STATISTICAL INFERENCE OF SEMIPARAMETRIC COX-AALEN TRANSFORMATION MODELS WITH FAILURE TIME DATA

by

Xi Ning

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2023

Approved by:

____
Dr. Yanqing Sun

____
Dr. Yinghao Pan

____
Dr. Qingning Zhou

____
Dr. Shen-En Chen

ABSTRACT

XI NING. Statistical Inference of Semiparametric Cox-Aalen Transformation Models with Failure Time Data. (Under the direction of DR. YANQING SUN and DR. YINGHAO PAN)

Failure time data are commonly encountered in epidemiological and biomedical studies, where the exact time of an event may be unknown or incomplete. Many semiparametric models have been developed in the literature to analyze such data; however, they may not always be effective in dealing with the diverse complexities that arise in practice. This motivated us to develop a more comprehensive class of semiparametric models for analyzing censored failure time data, with the ultimate goal of addressing the limitations of existing models and improving the accuracy of statistical inference.

In the first project, we propose a broad class of so-called Cox-Aalen transformation models that incorporate both multiplicative and additive covariate effects on the baseline hazard function through a transformation framework. The proposed model offers a high degree of flexibility and versatility, encompassing the Cox-Aalen model and transformation models as special cases. For right-censored data, we propose an estimating equation approach and devise an Expectation-Solving (ES) algorithm that involves fast and robust calculations. The resulting estimator is shown to be consistent and asymptotically normal via empirical process techniques. Moreover, the ES algorithm yields a computationally simple method for estimating the variance of both parametric and nonparametric estimators. Finally, we assess the performance of the proposed procedures by conducting extensive simulation studies and applying them in two randomized, placebo-controlled HIV prevention efficacy trials. The data example shows the utility of the proposed Cox-Aalen transformation models in enhancing statistical power for discovering covariate effects.

In the second project, we consider the regression analysis of the Cox-Aalen transformation models with partly interval-censored data, which comprise exact and interval-censored observations. We formulate a set of estimating equations and utilize an ES algorithm that guarantees stability and rapid convergence. Under regularity assumptions, we demonstrate that resulting estimators are consistent and asymptotically normal, and we propose the use of weighted bootstrapping techniques to estimate their variance consistently. To evaluate the proposed methods, we perform thorough simulation experiments and applied them to the analysis of the data from a randomized HIV/AIDS trial.

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to my advisors, Dr. Yanqing Sun and Dr. Yinghao Pan, for their exceptional guidance, infinite patience, and tremendous support throughout my Ph.D. journey in Biostatistics. Their vast knowledge and expertise in the field have been invaluable in helping me navigate the challenges and complexities of conducting this dissertation. Moreover, their steadfast commitment to advancing the frontiers knowledge in this area has been a constant source of inspiration to me, and has profoundly influenced my academic pursuits.

I would also like to extend my sincere thanks to my committee members, Dr. Qingning Zhou and Dr. Shen-En Chen, for their invaluable insights, comments, and suggestions throughout my research. Additionally, I am immensely grateful to Dr. Shaozhong Deng and Dr. Mohammad A. Kazemi for their support during my tenure as a teaching assistant in the Department of Mathematics and Statistics. My deep appreciation also goes to Dr. Peter B.Gilbert at Fred Hutchinson Cancer Center and the University of Washington for providing the datasets analyzed in this dissertation. I am thankful for all the wonderful people I met on this journey, especially my dear friends, who provided me with companionship and support, making this experience more manageable and enjoyable.

Furthermore, I would like to express my appreciation to the Graduate School for their financial support, including the 2022 Graduate School Summer/Teaching Fellowship Programs. I am also grateful to the National Institutes of Health NIAID grant R37-AI054165 and the National Science Foundation grant DMS-1915829 for their partial support of this dissertation.

Lastly, I wish to express my heartfelt gratitude to my parents, Zuogang Ning and Yanhua Cai, and my brother, Shuang Ning, for their unwavering love and sacrifices. Their constant support has enabled me to reach this point in my life. Additionally, I want to honor the memory of my beloved grandmother, Xiuzhen Xiao, whose love

and guidance were an integral part of my upbringing. Her values have helped shape the person I am today, and I will always carry her memory and the inspiration she provided with me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Failure time data, also known as survival data, is encountered in various areas or domains, such as medicine, social sciences, and finance. It represents the time between a specific starting point and the occurrence of an event of interest, such as death, machine failure, or disease development. Unlike traditional continuous or discrete data, failure time data is often censored, meaning that the event of interest may not have occurred for all subjects within the study at the time of analysis. Censoring can occur in several ways, with right-censored data being the most common. In right-censored data, the event of interest occurs after a certain observation time due to various reasons, such as the end of the study period or the loss of follow-up of the subjects. Another type of censoring is interval-censoring, which happens when the exact time of an event is not observed but is known to have occurred within a certain time interval. This can happen in situations such as periodic examinations or screenings, where the event of interest may have occurred between the last and current examination. Dealing with censored data imposes unique challenges for data analysis and interpretation in survival analysis.

Multiplicative and additive hazards models are two fundamental frameworks used to analyze censored failure time data, which assume that covariates affect the unspecified baseline hazard in a multiplicative and additive manner, respectively. In some cases, neither a strictly multiplicative nor additive model may be suitable. The multiplicative-additive hazards model offers a solution by integrating the strengths of both approaches, allowing the inclusion of covariates to have both multiplicative and additive effects. This model can capture complex relationships between covariates and the hazard function, including non-linear and interaction effects, and may provide improved predictions of the hazard function and survival probabilities.

The transformation model (Dabrowska and Doksum, 1988; Zeng and Lin, 2006) is alternative framework for analyzing failure time data, utilizing the methodology of transforming the failure time to improve the model fit or to satisfy modeling assumptions. One of the advantages of the transformation model is its flexibility in handling non-linear relationships between the covariates and the failure time variable. Additionally, the model can account for time-varying covariates and non-proportional hazards. However, it assumes a monotonic relationship between the transformed failure time variable and the covariates. In some cases, the relationship may not be monotonic, and a transformation model may not capture this relationship accurately.

This dissertation aims to enhance the existing literature on transformation models by proposing a more flexible and comprehensive approach. Specifically, we aim to develop a semiparametric transformation model that can handle a wider range of data scenarios and improve the accuracy of survival predictions. Chapter 2 presents the proposed model in detail, describing its structure and properties, and highlighting its advantages over existing models. In Chapter 3, we apply the proposed model to right-censored data, which is a common scenario in survival analysis where some observations have incomplete information on the event of interest. Additionally, we access the proposed model's performance against existing models, demonstrating its effectiveness in accurately estimating survival probabilities. Finally, we apply the proposed model and methods to simulation studies and two randomized HIV prevention efficacy trials. In Chapter 4, we extend the application of the proposed model to handle a mixture of exact and interval-censored observations, also known as the partly interval-censored data. We show how our proposed model can be adapted to handle this mixed type of data and demonstrate its performance via simulation studies and application to a randomized HIV/AIDS trial.

CHAPTER 2: THE SEMIPARAMETRIC COX-AALEN TRANSFORMATION MODELS

2.1 Introduction

Censored data are frequently encountered in epidemiological and biomedical studies. In the literature, the multiplicative and additive hazards models provide two principal frameworks for analyzing such data. The most popular multiplicative hazards model is the proportional hazards model (Cox, 1972), where the covariates are assumed to act multiplicatively on an unknown baseline hazard function. In contrast, the additive hazards models furnish an additive effect between the covariates and the baseline hazard function, providing a way to directly capture the increase or decrease in risk associated with the covariates (Aalen, 1980; Huffer and McKeague, 1991; Lin and Ying, 1994). Without prior domain knowledge, it is hard to determine which framework is preferable among multiplicative and additive hazards models. In fact, both models may often be used to complement each other and provide more complete insights. Therefore, various multiplicative-additive hazard models have been proposed to capture both multiplicative and additive effects (Lin and Ying, 1995; Martinussen and Scheike, 2002). In particular, Scheike and Zhang (2002) suggested a Cox-Aalen model by replacing the baseline hazard function in the Cox model with Aalen's additive model. The Cox-Aalen model has been studied for various types of censored data, e.g., right-censored (Scheike and Zhang, 2002), interval-censored (Boruvka and Cook, 2015), left-truncated and right-censored (Shen and Weng, 2018), left-truncated and mixed interval-censored (Shen and Weng, 2019), recurrent-event (Qu and Sun, 2019).

Transformation models have also received wide attention in the field of survival analysis. Dabrowska and Doksum (1988) introduced the class of linear transformation models, which includes the proportional hazards and proportional odds models (Pettitt, 1982; Bennett, 1983) as special cases. Estimators for this class of models were proposed by Dabrowska and Doksum (1988), Cheng et al. (1995), Fine et al. (1998), Chen et al. (2002), among others. Zeng and Lin (2006) later extended the linear transformation models to allow time-dependent covariates. Hereafter, we refer to this class of transformation models as Zeng and Lin's model to avoid confusion. There is rich literature investigating Zeng and Lin's model. Zeng and Lin (2006) proposed a nonparametric maximum likelihood estimator (NPMLE) in the presence of right-censored data. Zeng and Lin (2007) derived a system of self-consistent equations for the jump sizes of the baseline cumulative hazard function at exact failure times through an expectation-maximization (EM) algorithm. Chen (2009) showed that the self-consistent estimator derived in Zeng and Lin (2007) is asymptotically equivalent to a weighted Breslow-type estimator, which can be solved by a computationally-efficient iterative reweighting algorithm. More recently, Zeng et al. (2016) and Zhou et al. (2021) have investigated the nonparametric maximum likelihood estimation of Zeng and Lin's model in the context of interval-censored and partly interval-censored data, respectively.

However, one limitation of Zeng and Lin's model is that all covariate effects are assumed to be multiplicative within the transformation function. This assumption is too restrictive in some applications. For example, in an analysis of risk factors on mortality among patients with myocardial infarction, Scheike and Zhang (2003) showed that some covariates (e.g., ventricular fibrillation and congestive heart failure) have additive effects while others (e.g., age and sex) have multiplicative effects. In addition, they pointed out that naively treating all covariates as multiplicative led to incorrect results when predicting the survival probabilities. Another example is that in an HIV-1 vaccine study, HIV prevalence varies from region to region and by sex/gender; thus, the different regions/sex/gender subgroups have different baseline hazard functions (Corey et al., 2021). Moreover, a Kaplan-Meier plot shows that survival curves for different regions cross, suggesting an additive region effect. To the best of our knowledge, no existing work considers a class of semiparametric transformation models in which the baseline hazard function is allowed to depend on some potentially time-varying covariates additively. Therefore, it is desirable to provide a more powerful class of semiparametric transformation models that can accommodate both multiplicative and additive covariate effects under one unified framework.

The EM algorithm is a powerful tool for performing maximum likelihood estimation in the presence of latent variables or missing data (Dempster et al., 1977). In particular, various EM-type algorithms have been proposed to find NPMLE for semiparametric transformation models (Zeng and Lin, 2007; Liu and Zeng, 2013; Zeng et al., 2016; Zhou et al., 2021). In analogy to EM, Elashoff and Ryan (2004) proposed an expectation-solving (ES) algorithm that handles missing data for general estimating equations, greatly facilitating its application to a broader framework. When the complete-data estimating equations correspond to the score functions from the likelihood, the ES algorithm essentially reduces to the EM. The ES algorithm dramatically improves computational efficiency for solving estimating equations involving frailty or latent variables. For example, Johnson and Strawderman (2012) developed a smoothing expectation and substitution algorithm for the semiparametric accelerated failure time frailty model. Henderson and Rathouz (2018) considered an approximate EM procedure for a longitudinal latent class model for count data.

In this chapter, we propose a broad class of so-called Cox-Aalen transformation models that incorporate both multiplicative and additive covariate effects upon the baseline hazard function within a transformation. The proposed class of models is very flexible and contains Zeng and Lin's model and the Cox-Aalen model as special cases. However, the multiplicative-additive structure within the transformation and the need to estimate several nonparametric functions simultaneously impose additional challenges for model estimation. To alleviate such difficulties, we devise an ES-type algorithm, which iterates between an E-step wherein functions of complete data are replaced by their expectations and an S-step where these expected values are substituted into the complete-data estimating equations, which are then solved. More specifically, within the S-step, the high-dimensional parameters are updated using closed-form expressions, i.e., explicit calculations, while the low-dimensional parameters are updated via the Newton-Raphson method. Consequently, the proposed ES algorithm is fast and stable even under a high percentage censoring rate, as evidenced by our simulation studies and real data applications. Another attraction of our approach is that we provide simple variance estimators for both parametric and nonparametric estimates. Furthermore, the asymptotic properties of the proposed estimators are rigorously studied via modern empirical process techniques.

2.2 Model Description

Let $X(\cdot) = (X_1(\cdot), \ldots, X_q(\cdot))^\top$ and $Z(\cdot) = (Z_1(\cdot), \ldots, Z_d(\cdot))^\top$ denote $q \times 1$ and $d \times 1$ vectors of potentially time-varying covariates, and T denote the failure time of interest. We propose a broad class of so-called Cox-Aalen transformation models such that the cumulative hazard function for T conditional on covariates $X(\cdot)$ and $Z(\cdot)$ is represented as follows:

$$\Lambda(t \mid X(\cdot), Z(\cdot)) = G\left[\int_0^t \exp\{\beta^\top Z(s)\} d\Lambda_X(s)\right], \tag{2.1}$$

where β represents a $d \times 1$ vector of unknown regression parameters, $\Lambda_X(s) = \int_0^s X^\top(v)\alpha(v)dv$ is an unknown increasing function with $\alpha(v) = (\alpha_1(v), \ldots, \alpha_q(v))^\top$, and $G(\cdot)$ is a pre-specified transformation function that is strictly increasing and thrice continuously differentiable with $G(0) = 0$, $G'(0) > 0$ and $G(\infty) = \infty$. Here and hereafter, $G'(x) = dG(x)/dx$. In addition, let $A(t) = \int_0^t \alpha(s)ds = (A_1(t), \ldots, A_q(t))^\top$, where $A_j(t) = \int_0^t \alpha_j(v)dv$ for $j = 1, \ldots, q$. With X_1 fixed at 1, $\alpha_1(t)$ can be interpreted as a reference level of the risk. Generally, it is not meaningful to have that $X(\cdot)$ equals or is proportional to $Z(\cdot)$. For the choices of G, the class of frailty-induced transformations which is shown as below can be a useful tool:

$$G(x) = -\log \int_0^\infty \exp(-x\xi) f(\xi) d\xi, \tag{2.2}$$

where $f(\xi)$ is the density function of a non-negative random variable ξ with support $[0, \infty)$. The choice of the gamma density with unit mean and variance r for $f(\xi)$ yields the logarithmic transformations $G(x) = r^{-1}\log(1 + rx)$ $(r \geq 0)$ with $r = 0$ corresponding to $G(x) = x$. The class of Box-Cox transformations $G(x) = \{(1 + x)^\rho - 1\}/\rho$ can be obtained from the positive stable distribution with parameter $0 < \rho < 1$. Note that $G(x) = \log(1+x)$ is often considered as a member of the above class with $\rho = 0$. By treating the latent variable ξ as missing, the frailty-induced transformations are particularly useful in deriving EM-type algorithms (Zeng and Lin, 2007; Liu and Zeng, 2013; Zeng et al., 2016; Gao et al., 2018; Zhou et al., 2021). Some remarks regarding the Cox-Aalen transformation models are as follows:

Remark 2.1. *When* $\alpha_j(t) = 0$ $(j = 2, \ldots, q)$ *for any* t, *the right-hand side of* (2.1) *reduces to*

$$G\left[\int_0^t \exp\{\beta^\top Z(s)\} dA_1(s)\right]. \tag{2.3}$$

Hence, Zeng and Lin's model is a special case of the proposed models. Furthermore, when Z *is time-independent,* (2.3) *further reduces to the class of linear transformation models taking the form* $\log A_1(T) = -\beta^\top Z + \log G^{-1}(-\log \epsilon_0)$, *where* ϵ_0 *has a uniform distribution (Chen et al., 2002). Especially, the choices of* $G(x) = x$ *and* $G(x) = \log(1 + x)$ *yield the proportional hazards model and proportional odds model, respectively.*

Remark 2.2. *When* $G(x) = x$, *according to* (2.1), *the cumulative hazard function* *can be written as*

$$\int_0^t \left\{X^\top(s)\alpha(s)\right\} \exp\{\beta^\top Z(s)\}ds. \tag{2.4}$$

Thus, the conditional hazard function of T *is* $X^\top(t)\alpha(t)\exp\{\beta^\top Z(t)\}$. *Therefore, the Cox-Aalen model is a special case of the proposed models. In particular, when* $X_2, \ldots, X_q$ *represent levels in a set of factors, model* (2.4) *further reduces to the stratified Cox model (Kalbfleisch and Prentice, 2002).*

Remark 2.3. *Based on* (2.1)*, the odds of surviving beyond time* t *are*

$$\gamma(t \mid X, Z) = \frac{Pr(T > t \mid X, Z)}{Pr(T \leq t \mid X, Z)} = \{\Lambda_X(t)\}^{-1}\exp(-\beta^\top Z),$$

when $G(x) = \log(1 + x)$ *and* Z *are time-independent covariates. Consequently,*

$$\gamma(t \mid X, Z) = \gamma(t \mid X, Z_0)\exp\{-\beta^\top(Z - Z_0)\},$$

which is essentially a stratified proportional odds model.

As illustrated above, our proposed class of semiparametric models is very flexible and contains many popular models in survival analysis. To motivate our approach, we first set up the observed data likelihood and derive the NPMLE for a special case. Then, for more general situations, we propose estimating the parameters β and $A(\cdot)$ using estimating equations along with an easily-implemented ES algorithm.

CHAPTER 3: SEMIPARAMETRIC REGRESSION ANALYSIS OF THE COX-AALEN TRANSFORMATION MODELS WITH RIGHT-CENSORED DATA

3.1 Methods

3.1.1 Data Structure and Likelihood

For the ith individual, we let T_i and C_i be the failure time and censoring time, respectively. Let $\widetilde{T}_i = \min(T_i, C_i)$ be the observed time and define $\Delta_i = I(T_i \le C_i)$. Thus, $\Delta_i = 1$ indicates that the exact failure time for the ith individual was observed while $\Delta_i = 0$ implies censoring. For a random sample of n subjects, the observed data consists of $\mathcal{O}_i = \left\{\Delta_i, \widetilde{T}_i, X_i(t), Z_i(t), t \in [0, \tau]\right\}$ for $i = 1, \ldots, n$, where τ denotes the duration of the study. Moreover, we define $Y_i(t) = I(\widetilde{T}_i \ge t)$ and $N_i(t) = \Delta_i I(\widetilde{T}_i \le t)$.

Assume that T_i and C_i are conditionally independent given $X_i(\cdot)$ and $Z_i(\cdot)$. Under the proposed model (2.1), we can construct the likelihood function for the observed data as follows:

$$L_n(\beta, \Lambda_X) = \prod_{i=1}^{n} \left\{ \left(\Lambda'_{X_i}(\widetilde{T}_i) \exp\{\beta^\top Z_i(\widetilde{T}_i)\} G' \left[\int_0^{\widetilde{T}_i} \exp\{\beta^\top Z_i(s)\} d\Lambda_{X_i}(s) \right] \right)^{\Delta_i} \times \exp\left(-G\left[\int_0^{\widetilde{T}_i} \exp\{\beta^\top Z_i(s)\} d\Lambda_{X_i}(s) \right] \right) \right\}, \tag{3.1}$$

where $\Lambda'_X(\cdot)$ and $G'(\cdot)$ are the derivatives of $\Lambda_X(\cdot)$ and $G(\cdot)$, respectively. The likelihood (3.1) involves β and q infinite dimensional parameters A_j $(j = 1, \ldots, q)$, and it may not be concave in these parameters. Thus, the nonparametric maximum likelihood techniques are usually employed to restrict the parameter space.

3.1.2 Nonparametric Maximum Likelihood Estimation

To establish a simple and efficient estimation procedure, we adopt the idea in Zeng and Lin (2007) by treating ξ as a latent variable in the class of frailty-induced transformations (2.2). Note that model (2.1) is equivalent to the failure time T with

$$\Lambda(t \mid X(\cdot), Z(\cdot), \xi) = \xi \int_0^t \exp\{\beta^\top Z(s)\} d\Lambda_X(s), \tag{3.2}$$

because

$$\begin{aligned}
\Pr(T > t \mid X(\cdot), Z(\cdot)) &= E\left[\Pr\{T > t \mid X(\cdot), Z(\cdot), \xi\} \mid X(\cdot), Z(\cdot)\right] \\
&= E\left(\exp\left[-\xi \int_0^t \exp\left\{\beta^\top Z(s)\right\} d\Lambda_X(s)\right] \middle| X(\cdot), Z(\cdot)\right) \\
&= \int_0^\infty \exp\left[-\xi \int_0^t \exp\left\{\beta^\top Z(s)\right\} d\Lambda_X(s)\right] f(\xi) d\xi \\
&= \exp\left(-G\left[\int_0^t \exp\{\beta^\top Z(s)\} d\Lambda_X(s)\right]\right).
\end{aligned}$$

Based on model (3.2), it can be shown that the likelihood (3.1) is equivalent to the following

$$\prod_{i=1}^n \int_{\xi_i} \left(\left[\xi_i \Lambda'_{X_i}(\widetilde{T}_i) \exp\{\beta^\top Z_i(\widetilde{T}_i)\}\right]^{\Delta_i} \exp\left[-\xi_i \int_0^{\widetilde{T}_i} \exp\{\beta^\top Z_i(s)\} d\Lambda_{X_i}(s)\right] f(\xi_i)\right) d\xi_i.$$

Now we consider the nonparametric maximum likelihood estimation of β and $A(\cdot)$. Specifically, let $0 = t_0 < t_1 < \cdots < t_m < \infty$ denote the uniquely observed event times among n observations. Assume that the estimator for $A_j (j = 1, \ldots, q)$ is a step function with jump size a_{jk} at t_k for $k = 1, \ldots, m$ and define $a_{j0} = 0$. By observing that $d\Lambda_X(t) = X^\top(t) dA(t)$, thus, the estimator for Λ_X is a step function with jump size $X^\top(t_k) a_k$ at t_k where $a_k = (a_{1k}, \ldots, a_{qk})^\top$. Let $\mathcal{O}_i^C = \left\{\Delta_i, \widetilde{T}_i, X_i(t), Z_i(t), \xi_i, t \in [0, \tau]\right\}$ be the complete data for the ith subject.

The complete-data log-likelihood function can be written as

$$\sum_{i=1}^{n}\left\{\sum_{k=1}^{m}\left(\Delta_i I(\widetilde{T}_i = t_k)\left[\log\left\{\xi_i\left(X_{ik}^{\top}a_k\right)\right\} + \beta^{\top}Z_{ik}\right]\right)\right. \\ \left. -\xi_i \sum_{t_k\le \widetilde{T}_i}\exp(\beta^{\top}Z_{ik})(X_{ik}^{\top}a_k) + \log f(\xi_i)\right\}, \tag{3.3}$$

where $Z_{ik} = Z_i(t_k)$ and $X_{ik} = X_i(t_k)$.

To obtain the NPMLE of β and $A(\cdot)$, we propose an EM-type algorithm by treating ξ as missing data. In the E-step, we evaluate the posterior mean of ξ_i given the observed data, denoted by $\hat{E}(\xi_i)$. The detailed calculations are given in the next subsection (Section 3.1.3). In the M-step, we maximize the expectation of (3.3) conditional on the observed data. More specifically, we set the derivatives of the conditional expectation of (3.3) with respect to a_k ($k = 1, \ldots, m$) and β to zeros, respectively. Then one can solve for the estimates through the following equations:

$$\sum_{i=1}^{n}\left\{\Delta_i I(\widetilde{T}_i = t_k)\frac{X_{ik}}{X_{ik}^{\top}a_k} - I(\widetilde{T}_i \ge t_k)\hat{E}(\xi_i)\exp(\beta^{\top}Z_{ik})X_{ik}\right\} = 0, \quad \text{for } k = 1, \ldots, m \tag{3.4}$$

$$\sum_{i=1}^{n}\sum_{k=1}^{m}\left\{\Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \ge t_k)\hat{E}(\xi_i)(X_{ik}^{\top}a_k)\exp(\beta^{\top}Z_{ik})\right\}Z_{ik} = 0. \tag{3.5}$$

Notice that the dimension of unknown parameters $(a_1, \ldots, a_m, \beta)$ depends on m, which could be a large number when n is large or the censoring rate is low. Therefore, (3.4) and (3.5) are a system of high-dimensional nonlinear equations that is notoriously difficult to solve due to the curse of dimensionality. For a special case, i.e., X is a vector of design variables for categories, there exist explicit formulae for calculating the high-dimensional parameters a_k ($k = 1, \ldots, m$).

Here, we give further illustrations of the procedures in the M-step for the aforementioned special case. Let J be a categorical variable with q levels. Without loss of generality, we assume that D takes values in $\{1, \ldots, q\}$. Let $X = (1, X_2, \ldots, X_q)$

where $X_2, \ldots, X_q$ are group indicators, i.e., $X_2 = I(J = 2), \ldots, X_q = I(J = q)$. Here, $J = 1$ is considered as the reference group. We propose the following Gauss-Seidel method to jointly solve (3.4) and (3.5). Start with some initial values of the unknown parameters.

Step 1. Fix β, we update $a_k, (k = 1, \ldots, m)$ by solving (3.4). Note that for a fixed k, (3.4) can be written as

$$\begin{cases} \sum_{i=1}^n I(J_i = 1)\left\{\frac{\Delta_i I(\widetilde{T}_i = t_k)}{a_{1k}} - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\right\} = 0 \\ \sum_{i=1}^n I(J_i = 2)\left\{\frac{\Delta_i I(\widetilde{T}_i = t_k)}{a_{1k}+a_{2k}} - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\right\} = 0 \\ \vdots \\ \sum_{i=1}^n I(J_i = q)\left\{\frac{\Delta_i I(\widetilde{T}_i = t_k)}{a_{1k}+a_{qk}} - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\right\} = 0. \end{cases}$$

Hence, we obtain that

$$\begin{cases} a_{1k} = \frac{\sum_{i=1}^n \Delta_i I(J_i=1)I(\widetilde{T}_i=t_k)}{\sum_{i=1}^n I(J_i=1)I(\widetilde{T}_i\geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})} \\ a_{2k} = \frac{\sum_{i=1}^n \Delta_i I(J_i=2)I(\widetilde{T}_i=t_k)}{\sum_{i=1}^n \Delta_i I(J_i=2)I(\widetilde{T}_i\geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})} - a_{1k} \\ \vdots \\ a_{qk} = \frac{\sum_{i=1}^n \Delta_i I(J_i=q)I(\widetilde{T}_i=t_k)}{\sum_{i=1}^n I(J_i=q)I(\widetilde{T}_i\geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})} - a_{1k}. \end{cases} \tag{3.6}$$

Step 2. Fix $a_1, \ldots, a_m$, we update β by solving (3.5) using the Newton-Raphson method.

In the M-step, we iterate between steps 1 and 2 until convergence. The EM algorithm can be done when one alternates between E-step and M-step until convergence. However, such explicit formulae do not exist for more general scenarios; hence we consider an alternative estimating equation approach to overcome the aforementioned computational difficulties.

3.1.3 Estimating Equations

Following Elashoff and Ryan (2004), we develop an expectation-solving (ES) algorithm for model (2.1) in this section. We begin by constructing a system of complete-data estimating equations based on model (3.2), which has been shown to be equivalent to the proposed model (2.1). Note that the intensity for $N_i(t)$ is $Y_i(t)\xi_i \exp\{\beta^\top Z_i(t)\}X_i^\top(t)\alpha(t)$ if ξ_i is known. Let

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\xi_i \exp\{\beta_0^\top Z_i(s)\}X_i^\top(s)dA_0(s),$$

where (β_0, A_0) are the true values of (β, A). It is clear that $E\{X_i(t)dM_i(t)\} = 0$ for any $0 \le t \le \tau$ and $E\{\int_0^\tau Z_i(t)dM_i(t)\} = 0$. By treating ξ_i as missing, we consider the following complete-data estimating equations

$$\sum_{i=1}^n X_i(t)\left[dN_i(t) - Y_i(t)\xi_i \exp\{\beta^\top Z_i(t)\}X_i^\top(t)dA(t)\right] = 0 \qquad (0 \le t \le \tau), \qquad (3.7)$$

$$\sum_{i=1}^n \int_0^\tau Z_i(t)\left[dN_i(t) - Y_i(t)\xi_i \exp\{\beta^\top Z_i(t)\}X_i^\top(t)dA(t)\right] = 0. \qquad (3.8)$$

By the previous argued nonparametric techniques, i.e., the estimator for Λ_X is a step function with jump size $X^\top(t_k)a_k$ at t_k $(k = 1, \ldots, m)$, it is easy to note that (3.7) and (3.8) can be expressed as

$$\begin{cases} \sum_{i=1}^n \left\{\Delta_i I(\widetilde{T}_i = t_1) - I(\widetilde{T}_i \ge t_1)\xi_i(X_{i1}^\top a_1)\exp(\beta^\top Z_{i1})\right\} X_{i1} = 0 \\ \qquad\vdots \\ \sum_{i=1}^n \left\{\Delta_i I(\widetilde{T}_i = t_m) - I(\widetilde{T}_i \ge t_m)\xi_i(X_{im}^\top a_m)\exp(\beta^\top Z_{im})\right\} X_{im} = 0 \\ \sum_{i=1}^n \sum_{k=1}^m \left\{\Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \ge t_k)\xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\right\} Z_{ik} = 0. \end{cases} \qquad (3.9)$$

Write $\theta = (a_1^\top, \ldots, a_m^\top, \beta^\top)^\top$. We propose to estimate θ through an ES-type algorithm by treating ξ_i as missing. The ES algorithm iterates between an E-step wherein the functions of the complete data are replaced by their expectations, and an S-step where these expected values are substituted into the complete-data estimating equations (3.9), which are then solved. After specifying initial values of the unknown parameters θ, say $\theta^{(0)}$, the proposed ES algorithm iterates between the following two steps until convergence:

E-step. Evaluate the posterior means $\hat{E}(\xi_i)$. When $\Delta_i = 1$, the posterior density of ξ_i given the observed data $(\Delta_i = 1, T_i, X_i, Z_i)$ is proportional to $\xi_i \exp(-\xi_i S_{i1}) f(\xi_i)$, where $S_{i1} = \Delta_i \sum_{t_k \le T_i} (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik})$. Hence, we calculate

$$\hat{E}(\xi_i) = G^{'}(S_{i1}) - \frac{G^{''}(S_{i1})}{G^{'}(S_{i1})},$$

where $G'(\cdot)$ and $G''(\cdot)$ are the first and second derivatives of $G(\cdot)$, respectively. When $\Delta_i = 0$, the posterior density of ξ_i given the observed data $(\Delta_i = 0, C_i, X_i, Z_i)$ is proportional to $\exp(-\xi_i S_{i2}) f(\xi_i)$, where $S_{i2} = (1 - \Delta_i) \sum_{t_k \le C_i} (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik})$. One can obtain $\hat{E}(\xi_i) = G'(S_{i2})$. Therefore, the E-step can be summarized as

$$\hat{E}(\xi_i) = \Delta_i \left\{ G'(S_{i1}) - \frac{G''(S_{i1})}{G'(S_{i1})} \right\} + (1 - \Delta_i) G'(S_{i2}).$$

S-step. After replacing ξ by $\hat{E}(\xi_i)$, we solve (3.9) for θ. To this end, we propose the following nonlinear Gauss-Seidel method (Ortega and Rheinboldt, 1970; Ortega, 1972).

Step 1. Fix β, update a_k $(k = 1, \ldots, m)$ by solving

$$\begin{cases} \sum_{i=1}^n \left\{ \Delta_i I(\widetilde{T}_i = t_1) - I(\widetilde{T}_i \ge t_1) \hat{E}(\xi_i) (X_{i1}^\top a_1) \exp(\beta^\top Z_{i1}) \right\} X_{i1} = 0 \\ \vdots \\ \sum_{i=1}^n \left\{ \Delta_i I(\widetilde{T}_i = t_m) - I(\widetilde{T}_i \ge t_m) \hat{E}(\xi_i) (X_{im}^\top a_m) \exp(\beta^\top Z_{im}) \right\} X_{im} = 0. \end{cases} \tag{3.10}$$

Note that for fixed β, (3.10) is a system of linear equations in terms of a_k ($k = 1, \ldots, m$). The update of a_k is independent of updating a_j for $k \neq j$ ($k, j = 1, \ldots, m$). In particular, we have explicit formulae for updating a_k ($k = 1, \ldots, m$), i.e.,

$$a_k = \left\{ \sum_{i=1}^{n} I(\widetilde{T}_i \geq t_k) \hat{E}(\xi_i) \exp(\beta^\top Z_{ik}) X_{ik} X_{ik}^\top \right\}^{-1} \left\{ \sum_{i=1}^{n} \Delta_i I(\widetilde{T}_i = t_k) X_{ik} \right\}, \quad (3.11)$$

for $k = 1, \ldots, m$.

Step 2. Fix $a_1, \ldots, a_m$, we use the Newton-Raphson method to update β by solving the following equation:

$$\sum_{i=1}^{n} \sum_{k=1}^{m} \left\{ \Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k) \hat{E}(\xi_i)(X_{ik}^\top a_k) \exp(\beta^\top Z_{ik}) \right\} Z_{ik} = 0.$$

Within the S-step, we alternate between Step 1 and Step 2 until convergence is achieved. We employed a convergence criterion based on the absolute difference in the parameter estimates between two consecutive iterations. Specifically, we considered convergence to have been achieved when the absolute difference in the estimates fell below a predetermined threshold, such as 10^{-3}.

The E- and S-steps are iterated until convergence, and the resulting estimates are denoted by $\hat{\theta} = (\hat{a}_1^\top, \ldots, \hat{a}_m^\top, \hat{\beta}^\top)^\top$. A natural estimator of $A(t)$ is $\hat{A}(t) = \sum_{t_k \leq t} \hat{a}_k$ for $0 \leq t \leq \tau$. Moreover, recall that $A(t) = \int_0^t \alpha(s) ds$, hence we can estimate $\alpha(t)$, $0 \leq t \leq \tau$ via a kernel estimator

$$\hat{\alpha}(t) = \sum_{k=1}^{m} h^{-1} K\left(\frac{t - t_k}{h} \right) \hat{a}_k,$$

where $K(x)$ is the kernel function and h is the bandwidth. Throughout this paper, we choose the Epanechnikov kernel function, i.e., $K(x) = \frac{3}{4}\max\{1 - x^2, 0\}$.

The ES algorithm presented above has several advantageous characteristics. First, a closed-form formula for computing $\hat{E}(\xi_i)$ is obtained in the E-step. Second, the explicit calculation of the high-dimensional parameters a_k $(k = 1, \ldots, m)$ in the S-step avoids the need to solve a large system of nonlinear equations. Accordingly, the proposed ES algorithm performs stably and satisfactorily without calculating the inverse of any high-dimensional matrices. Third, when X is a vector of design variables for categories, the corresponding ES algorithm coincides with the EM algorithm proposed in Section 3.1.2 by observing that for fixed β, (3.4) and (3.10) share the same solution in terms of a_k $(k = 1, \ldots, m)$. This implies that the proposed ES estimator is also efficient under this special case. The justifications are presented in the next subsection. Similarly, it can be shown that the ES algorithm coincides with the EM algorithm when $X \equiv 1$, i.e., $q = 1$. Finally, we remark that (3.10) can be considered as a weighted version of (3.4), where each subject i receives a weight $X_{ik}^\top a_k$.

3.1.4 A Special Case

In this section, we show that when X is a vector of design variables for categories, the ES algorithm proposed in Section 3.1.3 coincides with the EM algorithm proposed in Section 3.1.2. To show this, we only need to show that for fixed β, equations (3.4) and (3.10) share the same solution in terms of a_k $(k = 1, \ldots, m)$. Note that for a fixed k, (3.10) can be written as

$$\sum_{i=1}^n \left\{ \Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik}) \right\} X_{ik} = 0,$$

which under this special case, is equivalent to

$$\begin{cases} \sum_{i=1}^n I(J_i = 1) \left\{ \Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik}) a_{1k} \right\} = 0 \\ \sum_{i=1}^n I(J_i = 2) \left\{ \Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik}) (a_{1k} + a_{2k}) \right\} = 0 \\ \vdots \\ \sum_{i=1}^n I(J_i = q) \left\{ \Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik}) (a_{1k} + a_{qk}) \right\} = 0. \end{cases} \tag{3.12}$$

It is easy to notice that (3.6) is the unique solution to (3.12) and thus also the unique solution to (3.10). In addition, we already showed in Section 3.1.2 that (3.6) is the unique solution to (3.4) under this special case. Thus, the ES and EM estimator coincide with each other when X is a vector of design variables for categories.

3.2 Variance Estimator

In this section, we provide easy-to-compute variance estimators for both the parametric estimates $\hat{\beta}$ and the nonparametric estimates $\hat{A}(t)$, $\hat{\alpha}(t)$. Note that $\hat{E}(\xi_i)$ is a function of the observed data $\mathcal{O}_i$ and the unknown parameter θ, i.e., $\hat{E}(\xi_i) = g(\mathcal{O}_i, \theta)$. Let $\mathcal{O}$ be the collection of $\mathcal{O}_i$ $(i = 1, \ldots, n)$, the proposed ES estimator is intrinsically equivalent to solving the following observed-data estimating equation: $U(\mathcal{O}, \theta) = 0$, where $U(\mathcal{O}, \theta) = (U_{a_1}, \ldots, U_{a_m}, U_\beta)$,

$$\begin{cases} U_{a_1} = \sum_{i=1}^n \left\{ \Delta_i I(\widetilde{T}_i = t_1) - I(\widetilde{T}_i \geq t_1) g(\mathcal{O}_i, \theta)(X_{i1}^\top a_1) \exp(\beta^\top Z_{i1}) \right\} X_{i1} \\ \vdots \\ U_{a_m} = \sum_{i=1}^n \left\{ \Delta_i I(\widetilde{T}_i = t_m) - I(\widetilde{T}_i \geq t_m) g(\mathcal{O}_i, \theta)(X_{im}^\top a_m) \exp(\beta^\top Z_{im}) \right\} X_{im} \\ U_\beta = \sum_{i=1}^n \sum_{k=1}^m \left\{ \Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k) g(\mathcal{O}_i, \theta)(X_{ik}^\top a_k) \exp(\beta^\top Z_{ik}) \right\} Z_{ik}. \end{cases} \tag{3.13}$$

Note that $U_{a_1}, \ldots, U_{a_m}, U_\beta$ also depend on the observed data $\mathcal{O}$ and the unknown parameter θ. Here, we compress the notation when there is no confusion. From (3.13), one can easily note that $U(\mathcal{O}, \theta)$ can be expressed as the sum of independent terms:

$$U(\mathcal{O}, \theta) = \sum_{i=1}^n U_i(\mathcal{O}_i, \theta).$$

Let $D(\mathcal{O}, \theta)$ be the derivative of $U(\mathcal{O}, \theta)$ with respect to θ, then the covariance matrix of $\hat{\theta}$ is consistently estimated by

$$D(\mathcal{O}, \theta)^{-1} \left\{ \sum_{i=1}^n U_i(\mathcal{O}_i, \theta) U_i^\top(\mathcal{O}_i, \theta) \right\} \{D(\mathcal{O}, \theta)^{-1}\}^\top \Bigg|_{\theta = \hat{\theta}}. \tag{3.14}$$

Therefore, the variance covariance matrix of $\hat{\beta}$ can be consistently estimated by the $d \times d$ lower right-hand corner of (3.14). The variance covariance matrix of $\hat{a}_k$ ($k = 1, \ldots, m$) can be consistently estimated by the $(qm) \times (qm)$ upper left-hand corner of (3.14).

In addition, recall that $\hat{A}(t) = \sum_{t_k \le t} \hat{a}_k$ and $\hat{\alpha}(t) = \sum_{k=1}^{m} h^{-1} K\left(\frac{t-t_k}{h}\right) \hat{a}_k$ for $0 \le t \le \tau$, the variance for $\hat{A}(t)$ and $\hat{\alpha}(t)$ are

$$\mathrm{Var}\left\{\hat{A}(t)\right\} = \sum_{t_k \le t} \sum_{t_j \le t} \mathrm{Cov}(\hat{a}_k, \hat{a}_j),$$

$$\mathrm{Var}\left\{\hat{\alpha}(t)\right\} = \sum_{k=1}^{m} \sum_{j=1}^{m} h^{-2} K\left(\frac{t-t_k}{h}\right) K\left(\frac{t-t_j}{h}\right) \mathrm{Cov}(\hat{a}_k, \hat{a}_j).$$

Variance estimators are obtained by replacing $\mathrm{Cov}(\hat{a}_k, \hat{a}_j)$ by $\widehat{Cov}(\hat{a}_k, \hat{a}_j)$ in the above expressions.

3.3 Asymptotic Properties

In this section, we present the asymptotic properties of the proposed ES estimator. Let $\phi(t) = G'(t)$, $\psi(t) = G''(t)/G'(t)$ and

$$\rho(t; \beta, A) = \int_0^t Y(s) \exp\{\beta^\top Z(s)\} X^\top(s) dA(s).$$

Hence, we express the posterior mean of ξ as

$$g(\tau; \beta, A) = \phi(\rho(\tau; \beta, A)) - \Delta \psi(\rho(\tau; \beta, A)).$$

Let P and $\mathbb{P}_n$ denote the true probability and denote the empirical measure, respectively. In addition, let $\theta = (\beta, A)$ be the parameters of interest and $\theta_0 = (\beta_0, A_0)$ be the true values of the parameters. Then the proposed ES estimator $\hat{\theta} = (\hat{\beta}, \hat{A})$ is essentially a Z-estimator solving the following observed-data estimating equation

$$\mathbb{P}_n\Phi(\beta, A)(t) \equiv \mathbb{P}_n \begin{pmatrix} \Phi_1(\beta, A) \\ \Phi_2(\beta, A)(t) \end{pmatrix} = 0,$$

for $0 \leq t \leq \tau$, where

$$\Phi_1(\beta, A) = \int_0^\tau \left[Z(t)dN(t) - Y(t)\exp\{\beta^\top Z(t)\}g(\tau; \beta, A)Z(t)X^\top(t)dA(t) \right],$$

$$\Phi_2(\beta, A)(t) = X(t)dN(t) - Y(t)\exp\{\beta^\top Z(t)\}g(\tau; \beta, A)X(t)X^\top(t)dA(t).$$

Let h be a function in $BV_1[0, \tau]$, where $BV_1[0, \tau]$ denotes the set of functions with total variation bounded by 1 on $[0, \tau]$. Define

$$\Phi_2(\beta, A)[h] = \int_0^\tau h(t) \left[X(t)dN(t) - Y(t)\exp\{\beta^\top Z(t)\}g(\tau; \beta, A)X(t)X^\top(t)dA(t) \right].$$

Similar to Gao et al. (2017) and van der Vaart and Wellner (1996a, Section 3.3.1), the proposed ES estimator $(\hat{\beta}, \hat{A})$ is equivalent to the root of the estimating equation

$$\mathbb{P}_n\Phi(\beta, A)[h] \equiv \mathbb{P}_n \begin{pmatrix} \Phi_1(\beta, A) \\ \Phi_2(\beta, A)[h] \end{pmatrix} = 0,$$

for all $h \in BV_1[0, \tau]$.

Write $\mathbf{\Psi}(\theta) = P\Phi(\beta, A)[h]$ and $\mathbf{\Psi}_n(\theta) = \mathbb{P}_n\Phi(\beta, A)[h]$. Note that $\mathbf{\Psi}(\theta)$ and $\mathbf{\Psi}_n(\theta)$ are actually h-dependent. Rigorously speaking, we should write $\mathbf{\Psi}(\theta)[h] = P\Phi(\beta, A)[h]$ and $\mathbf{\Psi}_n(\theta)[h] = \mathbb{P}_n\Phi(\beta, A)[h]$, but in the rest of the article, we suppress the letter h when there is no confusion. The proposed ES estimator is a Z-estimator that satisfies $\mathbf{\Psi}_n(\hat{\theta}) = 0$. To establish the asymptotic properties, we assume the following regularity conditions:

Condition 1. With probability one, $X(\cdot)$ and $Z(\cdot)$ have bounded total variation in

$[0, \tau]$.

Condition 2. Let $\mathcal{B}$ be a compact set of $\mathcal{R}^d$ and $BV[0, \tau]$ be the class of functions with bound variation over $[0, \tau]$. The true parameter (β_0, A_0) belongs to $\mathcal{B} \times BV^q[0, \tau]$ with β_0 an interior point of $\mathcal{B}$ and $A_0(t) = (A_{01}(t), \ldots, A_{0q}(t))^\top$ is continuous over $[0, \tau]$ with $A_0(0) = 0$. Here $BV^q[0, \tau]$ denotes the product space $BV[0, \tau] \times \cdots \times BV[0, \tau]$.

Condition 3. With probability one, there exists a positive constant a such that $P(Y(\tau) = 1 \mid X(\cdot), Z(\cdot)) > a$ and $PN^2(\tau) < \infty$. If there exists a vector Γ and a deterministic function $\Gamma_0(t)$ such that $\Gamma_0(t) + \Gamma^\top X(t) = 0$ with probability one, then $\Gamma_0(t) = 0$ and $\Gamma = 0$ for any $t \in [0, \tau]$.

Condition 4. The transformation function $G(x)$ is thrice continuously differentiable on the interval $[0, \infty)$ and satisfy the following: $G(0) = 0$, $G'(x) > 0$ and $G(\infty) = \infty$.

Condition 5. Let $\dot{\mathbf{\Psi}}_{\theta_0}$ be the Fréchet derivative of $\mathbf{\Psi}(\theta)$ with respect to θ at $\theta = \theta_0$. See (A.4) and (A.5) in Appendix A for detailed expressions of $\dot{\mathbf{\Psi}}_{\theta_0}$. We assume that $\dot{\mathbf{\Psi}}_{\theta_0}$ is an invertible map.

Remark 3.1. *Conditions 1 and 2 state the boundedness of the covariates and the compactness of the Euclidean parameter space, which are conventional conditions used in most regression analyses. Condition 3 ensures the existence and uniqueness of the jump sizes in* (3.11). *Condition 4 ensures that the function G is strictly increasing on* $[0, \infty)$. *Condition 5 is a classical condition for Z-estimators.*

Theorem 3.1. *Under Conditions* $1-5$, *the proposed ES estimator* $(\hat{\beta}, \hat{A})$ *is strongly consistent to* (β_0, A_0).

Theorem 3.2. *Under Conditions* $1-5$, $\sqrt{n}(\hat{\beta} - \beta_0, \hat{A} - A_0)$ *converges weakly to a zero-mean Gaussian process in the metric space* $\mathcal{R}^d \times \overline{lin}^q(BV_1[0, \tau])$.

Here, $\overline{\text{lin}}(BV_1[0, \tau])$ denotes the closed linear span for linear functionals of $BV_1[0, \tau]$. For each j $(j = 1, \ldots, q)$, A_j belongs to the Banach space $\overline{\text{lin}}(BV_1[0, \tau])$, where $A_j[h] =$

$\int h(t)dA_j(t)$ for $h \in BV_1[0,\tau]$. Thus, $A = (A_1, \ldots, A_q)^\top$ belongs to the Banach space $\overline{\text{lin}}^q(BV_1[0,\tau])$. Here, $\overline{\text{lin}}^q(BV_1[0,\tau])$ stands for the product space $\overline{\text{lin}}(BV_1[0,\tau]) \times \cdots \times \overline{\text{lin}}(BV_1[0,\tau])$. Detailed proofs of the above theorems are presented in Appendix A.

3.4 Simulation Studies

First, we conducted extensive simulation studies to assess the accuracy and reliability of the estimation and inference methods we proposed. Assume that the failure time T follows the Cox-Aalen transformation model given by:

$$\Lambda(t) = G\left[\int_0^t \exp\{\beta_1 Z_1(s) + \beta_2 Z_2\} d\Lambda_X(s)\right].$$

Here, $Z_1(t) = B_1 I(t \leq V) + B_2 I(t > V)$ is a time-dependent covariate where B_1 and B_2 are independent Ber(0.5), $V \sim$ Unif(0, 3), and $Z_2 \sim$ Unif(0, 1) is a time-independent covariate. We chose $\beta_1 = 0.5$ and $\beta_2 = -0.5$ as the values for the regression coefficients, and consider three different configurations for $\Lambda_X(s) = \int_0^s X^\top(v)dA(v)$, $A(t) = (A_1(t), \ldots, A_q(t))^\top$:

Scenario 1. $X = (1, X_2)^\top$ with $X_2 \sim$ Ber(0.4), $A_1(t) = \log(1 + t/4)$ and $A_2(t) = 0.1t$.

Scenario 2. $X = (1, X_2)^\top$ with $X_2 \sim$ Unif(0, 1), $A_1(t) = \log(1 + t/4)$ and $A_2(t) = 0.1t$.

Scenario 3. $X(t) = (1, X_2(t))^\top$ with $X_2(t) = B_3 + B_4 t$, where $B_3 \sim$ Unif(1, 2) and $B_4 \sim$ Unif(0.1, 0.5), $A_1(t) = \log(1 + t/4)$ and $A_2(t) = 0.1t$.

Scenario 4. Let J be a categorical variable that takes values in $\{1, 2, 3\}$ with equal probability. $X = (1, X_2, X_3)^\top$, where $X_2 = I(J = 2)$, $X_3 = I(J = 3)$, $A_1(t) = \log(1 + t/4)$, $A_2(t) = 0.1t$ and $A_3(t) = 0.05t$.

For function $G(\cdot)$, we consider the the class of logarithmic transformations $G(x) = r^{-1}\log(1 + rx)$ with $r = 0, 0.5$ and 1, where $r = 0$ yields the Cox-Aalen model. For all setups, we let $\tau = 1$ be the duration of the study. For each study participant, we generate one censoring time $C \sim$ Exponential(0.5). We set $\Delta = 1$ if $T \leq \min(C, \tau)$, and 0 otherwise. This process yields about $75\% \sim 85\%$ right-censored observations for $r = 0$, 0.5 and 1. We initialized the proposed ES algorithm for each simulated dataset with $\beta = 0$ and $a_k = (1/m, 0, \ldots, 0)$ for $k = 1, \ldots, m$. We also experimented with other initial values for β and a_k, but the resulting estimates were virtually indistinguishable from those obtained using the initial values specified above. We performed 1000 simulation replicates for each of the sample sizes $n = 200$, 500, and 800.

Table 3.1 presents the estimation results for β_1 and β_2 across Scenarios $1 - 4$. Despite the high censoring percentage, the results in Table 3.1 indicate that the proposed estimation methods exhibit favorable performance in multiple respects: (i) the proposed estimators are nearly unbiased; (ii) the estimated standard errors are in good agreement with the empirical standard errors; (iii) the empirical coverage probability of 95% confidence intervals are all close to the nominal 95% level; (iv) increasing the sample size reduces the bias and variability of the parameter estimator. Thus, our proposed estimating procedures are reliable for various Cox-Aalen transformation models.

In addition, Figure 3.1 presents the estimation results for the cumulative regression functions $A_1(\cdot)$ and $A_2(\cdot)$ in Scenario 1. The proposed estimators are again virtually unbiased and the estimated curves are able to capture the shapes of the true cumulative regression functions very well; the estimated standard errors are close to the empirical standard errors; the confidence intervals have reasonable coverage probabilities. Throughout all scenarios, we chose bandwidth $h = 0.1$ for the kernel estimator $\hat{\alpha}(t)$. Figure 3.2 shows the estimation results for $\alpha(\cdot)$ under Scenario 1. Figure 3.3 and 3.4 give the estimation results for $A(\cdot)$ and $\alpha(\cdot)$ under Scenario 2, respectively. Similarly, Figure 3.5 and 3.6 present the estimation results for $A(\cdot)$ and $\alpha(\cdot)$ under Scenario 4, respectively. These results further confirm the satisfactory performance

Table 3.1: Estimation results of the regression parameter β under scenarios 1 to 4

		$\beta_1 = 0.5$				$\beta_2 = -0.5$			
r	n	Bias	SE	SEE	CP	Bias	SE	SEE	CP
					Scenario 1				
0	200	0.003	0.350	0.340	0.950	−0.014	0.587	0.574	0.947
	500	0.007	0.212	0.212	0.952	−0.005	0.354	0.359	0.956
	800	−0.003	0.172	0.167	0.948	−0.001	0.279	0.285	0.951
0.5	200	−0.001	0.380	0.369	0.949	−0.014	0.639	0.630	0.950
	500	0.007	0.227	0.230	0.961	−0.002	0.388	0.395	0.957
	800	−0.004	0.185	0.181	0.951	−0.001	0.302	0.313	0.951
1	200	−0.002	0.402	0.402	0.956	−0.020	0.690	0.699	0.957
	500	0.008	0.246	0.252	0.957	0.002	0.415	0.437	0.968
	800	−0.005	0.198	0.198	0.946	−0.003	0.323	0.347	0.957
					Scenario 2				
0	200	0.004	0.337	0.334	0.954	−0.002	0.574	0.563	0.942
	500	0.006	0.208	0.209	0.962	−0.004	0.348	0.353	0.960
	800	−0.002	0.169	0.164	0.945	−0.006	0.275	0.280	0.949
0.5	200	−0.002	0.363	0.362	0.951	−0.014	0.624	0.620	0.948
	500	0.007	0.223	0.227	0.954	−0.000	0.384	0.389	0.963
	800	−0.003	0.184	0.179	0.946	−0.003	0.300	0.308	0.948
1	200	−0.005	0.389	0.396	0.957	−0.019	0.672	0.690	0.958
	500	0.007	0.240	0.248	0.957	0.002	0.412	0.432	0.965
	800	−0.004	0.197	0.195	0.944	−0.005	0.325	0.342	0.961
					Scenario 3				
0	200	0.005	0.288	0.287	0.958	−0.009	0.494	0.486	0.947
	500	0.002	0.180	0.180	0.949	−0.004	0.305	0.305	0.953
	800	−0.001	0.146	0.141	0.949	−0.009	0.245	0.241	0.950
0.5	200	0.001	0.319	0.321	0.954	−0.009	0.552	0.555	0.953
	500	0.001	0.198	0.202	0.956	−0.004	0.348	0.348	0.953
	800	−0.001	0.160	0.159	0.947	−0.009	0.276	0.276	0.942
1	200	0.010	0.343	0.365	0.964	−0.002	0.624	0.647	0.962
	500	−0.002	0.216	0.229	0.959	−0.002	0.380	0.403	0.963
	800	−0.006	0.175	0.180	0.963	−0.003	0.305	0.318	0.960
					Scenario 4				
0	200	0.010	0.339	0.334	0.951	−0.011	0.572	0.562	0.938
	500	0.005	0.211	0.209	0.948	−0.003	0.348	0.353	0.958
	800	−0.000	0.169	0.164	0.948	−0.000	0.280	0.279	0.948
0.5	200	0.007	0.366	0.362	0.946	−0.014	0.615	0.619	0.947
	500	0.003	0.228	0.227	0.950	−0.005	0.378	0.389	0.968
	800	0.001	0.183	0.179	0.950	−0.001	0.308	0.307	0.949
1	200	0.002	0.389	0.396	0.951	−0.008	0.657	0.690	0.960
	500	0.002	0.244	0.248	0.957	−0.002	0.412	0.433	0.967
	800	0.001	0.195	0.196	0.956	0.000	0.330	0.341	0.954

Note: Bias, bias of the parameter estimator; SE, empirical standard error of the parameter estimator; SEE, mean of the standard error estimator; CP, empirical coverage percentage of the 95% confidence interval.

Table 3.2: Estimation results of the regression parameter β with a misspecified $r = 0$ under the logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$

		$\beta_1 = 0.5$				$\beta_2 = -0.5$			
n	r_{true}	Bias	SE	SEE	CP	Bias	SE	SEE	CP
					Scenario 3				
500	0	0.002	0.180	0.180	0.949	−0.004	0.305	0.305	0.953
	0.5	−0.032	0.186	0.185	0.950	0.034	0.318	0.315	0.951
	1	−0.059	0.192	0.192	0.943	0.067	0.329	0.326	0.947
	1.5	−0.082	0.199	0.197	0.933	0.086	0.332	0.335	0.950
	2	−0.098	0.205	0.201	0.917	0.101	0.344	0.343	0.946
	2.5	−0.118	0.212	0.206	0.906	0.117	0.357	0.352	0.932
	3	−0.133	0.216	0.210	0.888	0.129	0.363	0.359	0.936
2000	0	0.001	0.088	0.089	0.951	−0.006	0.149	0.152	0.965
	0.5	−0.031	0.094	0.092	0.933	0.033	0.165	0.157	0.938
	1	−0.058	0.098	0.095	0.904	0.058	0.165	0.162	0.941
	1.5	−0.078	0.097	0.098	0.872	0.084	0.167	0.167	0.929
	2	−0.094	0.100	0.100	0.839	0.103	0.168	0.171	0.907
	2.5	−0.110	0.102	0.102	0.812	0.117	0.171	0.175	0.904
	3	−0.122	0.105	0.105	0.786	0.130	0.175	0.179	0.895

Note: Bias, bias of the parameter estimator; SE, empirical standard error of the parameter estimator; SEE, mean of the standard error estimator; CP, empirical coverage percentage of the 95% confidence interval.

of our proposed method in various numerical settings.

We also conducted simulation studies to investigate the robustness of the proposed estimator under the misspecification of the G function. The setups were the same as Scenario 3, and Table 3.2 reports the parameter estimation resultswith r misspecified as 0 while the data is generated from $r = r_{true}$. Here, r_{true} can be any value from $\{0, 0.5, 1, 1.5, 2, 2.5, 3\}$. It is easy to see that the misspecification of r values led to biased estimates and lower coverage probabilities than the nominal levels, even though the proposed variance estimators can accurately reflect the true variations.

Second, we show the advantages of the proposed model over Zeng and Lin's model through simulation studies. Specifically, we considered Scenario 5 as shown below, and generated the failure time T from the following Cox-Aalen transformation model

$$\Lambda(t \mid X, Z) = G\Big\{ \int_0^t \exp(\beta_1 Z_1) X^\top dA(s) \Big\}. \tag{3.15}$$

Scenario 5. $\beta_1 = 1$, $Z_1 \sim \text{Ber}(0.5)$ and $X(t) = (1, X_2)^\top$ with $X_2 \sim \text{Ber}(0.4)$. $A_1(t) = 0.1t + t^4/2$ and $A_2(t) = -2t^3/3 + t^2/2$.

Clearly, in model (3.15), Z_1 has a multiplicative effect while X_2 has an additive effect. If we naively treat all covariate effects as multiplicative and fit Zeng and Lin's model, we will obtain biased survival probability predictions. Figure 3.7 illustrates this bias. We also compared the predicted cumulative hazards from the proposed model and Zeng and Lin's model, displayed in Figure 3.8. When $Z_1 = 0$, the cumulative hazards for groups $X_2 = 0$ and $X_2 = 1$ intersect, indicating that the cumulative hazard in $X_2 = 1$ group is initially larger than the $X_2 = 0$ group, but becomes smaller later in the study. However, using Zeng and Lin's model, the cumulative hazard in $X_2 = 0$ is consistently larger than the group $X_2 = 1$. Hence, the proposed model can more accurately capture the complexity of the cumulative hazards when there exist additive covariate effects.

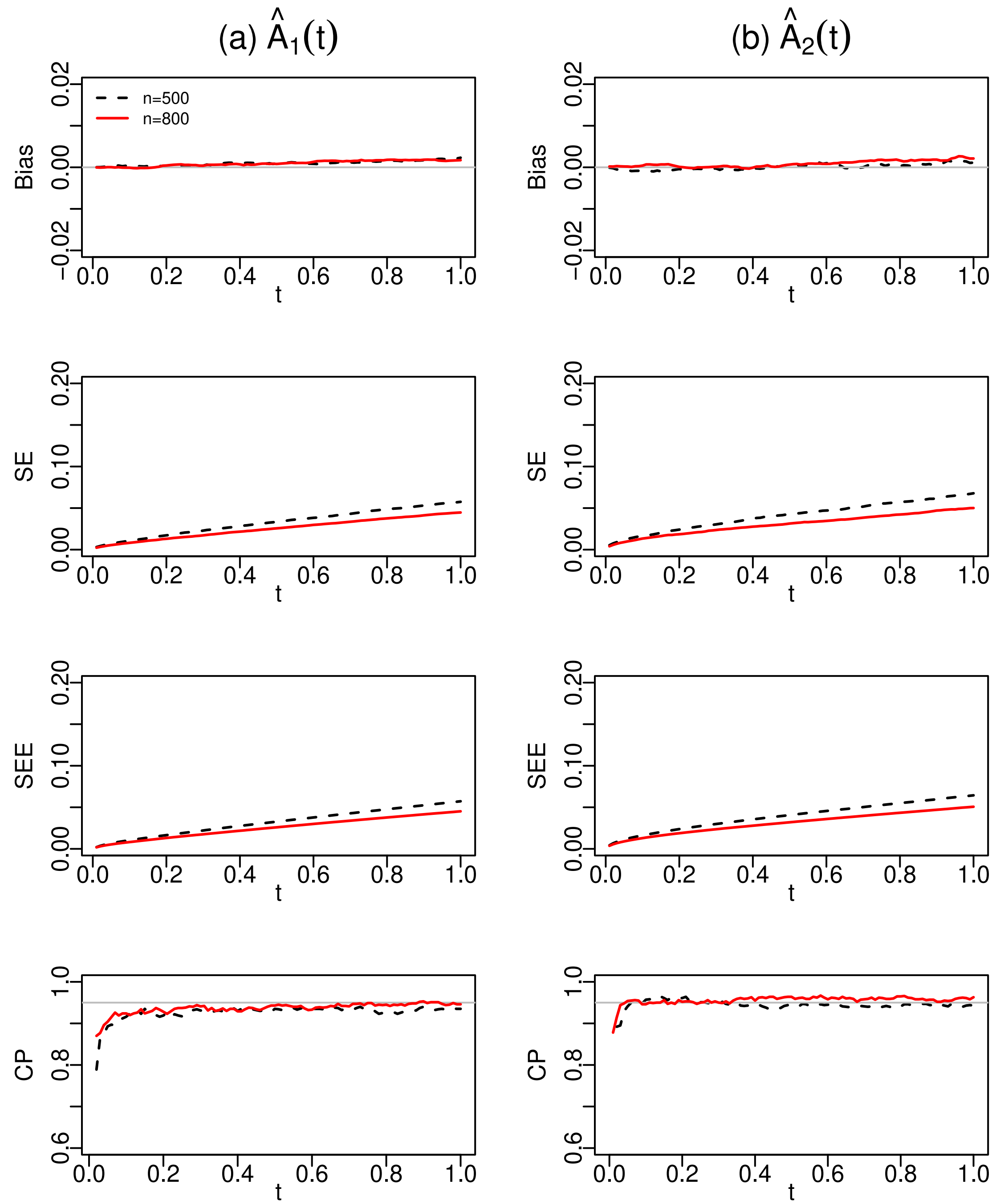

Figure 3.1: Estimation results for (a) $A_1(t) = \log(1 + t/4)$ and (b) $A_2(t) = 0.1t$ in Scenario 1 with $r = 0$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.

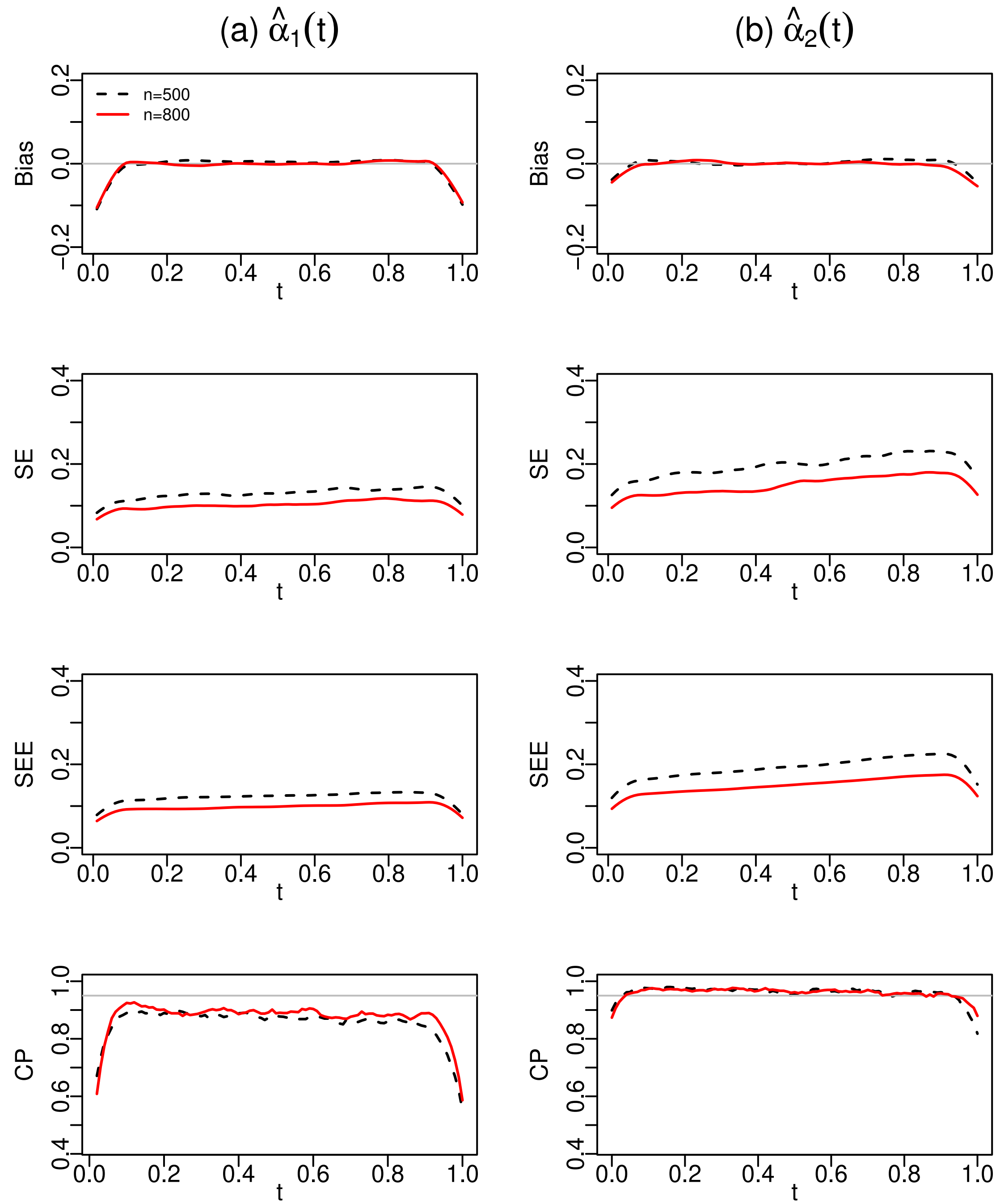

Figure 3.2: Estimation results for (a) $\alpha_1(t) = 1/(4+t)$ and (b) $\alpha_2(t) = 0.1$ in Scenario 1 with $r = 0$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.

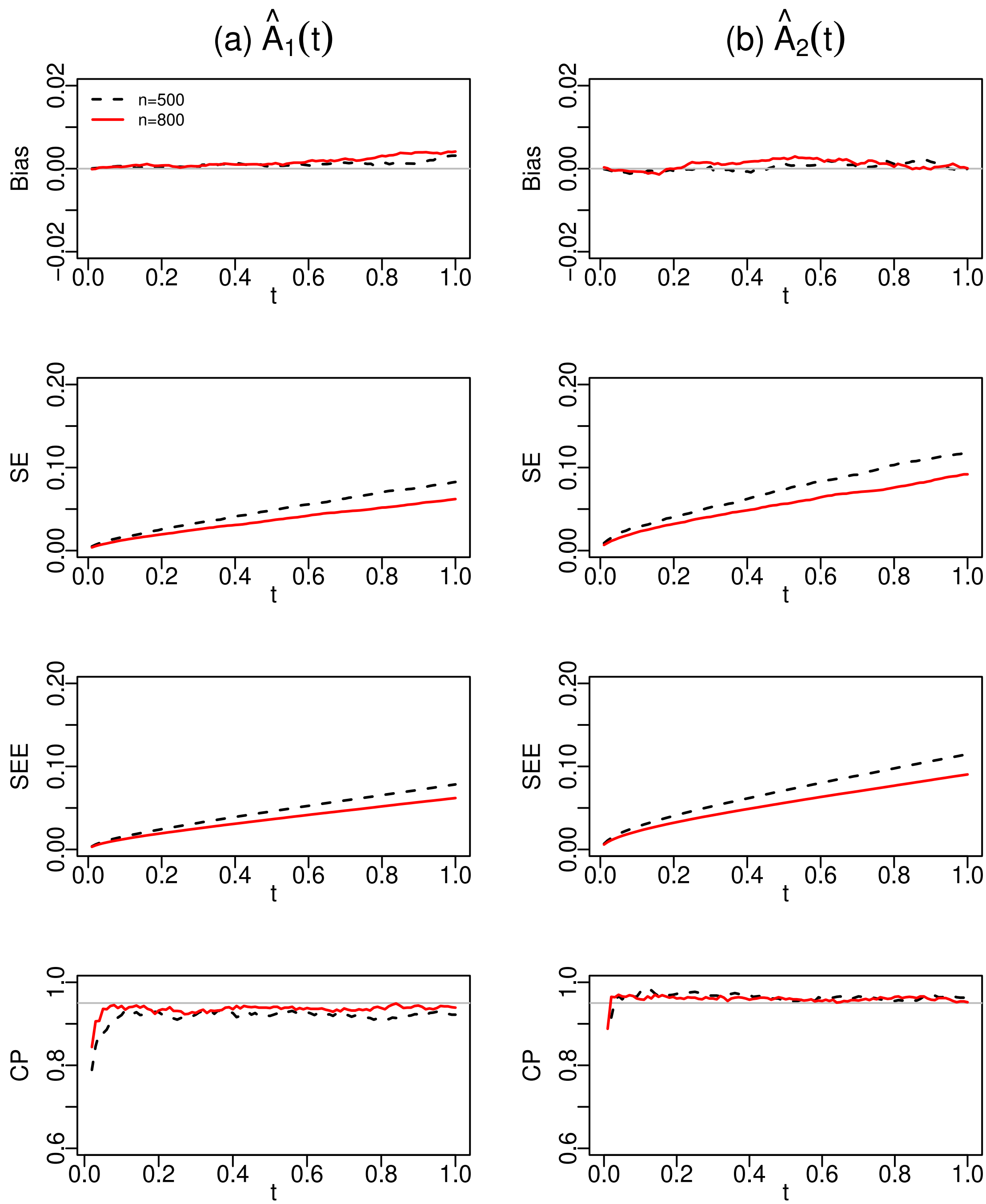

Figure 3.3: Estimation results for (a) $A_1(t) = \log(1 + t/4)$ and (b) $A_2(t) = 0.1t$ in Scenario 2 with $r = 0.5$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.

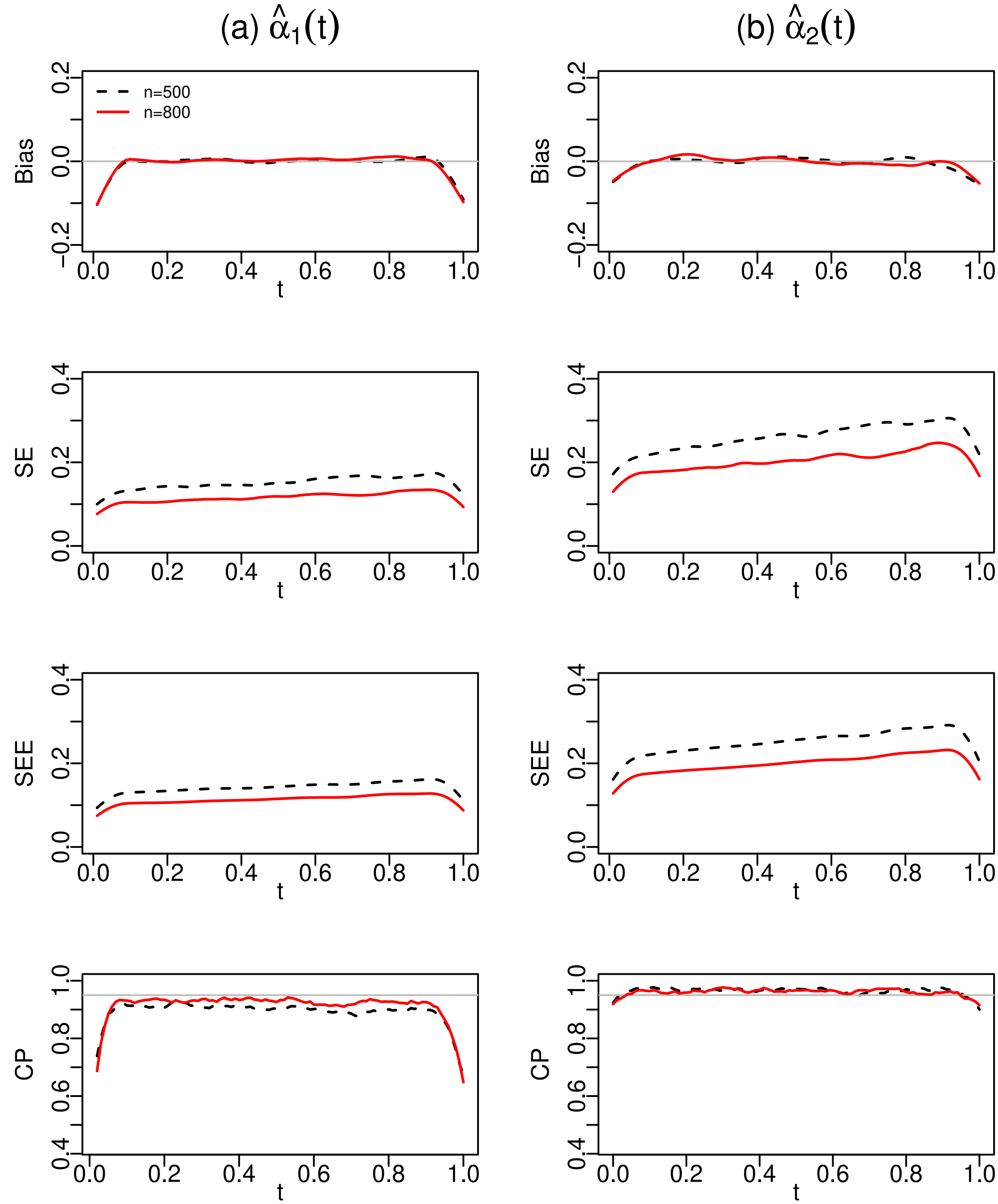

Figure 3.4: Estimation results for (a) $\alpha_1(t) = 1/(4+t)$ and (b) $\alpha_2(t) = 0.1$ in Scenario 2 with $r = 0.5$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.

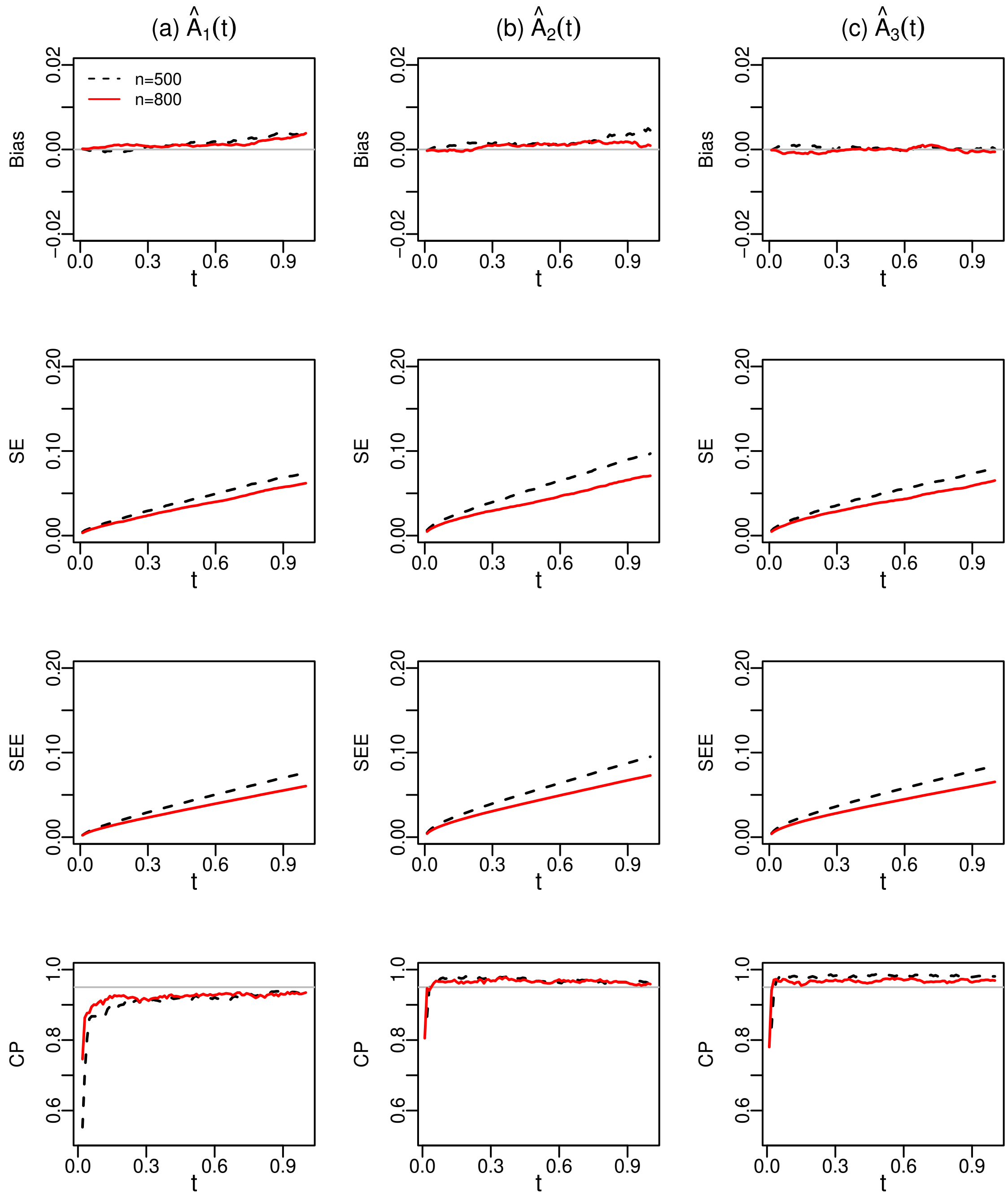

Figure 3.5: Estimation results for (a) $A_1(t) = \log(1 + t/4)$, (b) $A_2(t) = 0.1t$ and (c) $A_3(t) = 0.05t$ in Scenario 4 with $r = 1$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.

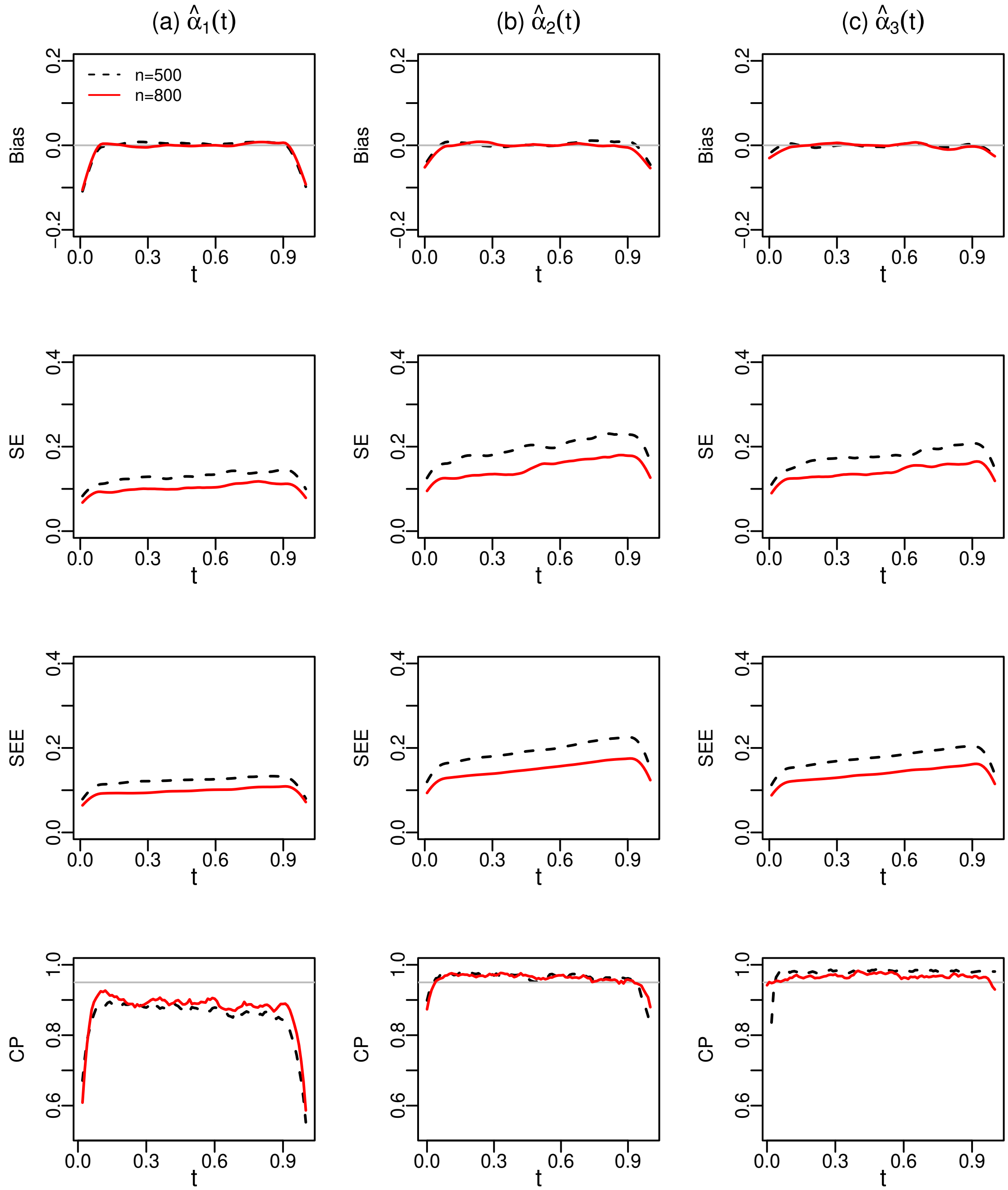

Figure 3.6: Estimation results for (a) $\alpha_1(t) = 1/(4+t)$, (b) $\alpha_2(t) = 0.1$ and (c) $\alpha_3(t) = 0.05$ in Scenario 4 with $r = 1$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.

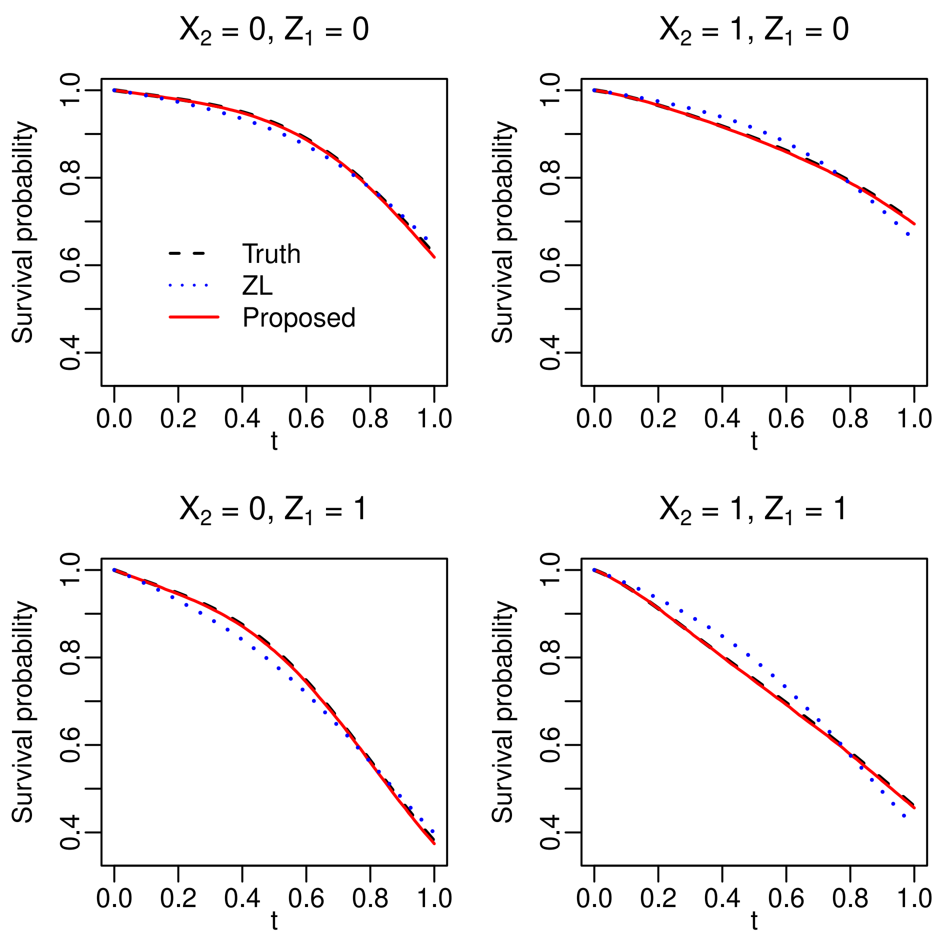

Figure 3.7: Predicted survival probabilities under Scenario 5 with $r = 1$ based on the proposed model and Zeng and Lin's model. Here, "ZL" stands for Zeng and Lin's model.

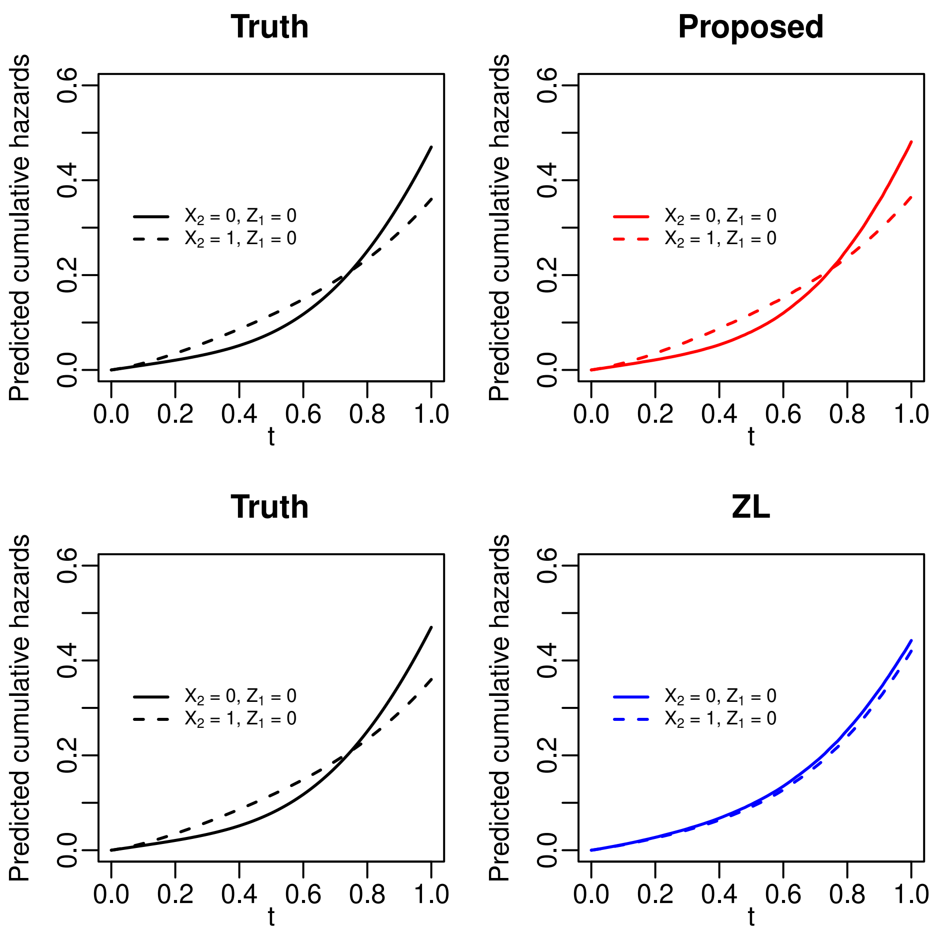

Figure 3.8: Predicted baseline cumulative hazard functions under Scenario 5 with $r = 1$ based on the proposed model and Zeng and Lin's model. Here, "ZL" stands for Zeng and Lin's model.

Lastly, we performed simulation studies to confirm that the log-likelihood values remain relatively stable across different r values when the percentage of censoring is high. Specifically, we assumed that the failure time T follows the Cox-Aalen transformation model:

$$\Lambda(t) = G\left\{\int_0^t \exp(\beta_1 Z_1) d\Lambda_X(s)\right\}, \tag{3.16}$$

where $\beta_1 = 1$, $Z_1 \sim \text{Unif}(0,1)$, $X = (1, X_2)^\top$ with $X_2 \sim \text{Ber}(0.5)$, $A_1(t) = t^2/2$ and $A_2 = 0.1t$. Let $\tau = 1$. We generate one censoring time $C \sim \text{Exponential}(b)$ such that $b = 0.3$ and $b = 7$ yield a censoring rate around 50% and 95%, respectively. We generate the data from the model (3.16) with $r = 0$ and then fit the generated data with r values in the interval $[0, 3]$ with a step size of 0.1. Note that $r = 0$ can be considered as the true model while other r values are misspecified. In Figure 3.9, we plot the average log-likelihood values across 200 replicates as a function of r. It is evident that the true value of $r = 0$ is indeed the one that maximizes the log-likelihood. However, under high censoring percentage, the log-likelihood changes very slowly with different values of r.

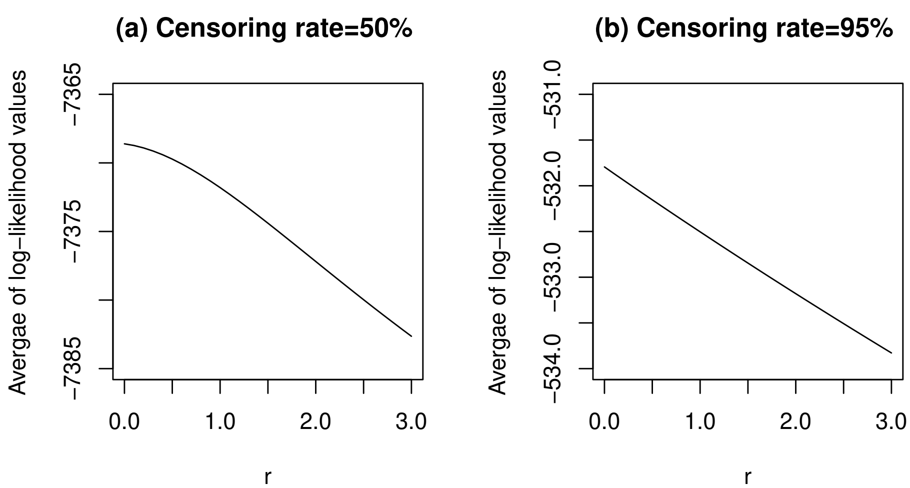

Figure 3.9: Average log-likelihood function at the final parameter estimates for model (3.16) with r values in the interval [0, 3] and a step size of 0.1.

3.5 Data Application

In this section, we apply the proposed model and methods to two harmonized randomized trials, HIV Vaccine Trials Network (HVTN) 704/HIV Prevention Trials Network (HPTN) 085 and HVTN 703/HPTN 081 (Corey et al., 2021), designed to determine whether a broadly neutralizing antibody (bnAb) can be used to prevent the acquisition of human immunodeficiency virus type 1 (HIV-1). The HVTN 704/HPTN 085 trial enrolled 2687 men who have sex with men and transgender persons in the Americas and Europe, and HVTN 703/HPTN 081 trial enrolled 1924 females in sub-Saharan Africa. For each trial, HIV-1 uninfected participants were randomly assigned in 1:1:1 ratio to receive infusions of a bnAb (VRC01) at a dose of 10 mg per kilogram of body weight (low-dose group), VRC01 at 30 mg per kilogram (high-dose group) or saline placebo, administered at 8-week intervals for 10 total infusions. The primary efficacy endpoint was diagnosis of HIV-1 infection by the week 80 trial visit, and HIV-1 testing was conducted at each 4-week trial visit starting at week 0. For participants acquiring HIV-1 infection, the diagnosis date was determined by the adjudicated diagnosis date based on validated assays (Corey et al., 2021). Participant follow-up is right-censored by the minimum of their last negative HIV-1 sample collection date and $\tau = 85.9$ weeks (Corey et al., 2021).

Among the 4559 HIV-1 negative participants from both trials, 1401 are in the USA and Switzerland, 1249 in Brazil and Peru, 1009 in South Africa, and 900 in other sub-Saharan African countries (Switzerland was pooled with the U.S. given few participants in Switzerland). We analyze the two trials pooled together, which is valid given the harmonized protocols such that essentially the study is one trial in two distinct study populations. There were a total of 174 HIV-1 infection diagnosis endpoints in the two trials pooled, including 60 out of 1520 participants in the low-dose group, 47 out of 1520 in the high-dose group, and 67 out of 1519 in the placebo group. The numbers of HIV-1 infection diagnosis endpoints by region are reported in

Table 3.3. Participants were categorized by age (in years old) into four groups, [17, 20], [21, 30], [31, 40] and [41, 52], with 540, 2651, 1102 and 266 participants, respectively.

Table 3.3: Summary statistics for the HIV-1 infections

Regions	Placebo	Low-dose	High-dose	Total
USA&Switzerland	9	8	6	23
Brazil&Peru	29	24	22	75
South Africa	16	16	11	43
Other sub-Saharan African countries	13	12	8	33
Total	67	60	47	174

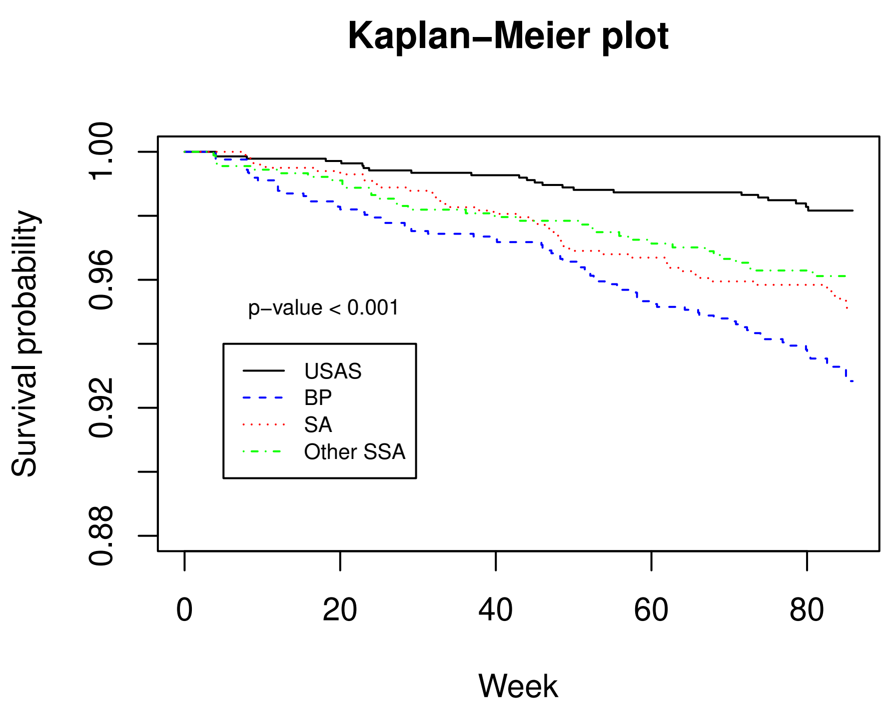

Figure 3.10: Kaplan-Meier plot for four regions in the full cohort. Here, "USAS", "BP", "SA" and "Other SSA" represent USA and Switzerland, Brazil and Peru, South Africa and other sub-Saharan African countries, respectively.

Figure 3.10 reveals that the risk of HIV-1 infection in different regions cross over. Therefore, without imposing proportional hazards for different regions, we consider the following Cox-Aalen transformation model to assess the association between treatment assignment, age, and region with the time since first infusion to HIV-1 infection diagnosis:

$$\Lambda(t \mid X, Z) = G\left\{\int_0^t \exp(\beta^\top Z) d\Lambda_X(s)\right\}, \tag{3.17}$$

where β is the unknown regression coefficients, and $\Lambda_X(s) = \int_0^s \left\{X^\top \alpha(v)\right\} dv = X^\top A(s)$ with $A(s) = (A_1(s), \cdots, A_4(s))^\top$. Here, $Z = (Z_1, Z_2, Z_3, Z_4, Z_5)^\top$, where Z_1 and Z_2 are indicators of being assigned to the low-dose and high-dose group, respectively, with the placebo group as the reference group; Z_3, Z_4, Z_5 are indicators of the age groups $[21, 30]$, $[31, 40]$ and $[41, 52]$, respectively, with $[17, 20]$ as the reference age group. In addition, let $X = (1, X_2, X_3, X_4)^\top$, where X_2, X_3, X_4 are indicators of participants from Brazil and Peru, South Africa, and other sub-Saharan African countries, respectively. The participants from USA and Switzerland are considered as the reference group.

To conduct our analysis, we employed model (3.17) with $G(x) = r^{-1}\log(1 + rx)$. We tested r values in the interval [0, 3] with a step size of 0.1 and chose the value of r that that yielded the maximum log-likelihood function at the final parameter estimates. Despite a high censoring rate of approximately 96.2%, we found that the log-likelihood was maximized at $r = 0$, as shown in Figure 3.11. Our simulation studies in Section 3.4 supported this finding by demonstrating that the log-likelihood values did not change significantly for different values of r due to the high censoring rate.

The lower panel of Table 3.4 shows the regression parameter estimates for the selected transformation function ($r = 0$). One can see that a high-dose VRC01 significantly lowers the risk of HIV-1 infections, while a low-dose VRC01 does not. Table 3.4 also reflects a significant correlation between age and HIV-1 infections, i.e., being older decreases the risk of HIV-1 infections. Figure 3.12 displays the estimated baseline cumulative hazard function for four different regions under the selected model. The risk of HIV-1 infection is the highest in Brazil and Peru, and lowest in the United States and Switzerland. The estimates for South Africa and other

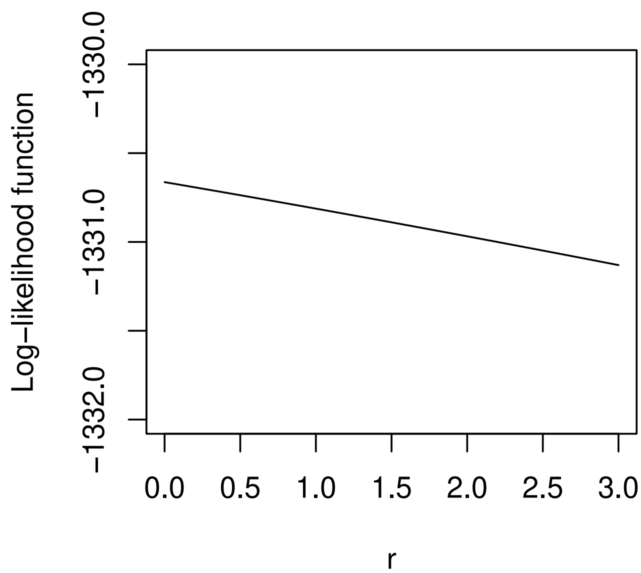

Figure 3.11: Log-likelihood function at the final parameter estimates in model (3.17) with r values in the interval [0, 3] and a step size of 0.1.

sub-Saharan African countries cross; in particular, South Africa has a lower risk at the beginning of the study but a higher risk later. Furthermore, Figure 3.13 depicts the estimated survival probabilities at sixteen distinct combinations of covariates, including age group and region. The figure provides additional evidence that supports our previously reported findings.

The four other panels of Table 3.4 (upper panels) show results from Zeng and Lin's model fit to each of the four geographic regions separately; this method was not applied to the full cohort (pooled) data because it cannot flexibly model the differences in baseline cumulative hazards and the diagnostics support lack of fit. In these results, the p-values for the effect of high-dose VRC01 markedly increase, and the coefficient estimates for the age group $[41, 52]$ are unstable because there are very few HIV-1 infection diagnosis endpoints in this age group in the three regions Brazil and Peru, South Africa, and other sub-Saharan African countries. Therefore, the results from the Cox-Aalen transformation modeling – which could be based on

Table 3.4: Regression analysis results in the HIV-1 trials under Zeng and Lin's model and the proposed model with $r = 0$

	USA/Switzerland			Brazil/Peru		
Covariates	Est	SE	p-value	Est	SE	p-value
Low-dose	−0.107	0.484	0.825	−0.167	0.276	0.545
High-dose	−0.437	0.524	0.404	−0.279	0.283	0.325
21 − 30	−0.454	0.630	0.472	−0.525	0.262	0.045
31 − 40	−2.709	1.141	0.018	−1.283	0.396	0.001
41 − 52	−1.152	0.903	0.202	−16.859	1.668	< 0.001
	South Africa			Other SSA		
Covariates	Est	SE	p-value	Est	SE	p-value
Low-dose	−0.025	0.354	0.943	−0.080	0.400	0.842
High-dose	−0.392	0.392	0.317	−0.509	0.449	0.258
21 − 30	−0.187	0.380	0.623	−0.480	0.498	0.335
31 − 40	−0.954	0.601	0.112	−0.719	0.586	0.220
41 − 52	−13.867	2.366	< 0.001	−13.871	2.626	< 0.001
	The Proposed Model					
Covariates	Est	SE	p-value			
Low-dose	−0.108	0.178	0.542			
High-dose	−0.363	0.190	0.056			
21 − 30	−0.429	0.187	0.022			
31 − 40	−1.219	0.274	< 0.001			
41 − 52	−1.989	0.721	0.006			

Note: Est and SE stand for the estimates of the regression parameters and the estimated standard errors, respectively. "Other SS" is for other sub-Saharan African countries. "USA/Switzerlan", "Brazil/Per", "South Afric", and "Other SSA" correspond to the estimation results when fitting Zeng and Lin's transformation models with the participants in those regions separately. "The Proposed Mode" corresponds to the estimation results when fitting the proposed Cox-Aalen transformation models with the full cohort.

the full cohort data through flexible specifications of the baseline cumulative hazard functions – provide new insights with improved precision and power beyond insights achieved from the application of Zeng and Lin's model.

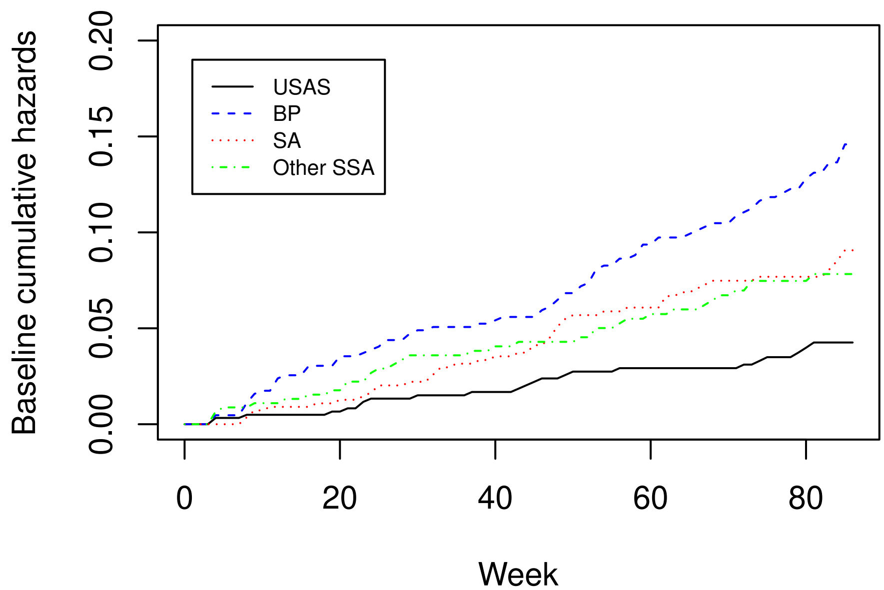

Figure 3.12: Estimated baseline cumulative hazard function, i.e., $\hat{\Lambda}(t \mid X, Z = 0)$ for four regions under the selected model ($r = 0$). Here,"USAS", "BP", "SA" and "Other SSA" represent USA and Switzerland, Brazil and Peru, South Africa and other sub-Saharan African countries, respectively.

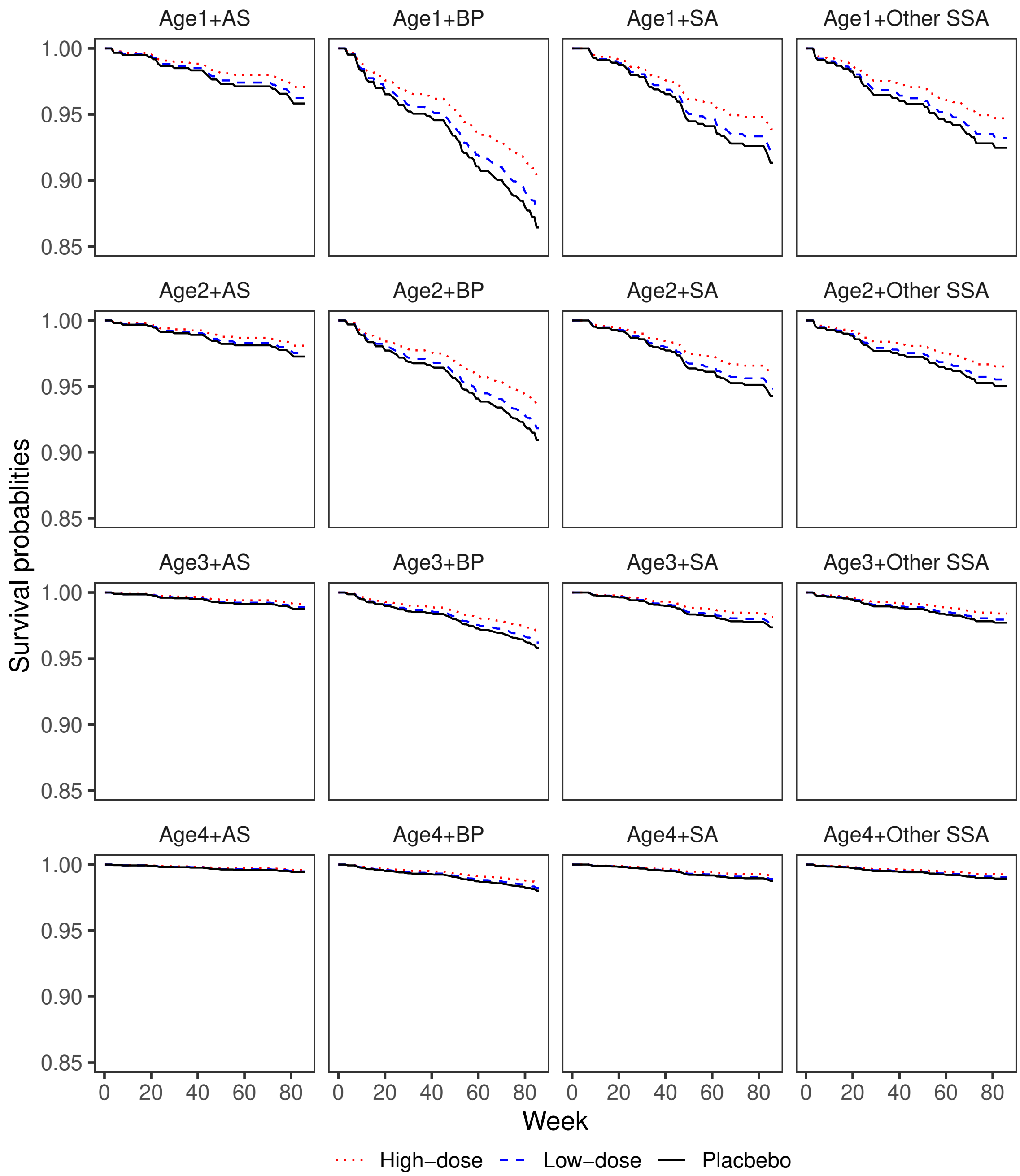

Figure 3.13: Predicted survival probabilities by analyzing different sets of covariates in the Cox-Aalen transformation model (3.17) with $r = 0$. Here, "Age1", "Age2", "Age3" and "Age4" stand for the age goups, $[17, 20]$, $[21, 30]$, $[31, 40]$ and $[41, 52]$, respectively. "AS", "BP", "SA" and "Other SSA" represent USA and Switzerland, Brazil and Peru, South Africa and other sub-Saharan African countries, respectively.

3.6 Discussion

In this chapter, we explored the use of frailty-induced transformations and developed a fast and stable ES algorithm to estimate the parametric and nonparametric components of the proposed Cox-Aalen transformation model with right-censored data, along with easy-to-compute variance estimators.

Elashoff and Ryan (2004) pointed out that an ES algorithm can be regarded as a block Newton-Gauss-Seidel algorithm (see Ortega 1972, p.146). Following Ortega (1972, p.147), an ES algorithm converges locally to the solution, $\hat{\theta}$, of $U(\theta) = 0$ if the Jacobian matrix $D = \partial U / \partial \theta$ is nonsingular at $\theta = \hat{\theta}$ and the largest eigenvalue of $D^{-1}(\hat{\theta})$ is less than 1. For general estimating equations, the two conditions above are difficult to verify in advance, especially for the second condition. Nevertheless, the matrix D is needed to calculate the variance of $\hat{\theta}$ in (3.14), and hence one can check the required conditions numerically.

In real data applications, we ascertain whether a covariate has a multiplicative or additive effect based on the following criteria. First, we may employ the underlying biological, physical meaning, or other domain knowledge for decision-making. Second, initial data exploration can be performed for each covariate, such as drawing the Kaplan-Meier (KM) plot. If the KM curves cross, forming an X-shape, this covariate should be modeled additively. Third, similar to Qu and Sun (2019), Yu et al. (2019), we may employ some AIC or BIC-based procedures. In particular, all possible combinations of covariate effects will be examined. However, it is easy to see that this is inefficient when there are many covariates. In addition, Scheike and Zhang (2003) proposed the supremum tests to determine the multiplicative and additive parts of the Cox-Aalen model. It would be valuable to explore whether similar testing procedures can be constructed for our proposed model. More theoretical and numerical studies are needed, which we leave for future work.

CHAPTER 4: SEMIPARAMETRIC REGRESSION ANALYSIS OF THE COX-AALEN TRANSFORMATION MODELS WITH PARTLY INTERVAL-CENSORED DATA

4.1 Introduction

Partly interval-censored data arise in epidemiological and biomedical studies when some failure times of interest are observed with exactitude while others are only known to have occurred within a specific time interval. This type of data combines features of both exact and interval-censored observations. Regression modeling of partly interval-censored data in the literature has primarily focused on the multiplicative and additive hazards models, which depend on whether the effects of covariates are multiplicative or additive. The proportional hazards model (Cox, 1972; Andersen and Gill, 1982) is the most popular multiplicative hazards model and has been extensively studied for modeling different types of censored data. For interval-censored data, maximum likelihood estimators have been proposed by several authors, including Finkelstein (1986), Huang (1996), Zhang et al. (2010), and Wang et al. (2016). For partly interval-censored data, Kim (2003) conducted the maximum likelihood estimation, and Pan et al. (2020) considered a Bayesian approach by introducing a nonhomogeneous Poisson process. However, the proportional hazards assumption cannot always be validated. Alternatively, the proportional odds model (Pettitt, 1982; Bennett, 1983) is prevalent in dealing with the non-proportionality. The literature on the use of this model for interval-censored data is extensive. Some examples include Huang and Rossini (1997), Rabinowitz et al. (2000), Zhu et al. (2021), and more references therein.

The additive hazards models have been advocated and utilized by several researchers, including Aalen (1980), Buckley (1984), Aalen (1989), Huffer and McKeague (1991), Lin and Ying (1994), and others. When analyzing interval-censored data, there are typically three types of statistical methods that are used: maximum likelihood estimation (Zeng et al., 2006; He et al., 2020), multiple imputation approach (Chen and Sun, 2010), and estimating equation approach (Wang et al., 2010). Researchers have also proposed models that combine both multiplicative and additive effects of covariates to enhance the modeling capacity (Lin and Ying, 1995; Martinussen and Scheike, 2002; Scheike and Zhang, 2002). The Cox-Aalen model (Scheike and Zhang, 2002) is an example of a multiplicative-additive hazards model that extends the proportional hazards model by replacing the baseline hazard function with Aalen's additive model. However, existing methods for interval-censored data (Boruvka and Cook, 2015; Shen and Weng, 2019) rely on the assumption of fixed covariates and cannot easily accommodate time-varying covariates. Furthermore, to the best of our knowledge, there is currently limited research on the use of the Cox-Aalen model for analyzing partly interval-censored data.

Recently, significant progress has been made in the development of transformation models (Zeng and Lin, 2006), referred to as Zeng and Lin's model in this dissertation. This class of models extends linear transformation models (Dabrowska and Doksum, 1988; Fine et al., 1998) by incorporating potentially time-variant covariates. For interval-censored data, Zhang et al. (2005) proposed an estimating equation approach for linear transformation models, and Zeng et al. (2016) developed a nonparametric maximum likelihood estimator (NPMLE) for Zeng and Lin's model using an EM algorithm. However, available inference methods for partly interval-censored data with transformation models are relatively limited due to the complex model and data structures involved. Some recent studies include the nonparametric maximum likelihood estimator by Zhou et al. (2021) and a Bayesian approach with a monotone spline approximation by Wang et al. (2022). Efficient semiparametric estimation for partly interval-censored data has also been studied for other models, such as the accelerated failure time model Gao et al. (2018). However, one limitation of Zeng and Lin's models is that they assume all covariate effects to be multiplicative. This assumption has been recognized as too restrictive by Scheike and Zhang (2003) and may result in prediction biases.

In this chapter, we focus on the regression analysis of partly interval-censored data through the class of Cox-Aalen transformation models. We begin by formulating the likelihood and deriving the NPMLE for a specific case. Subsequently, we propose a collection of estimating equations and develop an expectation-solving (ES) algorithm (Elashoff and Ryan, 2004), which iterates between an E-step, where functions of complete data are replaced by their expectations and an S-step where these expected values are substituted into the estimating equations. The asymptotic properties of the proposed estimators are thoroughly investigated. Additionally, we show that the variance of the estimators can be consistently estimated using the weighted bootstrap method. Finally, we demonstrate the proposed procedure's performance via simulation studies and a randomized HIV/AIDS trial analysis.

4.2 Methods

4.2.1 Data structure and Notation

For the ith individual subject to partly interval censoring, let T_i denote their failure time. If T_i is observed exactly, we set $\Delta_i = 1$. Otherwise, we set $\Delta_i = 0$ and denote the sequence of examination times for individual i as $U_{i1}, U_{i2}, \ldots, U_{iK_i}$, where $0 < U_{i1} < U_{i2} < \cdots < U_{iK_i} < \infty$, and K_i is the total number of examinations. For each event T_i, we define $(L_i, R_i]$ as the interval with the smallest possible width that encloses T_i. The left endpoint L_i of this interval is the maximum value of U_{ik} for $k = 0, 1, \ldots, K_i$ such that $U_{ik} \leq T_i$. Similarly, the right endpoint R_i is the minimum value of U_{ik} for $k = 1, 2, \ldots, K_i + 1$ such that $U_{ik} \geq T_i$. We set $U_{i0} = 0$ and $U_{i,K_i+1} = \infty$ to ensure that L_i and R_i are well-defined for all i. In addition, note that

$L_i = 0$ and $R_i = \infty$ correspond to left- and right-censored observations, respectively, while $0 < L_i < R_i < \infty$ corresponds to typically interval-censored observations. Let $X_i(\cdot)$ and $Z_i(\cdot)$ be $q \times 1$ and $d \times 1$ vectors of potentially time-varying covariates for individual i, respectively. Suppose we have a random sample of n individuals. Therefore, the observed partly interval-censored data comprise

$$\mathcal{O}_i = \{\Delta_i, \Delta_i T_i, (1-\Delta_i)L_i, (1-\Delta_i)R_i, X_i(\cdot), Z_i(\cdot)\} \quad \text{for } i = 1, \ldots, n. \tag{4.1}$$

4.2.2 Models and Likelihood

Recall that, for the proposed Cox-Aalen transformation model, the cumulative hazard function of failure time T_i conditional on $X_i(\cdot)$ and $Z_i(\cdot)$ takes the following form:

$$\Lambda(t \mid X_i(\cdot), Z_i(\cdot)) = G\left[\int_0^t \exp\{\beta^\top Z_i(s)\} d\Lambda_{X_i}(s)\right], \tag{4.2}$$

where β is a d-dimensional regression parameter, $\Lambda_{X_i}(s) = \int_0^t X_i^\top(v)\alpha(v)dv$ is an unspecified increasing function that depends on $X_i(\cdot)$ and a vector of unknown regression functions $\alpha(t) = (\alpha_1(t), \ldots, \alpha_q(t))^\top$. The function $G(\cdot)$ is pre-determined and satisfies the following properties: it is strictly increasing, thrice continuously differentiable, and $G(0) = 0$, $G'(0) > 0$, and $G(\infty) = \infty$. By setting the first element of $X_i(\cdot)$ to 1, the function $\alpha_1(t)$ can serve as a baseline or reference level for risk. In addition, we define $A(t) = \int_0^t \alpha(s)ds = (A_1(t), \ldots, A_q(t))^\top$, where $A_j(t) = \int_0^t \alpha_j(s)ds$ for $j = 1, \ldots, q$, as a vector of cumulative regression functions. For the choices of $G(\cdot)$, we again consider a class of frailty-induced transformation functions

$$G(x) = -\log \int_0^\infty \exp(-x\xi) f(\xi) d\xi, \tag{4.3}$$

where $f(\xi)$ is the the probability density function of a nonnegative random variable ξ on $[0, \infty)$. One widely used choice for $f(\xi)$ is the gamma density with mean 1 and variance r, which leads to $G(x) = r^{-1}\log(1+rx)$ for $r \geq 0$. Notably, when $r = 0$, $G(x)$ reduces to the identity function $G(x) = x$, and the Cox-Aalen transformation model (4.2) reduces to the Cox-Aalen model.

Suppose that K_i, $\{U_{ij}, j = 1, \ldots, K_i\}$ are independent of T_i conditional on $X_i(\cdot)$ and $Z_i(\cdot)$. Under model (4.2), the observed-data likelihood function based on (4.1) takes the form

$$\begin{aligned}\prod_{i=1}^{n} \Bigg(& \Lambda'_{X_i}(T_i) e^{\beta^\top Z_i(T_i)} G'\Big\{\int_0^{T_i} e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\Big\} \\ & \times \exp\Big[-G\Big\{\int_0^{T_i} e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\Big\}\Big]\Bigg)^{\Delta_i} \\ & \Bigg(\exp\Big[-G\Big\{\int_0^{L_i} e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\Big\}\Big] \\ & - \exp\Big[-G\Big\{\int_0^{R_i} e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\Big\}\Big]\Bigg)^{1-\Delta_i}.\end{aligned} \tag{4.4}$$

Here, the derivatives of $\Lambda_X(\cdot)$ and $G(\cdot)$ are denoted as $\Lambda'_X(\cdot)$ and $G'(\cdot)$, respectively. With the class of frailty-induced transformations (4.3), it is easy to show that the likelihood (4.4) is equivalent to

$$\begin{aligned}\prod_{i=1}^{n} & \left[\Lambda'_{X_i}(T_i) e^{\beta^\top Z_i(T_i)} \int_{\xi_i} \xi_i \exp\left\{-\xi_i \int_0^{T_i} e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\right\} f(\xi_i) d\xi_i\right]^{\Delta_i} \\ & \Bigg(\int_{\xi_i} \Bigg[\exp\Big\{-\xi_i \int_0^{L_i} e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\Big\} \\ & - \exp\Big\{-\xi_i \int_0^{R_i} e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\Big\}\Bigg] f(\xi_i) d\xi_i\Bigg)^{1-\Delta_i}.\end{aligned} \tag{4.5}$$

This class of frailty-induced transformations has the benefit of eliminating the need for the function $G(\cdot)$, allowing for the development of a computationally efficient EM algorithm by treating ξ_i as missing. This technique has been frequently applied in various studies, including those by Zeng and Lin (2007); Liu and Zeng (2013); Zeng et al. (2016); Gao et al. (2018); Zhou et al. (2021).

4.2.3 Nonparametric Maximum Likelihood Estimation

This subsection considers the nonparametric maximum likelihood estimator for β and $\Lambda_{X_i}(\cdot)$ in the Cox-Aalen transformation model. Specifically, let $0 = t_0 < t_1 < \cdots < t_m < \infty$ be the distinct time points in the set of T_i, L_i and $R_i < \infty$ for $i = 1, \ldots, n$. Moreover, we assume that the cumulative regression function $A_j(t)$ ($j = 1, \ldots, q$) is a step function with jump size a_{jk} at t_k for $k = 1, \ldots, m$ with $a_{j0} = 0$. By observing that $d\Lambda_{X_i}(t) = X_i^\top(t)dA(t)$, thus, Λ_{X_i} can be represented as a step function with a jump size $X_i^\top(t_k)a_k$ at time t_k, where $a_k = (a_{1k}, \ldots, a_{qk})^\top$. Then, we can express the likelihood function (4.5) as:

$$\begin{aligned}
&\prod_{i=1}^{n}\left[\Lambda_{X_i}\{T_i\}e^{\beta^\top Z_i(T_i)}\int_{\xi_i}\xi_i\exp\left\{-\xi_i\sum_{t_k\le T_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\right\}f(\xi_i)d\xi_i\right]^{\Delta_i}\\
&\quad\left(\int_{\xi_i}\left[\exp\left\{-\xi_i\sum_{t_k\le L_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\right\}\right.\right.\\
&\quad\left.\left.-\exp\left\{-\xi_i\sum_{t_k\le R_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\right\}\right]f(\xi_i)d\xi_i\right)^{1-\Delta_i}.
\end{aligned} \tag{4.6}$$

Here, $\Lambda_{X_i}\{T_i\}$ represents the jump size of Λ_{X_i} at time point T_i, $X_{ik} = X_i(t_k)$ and $Z_{ik} = Z_i(t_k)$. For simplicity, we rearrange (4.6) as

$$\begin{aligned}
&\prod_{i=1}^{n}\left(\int_{\xi_i}\exp\left\{-\xi_i\sum_{t_k<T_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\right\}\right.\\
&\quad\left.\times\,\xi_i\Lambda_{X_i}\{T_i\}e^{\beta^\top Z_i(T_i)}\exp\left[-\xi_i\Lambda_{X_i}\{T_i\}e^{\beta^\top Z_i(T_i)}\right]f(\xi_i)d\xi_i\right)^{\Delta_i}\\
&\quad\left(\int_{\xi_i}\exp\left\{-\xi_i\sum_{t_k\le L_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\right\}\right.\\
&\quad\left.\times\left[1-\exp\left\{-\xi_i\sum_{L_i<t_k\le R_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\right\}\right]^{I(R_i<\infty)}f(\xi_i)d\xi_i\right)^{1-\Delta_i}.
\end{aligned} \tag{4.7}$$

To simplify the process of maximizing the likelihood function (4.7), we adopt a set of independent Poisson variables, denoted as W_{ik} ($i = 1, \ldots, n; k = 1, \ldots, m$).

These variables are considered as latent variables with means $\xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})$ conditional on ξ_i. Other references that share similar ideas include Wang et al. (2016), Zeng et al. (2016), and Zhou et al. (2021).

Define $A_i = \Delta_i \sum_{t_k<T_i} W_{ik}$, $B_i = \Delta_i \sum_{t_k=T_i} W_{ik}$, $C_i = (1-\Delta_i)\sum_{t_k\le L_i} W_{ik}$ and $D_i = (1-\Delta_i)I(R_i<\infty)\sum_{t_k\le R_i} W_{ik}$ for $i = 1,\ldots,n$. Suppose that the observed data consist of

$$\begin{cases} (T_i, X_i, Z_i, A_i = 0, B_i = 1) & \text{if } \Delta_i = 1 \\ (L_i, R_i, X_i, Z_i, C_i = 0, D_i > 0) & \text{if } \Delta_i = 0, \end{cases} \tag{4.8}$$

for $i = 1,\ldots,n$. If $A_i = 0$ and $B_i = 1$, this indicates that A_i and B_i are known to be 0 and 1, respectively. Specifically, if $\Delta_i = 1$, then $W_{ik} = 0$ for all $t_k < T_i$ and $W_{ik} = 1$ for $t_k = T_i$. Similarly, if $C_i = 0$ and $D_i > 0$, this implies that C_i and D_i are known to be zero and positive, respectively. Specifically, if $\Delta_i = 0$, then $W_{ik} = 0$ for all $t_k \le L_i$, and at least one $W_{ik} \ge 1$ for $L_i < t_k \le R_i$ with $R_i < \infty$. By independent properties of W_{ik}, we can compute the conditional probability $Pr(A_i = 0|\xi_i) = \exp\big\{-\xi_i\sum_{t_k<T_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\big\}$. Applying this similar idea, the likelihood (4.7) can be represented with the data in (4.8) as

$$\prod_{i=1}^{n}\left\{\int_{\xi_i} Pr\Big(\sum_{t_k<T_i} W_{ik} = 0\Big|\xi_i\Big) Pr\Big(\sum_{t_k=T_i} W_{ik} = 1\Big|\xi_i\Big) f(\xi_i)d\xi_i\right\}^{\Delta_i}$$
$$\left[\int_{\xi_i} Pr\Big(\sum_{t_k\le L_i} W_{ik} = 0\Big|\xi_i\Big)\left\{1 - Pr\Big(\sum_{L_i<t_k\le R_i} W_{ik} = 0\Big|\xi_i\Big)\right\}^{I(R_i<\infty)} f(\xi_i)d\xi_i\right]^{1-\Delta_i}. \tag{4.9}$$

The parameter vector of interest is represented by $\theta = (a_1^\top,\ldots,a_m^\top,\beta^\top)^\top$. To estimate θ, we propose maximizing the likelihood function (4.9) using an EM algorithm that treats W_{ik} and ξ_i as complete data. Specifically, we aim to maximize the complete-data loglikelihood function in the EM algorithm, which is given by the following expression:

$$\sum_{i=1}^{n}\Bigg(\sum_{k=1}^{m} I(t_k \le R_i^*)\big[W_{ik}\log\big\{\xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\big\} - \xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik}) - \log W_{ik}!\big] + \log f(\xi_i)\Bigg), \tag{4.10}$$

where $R_i^* = \Delta_i T_i + (1-\Delta_i)\{L_i I(R_i=\infty) + R_i I(R_i<\infty)\}$.

In the E-step of the EM algorithm, we calculate the conditional expectation of (4.10). This step involves evaluating the posterior means of the latent variables W_{ik} and ξ_i given the observed data, which are denoted by $\hat{E}(W_{ik})$ and $\hat{E}(\xi_i)$, respectively. More details on this step are described in the next subsection. In the M-step, we maximize the conditional expectation of (4.10) with respect to the parameter vector $\theta = (a_1^\top, \ldots, a_m^\top, \beta^\top)^\top$. Specifically, we set the derivatives of the conditional expectation of (4.10) with respect to a_k ($k = 1, \ldots, m$) and β to zeros, respectively. Consequently, we solve for the estimates of a_k and β through the following equations:

$$\sum_{i=1}^{n} I(t_k \le R_i^*)\Big\{\frac{\hat{E}(W_{ik})}{X_{ik}^\top a_k} - \hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\Big\}X_{ik} = 0, \quad \text{for } k = 1, \ldots, m, \tag{4.11}$$

$$\sum_{i=1}^{n}\sum_{k=1}^{m} I(t_k \le R_i^*)\Big\{\hat{E}(W_{ik}) - \hat{E}(\xi_i)(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\Big\}Z_{ik} = 0. \tag{4.12}$$

Note that the dimension of θ is $(mq+d)$, which may exceed the sample size n of partly interval-censored data. Hence, (4.11) and (4.12) are a system of high-dimensional nonlinear equations that are unprecedentedly challenging and computationally expensive to solve. Under a special case that X_i represents levels in a set of factors, there exist explicit formulae for calculating the high-dimensional parameters a_k ($k = 1, \ldots, m$) when fixing β in (4.11). With those high-dimensional parameters fixed in (4.12), the low-dimensional parameter β can be solved via any root-finding algorithms, such as the Newton-Raphson method. Here, we give some further illustrations.

Let J be a categorical variable with q levels. We can assume, for the sake of simplicity and without affecting the generality of the analysis, that J takes values in $\{1,\ldots,q\}$. Let $X=(1,X_2,\ldots,X_q)$ where $X_2,\ldots,X_q$ are group indicators, i.e., $X_2=I(J=2),\ldots,X_q=I(J=q)$. Here, $J=1$ is considered as the reference group. We propose the following Gauss-Seidel method to jointly solve (4.11) and (4.12). Start with some initial values of the unknown parameters.

Step 1. Fix β, we update $a_k,(k=1,\ldots,m)$ by solving (4.11). Note that for a fixed k, equation (4.11) is equivalent to

$$\begin{cases} \sum_{i=1}^n I(J_i=1)I(t_k\le R_i^*)\left\{\frac{\hat{E}(W_{ik})}{a_{1k}}-\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\right\}=0 \\ \sum_{i=1}^n I(J_i=2)I(t_k\le R_i^*)\left\{\frac{\hat{E}(W_{ik})}{a_{1k}+a_{2k}}-\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\right\}=0 \\ \vdots \\ \sum_{i=1}^n I(J_i=q)I(t_k\le R_i^*)\left\{\frac{\hat{E}(W_{ik})}{a_{1k}+a_{qk}}-\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\right\}=0. \end{cases}$$

Hence, we obtain that

$$\begin{cases} a_{1k}=\frac{\sum_{i=1}^n I(J_i=1)I(t_k\le R_i^*)\hat{E}(W_{ik})}{\sum_{i=1}^n I(D_i=1)I(t_k\le R_i^*)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})} \\ a_{2k}=\frac{\sum_{i=1}^n I(J_i=2)I(t_k\le R_i^*)\hat{E}(W_{ik})}{\sum_{i=1}^n I(J_i=2)I(t_k\le R_i^*)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})}-a_{1k} \\ \vdots \\ a_{qk}=\frac{\sum_{i=1}^n I(J_i=q)I(t_k\le R_i^*)\hat{E}(W_{ik})}{\sum_{i=1}^n I(J_i=q)I(t_k\le R_i^*)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})}-a_{1k}. \end{cases} \tag{4.13}$$

Step 2. Fix $a_1,\ldots,a_m$, we update β by solving (4.12) using the Newton-Raphson method. We alternate between E- and M-steps until convergence is achieved.

In the case where X represents levels in a set of factors, coupled with $G(x)=x$, the model (4.2) reduces to the stratified Cox model (Kalbfleisch and Prentice, 2002), which serves as an alternative to accommodate the non-proportionality hazards assumption in the literature. Our methods fill the gap for estimating the stratified Cox model from partly interval-censored data, which has not been investigated yet, to the best of our knowledge. However, there are no explicit forms for a_k despite the aforementioned special case. Therefore, the computational challenges continue to present barriers to implementing the EM algorithm. To cope with more general situations, we adopt an estimating equation approach and employ an expectation-solving (ES) algorithm (Elashoff and Ryan, 2004) for feasible and simple computations.

4.2.4 Estimating Equations

In the previous subsection, we introduced a collection of independent latent variables W_{ik} such that $W_{ik}|\xi_i \sim \text{Poisson}\{(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\}$. Based on this, we construct the complete-data estimating equations $U(\theta) = (U_{a_1}, \ldots, U_{a_m}, U_\beta) = 0$, where

$$\begin{cases} U_{a_1} = \sum_{i=1}^n I(t_1 \le R_i^*) \left\{W_{i1} - \xi_i(X_{i1}^\top a_1)\exp(\beta^\top Z_{i1})\right\} X_{i1} \\ \vdots \\ U_{a_m} = \sum_{i=1}^n I(t_m \le R_i^*) \left\{W_{im} - \xi_i(X_{im}^\top a_m)\exp(\beta^\top Z_{im})\right\} X_{im} \\ U_\beta = \sum_{i=1}^n \sum_{k=1}^m I(t_k \le R_i^*) \left\{W_{ik} - \xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\right\} Z_{ik}. \end{cases} \tag{4.14}$$

Conditional expectation arguments easily establish that (4.14) is a system of unbiased estimating equations. For instance, notice that

$$\begin{aligned} & E\left[I(t_1 \le R_i^*) \left\{W_{i1} - \xi_i(X_{i1}^\top a_1)\exp(\beta^\top Z_{i1})\right\} X_{i1}\right] \\ & = E\left(E\left[I(t_1 \le R_i^*) \left\{W_{i1} - \xi_i(X_{i1}^\top a_1)\exp(\beta^\top Z_{i1})\right\} X_{i1} \mid \mathcal{O}_i, \xi_i\right]\right) \\ & = E\left[I(t_1 \le R_i^*) \left\{E(W_{i1}|\mathcal{O}_i, \xi_i) - \xi_i(X_{i1}^\top a_1)\exp(\beta^\top Z_{i1})\right\} X_{i1}\right] \\ & = 0. \end{aligned}$$

Hence, $E\left[\sum_{i=1}^n I(t_1 \le R_i^*) \left\{W_{i1} - \xi_i(X_{i1}^\top a_1)\exp(\beta^\top Z_{i1})\right\} X_{i1}\right] = 0$.

We propose to estimate θ through an ES algorithm by treating W_{ik} and ξ_i ($i = 1, \ldots, n; k = 1, \ldots, m$) as missing. The ES algorithm consists of two steps: an E-step, in which the missing data W_{ik} and ξ_i are imputed by their conditional expectations, and an S-step, which solves the conditional expectation of the estimating equations (4.14) given the observed data. Starting with initial values of parameters θ, the proposed ES algorithm proceeds through the following two steps iteratively until convergence:

E-step. Evaluate the posterior means $\hat{E}(W_{ik})$ and $\hat{E}(\xi_i)$ given the observed data. When $\Delta_i = 1$, the posterior density function of ξ_i given the observed data is proportional to $\xi_i \exp(-\xi_i S_{iT}) f(\xi_i)$, where $S_{iT} = \sum_{t_k \le T_i} (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik})$. Hence, we calculate

$$\hat{E}(\xi_i) = G^{'}(S_{iT}) - \frac{G^{''}(S_{iT})}{G^{'}(S_{iT})},$$

where $G'(x)$ and $G''(x)$ are the first and second derivatives of $G(\cdot)$ with respect to x, respectively. When $\Delta_i = 0$, the posterior density of ξ_i given the observed data is proportional to $\{\exp(-\xi_i S_{iL}) - \exp(-\xi_i S_{iR})\} f(\xi_i)$, where $S_{iL} = \sum_{t_k \le L_i} (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik})$ and $S_{iR} = \sum_{t_k \le R_i} (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik})$. We then obtain

$$\hat{E}(\xi_i) = \frac{\exp\{-G(S_{iL})\} G'(S_{iL}) - \exp\{-G(S_{iR})\} G'(S_{iR})}{\exp\{-G(S_{iL})\} - \exp\{-G(S_{iR})\}}.$$

For the posterior mean of W_{ik}, when $\Delta_i = 1$, we observe $(X_i, Z_i, A_i = 0, B_i = 1)$. Thus, $\hat{E}(W_{ik}) = 0$ for all $t_k < T_i$ and $\hat{E}(W_{ik}) = 1$ for $t_k = T_i$. When $\Delta_i = 0$, we observe $(L_i, R_i, X_i, Z_i, C_i = 0, D_i > 0)$. Thus, for any $t_k \le L_i$,

$$\hat{E}(W_{ik}) = E(W_{ik} | C_i = 0, D_i > 0, X_i, Z_i) = 0,$$

For $L_i < t_k \leq R_i$ with $R_i < \infty$,

$$
\begin{aligned}
\hat{E}(W_{ik}) &= E_{\xi_i}\left\{E(W_{ik}|\xi_i, C_i = 0, D_i > 0)|C_i = 0, D_i > 0\right\} \\
&= E_{\xi_i}\left[\frac{\xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})}{1-\exp\{-\xi_i(S_{iR}-S_{iL})\}}\middle| C_i = 0, D_i > 0\right] \\
&= \frac{(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})}{\exp\{-G(S_{iL})\}-\exp\{-G(S_{iR})\}}\int_0^\infty \frac{\xi_i\{e^{-\xi_i S_{iL}}-e^{-\xi_i S_{iR}}\}}{1-e^{-\xi_i(S_{iR}-S_{iL})}} f(\xi_i)d\xi_i \\
&= \frac{(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})}{\exp\{-G(S_{iL})\}-\exp\{-G(S_{iR})\}}\int_{\xi_i} \xi_i \exp(-\xi_i S_{iL}) f(\xi_i)d\xi_i \\
&= \frac{(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})}{\exp\{-G(S_{iL})\}-\exp\{-G(S_{iR})\}}\exp\{-G(S_{iL})\}G'(S_{iL}).
\end{aligned}
$$

S-step. Replacing W_{ik} and ξ_i with $\hat{E}(W_{ik})$ and $\hat{E}(\xi_i)$ in (4.14), we then solve for θ. Note that (4.14) is a large-dimensional nonlinear equation, and is not easy to be solved simultaneously. Thus, we propose the following nonlinear Gauss-Seidel method (Ortega and Rheinboldt, 1970; Ortega, 1972).

Step 1. Fix β, update a_k $(k = 1, \ldots, m)$ by solving

$$
\begin{cases}
\sum_{i=1}^n I(t_1 \leq R_i^*)\left\{\hat{E}(W_{i1}) - \hat{E}(\xi_i)(X_{i1}^\top a_1)\exp(\beta^\top Z_{i1})\right\}X_{i1} = 0 \\
\qquad\qquad\vdots \\
\sum_{i=1}^n I(t_m \leq R_i^*)\left\{\hat{E}(W_{im}) - \hat{E}(\xi_i)(X_{im}^\top a_m)\exp(\beta^\top Z_{im})\right\}X_{im} = 0.
\end{cases}
\tag{4.15}
$$

It is noted that for fixed β, equation (4.15) becomes linear with respect to a_k $(k = 1, \ldots, m)$. We easily calculate

$$
a_k = \left\{\sum_{i=1}^n I(t_k \leq R_i^*)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})X_{ik}X_{ik}^\top\right\}^{-1}\left\{\sum_{i=1}^n I(t_k \leq R_i^*)\hat{E}(W_{ik})X_{ik}\right\},
\tag{4.16}
$$

for $k = 1, \ldots, m$.

Step 2. Fix $a_1, \ldots, a_m$, we update β by solving the following equation via the

Newton-Raphson method:

$$\sum_{i=1}^{n}\sum_{k=1}^{m} I(t_k \le R_i^*) \left\{ \hat{E}(W_{ik}) - \hat{E}(\xi_i)(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik}) \right\} Z_{ik} = 0. \tag{4.17}$$

The S-step is declared convergent when the total absolute difference between the estimates obtained in two consecutive iterations falls below a predefined threshold, such as 10^{-3}.

We alternate between the E- and S-steps until convergence and denote the final estimates by $\hat{\theta} = (\hat{a}_1^\top, \ldots, \hat{a}_1^\top, \hat{\beta})^\top$. In this chapter, we employed a convergence criterion based on the largest absolute difference in the parameter estimates between two consecutive iterations. Specifically, we considered convergence to have been achieved when the largest absolute difference in the estimates fell below a predetermined threshold, such as 5×10^{-3}. We also explored the effect of using different threshold values and found that various small values led to similar results. A natural estimator for $A(t)$ is $\hat{A}(t) = \sum_{t_k \le t} \hat{a}_k$ for $0 \le t \le \tau$. Recall that $A(t) = \int_0^t \alpha(s)ds$, hence we can estimate $\alpha(t)$, $0 \le t \le \tau$ via a kernel estimator

$$\hat{\alpha}(t) = \sum_{k=1}^{m} h^{-1} K\left(\frac{t - t_k}{h}\right) \hat{a}_k,$$

where $K(x)$ is the kernel function and h is the bandwidth. Here, we choose the Epanechnikov kernel function, i.e., $K(x) = \frac{3}{4}\max\{1 - x^2, 0\}$.

The derived ES algorithm advances the maximization of the observed likelihood in several directions. First, closed expressions of the posterior means of W_{ik} and ξ are obtained in the E-step. Second, it only involves the low-dimensional parameter β in each iteration because of the explicit calculation of the high-dimensional parameters a_k ($k = 1, \ldots, m$) in the S-step. Third, when X is a vector of design variables for categories, the proposed ES estimator is efficient because the corresponding ES algorithm coincides with the EM algorithm proposed in Section 4.2.3 by observing that for fixed β, (4.11) and (4.15) share the same solution in terms of a_k ($k = 1, \ldots, m$). Similarly, when $X \equiv 1$, it can be shown that the ES algorithm coincides with the EM algorithms proposed in Zhou et al. (2021) with partly interval-censored data. Finally, we remark that (4.15) can be considered as a weighted version of (4.11), where each subject i receives a weight $X_{ik}^\top a_k$.

4.3 Variance Estimator

Note that $\hat{E}(W_{ik})$ and $\hat{E}(\xi_i)$ ($i = 1, \ldots, n; k = 1, \ldots, m$) can be represented as functions of the unknown parameter θ and the observed data $\mathcal{O}_i$. Write $\hat{E}(W_{i1}) = f_1(\mathcal{O}_i, \theta)$, $\ldots$, $\hat{E}(W_{im}) = f_m(\mathcal{O}_i, \theta)$ and $\hat{E}(\xi_i) = g(\mathcal{O}_i, \theta)$. Plugging those functions back in (4.15) and (4.17), the proposed ES algorithm is intrinsically solving a system of observed-data estimating equations. However, the high dimensionality of the parameter θ can pose computational challenges for estimating the variance of θ. Alternatively, we suggest employing the weighted bootstrap procedure, which works reasonably well in our settings.

Let $e_1, \ldots, e_n$ be i.i.d exponential random variables with mean one, which are independent of the observed data $\mathcal{O} = (\mathcal{O}_1, \ldots, \mathcal{O}_n)$ defined in Section (4.2.1). Let $\overline{e} = n^{-1} \sum_{i=1}^n e_i$ and $\widetilde{e}_i = e_i / \overline{e}$. In addition, let $\widetilde{U}(\theta) = (\widetilde{U}_{a_1}, \ldots, \widetilde{U}_{a_m}, \widetilde{U}_\beta)$ be the weighted version of $U(\theta)$ constructed in Section 4.2.4, where

$$\widetilde{U}_{a_k} = \sum_{i=1}^n I(t_k \leq R_i^*) \widetilde{e}_i \left\{ W_{ik} - \xi_i (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik}) \right\} X_{ik},$$

for $k = 1, \ldots, m$, and

$$\widetilde{U}_\beta = \sum_{i=1}^n \sum_{k=1}^m I(t_k \leq R_i^*) \widetilde{e}_i \left\{ W_{ik} - \xi_i (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik}) \right\} Z_{ik}.$$

Let $\widetilde{\theta} = (\widetilde{a}_1^\top, \ldots, \widetilde{a}_m^\top, \widetilde{\beta}^\top)^\top$ denotes estimator that solves the weighted estimating equation $\widetilde{U}(\theta) = 0$, which can be achieved through the proposed ES algorithm in

Section 4.2.4 only with trivial modifications. Specifically, in the E-step, the posterior means of W_{ik} and ξ_i $(i = 1, \ldots, n; k = 1, \ldots, m)$, denoted by $\widetilde{E}(W_{ik})$ and $\widetilde{E}(\xi_i)$, have the same expressions as $\hat{E}(W_{ik})$ and $\hat{E}(\xi_i)$ in Section 4.2.4. In the Step 1 of the S-step, for fixed $\widetilde{\beta}$, $\widetilde{a}_k$ $(k = 1, \ldots, m)$ can be updated via

$$\widetilde{a}_k = \left\{ \sum_{i=1}^{n} I(t_k \leq R_i^*)\widetilde{e}_i \widetilde{E}(\xi_i) \exp(\beta^\top Z_{ik}) X_{ik} X_{ik}^\top \right\}^{-1} \left\{ \sum_{i=1}^{n} I(t_k \leq R_i^*)\widetilde{e}_i \widetilde{E}(W_{ik}) X_{ik} \right\}.$$

In the Step 2 of the S-step, for fixed $\widetilde{a}_k$ $(k = 1, \ldots, m)$, we solve for $\widetilde{\beta}$ through the following

$$\sum_{i=1}^{n} \sum_{k=1}^{m} I(t_k \leq R_i^*)\widetilde{e}_i \left\{ \widetilde{E}(W_{ik}) - \widetilde{E}(\xi_i)(X_{ik}^\top a_k) \exp(\beta^\top Z_{ik}) \right\} Z_{ik} = 0.$$

Then one can iterate between the revised E-steps and S-steps until convergence. For each bootstrap replicate, we generate a set of bootstrap weights and run the revised ES algorithm to obtain $\widetilde{\theta}$. The variance of $\hat{\theta}$ can be estimated using the sample variance of the bootstrap replications, which are denoted as $\widetilde{\theta}$. The validity of the weighted bootstrap is proved in Appendix B.

4.4 Asymptotic Properties

We establish the asymptotic properties of the proposed estimators under the following regularity conditions:

Condition 1. With probability one, the vectors $X(t)$ and $Z(t)$ are uniformly bounded with uniformly bounded total variation over $[0, \tau]$. In addition, $X(t)$ is not identically zero for all $t \in [0, \tau]$.

Condition 2. Let $\mathcal{B}$ be a compact set of $\mathcal{R}^d$ and $BV[0, \tau]$ be the class of functions with bound variation over $[0, \tau]$. The true parameter (β_0, A_0) belongs to $\mathcal{B} \times BV^q[0, \tau]$ with β_0 an interior point of $\mathcal{B}$ and $A_0(t) = (A_{01}(t), \ldots, A_{0q}(t))^\top$ is continuously differentiable over $[0, \tau]$ with $A_0(0) = 0$. Here, $BV^q[0, \tau]$ denotes the product space

$BV[0, \tau] \times \cdots \times BV[0, \tau]$.

Condition 3. $0 < P(\Delta = 0) < 1$. For $\Delta = 0$, the number of monitoring times, K, is positive, and $E(K) < \infty$. In addition, there exists some constant $c > 0$ such that $P(U_{j+1} - U_j \geq c | K, X, Z, \Delta = 0) = 1$ ($j = 1, \ldots, K - 1$).

Condition 4. The transformation function G is thrice continuously differentiable on $[0, \infty)$ with $G(0) = 0$, $G'(x) > 0$ and $G(\infty) = \infty$.

Condition 5. If there exists a vector η and a deterministic function $\eta_0(t)$ such that $\eta_0(t) + \eta^\top X(t) = 0$ with probability one, then $\eta = 0$ for all $t \in [0, \tau]$ and $\eta_0(t) = 0$ for all $t \in [0, \tau]$.

Remark 4.1. *Conditions 1 and 2 state the boundedness of the covariates and the compactness of the Euclidean parameter space, which are standard used in survival analysis. Condition 3 is a conventional condition for partly interval-censored data, which requires that the smallest interval that brackets the failure time must be separated by at least c. Condition 3 also ensures that the proportion of exact observations is non-negligible. The smoothness condition for the joint density of* (U_j, U_{j+1}) *is used to prove the Donsker property of some function classes. Condition 4 ensures that the transformation function G is strictly increasing on* $[0, \infty)$*. Condition 5 ensures the existence and uniqueness of the jump sizes in* (4.16)*.*

Theorem 4.1. *Under Conditions* $1 - 5$*, the proposed ES estimator* $(\hat{\beta}, \hat{A})$ *is strongly consistent to* (β_0, A_0)*.*

Theorem 4.2. *Under Conditions* $1 - 5$*,* $\sqrt{n}(\hat{\beta} - \beta_0, \hat{A}(t) - A_0(t))$ *converges weakly to a zero-mean Gaussian process* $t \in [0, \tau]$*.*

Theorem 4.3. *Under Conditions* $1 - 5$*, the conditional distribution of* $\sqrt{n}(\widetilde{\theta} - \hat{\theta})$ *given the data converges weakly to the asymptotic distribution of* $\sqrt{n}(\hat{\theta} - \theta_0)$*.*

Detailed proofs of the above theorems are presented in Appendix B.

4.5 Simulation Studies

We conducted extensive simulation studies to examine the finite sample performance of the proposed estimators. The failure time T is generated from the following Cox-Aalen transformation model

$$\Lambda(t) = G\left[\int_0^t \exp\{\beta_1 Z_1(s) + \beta_2 Z_2\} d\Lambda_X(s)\right].$$

Here, $Z_1(t) = B_1 I(t \le V) + B_2 I(t > V)$ is a time-dependent covariate, with B_1 and B_2 being independent Ber(0.5) random variables and $V \sim \text{Unif}(0, 3)$. Additionally, $Z_2 \sim \text{Unif}(0, 1)$ is time-independent. To examine the effects of different transformations, we use a class of logarithmic transformations $G(x) = \log(1 + rx)/r$ with $r = 0, 0.5$, and 1, and note that $r = 0$ aligns with the Cox-Aalen model. We assigned the value of 0.5 to β_1 and the value of -0.5 to β_2 for all simulation scenarios, and considered following nnumerical scenarios for $\Lambda_X(s) = \int_0^s X^\top(u) dA(u)$ with $A(t) = (A_1(t), \ldots, A_q(t))^\top$:

Scenario 1. $X = (1, X_2)^\top$ with $X_2 \sim \text{Ber}(0.4)$, $A_1(t) = \log(1 + t/2)$ and $A_2(t) = 0.1t$.

Scenario 2. $X = (1, X_2)^\top$ with $X_2 \sim \text{Unif}(0, 1)$, $A_1(t) = \log(1 + t/2)$ and $A_2(t) = 0.1t$.

Scenario 3. Let J be a categorical variable that takes values in $\{1, 2, 3\}$ with equal probability, i.e., 1/3. Let $X = (1, X_2, X_3)^\top$, where $X_2 = I(J = 2)$, $X_3 = I(J = 3)$, $A_1(t) = \log(1 + t/2)$, $A_2(t) = 0.1t$ and $A_3(t) = 0.05t$.

For all setups, we let $\tau = 5$ be the duration of the study. For each study participant, we generate at least two monitoring times $U_1 \sim \text{Unif}(0, \tau/2)$ and $U_2 \sim \min\{0.1 + U_1 + \text{Unif}(0, \tau/2), \tau\}$. If $U_2 < \tau$, we continue generating the third monitoring time $U_3 \sim \min\{0.1 + U_2 + \text{Unif}(0, \tau/2), \tau\}$, and if $U_3 < \tau$, we generate one last monitoring time $U_4 \sim \min\{0.1 + U_3 + \text{Unif}(0, \tau/2), \tau\}$. Thus, the time axis $(0, \infty)$ is partitioned into at least three intervals and at most five intervals. We let $(L, R]$ be smallest interval that brackets the failure time T. In particular, if $R = \infty$, we set $\Delta = 0$. Otherwise, we generate $\Delta \sim \text{Ber}(\gamma)$. If $\Delta = 1$, the failure time is assumed to be exactly observed. Moreover, we can interpret γ as the proportion of failure observations that are exactly observed among those that are not right-censored. We investigate the impact of different levels of interval censoring by setting $\gamma = 0, 0.25, 0.5, 0.75$, or 1, where $\gamma = 0$ and $\gamma = 1$ correspond to purely interval- and right-censored data, respectively, and intermediate γ values generate partly interval-censored data. On average, the right-censoring rates range from 25% to 45% across all experimental setups.

For each experimental scenario, we initialized the proposed ES algorithm with $\beta = 0$ and $a_k = (1/m, 0, \ldots, 0)$ for $k = 1, \ldots, m$. The estimated parameters were found to be highly robust to different initial values. To estimate the variance of the proposed estimators, we used the weighted bootstrap procedure described in Section 4.3. The sample sizes were set to 200, 500, and 1000. We conducted 1000 simulation replicates for each scenario, and the variance estimate was obtained from 1000 bootstrap samples.

Table 4.1, 4.2, and 4.3 provide a summary of the parameter estimation results for Scenarios $1 - 3$. The simulation results demonstrate that the proposed methods perform exceptionally well in all experimental scenarios. The parameter estimators are nearly unbiased, the variance estimators obtained via weighted bootstrapping are very accurate, and the resulting confidence intervals exhibit correct coverage probabilities. As expected, the variance estimator tends to become smaller as the sample size increases and the proportion of exact observations grows. Figures 4.1, 4.2, and 4.3 provide evidence that the proposed estimation procedures for $A_1(\cdot)$ and $A_2(\cdot)$ have negligible bias, with the standard error of estimation (SEE) closely approximating the empirical standard error (SE), and a reasonable probability of coverage, further affirming their reliability. For the other values of γ, the estimated cumulative regression functions resemble those shown in Figure 4.1, and are consequently omitted from this dissertation.

4.6 Data Application

In this section, we considered a randomized trial, the AIDS Clinical Trials Group (ACTG) 175 trial, which enrolled a total of 2467 HIV-1-infected patients whose CD4 cell counts were from 200 to 500 cubic millimeter (Hammer et al., 1996). The trial included both patients who had received prior antiretroviral therapy (ART-experienced) and those who had not (ART-naive). The primary goal of the study was to compare the effectiveness of four antiretroviral regimens – zidovudine only, zidovudine and didanosine, zidovudine and zalcitabine, and didanosine only – in reducing mortality or AIDS morbidity among HIV-1-infected patients (Hammer et al., 1996). The study's primary endpoint was a composite outcome of death or development of AIDS, which has been historically used as a common measure of disease progression and treatment efficacy in HIV/AIDS research. The participants were randomly assigned to one of the four antiretroviral regimens, examined at weeks 2, 4, and 8, followed by assessments every 12 weeks. Their CD4 cell counts were measured at each of these follow-up visits, with the first measurement taken at week 8. Thus, if death occurs before the development of AIDS, the exact failures were determined. However, if AIDS occurs first, then we didn't know the exact time points instead of an interval due to the periodical examinations, i.e., interval-censored. Thus, the data consists of exact and interval-censored observations.

We excluded 10 participants without CD4 cell counts measurements, resulting in a total of 2457 HIV-1-infected participants in the full cohort, with 1396 and 1061 in the ART-experienced and ART-naive groups, respectively. Among the full cohort, there were a total of 306 cases observed, which represented 12.45% of the cohort. Within these cases, there were 230 AIDS events and 76 deaths, indicating 9.36% interval-censored observations and 3.09% exact observations. Of the cases observed, 215 occurred in the ART-experienced group, with 167 AIDS events and 48 deaths,

Table 4.1: Estimation results of the regression parameter β under Scenario 1

		$\beta_1 = 0.5$					$\beta_2 = -0.5$			
n	γ	Bias	SE	SEE	CP		Bias	SE	SEE	CP
						$r = 0$				
200	0	0.017	0.194	0.198	0.949		−0.012	0.299	0.300	0.956
	0.25	0.007	0.185	0.185	0.943		−0.002	0.292	0.293	0.955
	0.5	0.005	0.177	0.177	0.946		−0.002	0.290	0.290	0.956
	0.75	0.003	0.171	0.171	0.951		−0.003	0.290	0.288	0.951
	1	0.003	0.166	0.167	0.948		−0.000	0.287	0.287	0.954
500	0	0.006	0.116	0.119	0.965		0.003	0.184	0.186	0.954
	0.25	0.002	0.113	0.115	0.958		0.004	0.178	0.184	0.953
	0.5	0.001	0.110	0.111	0.950		0.004	0.177	0.183	0.954
	0.75	0.001	0.107	0.108	0.948		0.003	0.176	0.182	0.954
	1	0.002	0.105	0.106	0.949		0.003	0.175	0.181	0.951
1000	0	0.004	0.083	0.083	0.950		−0.009	0.130	0.131	0.951
	0.25	0.005	0.081	0.081	0.951		−0.006	0.129	0.130	0.947
	0.5	0.003	0.079	0.078	0.948		−0.005	0.128	0.129	0.945
	0.75	0.002	0.077	0.076	0.954		−0.005	0.129	0.129	0.950
	1	0.003	0.075	0.075	0.954		−0.004	0.128	0.128	0.950
						$r = 0.5$				
200	0	0.020	0.249	0.251	0.950		0.009	0.387	0.397	0.959
	0.25	0.003	0.235	0.228	0.948		0.015	0.378	0.385	0.960
	0.5	0.001	0.225	0.215	0.940		0.017	0.373	0.378	0.955
	0.75	−0.002	0.218	0.207	0.939		0.019	0.369	0.375	0.954
	1	−0.002	0.213	0.200	0.942		0.018	0.367	0.372	0.961
500	0	0.001	0.140	0.148	0.964		0.003	0.238	0.243	0.953
	0.25	0.001	0.136	0.140	0.954		0.004	0.237	0.240	0.944
	0.5	−0.001	0.132	0.134	0.960		0.006	0.235	0.237	0.948
	0.75	−0.001	0.127	0.130	0.956		0.007	0.233	0.236	0.948
	1	0.001	0.125	0.127	0.953		0.006	0.232	0.235	0.944
1000	0	−0.001	0.103	0.103	0.949		−0.003	0.171	0.171	0.947
	0.25	0.000	0.100	0.098	0.951		−0.001	0.170	0.170	0.951
	0.5	−0.002	0.097	0.095	0.942		0.000	0.169	0.168	0.952
	0.75	0.000	0.093	0.092	0.949		−0.002	0.169	0.167	0.945
	1	0.002	0.091	0.090	0.945		−0.002	0.169	0.167	0.953
						$r = 1$				
200	0	0.025	0.291	0.294	0.953		0.011	0.460	0.477	0.963
	0.25	0.005	0.270	0.260	0.946		0.023	0.446	0.455	0.956
	0.5	0.001	0.252	0.243	0.939		0.025	0.442	0.446	0.954
	0.75	−0.000	0.248	0.232	0.938		0.020	0.438	0.442	0.960
	1	−0.001	0.238	0.224	0.940		0.021	0.434	0.438	0.958
500	0	−0.004	0.163	0.171	0.963		0.009	0.287	0.290	0.945)
	0.25	−0.008	0.155	0.159	0.953		0.009	0.286	0.285	0.946)
	0.5	−0.008	0.149	0.152	0.949		0.013	0.283	0.281	0.940
	0.75	−0.003	0.144	0.147	0.958		0.012	0.278	0.279	0.948
	1	−0.002	0.140	0.142	0.962		0.010	0.278	0.277	0.941
1000	0	−0.002	0.118	0.118	0.948		0.003	0.207	0.203	0.944
	0.25	−0.004	0.114	0.112	0.951		0.002	0.206	0.201	0.941
	0.5	−0.004	0.111	0.108	0.946		0.004	0.205	0.199	0.939
	0.75	0.000	0.105	0.104	0.949		0.003	0.204	0.198	0.937
	1	0.001	0.103	0.101	0.943		0.000	0.203	0.197	0.945

Note: See the note to Table 3.1

Table 4.2: Estimation results of the regression parameter β under Scenario 2

		$\beta_1 = 0.5$				$\beta_2 = -0.5$			
n	γ	Bias	SE	SEE	CP	Bias	SE	SEE	CP
					$r = 0$				
200	0	−0.008	0.192	0.186	0.952	0.005	0.298	0.291	0.945
	0.25	0.001	0.180	0.179	0.954	0.014	0.294	0.288	0.943
	0.5	−0.005	0.176	0.173	0.945	0.007	0.288	0.286	0.946
	0.75	−0.001	0.172	0.169	0.945	0.008	0.286	0.285	0.944
	1	−0.001	0.169	0.166	0.945	0.009	0.283	0.285	0.944
500	0	0.006	0.116	0.118	0.957	0.002	0.184	0.185	0.956
	0.25	0.009	0.113	0.116	0.950	−0.004	0.181	0.189	0.956
	0.5	0.003	0.109	0.111	0.944	0.001	0.177	0.184	0.953
	0.75	0.000	0.106	0.108	0.948	0.003	0.175	0.181	0.954
	1	0.001	0.104	0.105	0.947	0.004	0.173	0.179	0.956
1000	0	0.003	0.083	0.083	0.948	−0.009	0.129	0.130	0.948
	0.25	0.006	0.081	0.081	0.954	−0.009	0.129	0.131	0.952
	0.5	0.004	0.078	0.078	0.949	−0.006	0.128	0.129	0.948
	0.75	0.004	0.077	0.076	0.946	−0.005	0.128	0.128	0.942
	1	0.004	0.076	0.074	0.948	−0.005	0.128	0.127	0.946
					$r = 0.5$				
200	0	−0.010	0.239	0.234	0.941	0.024	0.383	0.377	0.951
	0.25	−0.011	0.226	0.214	0.936	0.030	0.380	0.364	0.940
	0.5	−0.001	0.222	0.210	0.933	0.020	0.376	0.370	0.947
	0.75	−0.006	0.210	0.203	0.936	0.026	0.366	0.367	0.949
	1	−0.002	0.207	0.199	0.946	0.020	0.364	0.370	0.952
500	0	0.001	0.143	0.147	0.956	0.007	0.235	0.241	0.948
	0.25	0.012	0.141	0.141	0.956	−0.008	0.242	0.242	0.944
	0.5	0.002	0.133	0.133	0.950	0.002	0.235	0.236	0.947
	0.75	0.001	0.129	0.129	0.952	0.007	0.230	0.234	0.954
	1	−0.001	0.126	0.126	0.957	0.007	0.229	0.234	0.946
1000	0	-0.004	0.104	0.102	0.946	-0.000	0.170	0.169	0.948
	0.25	0.003	0.101	0.097	0.937	−0.004	0.172	0.169	0.945
	0.5	−0.000	0.097	0.094	0.937	0.001	0.169	0.167	0.945
	0.75	0.000	0.094	0.091	0.945	0.001	0.169	0.166	0.946
	1	0.002	0.091	0.089	0.949	−0.002	0.168	0.166	0.951
					$r = 1$				
200	0	−0.018	0.265	0.266	0.944	0.061	0.454	0.430	0.937
	0.25	−0.005	0.258	0.244	0.931	0.074	0.447	0.424	0.931
	0.5	−0.003	0.258	0.235	0.931	0.032	0.443	0.431	0.945
	0.75	−0.001	0.248	0.229	0.934	0.025	0.434	0.434	0.958
	1	−0.002	0.238	0.222	0.941	0.023	0.432	0.436	0.954
500	0	0.001	0.172	0.168	0.939	0.021	0.289	0.281	0.946
	0.25	−0.007	0.158	0.159	0.947	0.014	0.289	0.284	0.939
	0.5	−0.001	0.152	0.151	0.961	0.008	0.284	0.279	0.944
	0.75	−0.007	0.144	0.146	0.952	0.011	0.276	0.277	0.950
	1	−0.003	0.139	0.142	0.953	0.010	0.275	0.277	0.949
1000	0	−0.007	0.116	0.116	0.949	0.010	0.204	0.199	0.939
	0.25	0.006	0.112	0.111	0.939	0.003	0.204	0.200	0.941
	0.5	−0.002	0.110	0.107	0.944	0.007	0.203	0.198	0.942
	0.75	−0.001	0.104	0.103	0.953	0.006	0.202	0.197	0.944
	1	0.002	0.102	0.101	0.956	−0.000	0.201	0.197	0.947

Note: See the note to Table 3.1

Table 4.3: Estimation results of the regression parameter β under Scenario 3

		$\beta_1 = 0.5$				$\beta_2 = -0.5$			
n	γ	Bias	SE	SEE	CP	Bias	SE	SEE	CP
					$r = 0$				
200	0	0.016	0.200	0.205	0.952	−0.030	0.308	0.300	0.935
	0.25	0.001	0.190	0.186	0.950	−0.016	0.300	0.291	0.931
	0.5	−0.003	0.183	0.176	0.934	−0.012	0.294	0.287	0.940
	0.75	−0.007	0.174	0.170	0.944	−0.012	0.292	0.284	0.937
	1	−0.005	0.172	0.165	0.938	−0.009	0.289	0.282	0.939
500	0	0.001	0.118	0.119	0.955	−0.016	0.188	0.185	0.949
	0.25	−0.001	0.112	0.114	0.956	−0.010	0.187	0.183	0.947
	0.5	−0.001	0.109	0.110	0.955	−0.010	0.184	0.181	0.945
	0.75	−0.004	0.107	0.107	0.954	−0.010	0.182	0.180	0.949
	1	−0.003	0.103	0.104	0.955	−0.007	0.181	0.179	0.942
1000	0	0.003	0.083	0.083	0.948	-0.011	0.129	0.130	0.949
	0.25	0.003	0.081	0.080	0.940	−0.007	0.128	0.129	0.946
	0.5	0.002	0.078	0.077	0.946	−0.007	0.127	0.128	0.947
	0.75	0.003	0.077	0.076	0.945	−0.005	0.127	0.127	0.950
	1	0.004	0.076	0.074	0.946	−0.003	0.126	0.126	0.947
					$r = 0.5$				
200	0	0.016	0.248	0.263	0.965	−0.030	0.394	0.401	0.953
	0.25	−0.005	0.230	0.231	0.948	−0.013	0.380	0.383	0.958
	0.5	−0.011	0.222	0.214	0.940	−0.006	0.374	0.374	0.952
	0.75	−0.007	0.214	0.205	0.941	−0.008	0.370	0.369	0.955
	1	−0.008	0.210	0.197	0.933	−0.005	0.366	0.366	0.956
500	0	−0.003	0.149	0.149	0.945	−0.015	0.238	0.243	0.954
	0.25	−0.006	0.141	0.140	0.951	−0.013	0.238	0.239	0.949
	0.5	−0.006	0.135	0.134	0.941	−0.009	0.235	0.236	0.951
	0.75	−0.007	0.130	0.129	0.944	−0.008	0.231	0.234	0.957
	1	−0.006	0.126	0.126	0.945	−0.007	0.231	0.232	0.951
1000	0	0.003	0.104	0.102	0.955	−0.005	0.164	0.169	0.959
	0.25	0.002	0.099	0.098	0.955	−0.003	0.163	0.168	0.954
	0.5	0.001	0.096	0.094	0.953	0.002	0.160	0.166	0.960
	0.75	0.004	0.094	0.091	0.935	0.003	0.161	0.165	0.947
	1	0.004	0.091	0.089	0.943	0.003	0.161	0.164	0.958
					$r = 1$				
200	0	0.030	0.296	0.309	0.959	−0.013	0.479	0.482	0.949
	0.25	0.003	0.274	0.265	0.941	0.015	0.449	0.456	0.963
	0.5	−0.002	0.257	0.243	0.934	0.021	0.437	0.443	0.957
	0.75	−0.003	0.250	0.230	0.936	0.022	0.433	0.436	0.959
	1	−0.000	0.242	0.221	0.931	0.021	0.432	0.432	0.958
500	0	0.006	0.167	0.173	0.953	0.014	0.288	0.290	0.958
	0.25	−0.001	0.160	0.159	0.946	0.013	0.284	0.283	0.950
	0.5	−0.004	0.149	0.151	0.953	0.009	0.276	0.278	0.951
	0.75	−0.003	0.145	0.146	0.950	0.012	0.272	0.275	0.946
	1	−0.000	0.139	0.141	0.955	0.011	0.273	0.274	0.949
1000	0	0.000	0.117	0.118	0.956	0.002	0.203	0.201	0.950
	0.25	−0.000	0.111	0.111	0.948	0.001	0.200	0.199	0.947
	0.5	0.000	0.107	0.107	0.949	0.004	0.199	0.196	0.949
	0.75	0.002	0.103	0.103	0.952	0.006	0.200	0.195	0.952
	1	0.003	0.101	0.100	0.950	0.003	0.199	0.195	0.946

Note: See the note to Table 3.1

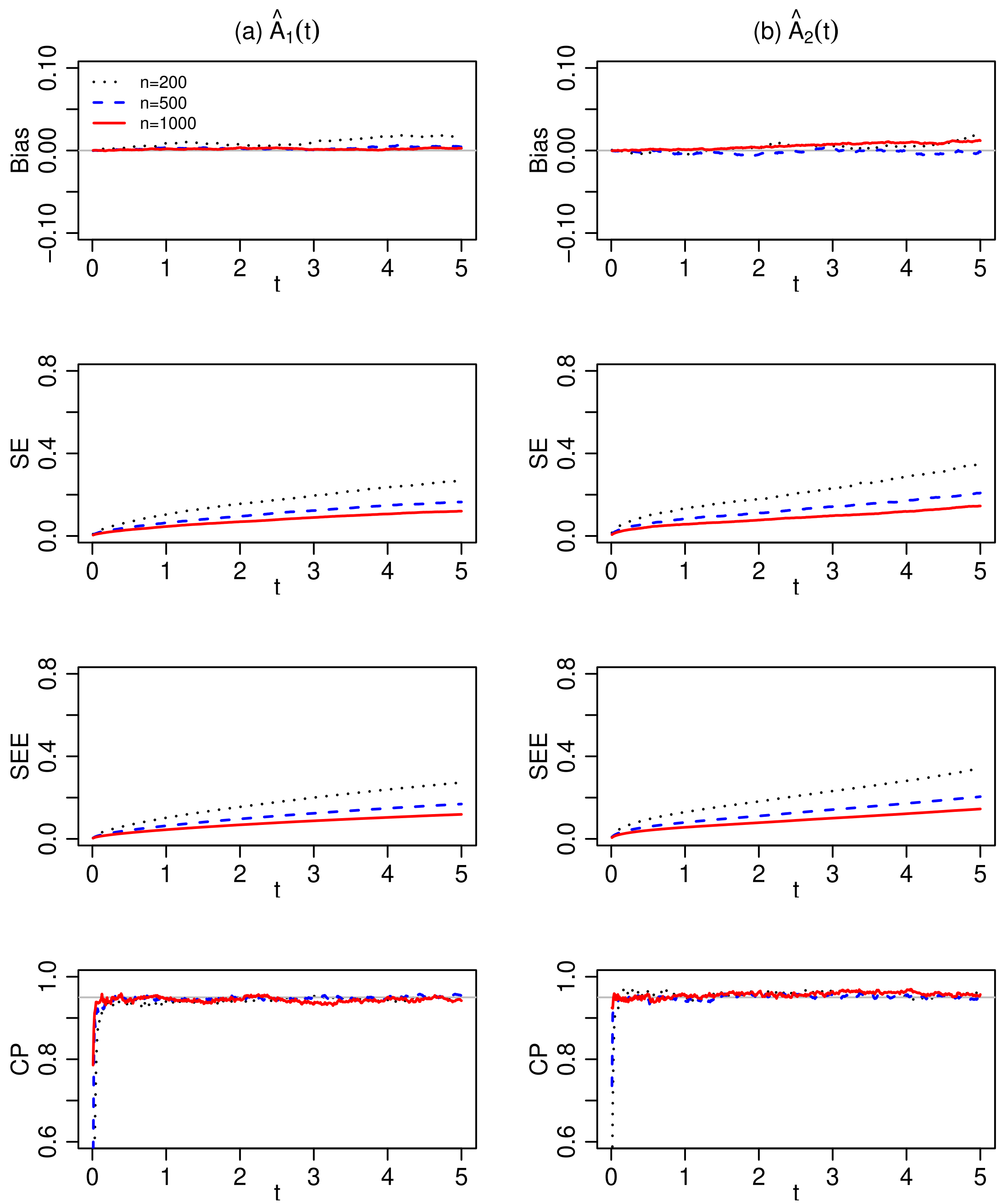

Figure 4.1: Estimation results for (a) $A_1(t) = \log(1 + t/2)$ and (b) $A_2(t) = 0.1t$ in Scenario 1 with $r = 0$. The dotted, dashed and solid lines represent $n = 200, 800$ and 1000, respectively. Bias, SE, SEE, and CP see the note to Table 4.1.

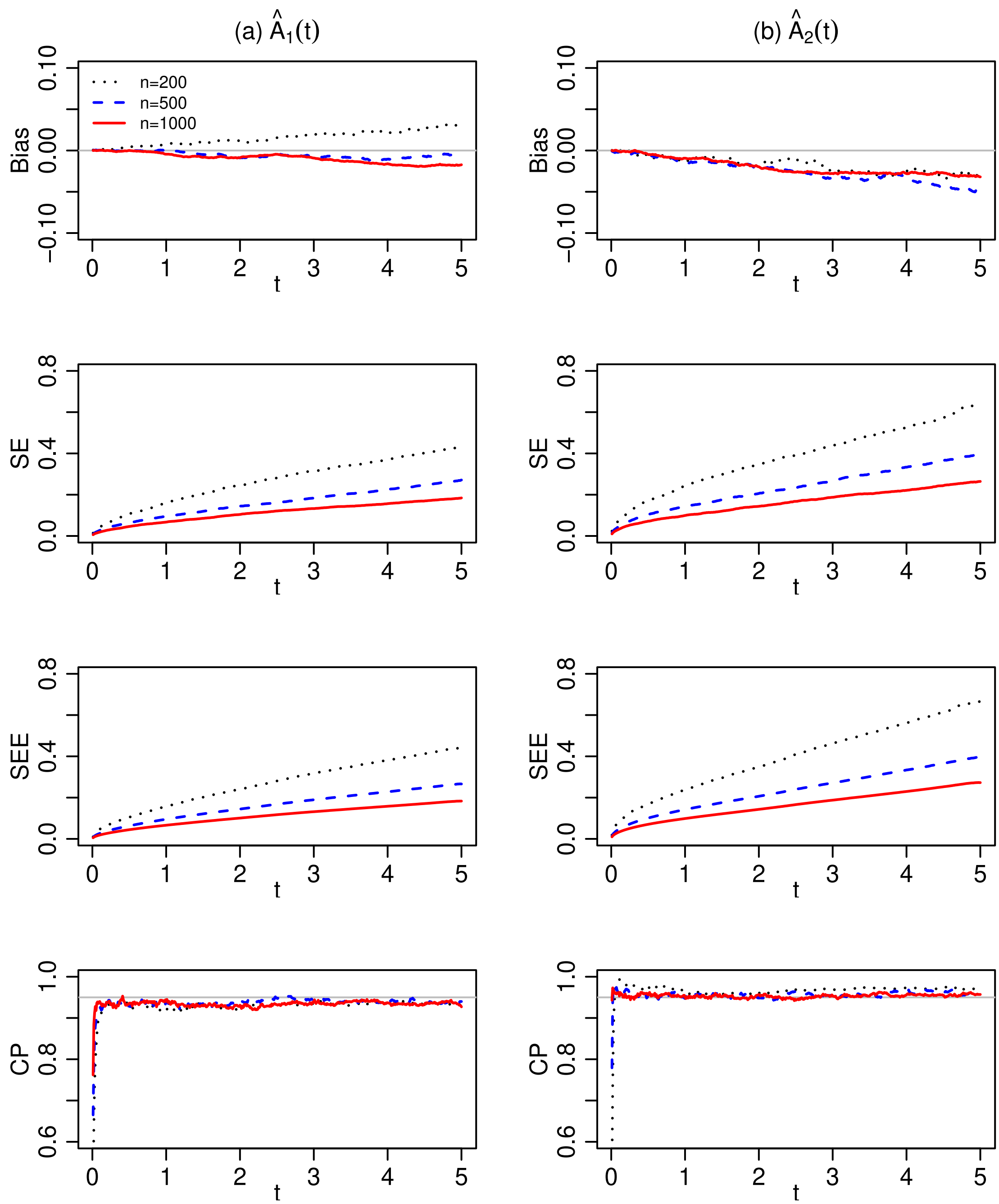

Figure 4.2: Estimation results for (a) $A_1(t) = \log(1 + t/2)$ and (b) $A_2(t) = 0.1t$ in Scenario 2 with $r = 0.5$. The dotted, dashed and solid lines represent $n = 200, 800$ and 1000, respectively. Bias, SE, SEE, and CP see the note to Table 4.1.

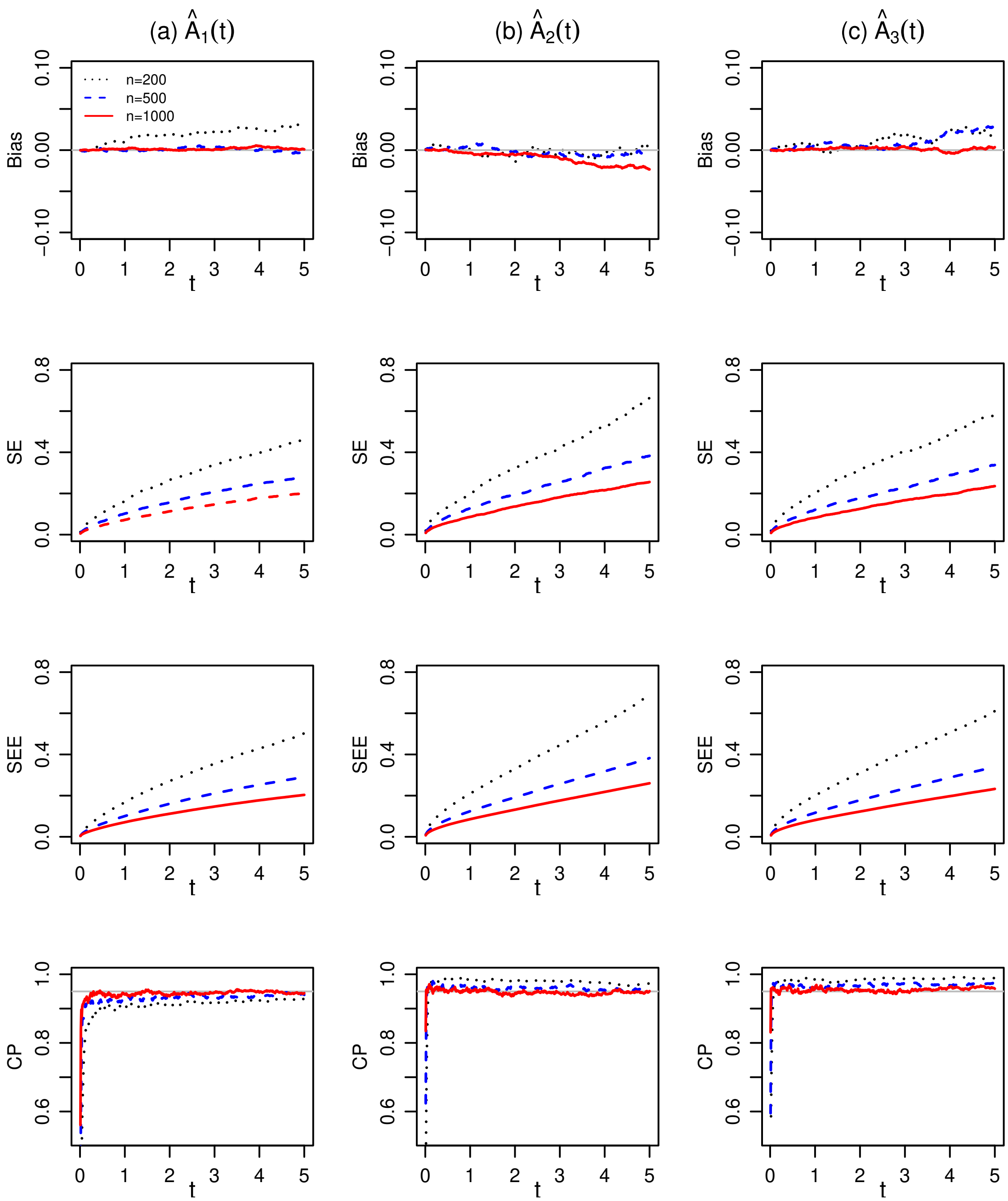

Figure 4.3: Estimation results for (a) $A_1(t) = \log(1 + t/2)$, (b) $A_2(t) = 0.1t$ and (c) $A_3(t) = 0.05t$ in Scenario 3 with $r = 1$. The dotted, dashed and solid lines represent $n = 200, 800$ and 1000, respectively. Bias, SE, SEE, and CP see the note to Table 4.1.

while 91 cases occurred in the ART-naive group, with 63 AIDS events and 28 deaths.

The following Cox-Aalen transformation model is used to assess the association between the time to AIDS or death and $\log_{10}$(CD4), treatment assignment (zidovudine + didanosine, zidovudine + zalcitabine, or didanosine), and ART experienced assignment, without assuming proportional hazards for patients who received prior ART versus those who did not. Moreover, we assume that the treatment effect is different between ART-experienced and ART-naive groups. Specifically, the model is defined as follows:

$$\Lambda(t \mid X, Z(\cdot)) = G\Big[\int_0^t \exp\Big\{\beta_1 Z_1(s) + \beta_2 Z_2 + \beta_3 (X_2 \cdot Z_2)\Big\} d\Lambda_X(s)\Big], \tag{4.18}$$

where β_1, β_2 and β_3 are unknown regression coefficients, $Z_1(s)$ is the time-dependent covariate $\log_{10}$(CD4), and Z_2 indicates whether or not the patient received treatment (1 for received, 0 for not received), $X = (1, X_2)^\top$, where X_2 is an indicator variable taking on values of 1 or 0 depending on whether the patient is in the ART-experienced or ART-naive group, respectively. In addition, $\Lambda_X(s) = \int_0^s X^\top dA(v)$, $A(v) = (A_1(v), A_2(v))^\top$.

We fitted model (4.18) with a logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$. To determine the optimal value of r, we plotted the loglikelihood against r, as shown in Figure 4.4. We considered values of r from 0 to 3 in increments of 0.1. Based on this plot, we selected $r = 2$ as the best-fit value. We present the estimation results for the selected model with $r = 2$ in Table 4.4, including estimates of the model parameters and their standard errors. For comparison, we also report the estimation results for $r = 0$ and $r = 1$. It is important to note that the choice of r can affect the model's performance, and it is advisable to consider a range of values and perform model selection based on various criteria.

Based on the all of the models considered, a lower value of $\log_{10}$(CD4) is associated

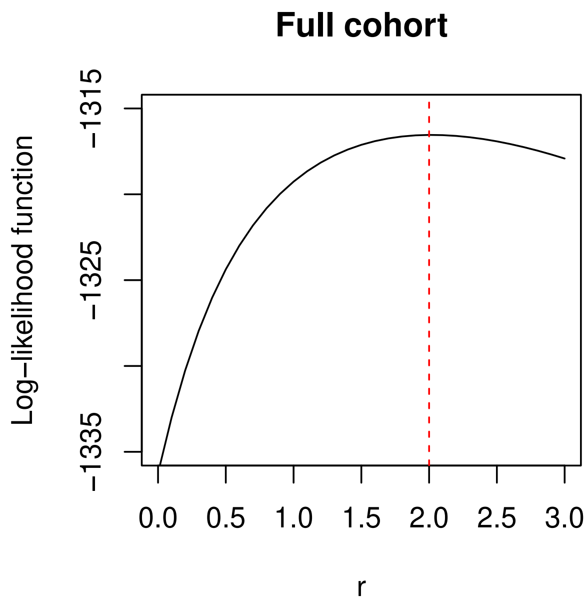

Figure 4.4: Log-likelihood function at the final parameter estimates in model (4.18) using $G(x) = r^{-1}\log(1+rx)$ with r values in the interval [0, 3] and a step size of 0.1.

Table 4.4: Regression analysis results for the ACTG 175 trial via the logarithmic transformation $G(x) = r^{-1}\log(1+rx)$

	$r=0$			$r=1$			Selected r		
Covariates	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
	The proposed model								
$\log_{10}$(CD4)	−2.749	0.119	0.000	−3.538	0.160	0.000	−4.050	0.187	0.000
Treatment	−0.524	0.224	0.019	−0.440	0.258	0.088	−0.388	0.281	0.167
Treatment × ART	0.621	0.281	0.027	0.678	0.335	0.043	0.641	0.369	0.082
	ART-naive								
$\log_{10}$(CD4)	−2.811	0.207	0.000	−3.607	0.310	0.000	−4.497	0.415	0.000
Treatment	−0.533	0.232	0.022	−0.439	0.262	0.094	−0.342	0.306	0.264
	ART-experienced								
$\log_{10}$(CD4)	−2.717	0.142	0.000	−3.503	0.193	0.000	−3.867	0.218	0.000
Treatment	0.094	0.169	0.578	0.232	0.212	0.274	0.245	0.230	0.287

Note: Est and SE stand for the estimates of the regression parameters and the estimated standard errors, respectively. Here, "ART" refers to the ART-naive group when ART = 0, and to the ART-experienced group when ART = 1. The selected r values for the proposed model, ART-naive group and ART-experienced group are 2, 2.8 and 1.7, respectively. Here, the estimation results for ART-naive and experienced groups are based on Zeng and Lin's model. SE is calculated via 1000 weighted bootstrapping samples.

with a significantly higher risk of time to death or AIDS onset, as shown in the upper panel of Table 4.4. The effect of treatment is significant under the model $r = 0$ and marginally significant under the model $r = 1$. Notably, the treatment effect differs between the ART-naive and experienced groups. The estimated coefficient for treatment is negative in the ART-naive group, indicating that patients in this group who receive treatment have a significantly lower risk of AIDS onset or death. In contrast, the estimated coefficient is positive in the ART-experienced group, suggesting that patients in this group who receive a placebo have a lower risk. Figure 4.5 supports these findings. Additionally, Figure 4.6 displays the estimated cumulative regression functions for patients in the ART-naive and ART-experienced groups. The plot reveals that the risk of death or AIDS onset varies over time and is not proportional within each group. In particular, patients in the ART-experienced group exhibit a higher risk than those in the ART-naive group at the start of the study and during an earlier stage. However, beyond this point, the risk becomes greater in the ART-naive group. One possible explanation for this pattern is drug resistance that may develop at an early stage in the ART-experienced group.

We also applied Zeng and Lin's model (Zeng and Lin, 2006) to the ART-naive and experienced groups separately to compare their results with those obtained using the proposed model. From the lower two panels of Table 4.4, we found that a lower value of $\log_{10}$(CD4) significantly reduces the risk of time to death or AIDS onset in both groups. However, for the ART-naive group, the p-value for the treatment under the selected Zeng and Lin's model is 0.264, which is higher than the p-value (0.167) for treatment obtained using the proposed model. Hence, the class of proposed Cox-Aalen transformation models, which can consider both ART-naive and experienced groups together by allowing for different baseline risk effects within each group, is suggested to be more flexible and has the potential to improve statistical power.

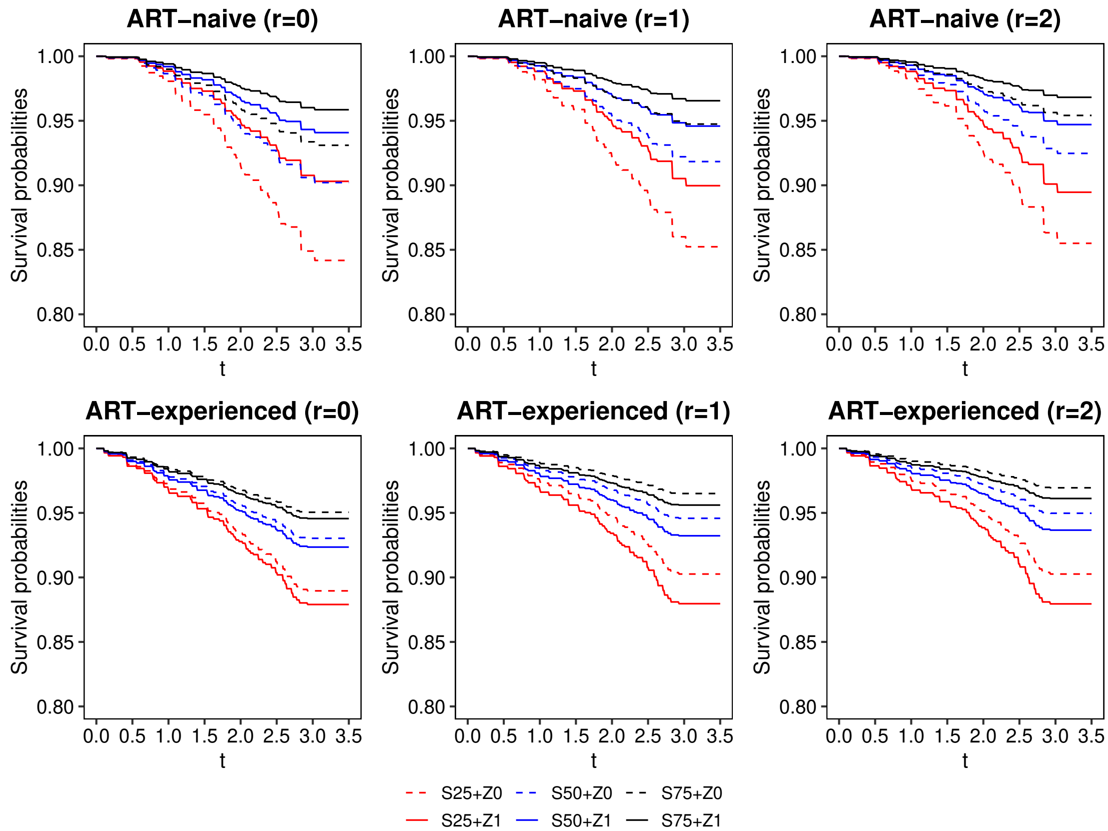

Figure 4.5: Estimated survival probabilities for ART experienced and naive groups. Here, S25, S50 and S75 represent the 25th, 50th and 75th percentile of $\log_{10}(\text{CD4})$, respectively. Moreover, Z0 and Z1 stand for the placebo and treatment groups, respectively.

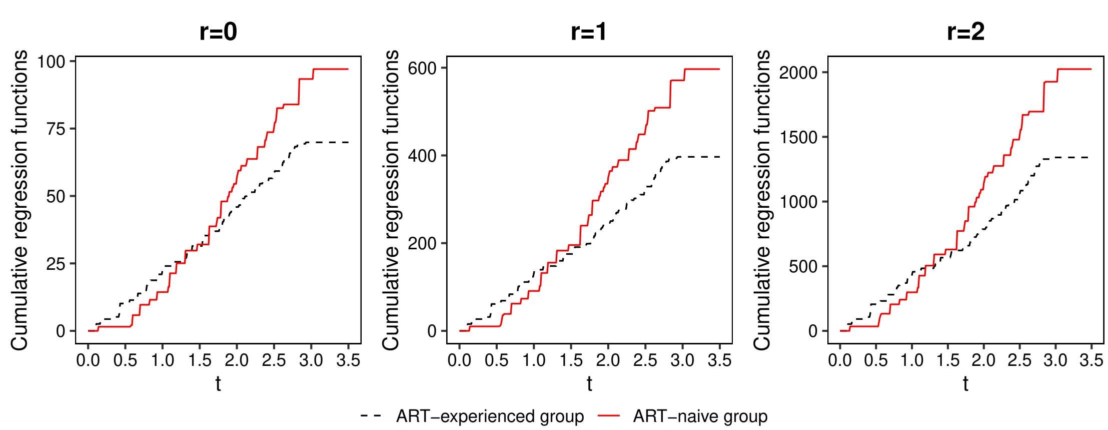

Figure 4.6: Estimated cumulative regression functions for ART-experienced and naive groups with $r = 0, 1,$ and 2 (selected), using $G(x) = r^{-1}\log(1 + rx)$.

4.7 Discussion

The Cox-Aalen transformation models are a versatile framework for modeling censored data, including the Cox-Aalen model (Scheike and Zhang, 2002) and the transformation models (Zeng and Lin, 2006) as special cases. Nevertheless, modeling partly interval-censored data within this class of regression models presents significant computational and theoretical challenges. To tackle these challenges, we proposed an estimating equation approach and derived a straightforward ES algorithm that enables fast and stable calculations. Moreover, we established the large sample properties of the proposed ES estimator by leveraging empirical process theory in a meticulous manner.

Previous studies on multiplicative hazards models or transformation models have focused primarily on maximum likelihood estimation. However, to the best of our knowledge, the maximum likelihood approach has not been extensively studied in the context of additive hazards models. Instead, ordinary least squares is commonly used to estimate the cumulative effect of covariates (Aalen, 1980, 1989; Huffer and McKeague, 1991; Scheike and Zhang, 2002). Nevertheless, performing maximum likelihood estimation with additive components can be highly challenging, even in the presence of right-censored data.

More recently, Boruvka and Cook (2015) explored semiparametric maximum likelihood estimation for the Cox-Aalen model with fixed covariates from interval-censored data. The proposed Cox-Aalen transformation model performed well in simulation studies with purely interval-censored data. However, it is important to note that the proposed model cannot be theoretically extended to purely interval-censored data. This limitation presents an interesting area for future research.

CHAPTER 5: CONCLUSIONS AND FUTURE WORK

This dissertation has presented a novel semiparametric transformation model that offers a more comprehensive and flexible approach to survival analysis. By incorporating both multiplicative and additive effects of covariates within the transformation, the proposed model can capture more complex relationships between covariates and survival time, resulting in improved accuracy of survival predictions. Through the application the proposed model to right-censored and partly interval-censored data scenarios, we have demonstrated its superior performance compared to existing models. The results from our simulation studies and real-world data analyses provide evidence of the model's effectiveness and highlight its potential for practical use in fields such as healthcare and social sciences.

Despite the promising results obtained in this dissertation, further research is necessary. Assessing the adequacy of the proposed model is also crucial because model misspecification affects the validity of inference and prediction accuracy. For instance, mis-specifying the transformation function can result in erroneous inferences. For Zeng and Lin's model, Chen et al. (2012) considered appropriate time-dependent residuals and constructed various graphical and numerical procedures for model assessment with censored data. In our analysis of the HIV-1 prevention trials and ACTG 175 trial, we use the log-likelihood to select the transformation function, even though the log-likelihood surface is relatively flat. Similar to Chen et al. (2012), we suggest constructing the cumulative sums of residuals over the argument of the transformation function to check the transformation form with censored data. Specifically, we propose using the function $W(x,t)$, defined as: $W(x,t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t I\left(\int_0^u Y_i(s)e^{\hat{\beta}^\top Z_i(s)}X_i^\top(s)d\hat{A}(s) \leq x\right)dM_i(u;\hat{\beta},\hat{A})$, where

$M_i(t;\beta,A) = N_i(t) - G\left\{\int_0^t Y_i(s)e^{\beta^\top Z_i(s)}X_i^\top(s)dA(s)\right\}$. A thorough theoretical and numerical investigation of model misspecification is still needed for the proposed model. We are currently pursuing this direction. Furthermore, it is crucial to note that the transformation form should also be checked for interval-censored data. One possible approach is to construct time-dependent residuals given the observed intervals. However, further examination and investigation are still required.

Another area for further research is to investigate the proposed model's performance in handling covariates with missingness. Incomplete data can significantly impact the accuracy of survival analysis, and it would be beneficial to evaluate the proposed model's performance under different types of missing data mechanisms for covariates. For example, if the covariate is missing at random, the inverse probability method can be used to estimate the missing covariate values based on the observed data and then include these imputed values in the survival analysis model. Ning et al. (2018) considered a class of weighted estimating equations for censored with missing covariates via linear transformation models to facilitate computation. Similar ideas may be applied to the proposed model, but further exploration and investigation are needed to assess the effectiveness and feasibility of this approach in the context of the proposed model.

REFERENCES

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

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in medicine*, 8(8):907–925.

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

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

Boruvka, A. and Cook, R. J. (2015). A Cox-Aalen model for interval-censored data. *Scandinavian Journal of Statistics*, 42(2):414–426.

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

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

Chen, L., Lin, D., and Zeng, D. (2012). Checking semiparametric transformation models with censored data. *Biostatistics*, 13(1):18–31.

Chen, L. and Sun, J. (2010). A multiple imputation approach to the analysis of interval-censored failure time data with the additive hazards model. *Computational statistics & data analysis*, 54(4):1109–1116.

Chen, Y.-H. (2009). Weighted Breslow-type and maximum likelihood estimation in semiparametric transformation models. *Biometrika*, 96(3):591–600.

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

Corey, L., Gilbert, P. B., Juraska, M., Montefiori, D. C., Morris, L., Karuna, S. T., Edupuganti, S., Mgodi, N. M., deCamp, A. C., Rudnicki, E., et al. (2021). Two randomized trials of neutralizing antibodies to prevent HIV-1 acquisition. *New England Journal of Medicine*, 384(11):1003–1014.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Dabrowska, D. M. and Doksum, K. A. (1988). Partial likelihood in transformation models with censored data. *Scandinavian journal of statistics*, pages 1–23.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Dudley, R. M. (2002). *Real analysis and probability*. Cambridge University Press,.

Elashoff, M. and Ryan, L. (2004). An EM algorithm for estimating equations. *Journal of Computational and Graphical Statistics*, 13(1):48–65.

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

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, pages 845–854.

Gao, F., Zeng, D., and Lin, D.-Y. (2017). Semiparametric estimation of the accelerated failure time model with partly interval-censored data. *Biometrics*, 73(4):1161–1168.

Gao, F., Zeng, D., and Lin, D.-Y. (2018). Semiparametric regression analysis of interval-censored data with informative dropout. *Biometrics*, 74(4):1213–1222.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, 335(15):1081–1090.

He, B., Liu, Y., Wu, Y., and Zhao, X. (2020). Semiparametric efficient estimation for additive hazards regression with case ii interval-censored survival data. *Lifetime Data Analysis*, 26(4):708–730.

Henderson, N. C. and Rathouz, P. J. (2018). AR (1) latent class models for longitudinal count data. *Statistics in medicine*, 37(29):4441–4456.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, 24(2):540–568.

Huang, J. and Rossini, A. (1997). Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*, 92(439):960–967.

Huffer, F. W. and McKeague, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, 86(413):114–129.

Johnson, L. M. and Strawderman, R. L. (2012). A smoothing expectation and substitution algorithm for the semiparametric accelerated failure time frailty model. *Statistics in Medicine*, 31(21):2335–2358.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure data.* Wiley.

Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):489–502.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer.

Lin, D. and Ying, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The annals of Statistics*, pages 1712–1734.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81(1):61–71.

Liu, X. and Zeng, D. (2013). Variable selection in semiparametric transformation models for right-censored data. *Biometrika*, 100(4):859–876.

Martinussen, T. and Scheike, T. H. (2002). A flexible additive multiplicative hazard model. *Biometrika*, 89(2):283–298.

Ning, Y., Yi, G., and Reid, N. (2018). A class of weighted estimating equations for semiparametric transformation models with missing covariates. *Scandinavian Journal of Statistics*, 45(1):87–109.

Ortega, J. M. (1972). *Numerical analysis: a second course.* SIAM.

Ortega, J. M. and Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables.* SIAM.

Pan, C., Cai, B., and Wang, L. (2020). A bayesian approach for analyzing partly interval-censored data under the proportional hazards model. *Statistical Methods in Medical Research*, 29(11):3192–3204.

Pettitt, A. N. (1982). Inference for the linear model using a likelihood based on ranks. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):234–243.

Polyanin, A. D. and Manzhirov, A. V. (2008). *Handbook of integral equations*. CRC press.

Qu, L. and Sun, L. (2019). The cox–aalen model for recurrent-event data with a dependent terminal event. *Statistica Neerlandica*, 73(2):234–255.

Rabinowitz, D., Betensky, R. A., and Tsiatis, A. A. (2000). Using conditional logistic regression to fit proportional odds models to interval censored data. *Biometrics*, 56(2):511–518.

Rudin, W. (1973). *Functional Analysis*. New York,NY: McGraw-Hill.

Scheike, T. H. and Zhang, M.-j. (2002). An additive–multiplicative Cox–Aalen regression model. *Scandinavian Journal of Statistics*, 29(1):75–88.

Scheike, T. H. and Zhang, M.-J. (2003). Extensions and applications of the Cox-Aalen survival model. *Biometrics*, 59(4):1036–1045.

Schwartz, J. T. (1969). *Nonlinear functional analysis*, volume 4. CRC Press.

Shen, P.-s. and Weng, L. N. (2018). The Cox–Aalen model for left-truncated and right-censored data. *Communications in Statistics-Theory and Methods*, 47(21):5357–5368.

Shen, P.-s. and Weng, L. N. (2019). The cox-aalen model for left-truncated and mixed interval-censored data. *Statistics*, 53(5):1152–1167.

van der Vaart, A. and Wellner, J. A. (1996a). *Weak convergence and empirical processes: with applications to statistics*. Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3. Cambridge university press.

van der Vaart, A. W. and Wellner, J. A. (1996b). *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media.

Wang, C., Jiang, J., and Song, X. (2022). Bayesian transformation models with partly interval-censored data. *Statistics in Medicine*, 41(7):1263–1279.

Wang, L., McMahan, C. S., Hudgens, M. G., and Qureshi, Z. P. (2016). A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics*, 72(1):222–231.

Wang, L., Sun, J., and Tong, X. (2010). Regression analysis of case ii interval-censored failure time data with the additive hazards model. *Statistica Sinica*, 20(4):1709.

Yu, G., Li, Y., Zhu, L., Zhao, H., Sun, J., and Robison, L. L. (2019). An additive–multiplicative mean model for panel count data with dependent observation and dropout processes. *Scandinavian Journal of Statistics*, 46(2):414–431.

Zeng, D., Cai, J., and Shen, Y. (2006). Semiparametric additive risks model for interval-censored data. *Statistica Sinica*, pages 287–302.

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

Zeng, D. and Lin, D. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):507–564.

Zeng, D., Mao, L., and Lin, D. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika*, 103(2):253–271.

Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the cox model with interval-censored data. *Scandinavian Journal of Statistics*, 37(2):338–354.

Zhang, Z., Sun, L., Zhao, X., and Sun, J. (2005). Regression analysis of interval-censored failure time data with linear transformation models. *Canadian Journal of Statistics*, 33(1):61–70.

Zhou, Q., Sun, Y., and Gilbert, P. B. (2021). Semiparametric regression analysis of partly interval-censored failure time data with application to an AIDS clinical trial. *Statistics in Medicine*.

Zhu, L., Tong, X., Cai, D., Li, Y., Sun, R., Srivastava, D. K., and Hudson, M. M. (2021). Maximum likelihood estimation for the proportional odds model with mixed interval-censored failure time data. *Journal of applied statistics*, 48(8):1496–1512.

APPENDIX A: PROOF OF THEOREMS IN CHAPTER 3

To establish the asymptotic properties, we assume the following regularity conditions:

Condition 1. With probability one, $X(\cdot)$ and $Z(\cdot)$ have bounded total variation in $[0, \tau]$.

Condition 2. Let $\mathcal{B}$ be a compact set of $\mathcal{R}^d$ and $BV[0,\tau]$ be the class of functions with bound variation over $[0,\tau]$. The true parameter (β_0, A_0) belongs to $\mathcal{B} \times BV^q[0,\tau]$ with β_0 an interior point of $\mathcal{B}$ and $A_0(t) = (A_{01}(t), \ldots, A_{0q}(t))^\top$ is continuous over $[0,\tau]$ with $A_0(0) = 0$. Here $BV^q[0,\tau]$ denotes the product space $BV[0,\tau] \times \cdots \times BV[0,\tau]$.

Condition 3. With probability one, there exists a positive constant a such that $P(Y(\tau) = 1 \mid X(\cdot), Z(\cdot)) > a$ and $PN^2(\tau) < \infty$. If there exists a vector Γ and a deterministic function $\Gamma_0(t)$ such that $\Gamma_0(t) + \Gamma^\top X(t) = 0$ with probability one, then $\Gamma_0(t) = 0$ and $\Gamma = 0$ for any $t \in [0,\tau]$.

Condition 4. The transformation function $G(x)$ is thrice continuously differentiable on the interval $[0,\infty)$ and satisfy the following: $G(0) = 0$, $G'(x) > 0$ and $G(\infty) = \infty$.

Condition 5. Assume $\dot{\mathbf{\Psi}}_{\theta_0}$ in (A.4) is an invertible map.

A.1 Proof of Consistency

Theorem A.1. *Under Conditions* $1-5$*, the proposed ES estimator* $(\hat{\beta}, \hat{A})$ *is strongly consistent to* (β_0, A_0).

Proof. Let $\phi(t) = G'(t)$, $\psi(t) = G''(t)/G'(t)$ and

$$\rho(t;\beta,A) = \int_0^t Y(s) e^{\beta^\top Z(s)} X^\top(s) dA(s).$$

Hence, the posterior mean of ξ can be written as

$$g(\tau;\beta,A) = \phi(\rho(\tau;\beta,A)) - \Delta\psi(\rho(\tau;\beta,A)).$$

Let P denote the true probability measure and $\mathbb{P}_n$ denote the empirical measure. Let $\theta = (\beta, A)$ and $\theta_0 = (\beta_0, A_0)$. Then the proposed ES estimator $\hat{\theta} = (\hat{\beta}, \hat{A})$ is essentially a Z-estimator solving the following observed-data estimating equation

$$\mathbb{P}_n \Phi(\beta, A)(t) \equiv \mathbb{P}_n \begin{pmatrix} \Phi_1(\beta, A) \\ \Phi_2(\beta, A)(t) \end{pmatrix} = 0, \tag{A.1}$$

for $0 \le t \le \tau$, where

$$\Phi_1(\beta, A) = \int_0^\tau \left\{ Z(t)dN(t) - Y(t)e^{\beta^\top Z(t)} g(\tau; \beta, A) Z(t) X^\top(t) dA(t) \right\}, \tag{A.2}$$

and

$$\Phi_2(\beta, A)(t) = X(t)dN(t) - Y(t)e^{\beta^\top Z(t)} g(\tau; \beta, A) X(t) X^\top(t) dA(t).$$

Let h be a function in $BV_1[0, \tau]$, where $BV_1[0, \tau]$ denotes the set of functions with total variation bounded by 1 on $[0, \tau]$. Define

$$\Phi_2(\beta, A)[h] = \int_0^\tau h(t) \left\{ X(t)dN(t) - Y(t)e^{\beta^\top Z(t)} g(\tau; \beta, A) X(t) X^\top(t) dA(t) \right\}. \tag{A.3}$$

Similar to Gao et al. (2017) and van der Vaart and Wellner (1996b, Section 3.3.1), the proposed ES estimator $(\hat{\beta}, \hat{A})$ is equivalent to the root of the estimating equation

$$\mathbb{P}_n \Phi(\beta, A)[h] \equiv \mathbb{P}_n \begin{pmatrix} \Phi_1(\beta, A) \\ \Phi_2(\beta, A)[h] \end{pmatrix} = 0,$$

for all $h \in BV_1[0, \tau]$. From (A.1), $\hat{A}$ is a step function with jumps at the observed failure time points t_k $(k = 1, \ldots, m)$. Write $\widetilde{h}(t) = \sum_{k=1}^m h(t_k) I(t_{k-1} < t \le t_k)$. Then the step function $\widetilde{h}$ can be written as a finite sum of simple functions, denoted as

$\widetilde{h}(t) = \sum_{k=1}^{m} \alpha_k I(t_{k-1} < t \leq t_k)$, where $\alpha_k = h(t_k)$. It is easy to see that $(\hat{\beta}, \hat{A})$ solves

$$\mathbb{P}_n \Phi_2(\hat{\beta}, \hat{A})[\widetilde{h}] = \sum_{k=1}^{m} \alpha_k \mathbb{P}_n \Phi_2(\hat{\beta}, \hat{A})(t_k) = 0.$$

The parameter of interest is $\theta = (\beta, A)$, where $A = (A_1, \ldots, A_q)^\top$. Let $\overline{\text{lin}}(BV_1[0,\tau])$ be the closed linear span for linear functionals of $BV_1[0,\tau]$. For each j ($j = 1, \ldots, q$), A_j is contained in the Banach space $\overline{\text{lin}}(BV_1[0,\tau])$, where $A_j[h] = \int h(t) dA_j(t)$ for $h \in BV_1[0,\tau]$. The corresponding norm is defined as $\|A_j\|_\rho = \sup_{\|h\|_{BV} \leq 1} |\int h(t) dA_j(t)|$, where $\|\cdot\|_{BV}$ is the bounded variation norm. Thus, $A = (A_1, \ldots, A_q)^\top$ is contained in the Banach space $\overline{\text{lin}}^q(BV_1[0,\tau])$ and we define $A[h] = \int h(t) dA(t) = (\int h(t) dA_1(t), \ldots, \int h(t) dA_q(t))^\top$ for $h \in BV_1[0,\tau]$. Here, $\overline{\text{lin}}^q(BV_1[0,\tau])$ stands for the product space $\overline{\text{lin}}(BV_1[0,\tau]) \times \cdots \times \overline{\text{lin}}(BV_1[0,\tau])$. Furthermore, the norm for A is defined as $\|A\|_{\mathcal{H}} = \sum_{j=1}^{q} \|A_j\|_\rho$ and the norm for θ is defined as $\|\theta\|_{\mathcal{V}} = \|\beta\|_d + \|A\|_{\mathcal{H}}$, where $\|\cdot\|_d$ is the Euclidean norm in $\mathcal{R}^d$ space. Hence, the function $\mathbb{P}_n \Phi(\beta, A)[h]$ is a map from $\mathcal{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$ to $\mathcal{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$.

Define $B_\delta(\beta_0, A_0) = \{(\beta, A) : \|\beta - \beta_0\|_d + \|A - A_0\|_{\mathcal{H}} < \delta\}$. We first show that the class of functions $\{\Phi(\beta, A)[h] : (\beta, A) \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\}$ is P-Donsker for some fixed $\delta > 0$. Since $Y(t)$ and $N(t)$ are either cadlag or caglad functions in $l^\infty[0,\tau]$, they are both Donsker by Lemma 4.1 in Kosorok (2008). Trivially, Conditions 1 and 2 indicate that $\{\beta \in \mathcal{B}\}$, $\{Z(t), t \in [0,\tau]\}$ and $\{X(t), t \in [0,\tau]\}$ are all Donsker classes, and therefore so is $\{\beta^\top Z(t), \beta \in \mathcal{B}, t \in [0,\tau]\}$ since the products of bounded Donsker classes are Donsker. The class $\{e^{\beta^\top Z(t)}, \beta \in \mathcal{B}, t \in [0,\tau]\}$ is also Donsker since exponentiation is Lipschitz continuous on compacts. On the other hand, we rewrite

$$\rho(\tau; \beta, A) = \int_0^\tau Y(s) e^{\beta^\top Z(s)} X^\top(s) dA(s) = \sum_{j=1}^{q} \int_0^\tau Y(s) e^{\beta^\top Z(s)} X_j(s) dA_j(s).$$

Following Zeng et al. (2016), for any $j = 1, \ldots, q$, if A_j is a monotone function, the class $\{\int_0^\tau Y(s)e^{\beta^\top Z(s)}X_j(s)dA_j(s) : (\beta, A) \in B_\delta(\beta_0, A_0)\}$ is a Donsker class because it is a convex hull of functions $\{Y(s)\exp\{\beta^\top Z(s)\}X_j(s)\}$. By Condition 2, A_j ($j = 1, \ldots, q$) can be expressed as the difference of pairs of monotonely increasing functions since it has bounded total variation over $[0, \tau]$. Thus, $\{\int_0^\tau Y(s)e^{\beta^\top Z(s)}X_j(s)dA_j(s) : (\beta, A) \in B_\delta(\beta_0, A_0)\}$ is Donsker because the sums of bounded Donsker classes are also Donsker from Example 2.10.7 in van der Vaart and Wellner (1996b). It follows immediately that $\{\rho(\tau; \beta, A) : (\beta, A) \in B_\delta(\beta_0, A_0)\}$ is also a Donsker class. Similarly, the class $\left\{\int_0^\tau Y(t)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA(t) : (\beta, A) \in B_\delta(\beta_0, A_0)\right\}$ is a Donsker class. By Condition 4, $G(x)$ is thrice continuously differentiable on $[0, \infty)$ and for any $x \in [0, \infty)$, $G'(x) > 0$, then

$$g(\tau; \beta, A) = G'(\rho(\tau; \beta, A)) - \Delta G''(\rho(\tau; \beta, A))/G'(\rho(\tau; \beta, A)),$$

is bounded for $(\beta, A) \in B_\delta(\beta_0, A_0)$. Moreover, $\{g(\tau; \beta, A) : (\beta, A) \in B_\delta(\beta_0, A_0)\}$ is a Donsker class due to the fact that any continuously differentiable function is locally Lipschitz and the preservation of the Donsker property under Lipschitz-continuous transformations by Theorem 9.31 in Kosorok (2008). Note that $\{h(\cdot) : h \in BV_1[0, \tau]\}$, $\{\int_0^\tau h(t)X(t)dN(t) : h \in BV_1[0, \tau]\}$ and

$$\left\{\int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t) : \quad (\beta, A) \in B_\delta(\beta_0, A_0), h \in BV_1[0, \tau]\right\},$$

are all Donsker classes. This follows because the class of functions with an upper bound of their total variations is Donsker by Example 19.11 and Theorem 19.5 of van der Vaart (1998). Under Conditions $1 - 4$, now it is clear that the following classes

$$\left\{\int_0^\tau Z(t)dN(t)\right\}, \left\{\int_0^\tau h(t)X(t)dN(t) : h \in BV_1[0,\tau]\right\},$$
$$\left\{g(\tau;\beta,A)\int_0^\tau Y(t)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA(t) : \theta \in B_\delta(\beta_0, A_0)\right\},$$
$$\left\{g(\tau;\beta,A)\int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t) : \theta \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\right\},$$

are all Donsker classes. Therefore, the class of function

$$\left\{\Phi(\beta, A)[h] : (\beta, A) \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\right\},$$

is P-Donsker as the sums of bounded Donsker classes are also Donsker.

To prove the local consistency of $\hat{\theta} = (\hat{\beta}, \hat{A})$, we use Theorem 1.20 (the implicit function theorem) in Schwartz (1969). For any $\theta = (\beta, A)$ in $B_\delta(\beta_0, A_0)$, write $\mathbf{\Psi}(\theta) = P\Phi(\beta, A)[h]$ and $\mathbf{\Psi}_n(\theta) = \mathbb{P}_n\Phi(\beta, A)[h]$. Note that $\mathbf{\Psi}(\theta)$ and $\mathbf{\Psi}_n(\theta)$ are actually h-dependent. Rigorously speaking, we should write $\mathbf{\Psi}(\theta)[h] = P\Phi(\beta, A)[h]$ and $\mathbf{\Psi}_n(\theta)[h] = \mathbb{P}_n\Phi(\beta, A)[h]$, but in the rest of the article, we suppress the letter h in both $\mathbf{\Psi}(\theta)[h]$ and $\mathbf{\Psi}_n(\theta)[h]$ when there is no confusion. The Fréchet derivative of $\mathbf{\Psi}(\theta)$ with respect to θ at $\theta = \theta_0$ can be derived using (A.2) and (A.3). In particular, the Fréchet derivative $\dot{\mathbf{\Psi}}_{\theta_0}(\theta - \theta_0)$ can be easily computed based on the weaker form

$$\dot{\mathbf{\Psi}}_{\theta_0}(\theta - \theta_0) = \left.\frac{d\mathbf{\Psi}(\theta_0 + \eta(\theta - \theta_0))}{d\eta}\right|_{\eta=0} = \begin{pmatrix} C_{11}(\beta - \beta_0) + C_{12}(A - A_0) \\ C_{21}(\beta - \beta_0) + C_{22}(A - A_0) \end{pmatrix}, \quad \text{(A.4)}$$

where

$$\begin{aligned}
C_{11}(\beta-\beta_0) &= B_1(\beta-\beta_0),\\
C_{12}(A-A_0) &= \int_0^\tau B_2(t)d(A-A_0),\\
C_{21}(\beta-\beta_0)[h] &= B_3[h](\beta-\beta_0),\\
C_{22}(A-A_0)[h] &= \int_0^\tau B_4[h](t)d(A-A_0).
\end{aligned} \tag{A.5}$$

Specifically,

$$\begin{aligned}
B_1 =& -E\left[g(\tau,\beta_0,A_0)\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)Z^\top(t)X^\top(t)dA_0(t)\right]\\
&-E\Bigg[\{\phi'(\rho(\tau;\beta_0,A_0))-\Delta\psi'(\rho(\tau;\beta_0,A_0))\}\\
&\quad\times\left\{\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\right\}^{\otimes 2}\Bigg],\\
B_2(t) =& -E\Bigg[\{\phi'(\rho(\tau;\beta_0,A_0))-\Delta\psi'(\rho(\tau;\beta_0,A_0))\}\\
&\quad\times\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\\
&\quad\times Y(t)e^{\beta_0^\top Z(t)}X^\top(t)\Bigg]\\
&-E\left[g(\tau;\beta_0,A_0)Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)\right],\\
B_3[h] =& -E\Bigg[\{\phi'(\rho(\tau;\beta_0,A_0))-\Delta\psi'(\rho(\tau;\beta_0,A_0))\}\\
&\quad\times\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}X(t)X^\top(t)h(t)dA_0(t)\\
&\quad\times\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z^\top(t)X^\top(t)dA_0(t)\Bigg]\\
&-E\left[g(\tau;\beta_0,A_0)\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}X(t)Z^\top(t)X^\top(t)h(t)dA_0(t)\right],\\
B_4[h](t) =& -E\Bigg[\{\phi'(\rho(\tau;\beta_0,A_0))-\Delta\psi'(\rho(\tau;\beta_0,A_0))\}\\
&\quad\times\left\{\int_0^\tau h(t)Y(t)e^{\beta_0^\top Z(t)}X(t)X^\top(t)dA_0(t)\right\}Y(t)e^{\beta_0^\top Z(t)}X^\top(t)\Bigg]\\
&-E\left\{h(t)Y(t)e^{\beta_0^\top Z(t)}g(\tau;\beta_0,A_0)X(t)X^\top(t)\right\}.
\end{aligned}$$

Here, $a^{\bigotimes 2} = aa^\top$ for any column vector a. It can be shown that $||\mathbf{\Psi}(\theta) - \mathbf{\Psi}(\theta_0) - \dot{\mathbf{\Psi}}_{\theta_0}(\theta - \theta_0)|| = o(||\theta - \theta_0||)$ as $\theta \to \theta_0$. Hence, $\mathbf{\Psi}(\theta)$ is Fréchet-differentiable at θ_0. The detailed calculations of the derivatives are given in Appendix A.3. Here we just present the corresponding results. Similarly, the Fréchet derivative of $\mathbf{\Psi}_n(\theta) = \mathbb{P}_n\Phi(\beta, A)[h]$ with respect to θ at $\theta = \theta_0$ can be derived and we use $\dot{\mathbf{\Psi}}_{\theta_0,n}$ to denote the corresponding derivative map. In particular, $\dot{\mathbf{\Psi}}_{\theta_0,n}$ can be obtained by replacing $\mathbf{\Psi}$ with $\mathbf{\Psi}_n$ in (A.4) and the expectations E in the terms B_1, $B_2(t)$, $B_3[h]$ and $B_4[h](t)$ with the empirical measure $\mathbb{P}_n$. Then one can obtain that $\|\mathbf{\Psi}_n(\theta) - \mathbf{\Psi}_n(\theta_0) - \dot{\mathbf{\Psi}}_{\theta_0,n}(\theta - \theta_0)\| = o(\|\theta - \theta_0\|)$ as $\theta \to \theta_0$. Hence, $\mathbf{\Psi}_n(\theta)$ is also Fréchet-differentiable at θ_0. Clearly, the maps $\dot{\mathbf{\Psi}}_\theta$ and $\dot{\mathbf{\Psi}}_{\theta,n}$ depend continuously on θ in $B_\delta(\beta_0, A_0)$.

Next, we show that $\dot{\mathbf{\Psi}}_{\theta_0,n}$ is invertible for larger enough n. Following the previous Donsker theory arguments, it can be shown that $\dot{\mathbf{\Psi}}_\theta(\theta^*)[h] - \dot{\mathbf{\Psi}}_{\theta,n}(\theta^*)[h] = o_p(1)$ uniformly in (θ, θ^*, h) in $B_\delta(\beta_0, A_0) \times \mathcal{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau]) \times BV_1[0,\tau]$ for some $\delta > 0$. By Condition 5, we know that $\dot{\mathbf{\Psi}}_{\theta_0}$ is invertible. Thus, there exists a constant $c_1 > 0$ such that $\|\dot{\mathbf{\Psi}}_{\theta_0}(\theta - \theta_0)\| \geq c_1\|\theta - \theta_0\|$ for any θ in $\mathcal{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$ by Lemma 6.16 in Kosorok (2008). Notice that there exists a positive constant c_2 such that

$$\left\| \frac{\dot{\mathbf{\Psi}}_{\theta_0,n}(\theta - \theta_0)}{\|\theta - \theta_0\|} \right\| = \left\| \dot{\mathbf{\Psi}}_{\theta_0,n}\left(\frac{\theta - \theta_0}{\|\theta - \theta_0\|}\right) \right\| = \left\| \dot{\mathbf{\Psi}}_{\theta_0}\left(\frac{\theta - \theta_0}{\|\theta - \theta_0\|}\right) + o_p(1) \right\| \geq c_1 + o_p(1) \geq c_2,$$

as $n \to \infty$ for any θ in $\mathcal{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$. Thus, $\|\dot{\mathbf{\Psi}}_{\theta_0,n}(\theta - \theta_0)\| \geq c_2\|\theta - \theta_0\|$ as $n \to \infty$ for any θ in $\mathcal{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$. Hence, $\dot{\mathbf{\Psi}}_{\theta_0,n}$ is invertible for larger enough n. In brief, we verified the three conditions, i.e., $\mathbf{\Psi}_n(\theta)$ is Fréchet-differentiable at θ_0, $\dot{\mathbf{\Psi}}_{\theta,n}$ depends continuously on θ in $B_\delta(\beta_0, A_0)$ and $\dot{\mathbf{\Psi}}_{\theta_0,n}$ is invertible for larger enough n. The implicit function theorem yields that for a sufficiently small $\delta > 0$, the map $\mathbf{\Psi}_n(\theta)$ is one-to-one from $B_\delta(\beta_0, A_0)$ onto a neighborhood of zero for large n.

Finally, we notice that $\{\Phi(\beta_0, A_0)[h] : h \in BV_1[0,\tau]\}$ is a Donsker class because

$$\{\Phi(\beta_0, A_0)[h] : h \in BV_1[0,\tau]\} \subset \{\Phi(\beta, A)[h] : (\beta, A) \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\}.$$

Hence, $\mathbb{P}_n\Phi(\beta_0, A_0)[h] - P\Phi(\beta_0, A_0)[h] = o_p(1)$, i.e., $\mathbf{\Psi}_n(\theta_0) - \mathbf{\Psi}(\theta_0) = o_p(1)$. The martingale properties and double expectations yield

$$\mathbf{\Psi}(\theta_0) = P\Phi(\beta_0, A_0)[h] = P\begin{pmatrix} \Phi_1(\beta_0, A_0) \\ \Phi_2(\beta_0, A_0)[h] \end{pmatrix} = 0.$$

Therefore, $\mathbf{\Psi}_n(\theta_0) = o_p(1)$. For an arbitrary small $\delta > 0$ and large n, by the implicit function theorem (Schwartz, 1969), there exists $\hat{\theta} = (\hat{\beta}, \hat{A})$ with $(\|\hat{\beta} - \beta_0\|_d + \|\hat{A} - A_0\|_{\mathcal{H}}) < \delta$ and $\mathbf{\Psi}_n(\hat{\theta}) = \mathbb{P}_n\Phi(\hat{\beta}, \hat{A})[h] = 0$ for any $h \in BV_1[0,\tau]$. This proves the consistency of $\hat{\theta} = (\hat{\beta}, \hat{A})$. □

A.2 Proof of Asymptotic Normality

Theorem A.2. *Under the conditions* $1-5$, $\sqrt{n}(\hat{\beta} - \beta_0, \hat{A} - A_0)$ *converges weakly to a zero-mean Gaussian process in the metric space* $\mathcal{R}^d \times \overline{lin}^q(BV_1[0,\tau])$.

Proof. We appeal to verify the conditions in Theorem 3.3.1 and Lemma 3.3.5 of van der Vaart and Wellner (1996b). Write

$$\begin{aligned} \mathbb{G}_n\Phi(\theta)[h] &= n^{1/2}\left\{\mathbb{P}_n\Phi(\theta)[h] - P\Phi(\theta)[h]\right\} \\ &= n^{1/2}\left\{\mathbf{\Psi}_n(\theta)[h] - \mathbf{\Psi}(\theta)[h]\right\}, \end{aligned}$$

where $\mathbf{\Psi}_n(\theta)[h] = \mathbb{P}_n\Phi(\theta)[h]$ and $\mathbf{\Psi}(\theta)[h] = P\Phi(\theta)[h]$. We prove the asymptotic normality of the proposed ES estimator by the following four steps:

(1) Show that $\mathbb{G}_n\Phi(\theta_0)[h] = n^{1/2}\left\{\mathbf{\Psi}_n(\theta_0)[h] - \mathbf{\Psi}(\theta_0)[h]\right\}$ converges in distribution to a tight random element $\mathcal{W}$ in $\mathcal{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$.

Under Conditions $1-4$, we have

$$\sup_{h\in BV_1[0,\tau]} \|\mathbf{\Psi}(\theta_0)[h]\| < \infty.$$

Because $\{\Phi(\theta_0)[h] : h \in BV_1[0,\tau]\}$ is a Donsker class, $\mathbb{G}_n\Phi(\theta_0)[h]$ converges weakly to a Gaussian process $\mathcal{W}$ in $\mathcal{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$.

(2) Verify $\mathbf{\Psi}(\theta)$ is Fréchet differentiable as a function of θ at $\theta=\theta_0$.

The Fréchet-differentiablility of $\mathbf{\Psi}(\theta)$ can be checked directly. In particular, the Fréchet derivative $\dot{\mathbf{\Psi}}_{\theta_0}(\theta-\theta_0)$ can be easily computed based on the weaker form

$$\dot{\mathbf{\Psi}}_{\theta_0}(\theta-\theta_0) = \frac{d\mathbf{\Psi}(\theta_0+\eta(\theta-\theta_0))}{d\eta}\bigg|_{\eta=0} = \begin{pmatrix} C_{11}(\beta-\beta_0)+C_{12}(A-A_0) \\ C_{21}(\beta-\beta_0)+C_{22}(A-A_0) \end{pmatrix},$$

where each of the components is given in (A.5). The detailed calculations are shown in Appendix B.3.

(3) To verify the condition (3.3.4) in van der Vaart and Wellner (1996b), it's sufficient to verify the conditions in Lemma 3.3.5 of van der Vaart and Wellner (1996b).

From Conditions $1-4$ and the results derived in the proof of Theorem 1, we have shown that $\{\Phi(\theta)[h] : \theta \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\}$ and $\{\Phi(\theta_0)[h] : h \in BV_1[0,\tau]\}$ both are Donsker classes. Thus,

$$\{\Phi(\theta)[h] - \Phi(\theta_0)[h] : \theta \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\}$$

is also a Donsker class for some $\delta > 0$. In view of the dominated convergence theorem, to show

$$\sup_{h\in BV_1[0,\tau]} P\{\Phi(\theta)[h] - \Phi(\theta_0)[h]\}^2 \to 0,$$

as $\theta \to \theta_0$, it is valid to show that $\Phi(\theta)[h]$ converges to $\Phi(\theta_0)[h]$ pointwise, uniformly in

h. This condition is satisfied because $h(t)$ has bounded total variation over $[0, \tau]$ and $\Phi(\theta)[h]$ is a continuous function over θ under Conditions $1-4$. Because $\hat{\theta}$ converges to θ_0 almost surely, it follows from Lemma 3.3.5 of van der Vaart and Wellner (1996b) that

$$\|\mathbb{G}_n(\Phi(\hat{\theta}) - \Phi(\theta_0))\| = o_{p^*}(1 + n^{1/2}\|\hat{\theta} - \theta_0\|), \tag{A.6}$$

where $o_{p^*}(1)$ denotes converging to zero in outer probability.

(4) Equation (A.6) can be written as

$$n^{1/2}(\mathbf{\Psi}_n - \mathbf{\Psi})(\hat{\theta}) - n^{1/2}(\mathbf{\Psi}_n - \mathbf{\Psi})(\theta_0) = o_{p^*}(1 + n^{1/2}\|\hat{\theta} - \theta_0\|).$$

By the definition of θ_0 and $\hat{\theta}$, $\mathbf{\Psi}(\theta_0) = 0$ and $\mathbf{\Psi}_n(\hat{\theta}) = 0$. It follows from Theorem 3.3.1 of van der Vaart and Wellner (1996b) that

$$n^{1/2}\dot{\mathbf{\Psi}}_{\theta_0}(\hat{\theta} - \theta_0) = -n^{1/2}(\mathbf{\Psi}_n - \mathbf{\Psi})(\theta_0) + o_{p^*}(1).$$

Finally, Condition 5 and the continuous mapping theorem give

$$n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow -\dot{\mathbf{\Psi}}_{\theta_0}^{-1}\mathcal{W}.$$

□

A.3 The Fréchet Derivative Map

This subsection provides details on the calculation of the Fréchet derivative of $\mathbf{\Psi}(\theta) = P\Phi(\beta, A)$. The Fréchet derivative of $P\Phi(\beta, A)$ at (β_0, A_0) is given by the map

$$(\beta - \beta_0, A - A_0) \to \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} \beta - \beta_0 \\ A - A_0 \end{pmatrix},$$

where

$$C_{11}(\beta - \beta_0) = B_1(\beta - \beta_0),$$
$$C_{12}(A - A_0) = \int_0^\tau B_2(t) d(A - A_0),$$
$$C_{21}(\beta - \beta_0)[h] = B_3[h](\beta - \beta_0),$$
$$C_{22}(A - A_0)[h] = \int_0^\tau B_4[h](t) d(A - A_0).$$

Specifically,

$$\begin{aligned} B_1 &= \left.\frac{\partial P\Phi_1(\beta, A)}{\partial \beta}\right|_{\beta=\beta_0, A=A_0} \\ &= -E\left[g(\tau, \beta_0, A_0) \int_0^\tau Y(t) e^{\beta_0^\top Z(t)} Z(t) Z^\top(t) X^\top(t) dA_0(t)\right] \\ &\quad - E\Bigg[\{\phi'(\rho(\tau; \beta_0, A_0)) - \Delta\psi'(\rho(\tau; \beta_0, A_0))\} \\ &\quad \times \left\{\int_0^\tau Y(t) e^{\beta_0^\top Z(t)} Z(t) X^\top(t) dA_0(t)\right\}^{\otimes 2} \Bigg]. \end{aligned}$$

Note that

$$\begin{aligned} &\left.\frac{\partial g(\tau; \beta, A + \eta A^*)}{\partial \eta}\right|_{\eta=0} \\ &= \left.\{\phi'(\rho(\tau; \beta, A + \eta A^*)) - \Delta\psi'(\rho(\tau; \beta, A + \eta A^*))\}\, \rho(\tau; \beta, A^*)\right|_{\eta=0} \\ &= \{\phi'(\rho(\tau; \beta, A)) - \Delta\psi'(\rho(\tau; \beta, A))\}\, \rho(\tau; \beta, A^*), \end{aligned}$$

where $\phi'(\cdot)$ and $\psi'(\cdot)$ are the first derivative of $\phi(\cdot)$ and $\psi(\cdot)$, respectively. Then,

$$
\begin{aligned}
&\left.\frac{\partial \Phi_1(\beta, A+\eta A^*)}{\partial \eta}\right|_{\eta=0} \\
&= -\int_0^\tau Y(t)e^{\beta^\top Z(t)} \frac{\partial g(\tau;\beta, A+\eta A^*)}{\partial \eta} Z(t)X^\top(t) d(A+\eta A^*)(t)\Big|_{\eta=0} \\
&\quad -\int_0^\tau Y(t)e^{\beta^\top Z(t)} g(\tau;\beta, A+\eta A^*) Z(t)X^\top(t) dA^*(t)\Big|_{\eta=0} \\
&= -\{\phi'(\rho(\tau;\beta,A)) - \Delta\psi'(\rho(\tau;\beta,A))\} \int_0^\tau Y(t)e^{\beta^\top Z(t)} Z(t)X^\top(t) dA(t) \\
&\quad \times \int_0^\tau Y(t)e^{\beta^\top Z(t)} X^\top(t) dA^*(t) \\
&\quad - g(\tau;\beta,A) \int_0^\tau Y(t)e^{\beta^\top Z(t)} Z(t)X^\top(t) dA^*(t).
\end{aligned}
$$

Thus, we can obtain

$$
\begin{aligned}
&\left.\frac{\partial P\Phi_1(\beta, A+\eta A^*)}{\partial \eta}\right|_{\eta=0, A^*=A-A_0, \beta=\beta_0, A=A_0} \\
&= -E\left[\{\phi'(\rho(\tau;\beta_0,A_0)) - \Delta\psi'(\rho(\tau;\beta_0,A_0))\} \int_0^\tau Y(t)e^{\beta_0^\top Z(t)} Z(t)X^\top(t) dA_0(t)\right. \\
&\quad \left.\times \int_0^\tau Y(t)e^{\beta_0^\top Z(t)} X^\top(t) d(A-A_0)(t)\right] \\
&\quad - E\left[g(\tau;\beta_0,A_0) \int_0^\tau Y(t)e^{\beta_0^\top Z(t)} Z(t)X^\top(t) d(A-A_0)(t)\right].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
B_2(t) = &-E\Big[\{\phi'(\rho(\tau;\beta_0,A_0)) - \Delta\psi'(\rho(\tau;\beta_0,A_0))\} \\
&\times \int_0^\tau Y(t)e^{\beta_0^\top Z(t)} Z(t)X^\top(t) dA_0(t) \\
&\times Y(t)e^{\beta_0^\top Z(t)} X^\top(t)\Big] \\
&- E\left\{g(\tau;\beta_0,A_0) Y(t)e^{\beta_0^\top Z(t)} Z(t)X^\top(t)\right\}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
B_3[h] &= \frac{\partial P\Phi_2(\beta, A)[h]}{\partial \beta}\bigg|_{\beta=\beta_0, A=A_0} \\
&= -E\bigg[\{\phi'(\rho(\tau;\beta_0, A_0)) - \Delta\psi'(\rho(\tau;\beta_0, A_0))\} \\
&\quad \times \int_0^\tau h(t)Y(t)e^{\beta_0^\top Z(t)} X(t)X^\top(t)dA_0(t) \\
&\quad \times \int_0^\tau Y(t)e^{\beta_0^\top Z(t)} Z^\top(t)X^\top(t)dA_0(t)\bigg] \\
&\quad - E\left[g(\tau;\beta_0, A_0)\int_0^\tau h(t)Y(t)e^{\beta_0^\top Z(t)}X(t)Z^\top(t)X^\top(t)dA_0(t)\right].
\end{aligned}
$$

Lastly,

$$
\begin{aligned}
&\frac{\partial \Phi_2(\beta, A+\eta A^*)[h]}{\partial \eta}\bigg|_{\eta=0} \\
&= -\int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}\frac{\partial g(\tau;\beta, A+\eta A^*)}{\partial \eta}X(t)X^\top(t)d(A+\eta A^*)(t)\bigg|_{\eta=0} \\
&\quad - \int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}g(\tau;\beta, A+\eta A^*)X(t)X^\top(t)dA^*(t)\bigg|_{\eta=0} \\
&= -\{\phi'(\rho(\tau;\beta, A)) - \Delta\psi'(\rho(\tau;\beta, A))\}\int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t) \\
&\quad \times \int_0^\tau Y(t)e^{\beta^\top Z(t)}X^\top(t)dA^*(t) \\
&\quad - g(\tau;\beta, A)\int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}X(t)X^\top(t)dA^*(t).
\end{aligned}
$$

Hence, $\frac{\partial P\Phi_2(\beta, A+\eta A^*)[h]}{\partial \eta}\Big|_{\eta=0, A^*=A-A_0, \beta=\beta_0, A=A_0} = \int_0^\tau B_4[h](t)d(A-A_0)(t)$, where

$$
\begin{aligned}
B_4[h](t) = &- E\bigg[\{\phi'(\rho(\tau;\beta_0, A_0)) - \Delta\psi'(\rho(\tau;\beta_0, A_0))\} \\
&\times \left\{\int_0^\tau h(t)Y(t)e^{\beta_0^\top Z(t)}X(t)X^\top(t)dA_0(t)\right\}Y(t)e^{\beta_0^\top Z(t)}X^\top(t)\bigg] \\
&- E\left\{h(t)Y(t)e^{\beta_0^\top Z(t)}g(\tau;\beta_0, A_0)X(t)X^\top(t)\right\}.
\end{aligned}
$$

APPENDIX B: PROOF OF THEOREMS IN CHAPTER 4

We establish the asymptotic properties of the proposed estimators under the following regularity conditions:

Condition 1 With probability one, the vectors $X(t)$ and $Z(t)$ are uniformly bounded with uniformly bounded total variation over $[0,\tau]$. In addition, $X(t)$ is not identically zero for all $t \in [0,\tau]$. Here, τ denotes the end of the study.

Condition 2 Let $\mathcal{B}$ be a compact set of $\mathcal{R}^d$ and $BV[0,\tau]$ be the class of functions with bound variation over $[0,\tau]$. The true parameter (β_0, A_0) belongs to $\mathcal{B} \times BV^q[0,\tau]$ with β_0 an interior point of $\mathcal{B}$ and $A_0(t) = (A_{01}(t), \ldots, A_{0q}(t))^\top$ is continuously differentiable over $[0,\tau]$ with $A_0(0) = 0$. Here, $BV^q[0,\tau]$ denotes the product space $BV[0,\tau] \times \cdots \times BV[0,\tau]$.

Condition 3 $0 < P(\Delta = 0) < 1$. For $\Delta = 0$, the number of monitoring times, K, is positive, and $E(K) < \infty$. In addition, there exists some constant $c > 0$ such that $P(U_{j+1} - U_j \geq c | K, X, Z) = 1$ $(j = 1, \ldots, K-1)$.

Condition 4 The transformation function G is thrice continuously differentiable on $[0,\infty)$ with $G(0) = 0$, $G'(x) > 0$ and $G(\infty) = \infty$.

Condition 5 If there exists a vector η and a deterministic function $\eta_0(t)$ such that $\eta_0(t) + \eta^\top X(t) = 0$ with probability one, then $\eta = 0$ for all $t \in [0,\tau]$ and $\eta_0(t) = 0$ for all $t \in [0,\tau]$.

B.1 Proof of Consistency

Theorem B.1. *Under Conditions* $1-5$*, the proposed ES estimator* $(\hat{\beta}, \hat{A})$ *is strongly consistent to* (β_0, A_0).

Proof. Let $\rho_0(t;\theta) = \int_0^t e^{\beta^\top Z(s)} X^\top(s) dA(s)$ and $\rho_1(t;\theta) = \int_0^t e^{\beta^\top Z(s)} Z(s) X^\top(s) dA(s)$. In addition, let $g_1(\cdot) = G'(\cdot) - G''(\cdot)/G'(\cdot)$. Note that the posterior mean of the latent

Poisson random variable W at any $t \in [0, \tau]$ can be written as

$$\Delta I(T = t) + (1 - \Delta)I(L < t \leq R)I(R < \infty)H_1(L, R; \theta) \exp\{\beta^\top Z(t)\}X^\top(t)dA(t),$$

where

$$H_1(L, R; \theta) = \frac{\exp[-G\{\rho_0(L; \theta)\}]G'\{\rho_0(L; \theta)\}}{\exp[-G\{\rho_0(L; \theta)\}] - \exp[-G\{\rho_0(R; \theta)\}]I(R < \infty)},$$

and the posterior mean of ξ can be written as

$$\hat{E}(\xi) = \Delta g_1\{\rho_0(T; \theta)\} + (1 - \Delta)H_2(L, R; \theta),$$

where

$$\begin{aligned} H_2(L, R; \theta) = &\frac{\exp[-G\{\rho_0(L; \theta)\}]G'\{\rho_0(L; \theta)\}}{\exp[-G\{\rho_0(L; \theta)\}] - \exp[-G\{\rho_0(R; \theta)\}]I(R < \infty)} \\ &- \frac{- \exp[-G\{\rho_0(R; \theta)\}]G'\{\rho_0(R; \theta)\}I(R < \infty)}{\exp[-G\{\rho_0(L; \theta)\}] - \exp[-G\{\rho_0(R; \theta)\}]I(R < \infty)}. \end{aligned}$$

Let P and $\mathbb{P}_n$ denote the true probability measure and empirical measure, respectively. The proposed ES estimator $\hat{\theta} = (\hat{\beta}, \hat{A})$ is essentially a Z-estimator solving the following observed-data estimating equation

$$\mathbb{P}_n \begin{pmatrix} \Phi_1(\theta) \\ \Phi_2(\theta)(t) \end{pmatrix} = 0 \quad \text{for all } t \in [0, \tau], \tag{B.1}$$

where

$$\begin{aligned} \Phi_1(\theta) = &\Delta Z(T) - \Delta g_1\{\rho_0(T; \theta)\} \int_0^\tau I(t \leq R^*)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA(t) \\ &+ (1 - \Delta)H_1(L, R; \theta) \int_0^\tau I(L < t \leq R < \infty)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA(t) \\ &- (1 - \Delta)H_2(L, R; \theta) \int_0^\tau I(t \leq R^*)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA(t), \end{aligned} \tag{B.2}$$

and

$$\begin{aligned}\Phi_2(\theta)(t) = {} & \Delta X(t)I(t=T) - \Delta g_1\{\rho_0(T;\theta)\}I(t \le R^*)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t) \\ & + (1-\Delta)H_1(L,R;\theta)I(L < t \le R < \infty)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t) \\ & - (1-\Delta)H_2(L,R;\theta)I(t \le R^*)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t).\end{aligned} \tag{B.3}$$

There are infinite number of estimating equations in (B.1). To resolve this, we consider

$$\begin{aligned}\Phi_1(\theta)[h_0] = {} & \Delta h_0^\top Z(T) - \Delta g_1\{\rho_0(T;\theta)\}\int_0^\tau I(t \le R^*)e^{\beta^\top Z(t)}h_0^\top Z(t)X^\top(t)dA(t) \\ & + (1-\Delta)H_1(L,R;\theta)\int_0^\tau I(L < t \le R < \infty)e^{\beta^\top Z(t)}h_0^\top Z(t)X^\top(t)dA(t) \\ & - (1-\Delta)H_2(L,R;\theta)\int_0^\tau I(t \le R^*)e^{\beta^\top Z(t)}h_0^\top Z(t)X^\top(t)dA(t),\end{aligned} \tag{B.4}$$

where $h_0 \in \mathcal{R}^d$, and

$$\begin{aligned}\widetilde{\Phi}_2(\theta)[h] = {} & \Delta h(T) \circ X(T) \\ & - \Delta g_1\{\rho_0(T;\theta)\}\int_0^\tau I(t \le R^*)e^{\beta^\top Z(t)}h(t) \circ X(t)X^\top(t)dA(t) \\ & + (1-\Delta)H_1(L,R;\theta)\int_0^\tau I(L < t \le R < \infty)e^{\beta^\top Z(t)}h(t) \circ X(t)X^\top(t)dA(t) \\ & - (1-\Delta)H_2(L,R;\theta)\int_0^\tau I(t \le R^*)e^{\beta^\top Z(t)}h(t) \circ X(t)X^\top(t)dA(t).\end{aligned} \tag{B.5}$$

Here, $h = (h_1, \ldots, h_q)^\top$ is a vector of q functions, where each h_j ($j = 1, \ldots, q$) belongs to the space of functions of bounded variation with a bound equal to 1, denoted as $BV_1[0,\tau]$. The notation $a \circ b$ represents the component-wise product of two vectors a and b of the same size. Hence, $h \in BV_1^q[0,\tau]$, where $BV_1^q[0,\tau]$ stands for the product space $BV_1[0,\tau] \times \cdots \times BV_1[0,\tau]$. It is easy to see that $\mathbb{P}_n\widetilde{\Phi}_2(\theta)[h] = 0$ for every

$h \in BV_1^q[0,\tau]$ is equivalent to

$$\Phi_2(\theta)[h] = 0 \quad \text{for every } h \in BV_1^q[0,\tau],$$

where

$$\begin{aligned}\Phi_2(\theta)[h] = \Delta \sum_{j=1}^{q} \Bigg[& h_j(T)X_j(T) \\ & - \Delta g_1\{\rho_0(T;\theta)\} \int_0^\tau I(t \le R^*) e^{\beta^\top Z(t)} h_j(t) X_j t) X^\top(t) dA(t) \\ & + (1-\Delta) H_1(L,R;\theta) \int_0^\tau I(L < t \le R < \infty) e^{\beta^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA(t) \\ & - (1-\Delta) H_2(L,R;\theta) \int_0^\tau I(t \le R^*) e^{\beta^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA(t) \Bigg].\end{aligned} \tag{B.6}$$

Therefore, the equations $\mathbb{P}_n\Phi_1(\theta) = 0$ and $\mathbb{P}_n\Phi_2(\theta)(t) = 0$ for every $t \in [0,\tau]$ in (B.1) are equivalent to equations $\mathbb{P}_n\Phi_1(\theta)[h_0] = 0$ for every $h_0 \in \mathcal{R}^d$ and $\mathbb{P}_n\Phi_2(\theta)[h] = 0$ for every $h \in BV_1^q[0,\tau]$, respectively. Several research studies have used similar techniques to develop and analyze Z-estimators, including the works of van der Vaart and Wellner (1996b, Section 3.3.1), Gao et al. (2017) and among others. Thus, the proposed ES estimator $\hat{\theta}$ is equivalent to the solution of the estimating equation

$$\mathbb{P}_n\Phi(\theta)[\widetilde{h}] \equiv \mathbb{P}_n\Phi_1(\theta)[h_0] + \mathbb{P}_n\Phi_2(\theta)[h], \quad \text{for every } \widetilde{h} = (h_0, h) \in \mathcal{R}^d \times BV_1^q[0,\tau]. \tag{B.7}$$

Now we introduce the parameter space that we consider. The parameter of interest is $\theta = (\beta, A)$, where $\beta \in \mathcal{R}^d$ and $A = (A_1, \ldots, A_q)$ consists of q infinite-dimensional cumulative regression functions. It is easy to see that $A = (A_1, \ldots, A_q)$ is in the Banach space $BV^q[0,\tau]$. The norm for A is defined as the summation of the norm of each component, i.e., $\|A\|_\rho = \sum_{j=1}^q \|A_j\|_v$, where $\|A_j\|_v$ is the sum of absolute value of $A_j(0)$ and the total variation of A_j on $[0,\tau]$. Finally, the norm for θ is defined as $\|\theta\|_\Theta = \|\beta\|_d + \|A\|_\rho$, where $\|\cdot\|_d$ is the Euclidean norm: $\|\beta\|_d = \sqrt{\sum_{j=1}^d \beta_j^2}$. Similarly, we define the norm $\|\widetilde{h}\|_{\mathcal{H}} = \|h_0\|_d + \sum_{j=1}^q \|h_j\|_v$, where $\|h_j\|_v$ is the sum of absolute value of $h_j(0)$ and the total variation of h_j on $[0,\tau]$. Let H be a subset of $\mathcal{R}^d \times BV_1^q[0,\tau]$ with $\|\widetilde{h}\|_{\mathcal{H}} \leq M < \infty$.

Let $\widetilde{h} = (h_0, h) \in H$. It is easy to note that the function $\mathbb{P}_n\Phi(\theta)[\widetilde{h}]$ is a map from $\mathcal{R}^d \times BV^q[0,\tau]$ to $l^\infty(H)$. Let $B_\delta(\theta_0) = \{\theta = (\beta, A) : \|\beta - \beta_0\|_d + \|A - A_0\|_\rho < \delta\}$, where $\delta > 0$. For any θ in $B_\delta(\theta_0)$, we write $\mathbf{\Psi}(\theta)[\widetilde{h}] = P\Phi(\theta)[\widetilde{h}]$ and $\mathbf{\Psi}_n(\theta)[\widetilde{h}] = \mathbb{P}_n\Phi(\theta)[\widetilde{h}]$. Note that $\mathbf{\Psi}(\theta)[\widetilde{h}]$ and $\mathbf{\Psi}_n(\theta)[\widetilde{h}]$ depend on $\widetilde{h}$. To suppress notations, we write $\mathbf{\Psi}(\theta)$ and $\mathbf{\Psi}_n(\theta)$ for $\mathbf{\Psi}(\theta)[\widetilde{h}]$ and $\mathbf{\Psi}_n(\theta)[\widetilde{h}]$, respectively, when there is no confusion. We prove the local consistency of $\hat{\theta} = (\hat{\beta}, \hat{A})$ by verifying the three conditions in Theorem 1.20 (the implicit function theorem) (Schwartz, 1969). These conditions are (1)$\mathbf{\Psi}_n(\theta)[\widetilde{h}]$ is Fréchet-differentiable in $B_\delta(\theta_0)$ with some $\delta > 0$; (2) the corresponding Fréchet derivative map depends continuously on θ in $B_\delta(\theta_0)$; (3) this map evaluated at θ_0 is a bounded linear map with a bounded linear inverse. We verified the first two conditions in Lemma 2 and the third condition in Lemma 4 in Appendix B.4.

By Lemma 1, the class $\{\Phi(\theta)[\widetilde{h}] : \theta \in B_\delta(\theta_0), \widetilde{h} \in H\}$ is a Donsker class for some $\delta > 0$. The class $\{\Phi(\theta_0)[\widetilde{h}] : \widetilde{h} \in H\}$ is also Donsker via Theorem 2.10.1 in van der Vaart and Wellner (1996b) because the latter class is a subset of the former class. By Donsker properties,

$$\mathbb{P}_n\Phi(\theta_0)[\widetilde{h}] - P\Phi(\theta_0)[\widetilde{h}] = o_p(1),$$

or equivalently, $\mathbf{\Psi}_n(\theta_0) - \mathbf{\Psi}(\theta_0) = o_p(1)$. In addition, $\mathbf{\Psi}(\theta_0) = P\Phi(\theta_0)[\widetilde{h}] = 0$ can be easily checked by double expectation properties. Therefore, $\mathbf{\Psi}_n(\theta_0) = o_p(1)$. By Lemma 2 and 4, we verified the three conditions of the implicit function theorem (Schwartz, 1969), and hence it yields that $\mathbf{\Psi}_n(\theta)$ is a one-to-one map from $B_\delta(\theta_0)$ onto a neighborhood of zero for large n and sufficiently small $\delta > 0$. As a result, for an arbitrary small $\delta > 0$ and large n, there exists $\hat{\theta} = (\hat{\beta}, \hat{A})$ with $(\|\hat{\beta} - \beta_0\|_d + \|\hat{A} -$

$A_0\|_\rho) < \delta$ and $\mathbf{\Psi}_n(\hat{\theta}) = \mathbb{P}_n\Phi(\hat{\theta})[\widetilde{h}] = 0$ for any $\widetilde{h} \in H$. This proves the consistency of $\hat{\theta} = (\hat{\beta}, \hat{A})$.

□

B.2 Proof of Asymptotic Normality

Theorem B.2. *Under Conditions* $1-5$, $\sqrt{n}(\hat{\beta} - \beta_0, \hat{A}(t) - A_0(t))$ *converges weakly to a zero-mean Gaussian process* $t \in [0, \tau]$.

Proof. We establish the asymptotic normality of $\hat{\theta} = (\hat{\beta}, \hat{A})$ by applying Theorem 3.3.1 and Lemma 3.3.5 of van der Vaart and Wellner (1996b) because the proposed ES estimator $\hat{\theta}$ is essentially a Z-estimator.

Let $\mathbb{G}_n\Phi(\theta)[\widetilde{h}] = n^{1/2}\left\{\mathbf{\Psi}_n(\theta)[\widetilde{h}] - \mathbf{\Psi}(\theta)[\widetilde{h}]\right\}$, where $\mathbf{\Psi}_n(\theta)[\widetilde{h}] = \mathbb{P}_n\Phi(\theta)[\widetilde{h}]$ and $\mathbf{\Psi}(\theta)[\widetilde{h}] = P\Phi(\theta)[\widetilde{h}]$. We begin by showing that $\mathbb{G}_n\Phi(\theta_0)[\widetilde{h}]$ converges in distribution to a tight random element $\mathcal{W}$ in $l^\infty(H)$. By Lemma 1, the class $\{\Phi(\theta)[\widetilde{h}] : \theta \in B_\delta(\theta_0), \widetilde{h} \in H\}$ is P-Donkser. It follows that the class $\{\Phi(\theta_0)[\widetilde{h}] : \widetilde{h} \in H\}$, as a subset of a Donsker class, is also Donsker (van der Vaart and Wellner, 1996b, Theorem 2.10.1). We note that the function $\Phi(\theta)[\widetilde{h}]$ involves the terms $g_1\{\rho_0(T;\theta)\}$, $H_1(L, R; \theta)$ and $H_2(L, R; \theta)$, of which the denominators are all bounded away from 0 as argued in Lemma 1. Then, under Conditions $1-5$, we have

$$\sup_{\widetilde{h}\in H} \|\mathbf{\Psi}(\theta_0)[\widetilde{h}]\| < \infty.$$

Hence, $\mathbb{G}_n\Phi(\theta_0)[\widetilde{h}] = n^{1/2}\left\{\mathbf{\Psi}_n(\theta_0)[\widetilde{h}] - \mathbf{\Psi}(\theta_0)[\widetilde{h}]\right\}$ converges weakly to a zero-mean Gaussian process $\mathcal{W}$ in $l^\infty(H)$.

By Lemma 2, the Fréchet-differentiablility of $\mathbf{\Psi}(\theta)$ at $\theta = \theta_0$ can be checked straightforwardly. In particular, we consider one-dimensional submodels $\eta \to \theta_0 +$

$\eta(\theta-\theta_0)$ and calculate the Fréchet derivative $\dot{\mathbf{\Psi}}_{\theta_0}(\theta-\theta_0)$ using the the weaker form

$$\begin{aligned}\dot{\mathbf{\Psi}}_{\theta_0}(\theta-\theta_0) =& \frac{d\mathbf{\Psi}(\theta_0+\eta(\theta-\theta_0))}{d\eta}\bigg|_{\eta=0} \\ =& Q_1^\top[\widetilde{h}](\beta-\beta_0)+\int_0^\tau Q_2[\widetilde{h}](t)d(A(t)-A_0(t)).\end{aligned}$$

Detailed calculations and expressions for $Q_1[\widetilde{h}]$ and $Q_2[\widetilde{h}](\cdot)$ are given in Lemma 2. In particular, $Q_1[\widetilde{h}]$ and $Q_2[\widetilde{h}](\cdot)$ are given by (B.13). Furthermore, Lemma 3 establishes the invertibility of the Fréchet derivative map $\dot{\mathbf{\Psi}}_{\theta_0}$.

Next, we verify condition (3.3.4) of Theorem 3.3.1 (van der Vaart and Wellner, 1996b), which is sufficient to verify the conditions in Lemma 3.3.5 of van der Vaart and Wellner (1996b). Since the classes $\left\{\Phi(\theta)[\widetilde{h}]:\theta\in B_\delta(\beta_0,A_0),\widetilde{h}\in H\right\}$ and $\{\Phi(\theta_0)[\widetilde{h}]:\widetilde{h}\in H\}$ are both Donsker classes, the class $\{\Phi(\theta)[\widetilde{h}]-\Phi(\theta_0)[\widetilde{h}]:\theta\in B_\delta(\beta_0,A_0),\widetilde{h}\in H\}$ is also P-Donsker for some $\delta>0$ because the sum of two bounded Donsker classes is still a Donsker class (van der Vaart and Wellner, 1996b, Example 2.10.7). Under Conditions $1-5$, it easy to note that $\Phi(\theta)[\widetilde{h}]$ is a continuous function over θ. In addition, $\widetilde{h}(t)$ has bounded total variation over $[0,\tau]$. Hence, $\Phi(\theta)[\widetilde{h}]$ converges to $\Phi(\theta_0)[\widetilde{h}]$ pointwise and uniformly in $\widetilde{h}$. By dominated convergence theorem,

$$\sup_{\widetilde{h}\in H} P\{\Phi(\theta)[\widetilde{h}]-\Phi(\theta_0)[\widetilde{h}]\}^2\to 0,$$

as $\theta\to\theta_0$ (van der Vaart and Wellner, 1996b, p. 317). The consistency of $\hat{\theta}$ has been shown in Theorem 3.1, i.e., $\hat{\theta}$ converges to θ_0 almost surely. Hence, applying Lemma 3.3.5 (van der Vaart and Wellner, 1996b), we have

$$\|\mathbb{G}_n(\Phi(\hat{\theta})-\Phi(\theta_0))\| = o_{p^*}(1+n^{1/2}\|\hat{\theta}-\theta_0\|), \tag{B.8}$$

where $o_{p^*}(1)$ denotes converging to zero in outer probability. Note that the equation

(B.8) can be written as

$$n^{1/2}(\mathbf{\Psi}_n - \mathbf{\Psi})(\hat{\theta}) - n^{1/2}(\mathbf{\Psi}_n - \mathbf{\Psi})(\theta_0) = o_{p^*}(1 + n^{1/2}\|\hat{\theta} - \theta_0\|).$$

In brief, we have showed that (1) $\mathbb{G}_n\Phi(\theta_0)[\widetilde{h}]$ converges in distribution to a tight random element $\mathcal{W}$; (2) the continuous invertibility of the operator $\dot{\mathbf{\Psi}}_{\theta_0}$; (3) condition (3.3.2) of Theorem 3.3.1 (van der Vaart and Wellner, 1996b); (4) $\mathbf{\Psi}(\theta_0) = 0$ and $\mathbf{\Psi}_n(\hat{\theta}) = 0$. The last statement is a trivial result by the definitions of θ_0 and $\hat{\theta}$. According to Theorem 3.3.1 of van der Vaart and Wellner (1996b), we obtain

$$n^{1/2}\dot{\mathbf{\Psi}}_{\theta_0}(\hat{\theta} - \theta_0) = -n^{1/2}(\mathbf{\Psi}_n - \mathbf{\Psi})(\theta_0) + o_{p^*}(1).$$

Finally, the continuous mapping theorem implies

$$n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow -\dot{\mathbf{\Psi}}_{\theta_0}^{-1}\mathcal{W}.$$

This proves the asymptotic normality of $\hat{\theta}$. □

B.3 Asymptotic properties of the weighted bootstrap variance estimator

The following theorem gives the asymptotic properties of the weighted bootstrap estimator, and hence validates the bootstrap procedure.

Theorem B.3. *Under Condition* $1-5$*, the conditional distribution of* $\sqrt{n}(\widetilde{\theta} - \hat{\theta})$ *given the data converges weakly to the asymptotic distribution of* $\sqrt{n}(\hat{\theta} - \theta_0)$

Proof. Let $e_1, \ldots, e_n$ be positive i.i.d random variables with a standard exponential distribution. Hence, $\mu = P(e_1) = 1 < \infty$, $\sigma^2 = \mathrm{var}(e_1) = 1 < \infty$ and $\|e_1\| < \infty$, where $\|e_1\| = \int_0^\infty \sqrt{P(|e_1 > x|)}dx$. The last inequality is satisfied because the $(2+\epsilon)$ moment of a standard exponential distribution exists for any $\epsilon > 0$ (Kosorok, 2008, p.20). In addition, we assume that $e_1, \ldots, e_n$ are independent of the observed data $\{\Delta_i, \Delta_i T_i, (1-\Delta_i)L_i, (1-\Delta_i)R_i, X_i, Z_i\}$.

Let $\widetilde{e}_i = e_i/\bar{e}$, where $\bar{e} = n^{-1}\sum_{i=1}^n e_i$. Let $\widetilde{\mathbb{P}}_n f = n^{-1}\sum_{i=1}^n \widetilde{e}_i f(\mathcal{O}_i)$ denote the weighted bootstrapped empirical process for any measurable function f. Let $\widetilde{\boldsymbol{\Psi}}_n$ be $\boldsymbol{\Psi}_n$ but with $\mathbb{P}_n$ replaced by $\widetilde{\mathbb{P}}_n$ and $\widetilde{\theta} = (\widetilde{\beta}, \widetilde{A})$ be the weighted bootstrap estimator that solves $\widetilde{\boldsymbol{\Psi}}_n(\theta) = 0$. Let $\widetilde{\boldsymbol{\Psi}}(\theta) = P\left(\widetilde{e}\cdot\Phi(\theta)[\widetilde{h}]\right)$, where $\widetilde{e}$ be a generic version of $\widetilde{e}_1$. By Lemma 1, the class of functions $\left\{\Phi(\theta)[\widetilde{h}] : \theta \in B_\delta(\theta_0), \widetilde{h} \in H\right\}$ is P-Donsker for some fixed $\delta > 0$. Hence, so is the class $\left\{\widetilde{e}\cdot\Phi(\theta)[\widetilde{h}] : \theta \in B_\delta(\theta_0), \widetilde{h} \in H\right\}$ via the multiplier central limit theorem (Kosorok, 2008, Theorem 10.1). We also note that $P\left(\widetilde{e}\cdot\Phi(\theta)[\widetilde{h}]\right) = P\left(\Phi(\theta)[\widetilde{h}]\right)$, which implies that $\widetilde{\boldsymbol{\Psi}}(\theta) = \boldsymbol{\Psi}(\theta)$. Trivially, the consistency of $\widetilde{\theta}$ holds by similar arguments in Theorem 3.1.

The weighted bootstrap empirical process is defined as

$$\widetilde{\mathbb{G}}_n\Phi(\theta)[\widetilde{h}] = n^{1/2}\left\{\widetilde{\mathbb{P}}_n\Phi(\theta)[\widetilde{h}] - \mathbb{P}_n\Phi(\theta)[\widetilde{h}]\right\}.$$

By the Taylor series expansion, we have

$$\begin{aligned} 0 &= \widetilde{\mathbb{P}}_n\Phi(\widetilde{\theta})[\widetilde{h}] - \widetilde{\mathbb{P}}_n\Phi(\hat{\theta})[\widetilde{h}] + \widetilde{\mathbb{P}}_n\Phi(\hat{\theta})[\widetilde{h}] - \mathbb{P}_n\Phi(\hat{\theta})[\widetilde{h}] \\ &= \left(\left.\frac{\partial\widetilde{\mathbb{P}}_n\Phi(\theta)[\widetilde{h}]}{\partial\theta}\right|_{\theta=\hat{\theta}}\right)(\widetilde{\theta}-\hat{\theta}) + (\widetilde{\mathbb{P}}_n - \mathbb{P}_n)\Phi(\hat{\theta})[\widetilde{h}] + o_p(\|\widetilde{\theta}-\theta_0\| + \|\hat{\theta}-\theta_0\|) \end{aligned} \tag{B.9}$$

By Theorem 2.6 of Kosorok (2008), the conditional distribution of $(\widetilde{\mathbb{P}}_n - \mathbb{P}_n)\Phi(\hat{\theta})[\widetilde{h}]$ given the data is asymptotically equivalent to the distribution of $(\mathbb{P}_n - P)\Phi(\hat{\theta})[\widetilde{h}]$ by the fact that $\mu = \sigma^2 = 1$ with a sequence of i.i.d standard exponential random variables. Hence, (B.9) can be written as

$$\begin{aligned} n^{1/2}\dot{\boldsymbol{\Psi}}_{\theta_0}(\widetilde{\theta}-\hat{\theta}) &= -n^{1/2}(\widetilde{\mathbb{P}}_n - \mathbb{P}_n)\Phi(\hat{\theta})[\widetilde{h}] + o_p(1) \\ &= -n^{1/2}(\mathbb{P}_n - P)\Phi(\hat{\theta})[\widetilde{h}] + o_p(1) \\ &= -\mathbb{G}_n\Phi(\theta_0)[\widetilde{h}] + o_p(1) \end{aligned}$$

Thus, Lemma 3 and the continuous mapping theorem give

$$n^{1/2}(\widetilde{\theta} - \hat{\theta}) \rightsquigarrow -\dot{\mathbf{\Psi}}_{\theta_0}^{-1}\mathcal{W}.$$

We conclude that $n^{1/2}(\widetilde{\theta} - \hat{\theta})$ converges to a zero-mean Gaussian process. Moreover, $n^{1/2}(\widetilde{\theta} - \hat{\theta})$ and $n^{1/2}(\hat{\theta} - \theta_0)$ have the same asymptotic distribution. □

B.4 Some useful Lemmas

Lemma 1. *Under Conditions* $1-5$, *the class of functions*

$$\left\{\Phi(\theta)[\widetilde{h}] : \theta \in B_\delta(\theta_0), \widetilde{h} \in H\right\}$$

is P-Donsker for some fixed $\delta > 0$.

Proof. We begin by rewriting (B.4) and (B.6) as follows

$$\begin{aligned}\Phi_1(\theta)[h_0] = h_0^\top \Big[&\Delta Z(T) - \Delta g_1\{\rho_0(T;\theta)\}\rho_1(T;\theta) \\ &+ (1-\Delta)H_1(L,R;\theta)I(R<\infty)\left\{\rho_1(R;\theta) - \rho_1(L;\theta)\right\} \\ &- (1-\Delta)H_2(L,R;\theta)\rho_1(R^*;\theta)\Big]\end{aligned} \tag{B.10}$$

and

$$\begin{aligned}\Phi_2(\theta)[h] = \sum_{j=1}^{q} \Big[&\Delta h_j(T)X_j(T) \\ &- \Delta g_1\{\rho_0(T;\theta)\} \int_0^\tau I(t \le T)e^{\beta^\top Z(t)}h_j(t)X_j(t)X^\top(t)dA(t) \\ &+ (1-\Delta)H_1(L,R;\theta) \int_0^\tau I(L < t \le R < \infty)e^{\beta^\top Z(t)}h_j(t)X_j(t)X^\top(t)dA(t) \\ &- (1-\Delta)H_2(L,R;\theta) \int_0^\tau I(t \le R^*)e^{\beta^\top Z(t)}h_j(t)X_j(t)X^\top(t)dA(t)\Big].\end{aligned} \tag{B.11}$$

To show the class $\left\{\Phi(\theta)[\widetilde{h}] = \Phi_1(\theta)[h_0] + \Phi_2(\theta)[h] : \theta \in B_\delta(\theta_0), \widetilde{h} = (h_0, h) \in H\right\}$ is P-Donsker for some $\delta > 0$, we need to show that each component is Donsker. Then the desired conclusion follows by the Donsker preservation properties (Kosorok, 2008, Corollary 9.32), i.e., the summation and multiplication of P-Donsker classes are also P-Donsker classes.

By Condition 1, $X(t)$ and $Z(t)$ are uniformly bounded with uniformly bounded total variations over $[0, \tau]$. Trivially, the classes $\{X(t) : t \in [0, \tau]\}$ and $\{Z(t) : t \in [0, \tau]\}$ are Donsker by Theorem 2.7.5 (van der Vaart and Wellner, 1996b) and Example 19.11 (van der Vaart, 1998). The classes $\{X(T)\}$ and $\{Z(T)\}$ are also Donsker classes because they are subsets of some Donsker classes. The classes $\{\Delta\}$ and $\{1 - \Delta\}$ are both P-Donsker because they are bounded and square-integrable (van der Vaart, 1998, p.270). Condition 2 indicates that the class $\{\beta \in \mathcal{B}\}$ is a Donsker class, and so is $\{\beta^\top Z(t) : t \in [0, \tau], \beta \in \mathcal{B}\}$ as the product of two bounded Donsker classes is also a Donsker class. The class $\{e^{\beta^\top Z(t)}, \beta \in \mathcal{B}, t \in [0, \tau]\}$ is P-Donsker since exponentiation is Lipschitz continuous on compacts. Note that

$$\rho_0(T; \theta) = \int_0^T e^{\beta^\top Z(s)} X^\top(s) dA(s) = \sum_{j=1}^q \int_0^T e^{\beta^\top Z(s)} X_j(s) dA_j(s).$$

Under Condition 2, each $A_j(t)$ $(j = 1, \ldots, q)$ has bounded total variation over $[0, \tau]$. By Theorem 7.2.4 in Dudley (2002), we can find two nondecreasing functions $A_{j1}(t)$ and $A_{j2}(t)$ such that $A_j(t) = A_{j1}(t) - A_{j2}(t)$. Thus,

$$\int_0^T e^{\beta^\top Z(s)} X_j(s) dA_j(s) = \int_0^T e^{\beta^\top Z(s)} X_j(s) dA_{j1}(s) - \int_0^T e^{\beta^\top Z(s)} X_j(s) dA_{j2}(s).$$

Following Zeng et al. (2016), the class $\{\int_0^T e^{\beta^\top Z(s)} X_j(s) dA_{j1}(s) : \theta \in B_\delta(\theta_0)\}$ is a Donsker class because it is a convex hull of functions $\{I(T \geq s) \exp\{\beta^\top Z(s)\} X_j(s)\}$. Likewise, the class $\{\int_0^T e^{\beta^\top Z(s)} X_j(s) dA_{j2}(s) : \theta \in B_\delta(\theta_0)\}$ is P-Donsker. Hence, the class $\{\int_0^T e^{\beta^\top Z(s)} X_j(s) dA_j(s) : \theta \in B_\delta(\theta_0)\}$ is P-Donsker because the sum of bounded

Donsker classes are also Donsker. It follows that the class $\{\Delta\rho_0(T;\theta) : \theta \in B_\delta(\theta_0)\}$ is a Donsker class. Similarly, the following classes

$$\{(1-\Delta)\rho_0(L;\theta) : \theta \in B_\delta(\theta_0)\}, \{(1-\Delta)\rho_0(R;\theta) : \theta \in B_\delta(\theta_0)\},$$
$$\{\Delta\rho_1(T;\theta) : \theta \in B_\delta(\theta_0)\}, \{(1-\Delta)\rho_1(L;\theta) : \theta \in B_\delta(\theta_0)\},$$
$$\{(1-\Delta)\rho_1(R;\theta) : \theta \in B_\delta(\theta_0)\}, \{(1-\Delta)\rho_1(R^*;\theta) : \theta \in B_\delta(\theta_0)\}$$

are all Donsker classes. By Condition 4, $G(x)$ is thrice continuously differentiable on $[0,\infty)$ and $G'(x) > 0$ for any $x \in [0,\infty)$, then the following functions

$$g_1[\rho_0(T;\theta)] = G'\{\rho_0(T;\theta)\} - \frac{G''\{\rho_0(T;\theta)\}}{G'\{\rho_0(T;\theta)\}}$$
$$\exp[-G\{\rho_0(L;\theta)\}]G'\{(\rho_0(L;\theta)\}$$
$$\exp[-G\{\rho_0(R;\theta)\}]G'\{(\rho_0(R;\theta)\}I(R<\infty)$$

are all bounded for any $\theta \in B_\delta(\theta_0)$. Notice that the denominators of $H_1(L,R;\theta)$ and $H_2(L,R;\theta)$ are $\exp[-G\{\rho_0(L;\theta)\}] - \exp[-G\{\rho_0(R;\theta)\}]I(R<\infty)$, which is bounded away from zero under Conditions $3-4$. Since any continuously differentiable function is locally Lipschitz, the classes $\{\Delta g_1\{\rho_0(T;\theta)\} : \theta \in B_\delta(\theta_0)\}$, $\{(1-\Delta)H_1(L,R;\theta) : \theta \in B_\delta(\theta_0)\}$ and $\{(1-\Delta)H_2(L,R;\theta) : \theta \in B_\delta(\theta_0)\}$ are all Donsker classes due to the preservation of the Donsker property under Lipschitz-continuous transformations by Theorem 9.31 (Kosorok, 2008). Now we conclude that the class $\{\Phi_1(\theta)[h_0] : \theta \in B_\delta(\theta_0), h_0 \in \mathcal{R}^d\}$ is a donsker class since $\Phi_1(\theta)[h_0]$ depends on h_0 linearly by Theorem 2.10.6 of van der Vaart and Wellner (1996b).

The class $\{h_j : h_j \in BV_1[0,\tau]\}$ $(j = 1,\ldots,q)$ is a Donsker class, according to Theorem 2.7.5 (van der Vaart and Wellner, 1996b) and Example 19.11 (van der Vaart, 1998). Thus, the class $\{h_j(T) : h_j \in BV_1[0,\tau]\}$ $(j = 1,\ldots,q)$ as a class of functions of T is also P-Donsker. Now we only need to show the class

$$\left\{\int_0^{(\cdot)} e^{\beta^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA(t) : \theta \in B_\delta(\theta_0), h_j \in BV_1[0,\tau]\right\},$$

is a Donsker class since finite summation of donsker classes are still Donkser class. This follows because the class of functions with an upper bound of their total variations is Donsker by Example 19.11 and Theorem 19.5 of van der Vaart (1998) under Conditions $1-5$. To this end, we conclude that under Conditions $1-5$, the class of function $\left\{\Phi(\theta)[\widetilde{h}] : \theta \in B_\delta(\theta_0), \widetilde{h} \in H\right\}$ is P-Donsker for some $\delta > 0$ because the sums and products of bounded Donsker classes are Donsker classes. □

Lemma 2. *Under Conditions* $1-5$*, the map* $\mathbf{\Psi} : \mathcal{R}^d \times BV^q[0,\tau] \to l^\infty(H)$ *is Fréchet-differentiable at* $\theta = \theta_0$*, with a derivative*

$$\dot{\mathbf{\Psi}}_{\theta_0}(\theta - \theta_0) = Q_1^\top[\widetilde{h}](\beta - \beta_0) + \int_0^\tau Q_2[\widetilde{h}](t) d(A(t) - A_0(t)), \tag{B.12}$$

where

$$\begin{aligned} Q_1[\widetilde{h}] &= B_1^\top h_0 + \sum_{j=1}^q \int_0^\tau B_{2,j}(t) h_j(t) dA_0(t), \\ Q_2[\widetilde{h}](t) &= h_0^\top B_3(t) + B_4[h](t). \end{aligned} \tag{B.13}$$

The expressions of B_1*,* $B_{2,j}(t)$*,* $B_3(t)$*, and* $B_4[h](t)$ *are given in* (B.25), (B.26), (B.27) *and* (B.28)*, respectively, in Appendix B.5. The map* $\mathbf{\Psi}_n$ *has the same properties and similar Fréchet derivative map at* $\theta = \theta_0$*, denoted as* $\dot{\mathbf{\Psi}}_{\theta_0,n}$ *by replacing the expectations* E *in the terms* B_1*,* $B_{2,j}(t)$*,* $B_3(t)$*, and* $B_4[h](t)$ *with the empirical measure* $\mathbb{P}_n$*. Furthermore, both* $\dot{\mathbf{\Psi}}_\theta$ *and* $\dot{\mathbf{\Psi}}_{\theta,n}$ *depend continuously on* θ*.*

Proof. The Fréchet derivative of $\mathbf{\Psi}(\theta)$ at $\theta = \theta_0$ can be derived using (B.10) and (B.11). Consider the one-dimensional submodels $\eta \to \theta_0 + \eta(\theta - \theta_0)$. The Fréchet derivative $\dot{\mathbf{\Psi}}_{\theta_0}(\theta-\theta_0)$ can be computed based on the weaker form

$$\begin{aligned}\dot{\mathbf{\Psi}}_{\theta_0}(\theta-\theta_0) &= \frac{d\mathbf{\Psi}(\theta_0+\eta(\theta-\theta_0))}{d\eta}\bigg|_{\eta=0} \\ &= Q_1^\top[\widetilde{h}](\beta-\beta_0) + \int_0^\tau Q_2[\widetilde{h}](t)d(A(t)-A_0(t)).\end{aligned} \tag{B.14}$$

It can be easy to show that $Q_1[\widetilde{h}] = B_1^\top h_0 + \sum_{j=1}^q \int_0^\tau B_{2,j}(t)h_j(t)dA_0(t)$ and $Q_2[\widetilde{h}](t) = h_0^\top B_3(t) + B_4[h](t)$, where B_1 is a $d\times d$ matrix, $B_{2,j}(\cdot)$ and $B_3(\cdot)$ are $d\times q$ matrices, and $B_4[h](\cdot)$ is a $1\times q$ vector of functions. The detailed calculations are provided in Appendix B.5. It can be shown that $||\mathbf{\Psi}(\theta)-\mathbf{\Psi}(\theta_0)-\dot{\mathbf{\Psi}}_{\theta_0}(\theta-\theta_0)|| = o(||\theta-\theta_0||)$ as $\theta\to\theta_0$. Hence, $\mathbf{\Psi}(\theta)$ is Fréchet-differentiable at θ_0. The Fréchet derivative of $\mathbf{\Psi}_n(\theta) = \mathbb{P}_n\Phi(\theta)[\widetilde{h}]$ with respect to θ at $\theta=\theta_0$, denoted as $\dot{\mathbf{\Psi}}_{\theta_0,n}$, can be derived closely. In particular, we replace $\mathbf{\Psi}$ with $\mathbf{\Psi}_n$ in (B.14) and the expectations E in the terms B_1, $B_{2,j}(t)$, $B_3(t)$, and $B_4[h](t)$ with the empirical measure $\mathbb{P}_n$ to obtain $\dot{\mathbf{\Psi}}_{\theta_0,n}$. Then one can show that $\|\mathbf{\Psi}_n(\theta)-\mathbf{\Psi}_n(\theta_0)-\dot{\mathbf{\Psi}}_{\theta_0,n}(\theta-\theta_0)\| = o(\|\theta-\theta_0\|)$ as $\theta\to\theta_0$. Hence, $\mathbf{\Psi}_n(\theta)$ is also Fréchet-differentiable at θ_0. Clearly, both maps $\dot{\mathbf{\Psi}}_\theta$ and $\dot{\mathbf{\Psi}}_{\theta,n}$ depend continuously on θ in $B_\delta(\theta_0)$. □

Lemma 3. *Under Conditions* $1-5$*, the map* $\dot{\mathbf{\Psi}}_{\theta_0}$ *is continuously invertible.*

Proof. To show the invertibility of the map $\dot{\mathbf{\Psi}}_{\theta_0}$, we can start by establishing the relationship between the maximum likelihood estimator and the proposed ES estimator. The observed-data likelihood function for a single subject takes the form

$$\begin{aligned}l(\beta,A) = \bigg(&\Lambda_X'(T)e^{\beta^\top Z(T)}G'\bigg\{\int_0^T e^{\beta^\top Z(s)}X^\top(s)dA(s)\bigg\} \\ &\times\exp\bigg[-G\bigg\{\int_0^T e^{\beta^\top Z(s)}X^\top(s)dA(s)\bigg\}\bigg]\bigg)^\Delta \\ &\bigg(\exp\bigg[-G\bigg\{\int_0^L e^{\beta^\top Z(s)}X^\top(s)dA(s)\bigg\}\bigg] \\ &-\exp\bigg[-G\bigg\{\int_0^R e^{\beta^\top Z(s)}X^\top(s)dA(s)\bigg\}\bigg]\bigg)^{1-\Delta},\end{aligned} \tag{B.15}$$

where $\Lambda_X^{'}(\cdot)$ and $G^{'}(\cdot)$ denote the derivatives of $\Lambda_X(\cdot)$ and $G(\cdot)$, respectively.

We consider the submodels $\beta_\eta = \beta + \eta h_0$, and $A_{j,\eta}(t) = \int_0^t (1 + \eta h_j(s))dA_j(s)$ $(j = 1, \dots, q)$, where $\widetilde{h} = (h_0, h) \in H$. Here, we use A_η to represent that each A_j is replaced with $A_{j,\eta}$. Thus, the derivatives of the observed data log-likelihood for a single subject along the submodels are

$$\begin{aligned}
\Phi_1^*(\theta)[h_0] &= \left.\frac{d \log l(\beta_\eta, A)}{d\eta}\right|_{\eta=0} \\
&= \Delta\left[\frac{G''\{\rho_0(T;\theta)\}}{G'\{\rho_0(T;\theta)\}} - G'\{\rho_0(T;\theta)\}\right] h_0^\top \rho_1(T;\theta) + \Delta h_0^\top Z(T) \\
&\quad + (1-\Delta) h_0^\top \frac{\exp[-G\{\rho_0(R;\theta)\}] G'\{\rho_0(R;\theta)\}\rho_1(R;\theta) I(R<\infty)\}}{\exp[-G\{\rho_0(L;\theta)\}] - \exp[-G\{\rho_0(R;\theta)\}] I(R<\infty)} \\
&\quad - (1-\Delta) h_0^\top \frac{\exp[-G\{\rho_0(L;\theta)\}] G'\{\rho_0(L;\theta)\}\rho_1(L;\theta)}{\exp[-G\{\rho_0(L;\theta)\}] - \exp[-G\{\rho_0(R;\theta)\}] I(R<\infty)}
\end{aligned} \tag{B.16}$$

and

$$\begin{aligned}
\Phi_2^*(\theta)[h] &= \left.\frac{d \log l(\beta, A_\eta)}{d\eta}\right|_{\eta=0} \\
&= \Delta\left[\frac{G''\{\rho_0(T;\theta)\}}{G'\{\rho_0(T;\theta)\}} - G'\{\rho_0(T;\theta)\}\right] \rho_2(T;\theta)[h] \\
&\quad + \Delta \frac{(h(T) \circ X(T))^\top A\{T\}}{X^\top(T) A\{T\}} \\
&\quad + (1-\Delta) \frac{\exp[-G\{\rho_0(R;\theta)\}] G'\{\rho_0(R;\theta)\}\rho_2(R;\theta)[h] I(R<\infty)\}}{\exp[-G\{\rho_0(L;\theta)\}] - \exp[-G\{\rho_0(R;\theta)\}] I(R<\infty)} \\
&\quad - (1-\Delta) \frac{\exp[-G\{\rho_0(L;\theta)\}] G'\{\rho_0(L;\theta)\}\rho_2(L;\theta)[h]}{\exp[-G\{\rho_0(L;\theta)\}] - \exp[-G\{\rho_0(R;\theta)\}] I(R<\infty)},
\end{aligned} \tag{B.17}$$

where $\rho_2(t;\theta)[h] = \sum_{j=1}^q \int_0^t e^{\beta^\top Z(s)} h_j(s) X_j(s) dA_j(s)$ and $A\{T\}$ denotes the jump size of $A(\cdot)$ at T. Let $\Phi^*(\theta)[\widetilde{h}] = \Phi_1^*(\theta)[h_0] + \Phi_2^*(\theta)[h]$. Therefore, the maximum likelihood estimator (β^*, A^*) can be obtained by solving the following equation

$$\mathbb{P}_n \Phi^*(\theta)[\widetilde{h}] \equiv \mathbb{P}_n \Phi_1^*(\theta)[h_0] + \mathbb{P}_n \Phi_2^*(\theta)[h] = 0 \quad \text{for every } \widetilde{h} = (h_0, h) \in H.$$

Recall that

$$\Phi_1(\theta)[h_0] = \Delta h_0^\top Z(T) - \Delta g_1\{\rho_0(T;\theta)\} \int_0^\tau I(t \le R^*) e^{\beta^\top Z(t)} h_0^\top Z(t) X^\top(t) dA(t)$$
$$+ (1-\Delta) H_1(L,R;\theta) \int_0^\tau I(L < t \le R < \infty) e^{\beta^\top Z(t)} h_0^\top Z(t) X^\top(t) dA(t)$$
$$- (1-\Delta) H_2(L,R;\theta) \int_0^\tau I(t \le R^*) e^{\beta^\top Z(t)} h_0^\top Z(t) X^\top(t) dA(t),$$

and

$$\Phi_2(\theta)[h] = \sum_{j=1}^q \Big[\Delta h_j(T) X_j(T)$$
$$- \Delta g_1\{\rho_0(T;\theta)\} \int_0^\tau I(t \le R^*) e^{\beta^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA(t)$$
$$+ (1-\Delta) H_1(L,R;\theta) \int_0^\tau I(L < t \le R < \infty) e^{\beta^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA(t)$$
$$- (1-\Delta) H_2(L,R;\theta) \int_0^\tau I(t \le R^*) e^{\beta^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA(t)\Big].$$

It can show that $\Phi_1(\theta)[h_0] = \Phi_1^*(\theta)[h_0]$ for every $h_0 \in \mathcal{R}^d$, and $\Phi_2(\theta)[h]$ can be rewritten as follows:

$$\Phi_2(\theta)[h] = \Delta\left[\frac{G''\{\rho_0(T;\theta)\}}{G'\{\rho_0(T;\theta)\}} - G'\{\rho_0(T;\theta)\}\right]\rho_3(T;\theta)[h] + \Delta \sum_{j=1}^q h_j(T) X_j(T)$$
$$+ (1-\Delta)\frac{\exp[-G\{\rho_0(R;\theta)\}] G'\{\rho_0(R;\theta)\}\rho_3(R;\theta)[h] I(R<\infty)\}}{\exp[-G\{\rho_0(L;\theta)\}] - \exp[-G\{\rho_0(R;\theta)\}] I(R<\infty)} \tag{B.18}$$
$$- (1-\Delta)\frac{\exp[-G\{\rho_0(L;\theta)\}] G'\{\rho_0(L;\theta)\}\rho_3(L;\theta)[h]}{\exp[-G\{\rho_0(L;\theta)\}] - \exp[-G\{\rho_0(R;\theta)\}] I(R<\infty)},$$

where $\rho_3(t;\theta)[h] = \sum_{j=1}^q \int_0^t e^{\beta^\top Z(s)}\{h_j(s) X_j(s)\} X^\top(s) dA(s)$ for each $h_j \in BV_1[0,\tau]$. Let $\Phi(\theta)[\widetilde{h}] = \Phi_1(\theta)[h_0] + \Phi_2(\theta)[h]$. The proposed ES-estimator essentially solves

$$\mathbb{P}_n \Phi(\theta)[\widetilde{h}] = \mathbb{P}_n \Phi_1(\theta)[h_0] + \mathbb{P}_n \Phi_2(\theta)[h] = 0, \text{for every } \widetilde{h} = (h_0, h) \in H.$$

Next, we establish the relationship between $\Phi(\theta)[\widetilde{h}]$ and $\Phi^*(\theta)[\widetilde{h}]$, where $\Phi(\theta)[\widetilde{h}]$

is the estimating equation constructed to estimate the parameters using an ES algorithm, and $\Phi^*(\theta)[\widetilde{h}]$ is the score equation calculated by taking the derivatives of the log-likelihood function along the submodels defined previously. Let $h_j^*(t) = (h_0/q, h_j(t)X_j(t), \ldots, h_j(t)X_j(t))$ for $j = 1, \ldots, q$. Then, it can be verified that

$$\Phi(\theta)[\widetilde{h}] = \sum_{j=1}^{q} \Phi^*(\theta)[h_j^*], \tag{B.19}$$

The Fréchet derivative map of $\mathbf{\Psi}(\theta) = P\Phi(\theta)$ at θ_0, denoted as $\dot{\mathbf{\Psi}}(\theta_0) = P\dot{\Phi}(\theta_0)$, takes the form shown in (B.12) in Lemma 2. Similarly, the Fréchet derivative map of $\mathbf{\Psi}^*(\theta) = P\Phi^*(\theta)$ at θ_0, denoted as $\dot{\mathbf{\Psi}}^*(\theta_0) = P\dot{\Phi}^*(\theta_0)$, can be calculated via a similar weaker form as shown in (B.14) using equations (B.16) and (B.17). The details have been omitted due to the similarity of the calculations involved.

From Lemma 2, $\dot{\mathbf{\Psi}}(\theta_0)[\widetilde{h}]$ is a linear operator. To show that $\dot{\mathbf{\Psi}}(\theta_0)[\widetilde{h}]$ is continuously invertible, it is sufficient to prove that $\dot{\mathbf{\Psi}}(\theta_0)[\widetilde{h}]$ is a one-to-one map and then it is invertible (Rudin, 1973). If $\widetilde{h} = 0$, it is easily to note that $\dot{\mathbf{\Psi}}(\theta_0)[\widetilde{h}] = 0$ for any θ in the neighborhood of θ_0. Additionally, from equation (B.19), we have the following

$$\dot{\mathbf{\Psi}}(\theta_0)[\widetilde{h}] = P\dot{\Phi}(\theta_0)[\widetilde{h}] = \sum_{j=1}^{q} P\dot{\Phi}^*(\theta_0)[h_j^*] = -\sum_{j=1}^{q} P\{\Phi^*(\theta_0)[h_j^*]\}^2, \tag{B.20}$$

where the last equality is followed by the loglikelihood property. Hence, $\dot{\mathbf{\Psi}}(\theta_0)[\widetilde{h}] = 0$ implies that, for each j ($j = 1, \ldots, q$), $\Phi^*(\theta_0)[h_j^*] = 0$ with probability 1. Then we show that $\Phi^*(\theta_0)[h_j^*] = 0$ ($j = 1, \ldots, q$) implies that $h_0 = 0$ and $h(\cdot) = 0$ for any t in $[0, \tau]$. By Condition 3, we choose $\Delta = 1$. When $j = 1$, $\Phi^*(\theta_0)[h_1^*] = 0$ with $h_1^*(t) = (h_0/q, h_1(t)X_1(t), \ldots, h_1(t)X_1(t))$, reduces to the following two equations:

$$h_0^\top\left(\left[\frac{G''\{\rho_0(T;\theta)\}}{G'\{\rho_0(T;\theta)\}} - G'\{\rho_0(T;\theta)\}\right]\rho_1(T;\theta) + Z(T)\right) = 0 \quad \text{for any } T,$$

and

$$\sum_{j=1}^{q}\int_0^T \left[\frac{G''\{\rho_0(T;\theta)\}}{G'\{\rho_0(T;\theta)\}} - G'\{\rho_0(T;\theta)\}\right] e^{\beta^\top Z(s)} h_1(s)X_1(s)X_j(s)dA_j(s)$$
$$+\sum_{j=1}^{q}\frac{h_1(T)X_1(T)X_j(T)A_j\{T\}}{X^\top(T)A\{T\}} = 0 \quad \text{for any } T.$$

Hence, we obtain that $h_0 = 0$ and

$$\int_0^T \left[\frac{G''\{\rho_0(T;\theta)\}}{G'\{\rho_0(T;\theta)\}} - G'\{\rho_0(T;\theta)\}\right] e^{\beta^\top Z(s)} h_1(s)X_1(s)X^\top(s)dA(s) + h_1(T)X_1(T) = 0. \tag{B.21}$$

which is a linear Volterra integral equation with a unique solution (Polyanin and Manzhirov, 2008). Note that (B.21) holds for any T. Hence, $h_1(t) = 0$ for any $t \in [0,\tau]$ since $X(t)$ is not identically zero for all $t \in [0,\tau]$ by Condition 1. By applying similar arguments for $j = 2,\ldots,q$, we obtain that $h_2(t) = \cdots = h_q(t) = 0$ for any $t \in [0,\tau]$. Consequently, $\dot{\mathbf{\Psi}}(\theta_0)[\widetilde{h}]$ is continuously invertible. □

Lemma 4. *Under Conditions* $1-5$*, the map* $\dot{\mathbf{\Psi}}_{\theta_0,n}$ *is invertible for larger enough* n*.*

Proof. The expressions of $\dot{\mathbf{\Psi}}_\theta$ are given in Lemma 2 and Appdenix B.5, see (B.12), (B.13), (B.14), (B.23), and (B.24). Following the similar steps as in the proof of Lemma 1, we can show that

$$\dot{\mathbf{\Psi}}_\theta(\theta^*)[h] - \dot{\mathbf{\Psi}}_{\theta,n}(\theta^*)[h] = o_p(1) \tag{B.22}$$

$(\theta,\theta^*,\widetilde{h})$ in $B_\delta(\beta_0,A_0)\times(\mathcal{R}^d\times BV^q[0,\tau])\times(\mathcal{R}^d\times BV_1^q[0,\tau])$ for some $\delta > 0$. By Lemma 3, the map $\dot{\mathbf{\Psi}}_{\theta_0}$ is invertible. Following Part (i) of Lemma 6.16 in Kosorok (2008), there exists a constant $c_1 > 0$ such that $\|\dot{\mathbf{\Psi}}_{\theta_0}(\theta-\theta_0)\| \geq c_1\|\theta-\theta_0\|$ for all θ in θ in $\mathcal{R}^d \times BV^q[0,\tau]$. Combining this with (B.22), there exists a positive constant

c_2 such that

$$\left\| \frac{\dot{\mathbf{\Psi}}_{\theta_0,n}(\theta-\theta_0)}{\|\theta-\theta_0\|} \right\| = \left\| \dot{\mathbf{\Psi}}_{\theta_0,n}\left(\frac{\theta-\theta_0}{\|\theta-\theta_0\|}\right) \right\| = \left\| \dot{\mathbf{\Psi}}_{\theta_0}\left(\frac{\theta-\theta_0}{\|\theta-\theta_0\|}\right) + o_p(1) \right\| \geq c_1 + o_p(1) \\ \geq c_2,$$

as $n \to \infty$ for any θ in $\mathcal{R}^d \times BV^q[0,\tau]$. Thus, $\|\dot{\mathbf{\Psi}}_{\theta_0,n}(\theta-\theta_0)\| \geq c_2\|\theta-\theta_0\|$ as $n \to \infty$ for any θ in θ in $\mathcal{R}^d \times BV^q[0,\tau]$. By applying Lemma 6.16 (Kosorok, 2008) again, $\dot{\mathbf{\Psi}}_{\theta_0,n}$ is invertible for larger enough n. □

B.5 The Fréchet Derivative Map

We provide the calculations of the Fréchet derivative of $\mathbf{\Psi}(\theta) = P\Phi(\beta, A)$ in this subsection. The Fréchet derivative of $\mathbf{\Psi}(\theta)$ at θ_0 is given by the map

$$\begin{aligned} \dot{\mathbf{\Psi}}_{\theta_0}(\theta-\theta_0) &= \left.\frac{d\mathbf{\Psi}(\theta_0+\eta(\theta-\theta_0))}{d\eta}\right|_{\eta=0} \\ &= Q_1^\top[\widetilde{h}](\beta-\beta_0) + \int_0^\tau Q_2[\widetilde{h}](t)d(A(t)-A_0(t)), \end{aligned} \tag{B.23}$$

where

$$\begin{aligned} Q_1[\widetilde{h}] &= B_1^\top h_0 + \sum_{j=1}^q \int_0^\tau B_{2,j}(t)h_j(t)dA_0(t), \\ Q_2[\widetilde{h}](t) &= h_0^\top B_3(t) + B_4[h](t). \end{aligned} \tag{B.24}$$

Here, B_1 is a $d \times d$ matrix, $B_{2,j}(\cdot)$ and $B_3(\cdot)$ are $d \times q$ matrices, and $B_4[h](\cdot)$ is $1 \times q$ vector of functions.

Then we show the calculations of the aforementioned terms in (B.24). To simplify the notations, let $g_2(x) = \exp\{-G(x)\}$ and $g_3(x) = \exp\{-G(x)\}G'(x)$. Then we can write

$$H_1(L,R;\theta) = \frac{g_3\{\rho_0(L;\theta)\}}{g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)}$$

and

$$H_2(L,R;\theta) = \frac{g_3\{\rho_0(L;\theta)\} - g_3\{\rho_0(R;\theta)\}I(R<\infty)}{g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)}.$$

Note that $g_2'(x) = -g_3(x)$ and $g_3'(x) = \exp\{-G(x)\}[G''(x) - \{G'(x)\}^2]$. In addition, since $g_1(x) = G'(x) - G''(x)/G'(x)$, we have that $g_1'(x) = G''(x) - G^{(3)}(x)/G'(x) + \{G''(x)/G'(x)\}^2$, where $G^{(3)}(x)$ is the thrice derivative of $G(x)$ with respect to x. Let $a^{\bigotimes 2} = aa^\top$ for any column vector a. B_1 can be easily obtained by taking the derivative of $P\Phi_1(\theta)$ with respect to β, i.e.,

$$\begin{aligned}
B_1 &= \left.\frac{\partial P\Phi_1(\theta)}{\partial\beta^\top}\right|_{\theta=\theta_0} \\
&= -E\left[\Delta g_1\{\rho_0(T;\theta_0)\}\int_0^T e^{\beta_0^\top Z(t)}\{Z(t)\}^{\bigotimes 2}X^\top(t)dA_0(t)\right] \\
&- E\left[\Delta g_1'\{\rho_0(T;\theta_0)\}\{\rho_1(T;\theta_0)\}^{\bigotimes 2}\right] \\
&+ E\left[(1-\Delta)\left\{\int_0^\tau I(L<t\le R<\infty)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\right\}M_1^\top(L,R;\theta_0)\right] \\
&+ E\left[(1-\Delta)H_1(L,R;\theta_0)\int_0^\tau I(L<t\le R<\infty)e^{\beta_0^\top Z(t)}\{Z(t)\}^{\bigotimes 2}X^\top(t)dA_0(t)\right] \\
&- E\left[(1-\Delta)\left\{\int_0^\tau I(t\le R^*)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\right\}M_2^\top(L,R;\theta_0)\right] \\
&- E\left[(1-\Delta)H_2(L,R;\theta_0)\int_0^\tau I(t\le R^*)e^{\beta_0^\top Z(t)}\{Z(t)\}^{\bigotimes 2}X^\top(t)dA_0(t)\right],
\end{aligned} \tag{B.25}$$

where

$$\begin{aligned}
M_1(L,R;\theta) &= \frac{\partial H_1(L,R;\theta)}{\partial\beta} \\
&= \frac{g_3'\{\rho_0(L;\theta)\}\rho_1(L;\theta)}{g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)} \\
&+ \frac{g_3\{\rho_0(L;\theta)\}\big[g_3\{\rho_0(L;\theta)\}\rho_1(L;\theta) - g_3\{\rho_0(R;\theta)\}\rho_1(R;\theta)I(R<\infty)\big]}{\big[g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)\big]^2},
\end{aligned}$$

and

$$
\begin{aligned}
M_2(L,R;\theta) &= \frac{\partial H_2(L,R;\theta)}{\partial \beta} \\
&= \frac{g_3'\{\rho_0(L;\theta)\}\rho_1(L;\theta) - g_3'\{\rho_0(R;\theta)\}\rho_1(R;\theta)I(R<\infty)}{g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)} \\
&\quad + \frac{g_3\{\rho_0(L;\theta)\} - g_3\{\rho_0(R;\theta)\}I(R<\infty)}{[g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)]^2} \\
&\quad \times \left[g_3\{\rho_0(L;\theta)\}\rho_1(L;\theta) - g_3\{\rho_0(R;\theta)\}\rho_1(R;\theta)I(R<\infty) \right].
\end{aligned}
$$

Moreover, we calculate

$$
\begin{aligned}
&\left.\frac{\partial P\Phi_2(\theta)[h]}{\partial \beta}\right|_{\theta=\theta_0} \\
&= -\sum_{j=1}^{q} \Bigg(E\left[\Delta g_1[\rho_0(T;\theta_0)] \int_0^T e^{\beta_0^\top Z(t)} X_j(t) h_j(t) Z(t) X^\top(t) dA_0(t)\right] \\
&- E\left[\Delta g_1'\{\rho_0(T;\theta_0)\}\left\{\int_0^T e^{\beta_0^\top Z(t)} X_j(t) h_j(t) X^\top(t) dA_0(t)\right\}\rho_1(T;\theta_0)\right] \\
&+ E\left[(1-\Delta)\left\{\int_0^\tau I(L<t\le R<\infty) e^{\beta_0^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA_0(t)\right\} M_1(L,R;\theta_0)\right] \\
&+ E\left[(1-\Delta) H_1(L,R;\theta_0) \int_0^\tau I(L<t\le R<\infty) e^{\beta_0^\top Z(t)} h_j(t) X_j(t) Z(t) X^\top(t) dA_0(t)\right] \\
&- E\left[(1-\Delta)\left\{\int_0^\tau I(t\le R^*) e^{\beta_0^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA_0(t)\right\} M_2(L,R;\theta_0)\right] \\
&- E\left[(1-\Delta) H_2(L,R;\theta_0) \int_0^\tau I(t\le R^*) e^{\beta_0^\top Z(t)} h_j(t) X_j(t) Z(t) X^\top(t) dA_0(t)\right] \Bigg).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
B_{2,j}(t) = &-E\left[\Delta g_1[\rho_0(T;\theta_0)]I(t \le T)e^{\beta_0^\top Z(t)}X_j(t)Z(t)X^\top(t)\right] \\
&- E\left[\Delta g_1'\{\rho_0(T;\theta_0)\}I(t \le T)e^{\beta_0^\top Z(t)}X_j(t)\rho_1(T;\theta_0)X^\top(t)\right] \\
&+ E\left[(1-\Delta)I(L < t \le R)I(R < \infty)e^{\beta_0^\top Z(t)}X_j(t)M_1(L,R;\theta_0)X^\top(t)\right] \\
&+ E\left[(1-\Delta)H_1(L,R;\theta_0)I(L < t \le R < \infty)e^{\beta_0^\top Z(t)}X_j(t)Z(t)X^\top(t)\right] \\
&- E\left[(1-\Delta)I(t \le R^*)e^{\beta_0^\top Z(t)}X_j(t)M_2(L,R;\theta_0)X^\top(t)\right] \\
&- E\left[(1-\Delta)H_2(L,R;\theta_0)I(t \le R^*)e^{\beta_0^\top Z(t)}X_j(t)Z(t)X^\top(t)\right].
\end{aligned}
\tag{B.26}
$$

Additionally, we can use $\rho_0(t;\beta,A) = \int_0^t e^{\beta^\top Z(s)}X^\top(s)dA(s)$, to derive the following result:

$$
\left.\frac{\partial \rho_0(t;\beta, A+\eta A^*)}{\partial \eta}\right|_{\eta=0} = \rho_0(t;\beta,A^*).
$$

By applying the chain rule, we obtain

$$
\begin{aligned}
&\left.\frac{\partial H_1(L,R;\beta,A+\eta A^*)}{\partial \eta}\right|_{\eta=0} \\
&= \frac{g_3'\{\rho_0(L;\theta)\}\rho_0(L;\beta,A^*)}{g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)} \\
&\quad + \frac{g_3\{\rho_0(L;\theta)\}\left[g_3\{\rho_0(L;\theta)\}\rho_0(L;\beta,A^*) - g_3\{\rho_0(R;\theta)\}\rho_0(R;\beta,A^*)I(R<\infty)\right]}{\left[g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)\right]^2} \\
&= M_3(L,R;\theta)\rho_0(L;\beta,A^*) - M_4(L,R;\theta)\rho_0(R;\beta,A^*),
\end{aligned}
$$

where

$$
\begin{aligned}
M_3(L,R;\theta) = &\frac{g_3'\{\rho_0(L;\theta)\}}{g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)} \\
&+ \frac{[g_3\{\rho_0(L;\theta)\}]^2}{\left[g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)\right]^2},
\end{aligned}
$$

and

$$
M_4(L,R;\theta) = \frac{g_3\{\rho_0(L;\theta)\}g_3\{\rho_0(R;\theta)\}I(R<\infty)}{\left[g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)\right]^2}.
$$

Using the same technique, we obtain

$$\begin{aligned}
&\left.\frac{\partial H_2(L,R;\beta,A+\eta A^*)}{\partial \eta}\right|_{\eta=0} \\
&= \frac{g_3'\{\rho_0(L;\theta)\}\rho_0(L;\beta,A^*) - g_3'\{\rho_0(R;\theta)\}\rho_0(R;\beta,A^*)I(R<\infty)}{g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)} \\
&\quad + \frac{\left[g_3\{\rho_0(L;\theta)\} - g_3\{\rho_0(R;\theta)\}I(R<\infty)\right] g_3\{\rho_0(L;\theta)\}\rho_0(L;\beta,A^*)}{\left[g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)\right]^2} \\
&\quad - \frac{\left[g_3\{\rho_0(L;\theta)\} - g_3\{\rho_0(R;\theta)\}I(R<\infty)\right] g_3\{\rho_0(R;\theta)\}\rho_0(R;\beta,A^*)I(R<\infty)}{\left[g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)\right]^2} \\
&= \left\{M_3(L,R;\theta) - M_4(L,R;\theta)\right\}\rho_0(L;\beta,A^*) - M_5(L,R;\theta)\rho_0(R;\beta,A^*),
\end{aligned}$$

where

$$\begin{aligned}
M_5(L,R;\theta) = &\frac{g_3'\{\rho_0(R;\theta)\}I(R<\infty)}{g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)} \\
&+ \frac{\left[g_3\{\rho_0(L;\theta)\} - g_3\{\rho_0(R;\theta)\}I(R<\infty)\right] g_3\{\rho_0(R;\theta)\}I(R<\infty)}{\left[g_2\{\rho_0(L;\theta)\} - g_2\{\rho_0(R;\theta)\}I(R<\infty)\right]^2}.
\end{aligned}$$

We calculate the following

$$\begin{aligned}
&\left.\frac{\partial \Phi_1(\beta,A+\eta A^*)}{\partial \eta}\right|_{\eta=0} \\
&= -\Delta g_1'\{\rho_0(T;\theta)\}\rho_1(T;\theta)\int_0^\tau I(t\le T)e^{\beta^\top Z(t)}X^\top(t)dA^*(t) \\
&\quad - \Delta g_1\{\rho_0(T;\theta)\}\int_0^\tau I(t\le T)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA^*(t) \\
&\quad + (1-\Delta)I(R<\infty)\left\{\rho_1(R;\theta) - \rho_1(L;\theta)\right\}M_3(L,R;\theta)\rho_0(L;\beta,A^*) \\
&\quad - (1-\Delta)I(R<\infty)\left\{\rho_1(R;\theta) - \rho_1(L;\theta)\right\}M_4(L,R;\theta)\rho_0(R;\beta,A^*) \\
&\quad + (1-\Delta)H_1(L,R;\theta)\int_0^\tau I(L<t\le R)I(R<\infty)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA^*(t) \\
&\quad - (1-\Delta)\rho_1(R^*;\theta)\left\{M_3(L,R;\theta) - M_4(L,R;\theta)\right\}\rho_0(L;\beta,A^*) \\
&\quad + (1-\Delta)\rho_1(R^*;\theta)M_5(L,R;\theta)\rho_0(R;\beta,A^*) \\
&\quad - (1-\Delta)H_2(L,R;\theta)\int_0^\tau I(t\le R^*)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA^*(t).
\end{aligned}$$

Hence,

$$
\begin{aligned}
B_3(t) = &-E\left[\Delta g_1'\{\rho_0(T;\theta_0)\}\rho_1(T;\theta_0)I(t \le T)e^{\beta_0^\top Z(t)}X^\top(t)\right] \\
&- E\left[\Delta g_1\{\rho_0(T;\theta_0)\}I(t \le T)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)\right] \\
&+ E\bigg[(1-\Delta)I(R<\infty)\left\{\rho_1(R;\theta_0)-\rho_1(L;\theta_0)\right\} \\
&\qquad \times M_3(L,R;\theta_0)I(t \le L)e^{\beta_0^\top Z(t)}X^\top(t)\bigg] \\
&- E\bigg[(1-\Delta)I(R<\infty)\left\{\rho_1(R;\theta_0)-\rho_1(L;\theta_0)\right\} \\
&\qquad \times M_4(L,R;\theta_0)I(t \le R)e^{\beta_0^\top Z(t)}X^\top(t)\bigg] \\
&+ E\left[(1-\Delta)H_1(L,R;\theta_0)I(L<t\le R<\infty)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)\right] \\
&- E\left[(1-\Delta)\rho_1(R^*;\theta_0)\left\{M_3(L,R;\theta_0)-M_4(L,R;\theta_0)\right\}I(t \le L)e^{\beta_0^\top Z(t)}X^\top(t)\right] \\
&+ E\left[(1-\Delta)\rho_1(R^*;\theta_0)M_5(L,R;\theta_0)I(t \le R)e^{\beta_0^\top Z(t)}X^\top(t)\right] \\
&- E\left[(1-\Delta)H_2(L,R;\theta_0)I(t \le R^*)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)\right].
\end{aligned}
\tag{B.27}
$$

Lastly, note that $\rho_3(t;\theta) = \sum_{j=1}^{q} \int_0^t e^{\beta^\top Z(s)} h_j(s) X_j(s) X^\top(s) dA(s)$.

$$
\begin{aligned}
&\left.\frac{\partial \Phi_2(\beta, A + \eta A^*)[h]}{\partial \eta}\right|_{\eta=0} \\
&= -\Delta g_1'\{\rho_0(T;\theta)\}\rho_3(T;\theta)[h] \int_0^\tau I(t \leq T) e^{\beta^\top Z(t)} X^\top(t) dA^*(t) \\
&- \Delta g_1\{\rho_0(T;\theta)\} \sum_{j=1}^{q} \int_0^\tau I(t \leq T) e^{\beta^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA^*(t) \\
&+ (1-\Delta) I(R < \infty) \{\rho_3(R;\theta)[h] - \rho_3(L;\theta)[h]\} M_3(L, R;\theta) \rho_0(L;\beta, A^*) \\
&\quad - (1-\Delta) I(R < \infty) \{\rho_3(R;\theta[h]) - \rho_3(L;\theta)[h]\} M_4(L, R;\theta) \rho_0(R;\beta, A^*) \\
&+ (1-\Delta) H_1(L, R;\theta) \sum_{j=1}^{q} \int_0^\tau I(L < t \leq R < \infty) e^{\beta^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA^*(t) \\
&- (1-\Delta) \rho_3(R^*;\theta)[h] \Big[\{M_3(L, R;\theta) - M_4(L, R;\theta)\} \rho_0(L;\beta, A^*)\Big] \\
&- (1-\Delta) \rho_3(R^*;\theta)[h] M_5(L, R;\theta) \rho_0(R;\beta, A^*) \\
&\quad - (1-\Delta) H_2(L, R;\theta) \sum_{j=1}^{q} \int_0^\tau I(t \leq R^*) e^{\beta^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA^*(t).
\end{aligned}
$$

Hence,

$$
\left.\frac{\partial P \Phi_2(\beta, A + \eta A^*)[h]}{\partial \eta}\right|_{\eta=0, A^*=A-A_0, \theta=\theta_0} = \int_0^\tau B_4[h](t) d(A - A_0)(t),
$$

where

$$
\begin{aligned}
&B_4[h](t) \\
&= -E\left[\Delta g_1'\{\rho_0(T;\theta_0)\}\rho_3(T;\theta_0)[h]I(t \le T)e^{\beta_0^\top Z(t)}X^\top(t)\right] \\
&- \sum_{j=1}^{q} E\left[\Delta g_1\{\rho_0(T;\theta_0)\}I(t \le T)h_j(t)e^{\beta_0^\top Z(t)}X_j(t)X^\top(t)\right] \\
&+ E\Bigg[(1-\Delta)I(R<\infty)\left\{\rho_3(R;\theta_0)[h] - \rho_3(L;\theta_0)[h]\right\} \\
&\times M_3(L,R;\theta_0)I(t \le L)e^{\beta_0^\top Z(t)}X^\top(t)\Bigg] \\
&- E\Bigg[(1-\Delta)I(R<\infty)\left\{\rho_3(R;\theta_0)[h] - \rho_3(L;\theta_0)[h]\right\} \\
&\times M_4(L,R;\theta_0)I(t \le R)e^{\beta_0^\top Z(t)}X^\top(t)\Bigg] \\
&+ \sum_{j=1}^{q} E\left[(1-\Delta)I(R<\infty)H_1(L,R;\theta_0)I(L<t\le R)e^{\beta_0^\top Z(t)}h_j(t)X_j(t)X^\top(t)\right] \\
&- E\left[(1-\Delta)\rho_3(R^*;\theta_0)[h]\left\{M_3(L,R;\theta_0) - M_4(L,R;\theta_0)\right\}I(t \le L)e^{\beta_0^\top Z(t)}X^\top(t)\right] \\
&+ E\left[(1-\Delta)\rho_3(R^*;\theta_0)[h]M_5(L,R;\theta_0)I(t \le R)e^{\beta_0^\top Z(t)}X^\top(t)\right] \\
&- \sum_{j=1}^{q} E\left[(1-\Delta)H_2(L,R;\theta_0)I(t \le R^*))e^{\beta_0^\top Z(t)}h_j(t)X_j(t)X^\top(t)\right].
\end{aligned}
\tag{B.28}
$$