Electronic Journal of Statistics
Vol. 19 (2025) 240–290
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2341

Regression analysis of semiparametric Cox-Aalen transformation models with partly interval-censored data

Xi Ning[1], Yanqing Sun[*2], Yinghao Pan[2], and Peter B. Gilbert[3,4]

[1] *Department of Statistics, Colby College,*
e-mail: xning@colby.edu

[2] *Department of Mathematics and Statistics, University of North Carolina at Charlotte,*
e-mail: yasun@charlotte.edu*;* ypan8@charlotte.edu

[3] *Department of Biostatistics, University of Washington,*

[4] *Vaccine and Infectious Disease and Public Health Sciences Divisions, Fred Hutchinson Cancer Center,*
e-mail: pgilbert@scharp.org

Abstract: Partly interval-censored data, comprising exact and interval-censored observations, are prevalent in biomedical, clinical, and epidemiological studies. This paper studies a flexible class of the semiparametric Cox-Aalen transformation models for regression analysis of such data. These models offer a versatile framework by accommodating both multiplicative and additive covariate effects and both constant and time-varying effects within a transformation, while also allowing for potentially time-dependent covariates. Moreover, this class of models includes many popular models such as the semiparametric transformation model, the Cox-Aalen model, the stratified Cox model, and the stratified proportional odds model as special cases. To facilitate efficient computation, we formulate a set of estimating equations and propose an Expectation-Solving (ES) algorithm that guarantees stability and rapid convergence. Under mild regularity assumptions, the resulting estimator is shown to be consistent and asymptotically normal. The validity of the weighted bootstrap is also established. A supremum test is proposed to test the time-varying covariate effects. Finally, the proposed method is evaluated through comprehensive simulations and applied to analyze data from a randomized HIV/AIDS trial.

MSC2020 subject classifications: Primary 62N02, 62N01.
Keywords and phrases: Cox-Aalen model, estimating equations, ES algorithm, hypothesis testing, partly interval-censored data, transformation models, time-dependent covariates, the ACTG 175 trial.

Received January 2024.

Contents

*Corresponding author: Yanqing Sun, e-mail: yasun@charlotte.edu

1. Introduction

Partly interval-censored data commonly arise in biomedical, clinical, and epidemiological studies. This type of data comprises a mixture of exact and interval-censored observations, with some failure times being exactly observed, whereas others are only known to fall within specific time intervals. Despite the influx of research on purely interval-censored data without exact observations (Finkelstein, 1986; Huang, 1996; Huang and Rossini, 1997; Cai and Betensky, 2003; Zeng, Cai and Shen, 2006; Sun, 2006; Wang et al., 2016), regression analysis of partly interval-censored data remains relatively underdeveloped, with a predominant focus on multiplicative hazards models; for instance, Kim (2003) considered maximum likelihood estimation, and Pan, Cai and Wang (2020) adopted a Bayesian approach by incorporating a nonhomogeneous Poisson process for the proportional hazards model (Cox, 1972; Andersen and Gill, 1982). In contrast to multiplicative hazards models, additive hazards models (Aalen, 1980, 1989; Lin and Ying, 1994) provide an important alternative by assuming an additive relationship between covariates and the baseline hazard function. Within this framework, Li and Ma (2019) proposed a maximum penalized likelihood method to analyze partly interval-censored data.

To enhance modeling capacity, researchers have also proposed models that combine both multiplicative and additive effects of covariates within a unified framework. Examples of such multiplicative-additive models include those proposed by Lin and Ying (1995), Martinussen and Scheike (2002), Scheike and Zhang (2002) and Zeng and Lin (2007). In particular, Scheike and Zhang (2002) introduced the Cox-Aalen model, which 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 readily accommodate time-varying covariates.

Recently, there has been increased attention towards transformation models. Zeng and Lin (2006) proposed a class of transformation models, which expands upon the linear transformation models (Dabrowska and Doksum, 1988; Fine, Ying and Wei, 1998) to account for time-varying covariates. In the sequel, we refer to this class of transformation models as the ZL model to avoid confusion. For interval-censored data, Zhang et al. (2005) proposed an estimating equation approach for linear transformation models, and Zeng, Mao and Lin (2016) developed a nonparametric maximum likelihood estimator for the ZL model using an EM algorithm. For partly interval-censored data, Wang, Jiang and Song (2022) proposed a Bayesian approach with a monotone spline approximation for linear transformation models, while Zhou, Sun and Gilbert (2021) studied nonparametric maximum likelihood estimation of the ZL model. It is worth noting that proportional hazards and proportional odds models are special cases of linear transformation and ZL models. Recently, Zhu et al. (2021) developed a maximum likelihood estimation procedure for fitting the proportional odds model to partly interval-censored data.

However, one limitation of the ZL model is its assumption of identical baseline hazard functions across individuals and that all covariate effects are multiplicative within the transformation function. Such an assumption is too restrictive in some applications. One example is the data analyses of the AIDS Clinical Trials Group 175 trial (Hammer et al., 1996; Zhou, Sun and Gilbert, 2021). This trial enrolled persons living with HIV-1 who received one of four antiretroviral regimens, with one regimen as the control and the remaining three regimens as the active treatment. The overall objective is to evaluate the association between antiretroviral treatment and the time to composite endpoint of CD4 cell count failure, AIDS, or death which is partly interval-censored. Participants were also categorized into two groups based on their prior antiretroviral therapy (ART) status: the ART-experienced and ART-naïve groups. Zhou, Sun and Gilbert (2021) considered two approaches: (1) fitting the ZL model to the full cohort, assuming a shared baseline hazard function for the ART-experienced and ART-naïve groups; (2) fitting the ZL model to the ART-experienced and ART-naïve groups separately. The first approach is questionable since participants in the ART-experienced group often develop drug resistance and thus have a different baseline hazard than the ART-naïve group. On the other hand, the second approach is valid but results in a loss of statistical power. It is desirable to develop new statistical methods for partly interval-censored data that can both incorporate heterogeneity and enhance statistical power.

In this paper, we study a flexible class of the Cox-Aalen transformation models of Ning et al. (2023) that allows the baseline hazard function to depend on covariates, thereby incorporating both multiplicative and additive covariate effects within a transformation. This flexibility enables us to capture more nuanced hazard structures and enhance statistical power. The proposed model includes many popular models, such as the ZL model, the Cox-Aalen model, the stratified Cox model and the stratified proportional odds model, as special cases. To the best of our knowledge, there is currently no research addressing the use of the Cox-Aalen transformation models for partly interval-censored data. To facilitate efficient computation, we formulate a set of estimating equations and employ an Expectation-Solving (ES) algorithm that guarantees stability and rapid convergence. In addition, we rigorously establish the asymptotic properties of the resulting estimators via modern empirical process techniques. The proposed estimator is a Z-estimator; therefore, we cannot directly rely on the consistency proofs provided in Zeng, Mao and Lin (2016) and Zhou, Sun and Gilbert (2021), as those proofs are specifically tailored to their M-estimators. Instead, we utilize the implicit function theorem (Schwartz, 1969, p. 15) to establish the consistency results. This theoretical development is interesting in its own right, requiring a delicate connection between the proposed estimator and a nonparametric maximum likelihood estimator, which constitutes an original contribution. A supremum test is proposed to test the time-varying covariate effects. Finally, we evaluate the performance of the proposed procedure through comprehensive simulation studies and apply it to the ACTG 175 trial.

2. Cox-Aalen transformation models

Let T be the failure time of interest, and $X(\cdot)$ and $Z(\cdot)$ be a $q\times 1$ and $d\times 1$ vector of potentially time-dependent covariates, respectively. We investigate a broad class of the Cox-Aalen transformation models which structures the cumulative hazard function for T, conditional on $X(\cdot)$ and $Z(\cdot)$, as:

$$\Lambda(t \mid X(\cdot), Z(\cdot)) = G\Big[\int_0^t \exp\{\beta^\top Z(s)\} d\Lambda_X(s)\Big], \tag{1}$$

where β is a $d\times 1$ vector of unknown regression parameters, $\Lambda_X(s) = \int_0^s X^\top(v)\alpha(v)\,dv$ is an unspecified increasing function which depends on $X(\cdot)$ and a vector of unknown regression functions $\alpha(t) = (\alpha_1(t), \dots, \alpha_q(t))^\top$, and $G(\cdot)$ is a predetermined transformation function that is strictly increasing and thrice continuously differentiable with $G(0) = 0$, $G'(0) > 0$ and $G(\infty) = \infty$. Hereafter, $G'(x) = dG(x)/dx$. Let $A(t) = \int_0^t \alpha(s)ds = (A_1(t), \dots, A_q(t))^\top$ be a vector of cumulative regression functions with $A_j(t) = \int_0^t \alpha_j(s)ds$ $(j = 1, \dots, q)$. In addition, the first component of $X(\cdot)$ is fixed at 1 so that $\alpha_1(t)$ can be interpreted as a reference level of the risk.

Model (1) has two main advantages due to its flexibility and generality. First, it accommodates both multiplicative and additive covariate effects within a transformation, i.e., the parametric component β quantifies the multiplicative effect of $Z(\cdot)$ while the nonparametric component $\alpha(\cdot)$ tracks the additive effect of $X(\cdot)$. As a result, the proposed model allows for different baseline cumulative hazards across individuals or groups. Note that, to ensure identifiability, $X(\cdot)$ and $Z(\cdot)$ should not overlap, i.e., for a single covariate, its effect is either multiplicative or additive, but not both. Second, it offers a wide range of flexible models, encompassing several popular models as special cases. For instance, when $G(x) = x$, the right-hand side of the model (1) simplifies to

$$\int_0^t \exp\{\beta^\top Z(s)\} X^\top(s) dA(s), \tag{2}$$

which indicates that the Cox-Aalen model (Scheike and Zhang, 2002) is a special case of proposed model. When X represents a vector designed for categories, model (2) further reduces to the stratified Cox model (Kalbfleisch and Prentice, 2002). Additionally, in the case where $X = 1$, i.e., $q = 1$, model (1) reduces to the ZL model (Zeng and Lin, 2006), where the cumulative hazard function takes the form

$$G\Big[\int_0^t \exp\{\beta^\top Z(s)\} dA_1(s)\Big]. \tag{3}$$

It is noteworthy that the choices $G(x) = x$ and $G(x) = \log(1 + x)$ in model (3) yield the proportional hazards model and proportional odds model, respectively.

Following the previous work of Zeng, Mao and Lin (2016) and Zhou, Sun and Gilbert (2021), we consider a class of frailty-induced transformation functions as

$$G(x) = -\log \int_0^\infty \exp(-x\xi) f(\xi) d\xi, \tag{4}$$

where $f(\xi)$ is the density function of a nonnegative random variable ξ on support $[0, \infty)$. A widely used choice of $f(\xi)$ is the gamma density with mean 1 and variance r, which generates the logarithmic transformations $G(x) = r^{-1} \log(1 + rx)$ $(r \geq 0)$. Notably, if $r = 0$, then $G(x) = x$. Another popular choice, i.e., the positive stable distribution with parameter ρ $(0 \leq \rho < 1)$, yields the class of Box-Cox transformations, $G(x) = \rho^{-1}\{(1 + x)^\rho - 1\}$. The incorporation of (4) into the ZL model has been widely employed across diverse contexts, driven by the development of EM algorithms (Liu and Zeng, 2013; Mao and Lin, 2017; Gao, Zeng and Lin, 2018).

3. Model estimation

3.1. Data and likelihood

Consider a random sample of n individuals subject to partly interval censoring. For the ith individual, we denote the failure time as T_i. If T_i is observed exactly, we set $\Delta_i = 1$. Otherwise, we set $\Delta_i = 0$, and denote the sequence of examination times undergone by 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. We set $U_{i0} = 0$ and $U_{i,K_i+1} = \infty$. In addition, for $\Delta_i = 0$, we define $(L_i, R_i]$ to be the smallest interval 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 $U_{ik} < T_i$. Similarly, the right endpoint R_i is the minimum value of U_{ik} for $k = 1, \ldots, K_i + 1$ such that $U_{ik} \geq T_i$. Note $L_i = 0$ and $R_i = \infty$ correspond to a left- and right-censored observation, respectively, while $0 < L_i < R_i < \infty$ indicates a typical interval-censored observation. Therefore, the observed data can be written as follows:

$$\mathcal{O}_i = \{\Delta_i, \Delta_i T_i, (1 - \Delta_i)L_i, (1 - \Delta_i)R_i, X_i, Z_i\} \quad (i = 1, \ldots, n), \tag{5}$$

where X_i and Z_i denote the covariates for the ith individual.

Suppose that the total number and the sequence of examination times are independent of the failure time conditional on the covariate histories. Under model (1), the observed-data likelihood function based on (5) takes the form

$$\begin{aligned}\prod_{i=1}^{n} \Big(&\Lambda'_{X_i}(T_i) e^{\beta^\top Z_i(T_i)} G'\Big\{ \int_0^{T_i} e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\Big\} \\ &\exp\Big[-G\Big\{\int_0^{T_i} e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\Big\}\Big]\Big)^{\Delta_i} \\ \Big(\exp\Big[-G\Big\{\int_0^{L_i} &e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\Big\}\Big] - \exp\Big[-G\Big\{\int_0^{R_i} e^{\beta^\top Z_i(s)} d\Lambda_{X_i}(s)\Big\}\Big]\Big)^{1-\Delta_i}\end{aligned} \tag{6}$$

where $\Lambda'_X(\cdot)$ and $G'(\cdot)$ denote the derivatives of $\Lambda_X(\cdot)$ and $G(\cdot)$, respectively. To motivate our approach, we start by considering the nonparametric maximum likelihood estimation of β and $\Lambda_X(\cdot)$, and then develop an EM algorithm tailored to a special case where X is designed for categorical data. To facilitate efficient computation, we propose an alternative estimating equation approach that is applicable to a broader range of scenarios. Intriguingly, these two approaches coincide in the aforementioned special case, as we will demonstrate later.

3.2. Nonparametric maximum likelihood estimation

In this subsection, we employ the nonparametric maximum likelihood estimation (NPMLE) method. Specifically, let $0 = t_0 < t_1 < \cdots < t_m < \infty$ denote the ordered unique values in the collection of $\{\Delta_i T_i, (1 - \Delta_i)L_i, (1 - \Delta_i)R_i I(R_i < \infty), i = 1, \ldots, n\}$. 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 $(k = 1, \ldots, m)$ with $a_{j0} = 0$. It is noted that $d\Lambda_{X_i}(t) = X_i^\top(t)dA(t)$, which implies that $\Lambda_{X_i}(t)$ is also a step function with jump size $X_i^\top(t_k)a_k$ at time t_k, and $\Lambda_{X_i}(0) = 0$. Here, $a_k = (a_{1k}, \ldots, a_{qk})^\top (k = 1, \ldots, m)$. Thus, the likelihood given in (6) can be written as

$$\prod_{i=1}^{n}\Big(\Lambda_{X_i}\{T_i\}e^{\beta^\top Z_i(T_i)}G'\Big\{\sum_{t_k\le T_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\Big\}$$
$$\exp\Big[-G\Big\{\sum_{t_k\le T_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\Big\}\Big]\Big)^{\Delta_i}$$
$$\Big(\exp\Big[-G\Big\{\sum_{t_k\le L_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\Big\}\Big]-\exp\Big[-G\Big\{\sum_{t_k\le R_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\Big\}\Big]\Big)^{1-\Delta_i}, \tag{7}$$

where $\Lambda_{X_i}\{T_i\}$ denotes the jump size of $\Lambda_{X_i}(\cdot)$ at T_i, $X_{ik}=X_i(t_k)$ and $Z_{ik}=Z_i(t_k)$.

Let $\theta=(\beta^\top,a_1^\top,\ldots,a_m^\top)^\top$ be the parameter of interest. Maximizing the likelihood function (7) with respect to θ based on the observed data (5) can be a daunting task due to the high dimensionality of the parameter space. Consequently, performing this maximization directly may not be feasible or practical. Based on the class of frailty-induced transformation functions given in (4), we easily obtain that $\exp\{-G(x)\}=\int_0^\infty \exp(-x\xi)f(\xi)d\xi$ and $\exp\{-G(x)\}G'(x)=\int_0^\infty \xi\exp(-x\xi)f(\xi)d\xi$. By utilizing these expressions, it can be shown that the likelihood function (7) is equivalent to

$$\prod_{i=1}^{n}\Big[\Lambda_{X_i}\{T_i\}e^{\beta^\top Z_i(T_i)}\int_{\xi_i}\xi_i\exp\Big\{-\xi_i\sum_{t_k\le T_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\Big\}f(\xi_i)d\xi_i\Big]^{\Delta_i}$$
$$\Big(\int_{\xi_i}\Big[\exp\Big\{-\xi_i\sum_{t_k\le L_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\Big\} \tag{8}$$
$$-\exp\Big\{-\xi_i\sum_{t_k\le R_i}(X_{ik}^\top a_k)e^{\beta^\top Z_{ik}}\Big\}\Big]f(\xi_i)d\xi_i\Big)^{1-\Delta_i}.$$

Next, we introduce latent variables W_{ik} $(i=1,\ldots,n;k=1,\ldots,m)$ which, conditional on ξ_i, are a class of independent Poisson variables with means $\xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})$. This approach is similar to those used in previous works, including Wang et al. (2016), Zeng, Mao and Lin (2016), and Zhou, Sun and Gilbert (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 $D_i=(1-\Delta_i)I(R_i<\infty)\sum_{L_i<t_k\le R_i}W_{ik}$ $(i=1,\ldots,n)$. Suppose that the observed data for individual i $(i=1,\ldots,n)$ consist of

$$\begin{cases}(T_i,X_i,Z_i,A_i=0,B_i=1) & \text{if } \Delta_i=1,\\ (L_i,R_i,X_i,Z_i,C_i=0,D_i>0) & \text{if } \Delta_i=0.\end{cases} \tag{9}$$

Notice that when $\Delta_i=1$, $A_i=0$ and $B_i=1$ indicate that $W_{ik}=0$ for $t_k<T_i$ and $W_{ik}=1$ for $t_k=T_i$. In addition, when $\Delta_i=0$, $C_i=0$ and $D_i>0$ imply that $W_{ik}=0$ for $t_k\le L_i$ and at least one $W_{ik}\ge 1$ for $L_i<t_k\le R_i$ with $R_i<\infty$. Using the independent properties of W_{ik} $(i=1,\ldots,n;k=1,\ldots,m)$, we calculate that $\Pr(A_i = 0 \mid \xi_i) = \exp\big\{-\xi_i \sum_{t_k<T_i}(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\big\}$. Applying this similar idea, the likelihood (8) can be represented with the data in (9) as

$$\prod_{i=1}^{n}\Big\{\int_{\xi_i}\Pr\Big(\sum_{t_k<T_i}W_{ik}=0\mid\xi_i\Big)\Pr\Big(\sum_{t_k=T_i}W_{ik}=1\mid\xi_i\Big)f(\xi_i)d\xi_i\Big\}^{\Delta_i}$$
$$\Big[\int_{\xi_i}\Pr\Big(\sum_{t_k\le L_i}W_{ik}{=}0\mid\xi_i\Big)\Big\{1{-}\Pr\Big(\sum_{L_i<t_k\le R_i}W_{ik}{=}0\mid\xi_i\Big)\Big\}^{I(R_i<\infty)}f(\xi_i)d\xi_i\Big]^{1-\Delta_i}. \tag{10}$$

Therefore, maximizing the likelihood function (7) based on the observed data in (5) is equivalent to maximizing the likelihood function (10) based on the data in (9).

We propose to maximize (10) through an EM algorithm by treating W_{ik} and ξ_i as missing data. The complete-data loglikelihood is given by

$$\begin{aligned}\sum_{i=1}^{n}\Big(\sum_{k=1}^{m}I(t_k\le R_i^*)\Big[&W_{ik}\log\{\xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\}\\ &-\xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})-\log W_{ik}!\Big]+\log f(\xi_i)\Big),\end{aligned} \tag{11}$$

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, we calculate the conditional expectation of the complete-data loglikelihood (11), given the observed data. This is equivalent to evaluating the posterior means of W_{ik} and ξ_i, denoted by $\hat{E}(W_{ik})$ and $\hat{E}(\xi_i)$, respectively. See the next subsection for details. In the M-step, we maximize the conditional expectation of (11) with respect to θ. Specifically, we set the derivatives of the conditional expectation of (11) with respect to a_k $(k=1,\dots,m)$ and β to zero, respectively, i.e.,

$$\sum_{i=1}^{n}I(t_k\le R_i^*)\Big\{\frac{\hat{E}(W_{ik})}{X_{ik}^\top a_k}-\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\Big\}X_{ik}=0 \quad (k=1,\dots,m), \tag{12}$$
$$\sum_{i=1}^{n}\sum_{k=1}^{m}I(t_k\le R_i^*)\Big\{\hat{E}(W_{ik})-\hat{E}(\xi_i)(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\Big\}Z_{ik}=0. \tag{13}$$

However, the dimension of θ could potentially match or exceed the sample size n due to the nature of the partly interval-censored data. As a result, solving equations (12) and (13) simultaneously becomes difficult and computationally intensive, as they constitute a large system of nonlinear equations. In the Appendix A.1, we demonstrate that explicit formulae can be derived for a_k $(k=1,\dots,m)$ for fixed β in (12), when X represents levels in a set of factors. Once these high-dimensional parameters are fixed in (13), the low-dimensional parameter β can be solved using any root-finding algorithm, such as the Newton-Raphson method. We iterate between E- and M-steps until the convergence criterion is achieved. Nevertheless, for more general scenarios, explicit forms for a_k $(k = 1, \ldots, m)$ are not readily available, and hence computational challenges still hinder the implementation of the EM algorithm. To overcome this, we adopt an estimating equation approach and develop an ES algorithm for simple computations.

3.3. Estimating equations

By exploiting the fact that W_{ik} $(i = 1, \ldots, n; k = 1, \ldots, m)$ conditional on ξ_i are independent Poisson random variables with mean $\xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})$, we construct a collection of complete-data estimating equations $U(\theta) = (U_{a_1}, \ldots, U_{a_m}, U_\beta) = 0$, where

$$\begin{cases} U_{a_1} = \sum_{i=1}^n I(t_1 \le R_i^*)\left\{W_{i1} - \xi_i(X_{i1}^\top a_1)\exp(\beta^\top Z_{i1})\right\}X_{i1} \\ \vdots \\ U_{a_m} = \sum_{i=1}^n I(t_m \le R_i^*)\left\{W_{im} - \xi_i(X_{im}^\top a_m)\exp(\beta^\top Z_{im})\right\}X_{im} \\ U_\beta = \sum_{i=1}^n \sum_{k=1}^m I(t_k \le R_i^*)\left\{W_{ik} - \xi_i(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\right\}Z_{ik}. \end{cases} \tag{14}$$

Conditional expectation arguments easily establish that (14) is a system of unbiased estimating equations.

We propose to estimate θ through an ES algorithm by treating W_{ik} and ξ_i as missing. Within the E-step, we compute the posterior means of W_{ik} and ξ_i given the observed data. Within the S-step, we solve the estimating equations (14) after replacing W_{ik} and ξ_i by corresponding conditional expectations. The ES algorithm is an extension of the EM algorithm, allowing us to handle general estimating equations beyond those derived from the loglikelihood (Elashoff and Ryan, 2004). The detailed calculations are outlined below:

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 $G''(x)$ is the second derivative of $G(\cdot)$ with respect to x. When $\Delta_i = 0$, it is easy to see that 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, $\hat{E}(W_{ik}) = E(W_{ik}|C_i = 0, D_i > 0, X_i, Z_i) = 0$, for $t_k \le L_i$. And for $L_i < t_k \le R_i$ with $R_i < \infty$,

$$\hat{E}(W_{ik}) \quad = \quad E_{\xi_i}\left\{E(W_{ik}|\xi_i, C_i = 0, D_i > 0) \mid C_i = 0, D_i > 0\right\}$$

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

S-step. Solve for θ using (14) with W_{ik} and ξ_i replaced by $\hat{E}(W_{ik})$ and $\hat{E}(\xi_i)$. Note that (14) is a large-dimensional nonlinear equation, which is not easily solved simultaneously. To circumvent this, we propose the following nonlinear Gauss-Seidel method (Ortega and Rheinboldt, 1970; Ortega, 1972). In Step 1, we fix β and update a_k $(k=1,\dots,m)$ by solving

$$
\begin{cases}
\sum_{i=1}^n I(t_1\le R_i^*)\left\{\hat{E}(W_{i1})-\hat{E}(\xi_i)(X_{i1}^\top a_1)\exp(\beta^\top Z_{i1})\right\}X_{i1}=0\\
\vdots\\
\sum_{i=1}^n I(t_m\le R_i^*)\left\{\hat{E}(W_{im})-\hat{E}(\xi_i)(X_{im}^\top a_m)\exp(\beta^\top Z_{im})\right\}X_{im}=0.
\end{cases}
\tag{15}
$$

It is important to note that for fixed β, the above system of equations is linear with respect to a_k $(k=1,\dots,m)$. We can therefore obtain

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

In Step 2, with $a_1,\dots,a_m$ fixed, we update β by solving the following equation via the Newton-Raphson method:

$$
\sum_{i=1}^n\sum_{k=1}^m I(t_k\le R_i^*)\left\{\hat{E}(W_{ik})-\hat{E}(\xi_i)(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\right\}Z_{ik}=0. \tag{16}
$$

The proposed ES algorithm iterates between the E- and S-steps until convergence. We compute the maximal relative change in the parameter estimates between two successive iterations to determine convergence, and the ES algorithm terminates when this value is below a small threshold, such as 5×10^{-3}. Additionally, the S-step alternates between Steps 1 and 2 until the sum of the absolute differences of the estimates at two successive iterations is less than a small positive number, such as 10^{-3}. Other small numbers were also tested, and similar results were obtained. We denote the final estimator as $\hat{\theta}=(\hat{\beta}^\top,\hat{a}_1^\top,\dots,\hat{a}_m^\top)^\top$. For estimating $A(t)$, a natural choice is the estimator $\hat{A}(t)=\sum_{t_k\le t}\hat{a}_k$.

There are several advancements in the proposed ES algorithm. First, it derives the closed-form expressions for the posterior means of W_{ik} and ξ_i in the E-step, and the high-dimensional parameters a_k $(k = 1, \dots, m)$ are updated explicitly in the S-step. These features make the algorithm highly versatile and applicable in various scenarios. Second, the resulting estimator is efficient when X is a vector of design variables for categories. This efficiency arises because, for fixed β, equations (12) and (15) share the same solution in terms of a_k $(k = 1, \dots, m)$: hence the ES algorithm coincides with the EM algorithm in Section 3.2. The details are provided in the Appendices. Similarly, when $X = 1$, i.e., $q = 1$, it can be shown that the proposed ES algorithm is equivalent to the EM algorithm proposed by Zhou, Sun and Gilbert (2021) for partly interval-censored data. Furthermore, when $G(x) = x$, model (1) yields the Cox-Aalen model. To the best of our knowledge, our methods provide a solution for estimating the Cox-Aalen model from partly interval-censored data with time-dependent covariates, thereby filling a gap in the existing literature. Lastly, it is worth mentioning that (15) can be interpreted as a weighted version of (12), where each subject i receives a weight $X_{ik}^\top a_k$.

3.4. Variance estimation

We propose a weighted bootstrap procedure to estimate the distribution and the variance of the proposed ES estimator. Let $e_1, \dots, e_n$ be i.i.d exponential random variables with mean one, which are independent of the observed data $\mathcal{O} = (\mathcal{O}_1, \dots, \mathcal{O}_n)$. Let $\overline{e} = n^{-1} \sum_{i=1}^n e_i$ and $\widetilde{e}_i = e_i / \overline{e}$. In addition, let $\widetilde{U}(\theta)$ be the weighted version of $U(\theta)$, where each subject i $(i = 1, \dots, n)$ in (14) receives weight $\widetilde{e}_i$. The final estimator that solves $\widetilde{U}(\theta) = 0$, denoted as $\widetilde{\theta} = (\widetilde{\beta}^\top, \widetilde{a}_1^\top, \dots, \widetilde{a}_m^\top)^\top$, can be obtained through the proposed ES algorithm in Section 3.3 with only trivial modifications. Specifically, in the E-step, we compute the posterior means of W_{ik} and ξ_i $(i = 1, \dots, n; k = 1, \dots, m)$, denoted as $\widetilde{E}(W_{ik})$ and $\widetilde{E}(\xi_i)$, respectively. These expressions are the same as $\hat{E}(W_{ik})$ and $\hat{E}(\xi_i)$ described in Section 3.3. In the S-step, with fixed β, we update a_k $(k = 1, \dots, m)$ using the explicit formula:

$$a_k = \Big\{ \sum_{i=1}^n I(t_k \le R_i^*) \widetilde{e}_i \widetilde{E}(\xi_i) \exp(\beta^\top Z_{ik}) X_{ik} X_{ik}^\top \Big\}^{-1} \Big\{ \sum_{i=1}^n I(t_k \le R_i^*) \widetilde{e}_i \widetilde{E}(W_{ik}) X_{ik} \Big\}.$$

Then, for fixed $a_1, \dots, a_m$, we solve for β using the Newton-Raphson method through the following equation:

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

One can generate a set of bootstrap weights for each bootstrap replicate and run the revised ES algorithm to obtain $\widetilde{\theta}$. The distribution of the ES estimator, and its variance in particular, can be estimated by, say, 1000 bootstrap replicates. The covariance matrix of $\hat{\theta}$ can be estimated by the sample variance of those $\widetilde{\theta}$'s.

4. Asymptotic theory

We establish the asymptotic properties of the proposed ES estimator and the validity of the weighted bootstrap 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 $[0,\tau]$. Here, τ denotes the duration of participant follow-up in the study, which is finite.*

Condition 2. *Let $\mathcal{B}$ be a compact set of $\mathbb{R}^d$ 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. $0 < Pr(\Delta = 0) < 1$. *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 $pr(U_{j+1} - U_j \geq c \mid K, X, Z, \Delta = 0) = 1$ $(j = 1, \ldots, K-1)$. The conditional densities of (U_j, U_{j+1}) given X, Z and K, denoted by $g_j(u, v|X, Z, K)$ $(j = 1, \ldots, K)$, have continuous second-order partial derivatives with respect to u and v when $v - u > c$ and are continuously differentiable with respect to X and Z.*

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 $\eta_0(t) + \eta^\top X(t) = 0$ for all $t \in [0,\tau]$ with probability one, then $\eta_0(t) = 0$ for all $t \in [0,\tau]$ and $\eta = 0$.*

Remark 1. *Conditions 1–2 state the boundedness of the covariates and the compactness of the Euclidean parameter space, which are standard in survival analysis. Condition 3 is a conventional assumption for interval-censored data, which requires that the two adjacent monitoring times are 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 a_k $(k = 1, \ldots, m)$.*

The parameter of interest is $\vartheta = (\beta, A)$, where $\beta \in \mathbb{R}^d$ and $A = (A_1, \ldots, A_q)$ consist 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 the absolute value of $A_j(0)$ and the total variation of A_j on $[0,\tau]$. Let $\mathcal{F} = \mathcal{B} \times BV^q[0,\tau]$. The norm for $\vartheta \in \mathcal{F}$ is defined as $\|\vartheta\|_\Theta = \|\beta\|_d + \|A\|_\rho$, where $\|\cdot\|_d$ is the Euclidean norm: $\|\beta\|_d = \sqrt{\sum_{j=1}^d \beta_j^2}$.

Let $\hat{\vartheta} = (\hat{\beta}, \hat{A})$ be the proposed ES estimator and $\widetilde{\vartheta} = (\widetilde{\beta}, \widetilde{A})$ the bootstrap estimator, where $\widetilde{A}(t) = \sum_{t_k \leq t} \widetilde{a}_k$ for $t \in [0,\tau]$. Let $\vartheta_0 = (\beta_0, A_0)$ be the true value of (β_0, A_0) under model (1). The asymptotic results of the proposed estimator are given in the following theorems.

Theorem 4.1. *Under Conditions 1–5, the proposed estimator* $(\hat{\beta}, \hat{A})$ *is strongly consistent to* (β_0, A_0) *in* $\mathcal{B} \times BV^q[0, \tau]$.

Theorem 4.2. *Under Conditions 1–5,* $n^{1/2}(\hat{\beta} - \beta_0, \hat{A} - A_0)$ *converges weakly to a zero-mean Gaussian process in* $\mathcal{B} \times BV^q[0, \tau]$.

Theorem 4.3. *Under Conditions 1–5, the conditional distribution of* $n^{1/2}(\widetilde{\beta} - \hat{\beta}, \widetilde{A} - \hat{A})$ *given the data converges weakly to the asymptotic distribution of* $n^{1/2}(\hat{\beta} - \beta_0, \hat{A} - A_0)$ *in* $\mathcal{B} \times BV^q[0, \tau]$.

Detailed proofs of the above theorems are presented in the Appendices.

Remark 2. *Note that the proposed estimator* $(\hat{\beta}, \hat{A})$ *is a Z-estimator. To prove Theorem 4.1, we cannot directly rely on the consistency proofs provided in Zeng, Mao and Lin (2016) and Zhou, Sun and Gilbert (2021), as those proofs are specifically tailored to their M-estimators. Instead, we utilize the implicit function theorem (Schwartz, 1969, p. 15) to establish the consistency results. One major challenge in proving Theorems 4.1–4.2 is verifying that the Fréchet derivative map corresponding to the proposed ES estimator is continuously invertible. By uncovering a connection between the proposed ES estimator and the NPMLE, we are able to address this challenge. See Lemma A.3 of the Appendices for details. The involved theoretical development is interesting in its own right.*

5. Model inference

Statistical inference for covariate effects is important for data applications. Hypothesis testing for the parametric effects can usually be conducted using Wald tests or Chi-square tests by contrasting the estimated effects with their estimated covariance matrix. However, hypothesis testing of the time-dependent effects requires additional work. We propose a supremum test procedure to test the regression function $A_j(t)$ under the null hypothesis $H_0 : A_j(t) \equiv 0$, $0 \le t \le \tau$, where $1 \le j \le q$. Testing of this hypothesis can be used to determine whether the baseline cumulative hazards vary between groups. We consider the supremum test statistic $S = \sup_{0 \le t \le \tau} |\sqrt{n}\hat{A}_j(t)|$. By Theorem 4.3, the critical value of the test can be approximated by using the weighted bootstrap procedure for large samples. Specifically, the critical value of the test at significance level α can be estimated by the $(1-\alpha)$ quantile of the bootstrap values of $S^* = \sup_{0 \le t \le \tau} |\sqrt{n}(\widetilde{A}_j(t) - \hat{A}_j(t))|$ based on, say 1000, weighted bootstrap samples. The null hypothesis is rejected if the test statistic S exceeds the critical value. This test procedure is readily adapted to test multiple regression functions.

6. Simulation studies

We conducted extensive simulations to access the finite sample performance of the proposed estimator. We generate T from the following Cox-Aalen transformation model

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

where $Z_1(t) = B_1 I(t \leq V) + B_2 I(t > V)$ with B_1 and B_2 being independent Ber(0.5), and $V \sim$ Unif(0, 3), $Z_2 \sim$ Unif(0, 1). We consider 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. We set $\beta_1 = 0.5$ and $\beta_2 = -0.5$, and consider three numerical settings 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 $X_2 \sim$ Ber(0.4), $A_1(t) = \log(1 + t/2)$ and $A_2(t) = 0.1t$.

Scenario 2. $X = (1, X_2)^\top$ with $X_2 \sim$ 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. $X = (1, X_2, X_3)^\top$, where $X_2 = I(J = 2)$, $X_3 = I(J = 3)$, $A_1(t) = \log(1 + t/2)$, $A_2(t) = 0.1t$ and $A_3(t) = 0.05t$.

We let $\tau = 5$ years be the duration of study follow-up, beyond which no examinations occurred. For each study participant, we generate at least two monitoring times $U_1 \sim$ Unif $(0, \tau/2)$ and $U_2 \sim \min\{0.1 + U_1 + \text{Unif}(0, \tau/2), \tau\}$. If $U_2 < \tau$, we proceed to generate 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 and at most five intervals. We let $(L, R]$ be the smallest interval that brackets the failure time T. In particular, if $R = \infty$, we set $\Delta = 0$. Otherwise, we generate $\Delta \sim$ Ber (γ). If $\Delta = 1$, the failure time is exactly observed. Additionally, γ is the proportion of exactly observed failure observations among those that are not right-censored. We consider four values for γ: 0.25, 0.5, 0.75, and 1. The value $\gamma = 1$ corresponds to purely right-censored data, while the other values simulate partially interval-censored data, including exact, left-, interval-, and right-censored observations. The right-censoring rates, on average, range from 25% to 45% across all setups. For each scenario, we applied the proposed ES algorithm by setting the initial value of β to 0 and the initial value of a_k $(k = 1, \ldots, m)$ to $(1/m, 0, \ldots, 0)$. The estimates were found to be robust with respect to the choice of the initial values. The variance was estimated using 1000 weighted bootstraps as described in Section 3.4. The sample size n was set to 200, 500, and 1000. We performed 1000 replicates to obtain reliable results.

The performance of $\hat{\beta}$ and $\hat{A}(t)$ at fixed points t is measured through the bias (Bias), empirical standard error of the estimator (SE), mean of the estimated standard errors (SEE), and the 95% empirical coverage probability (CP). Table 1 summarizes the parameter estimation results for Scenario 1. The proposed methods perform well across all values of γ. This is evidenced by several facts: the parameter estimators exhibit minimal bias, their bootstrapping variance estimators are quite close to the empirical variance, and the confidence intervals

Table 1

Estimation results for the regression parameter β under Scenario 1. Bias, SE, SEE, and CP stand, respectively, for the bias, empirical standard error, mean of the estimated standard errors, and empirical coverage probability of the 95% confidence interval. Each entry is based on 1000 simulations and 500 bootstraps.

			$\beta_1 = 0.5$				$\beta_2 = -0.5$			
r	n	γ	Bias	SE	SEE	CP	Bias	SE	SEE	CP
0	200	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	0.25	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	0.25	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
0.5	200	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	0.25	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	0.25	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
1	200	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	0.25	−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	0.25	−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

have proper coverage probabilities. As expected, the variance estimates tend to decrease as the sample size increases and the proportion of exact observations increases. Additionally, we display the estimation results for $A_1(\cdot)$ and $A_2(\cdot)$ in Figure 1, which further confirms the reliability of the proposed estimation procedures. The estimation results for Scenarios 2–3 can be found in Appendix A.4.

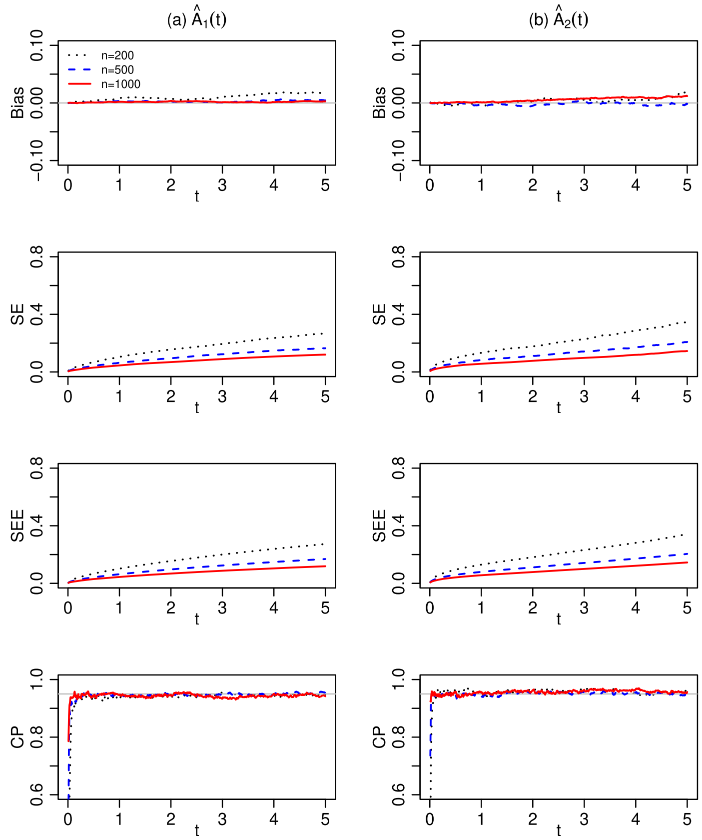

FIG 1. *Estimation results for (a)* $A_1(t) = \log(1 + t/2)$ *and (b)* $A_2(t) = 0.1t$ *in Scenario 1 with* $\gamma = 0.5$*, under the logarithmic transformation* $G(x) = r^{-1}\log(1 + rx)$ *with* $r = 0$*. The dotted, dashed and solid lines are for data sets with* $n = 200, 500, 1000$*, respectively. Bias, SE, SEE, and CP stand, respectively, for the bias, empirical standard error, mean of the estimated standard errors, and empirical coverage probability of the* 95% *confidence interval. The figures are based on 1000 simulations and 1000 bootstraps.*

The finite sample performance of the supremum test is examined under Scenario 1, except we set $A_2(t) = \kappa \cdot 0.1t$ with κ being a constant. We considered testing $A_2(t) = 0$. To examine the size of the test, we generated the data under the null model with $A_2(t) = 0$ ($\kappa = 0$). To examine the power of the test, we generated the data under the alternative models with $A_2(t) \neq 0$, using $\kappa = 0.5$, 1 and 1.5. We reject the null hypothesis at significance level 0.05 if the test statistic $S = \sup_{0\leq t\leq\tau} |\sqrt{n}\hat{A}_2(t)|$ is greater than 0.95 quantile of 500 bootstrap values of $S^* = \sup_{0\leq t\leq\tau} |\sqrt{n}(\widetilde{A}_2(t) - \hat{A}_2(t))|$. The size and power of the test is the percentage of rejection in 1000 simulations under the null model and alternative models, respectively. Table 2 shows satisfying finite sample performance. The size of the test is approximately 0.05, the model ($\kappa = 0$) with a larger r ($r = 2$) may require a larger sample size for better convergence. The power

Table 2

Simulation results for the size and power of the supremum test at significance level 0.05 under a similar setting as Scenario 1 except for $A_2(t) = \kappa \cdot 0.1t$, where $\kappa = 0$ is for the null model while $\kappa = 0.5$, 1 and 1.5 correspond to the alternatives. Each entry is based on 1000 simulations and 500 bootstraps.

n	κ	$r = 0$	$r = 1$	$r = 2$
500	0	0.047	0.024	0.017
	0.5	0.268	0.069	0.019
	1	0.746	0.234	0.063
	1.5	0.966	0.473	0.138
1000	0	0.053	0.045	0.029
	0.5	0.562	0.217	0.109
	1	0.978	0.626	0.333
	1.5	1.000	0.891	0.623

Table 3

Simulation results for estimation of the regression parameters with a misspecified $r = 0$ while the data is generated from r_{true} value for Scenario 1, under the logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$. Here, r_{true} can be any value from {0, 0.5, 1, 1.5, 2, 2.5, 3}. Bias, SE, SEE, and CP stand, respectively, for the bias, empirical standard error, mean of the estimated standard errors, and empirical coverage probability of the 95% confidence interval. Each entry is based on 1000 simulations and 1000 bootstraps.

			$\beta_1 = 0.5$				$\beta_2 = -0.5$			
γ	n	r_{true}	Bias	SE	SEE	CP	Bias	SE	SEE	CP
0.5	500	0	0.001	0.110	0.111	0.950	0.004	0.177	0.183	0.954
		0.5	−0.067	0.115	0.118	0.912	0.097	0.193	0.195	0.918
		1	−0.112	0.122	0.125	0.863	0.151	0.207	0.206	0.883
		1.5	−0.141	0.124	0.131	0.820	0.183	0.213	0.217	0.865
		2	−0.160	0.131	0.137	0.789	0.207	0.225	0.225	0.843
		2.5	−0.179	0.139	0.142	0.755	0.223	0.231	0.233	0.832
		3	−0.193	0.148	0.146	0.751	0.239	0.240	0.241	0.833
0.5	1000	0	0.003	0.079	0.078	0.948	−0.005	0.128	0.129	0.945
		0.5	−0.069	0.085	0.083	0.861	0.091	0.138	0.138	0.895
		1	−0.109	0.090	0.088	0.759	0.144	0.149	0.146	0.831
		1.5	−0.135	0.092	0.092	0.699	0.178	0.155	0.153	0.786
		2	−0.156	0.097	0.096	0.623	0.199	0.163	0.159	0.779
		2.5	−0.174	0.100	0.100	0.589	0.218	0.169	0.165	0.728
		3	−0.184	0.104	0.103	0.556	0.229	0.176	0.170	0.712

increases as the sample size and κ increase.

We also conducted simulation studies to investigate the sensitivity of the proposed estimator under the misspecification of the transformation function within the class of logarithmic transformation functions, i.e., $G(x) = r^{-1}\log(1+rx)$ $(r \geq 0)$. Table 3 reports the parameter estimation results under Scenario 1 with r misspecified as 0 while the data are generated from r_{true} value, with the proportion of exactly observed failure observations among those that are not right-censored being 50%, i.e., $\gamma = 0.5$. The r_{true} is taken from {0, 0.5, 1, 1.5, 2, 2.5, 3}. Biases are less than 0.1 for $r_{true} = 0.5$. The misspecification of r values led to increased biases and lower coverage probabilities than the nominal levels as r_{true} increases from 0.5 to 3.0. However, the proposed variance estimators track the true variations well.

Moreover, we demonstrate the better performance of the proposed method, which accommodates different baseline cumulative hazard functions, over the method in Zhou, Sun and Gilbert (2021) that assumes the same baseline cumulative hazard across all individuals, through a simulation example. Specifically, we generated the data from the Cox-Aalen transformation model with two different baseline cumulative hazard functions. Ignoring the difference and erroneously assuming the same cumulative hazard function, it results in biased estimation of the survival function and cumulative hazard. Therefore, the proposed method is more adept at capturing complex cumulative hazard functions. See Appendix A.4 for details.

7. Application

The AIDS Clinical Trials Group (ACTG) 175 trial enrolled a cohort of 2467 persons living with HIV-1 (PLWH) whose CD4 cell counts ranged from 200 to 500 per cubic millimeters (Hammer et al., 1996). These patients included those who had received antiretroviral therapy prior to the study (ART-experienced) and those who had not (ART-naïve). The primary objective of the study was to compare the efficacy of four antiretroviral regimens – zidovudine only, zidovudine + didanosine, zidovudine + zalcitabine, and didanosine only – in reducing mortality or AIDS morbidity among PLWH (Hammer et al., 1996). Patients in the trial were randomly assigned to one of the antiretroviral regimens and underwent examinations at weeks 2, 4, and 8, followed every 12 weeks thereafter. Their CD4 cell counts were measured within 30 days before randomization and at each follow-up visit from week 8 onwards. The primary endpoint of the study was a composite outcome defined as the first event among (1) CD4 failure: a CD4 cell count at or below 50 percent of the average of two pre-treatment/baseline counts that is confirmed by a second count obtained within 3 to 21 days; (2) AIDS: development of the acquired immunodeficiency syndrome (AIDS) defined by the 1987 Centers for Disease Control criteria (CDC, 1987); and (3) death. If the first event is death, it is observed exactly. If the first event is CD4 failure or AIDS, the exact failure time is unknown due to periodic examinations, resulting in an interval-censored observation. If none of the three event types occurred by the last examination, a right-censored observation is obtained.

After excluding 10 participants without CD4 cell count measurements, 2457 PLWH were included in the full cohort for analysis. Among them, 1396 (56.82%) were in the ART-experienced group, while 1061 (43.18%) were in the ART-naïve group. A total of 306 primary endpoints (12.45%) were recorded in the full cohort, with 230 CD4 failure or AIDS events (9.36%) and 76 deaths (3.09%). More specifically, we observed 215 cases in the ART-experienced group, with 167 CD4 failure or AIDS events and 48 deaths, while 91 cases occurred in the ART-naïve group, with 63 CD4 failure or AIDS events and 28 deaths.

We consider the following Cox-Aalen transformation model

$$\Lambda(t \mid X, Z(\cdot)) = G\Big[\int_0^t \exp\Big\{\beta_1 Z_1(s) + \beta_2 Z_2 + \beta_3 X_2 Z_2\Big\} d\Lambda_X(s)\Big], \qquad (17)$$

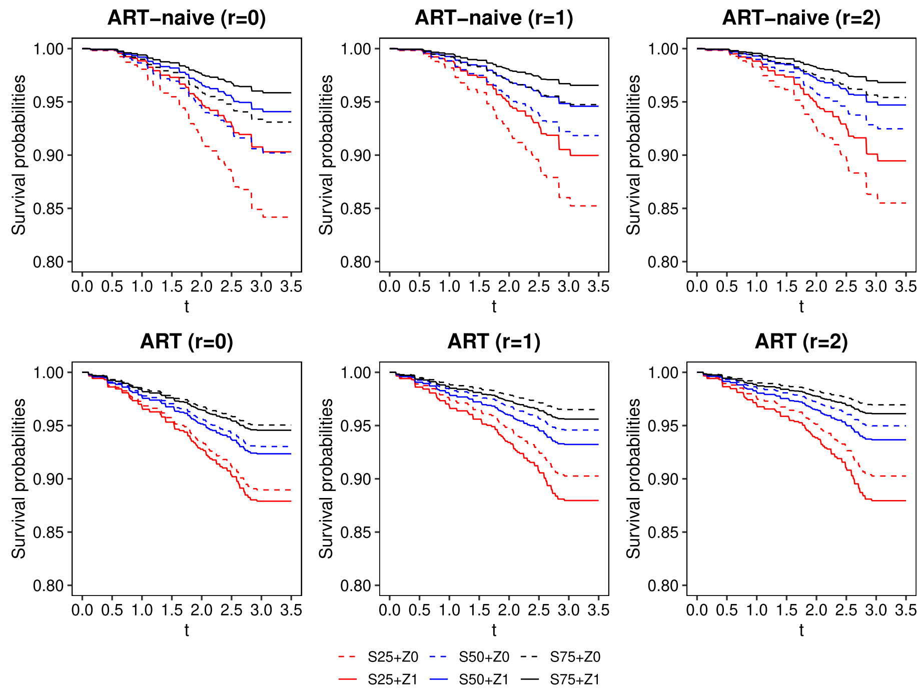

FIG 2. *Estimated survival probabilities for the ART-experienced and ART-naïve groups using the full cohort under model (17) for the ACTG 175 trial. Here, S25, S50 and S75 represent the 25th, 50th and 75th percentile of* $\log_{10}$ *(CD4), respectively. Moreover, Z0 and Z1 stand for the control and treatment groups, respectively.*

where β_1, β_2 and β_3 are unknown regression coefficients. In this model, $Z_1(s)$ is the time-dependent covariate $\log_{10}(\text{CD4})$, and Z_2 denotes the treatment indicator, i.e., 1 = three regimens pooled (zidovudine + didanosine, zidovudine + zalcitabine, didanosine only) and 0 = zidovudine only. In addition, $\Lambda_X(t) = A_1(t) + A_2(t)X_2$, where X_2 takes on values of 1 or 0 depending on whether the patient is in the ART-experienced or ART-naïve group, respectively. Since CD4 cell counts were only measured at scheduled times, linear interpolation was used to create CD4 cell count curves by connecting their values at measurement times. We fitted model (17) with the logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$. To choose the optimal value of r, we plotted the log-likelihood against r, with values of r from 0 to 3 in increments of 0.1, as shown in Figure 3. Based on this plot, we selected $r = 2$ as the best-fit value. Table 4 presents the estimation results for the selected model, including estimates of the model parameters and their standard errors. For comparison, we also report the estimation results for $r = 0$ and $r = 1$.

From the upper panel of Table 4, a lower value of $\log_{10}(\text{CD4})$ is associated with a significantly higher risk of the composite endpoint in all models considered. The treatment is significant under the model $r = 0$ and marginally

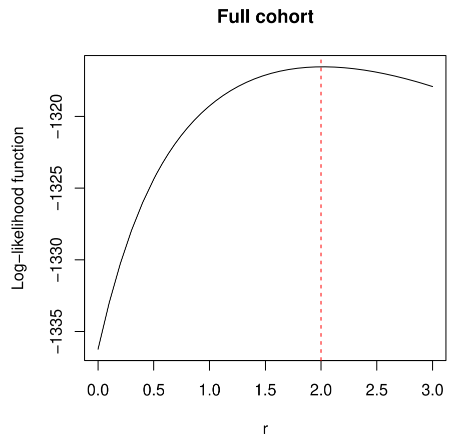

FIG 3. *Log-likelihood function for the estimated model under* $G(x) = r^{-1}\log(1 + rx)$ *with* r *values in the interval [0, 3] and a step size of 0.1.*

TABLE 4
Regression analysis results for the ACTG 175 trial. Est and SE stand for the estimates of the regression parameters and the estimated standard errors, respectively. The selected r values for the full cohort, ART-naïve group (using the ZL model) and ART-experienced (ART) group (using the ZL model) are 2, 2.8 and 1.7, respectively.

	$r = 0$			$r = 1$			Selected r		
Covariates	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
				Full cohort under model (17)					
$\log_{10}$ (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-naïve					
$\log_{10}$ (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					
$\log_{10}$ (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

significant under the model $r = 1$. Notably, the treatment effect varies between the ART-experienced and ART-naïve groups. The estimated coefficient for treatment is negative in the ART-naïve group, indicating that the treatment group ($Z_2 = 1$) has a significantly lower risk of the composite endpoint. In contrast, the estimated coefficient is positive in the ART-experienced group, suggesting that the control group ($Z_2 = 0$) has a lower risk. The estimated effect for the interaction term between Treatment and ART-experienced status is significant for $r = 0$ and $r = 1$ and marginally significant for $r = 2$. Figure 2 supports these findings.

We also conducted separate analyses of the ART-experienced and ART-naïve groups using the ZL model, and compared the results with those obtained using the proposed model. From the lower two panels of Table 4, we found that

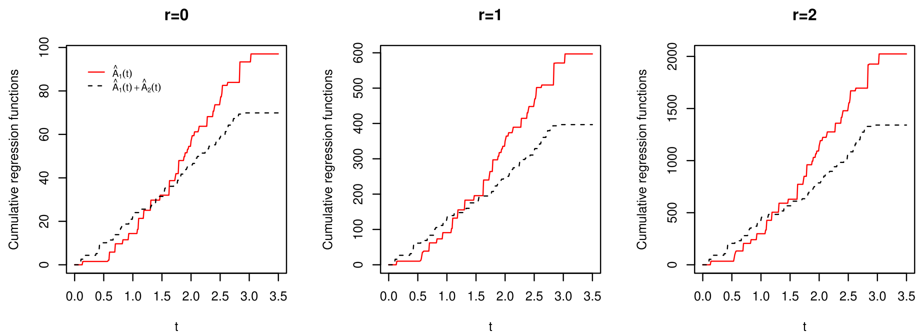

FIG 4. *Estimated cumulative regression functions for the ART-experienced and ART-naïve groups in the ACTG 175 trial, represented by* $\hat{A}_1(t) + \hat{A}_2(t)$ *and* $\hat{A}_1(t)$*, respectively. Here,* $G(x) = r^{-1}\log(1 + rx)$ *with* $r =$ *0, 1, and 2 (selected).*

the estimated effects of $\log_{10}$(CD4) and Treatment are similar, while the SE using the ZL model are larger with slightly larger p-values. However, one benefit of using the proposed model is the assessment of the interaction between Treatment and ART-experienced status. The result shows that the treatment effect is significantly different for the ART-experienced and ART-naïve groups. Additionally, Figure 4 provides further insight into the estimated baseline cumulative regression functions in the ART-experienced and ART-naïve groups using the proposed model. The figure illustrates that the risk of the composite endpointcrosses, i.e., patients in the ART-experienced group have a higher risk than those in the ART-naïve group at the beginning of the study and through an earlier stage; however, after a certain point, the risk is higher in the ART-naïve group. Given patients were not randomized to the ART-experienced vs. ART-naïve group, it is challenging to interpret the crossing risk curves. The explanation may have to do with the differential timing of drug resistance in the two groups and/or correlations of ART-experienced status with prognostic factors.

Although Figure 4 shows that $\hat{A}_2(t)$ is distinct from 0 for $r = 0$, 1 and 2, the hypothesis testing of $A_2(t) = 0$ does not indicate a significant difference in the cumulative baseline hazard functions for the ART-experienced and ART-naïve groups. The p-values are 0.338, 0.315 and 0.364, respectively, for $r = 0$, 1 and 2. The diagnostics plot given in Figure 5 shows $\hat{A}_2(t)$ versus 50 weighted bootstraps of $\tilde{A}_2(t) - \hat{A}_2(t)$ for $r = 0, 1$ and 2 (selected), which shows that $\hat{A}_2(t)$ does not deviate from the reference distribution generated by the bootstrap samples. The large variability in $\hat{A}_2(t)$ could have contributed to the lack of clear evidence for rejecting the null hypothesis. Nevertheless, the ACTG 175 trial example demonstrates the utility of the proposed statistical methods for Cox-Aalen transformation models with partly interval-censored data and its potential in enhancing statistical power for discovering treatment effects. Table 4 shows reduced estimation standard errors by allowing different baseline hazard functions for the ART-experienced and ART-naïve groups even though there is a lack of statistical significance.

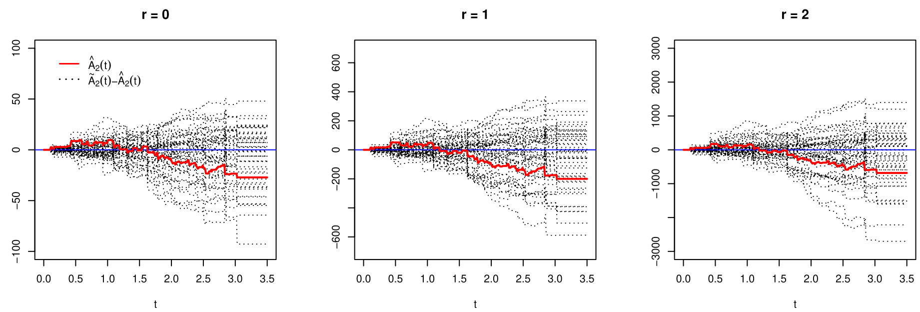

FIG 5. *A diagnosis plot that shows $\hat{A}_2(t)$ versus 50 weighted bootstraps of $\tilde{A}_2(t) - \hat{A}_2(t)$ for $r = 0, 1$ and 2 (selected) under the logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$ for the ACTG 175 trial.*

8. Discussion

The maximum likelihood approach has yet to be thoroughly explored for additive hazards models. The existing literature often uses the least squares principle (Aalen, 1980, 1989; Huffer and McKeague, 1991; Scheike and Zhang, 2002) to estimate the cumulative effect of covariates. Most recently, Boruvka and Cook (2015) proposed semiparametric maximum likelihood estimation for the Cox-Aalen model with fixed covariates from interval-censored data. Nevertheless, performing maximum likelihood estimation with additive components can be highly challenging, and we propose an alternative estimating equation approach that is fast and stable with theoretical guarantees. Note that the theoretical development in this paper is for partly interval-censored data. For purely interval-censored data, we conjecture that $\hat{A}$ would converge to A_0 slower than the parametric rate, and it would require new arguments to finish the proof. We leave this for future research.

For real data applications, one must assess the adequacy of the prespecified function G as misspecifying this function can result in erroneous inferences. Chen, Lin and Zeng (2012) considered appropriate time-dependent residuals and constructed various graphical and numerical procedures for model assessment. In the analysis of the ACTG 175 trial, we assumed that the function G is indexed by a parameter r, and selected the value that maximizes the estimated log-likelihood function. Model fitting based on this process of selecting r is known to generate the post-model-selection inference problem. Further research would be of interest to investigate other approaches such as sample-splitting and cross-fitting to allow valid inference (Zhang, Khalili and Asgharian, 2022). It would be worthwhile to develop formal diagnostic procedures to check the appropriateness of the G function and other model assumptions. In addition, determining the appropriate additive and multiplicative covariates can also be a challenging task, especially when prior knowledge is limited. One approach to address this issue is to evaluate different combinations of additive and multiplicative covariates, calculate the AIC for each model, and select the model with the lowest AIC as the best fit (Qu and Sun, 2019; Yu et al., 2019). However, rigorous evaluation methods are still needed, which are left for future work.

In addition, because the ART-experienced vs. ART-naïve group was not randomized, the analysis should adjust for potential confounders of the effect of ART-experienced status on the composite endpoint. Being ART experienced during the time of the ACTG 175 trial would likely mean receipt of zidovudine (AZT) monotherapy, for which drug resistance mutations rapidly develop. Therefore, the subsequent use of ART, especially AZT monotherapy, is hypothesized to place ART-experienced individuals at greater risk of the composite endpoint compared to ART-naïve individuals whose viral populations were not under selection pressure for the acquisition of resistance.

Lastly, while the proposed framework could potentially be extended to handle competing risks, this extension is not straightforward. Exploring this direction in future research would be of interest.

Appendices

In Section A.1, we give the details of the M-step for nonparametric maximum likelihood estimator (NPMLE) when X is a vector of design variables for categories. In Section A.2, we show that the ES algorithm proposed in Section 3.3 and the EM algorithm in Section 3.2 are equivalent when X is a vector of design variables for categories. In Section A.3, we provide detailed proofs of Theorems 4.1–4.3 in Section 4. Additional simulation results are given in Section A.4.

A.1. The proposed NPMLE in a special case

In Section 3.2, we derived a nonparametric maximum likelihood estimator (NPMLE) for the proposed model with partly interval-censored data. Specifically, the NPMLE can be obtained via an EM algorithm where we solve the following set of equations (12) and (13) in the M-step of the main paper.

Here, we demonstrate that (12) and (13) for the NPMLE can be efficiently solved in the special case when X is a vector of design variables for categories. Let J be a categorical variable with q levels. Without loss of generality, we assume that J takes values in $\{1,\dots,q\}$. Let $X = (1, X_2, \dots, X_q)$ where $X_2,\dots,X_q$ are group indicators, i.e., $X_2 = I(J = 2), \dots, X_q = I(J = q)$. Here, $J = 1$ is considered as the reference group. We propose the following Gauss-Seidel method to jointly solve (12) and (13). Start with some initial values of the unknown parameters.

Step 1. Fixing β, we update $a_k = (a_{1k},\dots,a_{qk})^\top (k = 1,\dots,m)$ by solving (12). Note that for a fixed k, (12) can be written as

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

Hence, we obtain that

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

Step 2. Fixing $a_1, \ldots, a_m$, we update β by solving (13) using the Newton-Raphson method. We iterate between Steps 1 and 2 until convergence.

A.2. Equivalence between the proposed ES and EM estimators in a special case

In this subsection, we show that when X is a vector of design variables for categories, the ES algorithm proposed in Section 3.3 coincides with the EM algorithm proposed in Section 3.2. To show this, it suffices to prove that for a fixed β, equations (12) and (15) in the main paper share the same solution in terms of a_k $(k = 1, \ldots, m)$.

Let J be a categorical variable with q levels, as defined in Appendix A.1. It is worth noting that for a fixed k, equation (15) for the ES estimator can be expressed as follows:

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

which in this special case is equivalent to

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

It is evident that (A.1) is the unique solution to (A.2) and, consequently, the unique solution to equation (15). Furthermore, in Section A.1, we demonstrated that (A.1) is also the unique solution to equation (12) in this particular case. Thus, when X represents a vector of design variables for categories, the ES and EM estimators coincide with each other.

A.3. Proofs of Theorems

Lemmas used in the proofs of Theorem 4.1–4.3 in Sections A.3.1–A.3.3 are given in Section A.3.4.

A.3.1. Proof of Theorem 4.1

For convenience, we may drop the i from $\{\Delta_i, T_i, L_i, R_i, X_i, Z_i, \xi_i, W_{ik}, t_{ik}\}$ when it does not cause confusion. Let

$$\rho_0(t;\vartheta) = \int_0^t \exp\{\beta^\top Z(s)\} X^\top(s) dA(s)$$

and

$$\rho_1(t;\vartheta) = \int_0^t \exp\{\beta^\top Z(s)\} Z(s) X^\top(s) dA(s).$$

In addition, let $g_1(\cdot) = G'(\cdot) - G''(\cdot)/G'(\cdot)$. Setting $t = t_k$, from the deviations in E-step in Section 3.3, the posterior mean of the latent Poisson random variable W_{ik} given the observed data can be written as

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

where

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

and the posterior mean of ξ can be written as

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

where

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

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

$$\mathbb{P}_n \begin{pmatrix} \Phi_1(\vartheta) \\ \Phi_2(\vartheta)(t) \end{pmatrix} = 0 \quad \text{for all } t \in [0, \tau], \tag{A.3}$$

where

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

and

$$\Phi_2(\vartheta)(t) = \Delta X(t)I(t=T) - \Delta I(t \le R^*)g_1\{\rho_0(T;\vartheta)\}e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t)$$
$$+ (1-\Delta)H_1(L,R;\vartheta)I(L<t\le R)I(R<\infty)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t)$$
$$- (1-\Delta)I(t\le R^*)H_2(L,R;\vartheta)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t). \qquad (A.5)$$

There are an infinite number of estimating equations in (A.3). To resolve this, we consider

$$\Phi_1(\vartheta)[h_0]$$
$$= \Delta h_0^\top Z(T) - \Delta g_1\{\rho_0(T;\vartheta)\}\int_0^\tau I(t\le R^*)e^{\beta^\top Z(t)}\{h_0^\top Z(t)\}X^\top(t)dA(t)$$
$$+(1-\Delta)H_1(L,R;\vartheta)\int_0^\tau I(L<t\le R)I(R<\infty)e^{\beta^\top Z(t)}\{h_0^\top Z(t)\}X^\top(t)dA(t)$$
$$-(1-\Delta)H_2(L,R;\vartheta)\int_0^\tau I(t\le R^*)e^{\beta^\top Z(t)}\{h_0^\top Z(t)\}X^\top(t)dA(t), \qquad (A.6)$$

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

$$\widetilde{\Phi}_2(\vartheta)[h]$$
$$= \Delta h(T)\circ X(T) - \Delta g_1\{\rho_0(T;\vartheta)\}\int_0^\tau I(t\le R^*)e^{\beta^\top Z(t)}\{h(t)\circ X(t)\}X^\top(t)dA(t)$$
$$+ (1-\Delta)H_1(L,R;\vartheta)\int_0^\tau I(L<t\le R)I(R<\infty)e^{\beta^\top Z(t)}\{h(t)\circ X(t)\}X^\top(t)dA(t)$$
$$- (1-\Delta)H_2(L,R;\vartheta)\int_0^\tau I(t\le R^*)e^{\beta^\top Z(t)}\{h(t)\circ X(t)\}X^\top(t)dA(t). \qquad (A.7)$$

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 all functions of bounded variation over $[0,\tau]$ 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 $\mathbb{P}_n\widetilde{\Phi}_2(\vartheta)[h]=0$ for every $h\in BV_1^q[0,\tau]$ is equivalent to

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

where

$$\Phi_2(\vartheta)[h]$$
$$= \sum_{j=1}^q \Delta\left[h_j(T)X_j(T) - g_1\{\rho_0(T;\vartheta)\}\int_0^\tau I(t\le R^*)e^{\beta^\top Z(t)}h_j(t)X_j(t)X^\top(t)dA(t)\right]$$
$$+\sum_{j=1}^q(1-\Delta)H_1(L,R;\vartheta)\int_0^\tau I(L<t\le R)I(R<\infty)e^{\beta^\top Z(t)}h_j(t)X_j(t)X^\top(t)dA(t)$$
$$-\sum_{j=1}^q(1-\Delta)H_2(L,R;\vartheta)\int_0^\tau I(t\le R^*)e^{\beta^\top Z(t)}h_j(t)X_j(t)X^\top(t)dA(t). \qquad (A.8)$$

Therefore, the equations $\mathbb{P}_n\Phi_1(\vartheta) = 0$ and $\mathbb{P}_n\Phi_2(\vartheta)(t) = 0$ for every $t \in [0, \tau]$ in (A.3) are equivalent to equations $\mathbb{P}_n\Phi_1(\vartheta)[h_0] = 0$ for every $h_0 \in \mathbb{R}^d$ and $\mathbb{P}_n\Phi_2(\vartheta)[h] = 0$ 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 (1996, Section 3.3.1) and Gao, Zeng and Lin (2017) among others. Thus, the proposed ES estimator $\hat{\vartheta}$ is equivalent to the solution of the estimating equation

$$\mathbb{P}_n\Phi(\vartheta)[\widetilde{h}] \equiv \mathbb{P}_n\Phi_1(\vartheta)[h_0]+\mathbb{P}_n\Phi_2(\vartheta)[h], \quad \text{for every } \widetilde{h} = (h_0, h) \in \mathbb{R}^d \times BV_1^q[0, \tau]. \tag{A.9}$$

Let $\mathcal{H} = \mathbb{R}^d \times BV^q[0, \tau]$. 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 $[0, \tau]$. In addition, let H be a subset of $\mathcal{H}$ with $\|\widetilde{h}\|_{\mathcal{H}} \le M < \infty$, and $\ell^\infty(H)$ be the collection of all bounded functions from H to $\mathbb{R}$.

Let $\widetilde{h} = (h_0, h) \in H$. It is easy to note that the function $\mathbb{P}_n\Phi(\vartheta)[\widetilde{h}]$ is a map from $\mathbb{R}^d \times BV^q[0, \tau]$ to $l^\infty(H)$. Let $B_\delta(\vartheta_0) = \{\vartheta = (\beta, A) : \|\beta - \beta_0\|_d + \|A - A_0\|_\rho < \delta\}$, where $\delta > 0$. For any ϑ in $B_\delta(\vartheta_0)$, we write $\Psi(\vartheta)[\widetilde{h}] = P\Phi(\vartheta)[\widetilde{h}]$ and $\Psi_n(\vartheta)[\widetilde{h}] = \mathbb{P}_n\Phi(\vartheta)[\widetilde{h}]$. Note that $\Psi(\vartheta)[\widetilde{h}]$ and $\Psi_n(\vartheta)[\widetilde{h}]$ depend on $\widetilde{h}$. To suppress notations, we write $\Psi(\vartheta)$ and $\Psi_n(\vartheta)$ for $\Psi(\vartheta)[\widetilde{h}]$ and $\Psi_n(\vartheta)[\widetilde{h}]$, respectively, when there is no confusion. We prove the local consistency of $\hat{\vartheta} = (\hat{\beta}, \hat{A})$ by verifying the three conditions in Theorem 1.20 (the implicit function theorem) (Schwartz, 1969). These conditions are (1)$\Psi_n(\vartheta)[\widetilde{h}]$ is Fréchet-differentiable in $B_\delta(\vartheta_0)$ with some $\delta > 0$; (2) the corresponding Fréchet derivative map depends continuously on ϑ in $B_\delta(\vartheta_0)$; (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 A.2 and the third condition in Lemma A.4 in Section A.3.4.

By Lemma A.1, the class $\{\Phi(\vartheta)[\widetilde{h}] : \vartheta \in B_\delta(\vartheta_0), \widetilde{h} \in H\}$ is a Donsker class for some $\delta > 0$. The class $\{\Phi(\vartheta_0)[\widetilde{h}] : \widetilde{h} \in H\}$ is also Donsker via Theorem 2.10.1 in van der Vaart and Wellner (1996) because the latter class is a subset of the former class. By Donsker properties,

$$\mathbb{P}_n\Phi(\vartheta_0)[\widetilde{h}] - P\Phi(\vartheta_0)[\widetilde{h}] = o_p(1),$$

or equivalently, $\Psi_n(\vartheta_0) - \Psi(\vartheta_0) = o_p(1)$. In addition, $\Psi(\vartheta_0) = P\Phi(\vartheta_0)[\widetilde{h}] = 0$ can be easily checked by double expectation properties. Therefore, $\Psi_n(\vartheta_0) = o_p(1)$. By Lemmas A.2 and A.4, we verified all three conditions of the implicit function theorem (Schwartz, 1969), and hence it yields that $\Psi_n(\vartheta)$ is a one-to-one map from $B_\delta(\vartheta_0)$ 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 $\hat{\vartheta} = (\hat{\beta}, \hat{A})$ with $(\|\hat{\beta} - \beta_0\|_d + \|\hat{A} - A_0\|_\rho) < \delta$ and $\Psi_n(\hat{\vartheta}) = \mathbb{P}_n\Phi(\hat{\vartheta})[\widetilde{h}] = 0$ for any $\widetilde{h} \in H$. This proves the consistency of $\hat{\vartheta} = (\hat{\beta}, \hat{A})$. $\square$

A.3.2. Proof of Theorem 4.2

We establish the asymptotic normality of $\hat{\vartheta} = (\hat{\beta}, \hat{A})$ by applying Theorem 3.3.1 and Lemma 3.3.5 of van der Vaart and Wellner (1996). Let

$$\mathbb{G}_n\Phi(\vartheta)[\widetilde{h}] = n^{1/2}\left\{\Psi_n(\vartheta)[\widetilde{h}] - \Psi(\vartheta)[\widetilde{h}]\right\},$$

where $\Psi_n(\vartheta)[\widetilde{h}] = \mathbb{P}_n\Phi(\vartheta)[\widetilde{h}]$ and $\Psi(\vartheta)[\widetilde{h}] = P\Phi(\vartheta)[\widetilde{h}]$. We begin by showing that $\mathbb{G}_n\Phi(\vartheta_0)[\widetilde{h}]$ converges in distribution to a tight random element $\mathcal{W}$ in $l^\infty(H)$. By Lemma A.1, the class $\{\Phi(\vartheta)[\widetilde{h}] : \vartheta \in B_\delta(\vartheta_0), \widetilde{h} \in H\}$ is P-Donkser. It follows that the class $\{\Phi(\vartheta_0)[\widetilde{h}] : \widetilde{h} \in H\}$, as a subset of a Donsker class, is also Donsker (van der Vaart and Wellner, 1996, Theorem 2.10.1). We note that the function $\Phi(\vartheta)[\widetilde{h}]$ involves the terms $g_1\{\rho_0(T;\vartheta)\}$, $H_1(L,R;\vartheta)$ and $H_2(L,R;\vartheta)$, of which the denominators are all bounded away from 0 as argued in Lemma A.1. Then, under Conditions 1–5, we have

$$\sup_{\widetilde{h}\in H} \|\Psi(\vartheta_0)[\widetilde{h}]\| < \infty.$$

Hence, $\mathbb{G}_n\Phi(\vartheta_0)[\widetilde{h}] = n^{1/2}\left\{\Psi_n(\vartheta_0)[\widetilde{h}] - \Psi(\vartheta_0)[\widetilde{h}]\right\}$ converges weakly to a zero-mean Gaussian process $\mathcal{W}$ in $l^\infty(H)$.

By Lemma A.2, the Fréchet-differentiablility of $\Psi(\vartheta)$ at $\vartheta = \vartheta_0$ can be checked straightforwardly. In particular, we consider one-dimensional submodels $\eta \to \vartheta_0 + \eta(\vartheta - \vartheta_0)$ and calculate the Fréchet derivative $\dot{\Psi}_{\vartheta_0}(\vartheta - \vartheta_0)$ using the the weaker form

$$\begin{aligned}\dot{\Psi}_{\vartheta_0}(\vartheta - \vartheta_0) &= \left.\frac{d\Psi(\vartheta_0 + \eta(\vartheta - \vartheta_0))}{d\eta}\right|_{\eta=0} \\ &= Q_1^\top[\widetilde{h}](\beta - \beta_0) + \int_0^\tau Q_2[\widetilde{h}](t)d(A(t) - A_0(t)).\end{aligned}$$

Detailed calculations and expressions for $Q_1[\widetilde{h}]$ and $Q_2[\widetilde{h}](\cdot)$ are given in Lemma A.2. In particular, $Q_1[\widetilde{h}]$ and $Q_2[\widetilde{h}](\cdot)$ are given by (A.13). Furthermore, Lemma A.3 establishes the invertibility of the Fréchet derivative map $\dot{\Psi}_{\vartheta_0}$.

Next, we verify condition (3.3.4) of Theorem 3.3.1 (van der Vaart and Wellner, 1996), which is sufficient to verify the conditions in Lemma 3.3.5 of van der Vaart and Wellner (1996). Since the classes $\left\{\Phi(\vartheta)[\widetilde{h}] : \vartheta \in B_\delta(\beta_0, A_0), \widetilde{h} \in H\right\}$ and $\{\Phi(\vartheta_0)[\widetilde{h}] : \widetilde{h} \in H\}$ are both Donsker classes, the class $\{\Phi(\vartheta)[\widetilde{h}] - \Phi(\vartheta_0)[\widetilde{h}] : \vartheta \in B_\delta(\beta_0, A_0), \widetilde{h} \in H\}$ 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, 1996, Example 2.10.7). Under Conditions 1–5, it easy to note that $\Phi(\vartheta)[\widetilde{h}]$ is a continuous function over ϑ. In addition, $\widetilde{h}(t)$ has bounded total variation over $[0,\tau]$. Hence, $\Phi(\vartheta)[\widetilde{h}]$ converges to $\Phi(\vartheta_0)[\widetilde{h}]$ pointwise and uniformly in $\widetilde{h}$. By the dominated convergence theorem,

$$\sup_{\widetilde{h}\in H} P\{\Phi(\vartheta)[\widetilde{h}] - \Phi(\vartheta_0)[\widetilde{h}]\}^2 \to 0,$$

as $\vartheta \to \vartheta_0$ (van der Vaart and Wellner, 1996, p. 317). The consistency of $\hat{\vartheta}$ has been proved, i.e., $\hat{\vartheta}$ converges to ϑ_0 almost surely. Hence, applying Lemma 3.3.5 (van der Vaart and Wellner, 1996), we have

$$\|\mathbb{G}_n(\Phi(\hat{\vartheta}) - \Phi(\vartheta_0))\| = o_{p^*}(1 + n^{1/2}\|\hat{\vartheta} - \vartheta_0\|), \tag{A.10}$$

where $o_{p^*}(1)$ denotes convergence to zero in outer probability. Note that equation (A.10) can be written as

$$n^{1/2}(\Psi_n - \Psi)(\hat{\vartheta}) - n^{1/2}(\Psi_n - \Psi)(\vartheta_0) = o_{p^*}(1 + n^{1/2}\|\hat{\vartheta} - \vartheta_0\|).$$

In brief, we have shown that (1) $\mathbb{G}_n\Phi(\vartheta_0)[\widetilde{h}]$ converges in distribution to a tight random element $\mathcal{W}$; (2) the continuous invertibility of the operator $\dot{\Psi}_{\vartheta_0}$; (3) condition (3.3.2) of van der Vaart and Wellner (1996, Theorem 3.3.1); (4) $\Psi(\vartheta_0) = 0$ and $\Psi_n(\hat{\vartheta}) = 0$. The last statement is a trivial result by the definitions of ϑ_0 and $\hat{\vartheta}$. According to Theorem 3.3.1 of van der Vaart and Wellner (1996), we obtain

$$n^{1/2}\dot{\Psi}_{\vartheta_0}(\hat{\vartheta} - \vartheta_0) = -n^{1/2}(\Psi_n - \Psi)(\vartheta_0) + o_{p^*}(1).$$

Finally, the continuous mapping theorem implies

$$n^{1/2}(\hat{\vartheta} - \vartheta_0) \rightsquigarrow -\dot{\Psi}_{\vartheta_0}^{-1}\mathcal{W}.$$

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

A.3.3. Proof of Theorem 4.3

Let $e_1, \ldots, e_n$ be positive i.i.d random variables with a standard exponential distribution. Hence, $\mu = E(e_1) = 1 < \infty$, $\sigma^2 = \text{var}(e_1) = 1 < \infty$ 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 $\mathcal{O}_i = \{\Delta_i, \Delta_i T_i, (1-\Delta_i)L_i, (1-\Delta_i)R_i, X_i, Z_i\}$ $(i = 1, \ldots, n)$.

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 $\widetilde{\Psi}_n$ be Ψ_n but with $\mathbb{P}_n$ replaced by $\widetilde{\mathbb{P}}_n$ and $\widetilde{\vartheta} = (\widetilde{\beta}, \widetilde{A})$ be the weighted bootstrap estimator that solves $\widetilde{\Psi}_n(\vartheta) = 0$. Let $\widetilde{\Psi}(\vartheta) = P\big(\widetilde{e} \cdot \Phi(\vartheta)[\widetilde{h}]\big)$, where $\widetilde{e}$ be a generic version of $\widetilde{e}_1$. By Lemma A.1, the class of functions $\big\{\Phi(\vartheta)[\widetilde{h}] : \vartheta \in B_\delta(\vartheta_0), \widetilde{h} \in H\big\}$ is P-Donsker for some fixed $\delta > 0$. Hence, so is the class $\big\{\widetilde{e} \cdot \Phi(\vartheta)[\widetilde{h}] : \vartheta \in B_\delta(\vartheta_0), \widetilde{h} \in H\big\}$ via the multiplier central limit theorem (Kosorok, 2008, Theorem 10.1). We also note that $P\big(\widetilde{e} \cdot \Phi(\vartheta)[\widetilde{h}]\big) = P\big(\Phi(\vartheta)[\widetilde{h}]\big)$, which implies $\widetilde{\Psi}(\vartheta) = \Psi(\vartheta)$. Trivially, the consistency of $\widetilde{\vartheta}$ holds by similar arguments in proving Theorem 4.1.

The weighted bootstrap empirical process is defined as

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

Applying the Taylor series expansion, we have

$$\begin{aligned}0 &= \widetilde{\mathbb{P}}_n\Phi(\widetilde{\vartheta})[\widetilde{h}] - \widetilde{\mathbb{P}}_n\Phi(\hat{\vartheta})[\widetilde{h}] + \widetilde{\mathbb{P}}_n\Phi(\hat{\vartheta})[\widetilde{h}] - \mathbb{P}_n\Phi(\hat{\vartheta})[\widetilde{h}] \\ &= \left(\frac{\partial\widetilde{\mathbb{P}}_n\Phi(\vartheta)[\widetilde{h}]}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)(\widetilde{\vartheta}-\hat{\vartheta}) + (\widetilde{\mathbb{P}}_n - \mathbb{P}_n)\Phi(\hat{\vartheta})[\widetilde{h}] + o_p(\|\widetilde{\vartheta}-\vartheta_0\| + \|\hat{\vartheta}-\vartheta_0\|)\end{aligned} \tag{A.11}$$

By Theorem 2.6 of Kosorok (2008), the conditional distribution of $(\widetilde{\mathbb{P}}_n - \mathbb{P}_n)\Phi(\hat{\vartheta})[\widetilde{h}]$ given the data is asymptotically equivalent to the distribution of $(\mathbb{P}_n - P)\Phi(\hat{\vartheta})[\widetilde{h}]$ by the fact that $\mu = \sigma^2 = 1$ with a sequence of i.i.d standard exponential random variables. Hence, (A.11) can be written as

$$\begin{aligned}n^{1/2}\dot{\Psi}_{\vartheta_0}(\widetilde{\vartheta}-\hat{\vartheta}) &= -n^{1/2}(\widetilde{\mathbb{P}}_n - \mathbb{P}_n)\Phi(\hat{\vartheta})[\widetilde{h}] + o_p(1) \\ &= -n^{1/2}(\mathbb{P}_n - P)\Phi(\hat{\vartheta})[\widetilde{h}] + o_p(1) \\ &= -\mathbb{G}_n\Phi(\vartheta_0)[\widetilde{h}] + o_p(1).\end{aligned}$$

Thus, Lemma A.3 and the continuous mapping theorem give

$$n^{1/2}(\widetilde{\vartheta}-\hat{\vartheta}) \rightsquigarrow -\dot{\Psi}_{\vartheta_0}^{-1}\mathcal{W}.$$

We conclude that $n^{1/2}(\widetilde{\vartheta}-\hat{\vartheta})$ converges to a zero-mean Gaussian process. Moreover, $n^{1/2}(\widetilde{\vartheta}-\hat{\vartheta})$ and $n^{1/2}(\hat{\vartheta}-\vartheta_0)$ have the same asymptotic distribution. □

A.3.4. Lemmas for Theorem 4.1–4.3

This subsection presents the lemmas used in the proof of Theorem 4.1–4.3 along with their proofs.

Lemma A.1. *Under Conditions 1–5, the class of functions* $\{\Phi(\vartheta)[\widetilde{h}] : \vartheta \in B_\delta(\vartheta_0), \widetilde{h} \in H\}$ *is P-Donsker for some fixed* $\delta > 0$.

Proof of Lemma A.1. To show the class

$$\left\{\Phi(\vartheta)[\widetilde{h}] = \Phi_1(\vartheta)[h_0] + \Phi_2(\vartheta)[h] : \vartheta \in B_\delta(\vartheta_0), \widetilde{h} = (h_0, h) \in H\right\}$$

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, 1996) 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 compact sets.

Note that

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

Under Condition 2, each $A_j(t)$ $(j=1,\dots,q)$ 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, Mao and Lin (2016), the class $\{\int_0^T e^{\beta^\top Z(s)} X_j(s) dA_{j1}(s) : \vartheta \in B_\delta(\vartheta_0)\}$ 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 $\{\int_0^T e^{\beta^\top Z(s)} X_j(s) dA_{j2}(s) : \vartheta \in B_\delta(\vartheta_0)\}$ is P-Donsker. Hence, the class $\{\int_0^T e^{\beta^\top Z(s)} X_j(s) dA_j(s) : \vartheta \in B_\delta(\vartheta_0)\}$ is P-Donsker because the sum of bounded Donsker classes is also Donsker. It follows that the class $\{\Delta\rho_0(T;\vartheta) : \vartheta \in B_\delta(\vartheta_0)\}$ is a Donsker class. Similarly, the following classes

$$\begin{aligned}
&\{(1-\Delta)\rho_0(L;\vartheta) : \vartheta \in B_\delta(\vartheta_0)\} \\
&\quad = \left\{(1-\Delta)\int_0^L e^{\beta^\top Z(s)} X^\top(s) dA(s) : \vartheta \in B_\delta(\vartheta_0)\right\} \\
&\{(1-\Delta)\rho_0(R;\vartheta) : \vartheta \in B_\delta(\vartheta_0)\} \\
&\quad = \left\{(1-\Delta)\int_0^R e^{\beta^\top Z(s)} X^\top(s) dA(s) : \vartheta \in B_\delta(\vartheta_0)\right\} \\
&\{\Delta\rho_1(T;\vartheta) : \vartheta \in B_\delta(\vartheta_0)\} \\
&\quad = \left\{\Delta\int_0^T e^{\beta^\top Z(s)} Z(s) X^\top(s) dA(s) : \vartheta \in B_\delta(\vartheta_0)\right\} \\
&\{(1-\Delta)\rho_1(L;\vartheta) : \vartheta \in B_\delta(\vartheta_0)\} \\
&\quad = \left\{(1-\Delta)\int_0^L e^{\beta^\top Z(s)} Z(s) X^\top(s) dA(s) : \vartheta \in B_\delta(\vartheta_0)\right\}
\end{aligned}$$

$$\{(1-\Delta)\rho_1(R;\vartheta):\vartheta\in B_\delta(\vartheta_0)\}$$
$$=\left\{(1-\Delta)\int_0^R e^{\beta^\top Z(s)}Z(s)X^\top(s)dA(s):\vartheta\in B_\delta(\vartheta_0)\right\}$$
$$\{(1-\Delta)\rho_1(R^*;\vartheta):\vartheta\in B_\delta(\vartheta_0)\}$$
$$=\left\{(1-\Delta)\int_0^{R^*} e^{\beta^\top Z(s)}Z(s)X^\top(s)dA(s):\vartheta\in B_\delta(\vartheta_0)\right\}$$

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 functions $g_1[\rho_0(T;\vartheta)]$, $\exp[-G\{\rho_0(L;\vartheta)\}]G'\{(\rho_0(L;\vartheta)\}$ and $\exp[-G\{\rho_0(R;\vartheta)\}]G'\{(\rho_0(R;\vartheta)\}I(R<\infty)$ are all bounded for any $\vartheta\in B_\delta(\vartheta_0)$. Notice that the denominators of

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

and

$$H_2(L,R;\vartheta)$$
$$=\frac{\exp[-G\{\rho_0(L;\vartheta)\}]G'\{\rho_0(L;\vartheta)\}-\exp[-G\{\rho_0(R;\vartheta)\}]G'\{\rho_0(R;\vartheta)\}I(R<\infty)}{\exp[-G\{\rho_0(L;\vartheta)\}]-\exp[-G\{\rho_0(R;\vartheta)\}]I(R<\infty)}$$

are

$$\exp[-G\{\rho_0(L;\vartheta)\}]-\exp[-G\{\rho_0(R;\vartheta)\}]I(R<\infty),$$

which is bounded away from zero under Conditions 3–4. Since any continuously differentiable function is locally Lipschitz, the classes $\{\Delta g_1\{\rho_0(T;\vartheta)\}:\vartheta\in B_\delta(\vartheta_0)\}$, $\{(1-\Delta)H_1(L,R;\vartheta):\vartheta\in B_\delta(\vartheta_0)\}$ and $\{(1-\Delta)H_2(L,R;\vartheta):\vartheta\in B_\delta(\vartheta_0)\}$ 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 $\{\Phi_1(\vartheta)[h_0]:\vartheta\in B_\delta(\vartheta_0),h_0\in\mathbb{R}^d\}$ is a Donsker class since $\Phi_1(\vartheta)[h_0]$ depends on h_0 linearly by Theorem 2.10.6 of van der Vaart and Wellner (1996).

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, 1996) 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

$$\left\{\int_0^{(\cdot)} e^{\beta^\top Z(t)}h_j(t)X_j(t)X^\top(t)dA(t):\vartheta\in B_\delta(\vartheta_0),h_j\in BV_1[0,\tau]\right\},$$

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 functions $\left\{\Phi(\vartheta)[\widetilde{h}] : \vartheta \in B_\delta(\vartheta_0), \widetilde{h} \in H\right\}$ is P-Donsker for some $\delta > 0$ because the sums and products of bounded Donsker classes are Donsker classes. □

Lemma A.2. *Under Conditions 1–5, the map* $\Psi\colon \mathbb{R}^d \times BV^q[0,\tau] \to l^\infty(H)$ *is Fréchet-differentiable at* $\vartheta = \vartheta_0$*, with derivative*

$$\dot{\Psi}_{\vartheta_0}(\vartheta - \vartheta_0) = Q_1^\top[\widetilde{h}](\beta - \beta_0) + \int_0^\tau Q_2[\widetilde{h}](t) d(A(t) - A_0(t)), \tag{A.12}$$

where

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

The expressions of B_1*,* $B_{2,j}(t)$*,* $B_3(t)$*, and* $B_4[h](t)$ *are given in* (A.25), (A.26), (A.27) *and* (A.28)*, respectively, in Section A.3.5. The map* Ψ_n *has the same properties and similar Fréchet derivative map at* $\vartheta = \vartheta_0$*, denoted as* $\dot{\Psi}_{\vartheta_0,n}$ *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* $\dot{\Psi}_\vartheta$ *and* $\dot{\Psi}_{\vartheta,n}$ *depend continuously on* ϑ*.*

Proof of Lemma A.2. Consider the one-dimensional submodels $\eta \to \vartheta_0 + \eta(\vartheta - \vartheta_0)$. The Fréchet derivative $\dot{\Psi}_{\vartheta_0}(\vartheta - \vartheta_0)$ can be computed based on the weaker form

$$\begin{aligned} \dot{\Psi}_{\vartheta_0}(\vartheta - \vartheta_0) &= \left.\frac{d\Psi(\vartheta_0 + \eta(\vartheta - \vartheta_0))}{d\eta}\right|_{\eta=0} \\ &= Q_1^\top[\widetilde{h}](\beta - \beta_0) + \int_0^\tau Q_2[\widetilde{h}](t) d(A(t) - A_0(t)). \end{aligned} \tag{A.14}$$

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 Section A.3.5. It can be shown that $||\Psi(\vartheta) - \Psi(\vartheta_0) - \dot{\Psi}_{\vartheta_0}(\vartheta - \vartheta_0)|| = o(||\vartheta - \vartheta_0||)$ as $\vartheta \to \vartheta_0$. Hence, $\Psi(\vartheta)$ is Fréchet-differentiable at ϑ_0. The Fréchet derivative of $\Psi_n(\vartheta) = \mathbb{P}_n \Phi(\vartheta)[\widetilde{h}]$ with respect to ϑ at $\vartheta = \vartheta_0$, denoted as $\dot{\Psi}_{\vartheta_0,n}$, can be derived closely. In particular, we replace Ψ with Ψ_n in (A.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 $\dot{\Psi}_{\vartheta_0,n}$. Then one can show that $\|\Psi_n(\vartheta) - \Psi_n(\vartheta_0) - \dot{\Psi}_{\vartheta_0,n}(\vartheta - \vartheta_0)\| = o(\|\vartheta - \vartheta_0\|)$ as $\vartheta \to \vartheta_0$. Hence, $\Psi_n(\vartheta)$ is also Fréchet-differentiable at ϑ_0. Clearly, both maps $\dot{\Psi}_\vartheta$ and $\dot{\Psi}_{\vartheta,n}$ depend continuously on ϑ in $B_\delta(\vartheta_0)$. □

Lemma A.3. *Under Conditions 1–5, the map* $\dot{\Psi}_{\vartheta_0}$ *is continuously invertible.*

Proof of Lemma A.3. To show the invertibility of the map $\dot{\Psi}_{\vartheta_0}$, we 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

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

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 $A_{j,\eta}(t) = \int_0^t (1 + \eta h_j(s)) dA_j(s)$ $(j = 1, \ldots, q)$, where $\widetilde{h} = (h_0, h) \in H$, $h = (h_1, \ldots, h_q)^\top$. Here, we use A_η to represent that each A_j is replaced with $A_{j,\eta}$. Then the derivatives of the observed data log-likelihood for a single subject along the submodels are

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

and

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

where $A\{T\}$ denotes the jump size of $A(\cdot)$ at T and

$$\rho_2(t;\vartheta)[h] = \sum_{j=1}^{q} \int_0^t e^{\beta^\top Z(s)} h_j(s) X_j(s) dA_j(s).$$

Let $\Phi^*(\vartheta)[\widetilde{h}] = \Phi_1^*(\vartheta)[h_0] + \Phi_2^*(\vartheta)[h]$, for any $\widetilde{h} = (h_0, h) \in H$. Therefore, the maximum likelihood estimator $(\hat{\beta}_{\text{MLE}}, \hat{A}_{\text{MLE}})$ can be obtained by solving the following equation

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

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

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

where

$$\rho_3(t;\vartheta)[h] = \sum_{j=1}^{q} \int_0^t e^{\beta^\top Z(s)} \{h_j(s) X_j(s)\} X^\top(s) dA(s).$$

Let $\Phi(\vartheta)[\widetilde{h}] = \Phi_1(\vartheta)[h_0] + \Phi_2(\vartheta)[h]$ for any $\widetilde{h} = (h_0, h) \in H$. Recall that the proposed ES-estimator solves

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

Next, we establish the relationship between $\Phi(\vartheta)[\widetilde{h}]$ and $\Phi^*(\vartheta)[\widetilde{h}]$, where $\Phi(\vartheta)[\widetilde{h}]$ is the estimating equation constructed to estimate the parameters using an ES algorithm, and $\Phi^*(\vartheta)[\widetilde{h}]$ is the score equation calculated by taking the derivatives of the log-likelihood function along the submodels. Let $\widetilde{h}_j^*(t) = (h_0/q, h_j^*(t))$, where $h_j^*(t) = (h_j(t)X_j(t), \ldots, h_j(t)X_j(t))$ for $j = 1, \ldots, q$. Now we verify that

$$\Phi(\vartheta)[\widetilde{h}] = \sum_{j=1}^{q} \Phi^*(\vartheta)[\widetilde{h}_j^*]. \tag{A.19}$$

To begin, we observe that:

$$\begin{aligned}\rho_2(t;\vartheta)[h_1^*] &= \sum_{j=1}^{q} \int_0^t e^{\beta^\top Z(s)} h_1(s) X_1(s) X_j(s) dA_j(s) \\ &= \int_0^t e^{\beta^\top Z(s)} h_1(s) X_1(s) X^\top(s) dA(s).\end{aligned}$$

Therefore, $\rho_3(t;\vartheta)[h] = \sum_{j=1}^q \rho_2(t;\vartheta)[h_j^*]$. Hence, $\Phi_2(\vartheta)[h] = \sum_{j=1}^q \Phi_2^*(\vartheta)[h_j^*]$, which implies (A.19) holds.

The Fréchet derivative map of $\Psi(\vartheta) = P\Phi(\vartheta)$ at ϑ_0, denoted as $\dot{\Psi}(\vartheta_0) = P\dot{\Phi}(\vartheta_0)$, takes the form shown in (A.13) in Lemma A.2. Similarly, the Fréchet derivative map of $\Psi^*(\vartheta) = P\Phi^*(\vartheta)$ at ϑ_0, denoted as $\dot{\Psi}^*(\vartheta_0) = P\dot{\Phi}^*(\vartheta_0)$, can be calculated via a similar weaker form as shown in (A.14) using equations (A.16) and (A.17). The details have been omitted due to the similarity of the calculations involved. From Lemma A.2, $\dot{\Psi}(\vartheta_0)[\widetilde{h}]$ is a linear operator. To show that $\dot{\Psi}(\vartheta_0)[\widetilde{h}]$ is continuously invertible, it is sufficient to prove that $\dot{\Psi}(\vartheta_0)[\widetilde{h}]$ is a one-to-one map and then it is invertible (Rudin, 1973). If $\widetilde{h} = 0$, it is easy to note that $\dot{\Psi}(\vartheta_0)[\widetilde{h}] = 0$ for any ϑ in the neighborhood of ϑ_0. Additionally, from equation (A.19), we have the following

$$\dot{\Psi}(\vartheta_0)[\widetilde{h}] = P\dot{\Phi}(\vartheta_0)[\widetilde{h}] = \sum_{j=1}^q P\dot{\Phi}^*(\vartheta_0)[\widetilde{h}_j^*] = -\sum_{j=1}^q P\{\Phi^*(\vartheta_0)[\widetilde{h}_j^*]\}^2, \tag{A.20}$$

where the last equality holds due to the loglikelihood property. Hence, $\dot{\Psi}(\vartheta_0)[\widetilde{h}] = 0$ implies that, for each j $(j = 1, \ldots, q)$, $\Phi^*(\vartheta_0)[\widetilde{h}_j^*] = 0$ with probability 1. Finally, we show that $\Phi^*(\vartheta_0)[\widetilde{h}_j^*] = 0$ $(j = 1, \ldots, q)$ implies that $h_0 = 0$ and $h(t) = (h_1(t), \ldots, h_q(t)) = 0$ for any t in $[0, \tau]$. We use Condition 3 and choose $\Delta = 1$. By (A.16) and (A.17), $\Phi^*(\vartheta_0)[\widetilde{h}_1^*] = 0$, where $\widetilde{h}_1^*(t) = (h_0/q, h_1^*(t))$ with $h_1^*(t) = (h_1(t)X_1(t), \ldots, h_1(t)X_1(t))$, implies the following equations:

$$h_0^\top \left(\left[\frac{G''\{\rho_0(T;\vartheta_0)\}}{G'\{\rho_0(T;\vartheta_0)\}} - G'\{\rho_0(T;\vartheta_0)\} \right] \rho_1(T;\vartheta_0) + Z(T) \right) = 0 \quad \text{for any } T,$$

and

$$\sum_{j=1}^q \int_0^T \left[\frac{G''\{\rho_0(T;\vartheta_0)\}}{G'\{\rho_0(T;\vartheta_0)\}} - G'\{\rho_0(T;\vartheta_0)\} \right] e^{\beta_0^\top Z(s)} h_1(s) X_1(s) X_j(s) dA_{0j}(s)$$
$$+ \sum_{j=1}^q \frac{h_1(T)X_1(T)X_j(T)A_{0j}\{T\}}{X^\top(T)A_0\{T\}} = 0 \quad \text{for any } T.$$

Hence, we obtain that $h_0 = 0$ and

$$\begin{aligned} \int_0^T & \left[\frac{G''\{\rho_0(T;\vartheta_0)\}}{G'\{\rho_0(T;\vartheta_0)\}} - G'\{\rho_0(T;\vartheta_0)\} \right] e^{\beta_0^\top Z(s)} h_1(s) X_1(s) X^\top(s) dA_0(s) \\ & + h_1(T)X_1(T) = 0, \end{aligned} \tag{A.21}$$

which is a linear Volterra integral equation with a unique solution (Polyanin and Manzhirov, 2008). Note that (A.21) holds for any T. Hence, $h_1(t) = 0$ for any $t \in [0, \tau]$ since $X(\cdot)$ does not have any component that is identically zero for all $t \in [0, \tau]$. 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, $\dot{\Psi}(\vartheta_0)[\widetilde{h}]$ is continuously invertible. $\square$

Lemma A.4. *Under Conditions 1–5, the map $\dot{\Psi}_{\vartheta_0,n}$ is invertible for larger enough n.*

Proof of Lemma A.4. The expressions of $\dot{\Psi}_{\vartheta}$ are given in Lemma A.2 and Section A.3.5 below; see (A.12), (A.13), (A.14), (A.23), and (A.24). Following the similar steps as in the proof of Lemma A.1, we can show that

$$\dot{\Psi}_{\vartheta}(\vartheta^*)[\widetilde{h}] - \dot{\Psi}_{\vartheta,n}(\vartheta^*)[\widetilde{h}] = o_p(1) \tag{A.22}$$

uniformly in $(\vartheta, \vartheta^*, \widetilde{h})$ in $B_\delta(\beta_0, A_0) \times (\mathbb{R}^d \times BV^q[0,\tau]) \times (\mathbb{R}^d \times BV_1^q[0,\tau])$ for some $\delta > 0$. By Lemma A.3, the map $\dot{\Psi}_{\vartheta_0}$ is invertible. Following Lemma 6.16 in Kosorok (2008), there exists a constant $c_1 > 0$ such that $\|\dot{\Psi}_{\vartheta_0}(\vartheta - \vartheta_0)\| \geq c_1\|\vartheta - \vartheta_0\|$ for all ϑ in $\mathbb{R}^d \times BV^q[0,\tau]$. Combining it with (A.22), there exists a positive constant c_2 such that

$$\begin{aligned}\left\|\frac{\dot{\Psi}_{\vartheta_0,n}(\vartheta - \vartheta_0)}{\|\vartheta - \vartheta_0\|}\right\| &= \left\|\dot{\Psi}_{\vartheta_0,n}\left(\frac{\vartheta - \vartheta_0}{\|\vartheta - \vartheta_0\|}\right)\right\| \\ &= \left\|\dot{\Psi}_{\vartheta_0}\left(\frac{\vartheta - \vartheta_0}{\|\vartheta - \vartheta_0\|}\right) + o_p(1)\right\| \geq c_1 + o_p(1) \geq c_2,\end{aligned}$$

as $n \to \infty$ for any ϑ in $\mathbb{R}^d \times BV^q[0,\tau]$. Thus, $\|\dot{\Psi}_{\vartheta_0,n}(\vartheta - \vartheta_0)\| \geq c_2\|\vartheta - \vartheta_0\|$ as $n \to \infty$ for any ϑ in $\mathbb{R}^d \times BV^q[0,\tau]$. By applying Lemma 6.16 (Kosorok, 2008) again, $\dot{\Psi}_{\vartheta_0,n}$ is invertible for larger enough n. □

A.3.5. Additional details on the Fréchet derivative map for Lemma A.2

This subsection provides additional details on the Fréchet derivatives used in Lemma A.2 and its proof. We provide the calculations of the Fréchet derivative of $\Psi(\vartheta) = P\Phi(\beta, A)$. The Fréchet derivative of $\Psi(\vartheta)$ at ϑ_0 is given by the map

$$\begin{aligned}\dot{\Psi}_{\vartheta_0}(\vartheta - \vartheta_0) &= \left.\frac{d\Psi(\vartheta_0 + \eta(\vartheta - \vartheta_0))}{d\eta}\right|_{\eta=0} \\ &= Q_1^\top[\widetilde{h}](\beta - \beta_0) + \int_0^\tau Q_2[\widetilde{h}](t)d(A(t) - A_0(t)),\end{aligned} \tag{A.23}$$

where

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

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 (A.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

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

and

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

We 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 $g_1'(x) = G''(x) - G^{(3)}(x)/G'(x) + \{G''(x)/G'(x)\}^2$, where $G^{(3)}(x)$ is the thrice derivative of $G(x)$ with respect to x. Let $a^{\otimes 2} = aa^\top$ for any column vector a. Since β belongs to the Euclidean space $\mathbb{R}^d$, B_1 can be easily obtained by taking the derivative of $P\Phi_1(\vartheta)$ with respect to β, i.e.,

$$\begin{aligned}
B_1 &= \left.\frac{\partial P\Phi_1(\vartheta)}{\partial \beta^\top}\right|_{\vartheta=\vartheta_0} \\
&= -E\bigg[\Delta g_1\{\rho_0(T;\vartheta_0)\}\int_0^T e^{\beta_0^\top Z(t)}\{Z(t)\}^{\otimes 2}X^\top(t)dA_0(t) \\
&\qquad + \Delta g_1'\{\rho_0(T;\vartheta_0)\}\{\rho_1(T;\vartheta_0)\}^{\otimes 2}\bigg] \\
&\quad + E\bigg[(1-\Delta)\bigg\{\int_0^\tau I(L<t\le R)I(R<\infty)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\bigg\} \\
&\qquad \times M_1^\top(L,R;\vartheta_0)\bigg] \\
&\quad + E\bigg[(1-\Delta)H_1(L,R;\vartheta_0)\int_0^\tau I(L<t\le R)I(R<\infty)e^{\beta_0^\top Z(t)}\{Z(t)\}^{\otimes 2} \\
&\qquad \times X^\top(t)dA_0(t)\bigg] \\
&\quad - E\bigg[(1-\Delta)\bigg\{\int_0^\tau I(t\le R^*)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\bigg\}M_2^\top(L,R;\vartheta_0)\bigg] \\
&\quad - E\bigg[(1-\Delta)H_2(L,R;\vartheta_0)\int_0^\tau I(t\le R^*)e^{\beta_0^\top Z(t)}\{Z(t)\}^{\otimes 2}X^\top(t)dA_0(t)\bigg],
\end{aligned} \tag{A.25}$$

where

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

and

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

Moreover, we calculate

$$
\begin{aligned}
&\frac{\partial P\Phi_2(\vartheta)[h]}{\partial \beta}\bigg|_{\vartheta=\vartheta_0} \\
&= -\sum_{j=1}^{q} E\bigg[\Delta g_1[\rho_0(T;\vartheta_0)] \int_0^T e^{\beta_0^\top Z(t)} X_j(t) h_j(t) Z(t) X^\top(t) dA_0(t) \bigg] \\
&\quad - \sum_{j=1}^{q} E\bigg[\Delta g_1'\{\rho_0(T;\vartheta_0)\} \bigg\{ \int_0^T e^{\beta_0^\top Z(t)} X_j(t) h_j(t) X^\top(t) dA_0(t) \bigg\} \rho_1(T;\vartheta_0) \bigg] \\
&\quad + \sum_{j=1}^{q} E\bigg[(1-\Delta) \bigg\{ \int_0^\tau I(L<t\le R) I(R<\infty) e^{\beta_0^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA_0(t) \bigg\} \\
&\qquad\qquad \times M_1(L, R; \vartheta_0) \bigg] \\
&\quad + \sum_{j=1}^{q} E\bigg[(1-\Delta) H_1(L, R; \vartheta_0) \int_0^\tau I(L<t\le R) I(R<\infty) e^{\beta_0^\top Z(t)} h_j(t) \\
&\qquad\qquad \times X_j(t) Z(t) X^\top(t) dA_0(t) \bigg] \\
&\quad - \sum_{j=1}^{q} E\bigg[(1-\Delta) \bigg\{ \int_0^\tau I(t\le R^*) e^{\beta_0^\top Z(t)} h_j(t) X_j(t) X^\top(t) dA_0(t) \bigg\} M_2(L, R; \vartheta_0) \bigg] \\
&\quad - \sum_{j=1}^{q} E\bigg[(1-\Delta) H_2(L, R; \vartheta_0) \int_0^\tau I(t\le R^*) e^{\beta_0^\top Z(t)} h_j(t) X_j(t) Z(t) X^\top(t) dA_0(t) \bigg].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
B_{2,j}(t) &= -E\left[\Delta g_1[\rho_0(T;\vartheta_0)] I(t\le T) e^{\beta_0^\top Z(t)} X_j(t) Z(t) X^\top(t) \right] \\
&\quad - E\left[\Delta g_1'\{\rho_0(T;\vartheta_0)\} I(t\le T) e^{\beta_0^\top Z(t)} X_j(t) \rho_1(T;\vartheta_0) X^\top(t) \right] \\
&\quad + E\left[(1-\Delta) I(L<t\le R) I(R<\infty) e^{\beta_0^\top Z(t)} X_j(t) M_1(L, R; \vartheta_0) X^\top(t) \right] \\
&\quad + E\left[(1-\Delta) H_1(L, R; \vartheta_0) I(L<t\le R) I(R<\infty) e^{\beta_0^\top Z(t)} X_j(t) Z(t) X^\top(t) \right]
\end{aligned}
$$

$$-E\left[(1-\Delta)I(t\leq R^*)e^{\beta_0^\top Z(t)}X_j(t)M_2(L,R;\vartheta_0)X^\top(t)\right]$$
$$-E\left[(1-\Delta)H_2(L,R;\vartheta_0)I(t\leq R^*)e^{\beta_0^\top Z(t)}X_j(t)Z(t)X^\top(t)\right]. \qquad \text{(A.26)}$$

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

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

where

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

and

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

Using the same technique, we obtain

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

where

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

Thus,

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

Hence,

$$
\begin{aligned}
B_3(t) = &- E\Big[\Delta g_1'\{\rho_0(T;\vartheta_0)\}\rho_1(T;\vartheta_0)I(t\le T)e^{\beta_0^\top Z(t)}X^\top(t)\Big] \\
&- E\Big[\Delta g_1\{\rho_0(T;\vartheta_0)\}I(t\le T)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)\Big] \\
&+ E\Big[(1-\Delta)I(R<\infty)\{\rho_1(R;\vartheta_0)-\rho_1(L;\vartheta_0)\} \\
&\qquad \times M_3(L,R;\vartheta_0)I(t\le L)e^{\beta_0^\top Z(t)}X^\top(t)\Big] \\
&- E\Big[(1-\Delta)I(R<\infty)\{\rho_1(R;\vartheta_0)-\rho_1(L;\vartheta_0)\} \\
&\qquad \times M_4(L,R;\vartheta_0)I(t\le R)e^{\beta_0^\top Z(t)}X^\top(t)\Big] \\
&+ E\Big[(1-\Delta)H_1(L,R;\vartheta_0)I(L<t\le R)I(R<\infty)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)\Big] \\
&- E\Big[(1-\Delta)\rho_1(R^*;\vartheta_0)\{M_3(L,R;\vartheta_0)-M_4(L,R;\vartheta_0)\} \\
&\qquad \times I(t\le L)e^{\beta_0^\top Z(t)}X^\top(t)\Big] \\
&+ E\Big[(1-\Delta)\rho_1(R^*;\vartheta_0)M_5(L,R;\vartheta_0)I(t\le R)e^{\beta_0^\top Z(t)}X^\top(t)] \\
&- E\Big[(1-\Delta)H_2(L,R;\vartheta_0)I(t\le R^*)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)\Big].
\end{aligned}
\tag{A.27}
$$

Lastly, since $\rho_3(t;\vartheta)[h] = \sum_{j=1}^q \int_0^t e^{\beta^\top Z(s)}h_j(s)X_j(s)X^\top(s)dA(s)$, we can obtain

$$
\left.\frac{\partial \Phi_2(\beta, A+\eta A^*)[h]}{\partial \eta}\right|_{\eta=0}
$$

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

Hence,

$$
\left.\frac{\partial P\Phi_2(\beta,A+\eta A^*)[h]}{\partial\eta}\right|_{\eta=0,A^*=A-A_0,\vartheta=\vartheta_0} = \int_0^\tau B_4[h](t)d(A-A_0)(t),
$$

where

$$
\begin{aligned}
&B_4[h](t)\\
&= -E\Big[\Delta g_1'\{\rho_0(T;\vartheta_0)\}\rho_3(T;\vartheta_0)[h]I(t\le T)e^{\beta_0^\top Z(t)}X^\top(t)\Big]\\
&\quad - \sum_{j=1}^q E[\Delta g_1\{\rho_0(T;\vartheta_0)\}I(t\le T)h_j(t)e^{\beta_0^\top Z(t)}X_j(t)X^\top(t)\Big]\\
&\quad + E\Big[(1-\Delta)I(R<\infty)\{\rho_3(R;\vartheta_0)[h]-\rho_3(L;\vartheta_0)[h]\}\\
&\qquad \times M_3(L,R;\vartheta_0)I(t\le L)e^{\beta_0^\top Z(t)}X^\top(t)\Big]\\
&\quad - E\Big[(1-\Delta)I(R<\infty)\{\rho_3(R;\vartheta_0)[h]-\rho_3(L;\vartheta_0)[h]\}\\
&\qquad \times M_4(L,R;\vartheta_0)I(t\le R)e^{\beta_0^\top Z(t)}X^\top(t)\Big]\\
&\quad + \sum_{j=1}^q E\Big[(1-\Delta)I(R<\infty)H_1(L,R;\vartheta_0)I(L<t\le R)e^{\beta_0^\top Z(t)}h_j(t)X_j(t)X^\top(t)\Big]\\
&\quad - E\Big[(1-\Delta)\rho_3(R^*;\vartheta_0)[h]\{M_3(L,R;\vartheta_0)-M_4(L,R;\vartheta_0)\}I(t\le L)e^{\beta_0^\top Z(t)}X^\top(t)\Big]\\
&\quad + E\Big[(1-\Delta)\rho_3(R^*;\vartheta_0)[h]M_5(L,R;\vartheta_0)I(t\le R)e^{\beta_0^\top Z(t)}X^\top(t)\Big]\\
&\quad - \sum_{j=1}^q E\Big[(1-\Delta)H_2(L,R;\vartheta_0)I(t\le R^*)e^{\beta_0^\top Z(t)}h_j(t)X_j(t)X^\top(t)\Big]. \qquad \text{(A.28)}
\end{aligned}
$$

Table A.1
Estimation results for the regression parameter β under Scenario 2. Bias, SE, SEE, and CP stand, respectively, for the bias, empirical standard error, mean of the estimated standard errors, and empirical coverage percentage of the 95% confidence interval. Each entry is based on 1000 simulations and 1000 bootstraps.

			$\beta_1 = 0.5$				$\beta_2 = -0.5$			
r	n	γ	Bias	SE	SEE	CP	Bias	SE	SEE	CP
0	200	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	0.25	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	0.25	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
0.5	200	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	0.25	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	0.25	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
1	200	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	0.25	−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	0.25	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

A.4. Additional simulation results

In this subsection, we provide additional simulation results for Scenarios 1–3 described in Section 6. Specifically, Tables A.1 and A.2 showcase the estimation

TABLE A.2

Estimation results for the regression parameter β under Scenario 3. Bias, SE, SEE, and CP stand, respectively, for the bias, empirical standard error, mean of the estimated standard errors, and empirical coverage percentage of the 95% confidence interval. Each entry is based on 1000 simulations and 1000 bootstraps.

			$\beta_1 = 0.5$				$\beta_2 = -0.5$			
r	n	γ	Bias	SE	SEE	CP	Bias	SE	SEE	CP
0	200	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	0.25	−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	0.25	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
0.5	200	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	0.25	−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	0.25	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
1	200	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	0.25	−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	0.25	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

results of β for Scenarios 2 and 3. Figures A.1 and A.2 display the estimation results for the cumulative regression functions corresponding to Scenarios 2–3, respectively.

Moreover, we considered Scenario 4 to compare the proposed approach, which

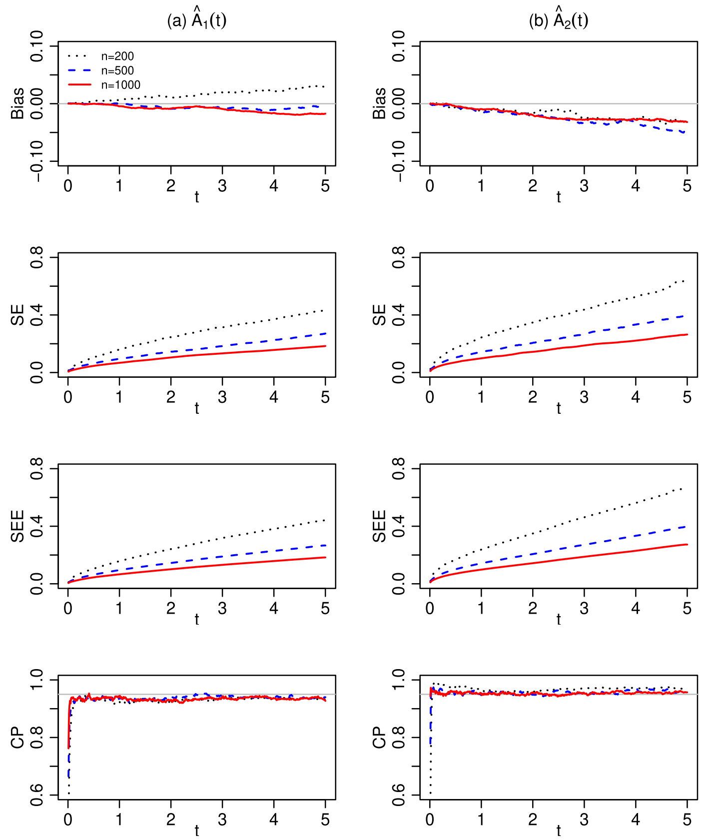

Figure A.1. *Estimation results for (a) $A_1(t) = \log(1+t/2)$ and (b) $A_2(t) = 0.1t$ in Scenario 2 with $\gamma = 0.5$, under the logarithmic transformation $G(x) = r^{-1}\log(1+rx)$ with $r = 0.5$. The dotted, dashed and solid lines are for data sets with $n = 200, 500, 1000$, respectively. Bias, SE, SEE, and CP stand, respectively, for the bias, empirical standard error, mean of the estimated standard errors, and empirical coverage probability of the 95% confidence interval. The figures are based on 1000 simulations and 1000 bootstraps.*

accounts for different baseline cumulative hazard functions, with the method in Zhou, Sun and Gilbert (2021), which assumes the same baseline cumulative hazard function for all individuals. In particular, in Scenario 4, the failure time T is generated from the following Cox-Aalen transformation model

$$\Lambda(t \mid X, Z) = G\Big\{\int_0^t \exp(\beta_1 Z_1) X^\top dA(s)\Big\}, \tag{A.29}$$

where $\beta_1 = 1$, $Z_1 \sim \text{Ber}(0.5)$ and $X(t) = (1, X_2)^\top$ with $X_2 \sim \text{Ber}(0.4)$. $A_1(t) = 0.16t + 0.0256t^2$ and $A_2(t) = 0.56t - 0.0384t^2$ for $0 \le t \le 5$. The partly interval censoring is generated the same as for Scenarios 1-3 in Section 6 with $\tau = 5$. We fitted model (A.29) by correctly specifying two different baseline cumulative

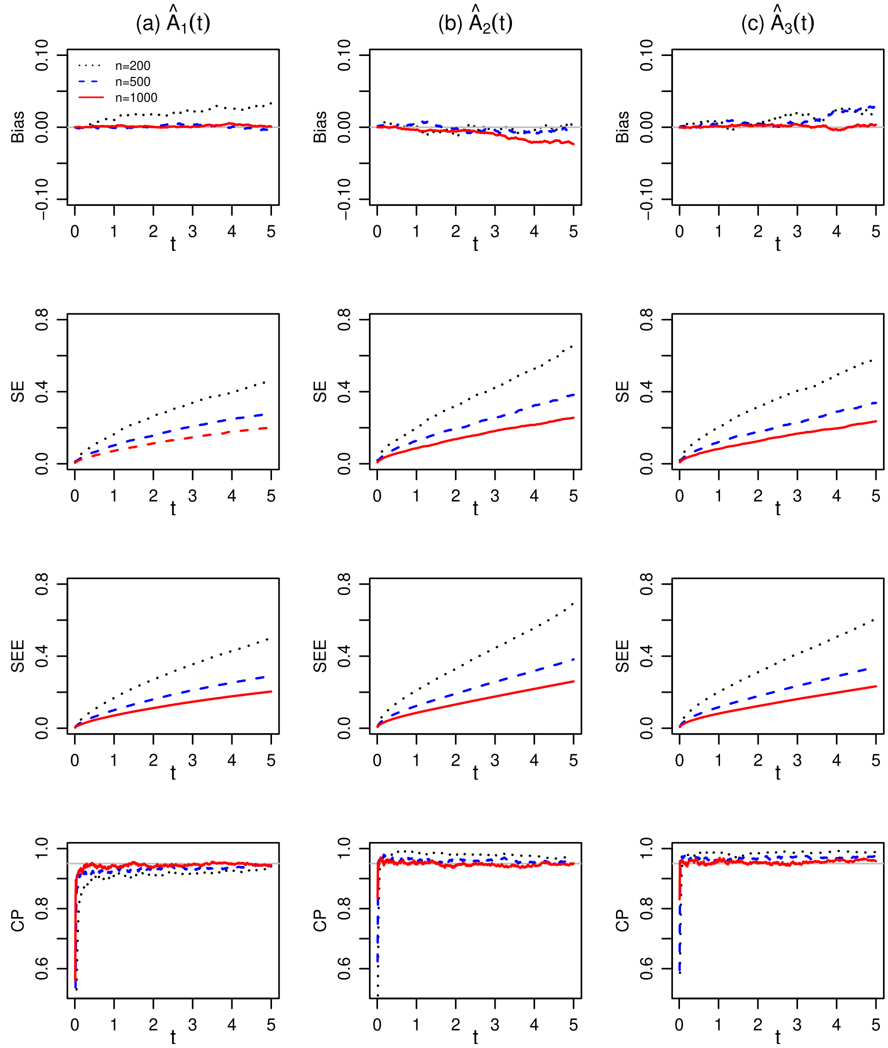

FIGURE A.2. *Estimation results for (a)* $A_1(t) = \log(1+t/2)$*, (b)* $A_2(t) = 0.1t$ *and (c)* $A_3(t) = 0.05t$ *in Scenario 3 with* $\gamma = 0.5$*, under the logarithmic transformation* $G(x) = r^{-1}\log(1+rx)$ *with* $r = 1$*. The dotted, dashed and solid lines are for data sets with* $n = 200, 500, 1000$*, respectively. Bias, SE, SEE, and CP stand, respectively, for the bias, empirical standard error, mean of the estimated standard errors, and empirical coverage probability of the* 95% *confidence interval. The figures are based on 1000 simulations and 1000 bootstraps.*

hazards (varying according to the value of X_2). As a comparison, we analyze the data using the method in Zhou, Sun and Gilbert (2021) that assumes the same baseline cumulative hazard across all individuals. Figures A.3 and A.4 show that ignoring the different cumulative hazard functions, as done in Zhou, Sun and Gilbert (2021), leads to biased predictions of survival probabilities

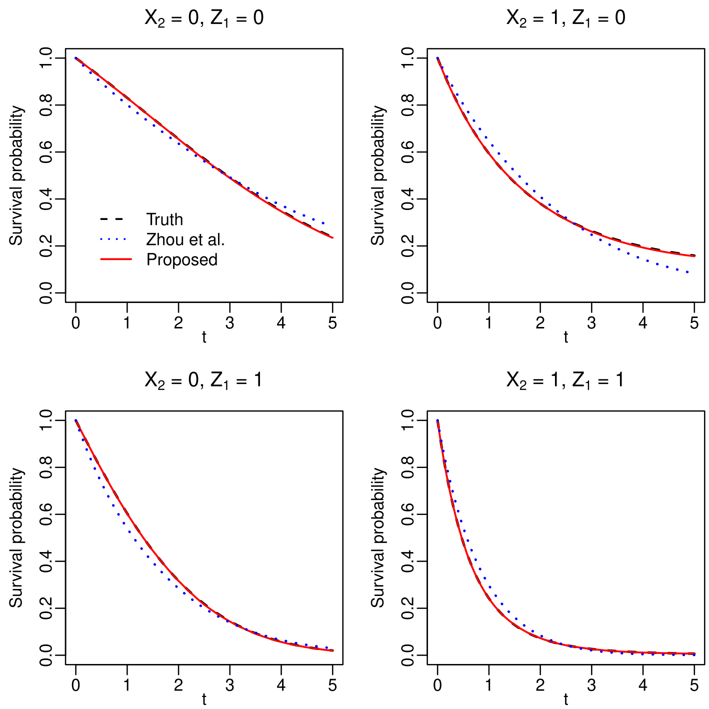

FIGURE A.3. *Predicted survival probabilities under Scenario 4 with* $r = 0$ *and* $n = 500$ *based on the proposed model and the method in Zhou, Sun and Gilbert (2021), when the proportion of exactly observed failure observations among those that are not right-censored is 50%, i.e.,* $\gamma = 0.5$.

and cumulative hazards, whereas the proposed method accurately captures the survival probabilities and cumulative hazards. Therefore, the proposed method is more adept at capturing complex scenarios.

Acknowledgments

The authors thank the editor for handling the paper and the associate editor and two referees for their comments that greatly improved this article.

Funding

This research was partially supported by the National Institutes of Health, National Institute of Allergy and Infectious Diseases [grant numbers UM1 AI68635 and R37 AI054165]. The research of Xi Ning was supported, in part, by the 2022 Graduate School Summer Fellowship Program from the University of North Carolina at Charlotte (UNC Charlotte). The research of Yanqing Sun was partially supported by the National Science Foundation [grant number DMS-1915829]

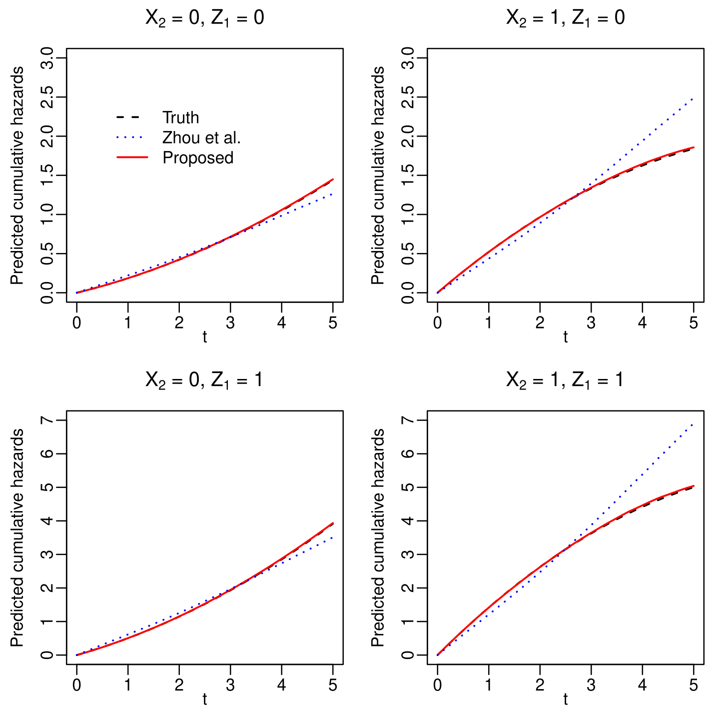

Figure A.4. *Predicted cumulative hazards under Scenario 4 with $r = 0$ and $n = 500$ based on the proposed model and the method in Zhou, Sun and Gilbert (2021), when the proportion of exactly observed failure observations among those that are not right-censored is 50%, i.e.,* $\gamma = 0.5$.

and the Reassignment of Duties fund provided by UNC Charlotte. Yinghao Pan's work was partially supported by the National Science Foundation [grant number DMS-2316868] and funds provided by UNC Charlotte.

References

Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical statistics and probability theory* 1–25. Springer. MR0577267

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine* **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* 1100–1120. MR0673646

Boruvka, A. and Cook, R. J. (2015). A Cox–Aalen Model for Interval-censored Data. *Scandinavian Journal of Statistics* **42** 414–426. MR3345112

CAI, T. and BETENSKY, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics* **59** 570–579. MR2004262

CDC (1987). Revision of the CDC surveillance case definition for acquired immunodeficiency syndrome. *MMWR* **36** 3S–15S.

CHEN, L., LIN, D. and ZENG, D. (2012). Checking semiparametric transformation models with censored data. *Biostatistics* **13** 18–31.

COX, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **34** 187–202. MR0341758

DABROWSKA, D. M. and DOKSUM, K. A. (1988). Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics* 1–23. MR0967953

DUDLEY, R. M. (2002). *Real analysis and probability*. Cambridge University Press. MR1932358

ELASHOFF, M. and RYAN, L. (2004). An EM algorithm for estimating equations. *Journal of Computational and Graphical Statistics* **13** 48–65. MR2044870

FINE, J., YING, Z. and WEI, L. (1998). On the linear transformation model for censored data. *Biometrika* **85** 980–986.

FINKELSTEIN, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* 845–854. MR0872963

GAO, F., ZENG, D. and LIN, D.-Y. (2017). Semiparametric estimation of the accelerated failure time model with partly interval-censored data. *Biometrics* **73** 1161–1168. MR3744530

GAO, F., ZENG, D. and LIN, D.-Y. (2018). Semiparametric regression analysis of interval-censored data with informative dropout. *Biometrics* **74** 1213–1222. MR3908139

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** 1081–1090.

HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics* **24** 540–568. MR1394975

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** 960–967. MR1482126

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** 114–129.

KALBFLEISCH, J. D. and PRENTICE, R. L. (2002). *The statistical analysis of failure data*. Wiley. MR1924807

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** 489–502. MR1983760

KOSOROK, M. R. (2008). *Introduction to empirical processes and semiparamet-*

ric inference. Springer. MR2724368

LI, J. and MA, J. (2019). Maximum penalized likelihood estimation of additive hazards models with partly interval censoring. *Computational Statistics & Data Analysis* **137** 170–180. MR3922200

LIN, D. Y. and YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81** 61–71. MR1279656

LIN, D. and YING, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The Annals of Statistics* 1712–1734. MR1370304

LIU, X. and ZENG, D. (2013). Variable selection in semiparametric transformation models for right-censored data. *Biometrika* **100** 859–876. MR3142337

MAO, L. and LIN, D. (2017). Efficient estimation of semiparametric transformation models for the cumulative incidence of competing risks. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* **79** 573. MR3611760

MARTINUSSEN, T. and SCHEIKE, T. H. (2002). A flexible additive multiplicative hazard model. *Biometrika* **89** 283–298. MR1913959

NING, X., PAN, Y., SUN, Y. and GILBERT, P. B. (2023). A semiparametic Cox–Aalen transformation model with censored data. *Biometrics* **79** 3111–3125. MR4680708

ORTEGA, J. M. (1972). *Numerical analysis: a second course*. SIAM. MR0403154

ORTEGA, J. M. and RHEINBOLDT, W. C. (1970). *Iterative solution of nonlinear equations in several variables*. SIAM. MR0273810

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** 3192–3204. MR4156848

POLYANIN, A. D. and MANZHIROV, A. V. (2008). *Handbook of integral equations*. CRC Press. MR2404728

QU, L. and SUN, L. (2019). The Cox–Aalen model for recurrent-event data with a dependent terminal event. *Statistica Neerlandica* **73** 234–255. MR3942103

RUDIN, W. (1973). *Functional Analysis*. McGraw-Hill Inc., New York. MR0365062

SCHEIKE, T. H. and ZHANG, M.-J. (2002). An additive–multiplicative Cox–Aalen regression model. *Scandinavian Journal of Statistics* **29** 75–88. MR1894382

SCHWARTZ, J. T. (1969). *Nonlinear functional analysis* **4**. CRC Press. MR0433481

SHEN, P.-S. and WENG, L. N. (2019). The Cox–Aalen model for left-truncated and mixed interval-censored data. *Statistics* **53** 1152–1167. MR3987595

SUN, J. (2006). *The statistical analysis of interval-censored failure time data*. Springer. MR2065763

VAN DER VAART, A. W. (1998). *Asymptotic statistics*. Cambridge University Press. MR1652247

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media. MR1385671

Wang, C., Jiang, J. and Song, X. (2022). Bayesian transformation models with partly interval-censored data. *Statistics in Medicine* **41** 1263–1279. MR4393423

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** 222–231. MR3500591

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** 414–431. MR3948562

Zeng, D., Cai, J. and Shen, Y. (2006). Semiparametric additive risks model for interval-censored data. *Statistica Sinica* **16** 287–302. MR2256093

Zeng, D. and Lin, D. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93** 627–640. MR2261447

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** 507–564. MR2370068

Zeng, D., Mao, L. and Lin, D. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika* **103** 253–271. MR3509885

Zhang, D., Khalili, A. and Asgharian, M. (2022). Post-model-selection inference in linear regression models: An integrated review. *Statistic Surveys* **16** 86–136. MR4390422

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** 61–70. MR2154998

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* **40** 4376–4394. MR4300119

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** 1496–1512. MR4261862