

Semiparametric inference for an accelerated failure time model with dependent truncation

Takeshi Emura¹ · Weijing Wang²

Received: 29 October 2013 / Revised: 29 April 2015 / Published online: 15 July 2015
© The Institute of Statistical Mathematics, Tokyo 2015

Abstract Truncated data are commonly seen in studies of biomedicine, epidemiology, astronomy and econometrics. Existing regression methods for analyzing left-truncated and right-censored data have been developed under the assumption that the lifetime variable of interest is independent of both truncation and censoring variables. In this article, we propose a semiparametric accelerated failure time model that incorporates both covariates and the truncation variable as regressors. The proposed model utilizes the truncation information in statistical modeling and hence allows for dependent truncation. For estimation, we develop a set of estimating equations constructed from the log-rank and quasi-independence test statistics. We show that the resulting estimators are consistent and asymptotically normal. We also propose an explicit formula for variance estimation based on a kernel method. Finite-sample performances of the estimators are studied by simulations. The proposed methodology is applied to analyze a real data for illustration.

Keywords Biased sampling · Censored regression · Left truncation · Log-rank test · Product-limit estimator · Quasi-independence · Survival analysis

✉ Takeshi Emura
takeshiemura@gmail.com

Weijing Wang
wjwang@stat.nctu.edu.tw

¹ Graduate Institute of Statistics, National Central University, Zhongli, Taiwan

² Institute of Statistics, National Chiao-Tung University, Hsin-Chu, Taiwan

1 Introduction

Truncated data are commonly seen in studies of biomedicine, epidemiology, astronomy and econometrics. Such data occur when the variables of interest can be observed only if their values satisfy certain criteria. Left truncation refers to the situation that the samples are available only when the variable of interest exceeds some threshold value. An example of such data is the lifetime data collected from the Channing house retirement community in Palo Alto, California (Hyde 1980). The data records residents' lifetime (age at death) and gender subject to the criteria that a resident had to live long enough to enter the retirement center. Thus, the entry age stands for the left-truncation variable. The data also include right censoring due to subjects' withdrawal or the end-of-study effect.

Most methods for analyzing left-truncated and right-censored data have been developed under the assumption that the lifetime variable is independent of both truncation and censoring variables. In contrast to independent censorship which is not a testable condition, there exist several nonparametric tests for verifying the quasi-independence assumption on the truncation variable (Tsai 1990; Jones and Crowley 1992; Chen et al. 1996; Martin and Betensky 2005; Emura and Wang 2010; Uña-Álvarez 2012; Rodríguez Gironde and de Uña-Álvarez 2012). From these articles, we see an interesting phenomenon that the null hypothesis of quasi-independence was often rejected in some common real-world examples. This further implies that the sampling selection criteria itself contains information about the variable of interest.

In this article, we propose a semiparametric accelerated failure time (AFT) model that incorporates both covariates and the truncation variable as regressors. As an important special case, the model includes the AFT model of Lai and Ying (1991) in which the lifetime variable is conditionally independent of truncation variable given covariates. A major goal of the proposed model is to allow dependent truncation, hereby utilizing the truncation information in statistical modeling.

The article is organized as follows. In Sect. 2, we present the proposed model and inference methods. In Sect. 3, we derive asymptotic properties of the proposed estimator and suggest methods for variance estimation. Simulations and data analysis are provided in Sects. 4 and 5, respectively. Concluding remarks are given in Sect. 6. Technical proofs are provided in the appendix.

2 Proposed methodology

Let Y^* be the logarithms of lifetime. Similarly, let T and C be the logarithms of left-truncation time and right-censoring time, respectively. Also, let \mathbf{X} be a p -dimensional covariate vector without the intercept term. To avoid the trivial case, we assume $\Pr(T \leq Y^*) > 0$. We also impose a usual assumption that left truncation precedes right censoring with probability one, i.e., $\Pr(T \leq C) = 1$. This assumption includes Type I censoring of the form $C = T + \tau$ for some $\tau > 0$, which often occurs in left-truncated data (Uña-Álvarez 2010).

In left-truncated and right-censored data, observed variables include $(T, Y, \Delta, \mathbf{X})$ subject to $T \leq Y^*$, where $Y = Y^* \wedge C$, $\Delta = I(Y^* \leq C)$, $a \wedge b = \min(a, b)$

and $I(\cdot)$ is the indicator function. Left-truncated and right-censored data consist of random replications of $(T, Y, \Delta, \mathbf{X})$, denoted as $(T_i, Y_i, \Delta_i, \mathbf{X}_i)$, subject to $T_i \leq Y_i$, for $i = 1, \dots, n$. Under this sampling scheme, the cumulative distribution function of $(T, Y, \Delta, \mathbf{X})$ is

$$F(t, y, \delta, \mathbf{x}) = \Pr(T \leq t, Y^* \wedge C \leq y, I(Y^* \leq C) \leq \delta, \mathbf{X} \leq \mathbf{x} | T \leq Y^*).$$

Our objective is to utilize the information of both \mathbf{X} and T , if dependent truncation exists, in modeling the behavior of Y^* .

2.1 Preliminary

Our proposal was motivated by the work of [Lai and Ying \(1991\)](#) who considered the following AFT model:

$$Y^* = \beta'_0 \mathbf{X} + \varepsilon, \tag{1}$$

where ε is the error term with an unspecified density $f_\varepsilon(\cdot)$. The p -dimensional parameter β_0 describes the covariate effect on the log-lifetime Y^* . Traditionally, it is assumed that ε is independent of (T, C, \mathbf{X}) ([Lai and Ying 1991](#)). It follows that Y^* and (T, C) are conditionally independent given \mathbf{X} . This implies that, after adjusting for the effect of \mathbf{X} , the truncation T contains no information about the lifetime Y^* . Accordingly, [Lai and Ying \(1991\)](#) proposed a rank-based estimation approach based on left-truncated and right-censored data $(T_i, Y_i, \Delta_i, \mathbf{X}_i)$, subject to $T_i \leq Y_i$, for $i = 1, \dots, n$.

We revisit the AFT regression procedure of [Lai and Ying \(1991\)](#). Note that the residual lifetime $e_i^Y(\beta) = Y_i - \beta' X_i$ is left truncated by $e_i^T(\beta) = T_i - \beta' X_i$. In terms of the residual scale, $R_i(\beta) = \sum_j I\{e_j^T(\beta) \leq e_i^Y(\beta) \leq e_j^Y(\beta)\}$ represents the number at risk for a subject i . [Lai and Ying \(1991\)](#) proposed the log-rank type estimating function

$$U_n(\beta) = \sum_{i=1}^n \Delta_i \phi_i(\beta) \left[\mathbf{X}_i - \frac{1}{R_i(\beta)} \sum_j \mathbf{X}_j I\{e_j^T(\beta) \leq e_i^Y(\beta) \leq e_j^Y(\beta)\} \right],$$

where $\phi_i(\beta)$ is a weight function. With the Gehan weight $\phi_i(\beta) = R_i(\beta)$, it is easy to see that $U_n(\beta)$ is equal to

$$U_n^G(\beta) = \sum_{i=1}^n \sum_{j=1}^n \Delta_i (\mathbf{X}_i - \mathbf{X}_j) I\{e_j^T(\beta) \leq e_i^Y(\beta) \leq e_j^Y(\beta)\}.$$

Under the special case of no truncation (i.e., $e_j^T(\beta) = -\infty$), the form of $U_n^G(\beta)$ reduces to the usual Gehan-type estimating function (Section 7.4.3 of [Kalbfleisch and Prentice 2002](#)) which is monotone in each component of β . This useful property has been applied by [Jin et al. \(2003\)](#) to solve the estimating equation using a linear programming method and construct a resampling-based variance estimator.

To facilitate further discussions, we derive an alternative expression for $U_n^G(\boldsymbol{\beta})$. Define $\text{sgn}(x)$ as -1 , 0 , or 1 if $x < 0$, $x = 0$, or $x > 0$, respectively. After some calculations provided in Appendix A1, we get

$$U_n^G(\boldsymbol{\beta}) = - \sum_{i < j} (\mathbf{X}_i - \mathbf{X}_j) \text{sgn}\{e_i^Y(\boldsymbol{\beta}) - e_j^Y(\boldsymbol{\beta})\} I\{\tilde{e}_{ij}^T(\boldsymbol{\beta}) \leq \tilde{e}_{ij}^Y(\boldsymbol{\beta})\} Q_{ij}(\boldsymbol{\beta}), \quad (2)$$

where $\tilde{e}_{ij}^T(\boldsymbol{\beta}) = e_i^T(\boldsymbol{\beta}) \vee e_j^T(\boldsymbol{\beta})$, $a \vee b \equiv \max(a, b)$, $\tilde{e}_{ij}^Y(\boldsymbol{\beta}) = e_i^Y(\boldsymbol{\beta}) \wedge e_j^Y(\boldsymbol{\beta})$, and

$$Q_{ij}(\boldsymbol{\beta}) = \Delta_i \Delta_j + \Delta_i (1 - \Delta_j) I\{e_i^Y(\boldsymbol{\beta}) < e_j^Y(\boldsymbol{\beta})\} \\ + (1 - \Delta_i) \Delta_j I\{e_i^Y(\boldsymbol{\beta}) > e_j^Y(\boldsymbol{\beta})\}.$$

Here, $Q_{ij}(\boldsymbol{\beta})$ indicates whether a pair (i, j) is orderable and $I\{\tilde{e}_{ij}^T(\boldsymbol{\beta}) \leq \tilde{e}_{ij}^Y(\boldsymbol{\beta})\}$ indicates whether the pair is comparable (Martin and Betensky 2005). The new expression in Eq. (2) is a U-statistic which will be used in subsequent discussions.

2.2 Proposed model

We consider relaxing the independent truncation assumption by utilizing the information of T in the following regression model:

$$Y^* = \boldsymbol{\beta}'_0 \mathbf{X} + \gamma_0 T + \varepsilon, \quad (3)$$

where ε is an error term with an unspecified density function $f_\varepsilon(\cdot)$. We assume that ε and (T, C, \mathbf{X}) are independent. Note that the parameter γ_0 explicitly measures the dependency of Y^* on T , conditional on the covariate \mathbf{X} , such that

$$\gamma_0 = \frac{\text{Cov}(Y^*, T | \mathbf{X})}{\text{Var}(T | \mathbf{X})}.$$

If $\gamma_0 = 0$ is assumed to be known, the model (3) reduces to the model (1). Hence, the new model is a broader version of Lai and Ying's model under an unknown population parameter γ_0 .

One can interpret the regression parameters $\boldsymbol{\beta}_0$ as the effect of the covariate \mathbf{X} on the lifetime Y^* , after adjusting for the value of truncation time T . For the Channing house example, the covariate "gender" is coded as $X = 1$ for male and $X = 0$ for female, and the truncation time T is the potential entry age. Then, the parameter $\boldsymbol{\beta}_0$ in the model (3) represents the difference of log-lifetime between male and female residents in the Channing house having the same potential entry age.

Without truncation, we will have $\Pr(T \leq Y^*) = 1$ which happens when all possible values of T fall below the lower boundary of Y^* . In the Channing house example where T denotes the logarithm of the age entering the study, "no truncation" refers to $T = -\infty$ which means that all subjects in the target population would be recruited into the study since they were just born. However, the real Channing house data are subject

to a truncation mechanism which yields a biased sample from the target population by imposing an additional constraint $T \leq Y^*$ with $\Pr(T \leq Y^*) < 1$. We will demonstrate that with appropriate estimation schemes, such sampling bias can be corrected even if the truncation mechanism is informative.

To further illustrate the proposed model (3), we generalize the model of [Martin and Betensky \(2005\)](#) to a regression setting in which the lifetime and truncation time, after the logarithm transformation, jointly follow a bivariate normal distribution

$$\begin{bmatrix} Y^* \\ T \end{bmatrix} \sim N \left(\begin{bmatrix} \beta'_0 \mathbf{X} \\ \mu_L \end{bmatrix}, \begin{bmatrix} \sigma_{Y^*}^2 & \rho \sigma_{Y^*} \sigma_T \\ \rho \sigma_{Y^*} \sigma_T & \sigma_T^2 \end{bmatrix} \right).$$

It follows that

$$Y^* | \mathbf{X}, T \sim N \left(\beta'_0 \mathbf{X} + \rho \frac{\sigma_{Y^*}}{\sigma_T} (T - \mu_L), \sigma_{Y^*}^2 (1 - \rho^2) \right).$$

This implies that the model (3) holds with $\gamma_0 = \rho \sigma_{Y^*} / \sigma_T$ and an independent error

$$\varepsilon \sim N \left(-\rho \frac{\sigma_{Y^*}}{\sigma_T} \mu_L, \sigma_{Y^*}^2 (1 - \rho^2) \right).$$

The parameter β_0 is interpreted as the population effect of \mathbf{X} on Y^* since $Y^* | \mathbf{X} \sim N(\beta'_0 \mathbf{X}, \sigma_{Y^*}^2)$ holds true. Nevertheless, it is wrong to fit the model (1) under the framework of [Lai and Ying \(1991\)](#) since the error $Y^* - \beta'_0 \mathbf{X}$ is clearly dependent on T . As we will see in our simulations, such a naïve method yields systematic bias for estimating the population parameter β_0 .

2.3 Estimation of regression coefficients

We need a set of estimating functions for the joint estimation of $\beta = (\beta_1, \dots, \beta_p)'$ and γ . Define $\varepsilon_i^Y(\beta, \gamma) = Y_i - \beta' X_i - \gamma T_i$ and $\varepsilon_i^T(\beta, \gamma) = T_i - \beta' X_i - \gamma T_i$. Note that $\varepsilon_i^Y(\beta, \gamma)$ is left-truncated by $\varepsilon_i^T(\beta, \gamma)$ satisfying $\varepsilon_i^T(\beta, \gamma) \leq \varepsilon_i^Y(\beta, \gamma)$. In light of Eq. (2), log-rank statistics with the Gehan weight can be modified as

$$\begin{aligned} S_n^{\text{Logrank}}(\beta, \gamma) &= - \sum_{i < j} (\mathbf{X}_i - \mathbf{X}_j) \text{sgn}\{\varepsilon_i^Y(\beta, \gamma) - \varepsilon_j^Y(\beta, \gamma)\} \\ &\quad \times I\{\widetilde{\varepsilon}_{ij}^T(\beta, \gamma) \leq \widetilde{\varepsilon}_{ij}^Y(\beta, \gamma)\} O_{ij}(\beta, \gamma), \end{aligned}$$

where $\widetilde{\varepsilon}_{ij}^T(\beta, \gamma) = \varepsilon_i^T(\beta, \gamma) \vee \varepsilon_j^T(\beta, \gamma)$, $\widetilde{\varepsilon}_{ij}^Y(\beta, \gamma) = \varepsilon_i^Y(\beta, \gamma) \wedge \varepsilon_j^Y(\beta, \gamma)$, and

$$\begin{aligned} O_{ij}(\beta, \gamma) &= \Delta_i \Delta_j + \Delta_i (1 - \Delta_j) I\{\varepsilon_i^Y(\beta, \gamma) < \varepsilon_j^Y(\beta, \gamma)\} \\ &\quad + (1 - \Delta_i) \Delta_j I\{\varepsilon_i^Y(\beta, \gamma) > \varepsilon_j^Y(\beta, \gamma)\}. \end{aligned}$$

Note that $O_{ij}(\beta, \gamma) = 1$ if a pair (i, j) is orderable and $O_{ij}(\beta, \gamma) = 0$, otherwise.

The second estimating function utilizes the assumption that $\varepsilon = Y^* - \beta'_0 \mathbf{X} - \gamma_0 T$ and $T - \beta'_0 \mathbf{X} - \gamma_0 T$ are independent. Motivated by the quasi-independence test based on the conditional Kendall's tau statistics (Martin and Betensky 2005), we set

$$\begin{aligned}
 S_n^{\text{Kendall}}(\boldsymbol{\beta}, \gamma) &= \sum_{i < j} \text{sgn}[\{\varepsilon_i^T(\boldsymbol{\beta}, \gamma) - \varepsilon_j^T(\boldsymbol{\beta}, \gamma)\}\{\varepsilon_i^Y(\boldsymbol{\beta}, \gamma) - \varepsilon_j^Y(\boldsymbol{\beta}, \gamma)\}] \\
 &\quad \times I\{\widetilde{\varepsilon}_{ij}^T(\boldsymbol{\beta}, \gamma) \leq \widetilde{\varepsilon}_{ij}^Y(\boldsymbol{\beta}, \gamma)\} O_{ij}(\boldsymbol{\beta}, \gamma).
 \end{aligned}$$

The estimator $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$ jointly solves $S_n^{\text{Logrank}}(\boldsymbol{\beta}, \gamma) = \mathbf{0}$ and $S_n^{\text{Kendall}}(\boldsymbol{\beta}, \gamma) = 0$.

It is important to note that both $S_n^{\text{Logrank}}(\boldsymbol{\beta}, \gamma)$ and $S_n^{\text{Kendall}}(\boldsymbol{\beta}, \gamma)$ are non-monotonic step functions in each components of $(\boldsymbol{\beta}, \gamma)$ [see the functional property of $S_n^{\text{Logrank}}(\boldsymbol{\beta}, \gamma)$ described in Appendix A2]. Accordingly the Newton-type algorithms and even the linear programming developed by Jin et al. (2003) are not directly applicable.

We suggest the Nelder–Mead simplex algorithm for minimizing non-differentiable functions (Nelder and Mead 1965). Define an objective function

$$\{ \|S_n^{\text{Logrank}}(\boldsymbol{\beta}, \gamma)\|_2 + |S_n^{\text{Kendall}}(\boldsymbol{\beta}, \gamma)|^2 \} / n^2,$$

where $\|\cdot\|_2$ is the Euclid (L_2) norm, defined to be $\|\mathbf{a}\|_2 \equiv \mathbf{a}'\mathbf{a}$ for a vector \mathbf{a} . Then, solving the estimating functions corresponds to minimizing the objective function.

In our numerical studies, we use R optim routine for implementing the Nelder–Mead algorithm. To ascertain the right solution, one needs to supply a reliable initial value. One possible suggestion is the initial value $(\hat{\boldsymbol{\beta}}^{LY}, 0)$ where $\hat{\boldsymbol{\beta}}^{LY}$ solves $S_n^{\text{Logrank}}(\boldsymbol{\beta}, 0) = \mathbf{0}$ and is the estimator of Lai and Ying (1991) with the Gehan weight. Here, $\hat{\boldsymbol{\beta}}^{LY}$ is found by minimizing $\|S_n^{\text{Logrank}}(\boldsymbol{\beta}, 0)\|_2 / n^2$ with the initial value $\boldsymbol{\beta} = \mathbf{0}$. Another possible initial value is $(\mathbf{0}, 0)$. In our simulations (Sect. 4), we use $(\hat{\boldsymbol{\beta}}^{LY}, 0)$ since it reaches the solutions through all the repetitions. In real data analysis, we suggest trying both methods.

2.4 Estimation of survival function

We consider estimation of the survival function $S_e(t) = \Pr(\varepsilon > t)$. Notice that $\{(\varepsilon_i^T(\hat{\boldsymbol{\beta}}, \hat{\gamma}), \varepsilon_i^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma}), \Delta_i); i = 1, \dots, n\}$ form approximately homogeneous (identically distributed) samples of left-truncated and right-censored data subject to $\varepsilon_i^T(\hat{\boldsymbol{\beta}}, \hat{\gamma}) \leq \varepsilon_i^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma})$. Therefore, we can estimate $S_e(t)$ by the product-limit estimator (Tsai et al. 1987)

$$\hat{S}_e(t; \hat{\boldsymbol{\beta}}, \hat{\gamma}) = \prod_{u \leq t} \left\{ 1 - \frac{\sum_j I(\varepsilon_j^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma}) = u, \Delta_j = 1)}{\sum_j I(\varepsilon_j^T(\hat{\boldsymbol{\beta}}, \hat{\gamma}) \leq u \leq \varepsilon_j^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma}))} \right\}.$$

Remark Since $\{(\varepsilon_i^T(\hat{\beta}, \hat{\gamma}), \varepsilon_i^Y(\hat{\beta}, \hat{\gamma}), \Delta_i)\}$ and $\{(\varepsilon_j^T(\hat{\beta}, \hat{\gamma}), \varepsilon_j^Y(\hat{\beta}, \hat{\gamma}), \Delta_j)\}$ are dependent for $i \neq j$, the asymptotic properties of $\hat{S}_\varepsilon(t; \hat{\beta}, \hat{\gamma})$ do not directly follow from those of the product-limit estimator. As for the consistency of $\hat{S}_\varepsilon(t; \hat{\beta}, \hat{\gamma})$, this dependency will be asymptotically negligible (Theorem 3). However, the dependency makes it difficult to derive the asymptotic normality of $\hat{S}_\varepsilon(t; \hat{\beta}, \hat{\gamma})$, and also influences the asymptotic variance.

3 Asymptotic analysis

3.1 Asymptotic theory

Let $\Theta \subset R^{p+1}$ be the parameter space for (β, γ) , and $(\beta_0, \gamma_0) \in \Theta$ be the true parameter value. Asymptotic properties of $(\hat{\beta}, \hat{\gamma})$ are constructed by expressing the estimating functions as a functional of the empirical distribution function,

$$F_n(t, y, \delta, \mathbf{x}) = \frac{1}{n} \sum_{i=1}^n I(T_i \leq t, Y_i \leq y, \Delta_i \leq \delta, \mathbf{X}_i \leq \mathbf{x}),$$

and then applying the asymptotic theory for empirical processes (e.g., [Van Der Vaart and Wellner 1996](#)). By straightforward calculations given in Appendix B.1,

$$\begin{aligned} \Phi(F_n; \beta, \gamma) &\equiv \frac{1}{n^2} \begin{bmatrix} S_n^{\text{Logrank}}(\beta, \gamma) \\ S_n^{\text{Kendall}}(\beta, \gamma) \end{bmatrix} \\ &= \iint \mathbf{h}\{(t_1, y_1, \delta_1, \mathbf{x}_1), (t_2, y_2, \delta_2, \mathbf{x}_2); \beta, \gamma\} dF_n(t_1, y_1, \delta_1, \mathbf{x}_1) \\ &\quad \times dF_n(t_2, y_2, \delta_2, \mathbf{x}_2), \end{aligned} \tag{4}$$

where \mathbf{h} is a deterministic function given by

$$\begin{aligned} &\mathbf{h}\{(t_1, y_1, \delta_1, \mathbf{x}_1), (t_2, y_2, \delta_2, \mathbf{x}_2); \beta, \gamma\} \\ &\equiv \frac{1}{2} \begin{bmatrix} -(\mathbf{x}_1 - \mathbf{x}_2) \text{sgn}\{y_1 - \beta' \mathbf{x}_1 - \gamma t_1 - (y_2 - \beta' \mathbf{x}_2 - \gamma t_2)\} \\ \text{sgn}\{(t_1 - \beta' \mathbf{x}_1 - \gamma t_1 - (t_2 - \beta' \mathbf{x}_2 - \gamma t_2))(y_1 - \beta' \mathbf{x}_1 - \gamma t_1 - (y_2 - \beta' \mathbf{x}_2 - \gamma t_2))\} \end{bmatrix} \\ &\times I\{(t_1 - \beta' \mathbf{x}_1 - \gamma t_1) \vee (t_2 - \beta' \mathbf{x}_2 - \gamma t_2) \leq (y_1 - \beta' \mathbf{x}_1 - \gamma t_1) \wedge (y_2 - \beta' \mathbf{x}_2 - \gamma t_2)\} \\ &\times \{\delta_1 \delta_2 + \delta_1 (1 - \delta_2) I(y_1 - \beta' \mathbf{x}_1 - \gamma t_1 < y_2 - \beta' \mathbf{x}_2 - \gamma t_2) \\ &\quad + (1 - \delta_1) \delta_2 I(y_1 - \beta' \mathbf{x}_1 - \gamma t_1 > y_2 - \beta' \mathbf{x}_2 - \gamma t_2)\}. \end{aligned}$$

Since $\Phi(F_n; \beta, \gamma) = \mathbf{0}$ holds at $(\beta, \gamma) = (\hat{\beta}, \hat{\gamma})$, one can regard $(\hat{\beta}, \hat{\gamma})$ as a Z-estimator based on the function $\Phi(F_n; \cdot) : \Theta \mapsto R^{p+1}$. As $\Phi(F_n; \beta, \gamma)$ is not differentiable with respect to (β, γ) , one cannot apply the classical technique based on a Taylor expansion of $\Phi(F_n; \beta, \gamma)$ around (β_0, γ_0) . Instead, we apply the empirical process theory for Z-estimators that accommodate the cases of non-differentiable functions (Sec. 3.3 of [Van Der Vaart and Wellner 1996](#)). The corresponding deterministic function is $\Phi(F; \cdot) : \Theta \mapsto R^{p+1}$ in which F_n is replaced by the true distribution F .

We impose the following conditions:

Assumption 1 There exists some constant $M > 0$ such that $\|\mathbf{X}\|_2 \leq M$ with probability one, where $\|\cdot\|_2$ is the Euclid (L_2) norm.

Assumption 2 For every $\varepsilon > 0$, $\inf_{\|(\boldsymbol{\beta}, \gamma) - (\boldsymbol{\beta}_0, \gamma_0)\|_2 \geq \varepsilon} \|\Phi(F; \boldsymbol{\beta}, \gamma)\|_2 > 0 = \|\Phi(F; \boldsymbol{\beta}_0, \gamma_0)\|_2$.

Assumption 1 is a requirement for the function \mathbf{h} to be bounded, which is sufficient for the functional $\Phi(\cdot; \boldsymbol{\beta}_0, \gamma_0)$ to be Hadamard differentiable. The same assumption is imposed by [Lai and Ying \(1991\)](#). Assumption 2 is often called ‘‘identifiability’’ ensuring that the true value $(\boldsymbol{\beta}_0, \gamma_0) \in \Theta$ is a well-separated point from $(\boldsymbol{\beta}, \gamma) \neq (\boldsymbol{\beta}_0, \gamma_0)$. This is the standard assumption for guaranteeing the consistency of M- or Z-estimator (Sec. 3.3 of [Van Der Vaart and Wellner, 1996](#); Sec. 5.2 of [Van Der Vaart 1998](#)).

Theorem 1 Under Assumptions 1, 2, the estimators $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$ converge in probability to the true parameter value $(\boldsymbol{\beta}_0, \gamma_0)$.

The proof of Theorem 1 is given in Appendix B.2.

Under Assumption 1, the functional $\Phi(\cdot; \boldsymbol{\beta}_0, \gamma_0)$ is Hadamard differentiable and hence the functional delta method applies (Sec. 20.2 of [Van Der Vaart 1998](#)). Then, we obtain the expression

$$\begin{aligned} & n^{1/2}\{\Phi(F_n; \boldsymbol{\beta}_0, \gamma_0) - \Phi(F; \boldsymbol{\beta}_0, \gamma_0)\} \\ &= n^{-1/2} \sum_{j=1}^n \phi_F\{(T_j, Y_j, \Delta_j, \mathbf{X}_j); \boldsymbol{\beta}_0, \gamma_0\} + o_P(1) \end{aligned} \quad (5)$$

where

$$\begin{aligned} & \phi_F\{(T_j, Y_j, \Delta_j, \mathbf{X}_j); \boldsymbol{\beta}, \gamma\} \\ &= 2 \int \mathbf{h}\{(t_1, y_1, \delta_1, \mathbf{x}_1), (T_j, Y_j, \Delta_j, \mathbf{X}_j); \boldsymbol{\beta}, \gamma\} dF(t_1, y_1, \delta_1, \mathbf{x}_1) \\ & \quad - 2 \iint \mathbf{h}\{(t_1, y_1, \delta_1, \mathbf{x}_1), (t_2, y_2, \delta_2, \mathbf{x}_2); \boldsymbol{\beta}, \gamma\} \\ & \quad \times dF(t_1, y_1, \delta_1, \mathbf{x}_1) dF(t_2, y_2, \delta_2, \mathbf{x}_2). \end{aligned} \quad (6)$$

We will establish the asymptotic normality under the additional conditions:

Assumption 3 The derivatives $\partial\Phi(F; \boldsymbol{\beta}, \gamma)/\partial(\boldsymbol{\beta}, \gamma)$ exist and continuous. Also, the matrix $\mathbf{A}_0 \equiv \partial\Phi(F; \boldsymbol{\beta}, \gamma)/\partial(\boldsymbol{\beta}, \gamma)|_{\boldsymbol{\beta}_0, \gamma_0}$ is non-singular.

Assumption 4 A class of functions

$$\mathfrak{S} = \{\phi_F(\cdot; \boldsymbol{\beta}, \gamma) - \phi_F(\cdot; \boldsymbol{\beta}_0, \gamma_0) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0, \gamma - \gamma_0\|_2 < \delta\}$$

is *F-Donsker* for some $\delta > 0$.

Assumption 3 insures the invertibility of the matrix \mathbf{A}_0 for the Z-estimator, which is essential not only for the asymptotic normality proof, but also for variance estimation (Sect. 3.2). In Assumption 4, whether the class of functions is Donsker depends on the “size” of \mathfrak{F} which may reduce as δ approaches to zero [see Section 19.2 of Van Der Vaart 1998]. As seen from Appendix B.3, Assumption 4 is a sufficient condition to verify the stochastic condition

$$\begin{aligned} & n^{1/2}\{\Phi(F_n; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \Phi(F; \hat{\boldsymbol{\beta}}, \hat{\gamma})\} - n^{1/2}\{\Phi(F_n; \boldsymbol{\beta}_0, \gamma_0) - \Phi(F; \boldsymbol{\beta}_0, \gamma_0)\} \\ & = o_P(1 + n^{1/2}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \hat{\gamma} - \gamma_0\|). \end{aligned}$$

Owing to the non-differentiability of $\Phi(F_n; \boldsymbol{\beta}, \gamma)$ with respect to $(\boldsymbol{\beta}, \gamma)$, the above stochastic condition is needed to regulate the asymptotic behavior of $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$.

Theorem 2 Under Assumptions 1, 2, 3, 4,

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \hat{\gamma} - \gamma_0)' = \mathbf{A}_0^{-1}n^{-1/2} \sum_{j=1}^n \phi_F\{(T_j, Y_j, \Delta_j, \mathbf{X}_j); \boldsymbol{\beta}_0, \gamma_0\} + o_P(1), \tag{7}$$

where ϕ_F is the mean zero random vector defined in Eq. (6) and \mathbf{A}_0 is defined in Assumption 3. Accordingly, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \hat{\gamma} - \gamma_0)$ converges weakly to a multivariate normal distribution with mean zero and covariance matrix $\mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}$, where $\mathbf{B}_0 \equiv E[\phi_F\{(T_j, Y_j, \Delta_j, \mathbf{X}_j); \boldsymbol{\beta}_0, \gamma_0\}\phi_F\{(T_j, Y_j, \Delta_j, \mathbf{X}_j); \boldsymbol{\beta}_0, \gamma_0\}']$.

The proof of Theorem 2 is given in Appendix B.3.

To prove the uniform consistency of the product-limit estimator $\hat{S}_\varepsilon(t; \hat{\boldsymbol{\beta}}, \hat{\gamma})$, we need an additional assumption. Let the interval $[a, b]$, where $a < b$, be the support of ε , where $a = \inf\{u; S_\varepsilon(u) < 1\}$ and $b = \sup\{u; S_\varepsilon(u) > 0\}$. Also, let

$$\begin{aligned} \pi_\varepsilon(s, \boldsymbol{\beta}, \gamma) &= E[I(T - \boldsymbol{\beta}'X - \gamma T \leq s) \\ &\quad \times S_\varepsilon\{s + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'X + (\gamma - \gamma_0)T\}S_C\{s + \boldsymbol{\beta}'X + \gamma T\}], \end{aligned}$$

where S_C is the survival function for C , and $E[\cdot]$ is taken over the joint distribution of $(T, \mathbf{X})|T \leq Y$. It is not difficult to see that $\sum_j I\{\varepsilon_j^T(\boldsymbol{\beta}, \gamma) \leq s \leq \varepsilon_j^Y(\boldsymbol{\beta}, \gamma)\}/n$ converges in probability to $\pi_\varepsilon(s, \boldsymbol{\beta}, \gamma)$. To establish the uniform consistency, it is natural to assume that $\pi_\varepsilon(s, \boldsymbol{\beta}, \gamma)$ is positive on $s \in [a, b]$, in the neighborhood of $(\boldsymbol{\beta}_0, \gamma_0)$.

Assumption 5 For some $\delta > 0$, $\inf_{\|(\boldsymbol{\beta}, \gamma) - (\boldsymbol{\beta}_0, \gamma_0)\| \leq \delta} \inf_{s \in [a, b]} \pi_\varepsilon(s, \boldsymbol{\beta}, \gamma) > 0$.

Theorem 3 Under Assumptions 1, 2 and 5, the product-limit estimator $\hat{S}_\varepsilon(t; \hat{\boldsymbol{\beta}}, \hat{\gamma})$ converges in probability to $S_\varepsilon(t)$, uniformly over $t \in [a, b]$.

The proof of Theorem 3 is given in Appendix B.4.

3.2 Variance estimation

Even for the usual AFT model in (1) without truncation, resampling based methods are often suggested for variance estimation (Jin et al. 2003). However, based on Theorem 2, we can directly estimate $\mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}$ by $\hat{\mathbf{A}}_0^{-1}\hat{\mathbf{B}}_0\hat{\mathbf{A}}_0^{-1}$ as defined below. We propose to estimate \mathbf{B}_0 by

$$\hat{\mathbf{B}}_0 = \sum_{j=1}^n \phi_{F_n}(T_j, Y_j, \Delta_j, \mathbf{X}_j; \hat{\boldsymbol{\beta}}, \hat{\gamma}) \phi_{F_n}(T_j, Y_j, \Delta_j, \mathbf{X}_j; \hat{\boldsymbol{\beta}}, \hat{\gamma})' / n,$$

where

$$\begin{aligned} &\phi_{F_n}(T_j, Y_j, \Delta_j, \mathbf{X}_j; \hat{\boldsymbol{\beta}}, \hat{\gamma}) \\ &= -2 \int \int \mathbf{h}\{(t_1, y_1, \delta_1, \mathbf{x}_1), (t_2, y_2, \delta_2, \mathbf{x}_2); \hat{\boldsymbol{\beta}}, \hat{\gamma}\} dF_n(t_1, y_1, \delta_1, \mathbf{x}_1) dF_n(t_2, y_2, \delta_2, \mathbf{x}_2) \\ &\quad + 2 \int \mathbf{h}\{(t_1, y_1, \delta_1, \mathbf{x}_1), (T_j, Y_j, \Delta_j, \mathbf{X}_j); \hat{\boldsymbol{\beta}}, \hat{\gamma}\} dF_n(t_1, y_1, \delta_1, \mathbf{x}_1) \\ &= n^{-1} \sum_{i=1}^n \left[\begin{array}{c} -(\mathbf{X}_i - \mathbf{X}_j) \text{sgn}\{\varepsilon_i^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma}) - \varepsilon_j^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma})\} \\ \text{sgn}\{(\varepsilon_i^T(\hat{\boldsymbol{\beta}}, \hat{\gamma}) - \varepsilon_j^T(\hat{\boldsymbol{\beta}}, \hat{\gamma}))(\varepsilon_i^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma}) - \varepsilon_j^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma}))\} \end{array} \right] \\ &\quad \times I\{\tilde{\varepsilon}_{ij}^T(\hat{\boldsymbol{\beta}}, \hat{\gamma}) \leq \tilde{\varepsilon}_{ij}^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma})\} O_{ij}(\hat{\boldsymbol{\beta}}, \hat{\gamma}). \end{aligned}$$

Estimation of the $(p + 1) \times (p + 1)$ matrix $\mathbf{A}_0 = \partial\Phi(F; \boldsymbol{\beta}, \gamma) / \partial(\boldsymbol{\beta}, \gamma) |_{\boldsymbol{\beta}_0, \gamma_0}$ is more challenging since $\Phi(F_n; \boldsymbol{\beta}, \gamma)$ is a step function in each component of $(\boldsymbol{\beta}, \gamma)$ and hence not differentiable. We propose to smooth out the jumps by a kernel density function K satisfying $K(u) = K(-u)$, $\int K(u)du = 1$ and $\int s^2 K(s)ds \equiv \mu_2(K) < \infty$. The resulting kernel estimator is $\hat{\mathbf{A}}_0 = [\hat{\mathbf{A}}_0^{(1)}(\hat{\boldsymbol{\beta}}, \hat{\gamma}; b_1), \dots, \hat{\mathbf{A}}_0^{(p+1)}(\hat{\boldsymbol{\beta}}, \hat{\gamma}; b_{p+1})]$, where

$$\begin{aligned} &\hat{\mathbf{A}}_0^{(k)}(\boldsymbol{\beta}, \gamma; b_k) \\ &= \int_{u \neq 0} \frac{1}{u} \Phi(F_n; \beta_1, \dots, \beta_k + u, \dots, \beta_p, \gamma) \frac{1}{b_k} K\left(\frac{u}{b_k}\right) du, \quad k = 1, \dots, p, \\ &\hat{\mathbf{A}}_0^{(p+1)}(\boldsymbol{\beta}, \gamma; b_{p+1}) = \int_{u \neq 0} \frac{1}{u} \Phi(F_n; \boldsymbol{\beta}, \gamma + u) \frac{1}{b_{p+1}} K\left(\frac{u}{b_{p+1}}\right) du, \end{aligned}$$

where $b_k > 0$ is the bandwidth. We particularly use the standard normal density for K .

Using the Taylor expansion under certain smoothness assumptions on $\Phi(F; \boldsymbol{\beta}, \gamma)$, the asymptotic mean square error is calculated as

$$\begin{aligned} \text{MSE}\{\hat{\mathbf{A}}_0^{(k)}(\hat{\boldsymbol{\beta}}, \hat{\gamma}; b_k)\} &= \text{Var}\{\hat{\mathbf{A}}_0^{(k)}(\hat{\boldsymbol{\beta}}, \hat{\gamma}; b_k)\} + \text{Bias}[\hat{\mathbf{A}}_0^{(k)}(\boldsymbol{\beta}, \gamma; b_k)]^2 \\ &\approx \frac{1}{nb_k} E \left[\frac{\partial}{\partial \beta_k} \phi_F\{(T_j, Y_j, \delta_j, \mathbf{X}_j); \boldsymbol{\beta}_0, \gamma_0\} \right]^2 \\ &\quad + \frac{b_k^4}{36} \left[\frac{\partial^3}{\partial \beta_k^3} \Phi(F; \boldsymbol{\beta}_0, \gamma_0) \right]^2 \mu_2(K)^2, \end{aligned}$$

The bandwidth that minimizes the preceding expression becomes

$$b_k^{\text{opt}} = \left(9E \left[\frac{\partial}{\partial \beta_k} \phi_F\{(T_j, Y_j, \delta_j, \mathbf{X}_j); \boldsymbol{\beta}_0, \gamma_0\} \right]^2 \left[\frac{\partial^3}{\partial \beta_k^3} \Phi(F; \boldsymbol{\beta}_0, \gamma_0) \right]^{-2} \mu_2(K)^{-2} \right)^{1/5} \frac{1}{n^{1/5}},$$

for $k = 1, \dots, p$. Thus, the optimal convergence rate is $b_k^{\text{opt}} = O(n^{-1/5})$.

Unfortunately, estimation of the unknown part of b_k^{opt} is even more difficult than the variance estimation. In addition, commonly used cross-validation schemes are computationally prohibitive. We instead use the idea of Silverman’s reference bandwidth under the normal kernel (Sheather 2004) and set

$$\hat{b}_k = 0.5 \min(S_k, \text{IQR}_k/1.34) n^{-1/5},$$

where S_k^2 is the sample variance and IQR_k is the sample inter-quartile range for the jumps of $\mathbf{S}_n^{\text{Logrank}}(\boldsymbol{\beta}, \gamma)$ with respect to β_k (Appendix C). Although this choice may be somewhat ad hoc, we prefer this approach owing to its computational simplicity and reasonable numerical performance. The bandwidth estimator \hat{b}_{p+1} is obtained by a similar fashion (Appendix C).

One can construct the standard error and confidence interval for $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$ using the proposed variance estimator. For instance, the standard error $se(\hat{\beta}_k)$ of $\hat{\beta}_k$ is the square root of the k th diagonal element of $\hat{\mathbf{A}}_0^{-1} \hat{\mathbf{B}}_0 \hat{\mathbf{A}}_0^{-1}$ divided by $n^{1/2}$. On the basis of the asymptotic normality (Theorem 2), a $(1 - \alpha) \times 100\%$ confidence interval becomes

$$[\hat{\beta}_k - z_{\alpha/2} se(\hat{\beta}_k), \hat{\beta}_k + z_{\alpha/2} se(\hat{\beta}_k)],$$

where $z_{\alpha/2}$ is the upper $(\alpha/2) \times 100\%$ point of $N(0, 1)$. The standard error and confidence interval for $\hat{\gamma}$ are derived similarly. If the confidence interval does not cover $\gamma = 0$, one can reject the independence assumption between T and Y^* given \mathbf{X} .

Estimation of the standard error for $\hat{S}_\varepsilon(t; \hat{\boldsymbol{\beta}}, \hat{\gamma})$ is more challenging. Note that the Greenwood-type variance formula produces serious under-estimation since $\{(\varepsilon_i^T(\hat{\boldsymbol{\beta}}, \hat{\gamma}), \varepsilon_i^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma}), \Delta_i)\}$ and $\{(\varepsilon_j^T(\hat{\boldsymbol{\beta}}, \hat{\gamma}), \varepsilon_j^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma}), \Delta_j)\}$ are not independent for $i \neq j$. A better alternative is the bootstrap. For $b = 1, \dots, B$, where B is usually 500 or 1000, a resample $\{(T_i^{(b)}, Y_i^{(b)}, \Delta_i^{(b)}, \mathbf{X}_i^{(b)}); (i = 1, \dots, n)\}$ is taken out with replacement from the observation. Then, for each b , we calculate the estimate

$\hat{S}_\varepsilon^{(b)}(t; \hat{\beta}^{(b)}, \hat{\gamma}^{(b)})$ based on the resample. The sample standard deviation of the B estimates can be used for the standard error for $\hat{S}_\varepsilon(t; \hat{\beta}, \hat{\gamma})$. Some simulations, not reported here, show that the bootstrap performs reasonably well.

4 Simulations

We conduct simulations to examine the performance of the proposed method and also to compare the proposed method with the method of [Lai and Ying \(1991\)](#).

4.1 Simulation design

We generate data under a bivariate normal model

$$\begin{bmatrix} Y^* \\ T \end{bmatrix} \sim N \left(\begin{bmatrix} \beta_0 X \\ -1 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right). \quad (8)$$

Here, the covariate X follows the uniform distribution on $[0,1]$. The censoring time C follows the conditional distribution $C^*|C^* > T$, where $C^* \sim N(1, 1)$. Then, left-truncated and right-censored data $(T_i, Y_i, \Delta_i, X_i)$ subject to $T_i \leq Y_i$, for $i = 1, \dots, n$, are generated.

Note that model (8) leads to

$$Y^* = \beta_0 X + \gamma_0 T + \varepsilon,$$

where $\varepsilon \sim N(\gamma_0, 1 - \gamma_0^2)$ and $\gamma_0 = \rho$. The corresponding survival function for ε is

$$S_\varepsilon(t) = 1 - \Phi \left\{ \frac{t - \gamma_0}{(1 - \gamma_0^2)^{1/2}} \right\}, \quad (9)$$

and the truncation probability is

$$\Pr(T \leq Y^*) = \int_0^1 \Phi \left[\frac{1 + \beta_0 u}{\{2(1 - \gamma_0)\}^{1/2}} \right] du,$$

where Φ is the cumulative distribution function for $N(0, 1)$. We examine six parameter configurations, namely $(\beta_0, \gamma_0) = (0, -0.3), (0, 0), (0, 0.3), (1, -0.3), (1, 0),$ and $(1, 0.3)$, which yield the truncation probability $\Pr(T \leq Y^*) = 0.73, 0.76, 0.80, 0.82, 0.85,$ and 0.89 , respectively. The corresponding censoring probability is $\Pr(C < Y^*|T \leq Y^*) = 0.28, 0.27, 0.24, 0.40, 0.38,$ and 0.36 , respectively. Two sample sizes are considered: $n = 150$ and 300 .

For simulated data, we compute $(\hat{\beta}, \hat{\gamma})$ by following the Nelder–Mead algorithm implemented in the R `optim` routine (Sect. 2.3). The variance estimator $\hat{\mathbf{A}}_0^{-1} \hat{\mathbf{B}}_0 \hat{\mathbf{A}}_0^{-1}$ for $(\hat{\beta}, \hat{\gamma})$ and the 95 % confidence intervals are then computed using the methods of

Sect. 3.2. We have checked that the routine successfully ascertains the right solution and that the matrix $\hat{\mathbf{A}}_0$ is invertible through the 500 replications (not shown).

4.2 Simulation results

Table 1 presents the results for estimating (β_0, γ_0) . In all the cases, the proposed estimator provides almost unbiased results. The standard deviation of the estimates decreases as the sample size increases from $n = 150$ to 300.

Table 1 also compares the standard deviations with the standard errors. For $n=150$, the standard error $\text{se}(\hat{\beta})$ tends to overestimate the standard deviation of $\hat{\beta}$. This upward bias is due to a few outlying variance estimates in the 500 replications, and hence does not affect the coverage probability much. The upward bias vanishes as the sample size increases from $n = 150$ to $n = 300$. In a similar fashion, the standard error $\text{se}(\hat{\gamma})$ is sometimes biased for the standard deviation of $\hat{\gamma}$ with $n = 150$, but it is nearly unbiased with $n = 300$. These results imply that the kernel approach to variance estimation works well for large samples. Accordingly the empirical coverage probabilities for the 95 % confidence intervals are all close to the nominal level.

Table 2 compares the performance between the proposed estimator and the estimator of Lai and Ying (1991) with the Gehan weight. The latter is obtained by solving $\mathbf{U}_n^G(\beta) = 0$, or equivalently solving $\mathbf{S}_n^{\text{Logrank}}(\beta, \gamma) = 0$ under $\gamma = 0$ (Sect. 2.3). As expected from our discussions in Sect. 2.2, Lai and Ying's estimator produces systematic bias for the cases of $\gamma_0 \neq 0$ (Table 2). The biases do not vanish as the sample size increases from $n = 150$ to $n = 300$, and the direction of bias is determined by the sign of γ_0 . If $\gamma_0 = 0$, both the proposed method and Lai and Ying's method are nearly unbiased, but the standard deviations of Lai and Ying's estimator are slightly smaller. However, the loss of efficiency in the proposed estimator is quite modest.

Table 3 examines the performance of $\hat{S}_\varepsilon(t; \hat{\beta}, \hat{\gamma})$ at selected values of t satisfying $S_\varepsilon(t) = 0.25, 0.50$, and 0.75 , which can be obtained from Eq. (9). The results show that the estimator $\hat{S}_\varepsilon(t; \hat{\beta}, \hat{\gamma})$ is nearly unbiased in all the cases. The standard deviation decreases as the sample size n increases from 150 to 300.

5 Data analysis

To illustrate the proposed methodology, we analyze the data collected from the Channing house retirement center available in Appendix 1 of Hyde (1980). The data consist of 462 subjects (97 men and 365 women) whose lifetime is left truncated by the entry age. Among them, 286 subjects are right censored, yielding the censoring proportion 0.62. The covariate "gender" is coded as $X = 1$ for male and $X = 0$ for female. We also include the truncation time (entry age) as a regressor. Our aim is to study the effects of gender and entry age jointly on the lifetime.

We fit the model $Y^* = \beta_0 X + \gamma_0 T + \varepsilon$ and obtain $\hat{\beta} = -0.030$ (se = 0.018) and $\hat{\gamma} = 0.26$ (se = 0.12). Here, the point estimates are obtained by the proposed algorithm (Sect. 2.3) with the initial values $(\beta = 0, \gamma = 0)$. The same estimates are obtained with the other initial values $(\hat{\beta}^{LY} = -0.036, \gamma = 0)$. The 95 % confidence interval for γ is (0.048, 0.468) which indicates that the entry age and lifetime are positively

Table 1 Simulation results for the proposed estimator $(\hat{\beta}, \hat{\gamma})$

(β_0, γ_0)	n	Estimation of β_0			Estimation of γ_0				
		Bias($\hat{\beta}$)	SD($\hat{\beta}$)	E{se($\hat{\beta}$)}	95 % Cov	Bias($\hat{\gamma}$)	SD($\hat{\gamma}$)	E{se($\hat{\gamma}$)}	95 % Cov
(0, -0.3)	150	0.049	0.384	0.414	0.954	0.001	0.185	0.185	0.930
	300	0.019	0.260	0.287	0.952	-0.001	0.119	0.122	0.950
(0, 0)	150	0.010	0.405	0.419	0.942	-0.012	0.158	0.175	0.958
	300	0.004	0.279	0.287	0.950	0.002	0.112	0.116	0.946
(0, 0.3)	150	-0.014	0.417	0.467	0.954	-0.063	0.171	0.297	0.966
	300	-0.016	0.285	0.292	0.936	-0.016	0.128	0.135	0.924
(1, -0.3)	150	0.021	0.369	0.400	0.968	-0.008	0.157	0.162	0.956
	300	-0.002	0.256	0.275	0.964	-0.001	0.112	0.112	0.964
(1, 0)	150	0.043	0.424	0.425	0.948	-0.015	0.158	0.160	0.966
	300	0.012	0.287	0.288	0.950	-0.008	0.105	0.108	0.962
(1, 0.3)	150	-0.001	0.375	0.410	0.964	0.004	0.167	0.144	0.946
	300	-0.008	0.255	0.278	0.960	-0.002	0.092	0.095	0.956

The average of biases (Bias), standard deviation (SD), the average of standard error estimate (E{se($\hat{\beta}$)}), and the empirical coverage probability of 95 % confidence interval (95 % Cov) based on 500 repetitions are reported

Table 2 Simulation results for comparing the proposed estimator with the estimator of Lai and Ying (1991)

(β_0, γ_0)	n	Proposed method			Lai and Ying (1991)		
		$E(\hat{\beta})$	Bias($\hat{\beta}$)	SD($\hat{\beta}$)	$E(\hat{\beta})$	Bias($\hat{\beta}$)	SD($\hat{\beta}$)
$(-1, -0.3)$	150	-0.939	0.061	0.392	-0.849	0.151	0.329
	300	-1.000	0.000	0.277	-0.871	0.129	0.238
$(-1, 0)$	150	-1.019	-0.019	0.460	-1.026	-0.026	0.443
	300	-1.004	-0.004	0.324	-1.008	-0.008	0.307
$(-1, 0.3)$	150	-1.099	-0.099	0.469	-1.256	-0.256	0.506
	300	-1.009	-0.009	0.317	-1.226	-0.226	0.358
$(1, -0.3)$	150	1.021	0.021	0.369	0.888	-0.112	0.336
	300	0.998	-0.002	0.256	0.883	-0.117	0.236
$(1, 0)$	150	1.043	0.043	0.424	1.027	0.027	0.411
	300	1.012	0.012	0.287	1.007	0.007	0.281
$(1, 0.3)$	150	0.999	-0.001	0.375	1.165	0.165	0.429
	300	0.992	-0.008	0.255	1.161	0.161	0.294

The average of estimates ($E(\hat{\beta})$), average of biases (Bias) and standard deviation (SD) are calculated based on 500 repetitions

Table 3 Simulation results for the proposed estimator $\hat{S}_\varepsilon(t; \hat{\beta}, \hat{\gamma})$

(β_0, γ_0)	n	$S_\varepsilon(t) = 0.25$		$S_\varepsilon(t) = 0.5$		$S_\varepsilon(t) = 0.75$	
		Bias	SD	Bias	SD	Bias	SD
$(0, -0.3)$	150	0.008	0.112	-0.003	0.151	-0.017	0.142
	300	0.003	0.079	-0.001	0.104	-0.009	0.095
$(0, 0)$	150	0.003	0.106	-0.004	0.142	-0.016	0.133
	300	0.003	0.072	0.002	0.099	-0.002	0.091
$(0, 0.3)$	150	-0.014	0.118	-0.038	0.160	-0.056	0.157
	300	-0.001	0.081	-0.008	0.119	-0.016	0.126
$(1, -0.3)$	150	0.002	0.102	-0.006	0.138	-0.015	0.129
	300	0.006	0.075	0.002	0.099	-0.004	0.089
$(1, 0)$	150	-0.004	0.102	-0.014	0.141	-0.021	0.133
	300	-0.002	0.072	-0.007	0.097	-0.009	0.089
$(1, 0.3)$	150	0.005	0.102	-0.002	0.128	-0.006	0.119
	300	0.002	0.069	-0.001	0.091	-0.003	0.084

The average of biases (Bias) and standard deviation (SD) based on 500 simulation repetitions are reported

associated (p – value = 0.016). This result agrees with Tsai’s quasi-independence test with the estimated conditional Kendall tau equal to 0.088 (p – value = 0.076) (see Emura and Wang 2010). The fitted value $\hat{\beta} = -0.030$ implies that women seemed to live longer than men but the gender difference is not statistically significant since the corresponding 95 % confidence interval $(-0.066, 0.007)$ covers the zero. This observation agrees with the previously reported results from the Cox proportional hazard

regression analysis under the independent left-truncation assumption [see Example 9.4 in p.313 of Klein and Moeschberger (2003)].

6 Concluding remarks

Existing regression models and methods developed for left-truncated and right-censored data focus on the covariate effect on lifetime under the assumption that lifetime is independent of truncation time. However, empirical evidence has shown that truncation may reveal useful information on the lifetime of interest (Tsai 1990; Jones and Crowley 1992; Chen et al. 1996; Martin and Betensky 2005; Emura and Wang 2010; Rodríguez Gironde and de Uña-Álvarez 2012). In this article, we explicitly include the truncation variable, along with other covariates, in the proposed semiparametric AFT regression model. We implement computing routines for the proposed method in R *depend.truncation* package which is freely available through the Comprehensive R Archive Network (CRAN) <http://cran.rproject.org/>.

Jones and Crowley (1992) also included the truncation variable as a covariate based on their models. However, their paper focused only on testing the assumption of quasi-independence without estimating the models.

Another approach for dependent truncation is based on copula models, where the dependence between truncation time and lifetime can be measured and estimated (Chaieb et al. 2006; Beaudoin and Lakhel-Chaieb 2008; Emura et al. 2011; Emura and Wang 2012; Emura and Murotani 2015). In the absence of covariates, these papers proposed semiparametric inference methods assuming a parametric copula with unspecified marginal distributions for the truncation time and lifetime. The copula-based models could be extended to include covariates along the line of Braekers and Veraverbeke (2005), Chen (2010), and Emura and Chen (2014) as studied under the competing risks setting. However, one difficulty comes from the joint estimation of the two marginal distribution functions, which are infinite dimensional. Compared with the copula approach, the proposed methodology is much easier to implement since the proposed estimating functions only involve regression parameters and avoid handling infinite dimensional parameters. This implies that there is enough information in the estimating functions to perform regression analysis.

Appendix A: Properties of the estimating functions

Appendix A1: Proof of Equation (2)

Without loss of generality, we assume that $e_i^Y(\boldsymbol{\beta}) \neq e_j^Y(\boldsymbol{\beta})$, $i \neq j$, holds true. By straightforward calculations,

$$\begin{aligned} U_n^G(\boldsymbol{\beta}) &= \sum_{i=1}^n \sum_{j=1}^n \Delta_i(\mathbf{X}_i - \mathbf{X}_j) I\{e_j^T(\boldsymbol{\beta}) \leq e_i^Y(\boldsymbol{\beta}) \leq e_j^Y(\boldsymbol{\beta})\} \\ &= \sum_{i=1}^n \sum_{j: e_i^Y(\boldsymbol{\beta}) < e_j^Y(\boldsymbol{\beta})} \Delta_i(\mathbf{X}_i - \mathbf{X}_j) I\{e_j^T(\boldsymbol{\beta}) \leq e_i^Y(\boldsymbol{\beta})\}. \end{aligned}$$

If a pair (i, j) satisfies $e_i^Y(\boldsymbol{\beta}) < e_j^Y(\boldsymbol{\beta})$, it holds $e_i^Y(\boldsymbol{\beta}) = \tilde{e}_{ij}^Y(\boldsymbol{\beta}) = e_i^Y(\boldsymbol{\beta}) \wedge e_j^Y(\boldsymbol{\beta})$. Thus,

$$U_n^G(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{j: e_i^Y(\boldsymbol{\beta}) < e_j^Y(\boldsymbol{\beta})} \Delta_i(\mathbf{X}_i - \mathbf{X}_j) I\{e_j^T(\boldsymbol{\beta}) \leq \tilde{e}_{ij}^Y(\boldsymbol{\beta})\}.$$

Since $\{e_j^T(\boldsymbol{\beta}) \leq \tilde{e}_{ij}^Y(\boldsymbol{\beta})\}$ holds if and only if $\{\tilde{e}_{ij}^T(\boldsymbol{\beta}) \leq \tilde{e}_{ij}^Y(\boldsymbol{\beta})\}$,

$$U_n^G(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{j: e_i^Y(\boldsymbol{\beta}) < e_j^Y(\boldsymbol{\beta})} \Delta_i(\mathbf{X}_i - \mathbf{X}_j) I\{\tilde{e}_{ij}^T(\boldsymbol{\beta}) \leq \tilde{e}_{ij}^Y(\boldsymbol{\beta})\}.$$

As long as a pair (i, j) satisfies $e_i^Y(\boldsymbol{\beta}) < e_j^Y(\boldsymbol{\beta})$, we obtain the identity

$$\begin{aligned} \Delta_i &= \Delta_i \Delta_j + \Delta_i(1 - \Delta_j) I\{e_i(\boldsymbol{\beta}) < e_j(\boldsymbol{\beta})\} + (1 - \Delta_i) \Delta_j I\{e_i(\boldsymbol{\beta}) > e_j(\boldsymbol{\beta})\} \\ &= Q_{ij}(\boldsymbol{\beta}). \end{aligned}$$

This leads to

$$\begin{aligned} U_n^G(\boldsymbol{\beta}) &= \sum_{i=1}^n \sum_{j: e_i^Y(\boldsymbol{\beta}) < e_j^Y(\boldsymbol{\beta})} (\mathbf{X}_i - \mathbf{X}_j) I\{\tilde{e}_{ij}^T(\boldsymbol{\beta}) \leq \tilde{e}_{ij}^Y(\boldsymbol{\beta})\} Q_{ij}(\boldsymbol{\beta}) \\ &= - \sum_{i < j} (\mathbf{X}_i - \mathbf{X}_j) \text{sgn}\{e_i^Y(\boldsymbol{\beta}) - e_j^Y(\boldsymbol{\beta})\} I\{\tilde{e}_{ij}^T(\boldsymbol{\beta}) \leq \tilde{e}_{ij}^Y(\boldsymbol{\beta})\} Q_{ij}(\boldsymbol{\beta}). \end{aligned}$$

□

Appendix A2: The functional behavior of $S_n^{\text{Logrank}}(\boldsymbol{\beta}, \gamma)$

We consider the case that $\mathbf{X} = X$ is one dimensional, and hence $\boldsymbol{\beta} = \beta$. For a given γ , $S_n^{\text{Logrank}}(\boldsymbol{\beta}, \gamma)$ is the sum of

$$L_{ij}(\beta) = (X_i - X_j) \text{sgn}\{\varepsilon_i^Y(\beta, \gamma) - \varepsilon_j^Y(\beta, \gamma)\} I\{\tilde{\varepsilon}_{ij}^T(\beta, \gamma) \leq \tilde{\varepsilon}_{ij}^Y(\beta, \gamma)\} O_{ij}(\beta, \gamma),$$

for $i < j$. Suppose $O_{ij}(\beta) = 1$. Otherwise $L_{ij}(\beta) = 0$ has no contribution to $S_n^{\text{Logrank}}(\boldsymbol{\beta}, \gamma)$. Note that the range of β that satisfies the condition $\tilde{\varepsilon}_{ij}^T(\beta, \gamma) \leq \tilde{\varepsilon}_{ij}^Y(\beta, \gamma)$ is always a closed interval. The function $\text{sgn}\{\varepsilon_i^Y(\beta, \gamma) - \varepsilon_j^Y(\beta, \gamma)\}$ is a step function that crosses with zero at $\beta = \{Y_i - Y_j - \gamma(T_i - T_j)\} / (X_i - X_j)$. Therefore, $L_{ij}(\beta)$ is non-monotonic step function which has at most three different jump points.

Appendix B: Proofs for asymptotic analysis

Appendix B.1: Proof of Equation (4)

By definition,

$$\begin{aligned}
 & 2n^{-2}S_n^{\text{Logrank}}(\boldsymbol{\beta}, \gamma) \\
 &= -n^{-2} \sum_{i,j} (\mathbf{X}_i - \mathbf{X}_j) \text{sgn}\{\varepsilon_i^Y(\boldsymbol{\beta}, \gamma) - \varepsilon_j^Y(\boldsymbol{\beta}, \gamma)\} I\{\tilde{\varepsilon}_{ij}^T(\boldsymbol{\beta}, \gamma) \leq \tilde{\varepsilon}_{ij}^Y(\boldsymbol{\beta}, \gamma)\} O_{ij}(\boldsymbol{\beta}, \gamma) \\
 &= - \iint [(\mathbf{x}_1 - \mathbf{x}_2) \text{sgn}\{y_1 - \boldsymbol{\beta}'\mathbf{x}_1 - \gamma t_1 - (y_2 - \boldsymbol{\beta}'\mathbf{x}_2 - \gamma t_2)\} \\
 &\quad \times I\{(t_1 - \boldsymbol{\beta}'\mathbf{x}_1 - \gamma t_1) \vee (t_2 - \boldsymbol{\beta}'\mathbf{x}_2 - \gamma t_2) \leq (y_1 - \boldsymbol{\beta}'\mathbf{x}_1 - \gamma t_1) \wedge (y_2 - \boldsymbol{\beta}'\mathbf{x}_2 - \gamma t_2)\} \\
 &\quad \times \{\delta_1 \delta_2 + \delta_1(1 - \delta_2)I(y_1 - \boldsymbol{\beta}'\mathbf{x}_1 - \gamma t_1 < y_2 - \boldsymbol{\beta}'\mathbf{x}_2 - \gamma t_2)\} \\
 &\quad + (1 - \delta_1)\delta_2 I(y_1 - \boldsymbol{\beta}'\mathbf{x}_1 - \gamma t_1 > y_2 - \boldsymbol{\beta}'\mathbf{x}_2 - \gamma t_2)] \\
 &\quad \times dF_n(t_1, y_1, \delta_1, \mathbf{x}_1) dF_n(t_2, y_2, \delta_2, \mathbf{x}_2).
 \end{aligned}$$

Also, using similar calculations for $2n^{-2}S_n^{\text{Kendall}}(\boldsymbol{\beta}, \gamma)$, we obtain Equation (4).

Appendix B.2: Proof for Theorem 1

The consistency proof follows from the general theory of Z-estimators. For Theorem 5.9 of [Van Der Vaart \(1998\)](#) to be applied, one needs to verify the following conditions:

- (i) $\sup_{(\boldsymbol{\beta}, \gamma) \in \Theta} \|\Phi(F_n; \boldsymbol{\beta}, \gamma) - \Phi(F; \boldsymbol{\beta}, \gamma)\|$ converges in probability to zero,
- (ii) For every $\varepsilon > 0$, $\inf_{\|(\boldsymbol{\beta}, \gamma) - (\boldsymbol{\beta}_0, \gamma_0)\|_2 \geq \varepsilon} \|\Phi(F; \boldsymbol{\beta}, \gamma)\|_2 > 0 = \|\Phi(F; \boldsymbol{\beta}_0, \gamma_0)\|_2$.

To prove (i), let $D\{(-\infty, \infty)^{p+3}\}$ be a collection of all right-continuous, bounded functions on $(-\infty, \infty)^{p+3}$ whose left limit exists such that $F, F_n \in D\{(-\infty, \infty)^{p+3}\}$. Define the norm $\sup_{x \in (-\infty, \infty)^{p+3}} |f(x)|$ for $f \in D\{(-\infty, \infty)^{p+3}\}$. Also, let $I^\infty(\Theta)$ be a collection of all bounded functions on Θ whose norm is defined by $\sup_{\theta \in \Theta} \|H(\theta)\|$ for $H \in I^\infty(\Theta)$. It is well-known from the Glivenko–Cantelli theorem that F_n converges in probability to F in the normed space $D\{(-\infty, \infty)^{p+3}\}$. Also, it can be shown that a map $D\{(-\infty, \infty)^{p+3}\} \mapsto I^\infty(\Theta)$, defined as

$$\begin{aligned}
 & F \mapsto \Phi(F; \boldsymbol{\beta}, \gamma) \\
 &= \iint \mathbf{h}\{(t_1, y_1, \delta_1, \mathbf{x}_1), (t_2, y_2, \delta_2, \mathbf{x}_2); \boldsymbol{\beta}, \gamma\} dF(t_1, y_1, \delta_1, \mathbf{x}_1) dF(t_2, y_2, \delta_2, \mathbf{x}_2),
 \end{aligned}$$

is continuous. By the continuous mapping theorem,

$$\begin{aligned}
 & \sup_{\Theta \in (\boldsymbol{\beta}, \gamma)} \left| \iint \mathbf{h}\{(t_1, y_1, \delta_1, \mathbf{x}_1), (t_2, y_2, \delta_2, \mathbf{x}_2); \boldsymbol{\beta}, \gamma\} dF_n(t_1, y_1, \delta_1, \mathbf{x}_1) dF_n(t_2, y_2, \delta_2, \mathbf{x}_2) \right. \\
 & \quad \left. - \iint \mathbf{h}\{(t_1, y_1, \delta_1, \mathbf{x}_1), (t_2, y_2, \delta_2, \mathbf{x}_2); \boldsymbol{\beta}, \gamma\} dF(t_1, y_1, \delta_1, \mathbf{x}_1) dF(t_2, y_2, \delta_2, \mathbf{x}_2) \right|
 \end{aligned}$$

converges in probability to zero, which implies (i). The validity of (ii) is due to Assumption 2. □

Appendix B.3: Proof for asymptotic normality

Note that the random vector $\phi_F\{(T_j, Y_j, \Delta_j, \mathbf{X}_j); \boldsymbol{\beta}, \gamma\}$ in Eq. (6) is bounded since \mathbf{h} is bounded under Assumption 1. Hence, under Assumption 1, the dominated convergence theorem shows

$$E[\phi_F\{(T, Y, \Delta, \mathbf{X}); \boldsymbol{\beta}, \gamma\} - \phi_F\{(T, Y, \Delta, \mathbf{X}); \boldsymbol{\beta}_0, \gamma_0\}]^2 \rightarrow 0,$$

as $(\boldsymbol{\beta}, \gamma) \rightarrow (\boldsymbol{\beta}_0, \gamma_0)$. By Lemma 3.3.5 of Van Der Vaart and Wellner 1996 and under Assumption 4, we have

$$\begin{aligned} & n^{1/2}\{\Phi(F_n; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \Phi(F; \hat{\boldsymbol{\beta}}, \hat{\gamma})\} - n^{1/2}\{\Phi(F_n; \boldsymbol{\beta}_0, \gamma_0) - \Phi(F; \boldsymbol{\beta}_0, \gamma_0)\} \\ & = o_P(1 + n^{1/2}\|\boldsymbol{\beta} - \boldsymbol{\beta}_0, \gamma - \gamma_0\|). \end{aligned}$$

Since $\Phi(F_n; \hat{\boldsymbol{\beta}}, \hat{\gamma}) = \Phi(F; \boldsymbol{\beta}_0, \gamma_0) = \mathbf{0}$, under Assumptions 3 and 4,

$$\begin{aligned} \mathbf{0} &= n^{1/2}\Phi(F_n; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - n^{1/2}\Phi(F; \boldsymbol{\beta}_0, \gamma_0) \\ &= n^{1/2}\{\Phi(F_n; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \Phi(F; \hat{\boldsymbol{\beta}}, \hat{\gamma})\} \\ &\quad + n^{1/2}\{\Phi(F; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \Phi(F; \boldsymbol{\beta}_0, \gamma_0)\} \\ &= n^{1/2}\{\Phi(F_n; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \Phi(F; \hat{\boldsymbol{\beta}}, \hat{\gamma})\} - n^{1/2}\{\Phi(F_n; \boldsymbol{\beta}_0, \gamma_0) - \Phi(F; \boldsymbol{\beta}_0, \gamma_0)\} \\ &\quad + n^{1/2}\{\Phi(F_n; \boldsymbol{\beta}_0, \gamma_0) - \Phi(F; \boldsymbol{\beta}_0, \gamma_0)\} \\ &\quad + n^{1/2}\{\Phi(F; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \Phi(F; \boldsymbol{\beta}_0, \gamma_0)\} \\ &= o_P(1 + n^{1/2}\|\boldsymbol{\beta} - \boldsymbol{\beta}_0, \gamma - \gamma_0\|) \\ &\quad + n^{-1/2} \sum_{j=1}^n \phi_F\{(T_j, Y_j, \Delta_j, \mathbf{X}_j); \boldsymbol{\beta}_0, \gamma_0\} \\ &\quad + \mathbf{A}_0 n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \hat{\gamma} - \gamma_0)', \end{aligned}$$

where the last equation uses Eq. (5) and the differentiability of $\Phi(F; \boldsymbol{\beta}, \gamma)$ at $(\boldsymbol{\beta}_0, \gamma_0)$. The preceding formula implies that $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$ is \sqrt{n} -consistent for $(\boldsymbol{\beta}_0, \gamma_0)$ and, therefore, Eq. (7) holds [see p.311 of Van Der Vaart and Wellner (1996)]. □

Appendix B.4: Consistency of the product-limit estimator

We show that $\sup_{t \in [a, b]} |\hat{S}_\varepsilon(t; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - S_\varepsilon(t)|$ converges in probability to zero. It is useful to write $S_\varepsilon(t) = \prod_{s \leq t} \{1 - d\Lambda_\varepsilon(s)\}$, where $\Lambda_\varepsilon(t) = -\log S_\varepsilon(t)$ is the cumulative hazard function. Similarly, $\hat{S}_\varepsilon(t; \boldsymbol{\beta}, \gamma) = \prod_{s \leq t} \{1 - d\hat{\Lambda}(s; \boldsymbol{\beta}, \gamma)\}$, where

$$d\hat{\Lambda}(s; \boldsymbol{\beta}, \gamma) = R(s; \boldsymbol{\beta}, \gamma)^{-1} \sum_j I(\varepsilon_j^Y(\boldsymbol{\beta}, \gamma) = s, \Delta_j = 1)$$

and $R(s; \boldsymbol{\beta}, \gamma) = \sum_j I\{\varepsilon_j^T(\boldsymbol{\beta}, \gamma) \leq s \leq \varepsilon_j^Y(\boldsymbol{\beta}, \gamma)\}$. Since the product integration is a continuous map, it suffices to show that $\sup_{t \in [a, b]} |\hat{\Lambda}_\varepsilon(t; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \Lambda_\varepsilon(t)|$ converges in probability to zero. Let $\Lambda_\varepsilon^*(t) = \int_{-\infty}^t J(s) d\Lambda_\varepsilon(s)$, where $J(s) = I\{R(s; \boldsymbol{\beta}_0, \gamma_0) > 0\}$, and

$$\begin{aligned} &\Lambda_\varepsilon(t, \boldsymbol{\beta}, \gamma) \\ &= \int_{-\infty}^t \frac{E[I(T - \boldsymbol{\beta}'\mathbf{X} - \gamma T) f_\varepsilon\{s + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{X} + (\gamma - \gamma_0)T\} S_C(s + \boldsymbol{\beta}'\mathbf{X} + \gamma T)]}{E[I(T - \boldsymbol{\beta}'\mathbf{X} - \gamma T) S_\varepsilon\{s + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{X} + (\gamma - \gamma_0)T\} S_C(s + \boldsymbol{\beta}'\mathbf{X} + \gamma T)]} ds. \end{aligned}$$

It is easy to see that

$$\Lambda_\varepsilon(t, \boldsymbol{\beta}_0, \gamma_0) = \int_{-\infty}^t f_\varepsilon(s) S_\varepsilon(s)^{-1} ds = \Lambda_\varepsilon(t).$$

Also, let $\Lambda_\varepsilon^*(t, \boldsymbol{\beta}, \gamma) = \int_{-\infty}^t J(s) d\Lambda_\varepsilon(s, \boldsymbol{\beta}, \gamma)$. Now, for δ from Assumption 5, we have

$$\begin{aligned} &|\hat{\Lambda}_\varepsilon(t; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \Lambda_\varepsilon(t)| \\ &\leq \sup_{\|(\boldsymbol{\beta}, \gamma) - (\boldsymbol{\beta}_0, \gamma_0)\| \leq \delta} |\hat{\Lambda}_\varepsilon(t; \boldsymbol{\beta}, \gamma) - \Lambda_\varepsilon^*(t; \boldsymbol{\beta}, \gamma)| + |\Lambda_\varepsilon^*(t; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \Lambda_\varepsilon^*(t)| \\ &\quad + |\Lambda_\varepsilon^*(t) - \Lambda_\varepsilon(t)|. \end{aligned}$$

Under Assumption 5, and by the Glivenko–Cantelli theorem, it follows that

$$\sup_{t \in [a, b]} \sup_{\|(\boldsymbol{\beta}, \gamma) - (\boldsymbol{\beta}_0, \gamma_0)\| \leq \delta} |\hat{\Lambda}_\varepsilon(t; \boldsymbol{\beta}, \gamma) - \Lambda_\varepsilon^*(t; \boldsymbol{\beta}, \gamma)|$$

converges in probability to zero. Since $\Lambda_\varepsilon(t, \boldsymbol{\beta}, \gamma)$ is continuous at $(\boldsymbol{\beta}_0, \gamma_0)$, the continuous mapping is applied to Theorem 1 to show that $\sup_{t \in [a, b]} |\Lambda_\varepsilon^*(t; \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \Lambda_\varepsilon^*(t)|$ converges in probability to zero. Note that $R(s; \boldsymbol{\beta}_0, \gamma_0)$ converges in probability to $\pi_\varepsilon(t, \boldsymbol{\beta}_0, \gamma_0) > 0$, uniformly over $t \in [a, b]$. Hence, $I\{R(s; \boldsymbol{\beta}_0, \gamma_0) = 0\}$ converges in probability to zero as well. Finally,

$$|\Lambda_\varepsilon^*(t) - \Lambda_\varepsilon(t)| = \left| \int_{-\infty}^t I\{R(s; \boldsymbol{\beta}_0, \gamma_0) = 0\} d\Lambda_\varepsilon(s) \right|,$$

converges in probability to zero, uniformly over $t \in [a, b]$. □

Appendix C: Bandwidth calculation

The jumps of $\mathbf{S}_n^{\text{Logrank}}(\boldsymbol{\beta}, \gamma)$ with respect to β_k given the other components satisfy

$$Y_i - \hat{\boldsymbol{\beta}}'_{(-k)} \mathbf{X}_{i(-k)} - \beta_k X_{i,k} - \hat{\gamma} T_i = Y_j - \hat{\boldsymbol{\beta}}'_{(-k)} \mathbf{X}_{j(-k)} - \beta_k X_{j,k} - \hat{\gamma} T_j,$$

where $X_{i,k}$ is k th component of \mathbf{X}_i , and $\mathbf{X}_{j(-k)}$ is the $k-1$ remaining vector. Accordingly, the jumps are identified at

$$\frac{Y_i - Y_j - \hat{\boldsymbol{\beta}}'_{(-k)}(\mathbf{X}_{i(-k)} - \mathbf{X}_{j(-k)}) - \hat{\gamma}(T_i - T_j)}{X_{i,k} - X_{j,k}} \approx \frac{\varepsilon_i^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma}) - \varepsilon_j^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma})}{X_{i,k} - X_{j,k}} \equiv \beta_{(i,j),k}^\#.$$

Hence, we define S_k^2 as the sample variance of all $\beta_{(i,j),k}^\#$, with $i < j$, satisfying $I\{\tilde{\varepsilon}_{ij}^T(\hat{\boldsymbol{\beta}}, \hat{\gamma}) \leq \tilde{\varepsilon}_{ij}^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma})\} O_{ij}(\hat{\boldsymbol{\beta}}, \hat{\gamma}) = 1$ and $X_{i,k} \neq X_{j,k}$. The IQR_k is defined similarly.

The bandwidth $b_{p+1} = 0.5 \min(S_{p+1}, \text{IQR}_{p+1}/1.34)n^{-1/5}$ is also obtained, where S_{p+1}^2 and IQR_{p+1} are calculated with $\gamma_{(i,j)}^\# \equiv \{\varepsilon_i^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma}) - \varepsilon_j^Y(\hat{\boldsymbol{\beta}}, \hat{\gamma})\}/(T_i - T_j)$.

Acknowledgments The authors thank the editor, associate editor and two reviewers for their helpful comments that substantially improve the manuscript. This work was supported by National Science Council of Taiwan under the grants: NSC 98-2118-M-009-003-MY2 (the second author) and NSC101-2118-M008-002-MY2 (the first author).

References

- Beaudoin, D., Lakhel-Chaieb, L. (2008). Archimedean copula model selection under dependent truncation. *Statistics in Medicine*, 27, 4440–4454.
- Braekers, R., Veraverbeke, N. (2005). A copula-graphic estimator for the conditional survival function under dependent censoring. *The Canadian Journal of Statistics*, 33, 429–447.
- Chaieb, L., Rivest, L.-P., Abdous, B. (2006). Estimating survival under a dependent truncation. *Biometrika*, 93, 655–669.
- Chen, Y.-H. (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *Journal of the Royal Statistical Society, Series B*, 72, 235–51.
- Chen, C.-H., Tsai, W.-Y., Chao, W.-H. (1996). The product-moment correlation coefficient and linear regression for truncated data. *Journal of the American Statistical Association*, 91, 1181–1186.
- de Uña-Álvarez, J. (2010). Nonparametric estimation of a conditional distribution from length-biased data. *Annals of the Institute of Statistical Mathematics*, 62, 323–341.
- de Uña-Álvarez, J. (2012). On the Markov three-state progressive model, *Recent Advances in System Reliability*. New York: Springer.
- Emura, T., Chen, Y.-H. (2014). Gene selection for survival data under dependent censoring: a copula-based approach. *Statistical Methods in Medical Research*. doi:10.1177/0962280214533378.
- Emura, T., Murotani, K. (2015). An algorithm for estimating survival under a copula-based dependent truncation model. *TEST*. doi:10.1007/s11749-015-0432-8.
- Emura, T., Wang, W. (2010). Testing quasi-independence for truncation data. *Journal of Multivariate Analysis*, 101, 223–239.
- Emura, T., Wang, W. (2012). Nonparametric maximum likelihood estimation for dependent truncation data based on copulas. *Journal of Multivariate Analysis*, 110, 171–188.
- Emura, T., Wang, W., Hung, H. N. (2011). Semi-parametric inference for copula models for truncated data. *Statistica Sinica*, 21, 349–367.
- Hyde, J. (1980). Survival analysis with incomplete observations. In: R. G. Miller, B. Efron, B. W. Brown, L. E. Moses (eds.), *Biostatistics Casebook* (pp. 31–46). New York: Wiley.
- Jin, Z., Lin, D. Y., Wei, L.-J., Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, 90, 341–353.
- Jones, M., Crowley, J. (1992). Nonparametric tests for the Markov model for survival data. *Biometrika*, 79, 513–522.

- Kalbfleisch, J. D., Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data* (2nd ed.). New York: Wiley.
- Klein, J. P., Moeschberger, M. L. (2003). *Survival analysis techniques for censored and truncated data*. New York: Springer.
- Lai, T.-L., Ying, Z. (1991). Rank regression methods for left-truncated and right-censored data. *Annals of Statistics*, *19*, 531–556.
- Martin, E. C., Betensky, R. A. (2005). Testing quasi-independence of failure and truncation via conditional Kendall's tau. *Journal of the American Statistical Association*, *100*, 484–492.
- Nelder, J. A., Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal*, *7*, 308–313.
- Rodríguez, Gironde M., de Uña-Álvarez, J. (2012). Testing Markovian in the three-state progressive model via future-past association. *Biometrical Journal*, *54*(2), 163–180.
- Sheather, S. (2004). Density estimation. *Statistical Science*, *19*, 588–597.
- Tsai, W.-Y. (1990). Testing the assumption of independence of truncation time and failure time. *Biometrika*, *77*, 169–177.
- Tsai, W.-Y., Jewell, N. P., Wang, M. C. (1987). A Note on the product-limit estimator under right censoring and left truncation. *Biometrika*, *74*, 883–886.
- Van Der Vaart, A. W. (1998). *Asymptotic Statistics, Cambridge Series in Statistics and Probabilistic Mathematics*. Cambridge: Cambridge University Press.
- Van Der Vaart, A. W., Wellner, J. A. (1996). *Weak Convergence and Empirical Process*. New York: Springer.