



Combined estimating equation approaches for the additive hazards model with left-truncated and interval-censored data

Tianyi Lu¹ · Shuwei Li¹ · Liuquan Sun²

Received: 15 April 2022 / Accepted: 1 March 2023 / Published online: 23 March 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Interval-censored failure time data arise commonly in various scientific studies where the failure time of interest is only known to lie in a certain time interval rather than observed exactly. In addition, left truncation on the failure event may occur and can greatly complicate the statistical analysis. In this paper, we investigate regression analysis of left-truncated and interval-censored data with the commonly used additive hazards model. Specifically, we propose a conditional estimating equation approach for the estimation, and further improve its estimation efficiency by combining the conditional estimating equation and the pairwise pseudo-score-based estimating equation that can eliminate the nuisance functions from the marginal likelihood of the truncation times. Asymptotic properties of the proposed estimators are discussed including the consistency and asymptotic normality. Extensive simulation studies are conducted to evaluate the empirical performance of the proposed methods, and suggest that the combined estimating equation approach is obviously more efficient than the conditional estimating equation approach. We then apply the proposed methods to a set of real data for illustration.

Keywords Additive hazards regression · Estimating equation · Interval censoring · Left truncation · Pairwise pseudo-likelihood

1 Introduction

Interval-censored failure time data arise frequently in many scientific studies including clinical trials, epidemiological surveys and cancer screening studies, and due to

✉ Shuwei Li
lishuwstat@163.com ; seslishuw@gzhu.edu.cn

¹ School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China

² Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

the periodical follow-ups in these situations, the failure time of interest cannot be monitored exactly but is only known to fall into a certain time interval. For example, in HIV/AIDS studies, urine samples are intermittently provided by the individuals at their clinical visits to examine the occurrence of the CMV shedding. Therefore, the CMV shedding time is only known to occur in a time interval formed by dates of the last negative urine test and the first positive urine test (Goggins and Finkelstein 2000). Another example of interval-censored data is given by Finkelstein (1986), where the failure event of interest is the occurrence of breast retraction among early breast cancer patients.

In practice, one may encounter several types of interval-censored data. One type of interval-censored data that has been discussed vastly in the literature is called case I interval-censored data or current status data (Sun 1999; Li et al. 2017, 2021; Ma et al. 2015). Such data arise when each individual under study is observed only once at one monitoring time and the failure event of interest is known only to occur before or after this monitoring time. In other words, the failure time of interest is either left- or right-censored rather than observed exactly, and the interval containing the failure time includes either zero or infinity. When there exist two or multiple observation times for each subject, the obtained data are usually referred to as case II or case K interval-censored data depending on the number of observations (Sun 2006). In this situation, the failure time of interest falls into a certain time interval, which may be given by the half-open interval or the interval with finite endpoints.

In many applications, the failure times of interest are often subject to left truncation under the cross-sectional sampling scheme, where only individuals that have not experienced the failure event at the study enrolment are included in the study. Therefore, the collected data suffer from the biased sampling since they are not representative for the whole population under study. For the analysis of left-truncated data, a large number of methods have been proposed under right censoring (Lai and Ying 1991; Wang et al. 1993; Gross and Lai 1996; Shen et al. 2009; Qin and Shen 2010; Huang and Qin 2013; Shen et al. 2017; Wu et al. 2018). For instance, Wang et al. (1993) proposed a partial likelihood-based estimation method for the proportional hazards model; Lai and Ying (1991) developed some rank-based estimators for the censored linear regression; Shen et al. (2009) assumed that the truncation times satisfy the stationary assumption and proposed some unbiased estimating equation approaches to estimate the transformation and accelerated failure time models; Huang and Qin (2013) and Wu et al. (2018) considered the general left-truncated data and proposed the pairwise likelihood-based methods for the additive and proportional hazards models, respectively.

Since interval-censored data are often subject to the left truncation in scientific fields, some methods have already been developed for their analysis (Pan and Chappell 1998, 2002; Kim 2003; Wang et al. 2015a, b; Gao and Chan 2019; Shen et al. 2019; Wang et al. 2021). In particular, for the left-truncated and case I interval-censored data, Kim (2003) and Wang et al. (2015b) proposed the conditional maximum likelihood estimation approaches for the proportional and additive hazards models, respectively. With respect to the left-truncated and general interval-censored data, Pan and Chappell (2002) proposed a marginal likelihood approach as well as a monotone maximum likelihood approach for the proportional hazards model; Wang et al. (2015a) proposed a conditional sieve maximum likelihood estimation method for the additive hazards

model; Shen et al. (2019) developed a conditional likelihood estimation approach for the general transformation models. It is worth noting that the aforementioned methods were established without incorporating the marginal distribution information of the observed truncation times, which may lose some estimation efficiency as anticipated. Corresponding to this, Gao and Chan (2019) proposed a nonparametric maximum likelihood estimation approach for the proportional hazards model under the stationary assumption for the truncation times. Motivated by the work of Huang and Qin (2013); Wang et al. (2021) further proposed a pairwise likelihood estimation approach for the proportional hazards model.

This study aims at investigating regression techniques for the left-truncated and interval-censored data with the additive hazards model. The motivations of considering the additive hazards model are mainly threefold. Firstly, the additive hazards model serves as an essential alternative to the commonly adopted proportional hazards model in survival analysis. This is because that the additive hazards model provides a different perspective of covariate effect on the failure time compared with the proportional hazards model and may be more plausible than the latter in some practical applications (Aranda-Ordaz 1983; Buckley 1984; Lin and Ying 1994; Lin et al. 1998). Secondly, unlike the proportional hazards model, it is easy to derive a tractable estimating equation for the additive hazards model under interval censoring by using the counting process formulations (Lin et al. 1998; Wang et al. 2010). Therefore, one can easily adapt the established estimating equations to accommodate left truncation. Thirdly, under the additive hazards model, the interpretation of the estimated covariate effects can be converted into that of the ratio of the estimated survival probabilities between different groups at a given time, which does not involve the baseline hazard function. Therefore, even without estimating the baseline hazard function, the additive hazards model with only the estimated regression parameters still has appealing interpretations as commented in the real data analysis of Sect. 6.

A major difficulty in estimating the additive hazards model with left-truncated and interval-censored data is how to develop a reliable and easily implemented estimation procedure. The first contribution of this work is to develop a simple conditional estimating equation method through modifying the risk sets of the estimating equation proposed by Lin et al. (1998) for analyzing the left-truncated and case I interval-censored data. Since the conditional estimating equation method is found to be very inefficient, we further improve its estimation efficiency by developing a combined estimating equation approach with the use of the marginal likelihood of the truncation times. A desirable feature of the two proposed estimating equations is that they have simple and tractable form, and do not involve any nuisance functions including the baseline hazard function in the additive hazards model and the density function of the truncation time. Therefore, they can be readily solved by the simple and standard optimization procedure, such as the Newton–Raphson algorithm.

Since the left-truncated and general interval-censored data are also frequently encountered in many applications and have complicated structure, the second contribution of the work is to develop two easily implemented methods to handle such data. Specifically, we first derive a conditional estimating equation approach by utilizing the working independence strategy for interval-censored data (Betensky et al. 2001; Zhu et al. 2008). Then, as above, a combined estimating equation approach by allow-

ing for the marginal likelihood of the truncation times is further proposed to improve the estimation efficiency of the former method. The numerical results obtained from extensive simulation studies imply that the proposed estimation methods can give reliable and stable performance. In particular, by utilizing the distribution information of the truncation times, the combined estimating equation approach can greatly improve the estimation efficiency of the conditional estimating equation method.

The rest of this article is arranged as follows. In Sect. 2, we introduce some notation, the assumed model as well as the data structure. In Sect. 3, we first consider the left-truncated and case I interval-censored data and develop the conditional and combined estimating equation approaches for the additive hazards model. In Sect. 4, we generalize the proposed estimation methods to the situation of left-truncated and general interval-censored data. In Sect. 5, we conduct extensive simulation studies to investigate the finite-sample performance of the proposed methods, followed by an application to a real dataset in Sect. 6. Some discussions and concluding remarks are given in Sect. 7.

2 Notation, model and data structure

Let T^* denote the underlying failure time of interest (i.e. the time to the failure event of interest), and \mathbf{Z}^* be a p -dimensional vector of covariates in the target population. Let A^* denote the underlying left truncation time (i.e. the time to the study enrolment) that is assumed to be independent of T^* . When T^* suffers from left truncation by A^* , we know that only subjects with $T^* \geq A^*$ are collected or sampled in the study. Let T , A , and \mathbf{Z} be the failure time, truncation time and covariate vector of a sampled subject (i.e. satisfying $T \geq A$), respectively. Then (T, A, \mathbf{Z}) has the same joint distribution as (T^*, A^*, \mathbf{Z}^*) given that $T^* \geq A^*$. Here it should be noted that the aforementioned independent assumption between T^* and A^* does not imply the independence between T and A since subject with larger A would have larger T under the left truncation scheme (Wu et al. 2018).

To describe the effects of \mathbf{Z}^* on T^* , we assume that T^* follows the additive hazards model with the conditional hazard function

$$\lambda(t | \mathbf{Z}^*) = \lambda(t) + \boldsymbol{\beta}^\top \mathbf{Z}^*, \tag{1}$$

where $\lambda(t)$ is the unknown baseline hazard function and $\boldsymbol{\beta}$ is a p -dimensional vector of regression parameters. Thus $\Lambda(t) = \int_0^t \lambda(s) ds$ is the cumulative baseline hazard function of T^* . Let f and S denote the density and survival functions of T^* , respectively, and let g be the density function of A^* . Then the joint density function (A, T) at (a, t) can be expressed as

$$h(a, t) = \frac{g(a)f(t)}{\int_0^\infty S(u)g(u)du} = \frac{f(t)}{S(a)} \times \frac{S(a)g(a)}{\int_0^\infty S(u)g(u)du} \quad (0 \leq a \leq t)$$

where $f(t)/S(a)$ is the conditional density of T given A and $S(a)g(a)/\int_0^\infty S(u)g(u)du$ is the marginal density of A (Huang and Qin 2013).

We first consider the situation of left-truncated and case I interval-censored data, in which each subject is examined only once for the occurrence of the event of interest at the observation or censoring time. Let C denote the observation or censoring time after the study enrolment, and assume that C is conditional independent of (A, T) given Z . Under case I interval censoring mechanism, T cannot be observed directly but is only known to be smaller than $A + C$ or larger than $A + C$. Define $\delta = I(A + C \leq T)$, that is, T is right-censored if $\delta = 1$ and left-censored otherwise. Then the observed data include n i.i.d. replicates of $\{C, \delta, A, Z\}$, namely, $\{C_i, \delta_i, A_i, Z_i; i = 1, \dots, n\}$.

3 Estimation procedure

In this section, we first develop a conditional estimating equation approach by adapting the estimating equation proposed by Lin et al. (1998) to account for the left truncation, and then develop a combined estimating equation approach to improve the estimation efficiency of the former method.

3.1 Conditional estimating equation approach

For each subject i , let $\tilde{C}_i = A_i + C_i$ and define the counting process $N_i(t) = \delta_i I(\tilde{C}_i \leq t)$, which jumps by one at time t when (i) subject i is monitored at time t , $\tilde{C}_i = t$, and (ii) subject i is still found to be failure-free at this time, that is, $T_i \geq t$. The process $N_i(t)$ can be regarded as right censoring at time t when subject i is monitored, $\tilde{C}_i = t$, and found to have experienced the failure at this time, $\delta_i = 0$, since this process always takes the value of 0 under this situation. By following the arguments given in Lin et al. (1998), we can derive the intensity model of $N_i(t)$, which takes the form

$$dH_i(t) = R_i(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{Z}_i t\} dH(t), \quad (2)$$

where $R_i(t) = I(A_i \leq t \leq \tilde{C}_i)$ is the at-risk process under left truncation, $dH(t) = \exp\{-\Lambda(t)\} d\Lambda_c(t)$, and $\Lambda_c(t)$ is the cumulative hazard function of \tilde{C}_i .

The above expression (2) suggests that

$$M_i(t) = N_i(t) - \int_0^t R_i(s) \exp\{-\boldsymbol{\beta}^\top \mathbf{Z}_i s\} dH(s) \quad (i = 1, \dots, n)$$

are martingales with respect to the σ -filtration: $\sigma\{N_i(s), R_i(s), \mathbf{Z}_i; s \leq t, i = 1, \dots, n\}$. Thus, one can make inference about $\boldsymbol{\beta}$ by applying the partial likelihood approach to (2). Specifically, the partial likelihood score function $\boldsymbol{\beta}$ is

$$\Phi(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_i(t), \quad (3)$$

and the observed information matrix for β is

$$I(\beta) = \sum_{i=1}^n \int_0^\tau \left\{ \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - \frac{S^{(1)}(\beta, t)^{\otimes 2}}{S^{(0)}(\beta, t)^2} \right\} dN_i(t), \tag{4}$$

where

$$S^{(k)}(\beta, t) = \sum_{j=1}^n R_j(t) \exp\{-\beta^\top \mathbf{Z}_j t\} (\mathbf{Z}_j t)^{\otimes k},$$

$\mathbf{a}^{\otimes 0} = 1, \mathbf{a}^{\otimes 1} = \mathbf{a}$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^\top$ for a column vector \mathbf{a} , and τ is a prespecified constant such that $P(A + C \geq \tau) > 0$.

After normalising, the conditional estimating equation for β is given by

$$U(\beta) = \frac{1}{n} \Phi(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right\} dN_i(t) = 0. \tag{5}$$

Solving $U(\beta) = 0$ yields the estimate of β , which is denoted by $\hat{\beta}_U$. Let β_0 be the true value of β . The following theorem shows the asymptotic property of $\hat{\beta}_U$, and the proof will be presented in the ‘‘Appendix A’’.

Theorem 1 *Under Conditions (C1)–(C3) given in the ‘‘Appendix A’’, we have that, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\beta}_U - \beta_0)$ converges in distribution to a zero-mean normal random vector with covariance matrix $\Sigma(\beta_0)^{-1}$, where $\Sigma(\beta_0)$ is the limit of $n^{-1}I(\beta_0)$ that can be consistently estimated by $n^{-1}I(\hat{\beta}_U)$.*

Note that the conditional estimating function proposed above is anticipated to loss some estimation efficiency since it does not incorporate the distribution information of the truncation times. To improve the efficiency, in what follows, we will provide a combined estimating equation approach by combining the conditional estimating equation and the pseudo-score-based estimating equation obtained from the pairwise information of the truncation times.

3.2 Combined estimating equation approach

To allow for the information provided by the truncation times, we adopt the pairwise likelihood method given in Kalbfleisch (1978) and Liang and Qin (2000). As in Huang and Qin (2013); Wu et al. (2018) and others, assume that the truncation time A does not depend on \mathbf{Z} and is not degenerate. Note that all observed truncation times satisfy $A < \tau$ in practical situations since we often set τ to be the maximum of the observed observation times. Note that, for $i < j$, by conditioning on having observed (A_i, A_j) but without knowing the order of A_i and A_j , the pairwise pseudo-likelihood of the observed (A_i, A_j) conditional on $A_i \leq \tau, A_j \leq \tau$ and $(\mathbf{Z}_i, \mathbf{Z}_j)$ is

$$\frac{\frac{S(A_i|Z_i)g(A_i)}{\int_0^\tau S(u|Z_i)g(u)du} \times \frac{S(A_j|Z_j)g(A_j)}{\int_0^\tau S(u|Z_j)g(u)du}}{\frac{S(A_i|Z_i)g(A_i)}{\int_0^\tau S(u|Z_i)g(u)du} \times \frac{S(A_j|Z_j)g(A_j)}{\int_0^\tau S(u|Z_j)g(u)du} + \frac{S(A_i|Z_j)g(A_i)}{\int_0^\tau S(u|Z_j)g(u)du} \times \frac{S(A_j|Z_i)g(A_j)}{\int_0^\tau S(u|Z_i)g(u)du}},$$

which equals

$$\frac{\exp(-\beta^\top Z_i A_i - \beta^\top Z_j A_j)}{\exp(-\beta^\top Z_i A_i - \beta^\top Z_j A_j) + \exp(-\beta^\top Z_j A_i - \beta^\top Z_i A_j)} = \frac{1}{1 + \exp\{\beta^\top (Z_i - Z_j)(A_i - A_j)\}}.$$

It is worth noting that the pairwise pseudo likelihood depends on the regression parameter β of the assumed additive hazards model (1) and may render some additional information to improve the estimation efficiency. Moreover, as a by-product, the pairwise likelihood method used above can enable us to eliminate the nuisance functions in the marginal distribution of the truncation times, which can greatly facilitate the subsequent estimation procedure. Define $\rho_{ij} = \rho(A_i, Z_i, A_j, Z_j) = (Z_i - Z_j)(A_i - A_j)$. Then the logarithm of the pairwise pseudo-likelihood and the normalised score function of β are given by

$$\sum_{1 \leq i < j \leq n} -\log \left\{ 1 + \exp(\beta^\top \rho_{ij}) \right\},$$

and

$$\psi(\beta) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \psi_{ij}(\beta), \tag{6}$$

respectively, where

$$\psi_{ij}(\beta) = \psi_{ij}(\beta; A_i, Z_i, A_j, Z_j) = \frac{-\rho_{ij}}{1 + \exp(-\beta^\top \rho_{ij})}.$$

Since $U(\beta)$ and $\psi(\beta)$ both can give consistent estimate of β , the combined estimating equation proposed is thus defined as

$$\epsilon(\beta) = U(\beta) + \psi(\beta) = 0. \tag{7}$$

Denote by $\hat{\beta}_\epsilon$ the estimate of β that solves the equation $\epsilon(\beta) = 0$. By following the lines of Huang and Qin (2013), we have the following theorem on the asymptotic property of $\hat{\beta}_\epsilon$, and relegate the proof to the ‘‘Appendix A’’.

Theorem 2 *Under Conditions (C1)–(C3) given in the ‘‘Appendix A’’, we can conclude that, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\beta}_\epsilon - \beta_0)$ converges to a normal distribution with mean zero and covariance matrix $[\Sigma(\beta_0) + V_2(\beta_0)]^{-1}[\Sigma(\beta_0) + V_1(\beta_0)][\Sigma(\beta_0) + V_2(\beta_0)]^{-1}$,*

where $\Sigma(\beta_0)$ is the limit of $n^{-1}I(\beta_0)$, $V_1(\beta_0) = 4E\{\psi_{12}(\beta_0)\psi_{13}(\beta_0)\}$ and $V_2(\beta_0) = -E\{\partial\psi_{12}(\beta_0)/\partial\beta\} = E[\rho_{12}^{\otimes 2} \exp(-\beta_0^\top \rho_{12})/\{1 + \exp(-\beta_0^\top \rho_{12})^2\}]$. In particular, $\Sigma(\beta_0)$ can be estimated by $n^{-1}I(\hat{\beta}_\epsilon)$, and $V_1(\beta_0)$ and $V_2(\beta_0)$ can be estimated by

$$\frac{4}{n-1} \sum_{i=1}^n \left\{ \frac{1}{n-1} \sum_{j=1, j \neq i}^n \psi_{ij}(\hat{\beta}_\epsilon) \right\}^{\otimes 2},$$

and

$$\frac{1}{n(n-1)} \sum_{i \neq j} \frac{\rho_{ij}^{\otimes 2} \exp(-\hat{\beta}_\epsilon^\top \rho_{ij})}{\{1 + \exp(-\hat{\beta}_\epsilon^\top \rho_{ij})\}^2},$$

respectively (Sen 1960).

4 Generations to the left-truncated and general interval-censored data

In this section, we extend the proposed methods above to the scenario of the left-truncated and general interval-censored data. Without loss of generality, assume that each subject is observed twice after the study enrolment, and the two monitoring times measured from the study enrolment are denoted by U and V , respectively, with $0 < U < V < \infty$. In this situation, we have left-censored T if $T - A < U$, right-censored T if $T - A \geq V$, and interval-censored T if $T - A \geq U$ and $T - A < V$, where A is the truncation time as defined above. Define $\delta^{(1)} = I(A + U \leq T)$ and $\delta^{(2)} = I(A + V \leq T)$.

Let $\{U_i, V_i, \delta_i^{(1)}, \delta_i^{(2)}, A_i, \mathbf{Z}_i; i = 1, \dots, n\}$ be n i.i.d. replicates of $\{U, V, \delta^{(1)}, \delta^{(2)}, A, \mathbf{Z}\}$. Define $\tilde{U}_i = A_i + U_i$ and $\tilde{V}_i = A_i + V_i$. As in Lin et al. (1998), for each i , define the counting process $N_i^{(1)}(t) = \delta_i^{(1)} I(\tilde{U}_i \leq t)$ with the intensity model

$$dH_i^{(1)}(t) = R_i^{(1)}(t) \exp\{-\beta^\top \mathbf{Z}_i t\} dH_1(t),$$

where $R_i^{(1)}(t) = I(A_i \leq t \leq \tilde{U}_i)$, $dH_1(t) = \exp\{-\Lambda(t)\} d\Lambda_U(t)$, and $\Lambda_U(t)$ is the cumulative hazard function of \tilde{U}_i . The above intensity model suggests that

$$M_i^{(1)}(t) = N_i^{(1)}(t) - \int_0^t R_i^{(1)}(s) \exp\{-\beta^\top \mathbf{Z}_i s\} dH_1(s) \quad (i = 1, \dots, n)$$

are martingales with respect to the σ -filtration: $\sigma\{N_i^{(1)}(s), R_i^{(1)}(s), \mathbf{Z}_i; s \leq t, i = 1, \dots, n\}$.

Similarly, we can also define the counting process related to \tilde{V}_i as $N_i^{(2)}(t) = \delta_i^{(2)} I(\tilde{V}_i \leq t)$, and its intensity model is given by

$$dH_i^{(2)}(t) = R_i^{(2)}(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{Z}_i t\} dH_2(t),$$

where $R_i^{(2)}(t) = I(\tilde{U}_i \leq t \leq \tilde{V}_i)$, $dH_2(t) = \exp\{-\Lambda(t)\} d\Lambda_V(t)$, and $\Lambda_V(t)$ is the cumulative hazard function of \tilde{V}_i . The above intensity model implies that

$$M_i^{(2)}(t) = N_i^{(2)}(t) - \int_0^t R_i^{(2)}(s) \exp\{-\boldsymbol{\beta}^\top \mathbf{Z}_i s\} dH_2(s) \quad (i = 1, \dots, n)$$

are martingales with respect to the σ -filtration: $\sigma\{N_i^{(2)}(s), R_i^{(2)}(s), \mathbf{Z}_i; s \leq t, i = 1, \dots, n\}$.

In what follows, we derive a conditional estimating equation approach for $\boldsymbol{\beta}$ under the conditional independent assumption between the examination and failure times given the covariates. The proposal below mainly utilizes the idea of working independence for interval-censored data (Betensky et al. 2001; Zhu et al. 2008), which treats the multiple examinations from a subject as the single examinations from different subjects. Therefore, the partial likelihood score function of $\boldsymbol{\beta}$ can be expressed as

$$\Phi(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S_1^{(1)}(\boldsymbol{\beta}, t)}{S_1^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_i^{(1)}(t) + \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S_2^{(1)}(\boldsymbol{\beta}, t)}{S_2^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_i^{(2)}(t), \tag{8}$$

where

$$S_1^{(k)}(\boldsymbol{\beta}, t) = \sum_{j=1}^n R_j^{(1)}(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{Z}_j t\} (\mathbf{Z}_j t)^{\otimes k},$$

$$S_2^{(k)}(\boldsymbol{\beta}, t) = \sum_{j=1}^n R_j^{(2)}(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{Z}_j t\} (\mathbf{Z}_j t)^{\otimes k},$$

$\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^\top$ for a column vector \mathbf{a} , and here τ is a prespecified constant such that $P(A + U \geq \tau) > 0$ and $P(A + V \geq \tau) > 0$.

The observed information matrix for $\boldsymbol{\beta}$ is

$$I(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{k=1}^2 \int_0^\tau \left\{ \frac{S_k^{(2)}(\boldsymbol{\beta}, t)}{S_k^{(0)}(\boldsymbol{\beta}, t)} - \frac{S_k^{(1)}(\boldsymbol{\beta}, t)^{\otimes 2}}{S_k^{(0)}(\boldsymbol{\beta}, t)^2} \right\} dN_i^{(k)}(t). \tag{9}$$

For each i , define

$$a_i(\boldsymbol{\beta}) = \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S_1^{(1)}(\boldsymbol{\beta}, t)}{S_1^{(0)}(\boldsymbol{\beta}, t)} \right\} dM_i^{(1)}(t)$$

$$\begin{aligned}
 &= \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S_1^{(1)}(\boldsymbol{\beta}, t)}{S_1^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_i^{(1)}(t) \\
 &\quad - \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S_1^{(1)}(\boldsymbol{\beta}, t)}{S_1^{(0)}(\boldsymbol{\beta}, t)} \right\} \frac{R_i^{(1)}(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{Z}_i t\}}{S_1^{(0)}(\boldsymbol{\beta}, t)} \left\{ \sum_{i=1}^n dN_i^{(1)}(t) \right\},
 \end{aligned}$$

and

$$\begin{aligned}
 b_i(\boldsymbol{\beta}) &= \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S_2^{(1)}(\boldsymbol{\beta}, t)}{S_2^{(0)}(\boldsymbol{\beta}, t)} \right\} dM_i^{(2)}(t) \\
 &= \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S_2^{(1)}(\boldsymbol{\beta}, t)}{S_2^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_i^{(2)}(t) \\
 &\quad - \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S_2^{(1)}(\boldsymbol{\beta}, t)}{S_2^{(0)}(\boldsymbol{\beta}, t)} \right\} \frac{R_i^{(2)}(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{Z}_i t\}}{S_2^{(0)}(\boldsymbol{\beta}, t)} \left\{ \sum_{i=1}^n dN_i^{(2)}(t) \right\}.
 \end{aligned}$$

Let $\hat{\boldsymbol{\beta}}_{U2}$ denote the estimate of $\boldsymbol{\beta}$ obtained by solving the normalised conditional estimating equation

$$U(\boldsymbol{\beta}) = \frac{1}{n} \Phi(\boldsymbol{\beta}) = 0. \tag{10}$$

We establish the asymptotic property of $\hat{\boldsymbol{\beta}}_{U2}$ in the following theorem and defer the detailed proof to the ‘‘Appendix B’’.

Theorem 3 *Under Conditions (C4)–(C6) given in the ‘‘Appendix B’’, we can conclude that, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_{U2} - \boldsymbol{\beta}_0)$ converges in distribution to a zero-mean normal random vector with covariance matrix $\Sigma_B(\boldsymbol{\beta}_0)^{-1} \Sigma_A(\boldsymbol{\beta}_0) \Sigma_B(\boldsymbol{\beta}_0)^{-1}$, where $\Sigma_A(\boldsymbol{\beta}_0)$ is the limit of*

$$\frac{1}{n} \sum_{i=1}^n [a_i(\boldsymbol{\beta}_0) + b_i(\boldsymbol{\beta}_0)][a_i(\boldsymbol{\beta}_0) + b_i(\boldsymbol{\beta}_0)]^\top,$$

and $\Sigma_B(\boldsymbol{\beta}_0)$ is the limit of $n^{-1}I(\boldsymbol{\beta}_0)$. In particular, $\Sigma_B(\boldsymbol{\beta}_0)$ can be estimated by $n^{-1}I(\hat{\boldsymbol{\beta}}_{U2})$, and $\Sigma_A(\boldsymbol{\beta}_0)$ can be estimated by

$$\frac{1}{n} \sum_{i=1}^n [a_i(\hat{\boldsymbol{\beta}}_{U2}) + b_i(\hat{\boldsymbol{\beta}}_{U2})][a_i(\hat{\boldsymbol{\beta}}_{U2}) + b_i(\hat{\boldsymbol{\beta}}_{U2})]^\top.$$

Under the left-truncated and general interval-censored data, the combined estimating equation for the additive hazards model can be defined as

$$\epsilon(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) + \psi(\boldsymbol{\beta}) = 0, \tag{11}$$

where $U(\beta)$ and $\psi(\beta)$ are given in (10) and (6), respectively. Let $\hat{\beta}_{\epsilon 2}$ be the estimate of β obtained by solving the equation $\epsilon(\beta) = 0$ (11). By following the arguments of Huang and Qin (2013), we can derive the following theorem on the asymptotic property of $\hat{\beta}$. The proof will be presented in the ‘‘Appendix B’’.

Theorem 4 *Under Conditions (C4)–(C6) given in the ‘‘Appendix B’’, we can conclude that, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\beta}_{\epsilon 2} - \beta_0)$ converges to a normal distribution with mean zero and covariance matrix $[\Sigma_B(\beta_0) + V_2(\beta_0)]^{-1}[\Sigma_A(\beta_0) + V_1(\beta_0)][\Sigma_B(\beta_0) + V_2(\beta_0)]^{-1}$, where $\Sigma_A(\beta_0)$ is the limit of*

$$\frac{1}{n} \sum_{i=1}^n [a_i(\beta_0) + b_i(\beta_0)][a_i(\beta_0) + b_i(\beta_0)]^T,$$

$\Sigma_B(\beta_0)$ is the limit of $n^{-1}I(\beta_0)$, $V_1(\beta_0) = 4E\{\psi_{12}(\beta_0)\psi_{13}(\beta_0)\}$ and $V_2(\beta_0) = -E\{\partial\psi_{12}(\beta_0)/\partial\beta\} = E[\rho_{12}^{\otimes 2} \exp(-\beta_0^T \rho_{12}) / \{1 + \exp(-\beta_0^T \rho_{12})^2\}]$.

In the above, $\Sigma_B(\beta_0)$ can be estimated by $n^{-1}I(\hat{\beta}_{\epsilon 2})$, $\Sigma_A(\beta_0)$ can be estimated by $\frac{1}{n} \sum_{i=1}^n [a_i(\hat{\beta}_{\epsilon 2}) + b_i(\hat{\beta}_{\epsilon 2})][a_i(\hat{\beta}_{\epsilon 2}) + b_i(\hat{\beta}_{\epsilon 2})]^T$, and $V_1(\beta_0)$ and $V_2(\beta_0)$ can be estimated by

$$\frac{4}{n-1} \sum_{i=1}^n \left\{ \frac{1}{n-1} \sum_{j=1, j \neq i}^n \psi_{ij}(\hat{\beta}_{\epsilon 2}) \right\}^{\otimes 2},$$

and

$$\frac{1}{n(n-1)} \sum_{i \neq j} \frac{\rho_{ij}^{\otimes 2} \exp(-\hat{\beta}_{\epsilon 2}^T \rho_{ij})}{\{1 + \exp(-\hat{\beta}_{\epsilon 2}^T \rho_{ij})\}^2},$$

respectively (Sen 1960).

5 Simulation studies

We conduct extensive simulation studies to evaluate the empirical performance of the proposed methods in finite samples. In the first study, we considered the left-truncated and case I interval-censored data, and assumed that there existed two covariates Z_1 and Z_2 , where Z_1 follows the Bernoulli distribution with a success probability of 0.5 and Z_2 follows the uniform distribution over (0, 1). The underlying failure time of interest T^* was then generated from the additive hazards model (1) with $\lambda(t) = 1$ and the true values of (β_1, β_2) being (0.5, 0.5). The underlying truncation time A^* was assumed to follow either the exponential distribution with mean 10 or the uniform distribution over (0, 100). Due to the presence of left truncation, only the pairs satisfying $T^* \geq A^*$ were kept in the simulated data, and thus $T = T^*$ and $A = A^*$ under this situation. The censoring time from the study enrolment C was generated from the exponential

distribution with mean $2/3$ and $\tilde{C} = C + A$, which yielded about 50% right censoring rates. We set $n = 100, 200, 400$ or 800 and used 1000 replications. To analyze the simulated data, we considered the following three competing methods: the combined estimating equation approach, the conditional estimating equation approach, and the naive method proposed by Lin et al. (1998) that ignores the existence of left truncation.

Tables 1 and 2 present the numerical results corresponding to the situations where the truncation time follows the exponential and uniform distributions, respectively. The tables include the estimation bias (Bias) given by the average of the estimates minus the true value, the sample standard errors (SSE) of the estimates, the average standard error estimates (SEE) and the 95% empirical coverage probabilities (CP). One can see from Tables 1 and 2 that the combined estimating equation approach can give unbiased estimates, the bias decreases when the sample size increases, the corresponding standard error estimates match the average standard error estimates well, and the empirical coverage probabilities are around the nominal value 95%, which implies that the normal approximation to the distribution of the estimates seems to be reasonable. As anticipated, the combined estimating equation approach proposed is much more efficient than the conditional estimating equation approach since the former utilized the distribution information of truncation times in the estimation procedure. In addition, it is worth noting that the naive method that ignores the existence of left truncation produced severely biased estimates. This finding can also be anticipated since the obtained left-truncated data are no longer representative for the whole population under study.

In the second study, we considered the scenario of the left-truncated and general interval-censored data. For each subject, we assumed that there existed two observation times \tilde{U} and \tilde{V} , where we set $\tilde{U} = A + U$, $\tilde{V} = A + V$, U follows uniform distribution over $(0, 0.5)$ and V follows uniform distribution over $(U + 0.3, U + 0.8)$. We kept the other simulation specifications being the same as above. Under this configuration, the left and right censoring rates are both approximately 33%. As in the first simulation study, we performed the analysis with the two proposed estimating equation approaches, and also considered the naive estimating equation method that ignores the existence of left truncation, which can be implemented with the proposed conditional estimating equation approach by setting $A = 0$. The simulation results given in Tables 3 and 4 correspond to the situations where the truncation time follows the exponential and uniform distributions, respectively, from which one can find the similar conclusions as above.

In the simulation studies above, we also tried to apply the generalized method of moments (GMM) (Hansen 1982) to further enhance the estimation efficiency. Let $G(\boldsymbol{\beta}) = (U(\boldsymbol{\beta})^\top, \psi(\boldsymbol{\beta})^\top)^\top$ and W be a positive-definite weight function matrix. A consistent estimator of $\boldsymbol{\beta}$ can be obtained through minimizing $G(\boldsymbol{\beta})^\top W^{-1} G(\boldsymbol{\beta})$, and the optimal matrix that gives the efficient estimation is $W = \text{var}\{G(\boldsymbol{\beta})\}$. However, the numerical results (not shown here) indicate that the GMM method does not lead to higher efficiency compared with the proposed combined estimating equation approach. As commented by Huang and Qin (2013), this phenomenon arises partially because the optimal weight function involves estimating the second moments of $U(\boldsymbol{\beta}_0)$ and $\psi(\boldsymbol{\beta}_0)$, which usually needs larger sample size to obtain the superiority of an efficient GMM estimator.

Table 1 Simulation results for the left-truncated and case I interval-censored data when the truncation time follows the exponential distribution, including the estimated bias (Bias), the sample standard errors (SSE) of the estimates, the average standard error estimates (SEE) and the 95% empirical coverage probabilities (CP)

n	Par	Combined EE				Conditional EE				Ignoring truncation			
		Bias	SSE	SEE	CP	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
100	β_1	0.0358	0.3346	0.3115	0.942	0.0403	0.7578	0.6935	0.950	-0.3056	0.7510	0.6778	0.914
	β_2	0.0261	0.5501	0.5194	0.945	0.0506	1.2302	1.1642	0.953	-0.3953	1.2053	1.1377	0.938
200	β_1	0.0182	0.2131	0.2127	0.950	0.0267	0.4835	0.4655	0.951	-0.3359	0.4619	0.4552	0.868
	β_2	0.0231	0.3663	0.3542	0.946	-0.0213	0.8050	0.7738	0.950	-0.3466	0.7657	0.7709	0.931
400	β_1	0.0156	0.1506	0.1503	0.955	0.0132	0.3318	0.3197	0.948	-0.3314	0.3212	0.3110	0.774
	β_2	0.0099	0.2484	0.2469	0.949	0.0270	0.5409	0.5308	0.942	-0.3761	0.5393	0.5245	0.880
800	β_1	0.0019	0.1055	0.1044	0.949	-0.0058	0.2225	0.2224	0.950	-0.3372	0.2139	0.2165	0.654
	β_2	0.0025	0.1701	0.1729	0.946	-0.0078	0.3662	0.3687	0.949	-0.3399	0.3711	0.3648	0.839

“Combined EE” denotes the combined estimating equation approach, “Conditional EE” denotes the conditional estimating equation approach, and “Ignoring truncation” denotes the estimating equation approach that ignores the existence of left truncation

Table 2 Simulation results for the left-truncated and case I interval-censored data when the truncation time follows the uniform distribution, including the estimated bias (Bias), the sample standard errors (SSE) of the estimates, the average standard error estimates (SEE) and the 95% empirical coverage probabilities (CP)

n	Par	Combined EE				Conditional EE				Ignoring truncation			
		Bias	SSE	SEE	CP	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
100	β_1	0.0342	0.3044	0.2982	0.945	0.0699	0.7416	0.7002	0.967	-0.3987	0.7111	0.6805	0.891
	β_2	0.0306	0.4933	0.4860	0.960	0.0522	1.2554	0.1747	0.946	-0.4085	1.2226	1.1514	0.937
200	β_1	0.0265	0.2014	0.2036	0.956	0.0375	0.4897	0.4685	0.946	-0.3488	0.4524	0.4529	0.864
	β_2	0.0279	0.3556	0.3352	0.944	0.0486	0.7873	0.7757	0.954	-0.3643	0.7807	0.7638	0.920
400	β_1	0.0126	0.1335	0.1427	0.960	0.0215	0.3260	0.3211	0.952	-0.3563	0.3164	0.3119	0.766
	β_2	0.0109	0.2327	0.2342	0.948	0.0295	0.5335	0.5335	0.960	-0.3734	0.5531	0.5249	0.860
800	β_1	0.0052	0.1070	0.0998	0.931	0.0102	0.2347	0.2232	0.942	-0.3613	0.2258	0.2172	0.584
	β_2	0.0066	0.1658	0.1644	0.955	0.0199	0.3860	0.3695	0.944	-0.3716	0.3686	0.3661	0.800

“Combined EE” denotes the combined estimating equation approach, “Conditional EE” denotes the conditional estimating equation approach, and “Ignoring truncation” denotes the estimating equation approach that ignores the existence of left truncation

Table 3 Simulation results for the left-truncated and general interval-censored data when the truncation time follows the exponential distribution, including the estimated bias (Bias), the sample standard errors (SSE) of the estimates, the average standard error estimates (SEE) and the 95% empirical coverage probabilities (CP)

n	Par	Combined EE				Conditional EE				Ignoring truncation			
		Bias	SSE	SEE	CP	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
100	β_1	0.0185	0.2678	0.2925	0.971	-0.0356	0.5945	0.5792	0.943	-0.3469	0.6368	0.5509	0.860
	β_2	0.0206	0.4703	0.4884	0.961	-0.0944	1.0162	0.9791	0.941	-0.3752	1.1237	0.9340	0.876
200	β_1	-0.0115	0.1905	0.2029	0.961	-0.0556	0.4013	0.3973	0.956	-0.3938	0.4115	0.3774	0.800
	β_2	-0.0188	0.3013	0.3395	0.973	-0.0678	0.6866	0.6755	0.946	-0.3980	0.7545	0.6448	0.868
400	β_1	-0.0093	0.1285	0.1427	0.972	-0.0509	0.2785	0.2779	0.951	-0.3451	0.2894	0.2637	0.780
	β_2	-0.0182	0.2160	0.2377	0.963	-0.0633	0.4511	0.4724	0.957	-0.3764	0.5043	0.4482	0.838
800	β_1	-0.0083	0.0893	0.1005	0.966	-0.0465	0.1890	0.1951	0.947	-0.3337	0.2091	0.1851	0.560
	β_2	-0.0031	0.1518	0.1681	0.966	-0.0430	0.3190	0.3321	0.962	-0.3799	0.3526	0.3145	0.746

“Combined EE” denotes the combined estimating equation approach, “Conditional EE” denotes the conditional estimating equation approach, and “Ignoring truncation” denotes the estimating equation approach that ignores the existence of left truncation

Table 4 Simulation results for the left-truncated and general interval-censored data when the truncation time follows the uniform distribution, including the estimated bias (Bias), the sample standard errors (SSE) of the estimates, the average standard error estimates (SEE) and the 95% empirical coverage probabilities (CP)

n	Par	Combined EE				Conditional EE				Ignoring truncation			
		Bias	SSE	SEE	CP	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
100	β_1	0.0124	0.2583	0.2798	0.969	-0.0575	0.5814	0.5813	0.957	-0.4208	0.6285	0.5553	0.846
	β_2	0.0261	0.4505	0.4676	0.966	-0.0116	0.9906	0.9906	0.948	-0.3315	1.0725	0.9476	0.910
200	β_1	-0.0095	0.1744	0.1949	0.975	-0.0540	0.3992	0.4010	0.956	-0.3737	0.4308	0.3810	0.798
	β_2	-0.0196	0.3039	0.3256	0.966	-0.0029	0.6926	0.6841	0.960	-0.3468	0.7466	0.6457	0.870
400	β_1	-0.0094	0.1283	0.1369	0.959	-0.0498	0.2640	0.2680	0.959	-0.3742	0.2840	0.2650	0.698
	β_2	-0.0121	0.2032	0.2277	0.972	-0.0487	0.4671	0.4766	0.953	-0.3999	0.5128	0.4502	0.814
800	β_1	-0.0090	0.0849	0.0965	0.969	-0.0507	0.1804	0.1969	0.965	-0.3797	0.1914	0.1862	0.467
	β_2	-0.0066	0.1432	0.1606	0.965	-0.0397	0.3104	0.3355	0.971	-0.3867	0.3465	0.3168	0.744

“Combined EE” denotes the combined estimating equation approach, “Conditional EE” denotes the conditional estimating equation approach, and “Ignoring truncation” denotes the estimating equation approach that ignores the existence of left truncation

Table 5 Analysis results of the MHCPS data, including the estimated covariate effect (Est), the standard error estimate (Std) and the p -value

Method	Est	Std	p -value
Combined EE	0.00165	0.00009	<0.001
Conditional EE	0.00057	0.00010	<0.001
Ignoring truncation	0.00050	0.00010	<0.001

Note: "Combined EE" denotes the combined estimating equation approach, "Conditional EE" denotes the conditional estimating equation approach, and "Ignoring truncation" denotes the estimating equation approach that ignores the existence of left truncation

6 Real data analysis

We now apply the proposed methods to a set of left-truncated and interval-censored data arising from the Massachusetts Health Care Panel Study (MHCPS) analyzed by Pan and Chappell (1998, 2002) and Gao and Chan (2019) and others. In 1975, the study enrolled individuals over the age of 65 in Massachusetts, and after that, three subsequent follow-ups were taken at the 1.25, 6, and 10 years after the enrolment to determine if individuals were still living actively, resulting in the interval-censored observation on the time to loss of active life. The main objective of the study was to evaluate the effect of gender (male or female) on the loss of active life for elderly people. For this, the failure time of interest is defined as the age at loss of active life. In addition, since only people who were active at the enrolment were included in the study, the time to the loss of active life was also subject to left truncation with the truncation time being the age at enrolment (Pan and Chappell 2002). Therefore, we had the left-truncated and interval-censored data on the time to loss of active life. After deleting a small proportion of unrealistic records of the raw data, a total of 1025 individuals with the age ranging from 65 to 97.3 were finally included in the current analysis. The right censoring rate is 45.8%.

In this MHCPS dataset, the data for the failure times of interest were given in the form of $\{[L_i, R_i); i = 1, \dots, n\}$ and contained a mixture of interval-censored and right-censored observations. To implement the proposed methods, we followed the strategy given in Wang et al. (2010) and made an adjustment: for subject i with $[L_i, R_i)$, we let $\tilde{V}_i = L_i$ and \tilde{U}_i to be the smallest observation time in the study if $R_i = \infty$; and we set $\tilde{U}_i = L_i$ and $\tilde{V}_i = R_i$ when $L_i > 0$ and $R_i < \infty$. Define $Z_i = 1$ if the i th subject was male and $Z_i = 0$ if this subject was female.

Table 5 presents the analysis results of the MHCPS data with the combined estimating equation approach, the conditional estimating equation approach, and the naive estimating equation method that ignores the existence of left truncation. They include the estimated covariate effect (Est), the standard error estimate (Std) and the p -value for testing the covariate effect being zero. One can find from Table 5 that the results obtained from the combined estimating equation approach suggest that males have significantly higher risk of losing active life than females, which is in accordance with the conclusion given by Gao and Chan (2019) under the proportional hazards model. In addition, one can find from Table 5 that the conditional and the naive estimating equation approaches give consistent conclusions, and it is apparent that they tend to

underestimate the covariate effect compared with the combined estimating equation approach. The attenuated covariate effect estimates obtained by the conditional and the naive estimating equation approaches may arise because the two approaches either fail to utilize the distribution information of the truncation times or ignore the existence of left truncation. In general, they both lead to larger estimation bias compared with the combined estimating equation approach as shown in the simulation studies.

As discussed before, under the assumed additive hazards model (1), the interpretation of the estimated covariate effect can be converted into that of the ratio of the estimated survival probabilities between different groups at a given time. For example, the estimated ratio of the probabilities of having the active life after 80 years old between male and female under the combined estimating equation approach is $S(80 | Z_i = 1)/S(80 | Z_i = 0) = \exp\{-\Lambda(80) - 0.00165 \times 80\} / \exp\{-\Lambda(80)\} \approx 0.876$. In contrast, the estimated ratios of the probabilities of having the active life after 80 years old between male and female are about 0.955 and 0.961 under the conditional and the naive estimating equation approaches, respectively.

7 Concluding remarks

Additive hazards model is an essential alternative to the commonly used proportional hazards model in failure time data analysis, and assumes an additive covariate effect on the hazard function of the failure time of interest. In this paper, we studied the additive hazards regression analysis of the left-truncated and interval-censored failure time data including the case I and general interval censoring schemes. By utilizing the order information of the pairwise truncation times, we developed the combined estimating equation approach for the estimation, which yielded more efficient estimators than the conditional estimating equation approach. The tractable estimating equations derived were very easy to solve with the routine root-finding method, such as Newton–Raphson algorithm. The asymptotic properties of the proposed estimators were established and numerical studies demonstrated the usefulness of the proposed methodology in finite samples.

In the preceding sections, we only considered time-independent covariates and it is apparent that one may encounter time-dependent covariates in some applications. For the latter case, one needs to redefine the model (1) as $\lambda(t | \mathbf{Z}^*(t)) = \lambda(t) + \boldsymbol{\beta}^\top \mathbf{Z}^*(t)$, where $\mathbf{Z}^*(t)$ is a p -dimensional vector of time-dependent covariates, and the estimation procedures proposed above can be readily adapted to handle this situation. Note that, in the proposed methods, we assumed that the failure times of interest are conditionally independent of the observation times given the covariates, which is usually referred to as non-informative censoring in the literature. However, this assumption may not hold in some applications, and it is helpful to extend the proposed methods to the case of informative censoring (Zhang et al. 2005; Ma et al. 2015; Li et al. 2017). Model checking is often of great interest and Ghosh (2003) proposed a formal goodness-of-fit test procedure for the additive hazards model under current status data or case I interval-censored data. Future work will be devoted to generalize the method proposed by Ghosh (2003) to the situation of left-truncated and interval-censored data. In addition, left-truncated and multivariate interval-censored failure

time data can also be encountered in many scientific fields and it is useful to extend the proposed methods to the analysis of such data.

Acknowledgements We are grateful to the Editor, the Associate Editor and two reviewers for their insightful comments and suggestions that greatly improved the article. Shuwei Li’s research was partially supported by the Science and Technology Program of Guangzhou of China (Grant No. 202102010512), National Statistical Science Research Project (Grant No. 2022LY041), National Nature Science Foundation of China (Grant No. 11901128) and Nature Science Foundation of Guangdong Province of China (Grant Nos. 2021A1515010044 and 2022A1515011901). Liuquan Sun’s research was partially supported by the National Natural Science Foundation of China (Grant No. 12171463).

Appendix A: Proofs of Theorems 1 and 2

Let $S^{(k)}(\beta, t) = \sum_{j=1}^n R_j(t) \exp\{-\beta^\top Z_j t\} (Z_j t)^{\otimes k}$ for $k = 0, 1, 2$. For a matrix A or a vector a , $\|A\| = \sup_{i,j} |a_{ij}|$ and $\|a\| = \sup_i |a_i|$, where a_{ij} is the (i, j) th element of A and a_i is the i th component of a . For a vector a , $|a| = (a^\top a)^{1/2}$. In what follows, “ \xrightarrow{P} ” and “ \xrightarrow{D} ” denote the convergence in probability and the convergence in distribution, respectively.

To establish the asymptotic properties of proposed estimators of β under left-truncated and case I interval-censored data, we need the following regularity conditions:

- (C1) $P(T^* > A^* | Z^*) > 0$, $P(A + C \geq \tau | Z) > 0$ and $P(R(t) = 1 | Z) > 0$ for all $t \in [A, \tau]$, where $R(t) = I(A \leq t \leq A + C)$.
- (C2) Z_i is bounded, and $H_0(\tau) < \infty$, where $H_0(t)$ is the true value of $H(t)$.
- (C3) The true regression vector β_0 lies in the interior of a compact set \mathcal{B} and there exist functions $s^{(k)}(\beta, t)$ with $k = 0, 1, 2$ defined on $\mathcal{B} \times (0, \tau]$ satisfying

- (a) $\sup_{\beta \in \mathcal{B}, t \in (0, \tau]} \|n^{-1} S^{(k)}(\beta, t) - s^{(k)}(\beta, t)\| \xrightarrow{P} 0$ as $n \rightarrow \infty$.
- (b) $s^{(0)}(\beta, t)$ is bounded away from 0 for $t \in (0, \tau]$.
- (c) For $k = 0, 1, 2$, $s^{(k)}(\beta, t)$ is a continuous function of β uniformly in $t \in (0, \tau]$, where $s^{(1)}(\beta, t) = \partial s^{(0)}(\beta, t) / \partial \beta$ and $s^{(2)}(\beta, t) = \partial^2 s^{(0)}(\beta, t) / \partial \beta \partial \beta^\top$.
- (d) For $\beta \in \mathcal{B}$, let $v(\beta, t) = s^{(2)}(\beta, t) / s^{(0)}(\beta, t) - e(\beta, t) e(\beta, t)^\top$ and $e(\beta, t) = s^{(1)}(\beta, t) / s^{(0)}(\beta, t)$. $\Sigma(\beta_0) = \int_0^\tau v(\beta_0, u) s^{(0)}(\beta_0, u) dH_0(u)$ is positive definite.

Proof of Theorem 1:

To prove the asymptotic properties of $\hat{\beta}_{UJ}$, we mainly follow the arguments given in Andersen and Gill (1982) and Kalbfleisch and Prentice (2002). Note that the objective function corresponding to the equation (3) is

$$l(\beta) = - \sum_{i=1}^n \int_0^\tau \left\{ \beta^\top Z_i u + \log[S^{(0)}(\beta, u)] \right\} dN_i(u).$$

For $t \leq \tau$, define

$$l(\boldsymbol{\beta}, t) = - \sum_{i=1}^n \int_0^t \left\{ \boldsymbol{\beta}^\top \mathbf{Z}_i u + \log[S^{(0)}(\boldsymbol{\beta}, u)] \right\} dN_i(u).$$

Consider the process

$$\begin{aligned} X(\boldsymbol{\beta}, t) &= n^{-1}(l(\boldsymbol{\beta}, t) - l(\boldsymbol{\beta}_0, t)) \\ &= -n^{-1} \sum_{i=1}^n \left\{ \int_0^t (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \mathbf{Z}_i u + \log \left[\frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}_0, u)} \right] \right\} dN_i(u). \end{aligned}$$

Then the compensator of $X(\boldsymbol{\beta}, t)$ is

$$\tilde{X}(\boldsymbol{\beta}, t) = -n^{-1} \int_0^t \left\{ (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top s^{(1)}(\boldsymbol{\beta}_0, u) + \log \left[\frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}_0, u)} \right] s^{(0)}(\boldsymbol{\beta}_0, u) \right\} dH_0(u).$$

Under Conditions (C1)–(C3), we can conclude that $\tilde{X}(\boldsymbol{\beta}, \tau)$ converges to a function of $\boldsymbol{\beta}$ denoted as $f(\boldsymbol{\beta})$, where

$$f(\boldsymbol{\beta}) = - \int_0^\tau \left\{ (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top s^{(1)}(\boldsymbol{\beta}_0, u) + \log \left[\frac{s^{(0)}(\boldsymbol{\beta}, u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} \right] s^{(0)}(\boldsymbol{\beta}_0, u) \right\} dH_0(u).$$

The first and second derivatives of $f(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ are given by

$$\partial f(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = - \int_0^\tau \left\{ s^{(1)}(\boldsymbol{\beta}_0, u) - s^{(1)}(\boldsymbol{\beta}, u) \frac{s^{(0)}(\boldsymbol{\beta}_0, u)}{s^{(0)}(\boldsymbol{\beta}, u)} \right\} dH_0(u),$$

and

$$\begin{aligned} \partial^2 f(\boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top &= \int_0^\tau \left\{ -\frac{s^{(2)}(\boldsymbol{\beta}, u)}{s^{(0)}(\boldsymbol{\beta}, u)} + \frac{s^{(1)}(\boldsymbol{\beta}, u)^{\otimes 2}}{s^{(0)}(\boldsymbol{\beta}, u)^2} \right\} s^{(0)}(\boldsymbol{\beta}_0, u) dH_0(u) \\ &= - \int_0^\tau v(\boldsymbol{\beta}, u) s^{(0)}(\boldsymbol{\beta}_0, u) dH_0(u) = -\Sigma(\boldsymbol{\beta}), \end{aligned}$$

respectively.

It is easy to find that $\partial f(\boldsymbol{\beta})/\partial \boldsymbol{\beta}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = 0$, and $\partial^2 f(\boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top$ is minus a positive definite matrix. Therefore, the unique maximum of $f(\boldsymbol{\beta})$ is at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. By following Andersen and Gill (1982), it can be verified that the predictable variation process of $X(\boldsymbol{\beta}, t) - \tilde{X}(\boldsymbol{\beta}, t)$ converges to 0 on interval $(0, \tau]$. Then by using the Lenglart’s inequality given in Andersen and Gill (1982), we can conclude that $X(\boldsymbol{\beta}, \tau) \xrightarrow{P} f(\boldsymbol{\beta})$ with probability one. Note that $\hat{\boldsymbol{\beta}}_U$ maximizes the function $X(\boldsymbol{\beta}, \tau)$. Applying the convex analysis to $X(\boldsymbol{\beta}, \tau)$ and $f(\boldsymbol{\beta})$ as Andersen and Gill (1982) leads to the conclusion that $\hat{\boldsymbol{\beta}}_U \rightarrow \boldsymbol{\beta}_0$ with probability one as $n \rightarrow \infty$.

By applying Taylor expansion about β_0 to $\Phi(\hat{\beta}_U)$, we have

$$0 = n^{-1/2}\Phi(\hat{\beta}_U) = n^{-1/2}\Phi(\beta_0) - n^{1/2}(\hat{\beta}_U - \beta_0)[n^{-1}I(\beta^*)]$$

where β^* is between $\hat{\beta}_U$ and β_0 and $I(\beta) = -\partial^2 l(\beta)/\partial\beta\partial\beta^\top$. Then the asymptotic distribution of $n^{1/2}(\hat{\beta}_U - \beta_0)$ can be established if we can show the probability limit of $n^{-1}I(\beta^*)$ and the asymptotic distribution of $\Phi(\beta_0)$.

By the consistency of $\hat{\beta}_U$ and the arguments given in (Andersen and Gill 1982), it can be shown show that $n^{-1}I(\beta^*) \xrightarrow{P} \Sigma(\beta_0)$ for any β^* that converges in probability to β_0 . Next, we prove that $n^{-1/2}\Phi(\beta_0) \xrightarrow{D} N[0, \Sigma(\beta_0)]$. Following Lin et al. (1998), for each i , we can conclude that $M_i(t)$ is martingale with respect to the σ -filtration: $\sigma\{N_i(s), R_i(s), \mathbf{Z}_i; s \leq t, i = 1, \dots, n\}$. Under Conditions (C2) and (C3), we know that $n^{-1/2}G_i^{(n)}(u) = n^{-1/2}\{\mathbf{Z}_i u - S^{(1)}(\beta_0, u)/S^{(0)}(\beta_0, u)\}$ is a vector of predictable process for each i . Simple algebraic manipulation shows that $n^{-1/2}\Phi(\beta_0, t) = \sum_{i=1}^n \int_0^t G_i^{(n)}(u)dM_i(u)$ is a martingale with the predictable covariation process

$$\begin{aligned} \langle n^{-1/2}\Phi(\beta_0) \rangle(t) &= \sum_{i=1}^n \int_0^t G_i^{(n)}(u)^\top G_i^{(n)}(u)R_i(u) \exp[-\beta_0^\top \mathbf{Z}_i u]dH_0(u) \\ &= \int_0^t V(\beta_0, u)S^{(0)}(\beta_0, u)dH_0(u). \end{aligned}$$

Under Condition (C3), we have

$$\langle n^{-1/2}\Phi(\beta_0) \rangle(t) \xrightarrow{P} \int_0^t v(\beta_0, u)s^{(0)}(\beta_0, u)dH_0(u).$$

Furthermore, note that Conditions (C2) and (C3) are sufficient to imply the Lindeberg condition of Rebolledo’s Central Limit Theorem (Andersen and Gill 1982). By applying the Rebolledo’s Central Limit Theorem, we can conclude that $n^{-1/2}\Phi(\beta_0) \xrightarrow{D} N[0, \Sigma(\beta_0)]$. Thus, $n^{1/2}(\hat{\beta}_U - \beta_0) \xrightarrow{D} N(0, \Sigma(\beta_0)^{-1})$.

Proof of Theorem 2:

To prove the asymptotic properties of $\hat{\beta}_\epsilon$ obtained by solving (7), we mainly follow the arguments given in Huang and Qin (2013). Note that the pairwise pseudo-likelihood of the observed (A_i, A_j) conditional on $A_i \leq \tau, A_j \leq \tau$ and $(\mathbf{Z}_i, \mathbf{Z}_j)$ is

$$\sum_{1 \leq i < j \leq n} -\log \left\{ 1 + \exp(\beta^\top \rho_{ij}) \right\},$$

which achieves the maximum at the true parameter value as $n \rightarrow \infty$. By applying the conditional Kullback–Leibler information inequality (Andersen 1970), the maximum pairwise pseudo-likelihood estimator $\hat{\beta}_\psi$ converges to β_0 with probability one. Note that $\psi(\beta)$ is a U-statistic of order two since $\psi_{ij}(\beta)$ is permutation-symmetric in the

augments (A_i, \mathbf{Z}_i) and (A_j, \mathbf{Z}_j) . By Conditions (C2) and Conditions (C3), we know that $E\{\psi_{ij}(\boldsymbol{\beta}_0)\} = 0$ and $E\{\psi_{ij}(\boldsymbol{\beta}_0)^{\otimes 2}\} < \infty$. By the projection method proposed by Hoeffding (1948), we know that $n^{1/2}\psi(\boldsymbol{\beta}_0)$ converges to the normal distribution with mean zero and covariance matrix $V_1 = 4E\{\psi_{12}(\boldsymbol{\beta}_0)^\top \psi_{13}(\boldsymbol{\beta}_0)\}$. Applying the delta method, $n^{1/2}(\hat{\boldsymbol{\beta}}_\psi - \boldsymbol{\beta}_0)$ converges to the normal distribution with mean zero and covariance matrix $V_2(\boldsymbol{\beta}_0)^{-1} V_1(\boldsymbol{\beta}_0) V_2(\boldsymbol{\beta}_0)^{-1}$, where $V_1(\boldsymbol{\beta}_0) = 4E\{\psi_{12}(\boldsymbol{\beta}_0) \psi_{13}(\boldsymbol{\beta}_0)\}$ and $V_2(\boldsymbol{\beta}_0) = -E\{\partial \psi_{12}(\boldsymbol{\beta}_0) / \partial \boldsymbol{\beta}\} = E[\rho_{12}^{\otimes 2} \exp(-\boldsymbol{\beta}_0^\top \rho_{12}) / \{1 + \exp(-\boldsymbol{\beta}_0^\top \rho_{12})^2\}]$.

Because $U(\boldsymbol{\beta})$ is linear of $\boldsymbol{\beta}$, the consistency of $\hat{\boldsymbol{\beta}}_\epsilon$ follows directly from the consistency of $\hat{\boldsymbol{\beta}}_\psi$. Since we have shown that $\hat{\boldsymbol{\beta}}_U$ and $\hat{\boldsymbol{\beta}}_\psi$ are asymptotic normal, the asymptotic normality of $\hat{\boldsymbol{\beta}}_\epsilon$ follows from the asymptotic independence of $U(\boldsymbol{\beta})$ and $\psi(\boldsymbol{\beta})$ (van der Vaart and Wellner 1996). Because $E[dM_i(t) \mid A_i, \mathbf{Z}_i] = 0$ and $\psi_{ij}(\boldsymbol{\beta}_0)$ only involves (A_i, \mathbf{Z}_i) and (A_j, \mathbf{Z}_j) , by double expectation we have that $E[U(\boldsymbol{\beta}_0)^\top \psi(\boldsymbol{\beta}_0)] = 0$ and $\text{var}[\epsilon(\boldsymbol{\beta}_0)] = \text{var}[U(\boldsymbol{\beta}_0)] + \text{var}[\psi(\boldsymbol{\beta}_0)]$. By the central limit theorem for U-statistics, $n^{1/2}\epsilon(\boldsymbol{\beta}_0)$ converges in distribution to a normal distribution with mean zero and covariance matrix $\text{var}[n^{1/2}U(\boldsymbol{\beta}_0)] + \text{var}[n^{1/2}\psi(\boldsymbol{\beta}_0)] = \Sigma(\boldsymbol{\beta}_0) + V_1(\boldsymbol{\beta}_0)$. Applying the Taylor expansion to $\epsilon(\boldsymbol{\beta})$, we can conclude that, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_\epsilon - \boldsymbol{\beta}_0)$ converges to a normal distribution with mean zero and covariance matrix $[\Sigma(\boldsymbol{\beta}_0) + V_2(\boldsymbol{\beta}_0)]^{-1} [\Sigma(\boldsymbol{\beta}_0) + V_1(\boldsymbol{\beta}_0)] [\Sigma(\boldsymbol{\beta}_0) + V_2(\boldsymbol{\beta}_0)]^{-1}$.

Appendix B: Proofs of Theorems 3 and 4

Define

$$S_1^{(k)}(\boldsymbol{\beta}, t) = \sum_{j=1}^n R_j^{(1)}(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{Z}_{jt}\} (\mathbf{Z}_{jt})^{\otimes k},$$

and

$$S_2^{(k)}(\boldsymbol{\beta}, t) = \sum_{j=1}^n R_j^{(2)}(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{Z}_{jt}\} (\mathbf{Z}_{jt})^{\otimes k},$$

for $k = 0, 1, 2$. To establish the asymptotic properties of the proposed estimators of $\boldsymbol{\beta}$ under left-truncated and general interval-censored data, we need the following regularity conditions:

- (C4) $P(T^* > A^* \mid \mathbf{Z}^*) > 0$, $P(A + U \geq \tau \mid \mathbf{Z}) > 0$, $P(A + V \geq \tau \mid \mathbf{Z}) > 0$, $P(R^{(1)}(t) = 1 \mid \mathbf{Z}) > 0$ and $P(R^{(2)}(t) = 1 \mid \mathbf{Z}) > 0$ for all $t \in [A, \tau]$, where $R^{(1)}(t) = I(A \leq t \leq A + U)$ and $R^{(2)}(t) = I(A \leq t \leq A + V)$.
- (C5) \mathbf{Z}_i is bounded. $H_{10}(\tau) < \infty$ and $H_{20}(\tau) < \infty$, where $H_{10}(t)$ and $H_{20}(t)$ are the true values of $H_1(t)$ and $H_2(t)$, respectively.
- (C6) The true regression vector $\boldsymbol{\beta}_0$ lies in the interior of a compact set \mathcal{B} and there exist functions $s_j^{(k)}(\boldsymbol{\beta}, t)$ with $k = 0, 1, 2$ and $j = 1, 2$ defined on $\mathcal{B} \times (0, \tau]$ satisfying

- (a) $\sup_{\beta \in \mathcal{B}, t \in (0, \tau]} \|n^{-1} S_j^{(k)}(\beta, t) - s_j^{(k)}(\beta, t)\| \xrightarrow{P} 0$ as $n \rightarrow \infty$ with $j = 1, 2$.
- (b) For $j = 1, 2$, $s_j^{(0)}(\beta, t)$ is bounded away from 0 for $t \in (0, \tau]$.
- (c) For $k = 0, 1, 2$, $s_j^{(k)}(\beta, t)$ is a continuous function of β uniformly in $t \in (0, \tau]$, where $s_j^{(1)}(\beta, t) = \partial s_j^{(0)}(\beta, t) / \partial \beta$ and $s_j^{(2)}(\beta, t) = \partial^2 s_j^{(0)}(\beta, t) / \partial \beta \partial \beta^\top$ with $j = 1, 2$.
- (d) For $\beta \in \mathcal{B}$, let $v_j(\beta, t) = s_j^{(2)}(\beta, t) / s_j^{(0)}(\beta, t) - e_j(\beta, t) e_j(\beta, t)^\top$ and $e_j(\beta, t) = s_j^{(1)}(\beta, t) / s_j^{(0)}(\beta, t)$ with $j = 1, 2$. $\Sigma(\beta_0) = \sum_{j=1}^2 \int_0^\tau v_j(\beta_0, u) s_j^{(0)}(\beta_0, u) dH_{j0}(u)$ is positive definite.

Proof of Theorem 3:

The objective function corresponding to the equation (8) is

$$l(\beta) = - \sum_{i=1}^n \int_0^\tau \left\{ \beta^\top Z_i t + \log[S_1^{(0)}(\beta, t)] \right\} dN_i^{(1)}(t) - \sum_{i=1}^n \int_0^\tau \left\{ \beta^\top Z_i t + \log[S_2^{(0)}(\beta, t)] \right\} dN_i^{(2)}(t).$$

Note that $\hat{\beta}_{U2}$ is obtained by maximizing $l(\beta)$ or solving the estimating Eq. (10), the consistency of $\hat{\beta}_{U2}$ can be established by applying the arguments in the proof of Theorem 1. Following Lin et al. (1998), for each i , we know that

$$M_i^{(1)}(t) = N_i^{(1)}(t) - \int_0^t R_i^{(1)}(u) \exp\{-\beta^\top Z_i u\} dH_{10}(u),$$

and

$$M_i^{(2)}(t) = N_i^{(2)}(t) - \int_0^t R_i^{(2)}(u) \exp\{-\beta^\top Z_i u\} dH_{20}(u)$$

are both martingales.

Also note that $n^{-1/2} \Phi(\beta)$ can be equivalently expressed as

$$n^{-1/2} \Phi(\beta) = n^{-1/2} \sum_{i=1}^n \{a_i(\beta) + b_i(\beta)\},$$

where $\Phi(\beta)$ is given by (8),

$$a_i(\beta) = \int_0^\tau \left\{ Z_i t - \frac{S_1^{(1)}(\beta, t)}{S_1^{(0)}(\beta, t)} \right\} dM_i^{(1)}(t),$$

and

$$b_i(\boldsymbol{\beta}) = \int_0^\tau \left\{ \mathbf{Z}_i t - \frac{S_2^{(1)}(\boldsymbol{\beta}, t)}{S_2^{(0)}(\boldsymbol{\beta}, t)} \right\} dM_i^{(2)}(t),$$

for $i = 1, \dots, n$.

We apply the Taylor expansion about $\boldsymbol{\beta}_0$ to $n^{-1/2}\Phi(\hat{\boldsymbol{\beta}}_{U2})$ and have

$$0 = n^{-1/2}\Phi(\hat{\boldsymbol{\beta}}_{U2}) = n^{-1/2}\Phi(\boldsymbol{\beta}_0) - n^{1/2}(\hat{\boldsymbol{\beta}}_{U2} - \boldsymbol{\beta}_0)[n^{-1}I(\boldsymbol{\beta}^*)],$$

where $\boldsymbol{\beta}^*$ is between $\hat{\boldsymbol{\beta}}_{U2}$ and $\boldsymbol{\beta}_0$ and $I(\boldsymbol{\beta}) = -\partial^2 l(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top$.

By applying the similar arguments in the proof of Theorem 1 and using Conditions (C4)–(C6), we have

$$n^{1/2}(\hat{\boldsymbol{\beta}}_{U2} - \boldsymbol{\beta}_0) \xrightarrow{D} N(0, \Sigma_B(\boldsymbol{\beta}_0)^{-1}\Sigma_A(\boldsymbol{\beta}_0)\Sigma_B(\boldsymbol{\beta}_0)^{-1}),$$

where $\Sigma_A(\boldsymbol{\beta}_0)$ is the probability limit of $\frac{1}{n}\sum_{i=1}^n [a_i(\boldsymbol{\beta}_0) + b_i(\boldsymbol{\beta}_0)][a_i(\boldsymbol{\beta}_0) + b_i(\boldsymbol{\beta}_0)]^\top$ and $\Sigma_B(\boldsymbol{\beta}_0)$ is the probability limit of $n^{-1}I(\boldsymbol{\beta}_0)$.

Proof of Theorem 4:

The proof follows the similar arguments in the proofs of Theorems 2 and 3, and we can conclude that the maximum pairwise pseudo-likelihood estimator $\hat{\boldsymbol{\beta}}_\psi$ converges to $\boldsymbol{\beta}_0$ with probability one, and $n^{1/2}(\hat{\boldsymbol{\beta}}_\psi - \boldsymbol{\beta}_0)$ converges to the normal distribution with mean zero and covariance matrix $V_2(\boldsymbol{\beta}_0)^{-1}V_1(\boldsymbol{\beta}_0)V_2(\boldsymbol{\beta}_0)^{-1}$, where $V_1(\boldsymbol{\beta}_0) = 4E\{\psi_{12}(\boldsymbol{\beta}_0)\psi_{13}(\boldsymbol{\beta}_0)\}$ and $V_2(\boldsymbol{\beta}_0) = -E\{\partial\psi_{12}(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}\} = E[\rho_{12}^{\otimes 2} \exp(-\boldsymbol{\beta}_0^\top \rho_{12}) / \{1 + \exp(-\boldsymbol{\beta}_0^\top \rho_{12})^2\}]$.

Then, by following the similar arguments as in the proof of Theorem 2, we know that $n^{1/2}\epsilon(\boldsymbol{\beta}_0)$ converges in distribution to a normal distribution with mean zero and covariance matrix $\text{var}[n^{1/2}U(\boldsymbol{\beta}_0)] + \text{var}[n^{1/2}\psi(\boldsymbol{\beta}_0)] = \Sigma_A(\boldsymbol{\beta}_0) + V_1(\boldsymbol{\beta}_0)$. Under Conditions (C4)–(C6), we apply the Taylor expansion about $\boldsymbol{\beta}_0$ to $\epsilon(\hat{\boldsymbol{\beta}}_{\epsilon 2})$, and can conclude that, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\epsilon 2} - \boldsymbol{\beta}_0)$ converges to a normal distribution with mean zero and covariance matrix $[\Sigma_B(\boldsymbol{\beta}_0) + V_2(\boldsymbol{\beta}_0)]^{-1}[\Sigma_A(\boldsymbol{\beta}_0) + V_1(\boldsymbol{\beta}_0)][\Sigma_B(\boldsymbol{\beta}_0) + V_2(\boldsymbol{\beta}_0)]^{-1}$.

References

Andersen EB (1970) Asymptotic properties of conditional maximum-likelihood estimators. *J R Stat Soc B* 32:283–301
 Andersen PK, Gill RD (1982) Cox’s regression model for counting processes: a large sample study. *Ann Stat* 10:1100–1120
 Aranda-Ordaz FJ (1983) An extension of the proportional-hazards model for grouped data. *Biometrics* 39:109–117
 Betensky RA, Rabinowitz D, Tsiatis AA (2001) Computationally simple accelerated failure time regression for interval censored data. *Biometrika* 88:703–711
 Buckley JD (1984) Additive and multiplicative models for relative survival rates. *Biometrics* 40:51–62
 Finkelstein DM (1986) A proportional hazards model for interval-censored failure time data. *Biometrics* 42:845–854

- Gao F, Chan KCG (2019) Semiparametric regression analysis of length-biased interval-censored data. *Biometrics* 75:121–132
- Ghosh D (2003) Goodness-of-fit methods for additive-risk models in tumorigenicity experiments. *Biometrics* 59:721–726
- Goggins WB, Finkelstein DM (2000) A proportional hazards model for multivariate interval-censored failure time data. *Biometrics* 56:940–943
- Gross ST, Lai TL (1996) Nonparametric estimation and regression analysis with left-truncated and right-censored data. *J Am Stat Assoc* 91:1166–1180
- Hansen LP (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50:1029–1954
- Hoëffding W (1948) A class of statistics with asymptotically normal distribution. *Ann Math Stat* 19:293–325
- Huang CY, Qin J (2013) Semiparametric estimation for the additive hazards model with left-truncated and right-censored data. *Biometrika* 100:877–888
- Kalbfleisch JD (1978) Likelihood methods and nonparametric tests. *J Am Stat Assoc* 73:167–170
- Kalbfleisch JD, Prentice RL (2002) *The statistical analysis of failure time data*. John Wiley & Sons, New Jersey
- Kim JS (2003) Efficient estimation for the proportional hazards model with left-truncated and “case 1” interval-censored data. *Stat Sin* 13:519–537
- Lai TL, Ying Z (1991) Rank regression methods for left-truncated and right-censored data. *Ann Stat* 19:531–556
- Li S, Hu T, Wang P, Sun J (2017) Regression analysis of current status data in the presence of dependent censoring with applications to tumorigenicity experiments. *Comput Stat Data Anal* 110:75–86
- Li S, Tian T, Hu T, Sun J (2021) A simulation-extrapolation approach for regression analysis of misclassified current status data with the additive hazards model. *Stat Med* 40:6309–6320
- Liang KY, Qin J (2000) Regression analysis under non-standard situations: a pairwise pseudolikelihood approach. *J R Stat Soc B* 62:773–786
- Lin D, Oakes D, Ying Z (1998) Additive hazards regression with current status data. *Biometrika* 85:289–298
- Lin DY, Ying Z (1994) Semiparametric analysis of the additive risk model. *Biometrika* 81:61–71
- Ma L, Hu T, Sun J (2015) Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika* 102:731–738
- Pan W, Chappell R (1998) Computation of the NPMLE of distribution functions for interval censored and truncated data with applications to the cox model. *Comput Stat Data Anal* 28:33–50
- Pan W, Chappell R (2002) Estimation in the Cox proportional hazards model with left-truncated and interval-censored data. *Biometrics* 58:64–70
- Qin J, Shen Y (2010) Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics* 66:382–392
- Sen PK (1960) On some convergence properties of U-statistics. *Calcutta Statist Assoc Bull* 10:1–18
- Shen PS, Chen HJ, Pan WH, Chen CM (2019) Semiparametric regression analysis for left-truncated and interval-censored data without or with a cure fraction. *Comput Stat Data Anal* 140:74–87
- Shen Y, Ning J, Qin J (2009) Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *J Am Stat Assoc* 104:1192–1202
- Shen Y, Ning J, Qin J (2017) Nonparametric and semiparametric regression estimation for length-biased survival data. *Lifetime Data Anal* 23:3–24
- Sun J (1999) A nonparametric test for current status data with unequal censoring. *J R Stat Soc B* 61:243–250
- Sun J (2006) *The statistical analysis of interval-censored failure time data*. Springer, New York
- van der Vaart AW, Wellner JA (1996) *Weak convergence and empirical processes*. Springer, New York
- Wang L, Sun J, Tong X (2010) Regression analysis of case II interval-censored failure time data with the additive hazards model. *Stat Sin* 20:1709–1723
- Wang MC, Brookmeyer R, Jewell NP (1993) Statistical models for prevalent cohort data. *Biometrics* 49:1–11
- Wang P, Li D, Sun J (2021) A pairwise pseudo-likelihood approach for left-truncated and interval-censored data under the Cox model. *Biometrics* 77:1303–1314
- Wang P, Tong X, Zhao S, Sun J (2015) Efficient estimation for the additive hazards model in the presence of left-truncation and interval censoring. *Stat Interface* 8:391–402
- Wang P, Tong X, Zhao S, Sun J (2015) Regression analysis of left-truncated and case i interval-censored data with the additive hazards model. *Commun Stat Theory Methods* 44:1537–1551

- Wu F, Kim S, Qin J, Saran R, Li Y (2018) A pairwise likelihood augmented Cox estimator for left-truncated data. *Biometrics* 74:100–108
- Zhang Z, Sun J, Sun L (2005) Statistical analysis of current status data with informative observation times. *Stat Med* 24:1399–1407
- Zhu L, Tong X, Sun J (2008) A transformation approach for the analysis of interval-censored failure time data. *Lifetime Data Anal* 14:167–178

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.