} \\
&\qquad + Y_i(t)\{Z_i(t) - \bar{Z}(t;\beta_0)\}\{dB_i(t;\beta_0)\frac{\partial}{\partial\beta}\widehat{R}_n(t;\beta_0) + \widehat{R}_n(t;\beta_0)\frac{\partial}{\partial\beta}dB_i(t;\beta_0)\}\Big].
\end{aligned}
$$

Step 4. Here we show the consistency of the variance estimators. This completes the proof of Theorem 3.2.1.

We first prove that $\widehat{R}_n(t;\widehat{\beta}_n)$ converges to the true value $R_0(t)$ in probability. Recall the martingale representation of $\widehat{R}_n(t;\beta_0) - R_0(t)$ described in step 1. The weak law of large numbers then implies that

$$
\widehat{R}_n(t;\beta_0) - R_0(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{t}\frac{D_n(u;\beta_0)dM_i(u)}{D_n(t;\beta_0)[n^{-1}\sum_{i=1}^{n} Y_i(u)\exp\{\beta_0 Z_i(u)\}]} \longrightarrow_p 0.
$$

In addition, because $\widehat{\beta}_n$ is consistent, it suffices by the continuous mapping theorem to show that $\widehat{R}_n(t;\widehat{\beta}_n) \to_p \widehat{R}_n(t;\beta_0)$. As a result, it can be shown that

$$
\widehat{R}_n(t;\widehat{\beta}_n) - R_0(t) = \{\widehat{R}_n(t;\widehat{\beta}_n) - \widehat{R}_n(t;\beta_0)\} + \{\widehat{R}_n(t;\beta_0) - R_0(t)\} \longrightarrow_p 0.
$$

Finally, the consistency of the variance estimators follows from the consistency of $\widehat{\beta}_n$ and $\widehat{R}_n(t;\widehat{\beta}_n)$, the continuous mapping theorem and the Slutsky's theorem. □

3.2.4 Hypothesis Testing

Based on the large sample properties of the estimator, the Wald test can be performed to examine the significance of regression parameters in the complementary log-log survival model. To test the hypothesis

$$
H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \qquad \text{versus} \qquad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0,
$$

the Wald statistic is of the form

$$T_n = \sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T \{\widehat{\boldsymbol{U}}_n^{-1} \widehat{\boldsymbol{V}}_n (\widehat{\boldsymbol{U}}_n^{-1})^T\}^{-1} \sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0).$$

Under the null hypothesis, T_n converges to a chi-squared distribution with p degrees of freedom, χ_p^2. Then a Wald test of size α is to reject $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ if $T_n > \chi_p^2(\alpha)$, where $\chi_p^2(\alpha)$ is the upper α quantile of the chi-squared distribution with p degrees of freedom. Also we can compute the p value by $p = \Pr(\chi_p^2 > T_n)$ and reject $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ if $p < \alpha$.

In practice, the interest often centers on testing the composite null hypothesis. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are of dimensions r and $p - r$, respectively. Consider the partition of $\widehat{\boldsymbol{U}}_n^{-1} \widehat{\boldsymbol{V}}_n (\widehat{\boldsymbol{U}}_n^{-1})^T$:

$$\widehat{\boldsymbol{U}}_n^{-1} \widehat{\boldsymbol{V}}_n (\widehat{\boldsymbol{U}}_n^{-1})^T = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_{11} and A_{22} are of dimensions $r \times r$ and $(p - r) \times (p - r)$, respectively. Let $\boldsymbol{\beta}_{10}$ be a fixed r-dimensional vector. To test the composite hypothesis,

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10} \qquad \text{versus} \qquad H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_{10},$$

the Wald statistic is given by

$$T_{1n} = \sqrt{n}(\widehat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_{10})^T A_{11}^{-1} \sqrt{n}(\widehat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_{10}).$$

Under the composite null hypothesis, $T_{1n} \to_d \chi_r^2(\alpha)$. We can reject $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ if $T_{1n} > \chi_r^2(\alpha)$, where $\chi_r^2(\alpha)$ is the upper α quantile of the chi-squared distribution with r degrees of freedom, or if $p = \Pr(\chi_r^2 > T_n) < \alpha$.

Let $\beta_i, i = 1, 2, \ldots, p$ be the ith component of vector $\boldsymbol{\beta}$ and $\widehat{\beta}_{in}$ be the corresponding estimator. We can construct a $100(1 - \alpha)\%$ confidence interval for β_i as

$$\widehat{\beta}_{in} - z_{\alpha/2} \times \sqrt{\frac{1}{n}\{\widehat{\boldsymbol{U}}_n^{-1} \widehat{\boldsymbol{V}}_n (\widehat{\boldsymbol{U}}_n^{-1})^T\}_{ii}},\ \widehat{\beta}_{in} + z_{\alpha/2} \times \sqrt{\frac{1}{n}\{\widehat{\boldsymbol{U}}_n^{-1} \widehat{\boldsymbol{V}}_n (\widehat{\boldsymbol{U}}_n^{-1})^T\}_{ii}}$$

where $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ quantile of a standard normal random variable, and $\{\widehat{\boldsymbol{U}}_n^{-1} \widehat{\boldsymbol{V}}_n (\widehat{\boldsymbol{U}}_n^{-1})^T\}_{ii}$ is the entry in the ith row and ith column of matrix $\widehat{\boldsymbol{U}}_n^{-1} \widehat{\boldsymbol{V}}_n (\widehat{\boldsymbol{U}}_n^{-1})^T$. Based on the confidence interval, a valid Wald test of size α is to reject $H_0 : \beta_i = 0$ if the $100(1 - \alpha)\%$ confidence interval for β_i does not contain zero.

3.3 Weighted Estimating Equation Estimator

3.3.1 Inference Procedures

More generally, we can define a class of weighted estimators, say $\widehat{\boldsymbol{\beta}}_{w,n}$, by solving the following estimating equations:

$$\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{W}_n(t)\boldsymbol{Z}_i(t)[dN_i(t)-Y_i(t)\,d\exp\{\widehat{\alpha}_n(t;\boldsymbol{\beta})+\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}]=\mathbf{0}$$

where $\boldsymbol{W}_n(t)$ is some prespecified weight function not depending on $\boldsymbol{\beta}$ and i. Recall

$$\widehat{\alpha}_n(t;\boldsymbol{\beta})=\log\Big[\frac{1}{D_n(t;\boldsymbol{\beta})}\int_0^t\Big\{D_n(u;\boldsymbol{\beta})\frac{\sum_{i=1}^n dN_i(u)}{P_n(u;\boldsymbol{\beta})}\Big\}\Big].$$

The derivation of $\widehat{\alpha}_n(t;\boldsymbol{\beta})$ is given in Section 3.2.

Apparently, $\widehat{\boldsymbol{\beta}}_n$ is one special case of $\widehat{\boldsymbol{\beta}}_{w,n}$ with $\boldsymbol{W}_n(t)=\mathbf{1}$. The procedure for obtaining $\widehat{\boldsymbol{\beta}}_{w,n}$ resembles the aforementioned procedure for $\widehat{\boldsymbol{\beta}}_n$. The Newton-Raphson method also works well in the case of the weighted estimating equation. The following theorem presents the large sample properties for $\widehat{\boldsymbol{\beta}}_{w,n}$. Proof of this theorem is similar to the proof of Theorem 3.2.1 and we omit details that are already shown in the proof of Theorem 3.2.1. To show the consistency of $\widehat{\boldsymbol{\beta}}_{w,n}$, we assume one additional regularity condition **C8**:

$$\begin{aligned}&\|\int_0^{\tau}\boldsymbol{w}(t)\frac{\partial\boldsymbol{u}(t;\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\mathrm{E}[dN(t)-Y(t)R(t;\boldsymbol{\beta}_0)\,d\exp\{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)\}]\\&+\int_0^{\tau}\mathrm{E}\Big(\boldsymbol{w}(t)Y(t)\{\boldsymbol{Z}(t)-\boldsymbol{u}(t;\boldsymbol{\beta}_0)\}\Big[d\exp\{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)\}\frac{\partial}{\partial\boldsymbol{\beta}}R(t;\boldsymbol{\beta}_0)\\&+R(t;\boldsymbol{\beta}_0)\frac{\partial}{\partial\boldsymbol{\beta}}d\exp\{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)\}\Big]^T\Big)\|>0,\end{aligned}$$

Theorem 3.3.1. *Under regularity conditions C1-C6 and C8, as $n\to\infty$, let $\boldsymbol{w}(t)$ be a deterministic function such that $\boldsymbol{W}_n(t)\to\boldsymbol{w}(t)$ almost surely, $\widehat{\boldsymbol{\beta}}_{w,n}$ is strongly consistent, and $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{w,n}-\boldsymbol{\beta}_0)$ converges weakly to a p-variate normal with mean zero and a covariance* *matrix* $\boldsymbol{U}_w^{-1}\boldsymbol{V}_w(\boldsymbol{U}_w^{-1})^T$*, where*

$$\begin{aligned}
\boldsymbol{U}_w &= \lim_{n\to\infty} -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\Big[\boldsymbol{W}_n(t)\frac{\partial\overline{\boldsymbol{Z}}(t;\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\{dN_i(t)-Y_i(t)\,\widehat{R}_n(t;\boldsymbol{\beta}_0)\,dB_i(t;\boldsymbol{\beta}_0)\} \\
&\quad +Y_i(t)\boldsymbol{W}_n(t)\{\boldsymbol{Z}_i(t)-\overline{\boldsymbol{Z}}(t;\boldsymbol{\beta}_0)\}\Big\{\frac{\partial\widehat{R}_n(t;\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}dB_i(t;\boldsymbol{\beta}_0) \\
&\quad +\widehat{R}_n(t;\boldsymbol{\beta}_0)\frac{\partial dB_i(t;\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\Big\}^T\Big], \\
\boldsymbol{V}_w &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{\xi}_{w,i}(t;\boldsymbol{\beta}_0)^{\otimes 2}Y_i(t)\{dR_0(t)B_i(t;\boldsymbol{\beta}_0)+R_0(t)\,dB_i(t;\boldsymbol{\beta}_0)\}, \\
\boldsymbol{\xi}_{w,i}(t;\boldsymbol{\beta}) &= \boldsymbol{W}_n(t)\{\boldsymbol{Z}_i(t)-\overline{\boldsymbol{Z}}(t;\boldsymbol{\beta})\}-\frac{D_n(t;\boldsymbol{\beta})}{P_n(t;\boldsymbol{\beta})} \\
&\quad \times\sum_{k=1}^{n}\int_t^{\tau}\Big[\boldsymbol{W}_n(u)\{\boldsymbol{Z}_k(u)-\overline{\boldsymbol{Z}}(u;\boldsymbol{\beta})\}\frac{Y_k(u)\,dB_k(u;\boldsymbol{\beta})}{D_n(u;\boldsymbol{\beta})}\Big].
\end{aligned}$$

Moreover, consistent estimators for $\boldsymbol{U}_w$ *and* $\boldsymbol{V}_w$ *are given respectively by*

$$\begin{aligned}
\widehat{\boldsymbol{U}}_{w,n} &= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\Big[\boldsymbol{W}_n(t)\frac{\partial\overline{\boldsymbol{Z}}(t;\widehat{\boldsymbol{\beta}}_{w,n})}{\partial\boldsymbol{\beta}}\{dN_i(t)-Y_i(t)\,\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_{w,n})\,dB_i(t;\widehat{\boldsymbol{\beta}}_{w,n})\} \\
&\quad +Y_i(t)\boldsymbol{W}_n(t)\{\boldsymbol{Z}_i(t)-\overline{\boldsymbol{Z}}(t;\widehat{\boldsymbol{\beta}}_{w,n})\}\Big\{\frac{\partial\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_{w,n})}{\partial\boldsymbol{\beta}}dB_i(t;\widehat{\boldsymbol{\beta}}_{w,n}) \\
&\quad +\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_{w,n})\frac{\partial dB_i(t;\widehat{\boldsymbol{\beta}}_{w,n})}{\partial\boldsymbol{\beta}}\Big\}^T\Big], \\
\widehat{\boldsymbol{V}}_{w,n} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{\xi}_{w,i}(t;\widehat{\boldsymbol{\beta}}_{w,n})^{\otimes 2}Y_i(t)\{d\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_{w,n})B_i(t;\widehat{\boldsymbol{\beta}}_{w,n})+\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_{w,n})\,dB_i(t;\widehat{\boldsymbol{\beta}}_{w,n})\}.
\end{aligned}$$

Also, for ease of presentation, we state the proof of Theorem 3.3.1 in a univariate setting, which can be generalized easily to the multivariate case.

Proof of Theorem 3.3.1. Define

$$S_{w,n}(\beta_0)=\sum_{i=1}^{n}\int_0^{\tau}W_n(t)Z_i(t)[dN_i(t)-Y_i(t)\{d\widehat{R}_n(t;\beta_0)B_i(t;\beta_0)+\widehat{R}_n(t;\beta_0)dB_i(t;\beta_0)\}].$$

First we establish a martingale representation of $S_{w,n}(\beta_0)$.

We can write

$$
\begin{aligned}
S_{w,n}(\beta_0) &= \sum_{i=1}^{n}\int_0^{\tau} W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}\{dN_i(t)-Y_i(t)\widehat{R}_n(t;\beta_0)\,dB_i(t;\beta_0)\}\\
&\quad -\sum_{i=1}^{n}\int_0^{\tau} W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}Y_i(t)\,d\widehat{R}_n(t;\beta_0)\,B_i(t;\beta_0)\\
&= \sum_{i=1}^{n}\int_0^{\tau} W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}\{dN_i(t)-Y_i(t)\widehat{R}_n(t;\beta_0)\,dB_i(t;\beta_0)\}\\
&\quad -\int_0^{\tau} W_n(t)d\widehat{R}_n(t;\beta_0)\sum_{i=1}^{n}\{Z_i(t)\,Y_i(t)\,B_i(t;\beta_0)\}+\int_0^{\tau} W_n(t)d\widehat{R}_n(t;\beta_0)\\
&\quad \times \frac{\sum_{i=1}^{n} Z_i(t)Y_i(t)B_i(t;\beta_0)}{\sum_{i=1}^{n} Y_i(t)B_i(t;\beta_0)}\sum_{i=1}^{n}\{Y_i(t)B_i(t;\beta_0)\}\\
&= \sum_{i=1}^{n}\int_0^{\tau} W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}\{dN_i(t)-Y_i(t)\widehat{R}_n(t;\beta_0)\,dB_i(t;\beta_0)\}\\
&= \sum_{i=1}^{n}\int_0^{\tau} W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}\big[dM_i(t)+Y_i(t)\,dB_i(t;\beta_0)\{R_0(t)-\widehat{R}_n(t;\beta_0)\}\big]\\
&\quad +\sum_{i=1}^{n}\int_0^{\tau} W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}Y_i(t)\,dR_0(t)\,B_i(t;\beta_0)\\
&= \sum_{i=1}^{n}\int_0^{\tau} W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}dM_i(t)+\sum_{i=1}^{n}\int_0^{\tau} W_n(t)\big[\{Z_i(t)-\bar{Z}(t;\beta_0)\}Y_i(t)\\
&\quad \times dB_i(t;\beta_0)\{R_0(t)-\widehat{R}_n(t;\beta_0)\}\big]+\int_0^{\tau} W_n(t)dR_0(t)\sum_{i=1}^{n}\{Z_i(t)\,Y_i(t)\,B_i(t;\beta_0)\}\\
&\quad -\int_0^{\tau} W_n(t)dR_0(t)\frac{\sum_{i=1}^{n} Z_i(t)Y_i(t)B_i(t;\beta_0)}{\sum_{i=1}^{n} Y_i(t)B_i(t;\beta_0)}\sum_{i=1}^{n}\{Y_i(t)B_i(t;\beta_0)\}\\
&= \sum_{i=1}^{n}\int_0^{\tau} W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}dM_i(t)+\sum_{i=1}^{n}\int_0^{\tau} W_n(t)\big[\{Z_i(t)-\bar{Z}(t;\beta_0)\}Y_i(t)\\
&\quad \times dB_i(t;\beta_0)\{R_0(t)-\widehat{R}_n(t;\beta_0)\}\big]\\
&= \sum_{i=1}^{n}\int_0^{\tau} W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}dM_i(t)-\sum_{i=1}^{n}\int_0^{\tau}\Big[W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}Y_i(t)\\
&\quad \times dB_i(t;\beta_0)\sum_{k=1}^{n}\int_0^{t}\Big\{\frac{D_n(u;\beta_0)}{D_n(t;\beta_0)\,P_n(u;\beta_0)}dM_k(u)\Big\}\Big].
\end{aligned}
$$

As a result of integration by parts,

$$
\begin{aligned}
&\sum_{i=1}^{n}\int_0^{\tau}\Big[W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}Y_i(t)\,dB_i(t;\beta_0)\sum_{k=1}^{n}\int_0^{t}\Big\{\frac{D_n(u;\beta_0)}{D_n(t;\beta_0)\,P_n(u;\beta_0)}dM_k(u)\Big\}\Big] \\
&= \int_0^{\tau}\bigg(\Big\{\sum_{k=1}^{n}\int_0^{t}\frac{D_n(u;\beta_0)}{P_n(u;\beta_0)}dM_k(u)\Big\}\times\Big[W_n(t)\sum_{i=1}^{n}\{Z_i(t)-\bar{Z}(t;\beta_0)\}\frac{Y_i(t)\,dB_i(t;\beta_0)}{D_n(t;\beta_0)}\Big]\bigg) \\
&= \Big\{\sum_{k=1}^{n}\int_0^{\tau}\frac{D_n(u;\beta_0)}{P_n(u;\beta_0)}dM_k(u)\Big\}\times\Big[\int_0^{\tau}W_n(t)\sum_{i=1}^{n}\{Z_i(t)-\bar{Z}(t;\beta_0)\}\frac{Y_i(t)\,dB_i(t;\beta_0)}{D_n(t;\beta_0)}\Big] \\
&\quad -\Big\{\sum_{k=1}^{n}\int_0^{0}\frac{D_n(u;\beta_0)}{P_n(u;\beta_0)}dM_k(u)\Big\}\times\Big[\int_0^{0}W_n(t)\sum_{i=1}^{n}\{Z_i(t)-\bar{Z}(t;\beta_0)\}\frac{Y_i(t)\,dB_i(t;\beta_0)}{D_n(t;\beta_0)}\Big] \\
&\quad -\int_0^{\tau}\bigg(\Big\{\sum_{k=1}^{n}\frac{D_n(t;\beta_0)}{P_n(t;\beta_0)}dM_k(t)\Big\}\times\Big[\int_0^{t}W_n(u)\sum_{i=1}^{n}\{Z_i(u)-\bar{Z}(u;\beta_0)\}\frac{Y_i(u)\,dB_i(u;\beta_0)}{D_n(u;\beta_0)}\Big]\bigg) \\
&= \int_0^{\tau}\bigg(\Big\{\sum_{i=1}^{n}\frac{D_n(t;\beta_0)}{P_n(t;\beta_0)}dM_i(t)\Big\}\times\Big[\int_t^{\tau}W_n(u)\sum_{k=1}^{n}\{Z_k(u)-\bar{Z}(u;\beta_0)\}\frac{Y_k(u)\,dB_k(u;\beta_0)}{D_n(u;\beta_0)}\Big]\bigg).
\end{aligned}
$$

Finally, we obtain a martingale representation of $S_{w,n}(\beta_0)$,

$$
\begin{aligned}
S_{w,n}(\beta_0) &= \sum_{i=1}^{n}\int_0^{\tau}W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}dM_i(t)-\sum_{i=1}^{n}\int_0^{\tau}\frac{D_n(t;\beta_0)}{P_n(t;\beta_0)} \\
&\quad \times\Big[\sum_{k=1}^{n}\int_t^{\tau}W_n(u)\{Z_k(u)-\bar{Z}(u;\beta_0)\}\frac{Y_k(u)\,dB_k(u;\beta_0)}{D_n(u;\beta_0)}\Big]dM_i(t) \\
&= \sum_{i=1}^{n}\int_0^{\tau}\Big[W_n(t)\{Z_i(t)-\bar{Z}(t;\beta_0)\}-\frac{D_n(t;\beta_0)}{P_n(t;\beta_0)}\sum_{k=1}^{n}\int_t^{\tau}W_n(u) \\
&\quad \times\{Z_k(u)-\bar{Z}(u;\beta_0)\}\frac{Y_k(u)\,dB_k(u;\beta_0)}{D_n(u;\beta_0)}\Big]dM_i(t) \\
&\equiv \sum_{i=1}^{n}\int_0^{\tau}\xi_{w,i}(t;\beta_0)dM_i(t).
\end{aligned}
$$

For arbitrary $\beta\neq\beta_0$, we have

$$
\begin{aligned}
&\frac{1}{n}S_{w,n}(\beta)-\frac{1}{n}S_{w,n}(\beta_0) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}W_n(t)\{Z_i(t)-\bar{Z}(t;\beta)\}\{dN_i(t)-Y_i(t)\widehat{R}_n(t;\beta)\,dB_i(t;\beta)\}
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau W_n(t)\{Z_i(t) - \bar{Z}(t;\beta_0)\}\{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} \\
&= \int_0^\tau \Big\{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\, dB_i(t;\beta)\Big\} W_n(t)\bar{Z}(t;\beta)\widehat{R}_n(t;\beta) - \int_0^\tau \Big\{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\, dB_i(t;\beta_0)\Big\} \\
&\quad \times W_n(t)\bar{Z}(t;\beta_0)\widehat{R}_n(t;\beta_0) - \int_0^\tau \Big\{\frac{1}{n}\sum_{i=1}^{n} Z_i(t)\, Y_i(t)\, dB_i(t;\beta)\Big\} W_n(t)\widehat{R}_n(t;\beta) \\
&\quad + \int_0^\tau \Big\{\frac{1}{n}\sum_{i=1}^{n} Z_i(t)\, Y_i(t)\, dB_i(t;\beta_0)\Big\} W_n(t)\widehat{R}_n(t;\beta_0) \\
&\quad - \int_0^\tau \Big\{\frac{1}{n}\sum_{i=1}^{n} dN_i(t)\Big\} W_n(t)\{\bar{Z}(t;\beta) - \bar{Z}(t;\beta_0)\} \\
&\longrightarrow_p \int_0^\tau \mathrm{E}[Y(t)\, d\exp\{\beta Z(t)\}] w(t)u(t;\beta)R(t;\beta) - \int_0^\tau \mathrm{E}[Y(t)\, d\exp\{\beta_0 Z(t)\}] \\
&\quad \times w(t)u(t;\beta_0)R(t;\beta_0) - \int_0^\tau \mathrm{E}[Z(t)Y(t)\, d\exp\{\beta Z(t)\}] w(t)R(t;\beta) \\
&\quad + \int_0^\tau \mathrm{E}[Z(t)Y(t)\, d\exp\{\beta_0 Z(t)\}] w(t)R(t;\beta_0) \\
&\quad - \int_0^\tau \mathrm{E}\{dN(t)\} w(t)\{u(t;\beta) - u(t;\beta_0)\} \equiv \Gamma_w(\beta).
\end{aligned}
$$

We know that $\mathrm{E}\{dM(t)|\mathcal{F}_{t-}\} = 0$ and

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \xi_{w,i}(t;\beta_0) \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big[W_n(t)\{Z_i(t) - \bar{Z}(t;\beta_0)\} - \frac{D_n(t;\beta_0)}{P_n(t;\beta_0)}\sum_{k=1}^{n}\int_t^\tau W_n(u)\{Z_k(u) - \bar{Z}(u;\beta_0)\} \\
&\quad \times \frac{Y_k(u)\, dB_k(u;\beta_0)}{D_n(u;\beta_0)}\Big] \\
&\longrightarrow_p w(t)\mathrm{E}\{Z(t)\} - w(t)u(t;\beta_0) - \frac{D(t;\beta_0)}{\mathrm{E}[Y(t)\exp\{\beta_0 Z(t)\}]}\int_t^\tau w(t)\mathrm{E}\Big[\{Z(s) - u(s;\beta_0)\} \\
&\quad \times \frac{Y(s)\, d\exp\{\beta_0 Z(s)\}]}{D(s;\beta_0)}\Big].
\end{aligned}
$$

It then follows from the weak law of large numbers that

$$
\frac{1}{n}S_{w,n}(\beta_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \xi_{w,i}(t;\beta_0) dM_i(t) \rightarrow_p 0.
$$

Therefore, $n^{-1}S_{w,n}(\beta) \rightarrow_p \Gamma_w(\beta)$.

Moreover, we obtain that

$$\frac{\partial}{\partial\beta}\{\frac{1}{n}S_{w,n}(\beta)\} = \frac{\partial}{\partial\beta}\{\frac{1}{n}S_{w,n}(\beta) - \frac{1}{n}S_{w,n}(\beta_0)\}$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\Big[W_n(t)\frac{\partial\bar{Z}(t;\beta)}{\partial\beta}\{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta)\,dB_i(t;\beta)\}$$

$$+ W_n(t)Y_i(t)\{Z_i(t) - \bar{Z}(t;\beta)\}\{dB_i(t;\beta)\frac{\partial}{\partial\beta}\widehat{R}_n(t;\beta) + \widehat{R}_n(t;\beta)\frac{\partial}{\partial\beta}dB_i(t;\beta)\}\Big],$$

$$\frac{\partial}{\partial\beta}\Gamma_w(\beta)$$

$$= -\int_0^{\tau} w(t)\frac{\partial u(t;\beta)}{\partial\beta} \times \mathrm{E}[dN(t) - Y(t)R(t;\beta)\,d\exp\{\beta Z(t)\}] - \int_0^{\tau}\mathrm{E}\Big(w(t)Y(t)$$

$$\times \{Z(t) - u(t;\beta)\}\Big[d\exp\{\beta Z(t)\}\frac{\partial}{\partial\beta}R(t;\beta) + R(t;\beta)\frac{\partial}{\partial\beta}d\exp\{\beta Z(t)\}\Big]\Big).$$

Apparently $\Gamma_w(\beta_0) = 0$. It is seen that $|\frac{\partial}{\partial\beta}\Gamma_w(\beta_0)| > 0$, by conditions (C6),(C8). It then follows that $\Gamma_w(\beta)$ is strictly monotone around a neighborhood of β_0, $U_w(\beta_0)$. By the Glivenko-Cantelli lemma, $\| \frac{\partial}{\partial\beta}\{\frac{1}{n}S_{w,n}(\beta)\} - \frac{\partial}{\partial\beta}\Gamma_w(\beta) \| \to_p 0$ uniformly in $\beta \in U_w(\beta_0)$. Also we have $n^{-1}S_{w,n}(\beta) \to_p \Gamma_w(\beta)$. As a result, for any $\epsilon > 0$, there exists sufficiently large n and $\nu > 0$, such that

$$S_{w,n}(\beta_0 - \epsilon) \times S_{w,n}(\beta_0 + \epsilon) < 0, \text{ and } \Pr\big\{\widehat{\beta}_{w,n} \in (\beta_0 - \epsilon, \beta_0 + \epsilon)\big\} > 1 - \nu,$$

which implies that $\widehat{\beta}_{w,n}$ is consistent.

It follows from the martingale representation of $S_{w,n}(\beta_0)$ and the martingale central limit theorem that $\frac{1}{\sqrt{n}}S_{w,n}(\beta_0)$ is asymptotically mean-zero normal with variance

$$V_w = \lim_{n\to\infty} < \frac{1}{\sqrt{n}}S_{w,n}(\beta_0), \frac{1}{\sqrt{n}}S_{w,n}(\beta_0) > (\tau)$$

$$= \lim_{n\to\infty} < \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\cdot}\xi_{w,i}(t;\beta_0)dM_i(t), \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\cdot}\xi_{w,i}(t;\beta_0)dM_i(t) > (\tau)$$

$$= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\xi_{w,i}^2(t;\beta_0)\,Y_i(t)\,d\Lambda_i(t)$$

$$= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\xi_{w,i}^2(t;\beta_0)\,Y_i(t)\{dR_0(t)\,B_i(t;\beta_0) + R_0(t)\,dB_i(t;\beta_0)\}.$$

Then by the Delta method and the consistency of $\widehat{\beta}_{w,n}$, it is seen that

$$\sqrt{n}(\widehat{\beta}_{w,n} - \beta_0) \longrightarrow_d \mathrm{N}(0, V_w/U_w^2),$$

where

$$\begin{aligned} U_w &= \lim_{n\to\infty} \frac{1}{n}\frac{\partial S_{w,n}(\beta_0)}{\partial \beta} \\ &= \lim_{n\to\infty} -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \Big[W_n(t)\frac{\partial \bar{Z}(t;\beta_0)}{\partial \beta}\{dN_i(t) - Y_i(t)\widehat{R}_n(t;\beta_0)\, dB_i(t;\beta_0)\} \\ &\qquad + W_n(t)Y_i(t)\{Z_i(t) - \bar{Z}(t;\beta_0)\}\{dB_i(t;\beta_0)\frac{\partial}{\partial \beta}\widehat{R}_n(t;\beta_0) + \widehat{R}_n(t;\beta_0)\frac{\partial}{\partial \beta}dB_i(t;\beta_0)\}\Big]. \end{aligned}$$

The consistency of the variance estimators follows from the consistency of $\widehat{\beta}_{w,n}$ and $\widehat{R}_n(t;\widehat{\beta}_{w,n})$, the continuous mapping theorem and the Slutsky's theorem. □

3.3.2 Choices of the Weight Function

One motivation of the weighted estimator is that $\widehat{\boldsymbol{\beta}}_{w,n}$ may gain substantial efficiency over $\widehat{\boldsymbol{\beta}}_n$ for appropriate choice of $\boldsymbol{W}_n(t)$. The optimal choice of $\boldsymbol{W}_n(t)$ results in the minimum asymptotic covariance of $\widehat{\boldsymbol{\beta}}_{w,n}$, which satisfies

$$\boldsymbol{w}(t)\, \boldsymbol{Z}(t) \propto \frac{\partial \log \lambda(t|\widetilde{\boldsymbol{Z}})}{\partial \boldsymbol{\beta}}\Big|_{\beta=\mathbf{0}}$$

[Bickel et al., 1993], where $\lambda(t|\widetilde{\boldsymbol{Z}})$ is the hazard function under model (3.1). $\frac{\partial}{\partial\beta}\log\lambda(t|\boldsymbol{\beta})$ is the score function for $\boldsymbol{\beta}$ expressed in terms of the hazard function $\lambda(t|\boldsymbol{\beta})$. This is a length-preserving transformation from the usual score function [Efron and Johnstone, 1990]. Bickel et al. [1993] has showed that the projection of the score function onto the class of unbiased linear estimating equations is the optimal one.

When covariates are all time-independent, we see that the unit weight is optimal and $\widehat{\boldsymbol{\beta}}_n$

reaches the semiparametric efficiency bound. Define $f'(t) = df(t)/dt$. We can write

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\beta}} \log \lambda(t|\widetilde{\boldsymbol{Z}}) \\
&= \frac{\partial}{\partial \boldsymbol{\beta}} \log[\exp\{\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}\}]' \\
&= \frac{\partial}{\partial \boldsymbol{\beta}} \log[\exp\{\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}\}\alpha'(t)] \\
&= \frac{\partial}{\partial \boldsymbol{\beta}}[\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z} + \log\{\alpha'(t)\}] \\
&= \boldsymbol{Z}.
\end{aligned}$$

Therefore,

$$\frac{\partial \log \lambda(t|\widetilde{\boldsymbol{Z}})}{\partial \boldsymbol{\beta}}\Big|_{\beta=\mathbf{0}} = \boldsymbol{Z} \Longrightarrow \boldsymbol{w}(t) = \mathbf{1}.$$

However, it is less straightforward to obtain the optimal weight in the presence of time-varying covariates. For example, in the case of a single time-varying covariate, the optimal weight is given by

$$w(t) = 1 - \frac{\log\{S_0(t)\} \times Z'(t)}{\lambda_0(t) \times Z(t)},$$

where $\lambda_0(t)$ is the baseline hazard function. It is seen that

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\beta}} \log \lambda(t|\widetilde{\boldsymbol{Z}}) \\
&= \frac{\partial}{\partial \boldsymbol{\beta}} \log[\exp\{\alpha(t) + \beta \times Z(t)\}]' \\
&= \frac{\partial}{\partial \boldsymbol{\beta}} \log[\exp\{\alpha(t) + \beta \times Z(t)\} \times \{\alpha'(t) + \beta \times Z'(t)\}] \\
&= \frac{\partial}{\partial \boldsymbol{\beta}}[\alpha(t) + \beta \times Z(t) + \log\{\alpha'(t) + \beta \times Z'(t)\}] \\
&= Z(t) + \frac{Z'(t)}{\alpha'(t) + \beta \times Z'(t)},
\end{aligned}$$

and then

$$
\begin{aligned}
&\frac{\partial \log \lambda(t|\widetilde{Z})}{\partial \beta}\Big|_{\beta=0} \\
&= Z(t) + \frac{Z'(t)}{\alpha'(t)} \\
&= Z(t) + \frac{Z'(t)}{[\log\{-\log S_0(t)\}]'} \\
&= Z(t) + \frac{-\log\{S_0(t)\} \times Z'(t)}{\lambda_0(t)}.
\end{aligned}
$$

The calculation of $w(t)$ involves the estimation of $\lambda_0(t)$. Some kernel estimates of $\lambda_0(t)$ may be used but are not stable with finite sample size [Tsiatis, 1990].

On the other hand, one may choose to assign more weight to specific observations. In particular, one can choose $\boldsymbol{W}_n(t)$ to be decreasing or increasing to emphasize the early or late observations. For example, a decreasing weight function can be the Kaplan-Meier estimates of the survival function.

3.4 Estimation of Survival Functions

3.4.1 Inference Procedures

Having obtained the estimators of the parameters in model (3.1), a natural next step is to study the inference procedures for the survival function. Under model (3.1), the survival function at a given time point t for an individual with covariate history $\widetilde{\boldsymbol{Z}}(t)$, say $\check{S}(t|\widetilde{\boldsymbol{Z}})$, is estimated by

$$
\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) = \exp[-\exp\{\widehat{\alpha}_n(t;\widehat{\boldsymbol{\beta}}_n) + \widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}(t)\}].
$$

$\widehat{\alpha}_n(t;\widehat{\boldsymbol{\beta}}_n)$ and $\widehat{\boldsymbol{\beta}}_n$ are consistent estimators as described in Section 3.2. Then the consistency of $\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})$ holds by virtue of the continuous mapping theorem and the Slutsky's theorem.

In the following, we show that $\sqrt{n}\big\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - \check{S}(t|\widetilde{\boldsymbol{Z}})\big\}$ converges weakly to a zero-mean normal distribution. Consider first the asymptotic distribution of $\log[-\log\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})\}]$. Apparently, the variation in $\log[-\log\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})\}]$ depends on the variability of both $\widehat{\alpha}_n(t;\widehat{\boldsymbol{\beta}}_n)$ and $\widehat{\boldsymbol{\beta}}_n$. The corresponding asymptotic distribution can be derived through the limiting joint distribution of $\widehat{\alpha}_n(t;\widehat{\boldsymbol{\beta}}_n)$ and $\widehat{\boldsymbol{\beta}}_n$. We present in Theorem 3.4.1 the asymptotic properties of $\log[-\log\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})\}]$.

Theorem 3.4.1. *Under the regularity conditions C1-C7, as $n \to \infty$,*

$$\sqrt{n}\big(\log[-\log\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})\}] - \log[-\log\{\check{S}(t|\widetilde{\boldsymbol{Z}})\}]\big) \longrightarrow_d \mathrm{N}\{0, \sigma^2(t)\},$$

$$\begin{aligned}
\text{where} \qquad \sigma^2(t) &= \frac{\sigma_1^2(t)}{R_0^2(t)} + 2\boldsymbol{\sigma}_{12}(t)^T\Big\{\frac{\boldsymbol{r}(t;\boldsymbol{\beta}_0)}{R_0^2(t)} + \frac{\boldsymbol{Z}(t)}{R_0(t)}\Big\} \\
&\quad + \Big\{\frac{\boldsymbol{r}(t;\boldsymbol{\beta}_0)}{R_0(t)} + \boldsymbol{Z}(t)\Big\}^T \boldsymbol{\sigma}_2^2 \Big\{\frac{\boldsymbol{r}(t;\boldsymbol{\beta}_0)}{R_0(t)} + \boldsymbol{Z}(t)\Big\}, \\
\sigma_1^2(t) &= \lim_{n\to\infty} \sum_{i=1}^n \int_0^t \Big[\frac{nD_n^2(u;\boldsymbol{\beta}_0)}{D_n^2(t;\boldsymbol{\beta}_0)P_n^2(u;\boldsymbol{\beta}_0)} Y_i(u)\{dR_0(u) \\
&\quad \times B_i(u;\boldsymbol{\beta}_0) + R_0(u)\, dB_i(u;\boldsymbol{\beta}_0)\}\Big], \\
\boldsymbol{\sigma}_{12}(t) &= \lim_{n\to\infty} \sum_{i=1}^n \int_0^t \Big[\frac{-D_n(u;\boldsymbol{\beta}_0)}{D_n(t;\boldsymbol{\beta}_0)P_n(u;\boldsymbol{\beta}_0)} \boldsymbol{U}^{-1}\boldsymbol{\xi}_i(u;\boldsymbol{\beta}_0) \\
&\quad \times Y_i(u)\{dR_0(u)\, B_i(u;\boldsymbol{\beta}_0) + R_0(u)\, dB_i(u;\boldsymbol{\beta}_0)\}\Big], \\
\boldsymbol{\sigma}_2^2 &= \boldsymbol{U}^{-1}\boldsymbol{V}(\boldsymbol{U}^{-1})^T, \\
\boldsymbol{r}(t;\boldsymbol{\beta}_0) &= \lim_{n\to\infty} \frac{\partial \widehat{R}_n(t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}.
\end{aligned}$$

Moreover, a consistent estimator for $\sigma^2(t)$ is given by $\widehat{\sigma}_n^2(t)$, where

$$\begin{aligned}
\widehat{\sigma}_n^2(t) &= \frac{\widehat{\sigma}_{1,n}^2(t)}{\widehat{R}_n^2(t;\widehat{\boldsymbol{\beta}}_n)} + 2\widehat{\boldsymbol{\sigma}}_{12,n}(t)^T\Big\{\frac{\widehat{\boldsymbol{r}}_n(t;\widehat{\boldsymbol{\beta}}_n)}{\widehat{R}_n^2(t;\widehat{\boldsymbol{\beta}}_n)} + \frac{\boldsymbol{Z}(t)}{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)}\Big\} \\
&\quad + \Big\{\frac{\widehat{\boldsymbol{r}}_n(t;\widehat{\boldsymbol{\beta}}_n)}{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)} + \boldsymbol{Z}(t)\Big\}^T \widehat{\boldsymbol{\sigma}}_{2,n}^2 \Big\{\frac{\widehat{\boldsymbol{r}}_n(t;\widehat{\boldsymbol{\beta}}_n)}{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)} + \boldsymbol{Z}(t)\Big\}, \\
\widehat{\sigma}_{1,n}^2(t) &= \sum_{i=1}^n \int_0^t \Big[\frac{nD_n^2(u;\widehat{\boldsymbol{\beta}}_n)\, Y_i(u)}{D_n^2(t;\widehat{\boldsymbol{\beta}}_n)\, P_n^2(u;\widehat{\boldsymbol{\beta}}_n)}\{d\widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n) \\
&\quad \times B_i(u;\widehat{\boldsymbol{\beta}}_n) + \widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n)\, dB_i(u;\widehat{\boldsymbol{\beta}}_n)\}\Big], \\
\widehat{\boldsymbol{\sigma}}_{12,n}(t) &= \sum_{i=1}^n \int_0^t \Big[\frac{-D_n(u;\widehat{\boldsymbol{\beta}}_n)}{D_n(t;\widehat{\boldsymbol{\beta}}_n)P_n(u;\widehat{\boldsymbol{\beta}}_n)} \widehat{\boldsymbol{U}}_n^{-1}\boldsymbol{\xi}_i(u;\widehat{\boldsymbol{\beta}}_n)Y_i(u) \\
&\quad \times\{d\widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n)\, B_i(u;\widehat{\boldsymbol{\beta}}_n) + \widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n)\, dB_i(u;\widehat{\boldsymbol{\beta}}_n)\}\Big], \\
\widehat{\boldsymbol{\sigma}}_{2,n}^2 &= \widehat{\boldsymbol{U}}_n^{-1}\widehat{\boldsymbol{V}}_n(\widehat{\boldsymbol{U}}_n^{-1})^T, \qquad \widehat{\boldsymbol{r}}_n(t;\widehat{\boldsymbol{\beta}}_n) = \frac{\partial \widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}}.
\end{aligned}$$

Proof of Theorem 3.4.1. By repeated use of the Taylor expansion, we can write

$$
\begin{aligned}
&\sqrt{n}\Big(\log[-\log\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})\}] - \log[-\log\{\check{S}(t|\widetilde{\boldsymbol{Z}})\}]\Big)\\
&= \sqrt{n}\Big[\log\{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)\} - \log\{R_0(t)\} + (\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T\boldsymbol{Z}(t)\Big]\\
&= \sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T\boldsymbol{Z}(t) + \frac{1}{R_0(t)}\sqrt{n}\{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n) - R_0(t)\} + o\{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n) - R_0(t)\}\\
&= \sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T\boldsymbol{Z}(t) + \frac{1}{R_0(t)}\sqrt{n}\{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n) - \widehat{R}_n(t;\boldsymbol{\beta}_0)\}\\
&\qquad + \frac{1}{R_0(t)}\sqrt{n}\{\widehat{R}_n(t;\boldsymbol{\beta}_0) - R_0(t)\} + o\{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n) - R_0(t)\}\\
&= \sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T\boldsymbol{Z}(t) + \frac{1}{R_0(t)}\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T\Big\{\frac{\partial\widehat{R}_n(t;\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\Big\} + o(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)\\
&\qquad + \frac{1}{R_0(t)}\sqrt{n}\{\widehat{R}_n(t;\boldsymbol{\beta}_0) - R_0(t)\} + o\{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n) - R_0(t)\}.\\
&= \sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T\Big\{\frac{1}{R_0(t)} \times \frac{\partial\widehat{R}_n(t;\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}} + \boldsymbol{Z}(t)\Big\} + o(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)\\
&\qquad + \frac{1}{R_0(t)}\sqrt{n}\{\widehat{R}_n(t;\boldsymbol{\beta}_0) - R_0(t)\} + o\{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n) - R_0(t)\}.
\end{aligned}
$$

In Section 3.2, we construct the martingale representations for $\widehat{R}_n(t;\boldsymbol{\beta}_0) - R_0(t)$ and $\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0$ respectively. Recall

$$
\begin{aligned}
\widehat{R}_n(t;\boldsymbol{\beta}_0) - R_0(t) &= \sum_{i=1}^{n}\int_0^t \Big\{\frac{D_n(u;\boldsymbol{\beta}_0)}{D_n(t;\boldsymbol{\beta}_0)\,P_n(u;\boldsymbol{\beta}_0)}dM_i(u)\Big\},\\
\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 &= \Big\{-\frac{\partial\boldsymbol{S}_n(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\Big\}^{-1}\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{\xi}_i(t;\boldsymbol{\beta}_0)dM_i(t).
\end{aligned}
$$

It then follows from the martingale central limit theorem that $\sqrt{n}\{\widehat{R}_n(t;\boldsymbol{\beta}_0) - R_0(t)\}$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ are asymptotically joint normal. In Theorem 3.2.1, we show that the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ is $\boldsymbol{\sigma}_2^2 = \boldsymbol{U}^{-1}\boldsymbol{V}(\boldsymbol{U}^{-1})^T$, where

$$
\begin{aligned}
\boldsymbol{V} &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{\xi}_i^{\otimes 2}(t;\boldsymbol{\beta}_0)\,Y_i(t)d\Lambda_i(t),\\
\boldsymbol{U} &= \lim_{n\to\infty}\frac{1}{n}\frac{\partial\boldsymbol{S}_n(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}.
\end{aligned}
$$

Using counting process arguments, we obtain the asymptotic variance of $\sqrt{n}\{\widehat{R}_n(t;\boldsymbol{\beta}_0) - R_0(t)\}$,

$$
\begin{aligned}
\sigma_1^2(t) &= \lim_{n\to\infty} < \sqrt{n}\{\widehat{R}_n(\cdot;\boldsymbol{\beta}_0) - R_0(\cdot)\}, \sqrt{n}\{\widehat{R}_n(\cdot;\boldsymbol{\beta}_0) - R_0(\cdot)\} > (t) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \int_0^t \frac{nD_n^2(u;\boldsymbol{\beta}_0)}{D_n^2(t;\boldsymbol{\beta}_0)P_n^2(u;\boldsymbol{\beta}_0)} Y_i(u)d\Lambda_i(u) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{D_n^2(u;\boldsymbol{\beta}_0)}{D_n^2(t;\boldsymbol{\beta}_0)\{\frac{1}{n}P_n(u;\boldsymbol{\beta}_0)\}^2} Y_i(u)d\Lambda_i(u) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{D_n^2(u;\boldsymbol{\beta}_0)}{D_n^2(t;\boldsymbol{\beta}_0)\{\frac{1}{n}P_n(u;\boldsymbol{\beta}_0)\}^2} Y_i(u)\{dR_0(u) \\
&\qquad \times B_i(u;\boldsymbol{\beta}_0) + R_0(u)\, dB_i(u;\boldsymbol{\beta}_0)\}.
\end{aligned}
$$

The asymptotic covariance between $\sqrt{n}\{\widehat{R}_n(t;\boldsymbol{\beta}_0) - R_0(t)\}$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ is given by

$$
\begin{aligned}
\boldsymbol{\sigma}_{12}(t) &= \lim_{n\to\infty} \sum_{i=1}^{n} \int_0^t \Big\{ \frac{D_n(u;\boldsymbol{\beta}_0)}{D_n(t;\boldsymbol{\beta}_0)\, P_n(u;\boldsymbol{\beta}_0)} \boldsymbol{U}^{-1}\boldsymbol{\xi}_i(u;\boldsymbol{\beta}_0)Y_i(u)d\Lambda_i(u) \Big\} \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_0^t \Big\{ \frac{D_n(u;\boldsymbol{\beta}_0)}{D_n(t;\boldsymbol{\beta}_0)\, \frac{1}{n}P_n(u;\boldsymbol{\beta}_0)} \boldsymbol{U}^{-1}\boldsymbol{\xi}_i(u;\boldsymbol{\beta}_0)Y_i(u)d\Lambda_i(u) \Big\} \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_0^t \Big[\frac{D_n(u;\boldsymbol{\beta}_0)}{D_n(t;\boldsymbol{\beta}_0)\, \frac{1}{n}P_n(u;\boldsymbol{\beta}_0)} \boldsymbol{U}^{-1}\boldsymbol{\xi}_i(u;\boldsymbol{\beta}_0)Y_i(u)\{dR_0(u) \\
&\qquad \times B_i(u;\boldsymbol{\beta}_0) + R_0(u)\, dB_i(u;\boldsymbol{\beta}_0)\} \Big],
\end{aligned}
$$

since

$$
\begin{aligned}
&\lim_{n\to\infty} < \sqrt{n}\{\widehat{R}_n(\cdot;\boldsymbol{\beta}_0) - R_0(\cdot)\}, \sum_{i=1}^{n} \int_0^{\cdot} -\frac{1}{\sqrt{n}}\boldsymbol{\xi}_i(u;\boldsymbol{\beta}_0)dM_i(u) > (t) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{-D_n(u;\boldsymbol{\beta}_0)}{D_n(t;\boldsymbol{\beta}_0)\, \frac{1}{n}P_n(u;\boldsymbol{\beta}_0)} \boldsymbol{\xi}_i(u;\boldsymbol{\beta}_0)Y_i(u)d\Lambda_i(u),
\end{aligned}
$$

and

$$
\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \Big\{ \frac{1}{n}\frac{\partial \boldsymbol{S}_n(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \Big\}^{-1} \sum_{i=1}^{n} \int_0^{\tau} -\frac{1}{\sqrt{n}}\boldsymbol{\xi}_i(t;\boldsymbol{\beta}_0)dM_i(t).
$$

Moreover, as $n \to \infty$, $o(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ and $o\{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n) - R_0(t)\}$ converge to zero respectively due to the consistency of $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)$. Therefore, we can show that $\sqrt{n}\big(\log[-\log\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})\}] -$

$\log[-\log\{\check{S}(t|\widetilde{\boldsymbol{Z}})\}]\big)$ converges weakly to a normal distribution with mean zero and variance,

$$\sigma^2(t) = \frac{\sigma_1^2(t)}{R_0^2(t)} + 2\boldsymbol{\sigma}_{12}(t)^T\Big\{\frac{\boldsymbol{r}(t;\boldsymbol{\beta}_0)}{R_0^2(t)} + \frac{\boldsymbol{Z}(t)}{R_0(t)}\Big\} + \Big\{\frac{\boldsymbol{r}(t;\boldsymbol{\beta}_0)}{R_0(t)} + \boldsymbol{Z}(t)\Big\}^T\boldsymbol{\sigma}_2^2\Big\{\frac{\boldsymbol{r}(t;\boldsymbol{\beta}_0)}{R_0(t)} + \boldsymbol{Z}(t)\Big\},$$

where

$$\begin{aligned}\boldsymbol{r}(t;\boldsymbol{\beta}_0) &= \lim_{n\to\infty}\frac{\partial\widehat{R}_n(t;\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}} \\ &= -\frac{\partial\, D(t;\beta_0)/\partial\beta}{D(t;\beta_0)}R(t;\beta_0) + \frac{1}{D(t;\beta_0)}\int_0^t\frac{\mathrm{E}\{dN(u)\}}{\mathrm{E}[Y(u)\exp\{\beta_0 Z(u)\}]} \\ &\quad\times\Big(\frac{\partial}{\partial\beta}D(u;\beta_0) - D(u;\beta_0)\frac{\mathrm{E}[Z(u)Y(u)\exp\{\beta_0 Z(u)\}]}{\mathrm{E}[Y(u)\exp\{\beta_0 Z(u)\}]}\Big).\end{aligned}$$

Consistency of the variance estimator $\widehat{\sigma}_n^2(t)$ follows from the consistency of $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)$, the continuous mapping theorem and the Slutsky's theorem. □

Then, an application of the Delta method leads to the asymptotic normality of $\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})$. The limiting variance of $\sqrt{n}\big\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - \check{S}(t|\widetilde{\boldsymbol{Z}})\big\}$ is of the form

$$\sigma_s^2(t) = \big[\check{S}(t|\widetilde{\boldsymbol{Z}})\log\{\check{S}(t|\widetilde{\boldsymbol{Z}})\}\big]^2\sigma^2(t).$$

We replace the unknown quantities with their sample estimators to obtain an estimator of $\sigma_s^2(t)$,

$$\widehat{\sigma}_{s,n}^2(t) = \big[\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})\log\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})\}\big]^2\widehat{\sigma}_n^2(t).$$

The following theorem summarizes the asymptotic results for $\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})$.

Theorem 3.4.2. *Under the regularity conditions C1-C7,* $\sqrt{n}\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - \check{S}(t|\widetilde{\boldsymbol{Z}})\}$ *converges weakly to a normal distribution with mean zero and variance* $\sigma_s^2(t)$ *as* $n\to\infty$*. Moreover, a consistent estimator for* $\sigma_s^2(t)$ *is given by* $\widehat{\sigma}_{s,n}^2(t)$*.*

Based on a normal approximation, a $100(1-\alpha)\%$ CI for the survival function at time t can be formed as

$$\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - z_{\alpha/2}\times\hat{\sigma}_{s,n}(t)/\sqrt{n},\ \widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) + z_{\alpha/2}\times\hat{\sigma}_{s,n}(t)/\sqrt{n},$$

where $z_{\alpha/2}$ is the upper $\frac{\alpha}{2}$ quantile of a standard normal random variable.

3.4.2 Hypothesis Testing

To test the hypothesis

$$H_0 : S(t|\widetilde{\boldsymbol{Z}}) = s_0 \qquad \text{versus} \qquad H_1 : S(t|\widetilde{\boldsymbol{Z}}) \neq s_0,$$

we can construct the Wald statistic

$$T_{s,n} = \frac{n(\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - s_0)^2}{\widehat{\sigma}^2_{s,n}(t)}.$$

Under the null hypothesis, $T_{s,n}$ converges to a chi-squared distribution with one degree of freedom, χ^2_1. Then a Wald test of size α is to reject $H_0 : S(t|\widetilde{\boldsymbol{Z}}) = s_0$ if $T_{s,n} > \chi^2_1(\alpha)$, where $\chi^2_1(\alpha)$ is the upper α quantile of the chi-squared distribution with one degree of freedom. Also we can compute the p value by $p = \Pr(\chi^2_1 > T_{s,n})$ and reject $H_0 : S(t|\widetilde{\boldsymbol{Z}}) = s_0$ if $p < \alpha$.

3.4.3 Monotonicity Issue

In practice, the estimated survival function $\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})$ may fail to be nonincreasing in t. We correct the monotonicity by replacing the problematic survival estimates with their smallest preceding values respectively [Hall and Wellner, 1980, Lin and Ying, 1994]. The modified estimator is defined as

$$\widehat{S}^*_n(t|\widetilde{\boldsymbol{Z}}) = \min_{s \leq t} \widehat{S}_n(s|\widetilde{\boldsymbol{Z}}).$$

In the following, we present a heuristic argument for the asymptotic equivalence between $\widehat{S}^*_n(t|\widetilde{\boldsymbol{Z}})$ and the original estimator $\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})$. First, by the Taylor expansion, it is seen that, for any $\delta > 0$,

$$\begin{aligned} & S(t|\widetilde{\boldsymbol{Z}}) - S(t+\delta|\widetilde{\boldsymbol{Z}}) \\ &= \{1 - S(t+\delta|\widetilde{\boldsymbol{Z}})\} - \{1 - S(t|\widetilde{\boldsymbol{Z}})\} \\ &= F(t+\delta|\widetilde{\boldsymbol{Z}}) - F(t|\widetilde{\boldsymbol{Z}}) \\ &= f(t^*|\widetilde{\boldsymbol{Z}}) \times \delta, \end{aligned}$$

where S is the survival function of the failure time T, F is the cumulative distribution function of T, f is the probability density function of T and $t^* \in (t, t+\delta)$. Then, we can write

$$\begin{aligned} &\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - \widehat{S}_n(t+\delta|\widetilde{\boldsymbol{Z}}) \\ = &\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - \widehat{S}_n(t+\delta|\widetilde{\boldsymbol{Z}}) + f(t^*|\widetilde{\boldsymbol{Z}}) \times \delta - S(t|\widetilde{\boldsymbol{Z}}) + S(t+\delta|\widetilde{\boldsymbol{Z}}) \\ = &\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - S(t|\widetilde{\boldsymbol{Z}})\} - \{\widehat{S}_n(t+\delta|\widetilde{\boldsymbol{Z}}) - S(t+\delta|\widetilde{\boldsymbol{Z}})\} + f(t^*|\widetilde{\boldsymbol{Z}}) \times \delta. \end{aligned}$$

For $n^{-2/3} \le \delta \le n^{-1/3}$ and some $\epsilon > 0$, we have

$$\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - \widehat{S}_n(t+\delta|\widetilde{\boldsymbol{Z}}) = f(t^*|\widetilde{\boldsymbol{Z}}) \times \delta + o_p(\delta^{1+\epsilon}) \ge 0.$$

For $\delta \ge n^{-1/3}$, we have

$$\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - \widehat{S}_n(t+\delta|\widetilde{\boldsymbol{Z}}) \ge f(t^*|\widetilde{\boldsymbol{Z}}) \times \delta + O_p(n^{-1/2}) > 0.$$

Therefore, for large n, we have

$$\widehat{S}_n(t+n^{-2/3}|\widetilde{\boldsymbol{Z}}) \le \widehat{S}^*_n(t|\widetilde{\boldsymbol{Z}}) \le \widehat{S}_n(t|\widetilde{\boldsymbol{Z}}).$$

Also, it is seen that

$$\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - \widehat{S}_n(t+n^{-2/3}|\widetilde{\boldsymbol{Z}}) = o_p(n^{-1/2}).$$

It then follows that

$$\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - \widehat{S}^*_n(t|\widetilde{\boldsymbol{Z}}) = o_p(n^{-1/2}).$$

As a result, the limiting distribution of $n^{1/2}\{\widehat{S}^*_n(t|\widetilde{\boldsymbol{Z}}) - S(t|\widetilde{\boldsymbol{Z}})\}$ and that of $n^{1/2}\{\widehat{S}_n(t|\widetilde{\boldsymbol{Z}}) - S(t|\widetilde{\boldsymbol{Z}})\}$ are the same.

3.4.4 Time-Independent Covariates

When covariates are all time-invariant, model 3.1 reduces to the Cox proportional hazards model. In this section, we examine the relationship between the proposed inference procedures for the survival function in the case of time-independent covariates and the estimation method of survival function using the Cox proportional hazards model.

Let $\widehat{S}_{cox}(t|\boldsymbol{Z})$ denote the estimator of the survival function given time-independent covariates $\boldsymbol{Z}$ under the Cox proportional hazards model. Tsiatis [1981] formulated an estimator for the asymptotic variance of $\widehat{S}_{cox}(t|\boldsymbol{Z})$, that is,

$$\widehat{var}\{\widehat{S}_{cox}(t|\boldsymbol{Z})\} = \{\widehat{S}_{cox}(t|\boldsymbol{Z})\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z})\}^2\Big[\sum_{i=1}^{l}\frac{1}{\big\{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_j)\big\}^2} + \boldsymbol{a}^T\widehat{var}(\widehat{\boldsymbol{\beta}})\boldsymbol{a}\Big],$$

where

$$\begin{aligned}
\boldsymbol{a} &= \sum_{i=1}^{l}\frac{\sum_{j\in R_i}\boldsymbol{Z}_j\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_j)}{\big\{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_j)\big\}^2} - \boldsymbol{Z}\frac{1}{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_j)},\\
\widehat{S}_{cox}(t|\boldsymbol{Z}) &= \exp\{-\exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z})\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})\},\\
\widehat{\Lambda}_0(t,\boldsymbol{\beta}) &= \int_0^t\frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)},
\end{aligned}$$

$\widehat{\boldsymbol{\beta}}$ is the Cox maximum partial likelihood estimator, $\widehat{var}(\widehat{\boldsymbol{\beta}})$ is the estimate of the asymptotic variance of $\widehat{\boldsymbol{\beta}}$, R_i is the risk set of subjects with observed survival times greater than or equal to the ith ordered failure time t_i, and l is such that $t_l < t, t_{l+1} \geq t$.

In the following, we show our estimator for the asymptotic variance of the survival estimate, $\hat{\sigma}^2_{s,n}(t)/n$, to be equivalent to that proposed by Tsiatis when covariates are all time-invariant. That is, $\hat{\sigma}^2_{s,n}(t)/n$ and $\widehat{var}\{\widehat{S}_{cox}(t|\boldsymbol{Z})\}$ are algebraically identical. First, note that, in the case of time-independent covariates,

$$\begin{aligned}
dB_i(t;\boldsymbol{\beta}) &= d\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i) = 0,\\
Q_n(t;\boldsymbol{\beta}) &= \sum_{i=1}^{n}Y_i(t)\,d\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i) = 0,
\end{aligned}$$

$$
\begin{aligned}
D_n(t;\boldsymbol{\beta}) &= \exp\Big\{\int_0^t \frac{Q_n(u;\boldsymbol{\beta})}{P_n(u;\boldsymbol{\beta})}\Big\} = \exp\Big\{\int_0^t \frac{Q_n(u;\boldsymbol{\beta})}{\sum_{i=1}^n Y_i(u)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)}\Big\} = 1,\\
\widehat{R}_n(t;\boldsymbol{\beta}) &= \frac{1}{D_n(t;\boldsymbol{\beta})}\int_0^t \Big\{D_n(u;\boldsymbol{\beta})\frac{\sum_{i=1}^n dN_i(u)}{P_n(u;\boldsymbol{\beta})}\Big\} = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{P_n(u;\boldsymbol{\beta})} = \widehat{\Lambda}_0(t,\boldsymbol{\beta}),\\
d\widehat{R}_n(t;\boldsymbol{\beta}) &= \frac{\sum_{i=1}^n dN_i(t)}{P_n(t;\boldsymbol{\beta})} = \frac{\sum_{i=1}^n dN_i(t)}{\sum_{i=1}^n Y_i(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)},\\
\boldsymbol{\xi}_i(t;\boldsymbol{\beta}) &= \boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t;\boldsymbol{\beta}) - \frac{D_n(t;\boldsymbol{\beta})}{P_n(t;\boldsymbol{\beta})}\sum_{k=1}^n \int_t^\tau \Big[\{\boldsymbol{Z}_k - \overline{\boldsymbol{Z}}(u;\boldsymbol{\beta})\}\frac{Y_k(u)\,dB_k(u;\boldsymbol{\beta})}{D_n(u;\boldsymbol{\beta})}\Big]\\
&= \boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t;\boldsymbol{\beta}) = \boldsymbol{Z}_i - \frac{\sum_{i=1}^n Y_i(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\boldsymbol{Z}_i}{\sum_{i=1}^n Y_i(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)},\\
N(t) &= \mathrm{I}(X \le t, \Delta = 1), \qquad Y(t) = \mathrm{I}(X \ge t).
\end{aligned}
$$

Then it is apparent that the proposed estimating function for $\boldsymbol{\beta}$ is identical to the Cox partial likelihood score function:

$$
\begin{aligned}
&\sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i[dN_i(t) - Y_i(t)\,d\exp\{\widehat{\alpha}_n(t;\boldsymbol{\beta}) + \boldsymbol{\beta}^T\boldsymbol{Z}_i\}]\\
&= \sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i[dN_i(t) - Y_i(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\,d\exp\{\widehat{\alpha}_n(t;\boldsymbol{\beta})\}]\\
&= \sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i[dN_i(t) - Y_i(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\,d\widehat{R}_n(t;\boldsymbol{\beta})]\\
&= \sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i\Big[dN_i(t) - Y_i(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\,\frac{\sum_{i=1}^n dN_i(t)}{\sum_{i=1}^n Y_i(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)}\Big]\\
&= \sum_{i=1}^k \Big\{\boldsymbol{Z}_{(i)} - \frac{\sum_{j\in R_i}\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j}{\sum_{j\in R_i}\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)}\Big\}.
\end{aligned}
$$

Denote $\boldsymbol{Z}_{(i)}$ the vector of covariates for the subject who fails at the ith ordered failure time t_i, $i = 1, 2, \ldots, k$. Thus our estimator for $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}_n$, is in the same form as the Cox maximum partial likelihood estimator. Therefore,

$$
\begin{aligned}
\widehat{S}_n(t|\boldsymbol{Z}) &= \exp\{-\exp(\widehat{\boldsymbol{\beta}}_n^T\boldsymbol{Z})\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)\}\\
&= \widehat{S}_{cox}(t|\boldsymbol{Z}).
\end{aligned}
$$

Furthermore, we can write

$$
\begin{aligned}
\widehat{\sigma}^2_{1,n}(t) &= \sum_{i=1}^{n} \int_0^t \Big[\frac{n D_n^2(u;\widehat{\boldsymbol{\beta}}_n)\, Y_i(u)}{D_n^2(t;\widehat{\boldsymbol{\beta}}_n)\, P_n^2(u;\widehat{\boldsymbol{\beta}}_n)} \{ d\widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n) B_i(u;\widehat{\boldsymbol{\beta}}_n) \\
&\qquad + \widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n)\, dB_i(u;\widehat{\boldsymbol{\beta}}_n)\} \Big] \\
&= \int_0^t \frac{n\, d\widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n)}{P_n^2(u;\widehat{\boldsymbol{\beta}}_n)} \Big\{ \sum_{i=1}^{n} Y_i(u) \exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_i) \Big\} \\
&= \int_0^t \frac{n}{P_n^2(u;\widehat{\boldsymbol{\beta}}_n)} P_n(u;\widehat{\boldsymbol{\beta}}_n) \frac{\sum_{i=1}^n dN_i(u)}{P_n(u;\widehat{\boldsymbol{\beta}}_n)} \\
&= \int_0^t \frac{n \sum_{i=1}^n dN_i(u)}{\Big\{ \sum_{i=1}^n Y_i(u) \exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_i) \Big\}^2} \\
&= \sum_{i=1}^{l} \frac{n}{\big\{ \sum_{j \in R_i} \exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j) \big\}^2},
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\boldsymbol{\sigma}}_{12,n}(t) &= \sum_{i=1}^{n} \int_0^t \Big[\frac{-D_n(u;\widehat{\boldsymbol{\beta}}_n)}{D_n(t;\widehat{\boldsymbol{\beta}}_n) P_n(u;\widehat{\boldsymbol{\beta}}_n)} \widehat{\boldsymbol{U}}_n^{-1} \boldsymbol{\xi}_i(u;\widehat{\boldsymbol{\beta}}_n) Y_i(u) \\
&\qquad \times \{ d\widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n)\, B_i(u;\widehat{\boldsymbol{\beta}}_n) + \widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n)\, dB_i(u;\widehat{\boldsymbol{\beta}}_n)\} \Big], \\
&= \int_0^t \Big[\frac{-d\widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n)}{P_n(u;\widehat{\boldsymbol{\beta}}_n)} \widehat{\boldsymbol{U}}_n^{-1} \Big\{ \sum_{i=1}^{n} \boldsymbol{\xi}_i(u;\widehat{\boldsymbol{\beta}}_n) Y_i(u) B_i(u;\widehat{\boldsymbol{\beta}}_n) \Big\} \Big] \\
&= \int_0^t \frac{-d\widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n)}{P_n(u;\widehat{\boldsymbol{\beta}}_n)} \widehat{\boldsymbol{U}}_n^{-1} \Big[\sum_{i=1}^{n} \{ \boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(u;\widehat{\boldsymbol{\beta}}_n) \} Y_i(u) B_i(u;\widehat{\boldsymbol{\beta}}_n) \Big] \\
&= \int_0^t \frac{-d\widehat{R}_n(u;\widehat{\boldsymbol{\beta}}_n)}{P_n(u;\widehat{\boldsymbol{\beta}}_n)} \widehat{\boldsymbol{U}}_n^{-1} \Big[\sum_{i=1}^{n} Y_i(u) \exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_i) \boldsymbol{Z}_i \\
&\qquad - \frac{\sum_{i=1}^n Y_i(u) \exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_i) \boldsymbol{Z}_i}{\sum_{i=1}^n Y_i(u) \exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_i)} \sum_{i=1}^{n} Y_i(u) \exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_i) \Big] \\
&= \boldsymbol{0},
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\boldsymbol{\sigma}}^2_{2,n} &= \widehat{\boldsymbol{U}}_n^{-1} \widehat{\boldsymbol{V}}_n (\widehat{\boldsymbol{U}}_n^{-1})^T \\
&= n\, \widehat{var}(\widehat{\boldsymbol{\beta}}_n),
\end{aligned}
$$

$$
\begin{aligned}
&\widehat{\boldsymbol{r}}_n(t;\widehat{\boldsymbol{\beta}}_n) + \widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)\,\boldsymbol{Z}\\
&= \frac{\partial}{\partial \boldsymbol{\beta}} \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{P_n(u;\widehat{\boldsymbol{\beta}}_n)} + \boldsymbol{Z} \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{P_n(u;\widehat{\boldsymbol{\beta}}_n)}\\
&= \int_0^t \frac{-\{\sum_{i=1}^n dN_i(u)\}\{\sum_{i=1}^n Y_i(u)\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_i)\boldsymbol{Z}_i\}}{\{\sum_{i=1}^n Y_i(u)\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_i)\}^2}\\
&\quad + \boldsymbol{Z} \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_i)}\\
&= -\sum_{i=1}^{l} \frac{\sum_{j\in R_i} \boldsymbol{Z}_j \exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j)}{\big\{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j)\big\}^2} + \boldsymbol{Z}\frac{1}{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j)}.
\end{aligned}
$$

As a result, we can represent $\widehat{\sigma}^2_{s,n}(t)/n$ as

$$
\begin{aligned}
\frac{\widehat{\sigma}^2_{s,n}(t)}{n} &= \frac{1}{n}\big[\widehat{S}_n(t|\boldsymbol{Z})\log\{\widehat{S}_n(t|\boldsymbol{Z})\}\big]^2 \widehat{\sigma}^2_n(t)\\
&= \frac{1}{n}\big\{\widehat{S}_n(t|\boldsymbol{Z})\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z})\big\}^2 \Bigg[\frac{\widehat{\sigma}^2_{1,n}(t)}{\widehat{R}^2_n(t;\widehat{\boldsymbol{\beta}}_n)} + 2\widehat{\boldsymbol{\sigma}}_{12,n}(t)^T\Bigg\{\frac{\widehat{\boldsymbol{r}}_n(t;\widehat{\boldsymbol{\beta}}_n)}{\widehat{R}^2_n(t;\widehat{\boldsymbol{\beta}}_n)}\\
&\quad + \frac{\boldsymbol{Z}}{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)}\Bigg\} + \Bigg\{\frac{\widehat{\boldsymbol{r}}_n(t;\widehat{\boldsymbol{\beta}}_n)}{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)} + \boldsymbol{Z}\Bigg\}^T \widehat{\boldsymbol{\sigma}}^2_{2,n}\Bigg\{\frac{\widehat{\boldsymbol{r}}_n(t;\widehat{\boldsymbol{\beta}}_n)}{\widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)} + \boldsymbol{Z}\Bigg\}\Bigg]\\
&= \frac{1}{n}\big\{\widehat{S}_n(t|\boldsymbol{Z})\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z})\big\}^2 \Bigg[\widehat{\sigma}^2_{1,n}(t) + 0 + \Big\{\widehat{\boldsymbol{r}}_n(t;\widehat{\boldsymbol{\beta}}_n)\\
&\quad + \widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)\boldsymbol{Z}\Big\}^T \widehat{\boldsymbol{\sigma}}^2_{2,n}\Big\{\widehat{\boldsymbol{r}}_n(t;\widehat{\boldsymbol{\beta}}_n) + \widehat{R}_n(t;\widehat{\boldsymbol{\beta}}_n)\boldsymbol{Z}\Big\}\Bigg]\\
&= \big\{\widehat{S}_n(t|\boldsymbol{Z})\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z})\big\}^2 \Bigg(\sum_{i=1}^{l} \frac{1}{\big\{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j)\big\}^2} + \Big[\sum_{i=1}^{l} \frac{\sum_{j\in R_i}\boldsymbol{Z}_j\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j)}{\big\{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j)\big\}^2}\\
&\quad - \boldsymbol{Z}\frac{1}{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j)}\Big]^T \widehat{var}(\widehat{\boldsymbol{\beta}}_n)\Big[\sum_{i=1}^{l} \frac{\sum_{j\in R_i}\boldsymbol{Z}_j\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j)}{\big\{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j)\big\}^2}\\
&\quad - \boldsymbol{Z}\frac{1}{\sum_{j\in R_i}\exp(\widehat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}_j)}\Big]\Bigg)\\
&= \widehat{var}\{\widehat{S}_{cox}(t|\boldsymbol{Z})\}.
\end{aligned}
$$

Chapter 4

SIMULATIONS

Simulation studies are carried out to assess the finite sample performance of the proposed methods. We consider simulation scenarios that differ in sample size, baseline survival distribution, censoring proportion, censoring distribution, value of regression parameters, number of regression parameters and distribution of covariate.

4.1 Time-Independent Covariates

We generate failure times from the following model:

$$\log[-\log\{S(t|Z)\}] = \log\{(0.2 \times t)^2\} + \beta \times Z, \tag{4.1}$$

where the baseline follows the Weibull distribution with shape parameter 2 and scale parameter 5, and Z follows the Unif(0,2) distribution. We simulate the censoring times from the Unif(0.5,c) distribution. The values of c are chosen such that the censoring rate is around 30%. The value of regression parameter β is set at 0.1, 0.5, 1, 2, or 3 and the sample size n varies from 100 to 1500. For each simulated data set, we calculate a Cox maximum partial likelihood estimate for β and an unweighted estimate following the proposed procedures. Table 4.1 summaries the simulation results for the estimators of β. Each entry of Table 4.1 is based on 1000 replications.

As expected, Table 4.1 shows the equivalence between our estimates and the Cox maximum partial likelihood estimates in the case of time-independent covariates. The simulation results for the two estimators are exactly the same. The estimating equation estimator has high efficiency when true beta value is small [Lin and Ying, 1994]. As shown in Table 4.1, the bias of beta estimates increases with the true beta value. When true beta value equals to

0.1, estimates are unbiased for all cases; when true beta value equals to 0.5 or 1, the biases of estimates are less than 0.005 for $n \geq 300$; when true beta value equals to 2, the biases are less than 0.005 for $n \geq 400$. When true beta value equals to 3, the bias reduces to -0.003 only when n is as large as 1500.

4.2 Time-Varying Covariates

We generate failure times from the following model:

$$\log[-\log\{S(t|Z)\}] = \log(0.5 \times t) + \beta \times Z \times \log(t), \tag{4.2}$$

where the baseline is exponential with rate parameter equal to 0.5. The failure times are censored at a fixed time point c and are subject to about 30% censoring. We simulate Z from the Unif(0,2) distribution. We consider various choices of regression parameter β and sample size n: $\beta = 0, 0.1, 0.2, 0.5$ and $n = 100, 200, 300, 500, 1000, 1500$. For each simulated data set, we calculate an unweighted estimate and a weighted estimate for β following the proposed procedures. The optimal weight function, $1 + 1/\log(t)$, is used in the estimation. As described in Section 3.3.2, in the case of a single time-varying covariate, the optimal weight is given by

$$w(t) = 1 - \frac{\log\{S_0(t)\} \times Z'(t)}{\lambda_0(t) \times Z(t)}.$$

Then under model (4.2), we can write the optimal weight as

$$w(t) = 1 + \frac{0.5 \times t \times \frac{Z}{t}}{0.5 \times Z \times \log(t)} = 1 + \frac{1}{\log(t)}.$$

Table 4.2 summaries the simulation results for the estimators of β. Each entry of Table 4.2 is based on 1000 replications. The unweighted estimates are unbiased in most cases. Some bias is present when $\beta = 0.5$, but the bias diminishes rapidly with sample size. From Table 4.2, the variance estimates for the unweighted estimator are in general agreement with the empirical variances and the empirical coverage probabilities of 95% confidence intervals (CIs) are close to the nominal level. Table 4.2 also suggests that the asymptotic approximations for the weighted estimator work well for practical sample sizes. The weighted

Table 4.1: Simulation results for the unweighted estimator $\hat{\beta}_n$ and the Cox maximum partial likelihood estimator $\hat{\beta}_{cox}$ under model (4.1).

		$\hat{\beta}_n$				$\hat{\beta}_{cox}$			
β	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
0.1	100	−0.002	0.213	0.208	0.944	−0.002	0.213	0.208	0.944
	200	0.001	0.142	0.144	0.959	0.001	0.142	0.144	0.959
	300	0.002	0.119	0.117	0.945	0.002	0.119	0.117	0.945
	400	−0.002	0.101	0.101	0.946	−0.002	0.101	0.101	0.946
	500	0.001	0.089	0.091	0.949	0.001	0.089	0.091	0.949
0.5	100	0.018	0.219	0.213	0.948	0.018	0.219	0.213	0.948
	200	0.013	0.156	0.148	0.939	0.013	0.156	0.148	0.939
	300	0.004	0.117	0.120	0.960	0.004	0.117	0.120	0.960
	400	−0.001	0.105	0.104	0.950	−0.001	0.105	0.104	0.950
	500	−0.001	0.090	0.093	0.953	−0.001	0.090	0.093	0.953
1	100	0.019	0.237	0.230	0.945	0.019	0.237	0.230	0.945
	200	0.007	0.161	0.159	0.947	0.007	0.161	0.159	0.947
	300	0.004	0.135	0.129	0.942	0.004	0.135	0.129	0.942
	400	0.005	0.114	0.111	0.944	0.005	0.114	0.111	0.944
	500	0.002	0.098	0.100	0.957	0.002	0.098	0.100	0.957
2	100	0.028	0.286	0.280	0.952	0.028	0.286	0.280	0.952
	200	0.011	0.193	0.194	0.951	0.011	0.193	0.194	0.951
	300	0.013	0.162	0.158	0.948	0.013	0.162	0.158	0.948
	400	0.003	0.136	0.136	0.953	0.003	0.136	0.136	0.953
	500	0.004	0.124	0.121	0.947	0.004	0.124	0.121	0.947
3	300	0.016	0.197	0.195	0.947	0.016	0.197	0.195	0.947
	400	0.011	0.166	0.168	0.961	0.011	0.166	0.168	0.961
	500	0.012	0.150	0.150	0.955	0.012	0.150	0.150	0.955
	1000	0.006	0.103	0.106	0.954	0.006	0.103	0.106	0.954
	1500	−0.003	0.083	0.086	0.958	−0.003	0.083	0.086	0.958

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval estimates appear to be more efficient than the unweighted estimates. The efficiencies of the unweighted estimates relative to the weighted ones are around 45%, regardless of true parameter value and sample size.

In addition, we evaluate the methods of Section 3.4 under a range of simulation scenarios. Failure times are generated from model (4.2) with Z taking values 0 or 1. The probability that Z equals 1, say $p1$, is set to be 0.2, 0.5 or 0.8. We choose $\beta = 0.1, 0.5, 1$. The censoring times are distributed uniformly on $(1, c)$. The values of c are chosen such that about twenty percent of the failure times are censored. The simulations are based on 1000 replications for $n = 200, 300$ or 500. Table 4.3-4.5 give the simulation results for the estimators of the survival probabilities at three time points $t = 0.5, 1.2, 1.8$ for $Z = 1$. The corresponding true survival probabilities are 0.792, 0.543, 0.385 when $\beta = 0.1$, 0.838, 0.518, 0.299 when $\beta = 0.5$, and 0.882, 0.487, 0.198 when $\beta = 1$. The true survival curves under model (4.2), constrained to the interval (0,10), are shown in Figure 4.1 for a range of values of β. From Figure 4.1, we can see a clear pattern of dramatic decline in survival probabilities at the earlier follow-up time, followed by much flatter curves. Table 4.6-4.8 give the simulation results for the estimators of the survival probabilities at $t = 0.5, 1.2, 1.8$ for $Z = 0$. When $Z = 0$, the survival estimates are the estimates for the baseline survival probabilities and the survival curve is independent of the value of β. Under model (4.2), the true baseline survival probabilities at $t = 0.5, 1.2, 1.8$ are 0.779, 0.549 and 0.407, respectively.

The survival estimators perform well under all the simulation settings examined. The estimates are virtually unbiased, the variance estimates agree well with the empirical variances, and the 95% CIs maintain their coverage probabilities near the nominal level. As shown in Table 4.3-4.8, the choices of $p1, n$ and t influence the size of the standard error (SE). We see a trend of SE increase at the later time, which can be explained by the decrease in the number of individuals at risk over time. As expected, SE decreases with sample size. For $Z = 1$ and given n, t and β, the larger value of $p1$ generally results in smaller SE. There are differences between $Z = 0$ and $Z = 1$ for the effect of $p1$ on SE, with the larger $p1$ yielding bigger SE in the simulation scenarios for $Z = 0$. This is not surprising because the estimation accuracy in

Table 4.2: Simulation results for the unweighted estimator $\hat{\beta}_n$ and the weighted estimator $\hat{\beta}_{w,n}$ with weight function $1 + 1/\log(t)$ under model (4.2).

		$\hat{\beta}_n$				$\hat{\beta}_{w,n}$				
β	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	RE.
0	100	−0.007	0.252	0.249	0.945	−0.006	0.173	0.168	0.941	47.1%
	200	0.002	0.178	0.174	0.955	<0.001	0.121	0.118	0.955	46.2%
	300	−0.002	0.144	0.142	0.942	<0.001	0.095	0.096	0.955	43.5%
	500	−0.002	0.112	0.109	0.944	<0.001	0.075	0.074	0.943	44.8%
	1000	−0.001	0.077	0.077	0.953	−0.002	0.052	0.052	0.951	45.6%
	1500	−0.002	0.061	0.063	0.946	<0.001	0.042	0.042	0.948	47.4%
0.1	100	0.007	0.297	0.282	0.955	0.005	0.189	0.188	0.961	40.5%
	200	0.002	0.202	0.195	0.948	0.003	0.132	0.131	0.950	42.7%
	300	0.003	0.160	0.158	0.944	0.001	0.108	0.106	0.946	45.6%
	500	0.002	0.123	0.121	0.950	0.001	0.083	0.081	0.943	45.5%
	1000	0.001	0.087	0.086	0.953	0.001	0.058	0.058	0.952	44.4%
	1500	−0.001	0.070	0.070	0.956	−0.001	0.047	0.047	0.949	45.1%
0.2	100	0.010	0.307	0.309	0.950	0.006	0.206	0.205	0.949	45.0%
	200	0.006	0.216	0.214	0.947	0.004	0.142	0.143	0.951	43.2%
	300	0.006	0.176	0.173	0.946	0.002	0.118	0.116	0.942	45.0%
	500	0.004	0.135	0.133	0.947	0.002	0.091	0.089	0.946	45.4%
	1000	0.003	0.095	0.094	0.956	0.002	0.064	0.063	0.950	45.4%
	1500	−0.001	0.076	0.076	0.955	−0.001	0.052	0.051	0.952	46.8%
0.5	100	0.023	0.406	0.391	0.950	0.015	0.279	0.263	0.942	47.2%
	200	0.019	0.270	0.268	0.947	0.007	0.178	0.184	0.957	43.5%
	300	0.016	0.220	0.217	0.950	0.007	0.151	0.150	0.951	47.1%
	500	0.013	0.166	0.167	0.954	0.004	0.119	0.116	0.941	51.4%
	1000	0.005	0.117	0.117	0.950	−0.003	0.082	0.082	0.952	49.1%
	1500	0.002	0.094	0.095	0.956	<0.001	0.067	0.067	0.949	50.8%

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; RE., relative efficiency of $\hat{\beta}_n$ vs. $\hat{\beta}_{w,n}$

survival estimates for $Z = k$, $k = 0, 1$, improves as the observed number of $Z = k$ increases.

Moreover, we conduct simulations with Weibull baseline survival. Failure times are simulated from the following model:

$$\log[-\log\{S(t|Z)\}] = \log\{(0.2 \times t)^2\} + \beta \times Z \times \log(t), \tag{4.3}$$

where the baseline is distributed Weibull with shape parameter 2 and scale parameter 5. We let Z be the uniform variable on (0,2). About thirty percent of the simulated failure times are censored and the censoring time follows the Unif(0.5,c) distribution. We set $\beta = 0.001, 0.1, 0.2, 0.5$ for $n = 100, 200, 300, 500, 1000$. Simulation results for the unweighted and optimal weighted estimates based on 1000 replications are reported in Table 4.9. Specifically, the weight function is given by

$$w(t) = 1 + \frac{\{(0.2 \times t)^2\} \times \frac{Z}{t}}{0.2^2 \times 2 \times t \times Z \times \log(t)} = 1 + \frac{0.5}{\log(t)}.$$

The results exhibited in Table 4.9 show that the performance of unweighted and weighted estimators is good in all simulation scenarios for model (4.3). The proposed inference procedures yield virtually unbiased estimates and proper coverage probabilities. There seems to be little advantage to the optimal weighted estimates over the unweighted estimates. The efficiencies of the unweighted estimates relative to the weighted estiamtes are all above 94%. By comparing Table 4.9 with Table 4.2, we see much improvement in the estimation accuracy when data come from the model with Weibull baseline. Either the bias or the SE in Table 4.9 is smaller than that in Table 4.2. The weight function is decreasing with time. For exponential baseline, most events occur at the earlier follow-up time, which results in larger values of weight function, and thus the efficiency of the estimators gets greatly improved.

We also simulate failure times from model (4.3) with Z taking values 0 or 1 to evaluate the proposed survival estimator. The simulated failure times are subject to about twenty percent uniform censoring. We employ the same setting as in Table 4.3-4.5: $\beta = 0.1, 0.5, 1$, $p1 = 0.2, 0.5, 0.8$, $n = 200, 300, 500$. Figure 4.2 provides plots of true survival probabilities under model (4.3) for subgroups with $Z = 0$ and $Z = 1$, respectively. In contrast to Figure 4.1, the

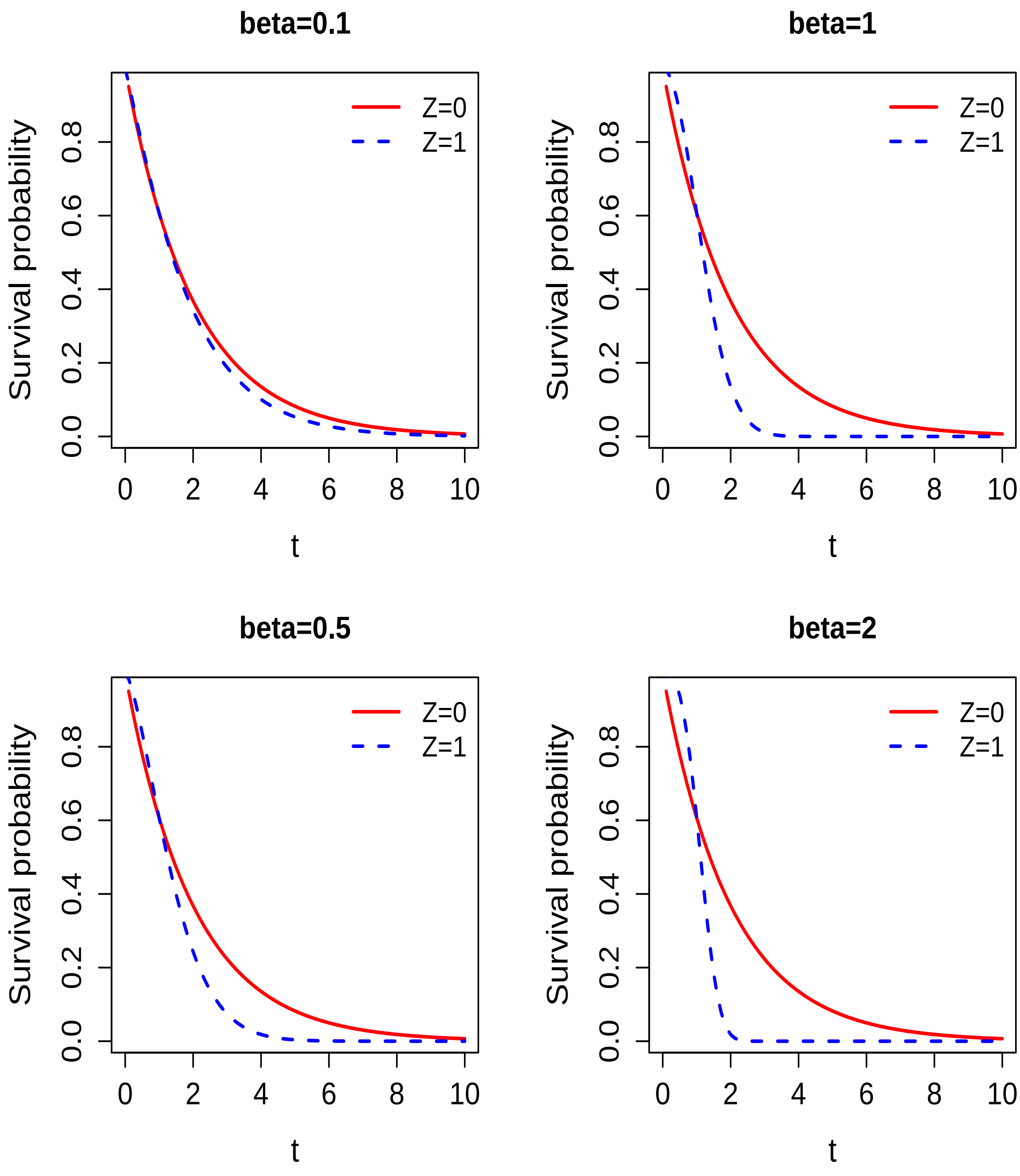

Figure 4.1: True survival curves under model (4.2) for subgroups with $Z = 1$ and $Z = 0$ respectively.

Table 4.3: Simulation results for the estimators of survival probabilities at three time points $t = 0.5, 1.2, 1.8$ under model (4.2) for $Z = 1$ and $\beta = 0.1$.

			$\boldsymbol{t} = 0.5$				$\boldsymbol{t} = 1.2$				$\boldsymbol{t} = 1.8$			
β	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
0.1	0.2	100	0.002	0.050	0.050	0.929	0.003	0.050	0.052	0.956	0.005	0.071	0.070	0.934
		200	<0.001	0.035	0.035	0.952	0.002	0.035	0.037	0.956	0.004	0.048	0.049	0.948
		300	0.001	0.029	0.028	0.941	<0.001	0.030	0.030	0.942	−0.001	0.040	0.040	0.957
		500	<0.001	0.021	0.022	0.947	<0.001	0.023	0.023	0.961	<0.001	0.030	0.031	0.948
	0.5	100	0.002	0.042	0.042	0.935	0.003	0.051	0.050	0.947	0.005	0.055	0.056	0.945
		200	<0.001	0.031	0.030	0.943	0.002	0.035	0.036	0.958	0.003	0.039	0.039	0.952
		300	<0.001	0.025	0.025	0.944	0.001	0.030	0.029	0.946	0.001	0.033	0.032	0.951
		500	<0.001	0.019	0.019	0.953	−0.001	0.022	0.023	0.960	<0.001	0.024	0.025	0.959
	0.8	100	0.002	0.040	0.041	0.935	0.002	0.047	0.050	0.959	0.005	0.050	0.052	0.959
		200	0.001	0.029	0.029	0.950	0.001	0.035	0.035	0.951	0.002	0.036	0.037	0.961
		300	<0.001	0.024	0.024	0.947	<0.001	0.030	0.029	0.939	0.001	0.031	0.030	0.939
		500	<0.001	0.018	0.018	0.958	−0.001	0.022	0.022	0.952	<0.001	0.023	0.023	0.956

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.4: Simulation results for the estimators of survival probabilities at three time points $t = 0.5, 1.2, 1.8$ under model (4.2) for $Z = 1$ and $\beta = 0.5$.

			$\boldsymbol{t} = 0.5$				$\boldsymbol{t} = 1.2$				$\boldsymbol{t} = 1.8$			
β	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
0.5	0.2	200	0.003	0.031	0.031	0.938	0.002	0.038	0.037	0.948	0.003	0.051	0.051	0.945
		300	0.001	0.026	0.025	0.942	0.001	0.031	0.031	0.949	0.001	0.042	0.042	0.939
		500	0.002	0.019	0.020	0.938	0.001	0.024	0.024	0.952	0.002	0.033	0.032	0.948
	0.5	200	<0.001	0.026	0.027	0.950	0.002	0.034	0.036	0.952	0.004	0.038	0.039	0.951
		300	<0.001	0.023	0.022	0.936	<0.001	0.029	0.029	0.947	0.001	0.032	0.032	0.947
		500	0.001	0.017	0.017	0.948	<0.001	0.023	0.023	0.949	0.001	0.025	0.025	0.948
	0.8	200	<0.001	0.025	0.026	0.952	<0.001	0.034	0.036	0.958	0.001	0.034	0.035	0.965
		300	−0.001	0.022	0.021	0.948	−0.001	0.028	0.029	0.956	0.001	0.028	0.029	0.956
		500	<0.001	0.017	0.016	0.949	<0.001	0.023	0.023	0.945	0.001	0.023	0.022	0.944

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.5: Simulation results for the estimators of survival probabilities at three time points $t = 0.5, 1.2, 1.8$ under model (4.2) for $Z = 1$ and $\beta = 1$.

			$\boldsymbol{t} = 0.5$				$\boldsymbol{t} = 1.2$				$\boldsymbol{t} = 1.8$			
β	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
1	0.2	200	0.003	0.027	0.027	0.929	0.004	0.039	0.038	0.944	0.005	0.051	0.051	0.942
		300	0.003	0.022	0.022	0.928	0.002	0.032	0.032	0.943	0.002	0.041	0.042	0.948
		500	0.001	0.016	0.017	0.940	0.001	0.024	0.024	0.954	0.002	0.032	0.033	0.952
	0.5	200	0.003	0.023	0.022	0.930	0.002	0.036	0.036	0.961	0.003	0.036	0.036	0.943
		300	0.003	0.019	0.018	0.931	0.001	0.031	0.030	0.949	0.002	0.029	0.030	0.954
		500	0.001	0.014	0.014	0.952	<0.001	0.023	0.023	0.958	0.001	0.022	0.023	0.949
	0.8	200	0.002	0.023	0.022	0.938	<0.001	0.036	0.036	0.946	0.003	0.033	0.033	0.951
		300	0.002	0.018	0.018	0.944	<0.001	0.030	0.029	0.947	<0.001	0.027	0.026	0.928
		500	0.001	0.014	0.014	0.947	−0.001	0.023	0.023	0.958	<0.001	0.021	0.020	0.943

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.6: Simulation results for the estimators of survival probabilities at three time points $t = 0.5, 1.2, 1.8$ under model (4.2) for $Z = 0$ and $\beta = 0.1$.

			$t = 0.5$				$t = 1.2$				$t = 1.8$			
β	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
0.1	0.2	200	0.001	0.030	0.030	0.952	0.001	0.034	0.035	0.956	0.002	0.036	0.036	0.957
		300	<0.001	0.025	0.024	0.946	0.001	0.030	0.029	0.940	0.001	0.031	0.030	0.943
		500	<0.001	0.019	0.019	0.949	<0.001	0.023	0.022	0.945	<0.001	0.023	0.023	0.953
	0.5	200	<0.001	0.031	0.032	0.960	0.001	0.035	0.035	0.956	0.002	0.040	0.039	0.945
		300	<0.001	0.027	0.026	0.940	<0.001	0.030	0.029	0.943	<0.001	0.033	0.032	0.952
		500	<0.001	0.020	0.020	0.954	<0.001	0.023	0.022	0.948	−0.001	0.025	0.025	0.946
	0.8	200	−0.002	0.039	0.039	0.951	0.001	0.035	0.036	0.961	0.003	0.048	0.049	0.941
		300	<0.001	0.032	0.031	0.944	<0.001	0.030	0.029	0.946	−0.002	0.040	0.040	0.949
		500	−0.001	0.024	0.024	0.945	<0.001	0.023	0.023	0.952	<0.001	0.031	0.031	0.951

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.7: Simulation results for the estimators of survival probabilities at three time points $t = 0.5, 1.2, 1.8$ under model (4.2) for $Z = 0$ and $\beta = 0.5$.

			$\boldsymbol{t} = 0.5$				$\boldsymbol{t} = 1.2$				$\boldsymbol{t} = 1.8$			
β	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
0.5	0.2	200	−0.001	0.031	0.030	0.951	<0.001	0.035	0.035	0.949	0.003	0.038	0.037	0.940
		300	<0.001	0.025	0.025	0.947	<0.001	0.030	0.029	0.939	0.001	0.031	0.030	0.936
		500	0.001	0.019	0.019	0.961	<0.001	0.022	0.022	0.946	0.001	0.022	0.024	0.966
	0.5	200	−0.002	0.032	0.033	0.958	−0.001	0.034	0.035	0.956	0.001	0.041	0.042	0.956
		300	−0.001	0.028	0.027	0.945	<0.001	0.030	0.029	0.943	<0.001	0.035	0.034	0.951
		500	<0.001	0.021	0.021	0.958	<0.001	0.022	0.022	0.949	0.001	0.026	0.026	0.960
	0.8	200	−0.006	0.042	0.043	0.950	<0.001	0.037	0.037	0.946	0.004	0.055	0.056	0.945
		300	−0.004	0.036	0.034	0.941	−0.001	0.031	0.030	0.946	<0.001	0.046	0.045	0.947
		500	−0.003	0.026	0.026	0.961	<0.001	0.023	0.023	0.951	0.003	0.036	0.035	0.943

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.8: Simulation results for the estimators of survival probabilities at three time points $t = 0.5, 1.2, 1.8$ under model (4.2) for $Z = 0$ and $\beta = 1$.

			$\boldsymbol{t} = 0.5$				$\boldsymbol{t} = 1.2$				$\boldsymbol{t} = 1.8$			
β	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
1	0.2	200	<0.001	0.032	0.031	0.947	0.001	0.035	0.035	0.947	0.003	0.039	0.038	0.943
		300	<0.001	0.026	0.025	0.937	<0.001	0.030	0.029	0.944	0.001	0.032	0.031	0.936
		500	−0.001	0.020	0.020	0.952	−0.001	0.022	0.022	0.953	<0.001	0.024	0.024	0.953
	0.5	200	−0.003	0.037	0.036	0.940	<0.001	0.036	0.036	0.952	0.004	0.046	0.046	0.944
		300	<0.001	0.030	0.029	0.945	−0.001	0.030	0.029	0.947	<0.001	0.038	0.037	0.941
		500	−0.002	0.022	0.022	0.962	−0.001	0.022	0.023	0.955	−0.001	0.028	0.029	0.953
	0.8	200	−0.009	0.049	0.048	0.948	−0.002	0.037	0.038	0.953	0.003	0.065	0.064	0.932
		300	−0.004	0.040	0.039	0.948	−0.001	0.031	0.031	0.951	<0.001	0.053	0.053	0.956
		500	−0.004	0.031	0.030	0.939	<0.001	0.024	0.024	0.950	0.002	0.042	0.041	0.949

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.9: Simulation results for the unweighted estimator $\hat{\beta}_n$ and the weighted estimator $\hat{\beta}_{w,n}$ with weight function $1 + 0.5/\log(t)$ under model (4.3).

		$\hat{\beta}_n$				$\hat{\beta}_{w,n}$				
β	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	RE.
0.001	100	0.002	0.129	0.125	0.944	0.002	0.127	0.123	0.945	96.9%
	200	−0.001	0.089	0.086	0.938	−0.002	0.087	0.085	0.940	95.6%
	300	0.001	0.067	0.069	0.964	0.001	0.067	0.068	0.961	1
	500	−0.001	0.052	0.053	0.948	−0.001	0.052	0.052	0.948	1
	1000	0.002	0.037	0.037	0.946	0.002	0.037	0.037	0.951	1
0.1	100	0.003	0.135	0.131	0.949	0.003	0.133	0.129	0.951	97.1%
	200	0.002	0.096	0.090	0.940	0.002	0.095	0.089	0.937	97.9%
	300	0.001	0.072	0.073	0.947	0.001	0.071	0.072	0.948	97.2%
	500	<0.001	0.056	0.056	0.951	−0.001	0.056	0.055	0.950	1
	1000	<0.001	0.039	0.039	0.960	<0.001	0.038	0.039	0.960	94.9%
0.2	100	0.009	0.144	0.139	0.950	0.008	0.141	0.137	0.954	95.9%
	200	0.004	0.099	0.096	0.947	0.003	0.098	0.095	0.952	98.0%
	300	0.003	0.079	0.077	0.952	0.002	0.078	0.076	0.953	97.5%
	500	−0.001	0.060	0.059	0.952	−0.001	0.059	0.059	0.957	96.7%
	1000	0.002	0.041	0.042	0.956	0.002	0.040	0.041	0.955	95.2%
0.5	100	0.016	0.174	0.167	0.950	0.012	0.170	0.164	0.950	95.5%
	200	0.007	0.120	0.115	0.938	0.005	0.117	0.113	0.941	95.1%
	300	0.006	0.090	0.093	0.947	0.004	0.089	0.092	0.950	97.8%
	500	0.002	0.074	0.071	0.942	<0.001	0.072	0.071	0.939	94.7%
	1000	0.002	0.051	0.050	0.947	0.001	0.050	0.050	0.947	96.1%

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; RE., relative efficiency of $\hat{\beta}_n$ vs. $\hat{\beta}_{w,n}$

survival curves shown in Figure 4.2 decrease gradually with time. Table 4.10-4.12 list the simulation results for the estimators of the survival probabilities at $t = 2, 3, 4$ for $Z = 1$ based on 1000 replications. For $Z = 1$, β's of $0.1, 0.5, 1$ correspond to true survival probabilities of $0.842, 0.797, 0.726$ respectively at $t = 2$, $0.669, 0.536, 0.340$ at $t = 3$, and $0.479, 0.278, 0.077$ at $t = 4$. As shown in Table 4.3-4.5 and Table 4.10-4.12, varying the baseline survival distribution has little effect on the pattern of simulation results for the survival estimates. The results in Table 4.10-4.12 indicate that the proposed estimation procedures provide reliable point estimates and CIs for the survival function. Some undercoverage is observed for CIs in the simulation scenarios when $\beta = 1$, $p1 = 0.2$ and $t = 4$: the empirical coverage probabilities range from 0.918 to 0.923. The low coverage probability may be due the large uncertainty in the tail of survival distribution and the fewer exposed samples.

4.3 Time-Varying and Time-Independent Covariates

Simulations were also done to assess the performance of the proposed methods for models including both the time-varying and time-independent covariates. We generate failure times from the following model:

$$\log[-\log\{S(t|Z)\}] = \log\{(0.2 \times t)^2\} + \beta_1 \times Z + \beta_2 \times Z \times \log(t), \tag{4.4}$$

where the baseline follows the Weibull distribution with shape parameter 2 and scale parameter 5, and Z takes values 1 or 0 with probability $p1, 1 - p1$. The censoring times are uniformly distributed in the interval 0.5 to c. We choose the values of c according to prespecified censoring rate, 20%. We select $(\beta_1, \beta_2) = (0.1, 0.05), (0.5, 0.05), (0.1, -0.2), (0.5, -0.2)$. All simulations are based on 1000 replications.

Table 4.13-4.16 list the simulation results of the unweighted and optimal weighted estimators of (β_1, β_2) for $p1 = 0.5, n = 200, 300, 500, 1000$. The optimal weight $\{1, 1 + 0.5/\log(t)\}$ does not improve the efficiency of beta estimates. The results for the unweighted estimates are the same as the results for the weighted estimates. From Table 4.13-4.16, we observe some bias in beta estimates for small sample size but the bias decreases as the sample size in-

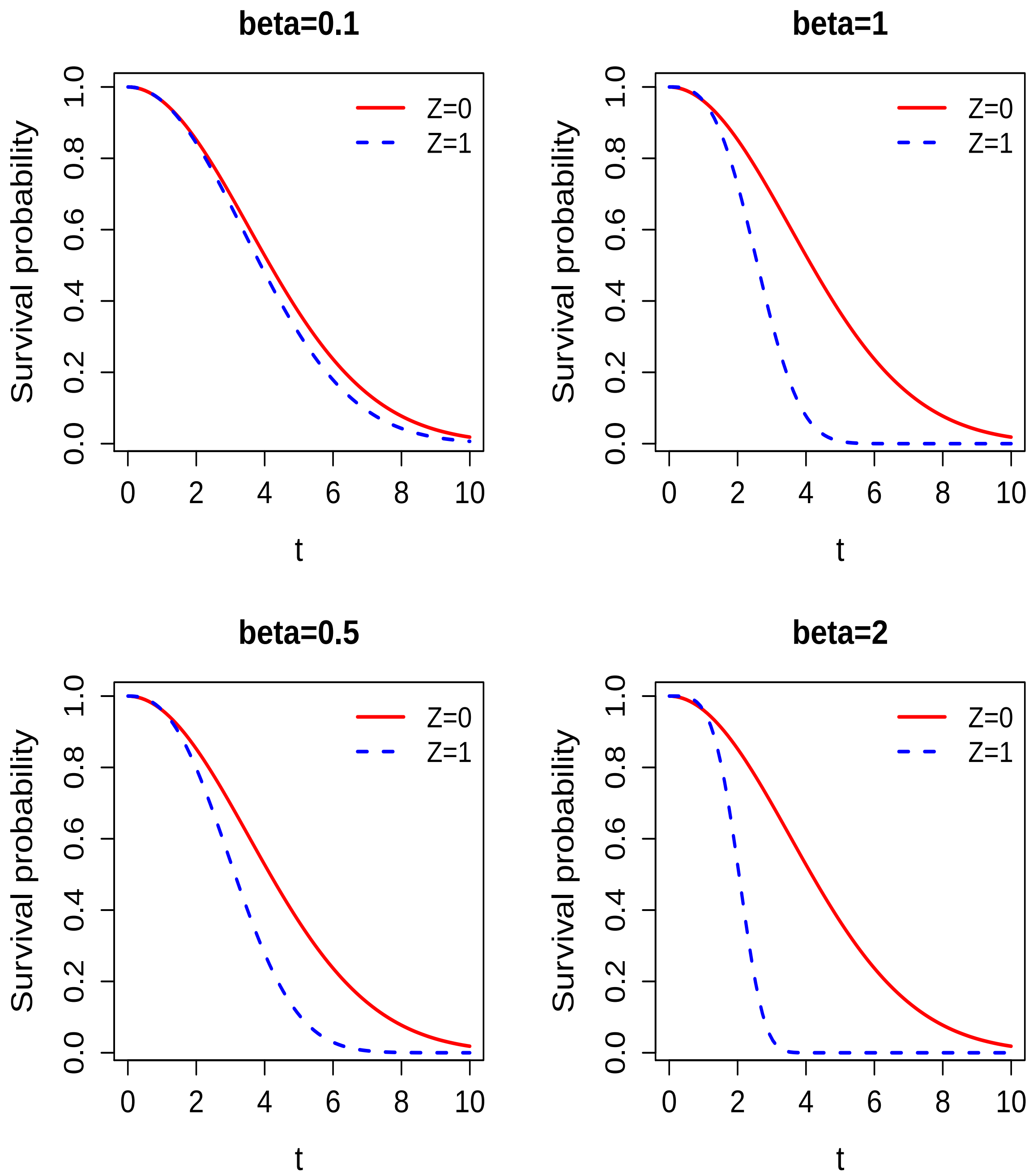

Figure 4.2: True survival curves under model (4.3) for subgroups with $Z = 1$ and $Z = 0$ respectively.

Table 4.10: Simulation results for the estimators of survival probabilities at three time points $t = 2, 3, 4$ under model (4.3) for $Z = 1$ and $\beta = 0.1$.

			$t = 2$				$t = 3$				$t = 4$			
β	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
0.1	0.2	100	<0.001	0.041	0.040	0.935	−0.004	0.068	0.064	0.930	−0.007	0.088	0.083	0.934
		200	0.001	0.029	0.028	0.935	<0.001	0.045	0.044	0.940	<0.001	0.059	0.058	0.939
		300	−0.001	0.024	0.023	0.944	−0.004	0.038	0.036	0.943	−0.005	0.049	0.047	0.939
		500	−0.001	0.018	0.018	0.941	−0.001	0.029	0.028	0.943	−0.002	0.039	0.037	0.940
	0.5	100	0.001	0.038	0.038	0.930	0.001	0.052	0.053	0.954	0.003	0.062	0.062	0.951
		200	<0.001	0.028	0.027	0.930	<0.001	0.039	0.038	0.946	0.001	0.046	0.044	0.939
		300	0.001	0.022	0.022	0.947	<0.001	0.030	0.031	0.948	0.002	0.035	0.036	0.949
		500	<0.001	0.017	0.017	0.960	0.001	0.023	0.024	0.960	0.002	0.026	0.028	0.970
	0.8	100	0.001	0.037	0.037	0.939	0.002	0.050	0.050	0.944	0.005	0.056	0.056	0.943
		200	0.001	0.027	0.026	0.942	0.002	0.035	0.035	0.943	0.003	0.040	0.039	0.952
		300	<0.001	0.021	0.022	0.959	0.001	0.029	0.029	0.952	0.003	0.033	0.032	0.948
		500	<0.001	0.017	0.017	0.950	<0.001	0.023	0.023	0.942	0.002	0.026	0.025	0.939

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.11: Simulation results for the estimators of survival probabilities at three time points $t = 2, 3, 4$ under model (4.3) for $Z = 1$ and $\beta = 0.5$.

			$\boldsymbol{t} = 2$				$\boldsymbol{t} = 3$				$\boldsymbol{t} = 4$			
β	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
0.5	0.2	200	<0.001	0.033	0.034	0.953	<0.001	0.054	0.054	0.956	0.001	0.062	0.060	0.941
		300	<0.001	0.028	0.028	0.949	−0.003	0.044	0.044	0.948	−0.002	0.049	0.049	0.945
		500	<0.001	0.021	0.022	0.955	−0.002	0.035	0.034	0.939	−0.001	0.040	0.038	0.932
	0.5	200	0.001	0.032	0.031	0.951	0.002	0.042	0.043	0.956	0.002	0.043	0.042	0.943
		300	0.002	0.026	0.026	0.938	0.001	0.034	0.035	0.957	0.001	0.034	0.035	0.954
		500	0.001	0.020	0.020	0.950	0.001	0.027	0.027	0.941	<0.001	0.027	0.027	0.942
	0.8	200	0.001	0.030	0.030	0.941	0.001	0.039	0.039	0.957	0.002	0.035	0.037	0.954
		300	<0.001	0.025	0.025	0.944	<0.001	0.033	0.032	0.946	0.002	0.030	0.030	0.945
		500	0.001	0.019	0.019	0.938	0.001	0.026	0.025	0.946	0.001	0.024	0.023	0.950

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.12: Simulation results for the estimators of survival probabilities at three time points $t = 2, 3, 4$ under model (4.3) for $Z = 1$ and $\beta = 1$.

			$t = 2$				$t = 3$				$t = 4$			
β	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
1	0.2	200	<0.001	0.044	0.042	0.931	−0.001	0.059	0.060	0.939	0.006	0.037	0.038	0.920
		300	<0.001	0.035	0.035	0.949	−0.003	0.050	0.049	0.940	0.002	0.031	0.030	0.918
		500	0.001	0.027	0.027	0.934	−0.001	0.039	0.038	0.950	0.001	0.024	0.024	0.923
	0.5	200	−0.001	0.037	0.037	0.945	−0.001	0.044	0.044	0.945	0.003	0.027	0.026	0.934
		300	<0.001	0.030	0.030	0.952	<0.001	0.036	0.036	0.948	0.002	0.022	0.021	0.930
		500	−0.001	0.024	0.023	0.941	−0.001	0.028	0.028	0.947	0.001	0.017	0.016	0.943
	0.8	200	−0.001	0.036	0.034	0.934	0.001	0.038	0.039	0.945	0.004	0.022	0.023	0.957
		300	−0.001	0.029	0.028	0.944	0.002	0.033	0.032	0.943	0.003	0.018	0.019	0.951
		500	0.001	0.022	0.022	0.946	0.002	0.024	0.024	0.950	0.003	0.014	0.014	0.954

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

creases. The corresponding CIs are reliable for all cases. The estimates for the time-covariate interaction β_2 appear to be more efficient than the estimates for the time-independent covariate β_1. The biases, as well as the variances, of the estimates for β_2 are smaller than that for β_1.

In addition, we estimate the survival probabilities at $t = 2, 4, 6$ for $Z = 1$ based on the data simulated from model (4.4) with $p1 = 0.2, 0.5, 0.8$, and report the results in Table 4.17-4.20. The true survival curves under model (4.4) for subsets of subjects with $Z = 0$ and $Z = 1$ respectively are shown in Figure 4.3. The pattern of curves in Figure 4.3 is similar to that in Figure 4.2. The true probabilities of surviving beyond time $2, 4, 6$ for $Z = 1$ are 0.833, 0.469, 0.175 when $(\beta_1, \beta_2) = (0.1, 0.05)$, 0.761, 0.323, 0.075 when $(\beta_1, \beta_2) = (0.5, 0.05)$, 0.857, 0.585, 0.329 when $(\beta_1, \beta_2) = (0.1, -0.2)$, and 0.795, 0.449, 0.190 when $(\beta_1, \beta_2) = (0.5, -0.2)$. Like the simulation results in previous sections, the survival estimator also performs quite well under model (4.4). The biases are all less than 0.005. We see close similarity between the empirical variances and the estimated variances. The head and tail of survival distribution are subject to greater uncertainty. The CI is not reliable for true survival probability close to 0 or 1 and small p_1. As one might expect, the coverage proportions are in the range 0.93-0.97 for nominal level 0.95. Undercoverage is seen for CIs when $p1 = 0.1$ and $t = 2, 6$. Nevertheless SEs vary with the choice of $t, n, p1$. We see a trend of SE decrease at the later follow-up time after the initial increase. Still, larger sample size provides reduction in SE. This also holds true for the choice of $p1$.

In model (4.4), the time-independent covariate Z and the time-varying covariate $Z \times \log(t)$ are highly correlated. Simulation experiments are also done for independent covariates. We consider the following model with Weibull baseline:

$$\log[-\log\{S(t|Z)\}] = \log\{(0.2 \times t)^2\} + \beta_1 \times Z_1 + \beta_2 \times Z_2 \times \log(t), \tag{4.5}$$

where Z_1 takes 0 or 1 with equal probability, Z_2 is uniformly distributed on the interval (1,3), and Z_1 and Z_2 are independent. The failure times generated from model (4.5) are subject to about 20% independent uniform censorship. For comparison, we use the same

Table 4.13: Simulation results for the unweighted estimator $(\hat{\beta}_{1,n}, \hat{\beta}_{2,n})$ and the weighted estimator $(\hat{\beta}_{w,1,n}, \hat{\beta}_{w,2,n})$ with weight function $(1, 1 + 0.5/\log(t))$ under model (4.4) for $(\beta_1, \beta_2) = (0.1, 0.05)$.

		$\hat{\beta}_n$				$\hat{\beta}_{w,n}$				
	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	RE.
$\beta_1 = 0.1$	200	−0.018	0.474	0.469	0.952	−0.018	0.474	0.469	0.952	1
	300	−0.016	0.383	0.380	0.958	−0.016	0.383	0.380	0.958	1
	500	0.006	0.297	0.290	0.937	0.006	0.297	0.290	0.937	1
	1000	−0.005	0.201	0.202	0.947	−0.005	0.201	0.202	0.947	1
$\beta_2 = 0.05$	200	0.010	0.264	0.263	0.960	0.010	0.264	0.263	0.960	1
	300	0.008	0.212	0.213	0.952	0.008	0.212	0.213	0.952	1
	500	−0.004	0.166	0.162	0.947	−0.004	0.166	0.162	0.947	1
	1000	0.001	0.112	0.113	0.950	0.001	0.112	0.113	0.950	1

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; RE., relative efficiency of $\hat{\beta}_n$ vs. $\hat{\beta}_{w,n}$

Table 4.14: Simulation results for the unweighted estimator $(\hat{\beta}_{1,n}, \hat{\beta}_{2,n})$ and the weighted estimator $(\hat{\beta}_{w,1,n}, \hat{\beta}_{w,2,n})$ with weight function $(1, 1 + 0.5/\log(t))$ under model (4.4) for $(\beta_1, \beta_2) = (0.5, 0.05)$.

		$\hat{\beta}_n$				$\hat{\beta}_{w,n}$				
	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	RE.
$\beta_1 = 0.5$	200	−0.006	0.466	0.459	0.950	−0.006	0.466	0.459	0.950	1
	300	−0.005	0.374	0.368	0.954	−0.005	0.374	0.368	0.954	1
	500	−0.001	0.289	0.282	0.949	−0.001	0.289	0.282	0.949	1
	1000	−0.003	0.195	0.197	0.944	−0.003	0.195	0.197	0.944	1
$\beta_2 = 0.05$	200	0.008	0.285	0.275	0.941	0.008	0.285	0.275	0.941	1
	300	0.009	0.223	0.219	0.952	0.009	0.223	0.219	0.952	1
	500	0.002	0.168	0.168	0.956	0.002	0.168	0.168	0.956	1
	1000	0.003	0.115	0.117	0.949	0.003	0.115	0.117	0.949	1

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; RE., relative efficiency of $\hat{\beta}_n$ vs. $\hat{\beta}_{w,n}$

Table 4.15: Simulation results for the unweighted estimator $(\hat{\beta}_{1,n}, \hat{\beta}_{2,n})$ and the weighted estimator $(\hat{\beta}_{w,1,n}, \hat{\beta}_{w,2,n})$ with weight function $(1, 1 + 0.5/\log(t))$ under model (4.4) for $(\beta_1, \beta_2) = (0.1, -0.2)$.

		$\hat{\beta}_n$				$\hat{\beta}_{w,n}$				
	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	RE.
$\beta_1 = 0.1$	200	0.032	0.476	0.466	0.936	0.032	0.476	0.466	0.936	1
	300	0.021	0.372	0.376	0.959	0.021	0.372	0.376	0.959	1
	500	0.008	0.300	0.288	0.945	0.008	0.300	0.288	0.945	1
	1000	0.001	0.201	0.202	0.944	0.001	0.201	0.202	0.944	1
$\beta_2 = -0.2$	200	−0.017	0.254	0.247	0.941	−0.017	0.254	0.247	0.941	1
	300	−0.014	0.198	0.199	0.954	−0.014	0.198	0.199	0.954	1
	500	−0.003	0.155	0.151	0.948	−0.003	0.155	0.151	0.948	1
	1000	<0.001	0.106	0.106	0.954	<0.001	0.106	0.106	0.954	1

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; RE., relative efficiency of $\hat{\beta}_n$ vs. $\hat{\beta}_{w,n}$

Table 4.16: Simulation results for the unweighted estimator $(\hat{\beta}_{1,n}, \hat{\beta}_{2,n})$ and the weighted estimator $(\hat{\beta}_{w,1,n}, \hat{\beta}_{w,2,n})$ with weight function $(1, 1 + 0.5/\log(t))$ under model (4.4) for $(\beta_1, \beta_2) = (0.5, -0.2)$.

		$\hat{\beta}_n$				$\hat{\beta}_{w,n}$				
	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	RE.
$\beta_1 = 0.5$	200	0.022	0.447	0.441	0.947	0.022	0.447	0.441	0.947	1
	300	0.009	0.359	0.356	0.959	0.009	0.359	0.356	0.959	1
	500	0.005	0.272	0.274	0.947	0.005	0.272	0.274	0.947	1
	1000	0.001	0.191	0.191	0.946	0.001	0.191	0.191	0.946	1
$\beta_2 = -0.2$	200	-0.017	0.253	0.245	0.945	-0.017	0.253	0.245	0.945	1
	300	-0.003	0.199	0.197	0.952	-0.003	0.199	0.197	0.952	1
	500	-0.005	0.153	0.151	0.950	-0.005	0.153	0.151	0.950	1
	1000	-0.002	0.105	0.105	0.954	-0.002	0.105	0.105	0.954	1

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; RE., relative efficiency of $\hat{\beta}_n$ vs. $\hat{\beta}_{w,n}$

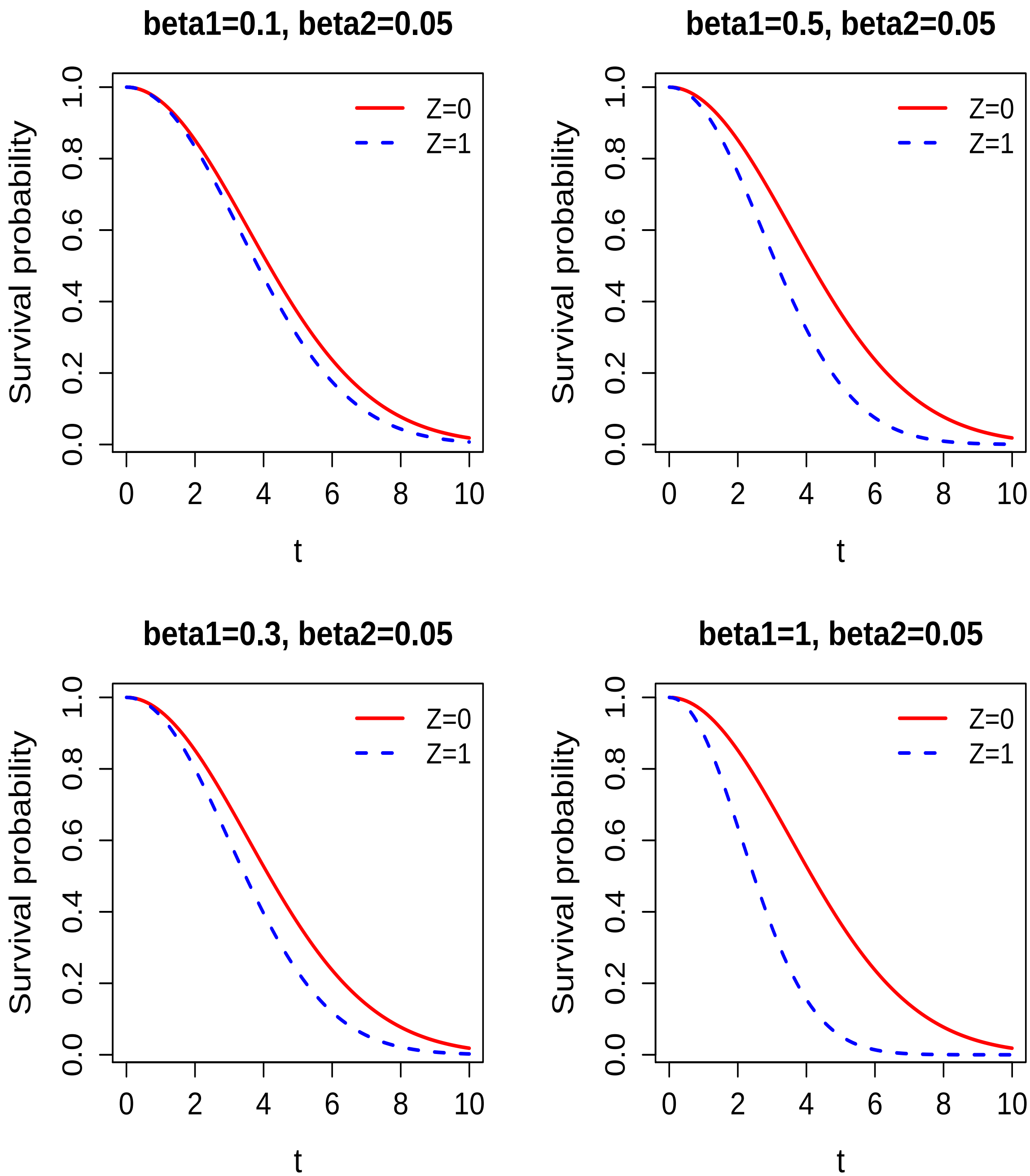

Figure 4.3: True survival curves under model (4.4) for subgroups with $Z = 1$ and $Z = 0$ respectively.

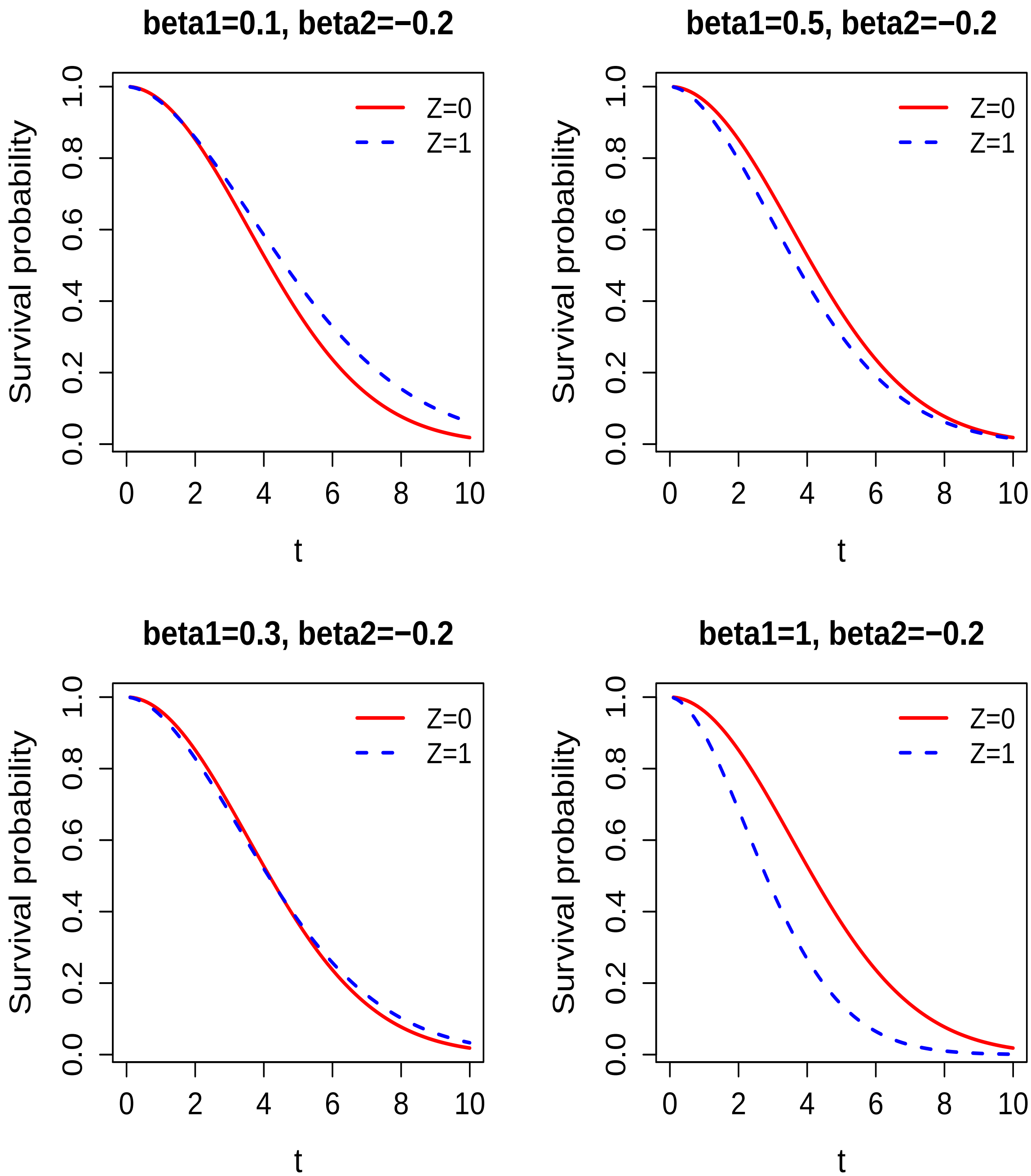

Figure 4.3: (continued)

Table 4.17: Simulation results for the estimators of survival probabilities at three time points $t = 2, 4, 6$ under model (4.4) for $Z = 1$ and $(\beta_1, \beta_2) = (0.1, 0.05)$.

			$\boldsymbol{t} = 2$				$\boldsymbol{t} = 4$				$\boldsymbol{t} = 6$			
$\boldsymbol{\beta}$	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
(0.1, 0.05)	0.2	200	<0.001	0.053	0.052	0.924	−0.001	0.073	0.072	0.935	−0.001	0.059	0.056	0.918
		300	0.002	0.043	0.042	0.929	0.001	0.062	0.058	0.929	<0.001	0.048	0.046	0.927
		500	0.002	0.032	0.032	0.934	0.002	0.045	0.045	0.948	0.002	0.038	0.036	0.939
	0.5	200	<0.001	0.036	0.035	0.924	0.001	0.050	0.048	0.949	0.002	0.041	0.039	0.932
		300	<0.001	0.029	0.028	0.935	<0.001	0.038	0.040	0.960	<0.001	0.031	0.032	0.946
		500	<0.001	0.022	0.022	0.949	−0.001	0.031	0.031	0.937	<0.001	0.025	0.025	0.946
	0.8	200	0.001	0.030	0.029	0.924	−0.001	0.039	0.041	0.962	0.005	0.033	0.033	0.947
		300	0.001	0.024	0.024	0.939	0.003	0.033	0.033	0.949	0.003	0.026	0.027	0.949
		500	<0.001	0.019	0.018	0.945	<0.001	0.025	0.026	0.955	0.002	0.020	0.021	0.960

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.18: Simulation results for the estimators of survival probabilities at three time points $t = 2, 4, 6$ under model (4.4) for $Z = 1$ and $(\beta_1, \beta_2) = (0.5, 0.05)$.

			$\boldsymbol{t} = 2$				$\boldsymbol{t} = 4$				$\boldsymbol{t} = 6$			
$\boldsymbol{\beta}$	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
(0.5, 0.05)	0.2	200	0.004	0.063	0.061	0.920	<0.001	0.067	0.067	0.943	0.001	0.039	0.038	0.899
		300	0.002	0.050	0.049	0.935	−0.004	0.053	0.055	0.947	−0.001	0.031	0.031	0.901
		500	0.004	0.037	0.038	0.951	0.001	0.043	0.042	0.945	<0.001	0.025	0.024	0.926
	0.5	200	0.001	0.041	0.041	0.940	0.001	0.045	0.046	0.943	0.003	0.027	0.027	0.930
		300	0.001	0.033	0.033	0.938	0.001	0.035	0.037	0.961	0.002	0.022	0.022	0.941
		500	0.001	0.026	0.026	0.948	0.001	0.029	0.029	0.948	0.001	0.018	0.017	0.934
	0.8	200	<0.001	0.033	0.034	0.954	−0.002	0.039	0.039	0.937	0.002	0.023	0.023	0.943
		300	<0.001	0.028	0.027	0.956	0.001	0.031	0.032	0.955	0.002	0.020	0.019	0.938
		500	<0.001	0.021	0.021	0.950	<0.001	0.024	0.025	0.961	0.001	0.014	0.015	0.955

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.19: Simulation results for the estimators of survival probabilities at three time points $t = 2, 4, 6$ under model (4.4) for $Z = 1$ and $(\beta_1, \beta_2) = (0.1, -0.2)$.

			$\boldsymbol{t} = 2$				$\boldsymbol{t} = 4$				$\boldsymbol{t} = 6$			
$\boldsymbol{\beta}$	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
(0.1, −0.2)	0.2	200	0.001	0.051	0.047	0.905	0.001	0.076	0.071	0.925	−0.002	0.075	0.070	0.926
		300	0.002	0.039	0.039	0.933	0.001	0.058	0.058	0.947	<0.001	0.061	0.057	0.928
		500	0.001	0.030	0.030	0.938	−0.001	0.046	0.045	0.944	−0.002	0.045	0.045	0.937
	0.5	200	0.001	0.033	0.032	0.932	0.002	0.049	0.047	0.938	0.003	0.050	0.048	0.928
		300	0.002	0.025	0.026	0.954	0.003	0.039	0.039	0.944	0.002	0.040	0.039	0.942
		500	<0.001	0.020	0.020	0.946	0.002	0.030	0.030	0.953	0.003	0.030	0.030	0.955
	0.8	200	<0.001	0.027	0.027	0.943	0.001	0.039	0.039	0.944	0.004	0.041	0.040	0.939
		300	0.001	0.022	0.022	0.941	0.001	0.032	0.032	0.951	0.002	0.033	0.032	0.947
		500	0.002	0.017	0.017	0.943	0.002	0.025	0.025	0.953	0.002	0.025	0.025	0.950

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

Table 4.20: Simulation results for the estimators of survival probabilities at three time points $t = 2, 4, 6$ under model (4.4) for $Z = 1$ and $(\beta_1, \beta_2) = (0.5, -0.2)$.

			$t = 2$				$t = 4$				$t = 6$			
$\boldsymbol{\beta}$	$p1$	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.
$(0.5, -0.2)$	0.2	200	-0.002	0.058	0.057	0.920	0.001	0.072	0.071	0.939	0.004	0.060	0.059	0.932
		300	<0.001	0.047	0.046	0.936	<0.001	0.059	0.058	0.936	-0.001	0.048	0.048	0.936
		500	-0.001	0.036	0.036	0.942	-0.001	0.045	0.045	0.952	<0.001	0.037	0.037	0.946
	0.5	200	0.001	0.038	0.038	0.944	<0.001	0.049	0.048	0.939	0.001	0.041	0.040	0.940
		300	-0.002	0.032	0.031	0.936	-0.001	0.040	0.040	0.953	0.001	0.032	0.033	0.945
		500	<0.001	0.023	0.024	0.949	0.001	0.031	0.031	0.951	0.001	0.026	0.026	0.947
	0.8	200	<0.001	0.031	0.031	0.950	<0.001	0.040	0.041	0.955	0.002	0.035	0.034	0.940
		300	0.001	0.026	0.026	0.945	0.002	0.034	0.033	0.944	0.003	0.028	0.028	0.951
		500	0.002	0.020	0.020	0.930	0.002	0.025	0.026	0.944	0.002	0.021	0.022	0.955

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; $p1$, the probability that Z equals 1

set of simulation parameters as in the above experiment for correlated covariates, and we also compute an unweighted estimate and a weighted estimate with optimal weight $(1, 1 + 0.5/\log(t))$ for each simulated sample. The key results are summarized in Table 4.21-4.24 and are satisfactory for practical sample size. The biases of estimates for (β_1, β_2) are all fairly small. This also holds true with respect to the biases of variance estimates. The CIs remain adequate coverage. The optimal weighted estimates are more efficient than the unweighted estimates for β_2, although the relative efficiency is close to 1. As compared to the results for model (4.4), the estimates have smaller biases and variances for (β_1, β_2) in model (4.5).

Table 4.21: Simulation results for the unweighted estimator $(\hat{\beta}_{1,n}, \hat{\beta}_{2,n})$ and the weighted estimator $(\hat{\beta}_{w,1,n}, \hat{\beta}_{w,2,n})$ with weight function $(1, 1 + 0.5/\log(t))$ under model (4.5) for $(\beta_1, \beta_2) = (0.1, 0.05)$.

		$\hat{\beta}_n$				$\hat{\beta}_{w,n}$				
	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	RE.
$\beta_1 = 0.1$	200	0.003	0.167	0.162	0.946	0.003	0.167	0.162	0.945	1
	300	−0.004	0.136	0.131	0.950	−0.004	0.136	0.131	0.950	1
	500	<0.001	0.102	0.101	0.943	<0.001	0.102	0.101	0.944	1
	1000	0.001	0.071	0.071	0.948	0.001	0.071	0.071	0.948	1
$\beta_2 = 0.05$	200	0.001	0.085	0.083	0.952	0.001	0.084	0.082	0.950	97.7%
	300	0.001	0.071	0.067	0.941	0.001	0.070	0.066	0.935	97.2%
	500	0.001	0.051	0.051	0.950	0.001	0.051	0.051	0.949	1
	1000	0.001	0.036	0.036	0.945	0.001	0.036	0.036	0.939	1

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; RE., relative efficiency of $\hat{\beta}_n$ vs. $\hat{\beta}_{w,n}$

Table 4.22: Simulation results for the unweighted estimator $(\hat{\beta}_{1,n}, \hat{\beta}_{2,n})$ and the weighted estimator $(\hat{\beta}_{w,1,n}, \hat{\beta}_{w,2,n})$ with weight function $(1, 1 + 0.5/\log(t))$ under model (4.5) for $(\beta_1, \beta_2) = (0.5, 0.05)$.

		$\hat{\beta}_n$				$\hat{\beta}_{w,n}$				
	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	RE.
$\beta_1 = 0.5$	200	0.007	0.167	0.166	0.943	0.007	0.167	0.166	0.944	1
	300	0.005	0.132	0.134	0.954	0.005	0.132	0.134	0.956	1
	500	0.003	0.105	0.103	0.947	0.003	0.105	0.103	0.946	1
	1000	0.003	0.073	0.073	0.951	0.003	0.073	0.073	0.951	1
$\beta_2 = 0.05$	200	−0.003	0.091	0.088	0.946	−0.003	0.089	0.087	0.943	95.7%
	300	0.003	0.071	0.071	0.946	0.003	0.070	0.070	0.952	97.2%
	500	<0.001	0.054	0.054	0.958	<0.001	0.053	0.054	0.953	96.3%
	1000	0.001	0.038	0.038	0.947	0.001	0.038	0.038	0.950	1

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; RE., relative efficiency of $\hat{\beta}_n$ vs. $\hat{\beta}_{w,n}$

Table 4.23: Simulation results for the unweighted estimator $(\hat{\beta}_{1,n}, \hat{\beta}_{2,n})$ and the weighted estimator $(\hat{\beta}_{w,1,n}, \hat{\beta}_{w,2,n})$ with weight function $(1, 1 + 0.5/\log(t))$ under model (4.5) for $(\beta_1, \beta_2) = (0.1, -0.2)$.

		$\hat{\beta}_n$				$\hat{\beta}_{w,n}$				
	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	RE.
$\beta_1 = 0.1$	200	0.004	0.160	0.162	0.955	0.004	0.159	0.161	0.955	98.8%
	300	0.002	0.132	0.131	0.955	0.002	0.132	0.131	0.956	1
	500	0.001	0.104	0.101	0.935	0.001	0.104	0.101	0.935	1
	1000	0.001	0.071	0.071	0.955	0.001	0.071	0.071	0.955	1
$\beta_2 = -0.2$	200	−0.004	0.068	0.066	0.943	−0.004	0.067	0.065	0.941	97.1%
	300	−0.005	0.057	0.053	0.934	−0.005	0.057	0.053	0.933	1
	500	0.001	0.040	0.041	0.952	0.001	0.040	0.040	0.954	1
	1000	−0.001	0.029	0.029	0.949	−0.001	0.029	0.028	0.948	1

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; RE., relative efficiency of $\hat{\beta}_n$ vs. $\hat{\beta}_{w,n}$

Table 4.24: Simulation results for the unweighted estimator $(\hat{\beta}_{1,n}, \hat{\beta}_{2,n})$ and the weighted estimator $(\hat{\beta}_{w,1,n}, \hat{\beta}_{w,2,n})$ with weight function $(1, 1 + 0.5/\log(t))$ under model (4.5) for $(\beta_1, \beta_2) = (0.5, -0.2)$.

		$\hat{\beta}_n$				$\hat{\beta}_{w,n}$				
	n	Bias	SE.	ESE.	CP.	Bias	SE.	ESE.	CP.	RE.
$\beta_1 = 0.5$	200	0.006	0.165	0.165	0.955	0.006	0.165	0.165	0.955	1
	300	0.005	0.131	0.133	0.951	0.005	0.131	0.133	0.952	1
	500	0.003	0.106	0.103	0.936	0.003	0.106	0.103	0.937	1
	1000	−0.003	0.072	0.072	0.949	−0.003	0.072	0.072	0.949	1
$\beta_2 = -0.2$	200	−0.002	0.073	0.070	0.943	−0.002	0.072	0.069	0.946	97.3%
	300	−0.001	0.056	0.056	0.961	−0.001	0.055	0.055	0.961	96.5%
	500	0.001	0.042	0.043	0.956	0.001	0.041	0.043	0.961	95.3%
	1000	−0.002	0.030	0.030	0.956	−0.002	0.029	0.030	0.956	93.4%

SE., empirical standard error; ESE., average of estimated standard error; CP., empirical coverage probability of 95% confidence interval; RE., relative efficiency of $\hat{\beta}_n$ vs. $\hat{\beta}_{w,n}$

Chapter 5

REAL EXAMPLES

In this chapter, we present two examples for illustrative purposes. We apply the proposed methods to two real clinical studies, including the well-known Mayo Clinic Primary Biliary Cirrhosis (PBC) trial, and a landmark randomized prevention trial in mother-to-child transmission of the Human Immunodeficiency Virus, namely HIVNET 012.

5.1 PBC Data

5.1.1 Data Description

The Mayo Clinic PBC data set is known to be a valuable resource in the liver disease research [Fleming and Harrington, 1991]. PBC is a rare but fatal chronic liver disease [Dickson et al., 1989]. Between January 1974 and May 1984, a randomized trial in PBC of the liver was conducted at Mayo Clinic. A total of 312 PBC patients were randomly assigned to the drug D-penicillamine (DPCA) or a placebo during the 10-year period. Complete demographic, clinical, biochemical and histologic records of the participants were collected at the time of randomization. The follow-up through 1986 found 125 deaths among the 312 study participants. The PBC data also include 106 PBC patients who were concurrently referred to Mayo Clinic but did not participate in the trial. Baseline laboratory results and time-to-event outcome through 1986 were collected from the additional 106 patients. The Appendix D.1 of Fleming and Harrington [1991] describes and lists the PBC data. The PBC data set is also available in the R package "survival". The PBC data set was first analyzed by the study investigators in 1986 [Dickson et al., 1989, Grambsch et al., 1989]. They did not find significant difference in the survival distribution between the DPCA and placebo groups and suggested combing the two groups in studying the association between the survival and baseline measurements.

Descriptive statistics are used to summarize the characteristics of patients at baseline and follow-up. For quantitative variable, mean, standard deviation (SD), median and range are calculated; for categorical variables, frequency and percentage are calculated. We report the descriptive results in Table 5.1 for the trial participants in each of the two treatment groups and the additional 106 non-participants. Among the 312 participants in the PBC clinical trial, 158 patients received the drug DPCA. The two treatment groups are similar with respect to baseline characteristics. The vast majority are female. The mean age at randomization is about 50 years with an SD of 10 years. About two thirds of patients are at the histologic stage 3 or 4. The distributions of clinical and biochemical variables are similar across treatment groups. As shown in Table 5.1, there is also little difference in the death rate and follow-up time across treatment groups. The 106 patients not included in the PBC trial were missing some clinical and biochemical measurements. All measured baseline characteristics of trial participants and non-participants are similar. There is a slightly higher proportion of death among trail participants compared to non-participants (40.1% vs. 34.0%). However, trial participants appear to have a longer follow-up time than the non-participants.

5.1.2 Data Analysis

We fit the following model to the PBC data set:

$$\log[-\log\{S(t|Z)\}] = \alpha(t) + \beta_1 \times Z + \beta_2 \times Z \times t, \tag{5.1}$$

where Z is baseline edema status (0 = no edema, 1 = edema despite diuretic therapy, untreated or successfully treated). We consider the failure time to be days from the randomization to death. The failure time was censored either because the participant received a liver transplant or because the participant was alive on the date of study analysis in July, 1986. In model (5.1), β_1 estimates the difference in $\log(-\log)$ survival probability at $t = 0$ between $Z = 1$ and $Z = 0$, and β_2 estimates the difference in the difference in $\log(-\log)$ survival

Table 5.1: PBC data. Characteristics of the PBC patients at baseline and follow-up.

	Trial participant			Non-participant
	DPCA (n=158)	Placebo (n=154)	All (n=312)	(n=106)
Demographic Characteristics				
Female, n (%)	137 (86.7)	139 (90.3)	276 (88.5)	98 (92.5)
Mean age at baseline (SD), year	51.42 (11.01)	48.58 (9.96)	50.02 (10.58)	52.87 (9.78)
Range of age at baseline, year	26.28 - 78.44	30.57 - 74.52	26.28 - 78.44	33.00 - 75.00
Clinical Measurement at Baseline				
Ascites, n (%)	14 (8.9)	10 (6.5)	28 (9.0)	NA
Hepatomegaly, n (%)	73 (46.2)	87 (56.5)	160 (51.3)	NA
Spiders, n (%)	45 (28.5)	45 (29.2)	90 (28.8)	NA
Edema, n (%)				
No edema	132 (83.6)	131 (85.1)	263 (84.3)	91 (85.8)
Untreated or successfully treated	16 (10.1)	13 (8.4)	29 (9.3)	15 (14.2)
Edema despite diuretic therapy	10 (6.3)	10 (6.5)	20 (6.4)	0
Histologic Stage at Baseline, n (%)				
Stage 1	12 (7.6)	4 (2.6)	16 (5.1)	5 (4.7)
Stage 2	35 (22.2)	32 (20.8)	67 (21.5)	25 (23.6)
Stage 3	56 (35.4)	64 (41.5)	120 (38.5)	35 (33.0)
Stage 4	55 (34.8)	54 (35.1)	109 (34.9)	35 (33.0)
NA	0	0	0	6 (5.7)

DPCA, D-penicillamine; SD, standard deviation

Table 5.1: (continued)

	Trial participant			Non-participant
	DPCA (n=158)	Placebo (n=154)	All (n=312)	(n=106)
Biochemical Measurement at Baseline				
Mean serum bilirubin (SD), mg/dl	2.873 (3.629)	3.649 (5.282)	3.256 (4.530)	3.117 (4.043)
Mean serum albumin (SD), g/dl	3.516 (0.443)	3.524 (0.396)	3.520 (0.420)	3.431 (0.435)
Mean urine copper (SD), ug/day	97.64 (90.59)	97.65 (80.49)	97.65 (85.61)	NA
Mean alkaline phosphatase (SD), U/l	2021 (2183)	1943 (2102)	1983 (2140)	NA
Mean SGOT (SD), U/ml	120.2 (54.5)	125.0 (58.9)	122.6 (56.7)	NA
Mean serum cholesterol (SD), mg/dl	365.0 (209.5)	373.9 (252.5)	369.5 (231.9)	NA
Mean triglycerides (SD), mg/dl	124.1 (71.5)	125.3 (58.5)	124.7 (65.1)	NA
Mean platelet count (SD)	258.8 (100.3)	265.2 (90.7)	261.9 (95.6)	241.7 (105.4)
Mean prothrombin time (SD), second	10.65 (0.85)	10.80 (1.14)	10.73 (1.00)	10.75 (1.08)
Status at endpoint, n (%)				
Dead	65 (41.2)	60 (39.0)	125 (40.1)	36 (34.0)
Transplant	10 (6.3)	9 (5.8)	19 (6.1)	6 (5.6)
Censored	83 (52.5)	85 (55.2)	168 (53.8)	64 (60.4)
Median time since randomization (range), day	1895 (41 - 4556)	1811 (51 - 4523)	1840 (41 - 4556)	1397 (41 - 4795)
Mean time since randomization (SD), day	2016 (1094)	1997 (1156)	2006 (1123)	1657 (1009)

probability between $Z = 1$ and $Z = 0$ associated with per unit change in t. The proportional hazards assumption in the Cox model implies that the difference in $\log(-\log)$ survival probability between groups is constant over time. A hypothesis test on the covaraite-time interaction in the complementary log-log survival model can serve as a goodness-of-fit tool for the Cox proportional hazards model. Note that model (5.1) with $\beta_2 = 0$ reduces to a standard Cox model. The estimates of regression coefficients and the corresponding confidence intervals (CIs) are computed using the methods of Section 3.2: $\widehat{\beta}_1 = 2.53$ (95% CI, 1.92 to 3.14), $\widehat{\beta}_2 = -4.84\text{e-}4$ (95% CI, $-6.76\text{e-}4$ to $-2.91\text{e-}4$). The coefficient of edema-time interaction is statistically significant, suggesting that there is evidence against the proportional hazards assumption. In Figure 5.1, we plot the $\log(-\log)$ Kaplan-Meier (KM) estimates as a graphical examination of the proportional hazards assumption for the edema covariate. The plot lends support to our finding that the standard Cox model might not be a perfect fit for the edema covariate.

Furthermore, we compute the survival estimates for groups categorized by baseline edema status under model (5.1) and present the estimated survival curves in Figure 5.2(A). Superimposed on the model-based survival curves in Figure 5.2 are the KM survival curves for edema groups. Model (5.1) fits the data well as can be seen in Figure 5.2(A) where the model-based curves and the KM curves match quite well. Moreover, we compare survival estimates under model (5.1) to one without the edema-time interaction. Figure 5.2(B) shows the estimated survival curves for the complementary log-log survival model containing only the edema covariate. It is clear from the curves shown in Figure 5.2 that model (5.1) provides a better fit than the edema covariate only model. The separation between model-based and KM curves for model (5.1) is less than that for the model without the edema-time interaction. Fitting model (5.1) also gives good CI estimation of the survival distribution. Figure 5.3 displays three types of 95% CIs of the survival distribution for each edema group. The dashed lines in Figure 5.3 represent the usual CI based on the KM estimator and the Greenwood formula. The calculated CIs from the fit of the models with and without the edema-time interaction are shown by solid and dotted lines, respectively. As shown in Figure 5.3, the

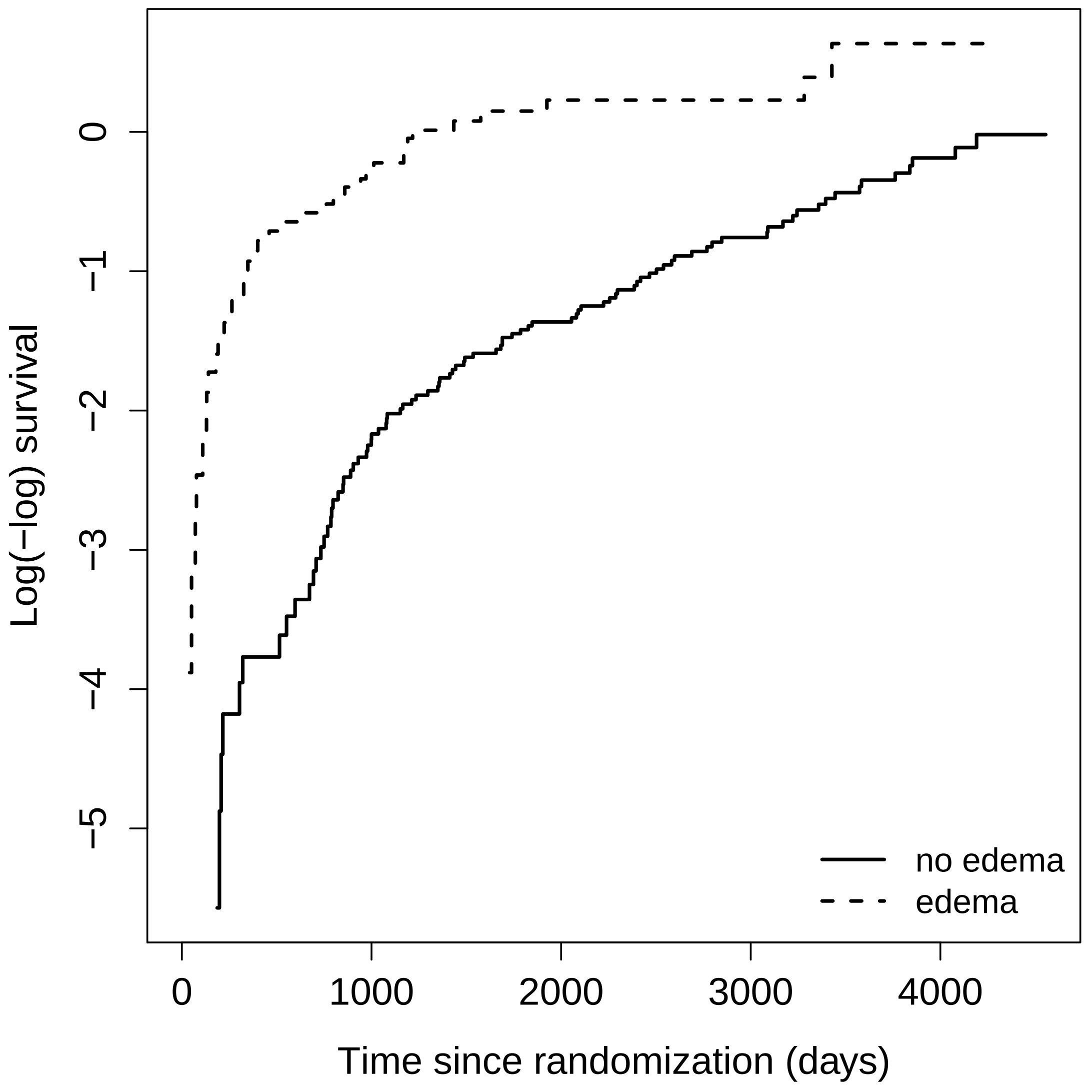

Figure 5.1: PBC Data. Log($-\log$) Kaplan-Meier estimates of patient survival for the no edema group and the edema group.

three types of CIs for the no edema group are close to one another; for the edema group, the CIs from model with interaction are close to the KM CIs while the separation between the CIs from model without interaction and the KM CIs is big.

We also fit model (5.1) with Z selected to be treatment arm (0 = placebo, 1 = DPCA). The resulting coefficients estimates are $\widehat{\beta}_1 = -0.14$ (95% CI, -0.78 to 0.50), $\widehat{\beta}_2$ =7.22e-5 (95% CI, -1.26e-4 to 2.71e-4). We do not find significant interaction, which suggests that the treatment effect might be constant throughout time. We then fit the Cox proportional hazards model with the treatment arm covariate, and obtain a coefficient estimate of 0.057 with a 95% CI (-0.294, 0.408). Based on the model fitting, there is little evidence for the effect of DPCA on the PBC patients' survival. Figure 5.4 shows the KM estimates of PBC patient survival, and there is no noticeable difference in $\log(-\log)$ survival between the DPCA and placebo groups. Model (5.1) with Z being treatment arm gives very similar survival estimates and CIs to those derived using the complementary log-log survival model containing only the treatment covariate. As seen in Figure 5.5, the estimated survival curves under model with and without treatment-time interaction coincide. This also holds true for the CI curves in Figure 5.6. From Figure 5.5 and 5.6, we also see that the model-based curves and the KM curves match well.

5.2 HIVNET 012 Data

5.2.1 Data Description

The HIVNET 012 randomized trial was conducted to assess the efficacy and safety of short-course nevirapine (NVP) versus short-course zidovudine (AZT) for the prevention of mother-to-child transmission of human immunodeficiency virus type-1 (HIV-1) in a breastfeeding population. A total of 626 HIV-1-infected women at more than 36 weeks' gestation were randomly assigned either NVP or AZT between November 1997 and April 1999 in Uganda and followed up until January 2001. A single dose of NVP was given orally to the women at the onset of labour. The oral NVP was also administrated only once to babies within

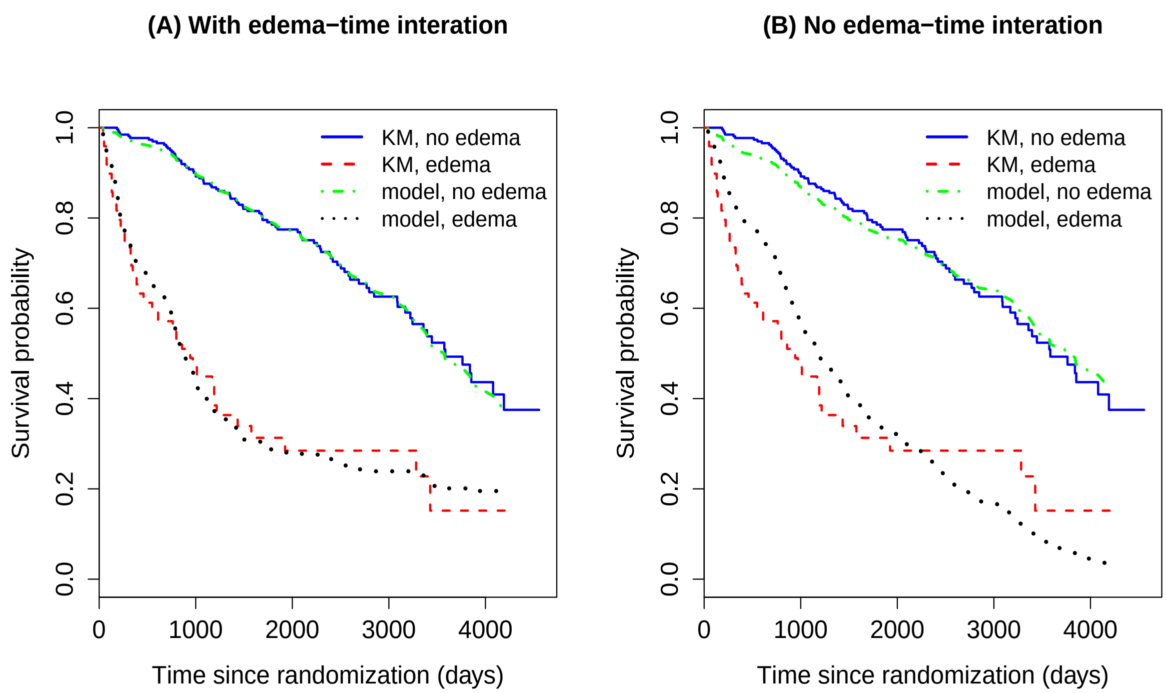

Figure 5.2: PBC Data. Kaplan-Meier and model-based estimates of survival functions for groups categorized by baseline edema status: (A) model containing edema and edema-time interaction; (B) model containing edema only.

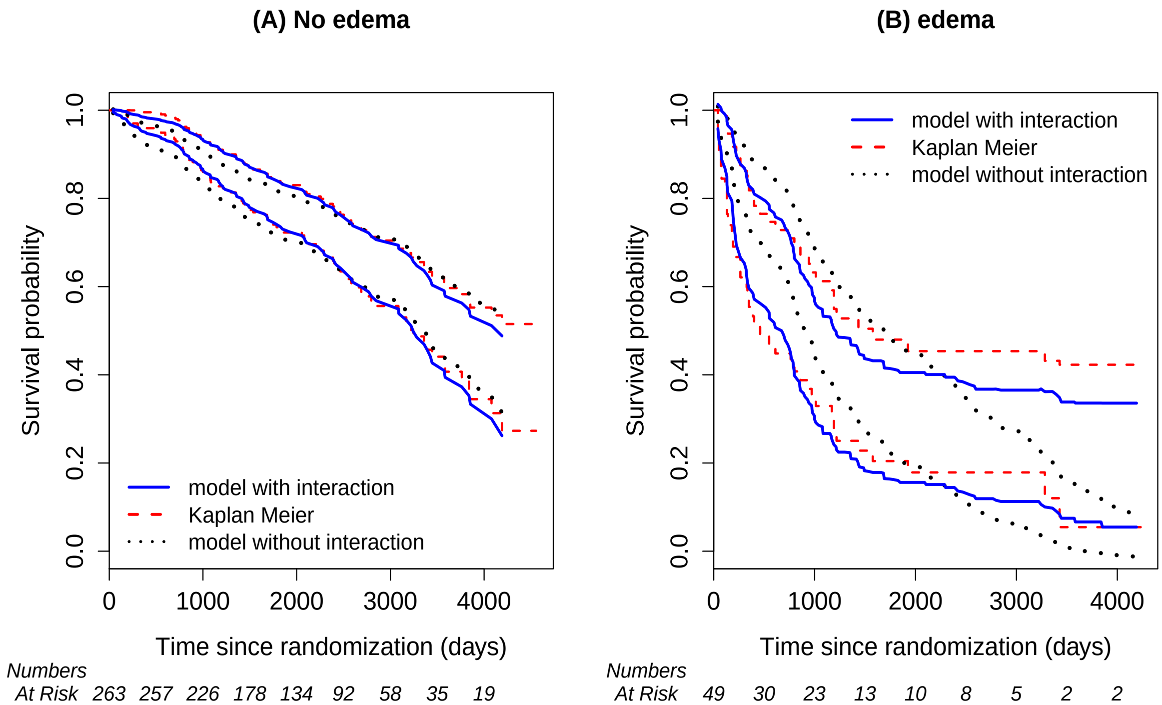

Figure 5.3: PBC Data. Comparison of three types of 95% confidence intervals for survival functions: (A) no edema group; (B) edema group. Kaplan-Meier, shown by dashed lines; model containing edema and edema-time interaction, shown by solid lines; model containing edema only, shown by dotted lines.

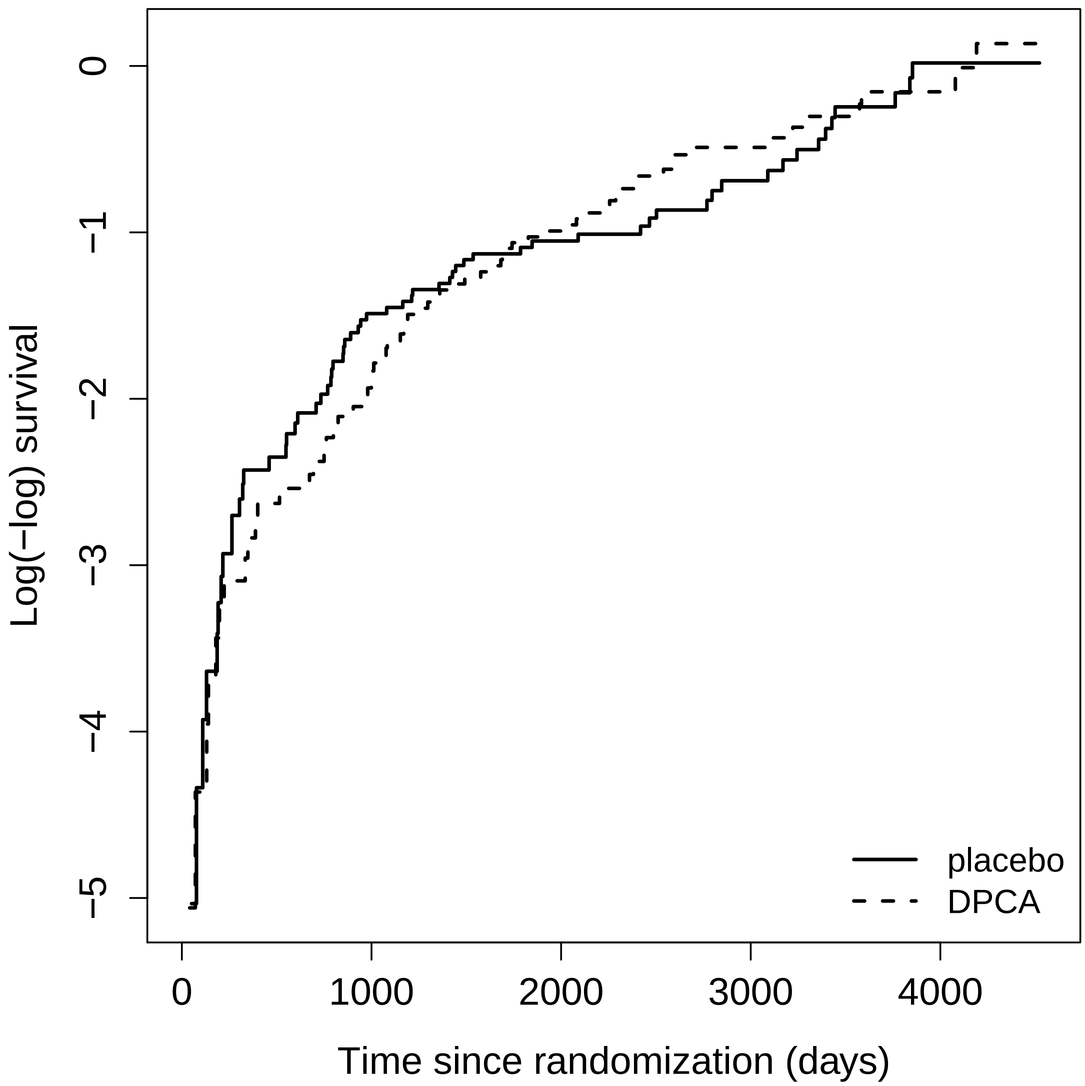

Figure 5.4: PBC Data. Log$(-\log)$ Kaplan-Meier estimates of patient survival for the DPCA group and the placebo group.

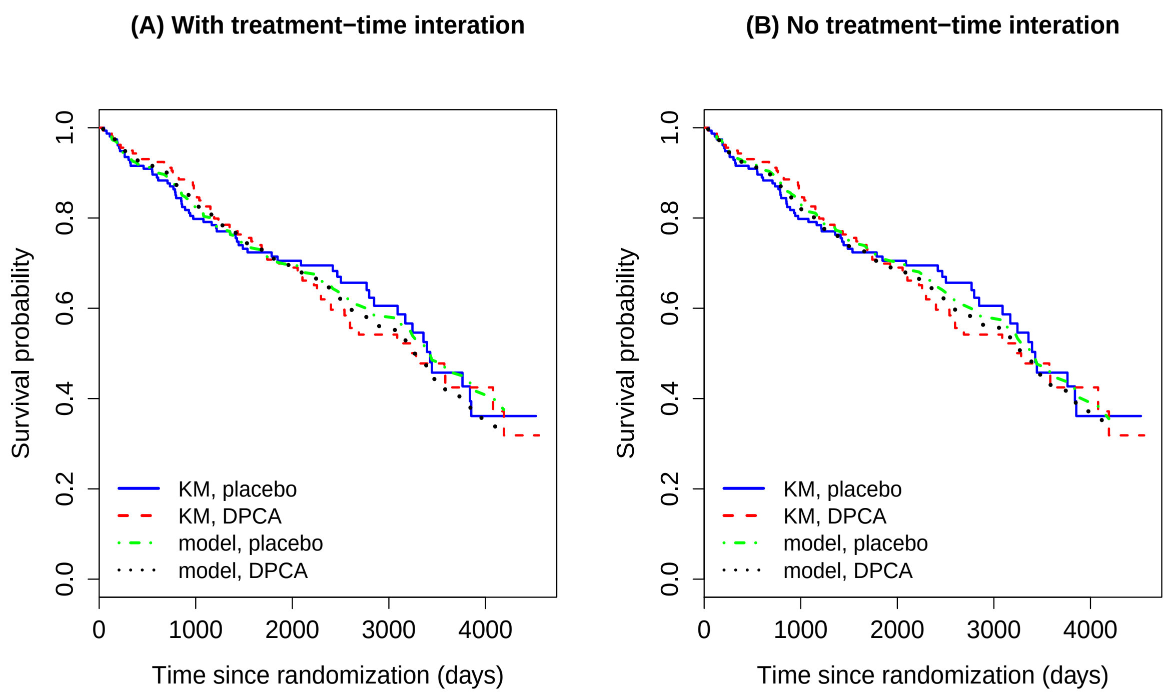

Figure 5.5: PBC Data. Kaplan-Meier and model-based estimates of survival functions for groups categorized by treatment arm: (A) model containing treatment and treatment-time interaction; (B) model containing treatment only.

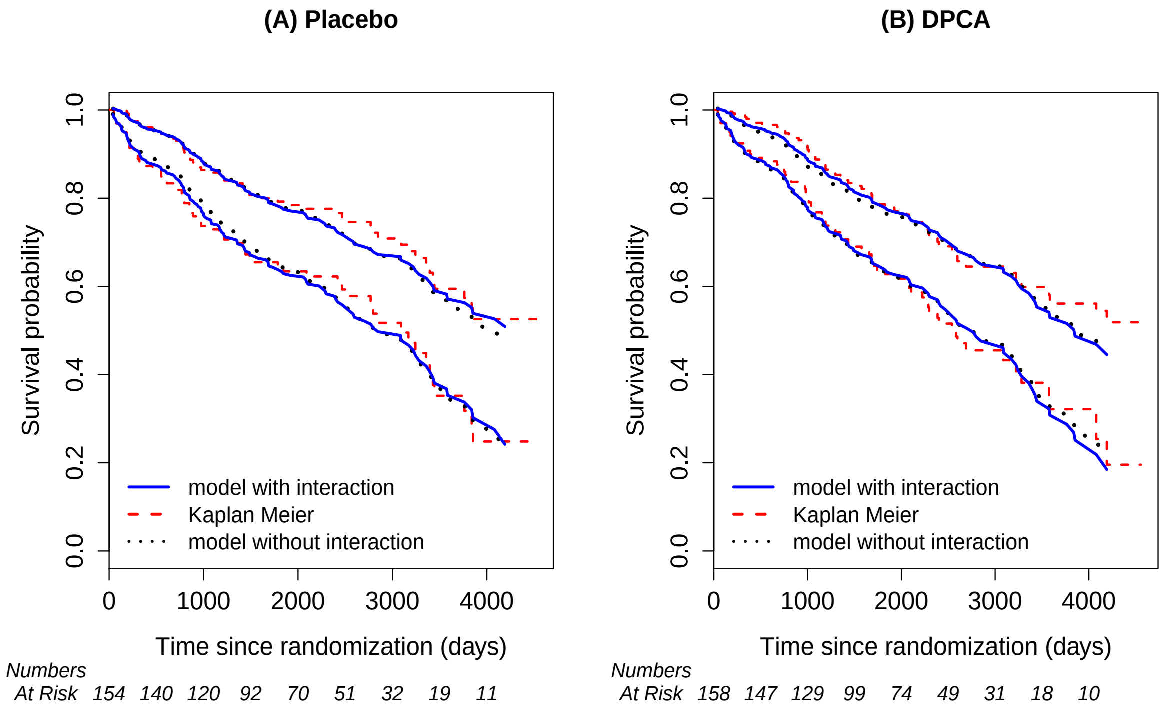

Figure 5.6: PBC Data. Comparison of three types of 95% confidence intervals for survival functions: (A) placebo group; (B) DPCA group. Kaplan-Meier, shown by dashed lines; model containing treatment and treatment-time interaction, shown by solid lines; model containing treatment only, shown by dotted lines.

72 hours after birth. The women in the AZT group took oral drug at the onset of labour and every three hours during labour; babies took oral AZT twice daily for seven days after birth. Complete demographic, medical and clinical records of the mothers and babies were collected before enrollment, on enrollment and over the follow-up period. The original trial results were reported by Guay et al. [1999] and Jackson et al. [2003] and they found that NVP was associated with reduced risk of HIV-1 transmission during the first 18 months of life compared with AZT.

Table 5.2 gives a description of infant and maternal characteristics by treatment arm. The study includes 619 mothers, with 308 in AZT group and 311 in NVP group. Table 5.2 shows that randomization is pretty fair in this study. Women on average have a gravidity history of three births. The distributions of maternal CD4+ counts and HIV-1 RNA viral load (VL) at baseline vary little across groups. The data set contains 632 babies, 25 of whom are twins or triplets. The descriptive statistics of all infants are similar to those of firstborn infants. For either AZT group or NVP group, approximately half are baby girls and the mean birthweight is around 3100 grams. The proportion of infant death through age 18 months is higher in the AZT group than that in the NVP group (13.3% vs. 9.6%). Also there is a higher proportion of HIV infection in babies of the AZT group compared to those of the NVP group (23.1% vs. 14.8%).

5.2.2 Data Analysis

While the trial showed a significant reduction of HIV-1-transmission risk among the infants born to their HIV-1-infected mothers taking the NVP, it is also important to understand whether or not the NVP would benefit the ultimate survival of the infants, at least through their early age 18 months. We first fit a Cox proportional hazards model with the treatment covariate. Under the Cox model, we estimate the hazard ratio for NVP-versus-AZT to be $e^{-0.37} = 0.69$, with a 95% CI $(0.43, 1.11)$, which does not show a statistically significant survival benefit of the NVP. We plot the $\log(-\log)$ KM survival estimates in Figure 5.7 to compare the 18-month infant survival between the two treatment arms. As shown in

Table 5.2: HIVNET 012 data. Characteristics of the mothers and babies by treatment arm.

	Treatment Arm	
	AZT (n=308)	NVP (n=311)
Maternal Characteristics		
Mean number of pregnancies (SD)	3.201 (1.685)	3.177 (1.630)
Twin or triplets birth delivery, n (%)	4 (1.3)	8 (2.6)
Mean CD4+ counts at baseline (SD)	463.3 (266.6)	481.5 (255.2)
Mean log10 viral load at baseline (SD)	4.385 (0.738)	4.351 (0.832)
Firstborn Infant Characteristics		
Baby girl, n (%)	156 (50.6)	154 (49.5)
Mean birthweight (SD), gram	3198 (460)	3080 (440)
Death within 18 months, n (%)	41 (13.3)	30 (9.6)
Mean time to death or censor (SD), month	15.63 (5.25)	16.49 (4.28)
Median time to death or censor (range), month	18 (0.033 - 18)	18 (0.033 - 18)
HIV infection within 18 months, n (%)	71 (23.1)	46 (14.8)
Mean time to HIV detection or censor (SD), day	391.2 (230.7)	440.1 (204.0)
Median time to HIV detection or censor (range),day	548 (1 - 548)	548 (1 - 548)

AZT, Zidovudine; NVP, Nevirapine; SD, standard deviation

Table 5.2: (continued)

	Treatment Arm	
	AZT (n=312)	NVP (n=320)
All Infant Characteristics		
Baby girl, n (%)	159 (51.0)	157 (49.1)
Mean birthweight (SD), gram	3189 (463)	3059 (460)
Death within 18 months, n (%)	42 (13.5)	33 (10.3)
Mean time to death or censor (SD), month	15.61 (5.28)	16.39 (4.45)
Median time to death or censor (range), month	18 (0.033 - 18)	18 (0.033 - 18)
HIV infection within 18 months, n (%)	71 (22.8)	47 (14.7)
Mean time to HIV detection or censor (SD), day	391.6 (230.6)	437.9 (205.9)
Median time to HIV detection or censor (range),day	548 (1 - 548)	548 (1 - 548)

AZT, Zidovudine; NVP, Nevirapine; SD, standard deviation

Figure 5.7, the AZT curve lies above the NVP curve over the entire observation period, but the difference between two treatment arms decreases over time. The trend of KM curves suggests that the NVP administration improved the infant survival through age 18 months compared to AZT but the survival benefit diminished over time. In fact, it is believed that the treatment effect of NVP might diminish as time progresses since only single-dose NVP was administered to the mothers and babies.

We consider only the data from the 619 firstborn babies and there are 308 firstborn babies in the AZT group. We fit four complementary log-log survival models to compare the 18-month infant survival between treatment groups. The response is time to death in months. 71 babies died within the 18-month follow-up, 514 babies were alive at age 18 months and other babies were censored at the time of their last HIV-1 test or lost to follow-up. The covariate of interest is the treatment group indicator of NVP versus AZT. Model I contains only the treatment indicator. Additionally, we include the treatment-time interaction in Model II and IV, and provide adjustment for birthweight and the base 10 logarithm of maternal VL at baseline in Model III and IV. The birthweight and baseline maternal VL are highly relevant to the infant's survival. The estimates of regression parameters from the model fitting are summarized in Table 5.3. Based on Model I and III, babies in the NVP group appear to experience survival benefit compared to those in the AZT group, though not significantly at the 0.05 level. As shown in Table 5.3, the treatment-time interaction in either Model II or IV is positive and significant, suggesting that the difference in $\log(-\log)$ survival between treatment groups decreases over time. We perform a Wald test on the coefficients of treatment indicator and treatment-time interaction to evaluate the significance of the treatment effect. The resultant p values are 0.040 and 0.021 for Model II and IV respectively, indicating that the NVP administration significantly improved the survival of babies born to HIV-1-infected mothers in less-developed countries compared with AZT. For illustration, Figure 5.8 shows the KM survival curves and the estimated survival curves at the median covariate values from Model IV for each treatment group, along with 95% CIs. The pattern of model-based curves are similar to that of KM curves. By overlaying Figure 5.8(B) over

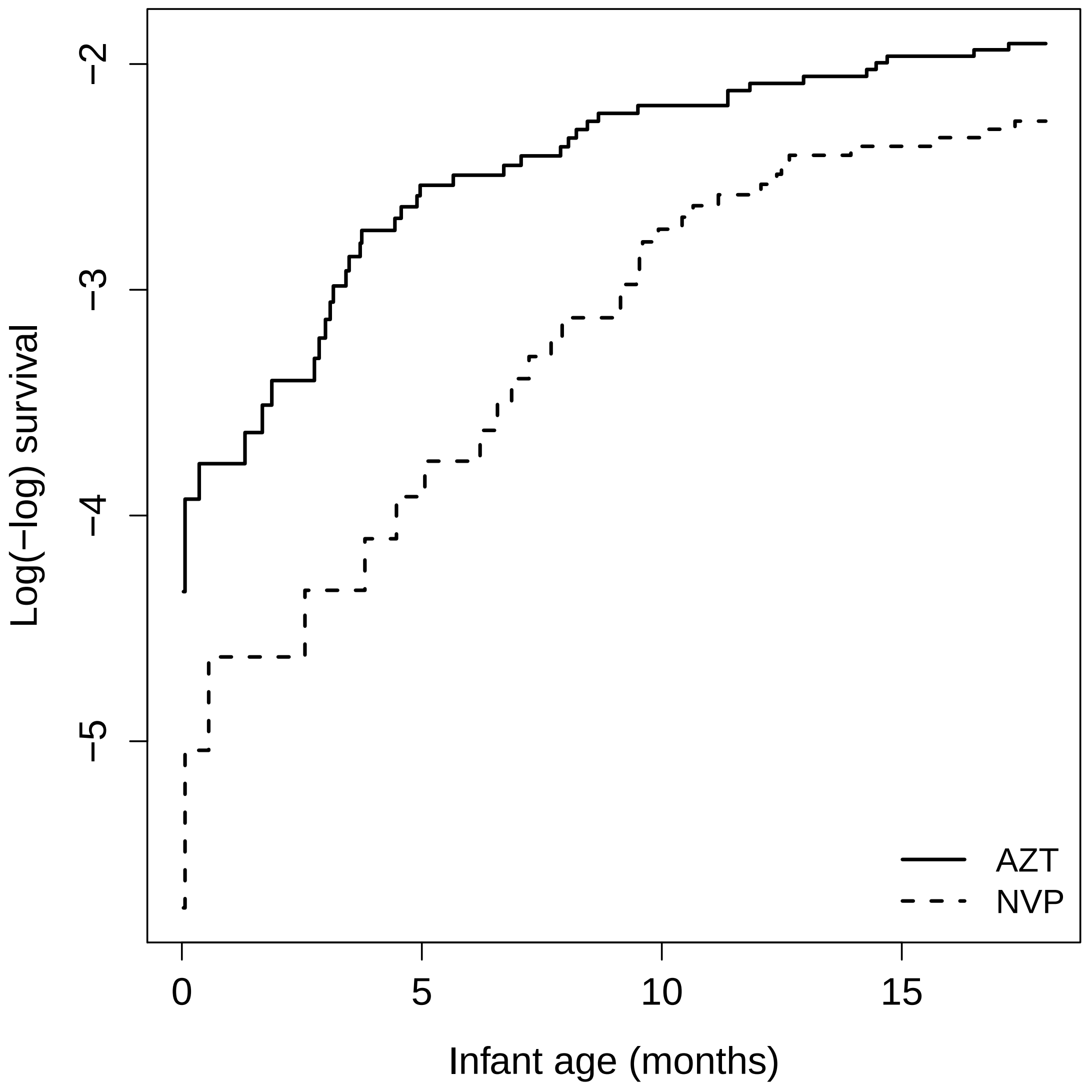

Figure 5.7: HIVNET 012 Data. Log(− log) Kaplan-Meier estimates of infant survival for the NVP group and the AZT group.

Figure 5.8(A), one can observe that the model-based curve is above the KM curve for either NVP or AZT. Figure 5.8(B) shows that infants in the NVP group with birthweight of 3100 grams and maternal VL of 25247 experience survival benefit compared to infants in the AZT group with birthweight of 3200 grams and maternal VL of 27800.

Table 5.3: HIVNET 012 data. Regression parameter estimates from fitting complementary log-log survival models.

Parameter	Est.	95% CI.	p
		Model I	
NVP	−0.37	(−0.84, 0.10)	0.124
		Model II	
NVP	−1.08	(−1.92, −0.24)	0.011
NVP × Time	4.18e-2	(1.31e-3, 8.22e-2)	0.043
		Model III	
NVP	−0.49	(−0.99, 0.01)	0.056
log10 VL	1.00	(0.64, 1.36)	<0.001
BW	−6.80e-4	(−1.25e-3, −1.07e-4)	0.020
		Model IV	
NVP	−1.24	(−2.13, −0.35)	0.006
log10 VL	1.00	(0.63, 1.36)	<0.001
BW	−6.68e-4	(−1.24e-3, −9.69e-5)	0.022
NVP × Time	4.51e-2	(1.49e-3, 8.86e-2)	0.043

Est., estimate; CI.,confidence interval; p, p value of the univariate Wald test; NVP, the treatment group indicator of NVP versus AZT; log10 VL, the base 10 logarithm of maternal HIV-1 RNA viral load at baseline; BW, birthweight in grams

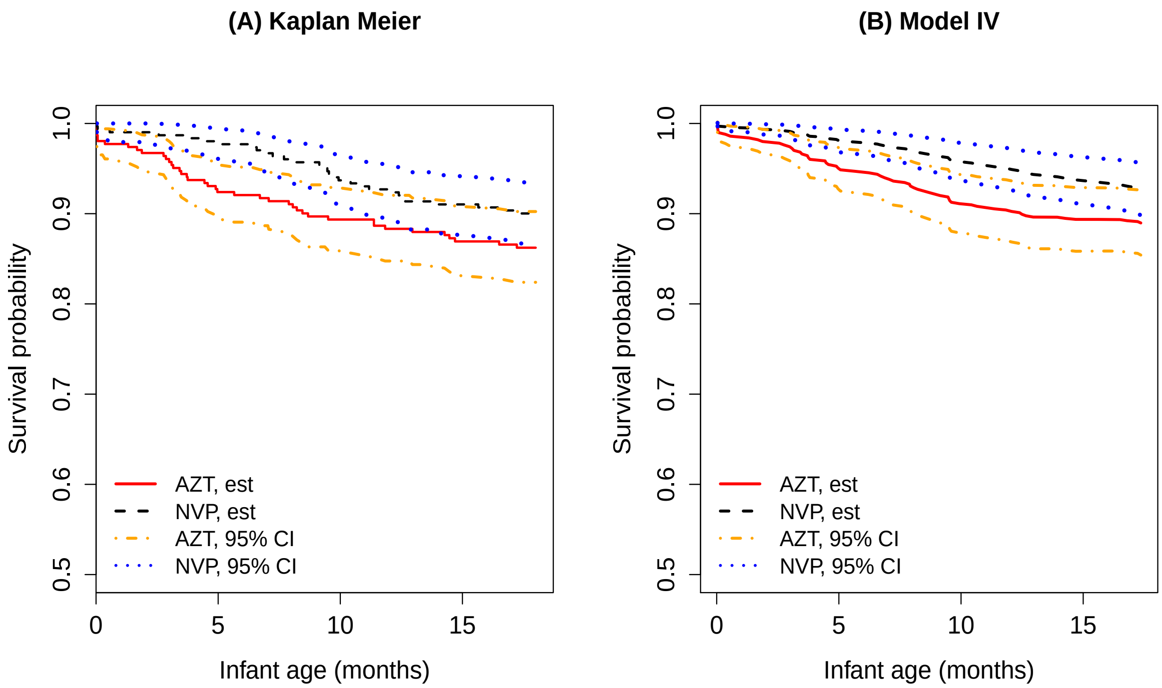

Figure 5.8: HIVNET 012 Data. (A) Kaplan-Meier estimates of survival functions for treatment groups, along with 95% confidence intervals. (B) Estimated survival curves from fitting Model IV: the solid and dash-dotted curves represent the point estimates and 95% confidence intervals of survival functions for babies in AZT group with birthweight of 3200 g and maternal viral load at baseline of 27800, respectively; the dashed and dotted curves represent the point estimates and 95% confidence intervals of survival functions for babies in NVP group with birthweight of 3100 g and maternal viral load at baseline of 25247, respectively.

Chapter 6

REMARKS

In this dissertation, we propose and study a novel complementary log-log survival model that includes the time-varying covariates. The proposed model is considered a natural expansion of the Cox proportional hazards model. In fact, when the covariates are all time-independent, the proposed model becomes the usual Cox model. In medical research, the Cox proportional hazards model has been widely used to assess a covariate's effect on censored time-to-event outcome, by estimating the associated hazard ratio. Per Lehamnn's alternative, the hazard ratio may be interpreted as "relative risk" for a time-independent covariate, e.g. treatment assignment in a randomized clinical trial. However, itself is not a direct measure of relative risk, particularly when the covariate is time-varying. One nice feature of the proposed model is its ability to provide direct interpretations of covariate effects on the survival function. Regression coefficients in the complementary log-log survival model are directly expressed in survival functions and can always be interpreted as relative risk, regardless of them being time-independent or time-varying. The addition of time-varying covariates leads to a more flexible model that can be of general use. The proposed model can also serve as a useful tool to check the proportional hazards assumption by assessing covariate-time interactions.

We develop semiparametric inference procedures for the model estimation. We estimate the regression parameters and the baseline survival function simultaneously from a set of martingale-based estimating equations. The resulting estimators possess desirable properties such as consistency, asymptotic normality and straightforward covariance estimation. Our methods yield the same results as the Cox proportional hazards model when covariates are all time-invariant. In the special case of time-independent covariates, our estimators for the model parameters are equivalent to the Breslow estimator for the cumulative baseline hazard function and the Cox maximum partial likelihood estimator; we also show the equivalence of our estimator for the asymptotic variance of the survival estimate to that proposed by Tsiatis [1981] using the Cox proportional hazards model. In addition, we develop weighted estimating equations that produce an efficient estimator for the regression coefficient in the proposed model. Moreover, we provide a point estimator and the corresponding pointwise confidence interval (CI) for the survival function at a given time point t for a subject with a particular set of covariates. In biomedical applications, there is often interest in estimating subject-specific survival curves. The proposed inference procedures for the estimation of regression parameters and survival functions are simple and reliable. We derive explicit expressions for the asymptotic variances of estimators that are straightforward for implementation.

The time-varying covariates in the model are limited to be "external". The external time-varying covariate is not dependent on an individual's survival process [Kalbfleisch and Prentice, 2002]. The internal time-varying covariate is related to the individual's survival status. Survival models with the internal time-varying covariates are not meaningful, because the existence of the internal time-varying covariates implies the subjects' survival status. In the example of HIVNET 012, the maternal CD4+ counts measured over the follow-up period is an external covariate since the values of CD4+ counts are generated by the mothers not by the infants; the infant CD4+ counts is considered "internal" since the values of CD4+ counts imply that the infants are still alive when the values are measured.

The model implicitly requires that $\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}(t)$ is increasing in t. One issue that arises in practice is that the estimated survival function $\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})$ may fail to be nonincreasing in t. We correct the monotonicity by replacing the problematic survival estimates with their smallest preceding values respectively. The modified estimator is defined as $\widehat{S}^*_n(t|\widetilde{\boldsymbol{Z}}) = \min_{s \leq t} \widehat{S}_n(s|\widetilde{\boldsymbol{Z}})$, which is asymptotically equivalent to the original estimator $\widehat{S}_n(t|\widetilde{\boldsymbol{Z}})$. The computation of proposed inference procedures involves the values of time-varying covariates at each observed failure time. This is not a problem for time-varying covariates of a known functional form, e.g. covariate-time interaction. However, in practice, it is rare to collect covariate measurements at all failure times for subjects over their entire observation periods. A reasonable remedy is to use some interpolation schemes [Fisher and Lin, 1999]. For example, one may impute the missing value using its previous observed covariate value.

The proposed pointwise CI of survival function maintains proper coverage at each fixed time point. In many applications, the confidence band is attractive as it is designed to guarantee the simultaneous coverage probability over the entire period of the data. Confidence bands for the Kaplan-Meier survival curve are usually based on transformation of the limiting process of the Nelson-Aalen estimator to the Brownian bridge [Hall and Wellner, 1980, Nair, 1984, Bie et al., 1987]. Lin et al. [1994] proposed a simulation approach to obtain confidence bands for the survival curves derived from the Cox proportional hazards model. The basic idea is to simulate a sample that has the same limiting distribution as the survival estimator. Due to the complicated limiting variance structure of our survival estimator, we may adapt the simulation procedures described by Lin et al. [1994] to our model setting for the construction of confidence bands. This will be developed in our future research.

The proposed weighted estimating equations may provide more efficient estimators for the regression parameters than the unweighted equations. However, the asymptotic covariance of the resultant estimators, even with optimal weights, may not achieve the semiparametric efficiency bound. In general, the nonparametric maximum likelihood estimators [Zeng and Lin, 2007] reach the semiparametric efficiency. In our future research, we shall develop the maximum likelihood estimation procedure for the complementary log-log survival model with time-varying covariates and compare its performance with the estimating equation procedures. We also assume a fixed sample of subjects throughout the study. In many clinical trials, patients may enter the study sequentially [Pocock, 1977]. Further research is needed to see if our proposed methods can be adapted to the group sequential designs.

The model proposed in this dissertation may be extended to a general transformation model by replacing the complementary log-log function by an unknown monotone function.

We assume the extended model as follows,

$$S(t|\widetilde{\boldsymbol{Z}}) = H\{\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}(t)\}, \tag{6.1}$$

where H is an unknown monotone function, $\alpha(t)$ is the baseline approximated by a Taylor expansion $(1+t+t^2)$, $\boldsymbol{Z}(t)$ is a p-dimensional vector of possibly time-varying covariates and $\boldsymbol{\beta}$ are the associated unknown regression parameters. Under model (6.1), the cumulative hazard function is of the form

$$\Lambda(t|\widetilde{\boldsymbol{Z}}) = -\log\{S(t|\widetilde{\boldsymbol{Z}})\} = -\log\left[H\{\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}(t)\}\right].$$

We plan to estimate the regression parameters from the martingale-based estimating equations, that is,

$$\sum_{i=1}^{n} \int_0^{\tau} [dN_i(t) + Y_i(t)\, d\log H\{\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}] \;=\; 0, \tag{6.2}$$

$$\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i(t)\, [dN_i(t) + Y_i(t)\, d\log H\{\alpha(t) + \boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}] \;=\; \mathbf{0}. \tag{6.3}$$

Apparently, we are not able to solve the equations straightforwardly and need to employ numerical approaches instead. A "naive" iterative program for computing the estimator $\widehat{\boldsymbol{\beta}}$ would suggest:

- Begin with an initial value $\boldsymbol{\beta}^{(0)}$
- Solve equations (6.2), (6.3) with $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$ to obtain $H^{(1)}(t, i, \boldsymbol{\beta})$, $t \in (0, \tau), i = 1, \ldots, n$
- Obtain $\boldsymbol{\beta}^{(1)}$ from $H^{(1)}(t, i, \boldsymbol{\beta})$, $t \in (0, \tau), i = 1, \ldots, n$
- Repeat until $\boldsymbol{\beta}^{(k)}$ satisfies the convergence criteria and set $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(k)}$

However, this "naive" program would be very complicated since the value of H depends on the values of $\boldsymbol{\beta}, t, i$, and the form of H is unknown. The difficulty lies in the determination of $\boldsymbol{\beta}$ value based on values of $H(t, i, \boldsymbol{\beta})$, $t \in (0, \tau), i = 1, \ldots, n$. The convergence may not be attained. Currently, we cannot solve this problem. More reasonable and efficient algorithms shall be explored in future.

BIBLIOGRAPHY

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

K. E. Atkinson. *An Introduction to Numerical Analysis*. New York: Wiley, 1989.

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

P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press, 1993.

O. Bie, O. Borgan, and K. Liestol. Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties. *Scandinavian Journal of Statistics*, 14:221–233, 1987.

N. E. Breslow. Discussion of paper of D. R. Cox. *Journal of the Royal Statistical Society, Series B*, 34:216–217, 1972.

I. V. Buckley and I. James. Linear regression with censored data. *Biometrika*, 66:429–436, 1979.

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

Y. Q. Chen, N. Hu, S. C. Cheng, P. Musoke, and L. P. Zhao. Estimating regression parameters in an extended proportional odds model. *Journal of the American Statistical Association*, 107:318–330, 2012.

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

S. C. Cheng, L. J. Wei, and Z. Ying. Predicting survival probabilities with semiparametric transformation models. *Journal of the American Statistical Association*, 92:227–235, 1997.

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

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

D. R. Cox and D. Oakes. *Analysis of Survival Data.* London: Chapman & Hall, 1984.

E. R. Dickson, P. M. Grambsch, T. R. Fleming, L. D. Fisher, and A. Langworthy. Prognosis in primary biliary cirrhosis: Model for decision making. *Hepatology*, 10:1–7, 1989.

B. Efron and I. M. Johnstone. Fisher's information in terms of the hazard rate. *Annals of Statistics*, 18:38–62, 1990.

L. D. Fisher and D. Y. Lin. Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health*, 20:145–157, 1999.

T. R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis.* New York: Wiley, 1991.

P. M. Grambsch, E. R. Dickson, M. Kaplan, G. Lesage, T. R. Fleming, and A. Langworthy. Extramural cross-validation of the Mayo primary biliary cirrhosis survival model establishes its generalizability. *Hepatology*, 10:846–850, 1989.

M. Greenwood. The natural duration of cancer. *Reports on Public Health and Medical Subjects*, 33:1–26, 1926.

L. A. Guay, P. Musoke, T. Fleming, D. Bagenda, M. Allen, C. Nakabiito, et al. Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial. *Lancet*, 354:795–802, 1999.

W. J. Hall and J. A. Wellner. Confidence bands for a survival curve from censored data. *Biometrika*, 67:133–143, 1980.

J. B. Jackson, P. Musoke, T. Fleming, L. A. Guay, D. Bagenda, M. Allen, et al. Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: 18-month follow-up of the HIVNET 012 randomised trial. *Lancet*, 362:859–868, 2003.

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data, 2nd edition*. New York: Wiley, 2002.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

D. Y. Lin and Z. Ying. Semiparametric analysis of the additive risk model. *Biometrika*, 81: 61–71, 1994.

D. Y. Lin, T. R. Fleming, and L. J. Wei. Confidence bands for survival curves under the proportional hazards model. *Biometrika*, 81:73–81, 1994.

R. Miller and J. Halpern. Regression with censored data. *Biometrika*, 69:521–531, 1982.

S. A. Murphy and A. W. Van der Vaart. On profile likelihood. *Journal of the American Statistical Association*, 95:449–465, 2000.

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

V. N. Nair. Confidence bands for survival functions with censored data: a comparative study. *Technometrics*, 26:265–275, 1984.

L. Peng and Y. Huang. Survival analysis with temporal covariate effects. *Biometrika*, 94: 719–733, 2007.

A. N. Pettitt. Inference for the linear model using a likelihood based on ranks. *Journal of the Royal Statistical Society, Series B*, 44:234–243, 1982.

S. J. Pocock. Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64:191–199, 1977.

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

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

A. W. Van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. New York: Springer Verlag, 1996.

L. J. Wei. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, 11:1871–1879, 1992.

S. Yang and R. L. Prentice. Semiparametric inference in the proportional odds regression model. *Journal of the American Statistical Association*, 94:125–136, 1999.

Z. Ying, L. J. Wei, and J. S. Lin. Prediction of survival probability based on a linear regression model. *Biometrika*, 79:205–209, 1992.

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

D. Zeng and D. Y. Lin. Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society, Series B*, 69:507–564, 2007.

VITA

Wenying Zheng was born in Zhoushan, Zhejiang, China. She received the Bachelor of Science in Statistics from Zhejiang University, China in 2007 and the Master of Science in Mathematics from University of Illinois at Chicago in 2009. She obtained her doctoral degree in Biostatistics from the Department of Biostatistics at University of Washington, Seattle in 2016.