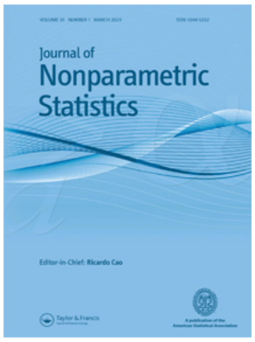

Journal of Nonparametric Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gnst20

Inference on semi-parametric transformation model with a pairwise likelihood based on left-truncated and interval-censored data

Yichen Lou, Peijie Wang & Jianguo Sun

JOURNAL OF NONPARAMETRIC STATISTICS
2023, VOL. 35, NO. 1, 38–55
https://doi.org/10.1080/10485252.2022.2138383

Inference on semi-parametric transformation model with a pairwise likelihood based on left-truncated and interval-censored data

Yichen Lou[a], Peijie Wang[a] and Jianguo Sun[b]

[a]School of Mathematics, Jilin University, Changchun, People's Republic of China; [b]Department of Statistics, University of Missouri, Columbia, MO, USA

ABSTRACT

Semi-parametric transformation models provide a general and flexible class of models for regression analysis of failure time data and many methods have been developed for their estimation. In particular, they include the proportional hazards and proportional odds models as special cases. In this paper, we discuss the situation where one observes left-truncated and interval-censored data, for which it does not seem to exist an established method. For the problem, in contrast to the commonly used conditional approach that may not be efficient, a pairwise pseudo-likelihood method is proposed to recover some missing information in the conditional method. The proposed estimators are proved to be consistent and asymptotically efficient and normal. A simulation study is conducted to assess the empirical performance of the method and suggests that it works well in practical situations. This method is illustrated by using a set of real data arising from an HIV/AIDS cohort study.

ARTICLE HISTORY
Received 11 February 2022
Accepted 14 October 2022

KEYWORDS
Efficient estimation; interval censoring; left truncation; pairwise pseudo-likelihood; semi-parametric transformation model

MATHEMATICS SUBJECT CLASSIFICATION
62N02

1. Introduction

Semi-parametric transformation models have become more and more popular in regression analysis of failure time data as they provide a general and flexible class of models that can cover various types of covariate effects (Fine, Ying, and Wei 1998; Chen, Jin, and Ying 2002; Zhang, Sun, Zhao, and Sun 2005). In particular, they include the proportional hazards model, the most commonly used one in the failure time data analysis, and the proportional odds model as special cases. In this paper, we discuss their estimation when one faces left-truncated and interval-censored failure time data.

Left truncation often occurs in many fields and one common situation where it happens is when the prevalent cohort sampling is used that recruits only the subjects who have experienced some initial event such as an infection or satisfied some certain conditions such as the people who have been infected by some virus (Pan and Chappell 1999; Wang, Li, and Sun 2021). A specific example is AIDS cohort studies that follow HIV-infected subjects and

CONTACT Peijie Wang wangpeijie@jlu.edu.cn School of Mathematics, Jilin University, Changchun 130012, People's Republic of China

Supplemental data for this article can be accessed here. https://doi.org/10.1080/10485252.2022.2138383

aim to investigate AIDS incubation or death risk (De Gruttola and Lagakos 1989). Interval censoring is an another feature of failure time data that one often has to deal with and means that the failure time of interest is known or observed only to belong to an interval instead of being observed exactly (Sun 2006). It is easy to see that interval-censored data include right-censored data as a special case (Kalbfleisch and Prentice 2011). There usually exist several types of interval-censored data, and one is the so-called case I or current-status data, for which each study subject is observed only once. Correspondingly, one may observe case II or K interval-censored data, meaning that there exist two observations or a sequence of observations, respectively, on each subject (Huang and Wellner 1997). In the following, we will focus on case K interval-censored failure time data.

A great deal of literature has been established for the analysis of interval-censored data and in particular, many authors have considered the application of parametric methods or models such as Weibull models and log-linear models. See, for instance, Marshall (1974), Hazelrig, Turner, and Blackstone (1982), Odell, Anderson, and D'Agostino (1992), and Li and Ma (2010). Many semi-parametric methods have also been developed for the analysis of interval-censored failure time data (Sun 2006; Kalbfleisch and Prentice 2011). For example, Huang (1996) and Huang and Wellner (1993) considered the nonparametric maximum likelihood estimation of the proportional hazards and odds models, and Rossini and Tsiatis (1996), Huang and Rossini (1997) and Shen (1998) developed some sieve maximum likelihood estimation procedures for the proportional odds model. Furthermore, to fit the semi-parametric transformation model to interval-censored data, Zhang et al. (2005) and Zhang and Zhao (2013) proposed an estimating equation method and an empirical likelihood inference approach, respectively, and Zeng, Mao, and Lin (2016) developed an EM-type algorithm. Wang, Zhao, Du, and Sun (2018) and Xu, Zhao, Hu, and Sun (2019) also discussed the same problem but under a more complicated type of interval censoring.

It is well known that left truncation can make the analysis much more complicated and the analysis that ignores it may yield biased estimation (Pan and Chappell 1999). Several methods have been proposed for regression analysis of left-truncated and interval-censored data. For example, Kim (2003) considered the fitting of the proportional hazards model, a particular case of the semiparametric transformation model, to left-truncated and current status data and Pan and Chappell (2002) and Shen (2014) discussed the same problem but for general interval-censored data. Also Wang, Tong, Zhao, and Sun (2015) investigated regression analysis of left-truncated data and interval-censored data arising from the additive hazards model. Under the semiparametric transformation model, Shen, Chen, Pan, and Chen (2019) proposed a computationally efficient EM algorithm for the left-truncated and interval-censored data without or with a cure fraction. All methods mentioned above as well as most of the existing methods for left-truncated data are conditional approaches given truncation times. As consequence, they may not be efficient. To address this or improve the efficiency, some composite or pairwise likelihood-based methods have been proposed under the proportional hazards or odds model (Huang and Qin 2012; Huang and Qin 2013; Wu, Kim, Qin, Saran, and Li 2018; Gao and Chan 2019). In particular, Wang et al. (2021) gave a pairwise likelihood approach for fitting the proportional hazards model to left-truncated and case II interval-censored data. In this paper, we will consider left-truncated and case K interval-censored data and generalise their method to semi-parametric transformation models.

The remainder of this paper is organised as follows. First we will introduce in Section 2 some notations and assumptions that will be used throughout the paper as well as the structure of observed data. The proposed pairwise pseudo-likelihood approach will then be presented in Section 3, and in the method, we will augment the conditional likelihood with a pairwise likelihood constructed based on truncation times. In Section 4, the asymptotic properties of the proposed estimators will be established. Section 5 gives some simulation results obtained from a simulation study conducted to assess the empirical performance of the proposed method, and they suggest that it works well and is efficient for practical situations. In Section 6, the proposed approach is illustrated by using a set of real data from an AIDS cohort study, and Section 7 contains some concluding remarks.

2. Notations and assumptions

Consider a failure time study and let T^* and Z^* denote the underlying failure time of interest and a $p \times 1$ vector of time-independent covariates. Suppose that there exists another random variable or an underlying truncation time denoted by A^* such that T^* is observed only if $T^* \geq A^*$. An example of this is that T^* represents the time to an event of interest such as the diagnosis of a disease, while A^* is the time to enter the study. In the following, we will assume that A^* is independent of T^*, and the underlying truncation time A^* is independent of the covariate Z^*.

To describe the covariate effects, we will assume that given Z^*, T^* follows the semi-parameter transformation model given by

$$\Lambda(t; Z^*) = G\left\{\exp(\beta^T Z^*)\Lambda_0(t)\right\} \tag{1}$$

in terms of the cumulative hazard function. In the above, β denotes a $p \times 1$ vector of regression coefficients, $G(\cdot)$ is a known, strictly increasing transformation function, and $\Lambda_0(\cdot)$ denotes an unknown increasing function. One major advantage of the model above is its flexibility as it includes many commonly used models as special cases. For example, it yields the proportional hazards model or the proportional odds model with the choice of $G(x) = x$ or $G(x) = \log(1 + x)$, respectively. Among others, one commonly used class of transformation functions is the logarithmic transformations given by $G(x) = r^{-1}\log(1 + rx)$, $r \geq 0$. Under model (1) and given Z^*, the survival functions of T^* can be written as $S(t \mid Z^*) = \exp\{-\Lambda(t; Z^*)\} = \exp[-G\{\exp(\beta^T Z^*)\Lambda_0(t)\}]$.

As mentioned above, it will be assumed that one only observes left-truncated data or the pair (A^*, T^*) with $A^* \leq T^*$, and we will use $(A, T, Z) \equiv (A^*, T^*, Z^*) \mid (A^* \leq T^*)$ to denote the realisation. Furthermore, we will assume that the study consists of n independent subjects and for subject i, there exists a sequence of observation times $U_{i0} = A_i < U_{i1} < \cdots < U_{iK_i}$, where K_i denotes the number of the observation times on subject i, $i = 1, \ldots, n$. In the following, it will be supposed that one can write U_{ij} as $U_{ij} = A_i + V_{ij}$ and the K_i and V_{ij}'s are independent of (A_i, T_i) given Z_i. Then the observed data have the form $\{\mathcal{O}_i = (A_i, U_{ij}, \delta_{ij} = I(U_{ij-1} < T_i \leq U_{ij}), Z_i, \tau_i, j = 1, \ldots, K_i);\ i = 1, \ldots, n\}$, where A_i, T_i and Z_i denote A, T and Z, respectively, associated with subject i and τ_i denotes a follow-up time for the ith subject that is assumed to be independent of T_i. Also it will be assumed that the main goal is to make inferences about β.

Let H denote the cumulative distribution function of A^*. Then under the assumptions above, one can derive that the full likelihood function is proportional to

$$
\begin{aligned}
\mathcal{L}_n^F &= \prod_{i=1}^{n} \frac{\{S(A_i \mid Z_i) - S(U_{i1} \mid Z_i)\}^{\delta_{i1}} \prod_{j=2}^{K_i} \{S(U_{ij-1} \mid Z_i) - S(U_{ij} \mid Z_i)\}^{\delta_{ij}}}{\int_0^\infty S(u \mid Z_i)\,\mathrm{d}H(u)} \\
&\quad \times S(U_{iK_i} \mid Z_i)^{1-\sum_{j=1}^{K_i}\delta_{ij}}\,\mathrm{d}H(A_i) \\
&= \prod_{i=1}^{n} \frac{\{S(A_i \mid Z_i) - S(U_{i1} \mid Z_i)\}^{\delta_{i1}} \prod_{j=2}^{K_i} \{S(U_{ij-1} \mid Z_i) - S(U_{ij} \mid Z_i)\}^{\delta_{ij}}}{S(A_i \mid Z_i)} \\
&\quad \times S(U_{iK_i} \mid Z_i)^{1-\sum_{j=1}^{K_i}\delta_{ij}} \times \prod_{i=1}^{n} \frac{S(A_i \mid Z_i)\,\mathrm{d}H(A_i)}{\int_0^\infty S(u \mid Z_i)\,\mathrm{d}H(u)} \\
&= \mathcal{L}_n^C \times \mathcal{L}_n^M.
\end{aligned}
$$

In the above, $\mathcal{L}_n^C$ denotes the conditional likelihood of $(U_{ij}, \delta_{ij}, i = 1, \ldots, n, j = 1, \ldots, K_i)$ given the (A_i, Z_i)'s and $\mathcal{L}_n^M$ the marginal likelihood of the A_i's given the Z_i's. For estimation of β, it is apparent that a natural way would be to maximise the full likelihood function $\mathcal{L}_n^F$. However, this would not be easy since $\mathcal{L}_n^F$ also involves two unknown functions $H(\cdot)$ and $\Lambda(\cdot)$. Another method would be to base the estimation on the conditional likelihood function $\mathcal{L}_n^C$ but as discussed above, this may not be efficient since it ignores the information about the parameters contained in $\mathcal{L}_n^M$. In the following, we propose an efficient composite approach.

3. An efficient pairwise likelihood approach

In this section, we will present the proposed pairwise likelihood estimation procedure and one key feature of the method is that one does not need to impose any parametric assumption on and estimate the underlying truncation distribution H. For $\Lambda(\cdot)$, by following Zhou, Hu, and Sun (2017) and others, we will employ the sieve approach to approximate it.

To derive the pairwise likelihood function, conditional on (Z_i, Z_j) and the order statistic of (A_i, A_j), the likelihood of the pair sample $\{(A_i, Z_i), (A_j, Z_j) : 1 \le i < j \le n\}$ has the form

$$
\frac{\frac{S(A_i|Z_i)\,\mathrm{d}H(A_i)}{\int_0^t S(u|Z_i)\,\mathrm{d}H(u)} \times \frac{S(A_j|Z_j)\,\mathrm{d}H(A_j)}{\int_0^t S(u|Z_j)\,\mathrm{d}H(u)}}{\frac{S(A_i|Z_i)\,\mathrm{d}H(A_i)}{\int_0^t S(u|Z_i)\,\mathrm{d}H(u)} \times \frac{S(A_j|Z_j)\,\mathrm{d}H(A_j)}{\int_0^t S(u|Z_j)\,\mathrm{d}H(u)} + \frac{S(A_i|Z_j)\,\mathrm{d}H(A_i)}{\int_0^t S(u|Z_j)\,\mathrm{d}H(u)} \times \frac{S(A_j|Z_i)\,\mathrm{d}H(A_j)}{\int_0^t S(u|Z_i)\,\mathrm{d}H(u)}} = \frac{1}{1 + R_{ij}(\beta, \Lambda_0)}.
$$

Also the generalised odds ratio $R_{ij}(\beta, \Lambda_0)$ can be rewritten as

$$
\begin{aligned}
R_{ij}(\beta, \Lambda_0) = \frac{S(A_i \mid Z_j) S(A_j \mid Z_i)}{S(A_i \mid Z_i) S(A_j \mid Z_j)} &= \exp\Big\{-G[\exp(\beta^T Z_j)\Lambda_0(A_i)] \\
&\quad - G[\exp(\beta^T Z_i)\Lambda_0(A_j)] + G[\exp(\beta^T Z_i)\Lambda_0(A_i)] + G[\exp(\beta^T Z_j)\Lambda_0(A_j)]\Big\}
\end{aligned}
$$

under the semi-parametric transformation model (1). Define

$$\mathcal{L}_n^P = \prod_{i<j} \frac{1}{1 + R_{ij}(\beta, \Lambda_0)},$$

and let $\theta = (\beta, \Lambda_0)$. To account for the different magnitudes of $\mathcal{L}_n^C$ and $\mathcal{L}_n^P$, for estimation of θ, we propose to consider the log pairwise pseudo-likelihood function

$$\begin{aligned}\ell_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\Bigg[&\delta_{i1}\log\{S(A_i \mid Z_i) - S(U_{i1} \mid Z_i)\} + \sum_{j=2}^{K_i}\delta_{ij}\log\{S(U_{ij-1} \mid Z_i) - S(U_{ij} \mid Z_i)\} \\ &+ \left(1 - \sum_{j=1}^{K_i}\delta_{ij}\right)\log\{S(U_{iK_i} \mid Z_i)\} - \log\{S(A_i \mid Z_i)\}\Bigg] \\ &- \frac{2}{n(n-1)}\sum_{i<j}\log\{1 + R_{ij}(\theta)\}.\end{aligned}$$

It is easy to see that the log pairwise pseudo-likelihood function $\ell_n(\theta)$ is a function of θ only or does not involve the unknown function $H(\cdot)$. Also $\ell_n(\theta)$ is formed from $\log \mathcal{L}_n^C$ and $\log \mathcal{L}_n^P$, which means that it retains the majority of the information about β and $\Lambda(\cdot)$ and thus yields more efficient estimation than $\ell_n^C(\theta)$ (Liang and Qin 2000). On the other hand, it still involves the unknown function $\Lambda_0(\cdot)$ and for this, by following Zhou et al. (2017) and others, we suggest to approximate $\Lambda_0(\cdot)$ by Bernstein polynomials before maximising $\ell_n(\theta)$. Specifically, define the parameter space $\Theta = \{\theta = (\beta, \Lambda_0) \in \mathcal{B} \otimes \mathcal{M}\}$, where $\mathcal{B} = \{\beta \in R^p : ||\beta|| \le M\}$ with M being a positive constant and $\mathcal{M}$ is the collection of all bounded and continuous nondecreasing, nonnegative functions over interval $[t_l, t_u]$ with $0 \le t_l < t_u < \infty$. Also define

$$B_k^*(t, m, t_l, t_u) = \binom{m}{k}\left(\frac{t - t_l}{t_u - t_l}\right)^k \left(1 - \frac{t - t_l}{t_u - t_l}\right)^{m-k},$$

Bernstein basis polynomials of degree $m = o(n^\nu)$, and the sieve space $\Theta_n = \{\theta_n = (\beta, \Lambda_n) \in \mathcal{B} \otimes \mathcal{M}_n\}$, where

$$\mathcal{M}_n = \left\{\Lambda_n(t) = \sum_{k=0}^{m}\phi_k^* B_k^*(t, m, t_l, t_u) : 0 \le \phi_0^* \le \cdots \le \phi_m^*, \sum_{0 \le k \le m}|\phi_k^*| \le M_n\right\}$$

for some $\nu \in (0, 1)$ and $M_n = o(n^a)$, which controls the size of the sieve space with $a \in (0, 1)$. In practice, $[t_l, t_u]$ is usually taken as the range of the A_i's, U_i's, and V_i's. To remove the constraint $0 \le \phi_0^* \le \cdots \le \phi_m^*$, define the re-parameterisation $\phi_0^* = \exp(\phi_0)$, $\phi_k^* = \sum_{i=0}^{k}\exp(\phi_i)$, $1 \le k \le m$ and let $B_i(t) = \sum_{k=i}^{m} B_k^*(t, m, t_l, t_u)$. Then we have

$$\Lambda_n(t) = \sum_{k=0}^{m}\sum_{i=0}^{k}\exp(\phi_i)B_k^*(t, m, t_l, t_u) = \sum_{k=0}^{m}\exp(\phi_k)B_k(t).$$

Let $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n)$ denote the estimator of θ given by maximising $\ell_n(\theta)$ over the sieve space Θ_n with the re-parameterisation above. One can obtain $\hat{\theta}_n$ by directly solving the equations

$$\frac{\partial \ell_n(\beta, \Lambda_n)}{\partial \beta} = 0, \quad \frac{\partial \ell_n(\beta, \Lambda_n)}{\partial \phi_k} = 0, \quad k = 0, \ldots, m.$$

Also for the implementation of the estimation approach proposed above, it is clear that one needs to determine the degree of Bernstein polynomials m, which controls the smoothness of the approximation, as well as the transformation function G in model (1). For this, we suggest to consider several different values of m and n^ν and various transformation functions such as those discussed in the next section, and then employ the AIC and chose the combination of (m, G) that minimised

$$AIC = -2n\ell_n(\hat{\theta}_n) + 2(p + m + 1).$$

Some more discussion and guidelines on the implementation as well as the selection of (m, G) are given below. For the determination of $\hat{\theta}_n$, in the numerical study below, the function tool fsolve in Matlab will be used.

4. Asymptotic properties

Now we will establish the asymptotic properties of $\hat{\theta}_n$ including the asymptotic consistency and normality as well as the convergence rate. For this, let $\theta_0 = (\beta_0, \Lambda_0)$ denote the true value of θ and define the distance between $\theta_1 = (\beta_1, \Lambda_1)$ and $\theta_2 = (\beta_2, \Lambda_2) \in \Theta$ as

$$d(\theta_1, \theta_2) = ||\theta_1 - \theta_2||_\theta = (||\beta_1 - \beta_2||^2 + ||\Lambda_1 - \Lambda_2)||_2^2)^{1/2},$$

where $\|\cdot\|$ is the Euclidean norm and

$$\|\Lambda_1 - \Lambda_2\|_2^2 = \sum_{j=1}^{K_i} E\{\Lambda_1(U_j) - \Lambda_2(U_j)\}^2 + E\{\Lambda_1(A) - \Lambda_2(A)\}^2.$$

First we will give the asymptotic consistency and convergence rate of $\hat{\theta}_n$ and then the asymptotic normality of $\hat{\beta}_n$ as $n \to \infty$.

Theorem 4.1 (Strong consistency): *Assume that Conditions A.1–A.5 given in the Appendix hold. Then we have that $d(\hat{\theta}_n, \theta_0)$ converges almost surely towards 0.*

Theorem 4.2 (Rate of convergence): *Assume that Conditions A.1–A.5 given in the Appendix hold. Then we have that*

$$d(\hat{\theta}_n, \theta_0) = O_p(n^{-\min\{(1-\nu)/2, \nu d/2\}}),$$

where $\nu \in (0, 1)$ such that $m = o(n^\nu)$ and d is defined in Condition A.1.

Theorem 4.3 (Asymptotic normality): *Assume that Conditions A.1–A.6 given in the Appendix hold. Then we have that*

$$n^{1/2}(\hat{\beta}_n - \beta_0) \rightarrow N(0, \Sigma),$$

where $\Sigma = I_*^{-1}(\beta_0)$, *denoting the semi-parametric efficient bound for* β *with respect to the log pairwise pseudo-likelihood function* $r_{ij}(\theta)$ *defined in the Appendix, with* $I_*(\beta_0)$ *given in the proof.*

The proof of the theorems above is sketched in the Supplementary Material. To see why the proposed method yields a more efficient estimator than the conditional method theoretically, note the fact that by following Huang and Qin (2013) and assuming that Λ_0 is known, the conditional estimator $\hat{\beta}_n^C$ is the solution to

$$Q_1(\beta) = \frac{1}{n}\frac{\partial \log \mathcal{L}_n^C}{\partial \beta} = 0,$$

while the proposed estimator $\hat{\beta}_n$ is the solution to

$$Q_1(\beta) + Q_2(\beta) = 0, \quad \text{where} Q_2 = \frac{2}{n(n-1)}\frac{\partial \log \mathcal{L}_n^P}{\partial \beta}.$$

Also note that $E\{Q_1(\beta_0)^T Q_1(\beta_0)\} = 0$. It follows by the projection method that one can show that $\sqrt{n}Q_1(\beta_0) \rightarrow N(0, V_{C_1})$ and $\sqrt{n}\{Q_1(\beta_0) + Q_2(\beta_0)\} \rightarrow N(0, V_{C_1} + V_{P_1})$ with V_{C_1} and $V_{C_1} + V_{P_1}$ denoting the asymptotic covariance matrices, respectively.

By using the delta method, we have that $\sqrt{n}(\hat{\beta}_n^C - \beta_0) \rightarrow N(0, \Sigma_C)$ and $\sqrt{n}(\hat{\beta}_n - \beta_0) \rightarrow N(0, \Sigma)$, where $\Sigma_C = (V_{C_2}V_{C_1}^{-1}V_{C_2})^{-1}$ and $\Sigma = \{(V_{C_2} + V_{P_2})(V_{C_1} + V_{P_1})^{-1}(V_{C_2} + V_{P_2})\}^{-1}$ with

$$V_{C_2} = -E\left\{\frac{\partial \ell_1^c(\theta)}{\partial \beta^T \partial \beta}\right\} \quad \text{and} \quad V_{P_2} = -E\left\{\frac{\partial \ell_{12}^p(\theta)}{\partial \beta^T \partial \beta}\right\},$$

where $\ell_1^c(\theta)$ and $\ell_{12}^p(\theta)$ are defined in Appendix. To see the relationship between Σ and Σ_C, define

$$D = 2V_{P_2}V_{C_1}^{-1}V_{C_2} + V_{P_2}V_{C_1}^{-1}V_{P_2} - (V_{C_2} + V_{P_2})V_{C_1}^{-1}(V_{C_1}^{-1} + V_{P_2}^{-1})^{-1}V_{C_1}^{-1}(V_{C_2} + V_{P_2}).$$

Then we have $(V_{C_2} + V_{P_2})(V_{C_1} + V_{P_2})^{-1}(V_{C_2} + V_{P_2}) = V_{C_2}V_{C_1}^{-1}V_{C_2} + D$ and $\Sigma = \Sigma_C - \Sigma_C(D^{-1} + \Sigma_C)^{-1}\Sigma_C$ if D be positive definite. In other words, $\hat{\beta}_n$ is more efficient than $\hat{\beta}_n^C$ when D is positive definite. Also, note from Theorem 4.2 that the choice of $v = 1/(1 + d)$ yields the optimal rate of convergence $n^{d/\{2(1+d)\}}$, which equals $n^{1/3}$ when $d = 2$ and improves as d increases.

For inference about β, it is clear that we need to estimate the covariance matrix of $\hat{\beta}_n$ and for this, a simple and direct approach would be to treat the problem as a parametric problem and then use, for example, the EM methods (Louis 1982). However, this would not give a valid estimation since the standard maximum likelihood theory does not apply here. We could also derive a consistent estimator of Σ and it will be seen from the supplementary material that this would not be straightforward. Corresponding to these, the following simple bootstrap procedure will be used (Efron and Tibshirani 1993), and the numerical studies below indicate that it works well. Let B be a prespecified positive integer and for each $b = 1, \ldots, B$, draw a simple random sample $\mathcal{O}^{(b)} = \{\mathcal{O}_i^{(b)}; i = 1, \ldots, n\}$ of size n with replacement from the observed data $\mathcal{O} = \{\mathcal{O}_i; i = 1, \ldots, n\}$. Let $\hat{\beta}_n^{(b)}$ denote the proposed estimator of β based on the resampled data set $\mathcal{O}^{(b)}$. Then one can estimate the covariance matrix of $\hat{\beta}_n$ by

$$\hat{\Sigma} = \frac{1}{B-1} \sum_{b=1}^{B} \left\{ \hat{\beta}_n^{(b)} - \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}_n^{(b)} \right\}^{\otimes 2}.$$

5. A simulation study

A simulation study was performed to evaluate the empirical performance of the pairwise pseudo-likelihood estimation procedure proposed in the previous sections. It was assumed that there exist two covariates Z_1 and Z_2 following the Bernoulli (0.5) and the uniform distribution over $(0, 1)$, respectively. The failure times were generated from the semi-parametric transformation model (1) with $G(x) = r^{-1} \log(1 + rx)$ $(r \geq 0)$ and $\Lambda_0(t) = t$. Also we generated the underlying truncation times A^* from either the exponential distribution or the uniform distribution with the mean depending on the truncation rate. For the given sample size n, the realisations of A^*, T^*, Z^* were generated until n subjects satisfied the sampling constraint $A^* \leq T^*$. Note that these n remaining subjects are referred to as the observations A, T and Z.

For the generation of the censoring intervals or observation times U_{ij} for each subject, we first assumed that there exists a sequence of pre-specified observation time points $t_1 < \cdots < t_k$ for all subjects as in most of medical follow-up studies. Then it was assumed that each subject is observed with probability p at each of the pre-specified time point t_j's independently, and for subject i, the U_{ij}'s were defined to be the observed pre-specified observation time points.

For the simulation results given below, it was assumed that $t_j = A + 0.1j$, $j = 1, \ldots, k = 10$, and $p = 0.8$, and they are based on $B = 20$ and $m = [n^{1/4}]$ with 500 replications. Also we set $n = 400$ with the truncation rate being 30%, 50%, 70% or 85% in the case of the truncation time following the exponential distribution and $n = 400$and600 with the truncation rate being 30%, 50% or 70% in the case of the truncation time following the uniform distribution. To assess the proposed method, we calculated the empirical bias (Bias) given by the average of the proposed estimates minus the true value, the sample standard error (SEE) of the estimates, the mean of the estimated standard error (ESE), the 95% empirical coverage probability (CP), and the relative efficiency (RE) defined as the ratio of the sample variance based on the conditional method to that based on the proposed method.

Table 1 provides the results on estimation of β given by the proposed method with the true value of (β_1, β_2) being $(0.5, 0.5)$, A^* following the exponential distribution, and $r = 0, 0.5$ and 1. For comparison, we also applied the conditional method and include the obtained results in the table. One can see from Table 1 that the proposed estimator seems to be unbiased and the bootstrap variance estimation procedure appears to work well. Also the CP results suggest that the normal approximation to the distribution of the estimator seems to be appropriate, and as expected, the proposed method is more efficient than the

Table 1. Simulation results with A^* following the exponential distribution.

			Proposed approach				Conditional approach				
Truncation rate	r	Par.	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP	RE
30%	0	β_1	0.0036	0.1095	0.1089	0.948	0.0035	0.1159	0.1144	0.940	1.12
		β_2	−0.0144	0.1868	0.1866	0.940	−0.0143	0.2033	0.1977	0.928	1.18
	0.5	β_1	−0.0007	0.1680	0.1661	0.940	0.0016	0.1855	0.1780	0.930	1.22
		β_2	0.01540	0.2856	0.2858	0.940	0.0184	0.3091	0.3087	0.936	1.17
	1	β_1	−0.0059	0.2234	0.2210	0.924	−0.0038	0.2472	0.2408	0.934	1.22
		β_2	0.0132	0.3902	0.3855	0.938	0.01450	0.4340	0.4196	0.930	1.24
50%	0	β_1	0.0010	0.1060	0.1021	0.932	0.0095	0.1188	0.1158	0.940	1.26
		β_2	0.0060	0.1797	0.1792	0.946	0.0085	0.2017	0.2013	0.930	1.26
	0.5	β_1	−0.0011	0.1733	0.1661	0.926	−0.0007	0.1981	0.1965	0.940	1.31
		β_2	0.0039	0.2903	0.2866	0.924	0.0043	0.3188	0.3360	0.948	1.21
	1	β_1	−0.0161	0.2590	0.2459	0.922	−0.0174	0.2969	0.2984	0.940	1.31
		β_2	0.0070	0.4189	0.4239	0.942	0.0313	0.4932	0.5078	0.932	1.39
70%	0	β_1	0.0038	0.0911	0.0945	0.958	0.0025	0.1166	0.1183	0.936	1.64
		β_2	0.0049	0.1645	0.1572	0.910	0.0115	0.2085	0.2001	0.922	1.61
	0.5	β_1	−0.0181	0.1692	0.1636	0.936	−0.0162	0.2243	0.2273	0.946	1.76
		β_2	−0.0008	0.2798	0.2861	0.946	0.0239	0.3966	0.3947	0.952	2.01
	1	β_1	0.0012	0.2803	0.2774	0.932	0.0073	0.3785	0.3948	0.952	1.82
		β_2	0.0013	0.4551	0.4655	0.950	0.0396	0.6918	0.6790	0.920	2.31
85%	0	β_1	0.0066	0.0921	0.0910	0.934	0.0089	0.1170	0.1210	0.948	1.61
		β_2	0.0106	0.1508	0.1483	0.944	0.0065	0.2107	0.2014	0.920	1.95
	0.5	β_1	0.0039	0.1715	0.1595	0.930	0.0109	0.2431	0.2429	0.928	2.01
		β_2	0.0188	0.2678	0.2782	0.952	0.0287	0.4054	0.4183	0.952	2.29
	1	β_1	0.0053	0.3146	0.2990	0.924	0.0246	0.4779	0.5220	0.936	2.31
		β_2	0.0208	0.4771	0.5118	0.954	0.0462	0.7961	0.8479	0.956	2.78

Bias: Empirical bias given by the average of the proposed estimates minus the true value; SSE: the sample standard error of the estimates; ESE: the mean of the estimated standard error; CP: the 95% empirical coverage probability; RE: the relative efficiency which is the ratio of the sample variance based on the conditional method to that based on the proposed method.

conditional method. In addition, the relative efficiency gain of the proposed estimator over the conditional method increased as the truncation rate increased.

Figure 1 provides an intuitive representation of the estimated baseline cumulative hazards function $\hat{\Lambda}_n$ for the situations corresponding to Table 1 with truncation rates of 30%, 50% and 70%, respectively. The figure shows that the Bernstein polynomial approximation seems to work well. The same conclusions are given in Table 2, which gives the estimation results obtained based on the same set-up as in Table 1 except A^* following the uniform distribution with different sample sizes. In addition, Table 2 indicates that when the sample size increased, the proposed estimator yielded better performance as expected.

Note that we considered the two-dimensional β in the set-ups above. By following the suggestion of a reviewer, we also studied the four-dimensional β with true value $\beta_0 = (0.5, -0.5, 1, -1)^T$ under the exponential truncation and the uniform truncation, respectively, along with $Z = (Z_1^T, Z_2^T, Z_3^T, Z_4^T)^T$. Here $Z_1 \sim \mathrm{Ber}(0.5)$, $Z_2 \sim \mathrm{U}(0, 1)$ and Z_3 and Z_4 follow the standard normal distributions with the covariance of 0.2. Table 3 presents the obtained results with the other set-ups being the same as Tables 1 and 2 and $n = 400$. One can see that they gave similar conclusions and again indicate that the proposed method works well.

A problem of practical interest is what is the consequence or impact if one ignores the truncation. To see this, we repeated the study giving Table 1 but only calculated the empirical bias for both the proposed method and the proposed method but assuming no

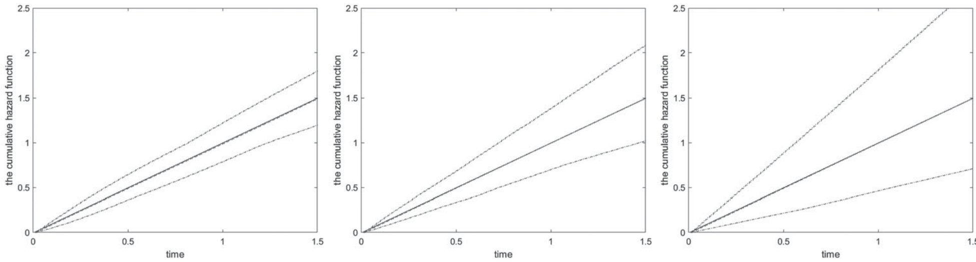

Figure 1. Estimate of the baseline cumulative hazards function $\Lambda(t)$ with truncation rate as 30% (Left), 50% (Middle) and 70% (Right): true function (solid), estimated function (dashed), 2.5% quantile and 97.5% quantile (dash-dot).

truncation and present the results in Table 4. Furthermore, Figure 2 shows the empirical biases for the estimation of β_1 and β_2 based on 85% truncation rate and various values for the transformation model parameter r. They both show that omitting the left truncation can lead to biased estimates.

6. An example

In this section, the proposed method is applied to a set of left-truncated and interval-censored data arising from a cohort of hemophiliacs concerning AIDS diagnosis discussed by Kim, De Gruttola, and Lagakos (1993) and Wang et al. (2021) among others. It consists of 257 subjects with Type A or B hemophilia who were at risk for infection by HIV through the contaminated blood factor that they received for their treatment. Among them, only 188 were found to be infected with the virus and 41 were subsequently diagnosed to have AIDS during the study. For each subject, the occurrence of the HIV infection was observed either exactly or within a narrow interval and for the latter case, for simplicity, a common way is to treat the midpoint of the observed interval as the actual HIV-1 infection time (Kim et al. 1993; Wang et al. 2021). Furthermore, for the AIDS diagnosis time, only interval-censored data are available with being left-truncated by the HIV infection time. For the analysis below, as others, we will focus on the 188 HIV infected subjects and are mainly interested on assessing the effects of the level of hemophilia treatment received and the age described below on the development of AIDS-related symptoms.

The study includes two factors or covariates whose effects on the AIDS diagnosis time were of interest. One is the treatment group indicator: the patients were divided into two groups, heavily and lightly treated, based on the amount of the contaminated blood received. The other is the age indicator, younger than 20 or not at the time of HIV infection. For the application of the proposed estimation procedure, define T_i to be the AIDS diagnosis time of the patient i from 1 January 1978, the beginning of the study, with the time scale being half a year. Also define $Z_{1i} = 0$ if the patient i belongs to the lightly treated group and 1 otherwise, and $Z_{2i} = 0$ the ith patient's age was below 20 and 1 otherwise.

To apply the proposed method, as in the simulation study, we set $G(x) = r^{-1} \log(1 + rx)$ $(r \geq 0)$ in model (1) and considered several values for r and m, the degree of Bernstein polynomials, around $m = [n^{1/4}]$. For the selection of the proper or optimal values of r and

Table 2. Simulation results with A^* following the uniform distribution.

				Proposed approach				Conditional approach				
Truncation rate	r	n	Par.	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP	RE
30%	0	400	β_1	0.0043	0.1119	0.1126	0.930	0.0038	0.1155	0.1149	0.922	1.07
			β_2	−0.0043	0.1935	0.1927	0.950	−0.0020	0.2015	0.1974	0.942	1.08
		600	β_1	0.0047	0.0962	0.0913	0.924	0.0061	0.0990	0.0934	0.928	1.06
			β_2	0.0095	0.1565	0.1568	0.942	0.0099	0.1576	0.1614	0.942	1.01
	0.5	400	β_1	0.0052	0.1821	0.1700	0.928	0.0042	0.1845	0.1765	0.938	1.03
			β_2	0.0129	0.3017	0.2943	0.938	0.0130	0.3123	0.3081	0.942	1.07
		600	β_1	0.0241	0.1444	0.1375	0.932	0.0257	0.1495	0.1432	0.932	1.07
			β_2	0.0053	0.2432	0.2399	0.936	0.0018	0.2562	0.2499	0.926	1.11
	1	400	β_1	0.0022	0.2511	0.2360	0.928	0.0007	0.2606	0.2536	0.932	1.08
			β_2	0.0258	0.3962	0.4072	0.944	0.0181	0.4195	0.4359	0.948	1.12
		600	β_1	−0.0091	0.1819	0.1931	0.962	−0.0153	0.1967	0.2059	0.956	1.17
			β_2	−0.0009	0.3312	0.3341	0.946	−0.0005	0.3526	0.3574	0.950	1.13
50%	0	400	β_1	0.0021	0.1094	0.1060	0.934	−0.0038	0.1167	0.1165	0.936	1.14
			β_2	−0.0010	0.1912	0.1825	0.914	−0.0072	0.2042	0.1985	0.934	1.14
		600	β_1	0.0111	0.0856	0.0872	0.932	0.0117	0.0921	0.0953	0.946	1.16
			β_2	0.0001	0.1501	0.1487	0.940	0.0009	0.1621	0.1625	0.940	1.17
	0.5	400	β_1	0.0038	0.1839	0.1753	0.934	0.0073	0.2042	0.2027	0.944	1.23
			β_2	0.0233	0.3027	0.3015	0.934	−0.0243	0.3451	0.3482	0.948	1.30
		600	β_1	−0.0115	0.1379	0.1440	0.946	−0.0082	0.1586	0.1651	0.944	1.32
			β_2	−0.0062	0.2584	0.2472	0.942	0.0049	0.2983	0.2811	0.930	1.33
	1	400	β_1	−0.0035	0.2644	0.2744	0.938	−0.0058	0.3130	0.3303	0.948	1.40
			β_2	−0.0039	0.4450	0.4655	0.948	0.0007	0.5394	0.5570	0.956	1.47
		600	β_1	0.0153	0.2196	0.2168	0.930	0.0239	0.2727	0.2595	0.930	1.54
			β_2	0.0027	0.3670	0.3775	0.944	−0.0169	0.4303	0.4458	0.950	1.37
70%	0	400	β_1	0.0096	0.0882	0.0926	0.948	0.0157	0.1188	0.1187	0.938	1.81
			β_2	0.0040	0.1596	0.1558	0.926	0.0054	0.2035	0.2020	0.934	1.63
		600	β_1	0.0020	0.0746	0.0745	0.944	0.0043	0.0965	0.0956	0.954	1.67
			β_2	0.0004	0.1315	0.1266	0.942	0.0104	0.1580	0.1633	0.954	1.44
	0.5	400	β_1	−0.0032	0.1667	0.1688	0.940	−0.0076	0.2301	0.2397	0.942	1.91
			β_2	0.0250	0.3004	0.2929	0.934	0.0397	0.4257	0.4160	0.930	2.01
		600	β_1	0.0055	0.1437	0.1391	0.942	0.0018	0.2000	0.1986	0.940	1.94
			β_2	−0.0199	0.2456	0.2371	0.926	−0.0173	0.3472	0.3333	0.932	2.00
	1	400	β_1	0.0009	0.2937	0.3007	0.954	0.0192	0.4162	0.4634	0.958	2.01
			β_2	0.0081	0.5268	0.5132	0.926	−0.0176	0.7450	0.8068	0.954	2.00
		600	β_1	0.0069	0.2504	0.2384	0.930	−0.0053	0.3480	0.3447	0.916	1.93
			β_2	−0.0278	0.4212	0.4137	0.938	−0.0341	0.5548	0.5919	0.950	1.73

The meaning of Bias, SSE, ESE, CP and RE are same as Table 1.

m, we employed the AIC and chose their combination that minimised

$$\mathrm{AIC} = -2n\ell_n(\hat{\theta}_n) + 2(p + m + 1).$$

It turned out that the optimal combination was $r = 0$, corresponding to the proportional hazards model, and $m = 4$, and the obtained results are given in Table 5 with $B = 20, 100$ or 200 for the variance estimation.

The results include the estimated covariate effects, their estimated standard errors and the p-values for testing the covariate effect being zero. For comparison, we also obtained and include in the table the results given by the conditional method. One can see from the table that all analysis results are consistent and suggest that there was no significant relationship between the age and the AIDS diagnosis time. In contrast, it appears that the patients in the heavily treated group had a significantly shorter time or higher risk of being diagnosed with AIDS than those in the lightly treated group.

Table 3. Simulation results with a larger dimensional covariate.

			Exponential truncation				Uniform truncation			
Truncation rate	r	Par.	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
30%	0	β_1	0.0006	0.1370	0.1361	0.934	0.0065	0.1352	0.1404	0.938
		β_2	0.0013	0.2301	0.2322	0.946	−0.0066	0.2440	0.2358	0.940
		β_3	0.0126	0.0921	0.0908	0.952	0.0167	0.0938	0.0932	0.932
		β_4	−0.0164	0.0904	0.0908	0.938	−0.0158	0.0916	0.0924	0.948
	0.5	β_1	0.0111	0.1899	0.1936	0.942	0.0016	0.2034	0.2045	0.938
		β_2	−0.0210	0.3401	0.3439	0.950	0.0148	0.3629	0.3497	0.920
		β_3	0.0054	0.1199	0.1233	0.950	0.0110	0.1286	0.1254	0.948
		β_4	−0.0118	0.1276	0.1207	0.920	−0.0225	0.1286	0.1273	0.920
	1	β_1	0.0083	0.2490	0.2572	0.940	−0.0016	0.2795	0.2720	0.942
		β_2	−0.0077	0.4239	0.4453	0.940	0.0044	0.4404	0.4684	0.960
		β_3	0.0068	0.1526	0.1583	0.944	0.0065	0.1217	0.1678	0.944
		β_4	−0.0112	0.1552	0.1569	0.940	−0.0049	0.1615	0.1690	0.948
50%	0	β_1	0.0113	0.1344	0.1334	0.930	0.0116	0.1397	0.1393	0.950
		β_2	−0.0197	0.2312	0.2243	0.932	−0.0034	0.2382	0.2376	0.936
		β_3	0.0154	0.0919	0.0927	0.938	0.0207	0.0972	0.0981	0.946
		β_4	−0.0166	0.0898	0.0937	0.938	−0.0140	0.0955	0.0977	0.946
	0.5	β_1	0.0091	0.1946	0.2019	0.948	−0.0121	0.2223	0.2165	0.926
		β_2	−0.0118	0.3506	0.3536	0.926	0.0096	0.3678	0.3750	0.960
		β_3	0.0063	0.1267	0.1287	0.926	0.0094	0.1370	0.1392	0.946
		β_4	−0.0043	0.1260	0.1302	0.928	−0.0090	0.1401	0.1391	0.936
	1	β_1	−0.0115	0.2859	0.2908	0.930	0.0138	0.3051	0.3229	0.964
		β_2	−0.0205	0.4878	0.5003	0.940	−0.0056	0.5426	0.5633	0.954
		β_3	0.0131	0.1822	0.1859	0.948	0.0383	0.2028	0.2154	0.936
		β_4	−0.0154	0.1703	0.1837	0.952	−0.0343	0.1932	0.2138	0.950
70%	0	β_1	0.0085	0.1235	0.1290	0.946	0.0146	0.1293	0.1340	0.942
		β_2	0.0045	0.2180	0.2197	0.950	0.0109	0.2230	0.2245	0.944
		β_3	0.0131	0.0937	0.0986	0.956	0.0155	0.1006	0.1006	0.928
		β_4	−0.0134	0.0996	0.0976	0.930	−0.0116	0.0994	0.1006	0.938
	0.5	β_1	0.0009	0.1994	0.2113	0.954	0.0105	0.2216	0.2272	0.944
		β_2	−0.0215	0.3465	0.3575	0.956	0.0061	0.3736	0.3881	0.954
		β_3	0.0124	0.1370	0.1400	0.934	0.0318	0.1379	0.1488	0.944
		β_4	−0.0107	0.1373	0.1373	0.928	−0.0237	0.1422	0.1503	0.954
	1	β_1	−0.0237	0.3211	0.3353	0.936	0.0294	0.3432	0.3872	0.950
		β_2	0.0462	0.5493	0.5812	0.950	0.0126	0.5985	0.6305	0.954
		β_3	0.0369	0.2075	0.2267	0.958	0.0241	0.2182	0.2181	0.962
		β_4	−0.0221	0.1952	0.2245	0.956	−0.0249	0.2224	0.2409	0.940

The meaning of Bias, SSE, ESE, CP and RE are same as Table 1.

Following the suggestion of a reviewer, we also performed the analysis by applying the proposed method but assuming a parametric model such as the Weibull model and the proportional odds model. Table 6 provides a summary of the results with $B = 20$ and gave similar conclusions as before except the proportional odds model, which does not provide a clear indication of the significance of treatment extent. To see the results graphically, Figure 3 displays the estimated survival functions given by the proposed method for the subjects in the four groups defined by the two covariates. Figure 4 presents the nonparametric maximum likelihood estimator (NPMLE) of the survival function given by the corrected Turnbull estimator proposed in Frydman (1994). The estimated survival functions again demonstrate that the covariate ‘Age’ had no influence while the two survival functions with different treatment groups show a non-ignorable difference.

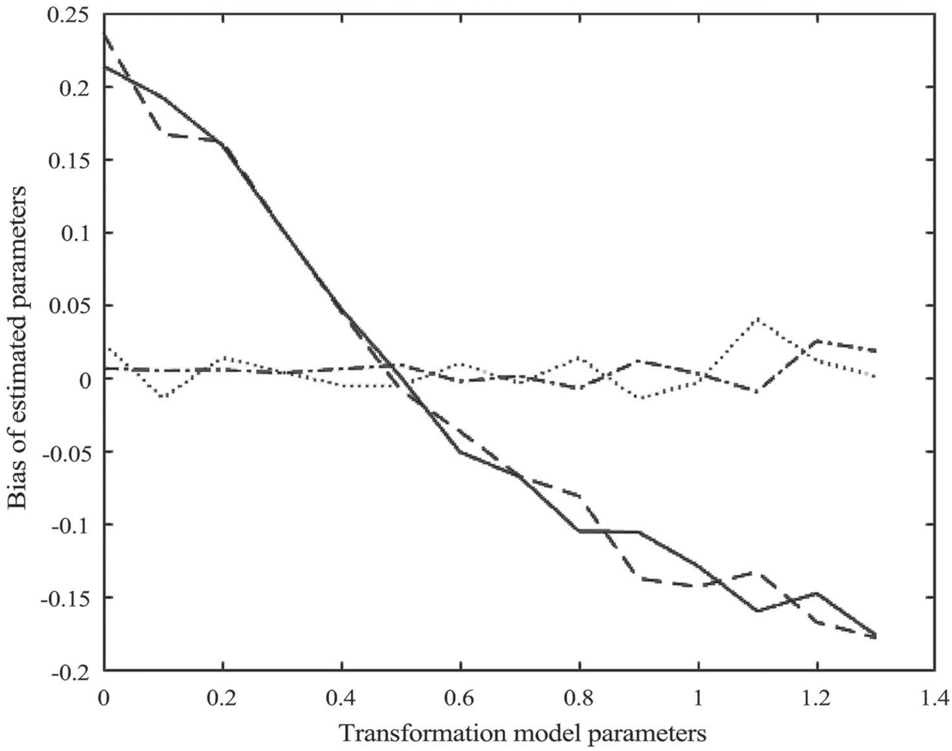

Figure 2. Empirical biases of β_1 (solid), β_2 (dashed) assuming no truncation and β_1 (dash-dot), β_2 (dotted) under proposed method with respect to r.

Table 4. Simulation results on the empirical bias with A^* following the exponential distribution.

Rate	r	Par.	Proposed approach	Ignoring truncation	Rate	r	Par.	Proposed approach	Ignoring truncation
30%	0	β_1	0.0128	0.0455	70%	0	β_1	−0.0073	0.1523
		β_2	0.0045	0.0379			β_2	0.0020	0.1691
	0.5	β_1	−0.0074	−0.0289		0.5	β_1	−0.0056	−0.0078
		β_2	0.0374	0.0162			β_2	0.0361	0.0375
	1	β_1	0.0036	−0.0745		1	β_1	0.0154	−0.1437
		β_2	0.0378	−0.0428			β_2	0.0356	−0.0998
50%	0	β_1	0.0119	0.0905	85%	0	β_1	0.0010	0.2037
		β_2	0.0029	0.0786			β_2	0.0374	0.2543
	0.5	β_1	0.0076	−0.0076		0.5	β_1	0.0194	0.0208
		β_2	−0.0064	−0.0216			β_2	−0.0155	−0.0412
	1	β_1	0.0214	−0.0868		1	β_1	0.0117	−0.1076
		β_2	−0.0353	−0.1248			β_2	0.0110	−0.0879

7. Conclusion

In this paper, we considered the estimation problem about the semi-parametric transformation models when one faces left-truncated and case K interval-censored failure data, for which there did not seem to exist an established inference procedure. As mentioned above, a key feature or advantage of models is their flexibility as they allow one to model various types of covariate effects and include many commonly used models such as the

Table 5. Analysis results for the AIDS cohort study.

		$r = 0, B = 20$		$r = 0, B = 100$		$r = 0, B = 200$	
Covariate	Index	Proposed approach	Conditional approach	Proposed approach	Conditional approach	Proposed approach	Conditional approach
Group	Estimator	0.7023	0.7361	0.7023	0.7361	0.7023	0.7361
	SD	0.2964	0.3026	0.3445	0.3473	0.3243	0.3284
	p value	0.0178	0.0150	0.0415	0.0340	0.0303	0.0250
Age	Estimator	0.0193	0.0641	0.0193	0.0641	0.0193	0.0641
	SD	0.2598	0.2661	0.3713	0.3755	0.3403	0.3486
	p value	0.9408	0.8096	0.9586	0.8645	0.9548	0.8541

SD: standard deviation of the estimator.

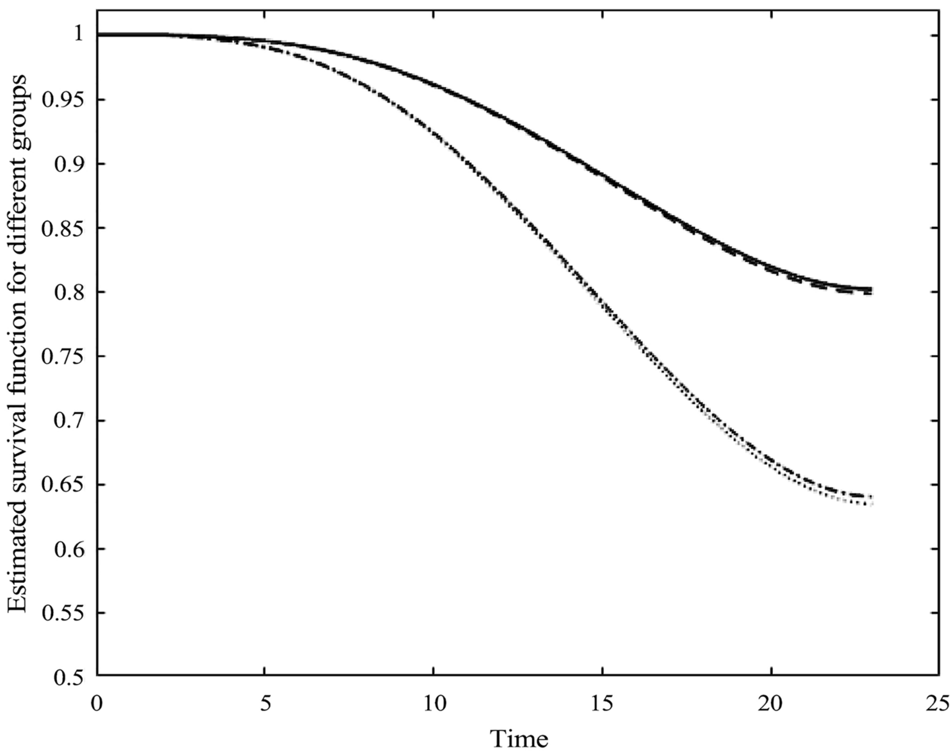

Figure 3. Estimates of the survival functions for different treatment groups and ages: lightly group with age below 20 (solid), lightly group with age over 20 (dashed), heavily group with age below 20 (dash-dot) and heavily group with age over 20 (dotted).

proportional hazards and odds models as special cases. For the problem, an effective pseudo-likelihood estimation method was proposed with the use of the sieve approach. The asymptotic properties of the resulting estimators were established and the conducted simulation study indicated that the proposed method works well under actual conditions. Compared with the traditional conditional approach, the proposed method made use of the relevant information in the marginal likelihood of truncation time and does not need the estimation of the unknown truncation distribution.

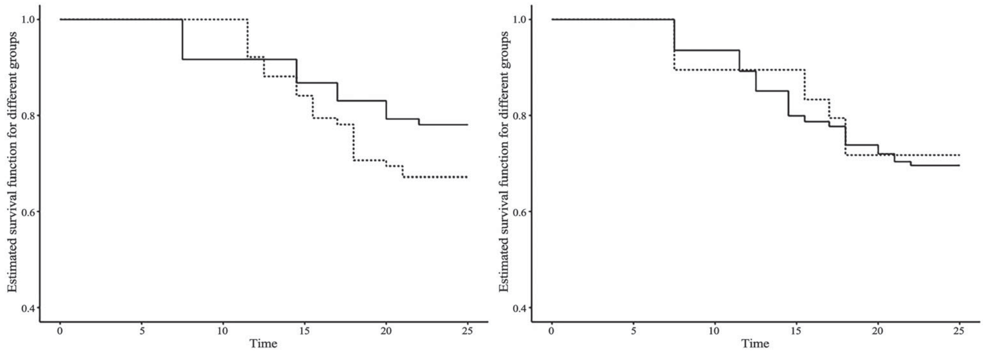

Figure 4. NPMLE estimates of the survival functions for different treatment groups (Left) and ages (Right): lightly group (solid), heavily group (dotted); age over 20 (solid), age below 20 (dotted).

Table 6. Comparison for the AIDS cohort study.

	Proportional hazards model			Proportional odds model			Weibull model		
Model Covariate	Estimator	SD	p-value	Estimator	SD	p-value	Estimator	SD	p-value
Group	0.7023	0.2964	0.0178	1.2260	0.9583	0.2008	0.7100	0.3458	0.0400
Age	0.0193	0.2598	0.9408	0.0031	0.7520	0.9967	0.0242	0.3714	0.9481

SD: standard deviation of the estimator.

In the previous sections, for simplicity, we have assumed that covariates do not have effects on the left truncated time or variable and it is apparent that this may not be true in practice. To address this, one could employ model (1) or other commonly used regression models and an estimation procedure similar to that proposed above could be developed. The same is true for the censoring mechanism or the observation process that generates the U_{ij}'s. Another assumption used in the above is that both left truncation variable and the observation process are independent of the failure time of interest. It is clear that this could be violated sometimes, or the observation process could be related to the failure time of interest. In other words, we have dependent or informative interval censoring (Sun 2006) and as a future research direction, It would be useful to generalise the proposed method to these situations although would not be straightforward as much more complicated modelling would be needed in general. Another research direction is to develop some goodness-of-fit tests for the selection of a particular transformation function or model in practice.

Acknowledgments

The authors wish to thank Associate Editor and two reviewers for their many helpful and useful comments and suggestions that greatly improved the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was partially supported by the National Nature Science Foundation of China (Grant Nos. 11801212, 12071176).

References

Chen, K., Jin, Z., and Ying, Z. (2002), 'Semiparametric Analysis of Transformation Models with Censored Data', *Biometrika*, 89, 659–668.

De Gruttola, V., and Lagakos, S. (1989), 'Analysis of Doubly-censored Survival Data, with Application to AIDS', *Biometrics*, 45, 1–11.

Efron, B., and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*, CRC Press.

Fine, J., Ying, Z., and Wei, L. (1998), 'On the Linear Transformation Model for Censored Data', *Biometrika*, 85, 980–986.

Frydman, H.A. (1994), 'A Note on Nonparametric Estimation of the Distribution Function from Interval-censored and Truncated Observations', *Journal of the Royal Statistical Society, Series B*, 56, 71–74.

Gao, F., and Chan, K.C.G. (2019), 'Semiparametric Regression Analysis of Length-biased Interval-censored Data', *Biometrics*, 75, 121–132.

Hazelrig, J.B., Turner, M.E., and Blackstone, E.H. (1982), 'Parametric Survival Analysis Combining Longitudinal and Cross-sectional-censored and Interval-censored Data with Concomitant Information', *Biometrics*, 38, 1–15.

Huang, J. (1996), 'Efficient Estimation for the Proportional Hazards Model with Interval Censoring', *The Annals of Statistics*, 24, 540–568.

Huang, C.Y., and Qin, J. (2012), 'Composite Partial Likelihood Estimation Under Length-biased Sampling, with Application to A Prevalent Cohort Study of Dementia', *Journal of the American Statistical Association*, 107, 946–957.

Huang, C.Y., and Qin, J. (2013), 'Semiparametric Estimation for the Additive Hazards Model with Left-truncated and Right-censored Data', *Biometrika*, 100, 877–888.

Huang, J., and Rossini, A. (1997), 'Sieve Estimation for the Proportional-odds Failure-time Regression Model with Interval Censoring', *Journal of the American Statistical Association*, 92, 960–967.

Huang, J., and Wellner, J.A. (1993), 'Regression Models with Interval Censoring', in *Probability Theory and Mathematical Statistics*, pp. 269–296.

Huang, J., and Wellner, J.A. (1997), 'Interval Censored Survival Data: A Review of Recent Progress', in *Biostatistics Proceedings of the First Seattle Symposium*, pp. 123–169.

Kalbfleisch, J.D., and Prentice, R.L. (2011), *The Statistical Analysis of Failure Time Data*, John Wiley & Sons.

Kim, J.S. (2003), 'Efficient Estimation for the Proportional Hazards Model with Left-truncated and Case 1 Interval-censored Data', *Statistica Sinica*, 13, 519–537.

Kim, M.Y., De Gruttola, V., and Lagakos, S. (1993), 'Analyzing Doubly Censored Data with Covariates with Application to AIDS', *Biometrics*, 49, 13–22.

Li, J., and Ma, S. (2010), 'Interval-censored Data with Repeated Measurements and a Cured Subgroup', *Journal of the Royal Statistical Society, Series C*, 59, 693–705.

Liang, K.Y., and Qin, J. (2000), 'Regression Analysis Under Non-standard Situations: A Pairwise Pseudo Likelihood Approach', *Journal of the Royal Statistical Society, Series B*, 62, 773–786.

Louis, T.A. (1982), 'Finding the Observed Information Matrix When Using the EM Algorithm', *Journal of the Royal Statistical Society, Series B*, 44, 226–233.

Marshall, M.L. (1974), 'Fitting the Two-term Mixed Exponential and Two-parameter Lognormal Distributions to Grouped and Censored Data', *Applied Statistics*, 23, 313–322.

Odell, P.M., Anderson, K.M., and D'Agostino, R.B. (1992), 'Maximum Likelihood Estimation for Interval-censored Data Using a Weibull-based Accelerated Failure Time Model', *Biometrics*, 48, 951–959.

Pan, W., and Chappell, R. (1999), 'A Note on Inconsistency of NPMLE of the Distribution Function from Left Truncated and Case I Interval-censored Data', *Lifetime Data Analysis*, 5, 281–291.

Pan, W., and Chappell, R. (2002), 'Estimation in the Cox Proportional Hazards Model with Left-truncated and Interval-censored Data', *Biometrics*, 58, 64–70.
Rossini, A., and Tsiatis, A. (1996), 'A Semiparametric Proportional Odds Regression Model for the Analysis of Current Status Data', *Journal of the American Statistical Association*, 91, 713–721.
Shen, X. (1998), 'Proportional Odds Regression and Sieve Maximum Likelihood Estimation', *Biometrika*, 85, 165–177.
Shen, P. (2014), 'Proportional Hazards Regression with Interval-censored and Left-truncated Data', *Journal of Statistical Computation and Simulation*, 84, 264–272.
Shen, P., Chen, H., Pan, W., and Chen, C. (2019), 'Semiparametric Regression Analysis for Left-truncated and Interval-censored Data without or with a Cure Fraction', *Computational Statistics & Data Analysis*, 140, 74–87.
Sun, J. (2006), *The Statistical Analysis of Interval-censored Failure Time Data*, New York: Springer.
Wang, P., Li, D., and 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., and Sun, J. (2015), 'Efficient Estimation for the Additive Hazards Model in the Presence of Left-truncation and Interval Censoring', *Statistics and Its Interface*, 8, 391–402.
Wang, P., Zhao, H., Du, M., and Sun, J. (2018), 'Inference on Semiparametric Transformation Model with General Interval-censored Failure Time Data', *Journal of Nonparametric Statistics*, 30, 758–773.
Wu, F., Kim, S., Qin, J., Saran, R., and Li, Y. (2018), 'A Pairwise Likelihood Augmented Cox Estimator for Left-truncated Data', *Biometrics*, 74, 100–108.
Xu, D., Zhao, S., Hu, T., and Sun, J. (2019), 'Regression Analysis of Informatively Interval-censored Failure Time Data with Semiparametric Linear Transformation Model', *Journal of Nonparametric Statistics*, 31, 663–679.
Zeng, D., Mao, L., and Lin, D. (2016), 'Maximum Likelihood Estimation for Semiparametric Transformation Models with Interval-censored Data', *Biometrika*, 103, 253–271.
Zhang, Z., Sun, L., Zhao, X., and Sun, J. (2005), 'Regression Analysis of Interval-censored Failure Time Data with Linear Transformation Models', *Canadian Journal of Statistics*, 33, 61–70.
Zhang, Z., and Zhao, Y. (2013), 'Empirical Likelihood for Linear Transformation Models with Interval-censored Failure Time Data', *Journal of Multivariate Analysis*, 116, 398–409.
Zhou, Q., Hu, T., and Sun, J. (2017), 'A Sieve Semiparametric Maximum Likelihood Approach for Regression Analysis of Bivariate Interval-censored Failure Time Data', *Journal of the American Statistical Association*, 112, 664–672.

Appendix. Technical details

Let $\ell_i^c(\theta)$ and $\ell_{i,j}^p(\theta)$ denote the log $\mathcal{L}_n^C$ and $\mathcal{L}_n^P$, respectively, corresponding to subject i and pair sample (i,j). For two independent observations O and O', denote $W = (O, O')$, then rewrite $\ell_n(\theta) = [2/\{n(n-1)\}] \sum_{1\le i<j\le n} r(\theta, W_{ij})$ with $W_{ij} = (O_i, O_j)$ and $r(\theta, W_{ij}) = r_{ij}(\theta)$, where $r_{ij}(\theta) = \{\ell_i^c(\theta) + \ell_j^c(\theta)\}/2 + \ell_{i,j}^p(\theta)$. The letter C represents a constant, and it does not necessarily represent the same value each time here and in the proof. Let $P^2 = P \otimes P$ denote the product probability measure. Then, we establish the asymptotic properties of $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n)$ under the following regularity conditions.

Condition A.1: The true value β, denoted by β_0, lies in the interior of a known compact set $\mathcal{B}$ in R^p. The true cumulative baseline hazard function $\Lambda_0(t)$ is dth continuously differentiable for $d \ge 2$ and strictly increasing on $[0, \tau]$ with $\Lambda_0(0) = 0$.

Condition A.2: The covariate vector Z has bounded support in R^p and $cov(Z)$ is nonsingular.

Condition A.3: The number of monitoring times, K, is positive, and $E(K) < \infty$. With probability 1, there exists a constant $\delta > 0$ such that $P(A^* \le U_1 \mid Z) > \delta$ and $P(U_K \ge \tau \mid Z) \ge \delta$. In addition, there exists a positive η such that $P(U_j - U_{j-1} > \eta) = 1$, where $j = 2, \ldots, K$. Furthermore the union support of A, U_j is contained in an interval $[a, b]$, where $0 < a < b < \infty$ with $0 < \Lambda_0(a) < \Lambda_0(b) < \infty$ and $j = 1, \ldots, K$.

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

Condition A.5: For every θ in a neighbourhood of θ_0, $P^2\{r(\theta, W) - r(\theta_0, W)\} \leq -Cd^2(\theta, \theta_0)$.

Condition A.6: $0 < P^2\{r'(\theta_0, W)[\iota]\} < \infty$ for all $\iota \neq 0, \iota \in V$, where V denotes a linear span of $\Theta - \theta_0$; For $\theta \in \{\theta \in \Theta, d(\theta, \theta_0) = O(\delta_n)\}$, $P^2\{r''(\theta, W)[\theta - \theta_0, \theta - \theta_0] - r''(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} = O(\delta_n^3)$ and $\delta_n^3 = o(n^{-1})$.