Statistics and Its Interface Volume 12 (2019) 77–92

A semiparametric linear transformation model for general biased-sampling and right-censored data

Wenhua Wei*, Yong Zhou†, and Alan T. K. Wan‡

The semiparametric linear transformation (SLT) model is a useful alternative to the proportional hazards ([9]) and proportional odds ([4]) models for studying the dependency of survival time on covariates. In this paper, we consider the SLT model for biased-sampling and right-censored data, a feature commonly encountered in clinical trials. Specifically, we develop an unbiased estimating equations approach based on counting process for the simultaneous estimation of unknown coefficients and handling of sampling bias. We establish the consistency and the asymptotic normality of the proposed estimator, and provide a closed form expression for the estimator's covariance matrix that can be consistently estimated by a plug-in method. In a simulation study, we compare the finite sample properties of the proposed estimator with those of existing methods that either assumes that the sampling bias is of the length-bias type, or ignores the sampling bias altogether. The proposed method is further illustrated by two real clinical datasets.

AMS 2000 subject classifications: Primary 62N01; secondary 62F12.
Keywords and phrases: Biased-sampling, Estimating equation, Right-censoring, Semiparametric linear transformation model.

1. INTRODUCTION

Biased-sampling arises when there is a preferential selection of observations from the population. Ignoring the sampling bias will lead to biased estimators of the parameters because the sample is no longer a faithful representation of the target population. Left truncation, which often occurs in survival studies, is well-known to produce biased samples. There is an extensive literature on methods for handling left-truncated data. See, for example, [20, 21, 40, 42], among others. In particular, when the left-truncated variable is uniformly distributed, the variable is said to be length-biased. Studies focusing on inference under length-biased sampling include [2, 3, 14, 15, 16, 27, 30, 31, 32, 34, 35, 39, 43], among others.

Suppose that T is the observed failure time in a survival study. Under length-biased sampling ([15, 23, 38]), $T = A + V$, where A is the left truncation or backward recurrence time, measured from the onset of event (e.g., the beginning of disease) to the time of examination, and V is the residual survival or forward recurrence time, measured from the time of examination to the event of interest (e.g., death). Typically, V is also right-censored by some censoring time C. The total censoring time is thus $A + C$, and censoring is informative because the failure and censoring times share the common backward recurrence time A. While length-biased sampling has been extensively studied, it is just one type of biased-sampling encountered in practice. [22] gave an example of biased-sampling that does not fall within the category of length-biased sampling. They considered the situation where individuals are randomly asked to provide data on a variable X with density $f(x)$. These individuals may or may not respond, and let $w(x)$ be the probability that an individual responds with $w(\cdot)$ being a known function. Hence X has a conditional density that is proportional to $w(x)f(x)$. The observations of X thus constitute a biased sample, but they are not length-biased because there is no observable backward recurrence time A. For estimation under general biased-sampling, [41] considered a proportional hazards (PH) model and proposed a pseudo-partial likelihood approach. [41] assumed that the the density function $g(t|Z, R(t))$ of the observed biased data is

$$g(t|Z,R) = W(t,Z)f(t|R(\cdot))/\alpha(Z,R), \tag{1}$$

where $W(t,Z)$ is a known nonnegative weight function, $R(t)$ is a time-dependent covariate, Z is a time-independent covariate, $f(t|R(\cdot))$ denotes a conditional density function given $R(s)$ for $s \le t$ and $\alpha(Z,R)$ is a normalising constant making $g(t|Z,R)$ a genuine density function. The density function in (1) includes those under left-truncated and length-biased sampling as special cases with $W(t,Z) = I(t \ge Z)$ and $W(t,Z) = t$, respectively. For these latter two cases, $W(t,Z)$ in the density function is completely determined by the distribution of the truncated variable; as well, the truncated and biased variables are both observed. On the other hand, under general biased-sampling (1), only the biased variable is observed but the truncated variable is latent or missing. This is the major difference between (1) and

*Corresponding author. Assistant Professor at Shanghai University of Finance and Economics.
†Professor at Academy of Mathematics and Systems Science.
‡Professor at City University of Hong Kong.

left-truncated or length-biased sampling. In other words, the left-truncated and length-biased data can be treated as complete data, whereas the general biased-sampling data are incomplete data with missing truncated variable. Thus, under general biased-sampling, the data are less informative ([41]), and consequently, existing approaches for analysing left-truncated and length-biased data are inapplicable and new methods are needed.

The purpose of this paper is to examine the covariate effects on a simplified version of the general biased-sampling setup of (1), given by

$$g_T(t|Z) = W(t,Z)f_{T^*}(t|Z)/u(Z), \tag{2}$$

where $g_T(t|Z)$ is the conditional density function of T, an observed nonnegative random variable conditional on Z, the time-independent covariate, $W(t,Z)$ is a known nonnegative weight function, $f_{T^*}(t|Z)$ is the conditional density function of the unbiased sample T^* given Z, and $u(Z) = \int_0^\infty W(t,Z)f_{T^*}(t|Z)dt < \infty$ is a normalising constant. We assume that the data are subject to both biased-sampling and right-censoring.

For the analysis of covariate effects on survival time, the PH and proportional odds (PO) models are by far the most popular methods. Although these models make no parametric assumptions about the baseline hazards, they assume multiplicative covariate effects on either the hazard function or the odds of survival beyond t, which may be difficult to justify in some situations. An alternative modelling framework that has gained popularity in recent years is the semiparametric linear transformation (SLT) model, which relates an unknown transformation of the survival time linearly to the covariates, and includes the PH and PO models as special cases ([7]). The SLT model has been extensively studied for the case of right-censored data. In particular, [7] proposed an inverse probability weighting approach to estimate the coefficients. Although their procedure avoids the estimation of the infinite dimensional transformation function, it assumes independence between the censoring variable and the covariates. [5] relaxed this covariate-independent censoring assumption and suggested a martingale-based approach for the simultaneous estimation of finite and infinite dimensional parameters. There have been various refinements and modifications of approaches in [5] and [7] to other data types, including the work of [8, 12, 18, 19, 24, 25, 44], among others. [45] considered a broader class of transformation models that can accommodate time-varying covariates and recurrent events. In addition, [6] and [26] studied the varying-coefficient and partially linear transformation models respectively, both being extensions of the SLT model.

As mentioned above, [41] considered the data assumption (1) in conjunction with the PH model. In addition, [28] considered the biased-sampling scheme in (2) but assumed the data were uncensored, and used the accelerated failure time (AFT) model as their basis of analysis. [18] studied the SLT model for general biased-sampling data for the case where the data are sampled with bias after having been subject to right-censoring. Furthermore, [8, 23, 44] considered the SLT model under the length-biased data. To the best of our knowledge, the biased-sampling scheme (2) coupled with right-censored data under the SLT model is unexplored, and in this paper we take steps in this direction. The major difference between our and [18] work is that we consider data that are subject to right-censoring *after* they have been sampled with bias, whereas [18] considered the opposite. As noted by [41], the case we consider is more complex and difficult to handle. It is also more commonly encountered than the case considered by [18] - in clinical trials, usually the biased sample data due to preferential selection first arises and then the sample data are subject to right-censoring due to loss of follow-up, and not the other way round.

The remainder of this paper is organised as follows. Section 2 begins with a presentation of the model and notations. The same section also contains a derivation of the proposed estimation method and the description of an iterative algorithm for computing the estimator. Section 3 develops an asymptotic theory of the proposed method. Simulation results on the finite sample properties of the method are reported in Section 4. Section 5 contains applications of the method to two data sets on heart transplant and dementia. Section 6 contains some concluding remarks. The proof of our main theorem is given in the Appendix.

2. MODEL, NOTATIONS, ESTIMATION METHODOLOGY AND COMPUTATIONAL ALGORITHM

2.1 Model and notations

Let T^* be the unbiased failure time variable of interest, and Z be a $p \times 1$ dimensional time-independent covariate. Denote the conditional density function and survival function of T^* given Z as $f_{T^*}(t|Z)$ and $S_{T^*}(t|Z)$ respectively. We assume the following SLT model for T^*:

$$H(T^*) = -Z^T\beta + \epsilon, \tag{3}$$

where $H(\cdot)$ is an unknown monotonic increasing transformation function, β is an unknown $p \times 1$ vector of regression coefficients, and ϵ is an error term with a known distribution and independent of Z. We use $\lambda_\epsilon(t)$ and $\Lambda_\epsilon(t)$ to denote the hazard and cumulative hazard functions of ϵ, respectively. The special cases of the PH and PO models result when ϵ follows the extreme value and standard logistic distributions ([7]), respectively. Moreover, model (3) generalizes the Box-Cox transformation models when ϵ follows the standard normal distribution in the absence of right-censoring ([26]).

We assume that T^* is subject to both the biased-sampling scheme (2) and right-censoring. Let (2) be the density function of T, the biased-sampling variable, and C be the right-censored variable with survival function $S_C(\cdot)$. The observed data set $\{(X_i, \delta_i, Z_i), i = 1, \ldots, n\}$ is a sample of n independently and identically distributed (i.i.d.) realizations from the population (X, δ, Z), where $X = \min(T, C), \delta = I(T \le C)$, and $I(\cdot)$ is an indicator function. We assume throughout our analysis that C and Z are independent, and for any given Z, T and C are mutually independent. Hence the likelihood function of the observed data set is proportional to

$$\prod_{i=1}^{n} g_T(X_i|Z_i)^{\delta_i}\{1 - G_T(X_i|Z_i)\}^{1-\delta_i},$$

with $G_T(t|Z)$ being the corresponding cumulative distribution function of $g_T(t|Z)$. Next, we consider an unbiased estimating equations approach for estimating β and $H(\cdot)$ in model (3) based on the observed data set.

2.2 Estimating methodology

Our approach of estimation is in the spirits of [5] and [41]. We first discuss the uncensored case to facilitate understanding. Define $N^*(t) = I(T \le t)$ and $Y^*(t) = I(T \ge t)$. Motivated by [41], a mean zero process for the uncensored biased-sampling data is $M^*(t) = N^*(t) - \int_0^t Y^*(s)\frac{W(s,Z)}{W(T,Z)}d\Lambda_\epsilon(H(s) + Z^T\beta)$, obtained by taking expectation of the mean zero process for left-truncated data ([37]) conditional on the observations. For any nonnegative $W(t, Z)$ that satisfies $u(Z) = \int_0^\infty W(t, Z)f_{T^*}(t|Z)dt < \infty$, it can be easily verified that

$$\begin{aligned}
&E\{M^*(t)\}\\
=\;& E\Big\{N^*(t) - \int_0^t Y^*(s)\frac{W(s,Z)}{W(T,Z)}d\Lambda_\epsilon(H(s) + Z^T\beta)\Big\}\\
=\;& E_Z\Big\{E\Big[N^*(t) - \int_0^t Y^*(s)\frac{W(s,Z)}{W(T,Z)}\\
&\times d\Lambda_\epsilon(H(s) + Z^T\beta)\Big|Z\Big]\Big\}\\
=\;& E_Z\Big\{E[I(T \le t)|Z] - \int_0^t E\Big[\frac{Y^*(s)}{W(T,Z)}\Big|Z\Big]\\
&\times W(s,Z)d\Lambda_\epsilon(H(s) + Z^T\beta)\Big\}\\
=\;& E_Z\Big\{\int_0^t \frac{W(s,Z)f_{T^*}(s|Z)}{u(Z)}ds - \int_0^t\int_s^\infty \frac{W(u,Z)}{W(u,Z)}\\
&\times \frac{f_{T^*}(u|Z)}{u(Z)}du W(s,Z)d\Lambda_\epsilon(H(s) + Z^T\beta)\Big\}\\
=\;& E_Z\Big\{\int_0^t \frac{W(s,Z)f_{T^*}(s|Z)dt}{u(Z)} - \int_0^t \frac{W(s,Z)S_{T^*}(s|Z)}{u(Z)}\\
&\times d\Lambda_\epsilon(H(s) + Z^T\beta)\Big\} = 0,
\end{aligned}$$

where the notation $E_Z(\cdot)$ means that the expectation is taken with respect to Z. Similar to [5], the estimating equations for the uncensored biased-sampling data $\{(T_i, Z_i), i = 1, \ldots, n\}$ are:

$$(4)\quad \sum_{i=1}^{n}\Big\{dN_i^*(t) - Y_i^*(t)\frac{W(t,Z_i)}{W(T_i,Z_i)}d\Lambda_\epsilon(H(t) + Z_i^T\beta)\Big\} = 0$$

and

$$(5)\quad \sum_{i=1}^{n}\int_0^\infty Z_i\Big\{dN_i^*(t) - Y_i^*(t)\frac{W(t,Z_i)}{W(T_i,Z_i)}d\Lambda_\epsilon(H(t) + Z_i^T\beta)\Big\} = 0,$$

where $H(\cdot)$ is a nondecreasing function satisfying $H(0) = -\infty$, and $N_i^*(t)$ and $Y_i^*(t), i = 1, \ldots, n$, are the corresponding sample analogues of $N^*(t)$ and $Y^*(t)$, respectively. Note that equation (4) is a difference equation for identifying of the transformation function $H(\cdot)$ when β is fixed, whereas equation (5) is for the purpose of identifying β.

In the presence of right-censoring, let $N(t) = I(X \le t, \delta = 1)$ be the counting process for the failure, and $Y(t) = I(X \ge t)$ be the at-risk indicator. Under the conditional independent censoring assumption, i.e., T and C are independent given the covariate Z, the conditional joint probability of observing a failure at time t is

$$\begin{aligned}
P(X \in (t, t+dt), \delta = 1|Z) &= P(T \in (t, t+dt), T \le C|Z)\\
&= g_T(t|Z)S_C(t)dt.
\end{aligned}$$

By the mean zero property of $M^*(t)$, it makes strong sense to consider the process $M(t) = N(t) - \int_0^t \delta Y(s)\frac{W(s,Z)}{W(X,Z)}\frac{S_C(s)}{S_C(X)}d\Lambda_\epsilon(H(s) + Z^T\beta)$ for the right-censored case. Straightforward calculations yield

$$\begin{aligned}
&E[M(t)|Z]\\
=\;& E[N(t)|Z] - E\Big[\int_0^t \delta Y(s)\frac{W(s,Z)}{W(X,Z)}\frac{S_C(s)}{S_C(X)}\\
&\times d\Lambda_\epsilon(H(s) + Z^T\beta)\Big|Z\Big]\\
=\;& \int_0^t g_T(s|Z)S_C(s)ds - \int_0^t W(s,Z)S_C(s)\\
&\times E\Big[\frac{\delta Y(s)}{W(X,Z)S_C(X)}\Big|Z\Big]d\Lambda_\epsilon(H(s) + Z^T\beta)\\
=\;& \int_0^t g_T(s|Z)S_C(s)ds - \int_0^t W(s,Z)S_C(s)\\
&\times \int_s^\infty \frac{g_T(v|Z)S_C(v)}{W(v,Z)S_C(v)}dv d\Lambda_\epsilon(H(s) + Z^T\beta)\\
=\;& \int_0^t \frac{W(s,Z)f_{T^*}(s|Z)}{u(Z)}S_C(s)ds\\
&- \int_0^t \frac{W(s,Z)S_C(s)S_{T^*}(s|Z)}{u(Z)}d\Lambda_\epsilon(H(s) + Z^T\beta) = 0.
\end{aligned}$$

Thus, we have $E[M(t)] = E_Z\{E[M(t)|Z]\} = 0$.

When the survival function $S_C(\cdot)$ is known, the unbiased estimating equations for $H(\cdot)$ and β can be constructed similar to (4) and (5). In practice, however, $S_C(\cdot)$ is always unknown and a consistent estimator is needed to substitute for

$S_C(\cdot)$. Under the covariate-independent censoring assumption, a natural consistent estimator of $S_C(\cdot)$ is the following Kaplan-Meier estimator ([17]):

$$\hat{S}_C(t) = \prod_{s\leq t}\Big\{1 - \frac{\Delta N_C(s)}{\overline{Y}(s)}\Big\},$$

where $\Delta N_C(t) = N_C(t) - N_C(t-)$, $N_C(t) = \sum_{i=1}^n N_i^C(t)$, $N_i^C(t) = I(X_i \leq t, \delta_i = 0)$, $\overline{Y}(t) = \sum_{i=1}^n Y_i(t)$, and $Y_i(t) = I(X_i \geq t)$. Replacing $S_C(\cdot)$ by $\hat{S}_C(\cdot)$ in $M(t)$ results in the following asymptotically unbiased estimating equations for $H(\cdot)$ and β:

$$(6) \qquad \sum_{i=1}^n \Big\{dN_i(t) - \delta_i Y_i(t)\frac{W(t,Z_i)}{W(X_i,Z_i)}\frac{\hat{S}_C(t)}{\hat{S}_C(X_i)} \times d\Lambda_\epsilon(H(t) + Z_i^T\beta)\Big\} = 0$$

and

$$(7) \qquad U(\beta, H) = \sum_{i=1}^n \int_0^\tau Z_i\Big\{dN_i(t) - \delta_i Y_i(t)\frac{W(t,Z_i)}{W(X_i,Z_i)} \times \frac{\hat{S}_C(t)}{\hat{S}_C(X_i)} d\Lambda_\epsilon(H(t) + Z_i^T\beta)\Big\} = 0,$$

where $N_i(t) = I(X_i \leq t, \delta_i = 1), Y_i(t) = I(X_i \geq t)$ for $i = 1, \ldots, n$, and $H(\cdot)$ is also a nondecreasing function that satisfies $H(0) = -\infty$ and $\tau = \inf\{t : Pr(X > t) = 0\}$.

The assumption of covariate-independent censoring may be too rigid in practice and this constitutes the main disadvantage of our method. When this assumption does not hold, estimating equations similar to (6) and (7) may be constructed in the same manner but with $S_C(t)$ replaced by a consistent estimator under that corresponding case. Specifically, when Z is discrete and takes on only finite values, for any fixed value $Z = z$, we can estimate $S_C(\cdot|z)$ by the Kaplan-Meier estimator $\hat{S}_C(\cdot|z)$ based on the observations for which $Z = z$ in $\{(X_i, \delta_i, Z_i), i = 1, \ldots, n\}$ ([7]). On the other hand, if Z is continuous, we can relate C to Z by an auxiliary model, such as the PH, AFT or additive hazards models, and then obtain a consistent estimator of $S_C(\cdot|Z)$. The simulation results reported in Section 4 indicate that the performance of the proposed estimator constructed under the covariate-independent censoring assumption is quite robust when the dependence exists between C and Z. This indicates that our method can be applied to a wide range of real data.

2.3 Computational algorithm

Since the estimating equations under the uncensored and right-censored cases share a very similar structure and both can be solved by the same algorithm, we only provide the iterative steps for the right-censored case here.

Let $0 < t_1 < \cdots < t_K < \infty$ denote the observed K uncensored failure times within the data set $\{(X_i, \delta_i, Z_i), i = 1, \ldots, n\}$. For any fixed β, $H(\cdot)$ is uniquely determined by the estimating equation (6) as a nondecreasing step function with jumps occurring only at the uncensored failure times $t_j, j = 1, \ldots, K$. Thus, for all $t < t_1$, $H(t) \equiv -\infty$, and the estimation of $H(\cdot)$ is only required at the points $t_1, \ldots, t_K$.

Note that the estimating equation (6) can be equivalently expressed by the following K equations:

$$(8) \qquad \begin{cases} \sum_{i=1}^n \delta_i Y_i(t_1)\frac{W(t_1,Z_i)}{W(X_i,Z_i)}\frac{\hat{S}_C(t_1)}{\hat{S}_C(X_i)}\Lambda_\epsilon(H(t_1) \\ \qquad + Z_i^T\beta) = \sum_{i=1}^n dN_i(t_1), \\ \sum_{i=1}^n \delta_i Y_i(t_2)\frac{W(t_2,Z_i)}{W(X_i,Z_i)}\frac{\hat{S}_C(t_2)}{\hat{S}_C(X_i)}\lambda_\epsilon(H(t_2-) \\ \qquad + Z_i^T\beta)\Delta H(t_2) = \sum_{i=1}^n dN_i(t_2), \\ \qquad \vdots \\ \sum_{i=1}^n \delta_i Y_i(t_K)\frac{W(t_K,Z_i)}{W(X_i,Z_i)}\frac{\hat{S}_C(t_K)}{\hat{S}_C(X_i)}\lambda_\epsilon(H(t_K-) \\ \qquad + Z_i^T\beta)\Delta H(t_K) = \sum_{i=1}^n dN_i(t_K), \end{cases}$$

where $\Delta H(t) = H(t) - H(t-)$. Thus, we can compute the estimators of $(\beta, H(\cdot))$ via the following iterative steps:

Step 0: Choose an initial value for β and denote it as $\beta^{(0)}$, where $\beta^{(0)} = (\beta_1^{(0)}, \ldots, \beta_p^{(0)})$.

For the m-th iteration, the estimators are updated based on $(\beta^{(m-1)}, H^{(m-1)}(\cdot))$ as follows.

Step 1: The estimators of $H^{(m)}(t_j)$ $j = 1, \ldots, K$ are obtained by solving the equations in (8). Specifically, $H^{(m)}(t_1)$ is obtained by solving

$$\sum_{i=1}^n \delta_i Y_i(t_1)\frac{W(t_1,Z_i)}{W(X_i,Z_i)}\frac{\hat{S}_C(t_1)}{\hat{S}_C(X_i)}\Lambda_\epsilon(H(t_1) + Z_i^T\beta^{(m-1)}) = \sum_{i=1}^n dN_i(t_1).$$

Similarly, $H^{(m)}(t_j), j = 2, \ldots, K$, are obtained one by one by solving the remaining $K-1$ equations in (8). Denote the resultant estimators as $H^{(m)}(\cdot) = (H^{(m)}(t_1), \ldots, H^{(m)}(t_K))$.

Step 2: The estimator of β is obtained by solving

$$\sum_{i=1}^n \int_0^\tau Z_i\Big\{dN_i(t) - \delta_i Y_i(t)\frac{W(t,Z_i)}{W(X_i,Z_i)}\frac{\hat{S}_C(t)}{\hat{S}_C(X_i)} \times d\Lambda_\epsilon(H^{(m)}(t) + Z_i^T\beta)\Big\} = 0.$$

As $H^{(m)}(\cdot)$ is a step function with K positive jumps, solving the above equation is equivalent to finding a solution to

$$\sum_{i=1}^n\sum_{j=1}^K Z_i\Big\{\Delta N_i(t_j) - \delta_i Y_i(t_j)\frac{W(t_j,Z_i)}{W(X_i,Z_i)}\frac{\hat{S}_C(t_j)}{\hat{S}_C(X_i)} \times \lambda_\epsilon(H^{(m)}(t_j-) + Z_i^T\beta)\Delta H^{(m)}(t_j)\Big\} = 0,$$

where $\Delta N_i(t_j) = N_i(t_j) - N_i(t_j-), \Delta H^{(m)}(t_j) = H^{(m)}(t_j) - H^{(m)}(t_j-)$.

Repeat Steps 1 and 2 alternately until the estimators converge. Our convergence criterion is based on the following l_2-norm:

$$\Delta^{(m)} = \Big\{ \sum_{j=1}^{p} (\beta_j^{(m)} - \beta_j^{(m-1)})^2 + \sum_{j=1}^{K} (H^{(m)}(t_j) - H^{(m-1)}(t_j))^2 \Big\}^{\frac{1}{2}}.$$

The algorithm ends when $\Delta^{(m)}$ is less than a prescribed threshold value, say, 10^{-3}. Denote the final estimators as $(\hat{\beta}, \hat{H}(\cdot, \hat{\beta}))$.

3. ASYMPTOTIC PROPERTIES OF THE PROPOSED ESTIMATOR

We write the estimators as $(\hat{\beta}_n, \hat{H}_n(t, \hat{\beta}_n))$ to indicate their dependence on the sample size n, and denote the true values of β and $H(\cdot)$ as β_0 and $H_0(\cdot)$, respectively. For the purpose of simplifying the analysis, we define the following quantities for any s and $t \in (0, \tau]$:

$$
\begin{aligned}
B_1(t) &= E\Big\{\delta Y(t) \frac{W(t,Z)S_C(t)}{W(X,Z)S_C(X)} \dot{\lambda}_\epsilon(H_0(t) + Z^T\beta_0)\Big\},\\
B_2(t) &= E\Big\{\delta Y(t) \frac{W(t,Z)S_C(t)}{W(X,Z)S_C(X)} \lambda_\epsilon(H_0(t) + Z^T\beta_0)\Big\},\\
B(t,s) &= \exp\Big\{ \int_s^t \frac{B_1(u)}{B_2(u)} dH_0(u)\Big\},\\
B_1^Z(t) &= E\Big\{Z\delta Y(t) \frac{W(t,Z)S_C(t)}{W(X,Z)S_C(X)} \dot{\lambda}_\epsilon(H_0(t) + Z^T\beta_0)\Big\},\\
B_2^Z(t) &= E\Big\{Z\delta Y(t) \frac{W(t,Z)S_C(t)}{W(X,Z)S_C(X)} \lambda_\epsilon(H_0(t) + Z^T\beta_0)\Big\},\\
z(t) &= \frac{1}{B_2(t)}\Big\{B_2^Z(t) + \int_t^\tau \Big[B_1^Z(s) - \frac{B_2^Z(s)B_1(s)}{B_2(s)}\Big] \\
&\quad \times B(t,s) dH_0(s)\Big\},\\
D(t) &= E\Big\{ \int_0^\tau [Z - z(s)]\delta Y(s) \frac{W(s,Z)S_C(s)}{W(X,Z)S_C(X)} \{I(s \geq t) \\
&\quad - I(X \geq t)\} d\Lambda_\epsilon(H_0(s) + Z^T\beta_0)\Big\}/\pi(t),\\
\Sigma_* &= E\Big\{ \int_0^\tau \{Z - z(t)\} Z^T \delta Y(t) \frac{W(t,Z)S_C(t)}{W(X,Z)S_C(X)} \\
&\quad \times \dot{\lambda}_\epsilon(H_0(t) + Z^T\beta_0) dH_0(t)\Big\},\\
\Sigma^* &= E\Big\{ \int_0^\tau \{Z - z(t)\} dM(t) + \int_0^\tau D(t) dM^C(t)\Big\}^{\otimes 2},
\end{aligned}
$$

where $\dot{\lambda}_\epsilon(t)$ is the derivative of $\lambda_\epsilon(t)$ and $a^{\otimes 2} = aa^T$ for any vector a, $\pi(t) = P(X \geq t), M^C(t) = N^C(t) - \int_0^t Y(s) d\Lambda_C(s), N^C(t) = I(X \leq t, \delta = 0)$, and $\Lambda_C(\cdot)$ is the cumulative hazard function of the censoring variable C.

The asymptotic properties of the proposed estimator $\hat{\beta}_n$, including its consistency and asymptotic normality, are summarised in the following theorem, the proof of which is given in the Appendix.

Theorem 3.1. (Asymptotic Properties of $\hat{\beta}_n$) *Assume that conditions (C1)-(C6) in the Appendix are satisfied, then there exists a unique $\hat{\beta}_n$ within a neighborhood of β_0 for all large n. That is, as $n \to \infty$, we have*

$$\hat{\beta}_n \overset{\mathcal{P}}{\to} \beta_0,$$

where $\overset{\mathcal{P}}{\to}$ denotes convergence in probability. Also, as $n \to \infty$, we have

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \overset{\mathcal{D}}{\to} N(0, \Sigma),$$

where $\Sigma = \Sigma_*^{-1}\Sigma^*(\Sigma_*^{-1})^T$, and $\overset{\mathcal{D}}{\to}$ denotes convergence in distribution. In other words, $\hat{\beta}_n$ has an asymptotic normal distribution. In addition, Σ can be consistently estimated by a plug-in estimator $\hat{\Sigma} = \hat{\Sigma}_*^{-1}\hat{\Sigma}^*(\hat{\Sigma}_*^{-1})^T$, where

$$
\begin{aligned}
\hat{\Sigma}_* &= \frac{1}{n}\sum_{i=1}^n \Big\{ \int_0^\tau \{Z_i - \hat{z}(t)\} Z_i^T \delta_i Y_i(t) \frac{W(t,Z_i)}{W(X_i,Z_i)} \\
&\quad \times \frac{\hat{S}_C(t)}{\hat{S}_C(X_i)} \dot{\lambda}_\epsilon(\hat{H}_n(t,\hat{\beta}_n) + Z_i^T\hat{\beta}_n) d\hat{H}_n(t,\hat{\beta}_n)\Big\},\\
\hat{\Sigma}^* &= \frac{1}{n}\sum_{i=1}^n \Big\{ \int_0^\tau \{Z_i - \hat{z}(t)\} d\hat{M}_i(t) \\
&\quad + \int_0^\tau \hat{D}(t) d\hat{M}_i^C(t)\Big\}^{\otimes 2},\\
\hat{z}(t) &= \frac{1}{\hat{B}_2(t)}\Big\{\hat{B}_2^Z(t) + \int_t^\tau \Big[\hat{B}_1^Z(s) - \frac{\hat{B}_2^Z(s)\hat{B}_1(s)}{\hat{B}_2(s)}\Big] \\
&\quad \times \hat{B}(t,s) d\hat{H}_n(s,\hat{\beta}_n)\Big\},\\
\hat{D}(t) &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau [Z_i - \hat{z}(s)]\delta_i Y_i(s) \frac{W(s,Z_i)\hat{S}_C(s)}{W(X_i,Z_i)\hat{S}_C(X_i)} \\
&\quad \times \{I(s \geq t) - I(X_i \geq t)\} \\
&\quad \times d\Lambda_\epsilon(\hat{H}_n(s,\hat{\beta}_n) + Z_i^T\hat{\beta}_n)/\hat{\pi}(t),\\
\hat{M}_i(t) &= N_i(t) - \int_0^t \delta_i Y_i(s) \frac{W(s,Z_i)}{W(X_i,Z_i)} \frac{\hat{S}_C(s)}{\hat{S}_C(X_i)} \\
&\quad \times d\Lambda_\epsilon(\hat{H}_n(s,\hat{\beta}_n) + Z_i^T\hat{\beta}_n),\\
\hat{M}_i^C(t) &= N_i^C(t) - \int_0^t Y_i(s) d\hat{\Lambda}_C(s),\\
\hat{\pi}(t) &= \frac{1}{n}\overline{Y}(t), \quad \hat{\Lambda}_C(t) = \int_0^t \frac{dN_C(s)}{\overline{Y}(s)},\\
\hat{B}_1(t) &= \frac{1}{n}\sum_{i=1}^n \Big\{\delta_i Y_i(t) \frac{W(t,Z_i)\hat{S}_C(t)}{W(X_i,Z_i)\hat{S}_C(X_i)} \\
&\quad \times \dot{\lambda}_\epsilon(\hat{H}_n(t,\hat{\beta}_n) + Z_i^T\hat{\beta}_n)\Big\},
\end{aligned}
$$

$$\hat{B}_2(t) = \frac{1}{n}\sum_{i=1}^{n}\Big\{\delta_i Y_i(t)\frac{W(t,Z_i)\hat{S}_C(t)}{W(X_i,Z_i)\hat{S}_C(X_i)} \times\lambda_\epsilon(\hat{H}_n(t,\hat{\beta}_n)+Z_i^T\hat{\beta}_n)\Big\},$$

$$\hat{B}(t,s) = \exp\Big\{\int_s^t \frac{\hat{B}_1(u)}{\hat{B}_2(u)}d\hat{H}_n(u,\hat{\beta}_n)\Big\},$$

$$\hat{B}_1^Z(t) = \frac{1}{n}\sum_{i=1}^{n}\Big\{Z_i\delta_i Y_i(t)\frac{W(t,Z_i)\hat{S}_C(t)}{W(X_i,Z_i)\hat{S}_C(X_i)} \times\dot{\lambda}_\epsilon(\hat{H}_n(t,\hat{\beta}_n)+Z_i^T\hat{\beta}_n)\Big\},$$

$$\hat{B}_2^Z(t) = \frac{1}{n}\sum_{i=1}^{n}\Big\{Z_i\delta_i Y_i(t)\frac{W(t,Z_i)\hat{S}_C(t)}{W(X_i,Z_i)\hat{S}_C(X_i)} \times\lambda_\epsilon(\hat{H}_n(t,\hat{\beta}_n)+Z_i^T\hat{\beta}_n)\Big\}$$

for any $s,t\in[0,\tau]$.

In the Appendix, we also show the consistency of $\hat{H}_n(\cdot,\hat{\beta}_n)$ and the asymptotic normality of $\sqrt{n}\{\hat{H}_n(\cdot,\hat{\beta}_n)-H_0(\cdot)\}$. We do not state these results here as main theorems as we are primarily interested in the estimation of β and hence consider $H(\cdot)$ as a nuisance parameter only.

4. SIMULATION RESULTS

In this section, we conduct simulations to evaluate the performance of the proposed estimator in small sample, and draw comparisons with two other estimators, the estimator of [5] which accounts for right-censoring but ignores biased-sampling, and the combined estimating equation (CEE) estimator of [8] which assumes that the data are length-biased and right-censored. For the implementation of the iterative algorithm, we set the threshold value for convergence to 10^{-3}. It is found that our algorithm always converges with this threshold.

The unbiased data T^* are generated from the transformation model (3) with two independent covariates Z_1 and Z_2, generated from the $U(0,1)$ and $Bernoulli(0.5)$ distributions, respectively. We let the hazard function of ϵ be $\lambda_\epsilon(t)=\frac{\exp(t)}{1+r*\exp(t)}$ with $r=0,1,2$ ([11]). This hazard function leads to the PH and PO models when $r=0$ and $r=1$, respectively. We set the transformation function to $H(t)=\log(t)$ when $r=0$, and $H(t)=2r\log(t)$ when $r=1,2$. The true values β_0 are set to $(1,-1)^T$ or $(-1,1)^T$. For purposes of simplicity and to facilitate comparisons with the CEE estimator in [8], we set the biased weight function to $W(t,Z)\equiv t$. With this weight function, T, the biased data, can be generated by a similar procedure to that for generating length-biased data described in [38]. The only difference is that A, the backward recurrence time, is available for computing the CEE estimator in [8], but it is unavailable for the computation of the estimators in this paper and in [5]. In other words, the CEE estimator is able to utilise more information. We consider both the uncensored and right-censored cases. For the latter, we generate the censoring variable from the $EXP(c)$ distribution, where c is chosen in order for the censoring percentages (CR) to be 20% or 40%. We set the sample size n to 100 or 200, and the number of replications to 500.

The simulation results are reported in Tables 1-4 on pages 83-84. In the tables, our proposed estimator, and the estimators in [5] and [8] are labelled as "Proposed", "Chen" and "CEE" respectively. We evaluate the estimators' performance in terms of bias magnitude (BIAS), standard error (SE), estimated standard deviation (SD) and the coverage probability (CP) of the corresponding nominal 95% level confidence interval. The SDs of all three estimators are calculated via their corresponding plug-in estimators of the variances. In all cases, our proposed estimator results in biases of negligible magnitude, and are comparable to the biases of the CEE estimator, whereas the biases of Chen's estimator are relatively large. This result comes as no surprise since Chen's method does not take the biased-sampling aspect of the data into account. The SDs and SEs associated with our proposed estimator are very generally close, indicating that the plug-in estimator of the variance is consistent. Although in most cases, the SDs and SEs of the proposed estimator are larger than their CEE-based counterparts, this should not count against our method because the CEE estimator uses information from the backward recurrence time A, while our method does not utilise this information. Moreover, in many cases, the CEE-based SDs and SEs have a large discrepancy, indicating the likelihood of inconsistency of the plug-in estimator of the variance of the CEE estimator. This discrepancy between the SDs and SEs can sometimes result in a relatively large deviation in the CPs obtained by the CEE method from the nominal 0.95 level, whereas our estimator generally yields CPs that are very close to 0.95. In a large number of cases, Chen's method results in CPs that differ very considerably from 0.95. Finally, an increase in CR generally has the effect of worsening all estimators' performance, and an increase in sample size generally improves all estimators' performance, *ceteris paribus*.

As stated in Section 2.2, the proposed estimator is developed under the covariate-independent censoring assumption. This independence assumption may be a major drawback of our method. Here, we also examine the covariate-dependent censoring case to gain some insights of the extent to which the independence assumption impacts the results when this assumption does not hold. For this purpose, we generate the censoring variable based on $-Z_1-Z_2+EXP(c)$, and let all other parameters be the same as before. Table 5 on page 85 and Table 6 on page 85 report the results for the case of $\beta_0=(1,-1)^T$. The results are very comparable to those reported in Table 1 on page 83 and Table 2 on page 83, where covariate-independent censoring is assumed. In addition, we find that the simulation results

Table 1. Simulation results for $\beta_0 = (1, -1)^T$ with $n = 100$

			$r = 0$			$r = 1$			$r = 2$		
		CR(%)	Proposed	Chen	CEE	Proposed	Chen	CEE	Proposed	Chen	CEE
$\beta_{01} = 1$	Bias	0%	0.0260	0.4918	0.0254	-0.0143	-0.2799	-0.0817	0.0111	-0.3506	-0.0594
		20%	0.0165	0.5095	-0.0290	0.0035	-0.2322	-0.0660	0.0693	-0.3088	-0.0305
		40%	0.0064	0.5466	-0.0631	-0.0014	-0.1773	-0.0212	0.0534	-0.3369	-0.0850
	SE	0%	0.2995	0.4222	0.3030	0.7453	0.7314	0.7726	1.2343	1.0628	1.2111
		20%	0.3631	0.4600	0.3284	0.9821	0.7642	0.8886	1.5218	1.0481	1.3463
		40%	0.4545	0.4940	0.3765	1.1729	0.7539	0.8748	1.6919	1.0246	1.3194
	SD	0%	0.2748	0.3831	0.2862	0.8827	0.7156	1.0154	1.4296	1.0746	2.0848
		20%	0.3201	0.4230	0.3167	0.9896	0.7169	1.1074	1.5817	1.0525	2.1661
		40%	0.3864	0.4760	0.3654	1.1188	0.7430	1.1878	1.7875	1.0709	2.1182
	CP(95%)	0%	94.0	73.4	94.0	98.0	91.8	96.6	97.4	92.6	93.4
		20%	92.4	77.4	94.6	94.4	92.4	96.0	95.4	92.8	95.2
		40%	91.6	79.6	93.0	92.2	93.2	96.8	95.2	94.4	94.2
$\beta_{02} = -1$	Bias	0%	-0.0215	-0.5125	-0.0223	-0.0217	0.2474	0.0183	0.0245	0.3960	0.0985
		20%	-0.0446	-0.5483	-0.0022	0.0052	0.2239	0.0600	0.0125	0.3910	0.0923
		40%	-0.0390	-0.5608	0.0401	0.0035	0.2147	0.0507	-0.0438	0.3373	0.0407
	SE	0%	0.2089	0.2928	0.2128	0.4741	0.4422	0.5018	0.7024	0.6134	0.6948
		20%	0.2330	0.2972	0.2206	0.5635	0.4340	0.4929	0.8173	0.5592	0.7011
		40%	0.2866	0.3364	0.2550	0.6804	0.4526	0.5335	1.0295	0.6559	0.8146
	SD	0%	0.1923	0.2620	0.1986	0.4900	0.4276	0.3833	0.7790	0.6285	0.7782
		20%	0.2178	0.2798	0.2169	0.5488	0.4234	0.4165	0.8672	0.6144	0.7489
		40%	0.2474	0.3060	0.2391	0.6313	0.4384	0.4482	0.9803	0.6238	0.7355
	CP(95%)	0%	94.6	51.8	93.6	95.8	88.8	87.8	96.6	91.6	93.6
		20%	94.0	50.4	95.0	93.2	90.4	91.2	96.4	92.6	94.4
		40%	93.6	55.4	91.6	92.0	92.0	92.2	92.2	92.2	90.6

Table 2. Simulation results for $\beta_0 = (1, -1)^T$ with $n = 200$

			$r = 0$			$r = 1$			$r = 2$		
		CR(%)	Proposed	Chen	CEE	Proposed	Chen	CEE	Proposed	Chen	CEE
$\beta_{01} = 1$	Bias	0%	0.0100	0.4654	0.0112	-0.0391	-0.2951	-0.0594	0.0071	-0.3807	-0.0597
		20%	0.0007	0.4904	-0.0353	0.0118	-0.2380	0.0052	0.0556	-0.3500	-0.0044
		40%	-0.0043	0.5172	-0.0803	-0.0506	-0.2337	-0.0726	-0.0950	-0.4230	-0.0948
	SE	0%	0.1894	0.2680	0.1929	0.5301	0.4769	0.5540	0.8449	0.7477	0.8483
		20%	0.2293	0.2952	0.2171	0.7465	0.5310	0.6360	1.1096	0.7782	0.9629
		40%	0.3131	0.3715	0.2613	0.8519	0.5232	0.6600	1.2983	0.7714	1.0073
	SD	0%	0.1944	0.2684	0.1987	0.5937	0.5023	0.7218	0.9544	0.7540	1.1482
		20%	0.2278	0.2945	0.2191	0.6946	0.5029	0.7855	1.0993	0.7375	1.2753
		40%	0.2805	0.3335	0.2519	0.8159	0.5212	0.8359	1.2590	0.7490	1.4167
	CP(95%)	0%	94.8	60.0	95.2	96.8	92.0	96.4	98.0	92.4	95.6
		20%	95.0	63.2	94.8	93.0	91.8	96.6	95.2	92.2	95.6
		40%	93.0	66.4	91.6	92.2	92.4	96.8	94.4	89.0	93.8
$\beta_{02} = -1$	Bias	0%	-0.0159	-0.4936	-0.0172	-0.0033	0.2717	0.0305	-0.0516	0.3485	-0.0094
		20%	-0.0080	-0.5012	0.0273	0.0149	0.2666	0.0490	-0.0451	0.3295	0.0050
		40%	-0.0232	-0.5241	0.0640	-0.0059	0.2077	0.0187	0.0287	0.3298	-0.0039
	SE	0%	0.1385	0.1907	0.1426	0.3264	0.3041	0.3385	0.5120	0.4480	0.5192
		20%	0.1597	0.2140	0.1553	0.4142	0.3265	0.3728	0.5958	0.4233	0.5430
		40%	0.1970	0.2334	0.1745	0.4676	0.3221	0.3744	0.7545	0.4354	0.5812
	SD	0%	0.1372	0.1824	0.1400	0.3391	0.3012	0.2694	0.5251	0.4402	0.4039
		20%	0.1516	0.1944	0.1491	0.3934	0.2990	0.2907	0.6044	0.4297	0.4380
		40%	0.1762	0.2131	0.1643	0.4604	0.3089	0.3155	0.7013	0.4360	0.5120
	CP(95%)	0%	94.8	23.0	95.2	96.8	85.0	88.4	95.4	86.6	88.0
		20%	93.4	28.8	93.8	93.8	82.4	89.8	94.8	88.6	89.2
		40%	93.4	32.4	91.0	93.4	88.6	89.6	92.6	88.4	89.0

Table 3. Simulation results for $\beta_0 = (-1, 1)^T$ with $n = 100$

			$r=0$			$r=1$			$r=2$		
		CR(%)	Proposed	Chen	CEE	Proposed	Chen	CEE	Proposed	Chen	CEE
$\beta_{01} = -1$	Bias	0%	-0.0275	-0.4939	-0.0324	-0.0224	0.2024	0.0194	-0.0419	0.3028	0.0424
		20%	-0.0507	-0.5399	-0.0042	-0.0180	0.1951	0.0116	0.0415	0.3598	0.0713
		40%	0.0145	-0.4857	0.0932	0.0277	0.2428	0.0854	0.0950	0.3885	0.1030
	SE	0%	0.3001	0.4266	0.3049	0.7749	0.7169	0.8206	1.2827	1.1029	1.3048
		20%	0.3453	0.4474	0.3219	0.9113	0.7241	0.8197	1.5345	1.0354	1.2966
		40%	0.4444	0.5052	0.3807	1.0970	0.7502	0.9095	1.8106	1.0981	1.4334
	SD	0%	0.2751	0.3835	0.2850	0.8239	0.6953	0.7458	1.3289	1.0430	1.9591
		20%	0.3227	0.4212	0.3188	0.9266	0.6919	0.8299	1.4860	1.0221	1.7409
		40%	0.3838	0.4739	0.3616	1.0411	0.7199	0.8792	1.6547	1.0377	3.2623
	CP(95%)	0%	92.6	74.6	92.8	95.6	93.8	92.0	95.0	93.2	90.2
		20%	93.2	76.2	94.8	95.2	93.0	93.0	94.0	94.0	88.8
		40%	92.4	82.8	93.6	92.4	92.8	93.0	92.2	92.6	91.8
$\beta_{02} = 1$	Bias	0%	0.0321	0.5225	0.0341	-0.0243	-0.2554	-0.0815	0.0425	-0.3560	-0.0099
		20%	0.0148	0.5273	-0.0231	-0.0221	-0.2399	-0.0793	0.0124	-0.3560	-0.0446
		40%	0.0338	0.5665	-0.0501	-0.0306	-0.1813	-0.0688	0.0495	-0.2991	-0.0092
	SE	0%	0.2179	0.2996	0.2238	0.4706	0.4469	0.4656	0.7313	0.6248	0.7409
		20%	0.2121	0.2831	0.2132	0.6129	0.4349	0.5433	0.8799	0.5959	0.7413
		40%	0.2593	0.3164	0.2356	0.6800	0.4462	0.5194	1.0937	0.6711	0.8688
	SD	0%	0.1941	0.2621	0.1996	0.5276	0.4387	0.5812	0.8491	0.6489	1.1799
		20%	0.2128	0.2777	0.2122	0.5950	0.4370	0.6345	0.9363	0.6348	1.1019
		40%	0.2455	0.3057	0.2366	0.6615	0.4525	0.6641	1.0464	0.6471	1.6717
	CP(95%)	0%	93.0	50.4	92.8	96.8	89.2	96.6	97.4	91.6	95.6
		20%	95.2	55.2	94.2	93.8	91.8	95.0	93.8	92.8	95.0
		40%	93.2	56.4	94.4	94.6	92.4	97.2	92.6	90.6	95.2

Table 4. Simulation results for $\beta_0 = (-1, 1)^T$ with $n = 200$

			$r=0$			$r=1$			$r=2$		
		CR(%)	Proposed	Chen	CEE	Proposed	Chen	CEE	Proposed	Chen	CEE
$\beta_{01} = -1$	Bias	0%	-0.0103	-0.4712	-0.0104	-0.0260	0.2910	0.0143	-0.0352	0.3565	0.0189
		20%	-0.0059	-0.4885	0.0383	0.0203	0.2549	0.0315	0.0107	0.3907	0.0743
		40%	0.0177	-0.4910	0.1018	0.0089	0.2233	0.0389	0.0057	0.3160	-0.0452
	SE	0%	0.1988	0.2831	0.1998	0.5624	0.5003	0.5796	0.8276	0.7153	0.8369
		20%	0.2342	0.3022	0.2231	0.6731	0.4898	0.5980	1.0754	0.7207	0.9586
		40%	0.2984	0.3259	0.2382	0.8336	0.5183	0.6563	1.2460	0.7686	0.9733
	SD	0%	0.1938	0.2682	0.1981	0.5661	0.4954	0.5047	0.9001	0.7397	0.8071
		20%	0.2275	0.2954	0.2195	0.6588	0.4921	0.5395	1.0295	0.7232	0.8969
		40%	0.2794	0.3346	0.2504	0.7768	0.5107	0.6010	1.1947	0.7313	0.9636
	CP(95%)	0%	94.8	59.0	94.2	94.2	91.4	92.6	96.6	92.0	91.8
		20%	94.4	62.4	93.4	94.2	92.4	92.6	93.4	92.0	90.0
		40%	92.6	69.4	93.8	92.2	92.6	93.2	93.0	93.2	90.8
$\beta_{02} = 1$	Bias	0%	0.0096	0.4862	0.0083	-0.0083	-0.2859	-0.0420	0.0159	-0.3735	-0.0340
		20%	0.0149	0.5066	-0.0224	0.0135	-0.2460	-0.0328	0.0343	-0.3552	-0.0104
		40%	0.0085	0.5188	-0.0701	-0.0016	-0.2150	-0.0076	-0.0668	-0.3560	-0.0186
	SE	0%	0.1420	0.1991	0.1438	0.3284	0.3131	0.3375	0.4854	0.4126	0.4932
		20%	0.1664	0.2239	0.1656	0.3992	0.2924	0.3533	0.6130	0.4191	0.5439
		40%	0.1925	0.2347	0.1678	0.4890	0.3259	0.3823	0.7721	0.4416	0.5823
	SD	0%	0.1365	0.1818	0.1392	0.3587	0.3053	0.4158	0.5565	0.4473	0.6347
		20%	0.1520	0.1941	0.1493	0.4122	0.3040	0.4397	0.6378	0.4382	0.6874
		40%	0.1773	0.2129	0.1649	0.4831	0.3140	0.4807	0.7428	0.4441	0.7438
	CP(95%)	0%	94.8	25.6	95.6	97.0	84.0	97.6	97.2	88.4	97.0
		20%	93.6	28.8	91.8	95.8	87.2	97.4	95.6	86.0	97.8
		40%	94.2	31.0	91.6	93.2	89.2	98.4	92.8	86.2	97.4

Table 5. Simulation results for $\beta_0 = (1, -1)^T$ under covariate-dependent censoring with $n = 100$

			$r=0$			$r=1$			$r=2$		
		CR(%)	Proposed	Chen	CEE	Proposed	Chen	CEE	Proposed	Chen	CEE
$\beta_{01}=1$	Bias	20%	0.0249	0.5017	-0.0135	0.0384	-0.2385	-0.0325	0.0268	-0.3091	-0.0548
		40%	0.0229	0.5411	-0.0358	0.0454	-0.1943	-0.0357	-0.0336	-0.3931	-0.0987
	SE	20%	0.3480	0.4561	0.3270	0.9405	0.7399	0.9070	1.4915	1.1529	1.3045
		40%	0.4242	0.5170	0.3846	1.1235	0.7775	0.9404	1.7417	1.1668	1.3683
	SD	20%	0.3174	0.4245	0.3191	0.9868	0.7287	1.1088	1.6232	1.0837	2.1373
		40%	0.3756	0.4877	0.3660	1.1403	0.7807	1.2506	1.8498	1.1476	2.7439
	CP(95%)	20%	94.6	76.2	95.2	95.8	94.0	95.6	96.4	91.8	92.4
		40%	91.4	79.4	93.4	94.0	94.4	95.4	95.6	93.4	95.4
$\beta_{02}=-1$	Bias	20%	-0.0212	-0.5150	0.0132	0.0267	0.2433	0.0940	-0.0245	0.3520	0.0584
		40%	-0.0281	-0.5361	0.0414	0.0308	0.2022	0.0588	0.0210	0.3332	0.0893
	SE	20%	0.2311	0.3115	0.2285	0.5812	0.4636	0.5342	0.8652	0.6165	0.7732
		40%	0.2820	0.3514	0.2623	0.6533	0.4639	0.5370	0.9825	0.6519	0.8019
	SD	20%	0.2140	0.2807	0.2157	0.5596	0.4312	0.4290	0.8682	0.6302	0.7541
		40%	0.2429	0.3075	0.2407	0.6344	0.4541	0.4683	0.9790	0.6595	0.9256
	CP(95%)	20%	93.8	56.2	92.0	93.4	89.2	91.0	95.0	92.6	91.4
		40%	91.4	62.2	92.2	93.8	90.8	92.2	94.8	91.4	92.2

Table 6. Simulation results for $\beta_0 = (1, -1)^T$ under covariate-dependent censoring with $n = 200$

			$r=0$			$r=1$			$r=2$		
		CR(%)	Proposed	Chen	CEE	Proposed	Chen	CEE	Proposed	Chen	CEE
$\beta_{01}=1$	Bias	20%	0.0207	0.5016	-0.0130	0.0274	-0.2504	0.0017	0.0039	-0.3701	-0.0363
		40%	-0.0100	0.4992	-0.0669	-0.0757	-0.2499	-0.0752	-0.0031	-0.3124	-0.0040
	SE	20%	0.2297	0.3026	0.2205	0.6730	0.5190	0.6056	1.0293	0.7340	0.8981
		40%	0.2942	0.3480	0.2466	0.8573	0.5745	0.7023	1.2795	0.7856	0.9957
	SD	20%	0.2251	0.2971	0.2215	0.6943	0.5136	0.7872	1.0925	0.7579	1.2653
		40%	0.2711	0.3386	0.2533	0.8273	0.5474	0.8620	1.2755	0.7989	1.4317
	CP(95%)	20%	93.6	63.2	94.2	95.0	91.4	97.6	96.0	92.4	96.2
		40%	93.2	70.8	94.6	92.4	91.4	95.2	95.0	94.4	96.4
$\beta_{02}=-1$	Bias	20%	0.0037	-0.4727	0.0313	-0.0314	0.2482	0.0029	0.0192	0.3819	0.0320
		40%	-0.0076	-0.5048	0.0515	-0.0235	0.2284	0.0084	0.0075	0.3447	0.0188
	SE	20%	0.1502	0.2006	0.1463	0.3941	0.3185	0.3661	0.6344	0.4669	0.5724
		40%	0.1863	0.2346	0.1697	0.4739	0.3303	0.4086	0.7364	0.4836	0.5949
	SD	20%	0.1508	0.1941	0.1501	0.3886	0.3035	0.2864	0.6040	0.4405	0.4453
		40%	0.1738	0.2149	0.1673	0.4585	0.3191	0.3227	0.6945	0.4598	0.5019
	CP(95%)	20%	94.2	33.6	95.2	93.2	85.8	87.4	93.2	85.0	89.2
		40%	94.8	37.2	91.6	93.4	90.2	90.6	92.4	87.8	92.2

of the proposed estimator are comparable to those of the CEE estimator, which requires no assumption of covariate-independent censoring. These results may be taken as an indication that our proposed method is robust with respect to dependence of C on Z. Similar results are observed for the case of $\beta_0 = (-1, 1)^T$, and they are omitted here to conserve space.

To sum up, the simulation results demonstrate that the proposed estimator performs well under general biased-sampling and right-censored situations, and is robust with respect to the censoring mechanism. The proposed iterative algorithm also works effectively.

5. REAL DATA EXAMPLES

5.1 Application to the Stanford heart transplant data

This real data example uses the Stanford heart transplant data given in [10] and subsequently updated by [29]. The original data set contains information on the ages, pre-transplant waiting time, censoring indices, transplant status, observed survival time, and three mismatch scores of 103 potential heart transplant patients who participated in the Stanford heart transplant program between October 1967 and April 1974. The updated data set, constructed by

Table 7. Estimation Results for the Stanford heart transplant data

		r=0	r=1	r=2
	Est	-0.1368	-0.2533	-0.4124
β_1	SD	0.0535	0.0839	0.1158
	CI(95%)	(-0.2417, -0.0319)	(-0.4177, -0.0889)	(-0.6394, -0.1853)
	Est	0.0019	0.0035	0.0057
β_2	SD	0.0007	0.0011	0.0018
	CI(95%)	(0.0005, 0.0033)	(0.0014, 0.0056)	(0.0021, 0.0093)

Note: β_1 and β_2 are the regression coefficients for "age" and "age^2", respectively.

[29], contains observations up to February 1980 for a total of 184 patients who have received heart transplants. It provides information on the survival time, censoring indicators in February 1980, ages at the first transplant and the T5 mismatch scores for these patients. However, the pre-transplant waiting time is unreported in the new data set.

[29] utilized four regression techniques for right-censored data to analyse their data. As the pre-transplant waiting time is missing, the updated data can be treated as the biased-sampling and right-censored data, with the distribution function of the pre-transplant waiting time used as the weight function ([41]). Based on the original 103 observations, [41] estimated the weight function as $W(t, Z) = 1 - \exp(-0.027t^{0.925})$. We use the same weight function here and assume that it is known *a priori*. Following [29, 41], we remove 27 of the 184 observations for which the T5 mismatch scores are missing and 5 observations for which the survival time is less than 10 days from the data set. This results in a sample of 152 observations with a 36% right-censoring rate.

Following [41], we consider two covariates "age" and "age^2" in the regression. We select the same parameter r as in Section 4. The estimated coefficients (EST), estimated standard deviations (SD) and estimated 95% confidence intervals (CI) for the regression coefficients are reported in Table 7 on page 86.

The results based on our proposed method for $r = 0$ are very similar to those reported by [29, 41] under the PH model. The covariates "age" and "age^2" are significantly related to the survival time as the 95% confidence intervals of both β_1 and β_2 do not contain 0. The results for $r = 1$ and $r = 2$ are also very similar to those obtained for $r = 0$, indicating no major difference in results based on the three special cases of the SLT model.

5.2 Application to the dementia data

Our second real data example uses the dementia data from the Canadian Study of Health and Aging that comprise observations of 1132 patients diagnosed with dementia in 1991 and subsequently followed until 1996. The same data were used by [8, 23, 38] in their studies. The data provide information on the dementia types: possible Alzheimer's, probable Alzheimer's or vascular dementia. The data also contain the date of onset, date of screening, date of death or censoring, and the censoring indicators of the patients. We remove observations with missing date of onset or missing dementia type, leaving 818 observations in the data set for analysis. These include 393 observations of patients with probable Alzheimer's, 252 with possible Alzheimer's, and 173 with vascular dementia. The right-censoring rate for this sample is 22%.

Our primary interest here is to study the survival difference among the three disease types. An application of the stationarity test of [1] to the observations confirms that the data satisfy the stationarity assumption. We thus treat the data as length-biased and set the weight function to $W(t, Z) = t$. Similar to [38], we find no evidence of dependence between the censoring variable and the dementia type. We take probable Alzheimer's as the reference type and represent possible Alzheimer's and vascular dementia by two dummy variables. The parameter r are selected as in Section 4. The estimation results of our proposed estimator and the CEE estimator in [8] are reported in Table 8 on page 87.

From Table 8, we find that the estimated coefficients by the proposed method are negative for β_1 and positive for β_2 under all three choices of r, implying that relative to probable Alzheimer's, possible Alzheimer's and vascular dementia are associated with lower and higher risk of death, respectively. However, in all cases the 95% confidence intervals of the coefficients contain 0. The survival difference between probable Alzheimer's and possible Alzheimer's, and between probable Alzheimer's and vascular dementia, are therefore not statistically significant. The same conclusions may be drawn from the results based on the CEE estimator also shown in Table 8. Our findings concur with the conclusions reached by [38] even though our analysis ignores informative censoring, which is usually considered as a salient feature of the data in other studies. This demonstrates the robustness of our approach.

6. CONCLUDING REMARKS

In this paper, we have proposed an estimation method in conjunction with the SLT model when the data are subject to general biased-sampling and right-censoring. We show that the proposed estimator possesses good large sample

Table 8. *Estimation Results for the dementia data*

		Proposed			CEE		
		r=0	r=1	r=2	r=0	r=1	r=2
	Est	-11.31	-31.67	-59.17	-9.95	-28.07	-52.49
β_1	SD	8.28	23.37	43.41	6.69	17.52	30.76
	CI(95%)	(-27.55, 4.93)	(-77.48, 14.14)	(-144.25, 25.90)	(-23.06,3.17)	(-62.41,6.27)	(-112.77,7.79)
	Est	6.45	18.49	35.04	7.68	21.50	41.38
β_2	SD	10.53	29.57	57.70	8.25	23.02	47.26
	CI(95%)	(-14.18, 27.09)	(-39.46, 76.44)	(-78.07, 148.14)	(-8.49, 23.84)	(-23.62, 66.62)	(-51.25,134.01)

Note: β_1 and β_2 are the regression coefficients for possible Alzheimer's and vascular dementia, respectively. All the quantities are multiplied by 10^2.

asymptotic properties, can be easily implemented by an iterative algorithm, and fares well in terms of finite sample efficiency compared to the method in [5] that neglects biased-sampling, and the CEE method in [8] which assumes that the data are length-biased. Our proposed estimator also demonstrates robustness when the assumption of covariate-independent censoring is violated or informative censoring is ignored when the data are length-biased. Work in progress by the authors considers the broader class of transformation models ([45]) to incorporate the time-dependent covariates and the varying coefficient transformation models ([6]) for more general cases.

7. APPENDIX

This appendix provides the proofs of the main results given in Section 3. Let $\|a\|$ denote the Euclidean norm for a vector a and $\|f\|$ the supremum norm for a function f, that is, $\|f\| = \sup_{t\in[0,\tau]} |f(t)|$. Our proofs of results require the following conditions:

(C1) The true parameter β_0 belongs to the interior of the parameter space $\mathcal{B}$, where $\mathcal{B}$ is a compact set in $\mathbb{R}^p$.

(C2) The covariate Z is a $p\times 1$ dimensional bounded vector not contained in a $(p-1)$ dimensional hyperplane.

(C3) τ is finite with $P(T > \tau) > 0$ and $P(C > \tau) > 0$. The survival function $S_C(\cdot)$ of the censoring variable C is continuous, satisfying $P(C > T) > 0$.

(C4) $\lambda_\epsilon(t)$ is positive and bounded and $\dot\lambda_\epsilon(t)$ is bounded and continuously differentiable on $(-\infty, m)$ for any finite constant m; as well, $\lim_{t\to-\infty} \lambda_\epsilon(t) = 0$.

(C5) $H_0(t)$ has a continuous and positive derivative $\dot H_0(t)$ on $[0, \tau]$.

(C6) Σ_* and Σ^* are nonsingular matrices.

Condition (C1) that requires the compactness of the finite dimensional parameter space $\mathcal{B}$ is conventional in the literature. Condition (C2) is a mild condition for the boundedness of the covariate Z and the nonsingularity of ZZ^T. Condition (C3) suggests the right-censoring can occur after failure times in order to avoid the complex technical issues concerning tail behaviour. Condition (C4) is usually satisfied for distributions of ϵ commonly assumed for transformation models, and (C5) is a general assumption for transformation models. Condition (C6) restricts Σ^* and Σ_* to be nonsingular, which is a necessary condition for establishing the asymptotic normality of $\hat\beta_n$ and usually holds for specific families that are commonly used for the transformation models; e.g., the PH and PO models ([18]).

Proof of Theorem 1. Our approach is similar to the approaches taken by [5, 18, 25]. Here, we only present the main results and divide them into three parts.

Part 1: Proof of the consistency of $\hat\beta_n$ and $\hat H_n(t, \hat\beta_n)$.

For any fixed β, let $\hat H_n(t, \beta)$ be the function representing the unique solution of the estimating equation (6). Write $\hat H_{n0}(t) = \hat H_n(t, \beta_0)$. We first show that $\hat H_{n0}(t)$ converges to $H_0(t)$, namely, as $n \to \infty, \sup\{|\hat H_{n0}(t) - H_0(t)| : t \in (0, \tau]\} \to 0$ in probability. Due to the monotonicity of $\hat H_{n0}(t)$ on $t \in (0, \tau]$, it suffices to show that the limiting function of $\{\hat H_{n0}(t)\}$ is unique and equal to $H_0(t)$. Let $\tilde H(t)$ be any limit of the sequence $\{\hat H_{n0}(t)\}$, where $\tilde H(t)$ is a function on $t \in (0, \tau]$. We only have to show $\tilde H(t) = H_0(t)$.

Now, consider the estimating equation (6). It follows from the law of large numbers and the uniform consistency of the Kaplan-Meier estimator $\hat S_C(t)$ ([13], Theorem 3.4.2, p. 115) that

$$E[N(t)] = \int_0^t E\Big\{\delta Y(s)\frac{W(s,Z)S_C(s)}{W(X,Z)S_C(X)} \times\lambda_\epsilon(\tilde H(s) + Z^T\beta_0)\Big\}d\tilde H(s),$$

which implies that $\tilde H(t)$ is differentiable and hence must satisfy

$$\frac{d\tilde H(t)}{dt} = \frac{dE[N(t)]}{dt}\Big(E\Big\{\delta Y(t)\frac{W(t,Z)S_C(t)}{W(X,Z)S_C(X)} \times\lambda_\epsilon(\tilde H(t) + Z^T\beta_0)\Big\}\Big)^{-1}. \quad (9)$$

On one hand, (9) is a Cauchy problem which always results in a unique solution under local smoothness assumptions ([36], Theorem 3.4.2, p. 40). On the other hand, $H_0(t)$ satisfies (9). Thus, we have $\tilde H(t) = H_0(t)$ and the sequence $\{\hat H_{n0}(t)\}$ converges to $H_0(t)$ on $t \in (0, \tau]$.

For any t in a compact subset of the interior of the support of X, we will show in Part 2 that the derivative of $\hat{H}_n(t,\beta)$ with respect to β is bounded in the neighborhood of β_0. Thus, $\hat{H}_n(t,\beta_n) \to \hat{H}_{n0}(t)$ provided that $\beta_n \to \beta_0$. As $\hat{H}_{n0}(t) \to H_0(t)$, we have $\hat{H}_n(t,\hat{\beta}_n) \to H_0(t)$ provided that $\hat{\beta}_n$ is a consistent estimator of β_0.

We now prove the consistency of $\hat{\beta}_n$. Using the uniform law of large numbers ([33]) and the uniform consistency of $\hat{S}_C(t)$, we can show that for β in a neighborhood of β_0, $n^{-1}U(\beta,\hat{H}_n(\cdot,\beta))$ converges uniformly to a non-random function $u(\beta)$, and $n^{-1}\frac{\partial U(\beta,\hat{H}_n(\cdot,\beta))}{\partial\beta}$ converges uniformly to $\dot{u}(\beta)$, the first derivative of $u(\beta)$. We will show in Part 2 that $\dot{u}(\beta_0) = -\Sigma_*$. As we assume $-\Sigma_*$ to be nonsingular, $n^{-1}\frac{\partial U(\beta,\hat{H}_n(\cdot,\beta))}{\partial\beta}$ is nonsingular in a neighborhood of β_0. Thus there exists boundary values r and R that are bounded away from 0 and ∞ in probability such that for any β_1 and β_2 in a neighborhood of β_0, we have

$$\begin{aligned} r\|\beta_1-\beta_2\| &\le \|U(\beta_1,\hat{H}_n(\cdot,\beta_1)) - U(\beta_2,\hat{H}_n(\cdot,\beta_2))\| \\ &\le R\|\beta_1-\beta_2\|. \end{aligned}$$

Combining condition (C1) and the fact that $U(\hat{\beta}_n,\hat{H}_n(\cdot,\hat{\beta}_n)) = 0$ and $U(\beta_0,\hat{H}_{n0}(\cdot)) \to u(\beta_0) = 0$ as $n\to\infty$, it follows that there is a neighborhood of β_0 such that $\hat{\beta}_n$ exists and is unique. Therefore, $\hat{\beta}_n \to \beta_0$ in probability, and thus we have $\hat{H}_n(t,\hat{\beta}_n) \to H_0(t)$.

Part 2: Proof of $\frac{1}{n}\frac{\partial U(\beta,\hat{H}_n(\cdot,\beta))}{\partial\beta}$ converging to $-\Sigma_*$ at $\beta=\beta_0$ in probability.

For any $t>0$ and $x\in(-\infty,\infty)$, let us define

$$\lambda^*\{H_0(t)\} = B(t,a), \quad \Lambda^*(x) = \int_b^x \lambda^*(s)ds,$$

where $a>0$ and b are finite fixed numbers chosen to guarantee the finiteness of the above integrals. It is clear that $B(t,s) = \frac{\lambda^*\{H_0(t)\}}{\lambda^*\{H_0(s)\}}$ and $d\lambda^*\{H_0(t)\} = \lambda^*\{H_0(t)\}\frac{B_1(t)}{B_2(t)}dH_0(t)$. For simplicity, denote

$$A_i^\circ(t) = \delta_i Y_i(t)\frac{W(t,Z_i)S_C(t)}{W(X_i,Z_i)S_C(X_i)}$$

and

$$A_i(t) = \delta_i Y_i(t)\frac{W(t,Z_i)\hat{S}_C(t)}{W(X_i,Z_i)\hat{S}_C(X_i)}, \quad i=1,\dots,n.$$

Mimicking step 2 of [18] (p. 225) and recognizing that $\hat{H}_n(t,\beta_0)$ converges to $H_0(t)$ and condition (C4), we obtain

$$\frac{\partial\hat{H}_n(t,\beta)}{\partial\beta}\Big|_{\beta=\beta_0} = -\int_0^t \frac{B(s,t)}{B_2(s)}B_1^Z(s)dH_0(s) + o_p(1)$$

and

$$\begin{aligned} d\frac{\partial\hat{H}_n(t,\beta)}{\partial\beta}\Big|_{\beta=\beta_0} &= -B_2^{-1}(t)\Big[B_1^Z(t) + B_1(t) \\ &\qquad \times\frac{\partial\hat{H}_n(t,\beta)}{\partial\beta}\Big|_{\beta=\beta_0} + o_p(1)\Big]dH_0(t). \end{aligned}$$

Based on these results, we have

$$\begin{aligned} &\frac{1}{n}\frac{\partial U(\beta,\hat{H}_n(\cdot,\beta))}{\partial\beta}\Big|_{\beta=\beta_0} \\ &= -\frac{1}{n}\sum_{i=1}^n\int_0^\tau Z_iA_i(t)\lambda_\epsilon(\hat{H}_n(t,\beta_0)+Z_i^T\beta_0) \\ &\quad \times d\frac{\partial\hat{H}_n(t,\beta)}{\partial\beta}\Big|_{\beta=\beta_0} - \frac{1}{n}\sum_{i=1}^n\int_0^\tau Z_iA_i(t)\Big(\frac{\partial\hat{H}_n(t,\beta)}{\partial\beta}\Big|_{\beta=\beta_0} \\ &\quad +Z_i\Big)^T\dot{\lambda}_\epsilon(\hat{H}_n(t,\beta_0)+Z_i^T\beta_0)d\hat{H}_n(t,\beta_0) \\ &= -\frac{1}{n}\sum_{i=1}^n\int_0^\tau\Big[Z_i - \frac{B_2^Z(t)}{B_2(t)}\Big]Z_i^TA_i^\circ(t)\dot{\lambda}_\epsilon(H_0(t)+Z_i^T\beta_0) \\ &\quad \times dH_0(t) + \int_0^\tau\Big[B_1^Z(t) - \frac{B_1(t)B_2^Z(t)}{B_2(t)}\Big] \\ &\quad \times\int_0^t\frac{B(s,t)}{B_2(s)}B_1^Z(s)dH_0(s)dH_0(t) + o_p(1) \\ &= -\frac{1}{n}\sum_{i=1}^n\int_0^\tau[Z_i - z(t)]Z_i^TA_i^\circ(t)\dot{\lambda}_\epsilon(H_0(t)+Z_i^T\beta_0)dH_0(t) \\ &\quad +o_p(1), \end{aligned}$$

which yields the desired result that $\frac{1}{n}\frac{\partial U(\beta,\hat{H}_n(\cdot,\beta))}{\partial\beta}\Big|_{\beta=\beta_0}$ converges to $-\Sigma_*$ in probability.

Part 3: Proof of the asymptotic normality of $\hat{\beta}_n$ and $\hat{H}_n(t,\hat{\beta}_n)$.

Combining the definition of the process $M(t)$ and estimating equation (6), and following step 2 of [5] (p. 666), we obtain

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n dM_i(t) &= \frac{1}{n}\sum_{i=1}^n[A_i(t)d\Lambda_\epsilon(\hat{H}_n(t,\beta_0)+Z_i^T\beta_0) \\ &\qquad -A_i^\circ(t)d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)] \\ &= \frac{1}{n}\sum_{i=1}^n A_i^\circ(t)[d\Lambda_\epsilon(\hat{H}_n(t,\beta_0)+Z_i^T\beta_0) \\ &\qquad -d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)] + \frac{1}{n}\sum_{i=1}^n[A_i(t) \\ &\qquad -A_i^\circ(t)]d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0) + o_p(n^{-\frac{1}{2}}) \\ &= \frac{1}{n}\sum_{i=1}^n A_i^\circ(t)d\Big\{\frac{\lambda_\epsilon(H_0(t)+Z_i^T\beta_0)}{\lambda^*\{H_0(t)\}} \\ &\qquad \times(\Lambda^*(\hat{H}_n(t,\beta_0)) - \Lambda^*(H_0(t)))\Big\} \end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{n}\sum_{i=1}^{n}[A_i(t)-A_i^\circ(t)]d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)\\
&+o_p(n^{-\frac{1}{2}})\\
=\;& \frac{B_2(t)}{\lambda^*\{H_0(t)\}}d(\Lambda^*(\hat{H}_n(t,\beta_0))-\Lambda^*(H_0(t)))\\
&+(\Lambda^*(\hat{H}_n(t,\beta_0))-\Lambda^*(H_0(t)))\\
&\times\Big\{\frac{B_1(t)-B_2(t)\frac{B_1(t)}{B_2(t)}}{\lambda^*\{H_0(t)\}}\Big\}dH_0(t)\\
&+\frac{1}{n}\sum_{i=1}^{n}[A_i(t)-A_i^\circ(t)]d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)\\
&+o_p(n^{-\frac{1}{2}})\\
=\;& \frac{B_2(t)}{\lambda^*\{H_0(t)\}}d(\Lambda^*(\hat{H}_n(t,\beta_0))-\Lambda^*(H_0(t)))\\
&+\frac{1}{n}\sum_{i=1}^{n}[A_i(t)-A_i^\circ(t)]d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)\\
&+o_p(n^{-\frac{1}{2}}).
\end{aligned}
$$

This leads to

$$
\begin{aligned}
(10)\;&\Lambda^*(\hat{H}_n(t,\beta_0))-\Lambda^*(H_0(t))\\
=\;& \frac{1}{n}\sum_{i=1}^{n}\int_0^t\frac{\lambda^*\{H_0(s)\}}{B_2(s)}[A_i^\circ(s)-A_i(s)]d\Lambda_\epsilon(H_0(s)+Z_i^T\beta_0)\\
&+\frac{1}{n}\sum_{i=1}^{n}\int_0^t\frac{\lambda^*\{H_0(s)\}}{B_2(s)}dM_i(s)+o_p(n^{-\frac{1}{2}}).
\end{aligned}
$$

Now, because

$$
\begin{aligned}
&n^{-\frac{1}{2}}U(\beta_0,\hat{H}_n(\cdot,\beta_0))\\
=\;& n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i[dN_i(t)-A_i(t)d\Lambda_\epsilon(\hat{H}_n(t,\beta_0)+Z_i^T\beta_0)]\\
=\;& n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i[dN_i(t)-A_i^\circ(t)d\Lambda_\epsilon(\hat{H}_n(t,\beta_0)+Z_i^T\beta_0)]\\
&+n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i[A_i^\circ(t)-A_i(t)]d\Lambda_\epsilon(\hat{H}_n(t,\beta_0)+Z_i^T\beta_0)\\
=:\;& I_1+I_2,
\end{aligned}
$$

where

$$
\begin{aligned}
I_1 \;=\;& n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i[dN_i(t)-A_i^\circ(t)\\
&\times d\Lambda_\epsilon(\hat{H}_n(t,\beta_0)+Z_i^T\beta_0)]\\
=\;& n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i dM_i(t)-n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_iA_i^\circ(t)\\
&\times d[\Lambda_\epsilon(\hat{H}_n(t,\beta_0)+Z_i^T\beta_0)-\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)]\\
=\;& -n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_iA_i^\circ(t)d\Big\{\frac{\lambda_\epsilon(H_0(t)+Z_i^T\beta_0)}{\lambda^*\{H_0(t)\}}\\
&\times[\Lambda^*(\hat{H}_n(t,\beta_0))-\Lambda^*(H_0(t))]\Big\}\\
&+n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i dM_i(t)+o_p(1)\\
=\;& -n^{-\frac{1}{2}}\int_0^\tau\frac{B_2^Z(t)}{\lambda^*\{H_0(t)\}}\Big\{\sum_{i=1}^{n}\frac{\lambda^*\{H_0(t)\}}{B_2(t)}[dM_i(t)\\
&+(A_i^\circ(t)-A_i(t))d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)]\Big\}\\
&-n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau\frac{\lambda^*\{H_0(t)\}}{B_2(t)}\int_t^\tau\frac{B_1^Z(s)-B_2^Z(s)\frac{B_1(s)}{B_2(s)}}{\lambda^*\{H_0(s)\}}\\
&\times dH_0(s)[dM_i(t)+(A_i^\circ(t)-A_i(t))\\
&\times d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)]+n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i dM_i(t)\\
&+o_p(1)\\
=\;& -n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau\frac{B_2^Z(t)}{B_2(t)}dM_i(t)-n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau\frac{B_2^Z(t)}{B_2(t)}\\
&[A_i^\circ(t)-A_i(t)]d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)\\
&-n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau\frac{1}{B_2(t)}\int_t^\tau(B_1^Z(s)-\frac{B_2^Z(s)B_1(s)}{B_2(s)})\\
&\times B(t,s)dH_0(s)dM_i(t)-n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau\frac{A_i^\circ(t)-A_i(t)}{B_2(t)}\\
&\times\int_t^\tau(B_1^Z(s)-\frac{B_2^Z(s)B_1(s)}{B_2(s)})B(t,s)dH_0(s)\\
&\times d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)+n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i dM_i(t)\\
&+o_p(1)\\
=\;& n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau[Z_i-z(t)]dM_i(t)-n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau[A_i^\circ(t)\\
&-A_i(t)]z(t)d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)+o_p(1),
\end{aligned}
$$

and

$$
\begin{aligned}
I_2 \;=\;& n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i[A_i^\circ(t)-A_i(t)]d\Lambda_\epsilon(\hat{H}_n(t,\beta_0)+Z_i^T\beta_0)\\
=\;& n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i[A_i^\circ(t)-A_i(t)]d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)\\
&+o_p(1),
\end{aligned}
$$

we have

$$
\begin{aligned}
&n^{-\frac{1}{2}}U(\beta_0,\hat{H}_n(\cdot,\beta_0))\\
=\;& n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau[Z_i-z(t)]dM_i(t)-n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau[A_i^\circ(t)\\
&-A_i(t)]z(t)d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)+n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau Z_i
\end{aligned}
$$

$$\times[A_i^\circ(t)-A_i(t)]d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)+o_p(1)$$
$$= n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^\tau [Z_i-z(t)][dM_i(t)+(A_i^\circ(t)-A_i(t))$$
$$\times d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)]+o_p(1).$$

From [13], $\frac{S_C(\cdot)-\hat{S}_C(\cdot)}{S_C(\cdot)}$ has the following martingale integral representation:

$$\frac{S_C(t)-\hat{S}_C(t)}{S_C(t)}=\int_0^\tau \frac{I(s\le t)}{\overline{Y}(s)}dM_C(s),$$

where

$$M_C(t)=\sum_{i=1}^n M_i^C(t),\ \ M_i^C(t)=N_i^C(t)-\int_0^t Y_i(s)d\Lambda_C(s).$$

Thus, we have

$$A_i^\circ(t)-A_i(t)=A_i^\circ(t)\Big\{\int_0^\tau \frac{I(s\le t)-I(s\le X_i)}{\overline{Y}(s)}dM_C(s)\Big\},$$

and hence

$$n^{-\frac{1}{2}}U(\beta_0,\hat{H}_n(\cdot,\beta_0))$$
$$= n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^\tau [Z_i-z(t)]\Big\{dM_i(t)+A_i^\circ(t)$$
$$\times\Big\{\int_0^\tau \frac{I(s\le t)-I(s\le X_i)}{\overline{Y}(s)}dM_C(s)\Big\}$$
$$\times d\Lambda_\epsilon(H_0(t)+Z_i^T\beta_0)\Big\}+o_p(1)$$
$$= n^{-\frac{1}{2}}\sum_{i=1}^n\Big\{\int_0^\tau [Z_i-z(t)]dM_i(t)+\int_0^\tau\frac{1}{n}\sum_{j=1}^n\int_0^\tau [Z_j$$
$$-z(s)]A_j^\circ(s)\frac{I(s\ge t)-I(X_j\ge t)}{\frac{1}{n}\overline{Y}(t)}d\Lambda_\epsilon(H_0(s)+Z_j^T\beta_0)$$
$$\times dM_i^C(t)\Big\}+o_p(1)$$
$$= n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^\tau [Z_i-z(t)]dM_i(t)$$
$$+n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^\tau D(t)dM_i^C(t)+o_p(1),$$

which is a sum of independent mean zero random vectors. Hence it follows from the Central Limit Theorem that $n^{-\frac{1}{2}}U(\beta_0,\hat{H}_n(\cdot,\beta_0))\overset{\mathcal{D}}{\to}N(0,\Sigma^*)$. By the Taylor series expansion, we have

$$n^{-\frac{1}{2}}\{U(\hat\beta_n,\hat{H}_n(\cdot,\hat\beta_n))-U(\beta_0,\hat{H}_n(\cdot,\beta_0))\}$$
$$= \frac{1}{n}\frac{\partial U(\beta,\hat{H}_n(\cdot,\beta))}{\partial\beta}\Big|_{\beta=\beta_0}\sqrt{n}(\hat\beta_n-\beta_0)+o_p(1).$$

Using results from Part 2 and recognizing that $U(\hat\beta_n,\hat{H}_n(\cdot,\hat\beta_n))=0$, it follows that

$$\sqrt{n}(\hat\beta_n-\beta_0)\overset{\mathcal{D}}{\to}N(0,\Sigma_*^{-1}\Sigma^*(\Sigma_*^{-1})^T).$$

The proof of the consistency of the variance estimator is straightforward and we omit it here.

To prove the weak convergence of $\sqrt{n}\{\hat{H}_n(t,\hat\beta_n)-H_0(t)\}$, note that

$$\sqrt{n}\{\hat{H}_n(t,\hat\beta_n)-H_0(t)\}$$
$$= \sqrt{n}\{\hat{H}_n(t,\hat\beta_n)-\hat{H}_n(t,\beta_0)\}+\sqrt{n}\{\hat{H}_n(t,\beta_0)-H_0(t)\}$$
$$= \sqrt{n}\frac{\partial\hat{H}_n(t,\beta)}{\partial\beta}\Big|_{\beta=\beta_0}(\hat\beta_n-\beta_0)+\sqrt{n}\{\hat{H}_n(t,\beta_0)-H_0(t)\}$$
$$+o_p(1)$$
$$= -\frac{1}{\sqrt{n}}\int_0^t\frac{B(s,t)}{B_2(s)}B_1^Z(s)dH_0(s)\Sigma_*^{-1}\sum_{i=1}^n\Big\{\int_0^\tau [Z_i-z(t)]$$
$$\times dM_i(t)+\int_0^\tau D(t)dM_i^C(t)\Big\}$$
$$+\sqrt{n}\{\hat{H}_n(t,\beta_0)-H_0(t)\}+o_p(1).$$

By (10), we have

$$\sqrt{n}\{\hat{H}_n(t,\beta_0)-H_0(t)\}$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t\frac{B(s,t)}{B_2(s)}[dM_i(s)+(A_i^\circ(s)-A_i(s))$$
$$\times d\Lambda_\epsilon(H_0(s)+Z_i^T\beta_0)]+o_p(1)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t\frac{B(s,t)}{B_2(s)}\Big\{dM_i(s)+A_i^\circ(s)$$
$$\times\int_0^\tau\frac{I(u\le s)-I(u\le X_i)}{\overline{Y}(u)}dM_C(u)$$
$$\times d\Lambda_\epsilon(H_0(s)+Z_i^T\beta_0)\Big\}+o_p(1)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t\frac{B(s,t)}{B_2(s)}dM_i(s)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t\frac{\int_u^t\frac{B(s,t)}{B_2(s)}A_i^\circ(s)d\Lambda_\epsilon(H_0(s)+Z_i^T\beta_0)}{\overline{Y}(u)}$$
$$\times dM_C(u)-\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t\frac{B(s,t)}{B_2(s)}A_i^\circ(s)$$
$$\times d\Lambda_\epsilon(H_0(s)+Z_i^T\beta_0)\int_0^{X_i}\frac{1}{\overline{Y}(u)}dM_C(u)+o_p(1).$$

Consequently,

$$\sqrt{n}\{\hat{H}_n(t,\hat\beta_n)-H_0(t)\}$$
$$= -\frac{1}{\sqrt{n}}\int_0^t\frac{B(s,t)}{B_2(s)}B_1^Z(s)dH_0(s)\Sigma_*^{-1}\sum_{i=1}^n\Big\{\int_0^\tau [Z_i$$
$$-z(t)]dM_i(t)+\int_0^\tau D(t)dM_i^C(t)\Big\}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \frac{B(s,t)}{B_2(s)}dM_i(s)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \frac{\int_u^t \frac{B(s,t)}{B_2(s)}A_i^\circ(s)d\Lambda_\epsilon(H_0(s)+Z_i^T\beta_0)}{\overline{Y}(u)}$$
$$\times dM_C(u)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \frac{B(s,t)}{B_2(s)}A_i^\circ(s)$$
$$\times d\Lambda_\epsilon(H_0(s)+Z_i^T\beta_0)\int_0^{X_i}\frac{1}{\overline{Y}(u)}dM_C(u)+o_p(1).$$

Using an argument similar to that used in step 3 in [18] (p. 226), we can show that $\sqrt{n}\{\hat{H}_n(t,\hat{\beta}_n)-H_0(t)\}$ converges weakly to a tight Gaussian process. This completes the proof of the theorem. □

ACKNOWLEDGEMENTS

Wei's work was supported by a research fund from the Shanghai University of Finance and Economics (No. 2017110070). Zhou's work was supported by the State Key Program in the Major Research Plan of National Natural Science Foundation of China (No. 91546202), the State Key Program of National Natural Science Foundation of China (No. 71331006), and Innovative Research Team of Shanghai University of Finance and Economics (No. IRTSHUFE13122402). Wan's work was supported by a Theme-Based Research Scheme (No. T32-102/14N) and a General Research Fund (No. 9042086), both from the Hong Kong Research Grants Council, and a strategic grant from the City University of Hong Kong (No. 7004786). We thank the editor, the associate editor and two referees for helpful comments on an earlier draft of the paper. The usual disclaimer applies.

Received 15 April 2017

REFERENCES

[1] Addona, V. and Wolfson, D. B. (2006). A formal test for stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Analysis* **12** 267–284. MR2328577

[2] Asgharian, M., M'Lan, C. E., and Wolfson, D. B. (2002). Length-biased sampling with right censoring: an unconditional approach. *Journal of the American Statistical Association* **97** 201–209. MR1947280

[3] Asgharian, M. and Wolfson, D. B. (2005). Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *The Annals of Statistics* **33** 2109–2131. MR2211081

[4] Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine* **2** 273–277.

[5] Chen, K. N., Jin, Z. Z., and Ying, Z. L. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89** 659–668. MR1929170

[6] Chen, K. N. and Tong, X. W. (2010). Varying coefficient transformation models with censored data. *Biometrika* **97** 969–976. MR2746165

[7] Cheng, S. C., Wei, L. J., and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82** 835–845. MR1380818

[8] Cheng, Y. J. and Huang, C. Y. (2014). Combined estimating equation approaches for semiparametric transformation models with length-biased survival data. *Biometrics* **70** 608–618. MR3261780

[9] Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of Royal Statistical Society: Series B* **34** 187–220. MR0341758

[10] Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* **72** 27–36.

[11] Dabrowska, D. M. and Doksum, K. A. (1988). Estimation and testing in the two-sample generalized odds-rate model. *Journal of the American Statistical Association* **83** 744–749. MR0963802

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

[13] Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York. MR1100924

[14] Huang, C. Y. and Qin, J. (2011). Nonparametric estimation for length-biased and right-censored data. *Biometrika* **98** 177–186. MR2804218

[15] 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. MR3010882

[16] Huang, C. Y., Qin, J., and Follmann, D. A. (2012). A maximum pseudo-profile likelihood estimator for the Cox model under length-biased sampling. *Biometrika* **99** 199–210. MR2899673

[17] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53** 457–481. MR0093867

[18] Kim, J. P., Lu, W. B., Sit, T., and Ying, Z. L. (2013). A unified approach to semiparametric transformation models under general biased sampling schemes. *Journal of the American Statistical Association* **108** 217–227. MR3174614

[19] Kong, L., Cai, J. W., and Sen, P. K. (2004). Weighted estimating equations for semiparametric transformation models with censored data from a case-cohort design. *Biometrika* **91** 305–319. MR2081303

[20] Lai, T. L. and Ying, Z. L. (1991a). Estimating a distribution function with truncated and censored data. *The Annals of Statistics* **19** 417–442. MR1091860

[21] Lai, T. L. and Ying, Z. L. (1991b). Rank regression methods for left-truncated and right-censored data. *The Annals of Statistics* **19** 531–556. MR1105835

[22] Li, G. and Qin, J. (1998). Semiparametric likelihood-based inference for biased and truncated data when the total sample size is known. *Journal of the Royal Statistical Society, Series B* **60** 243–254. MR1625616

[23] Liu, H., Qin, J., and Shen, Y. (2012). Imputation for semiparametric transformation models with biased-sampling data. *Lifetime Data Analysis* **18** 470–503. MR2973776

[24] Lu, W. B. and Tsiatis, A. A. (2006). Semiparametric transformation models for the case-cohort study. *Biometrika* **93** 207–214. MR2277751

[25] Lu, W. B. and Ying, Z. L. (2004). On semiparametric transformation cure models. *Biometrika* **91** 331–343. MR2081305

[26] Lu, W. B. and Zhang, H. H. (2010). On estimation of partially linear transformation models. *Journal of the American Statistical Association* **105** 683–691. MR2724852

[27] Luo, X. D. and Tsai, W. Y. (2009). Nonparametric estimation for right-censored length-biased data: a pseudo-partial likelihood approach. *Biometrika* **96** 873–886. MR2767276

[28] Mandel, M. and Ritov, Y. (2010). The accelerated failure time model under biased sampling. *Biometrics* **66** 1306–1308. MR2758519

[29] Miller, R. and Halpern, J. (1982). Regression with censored

data. *Biometrika* **69** 521–531. MR0695199

[30] Ning, J., Qin, J., and Shen, Y. (2011). Buckley-James-type estimator with right-censored and length-biased data. *Biometrics* **67** 1369–1378. MR2872387

[31] Ning, J., Qin, J., and Shen, Y. (2014a). Semiparametric accelerated failure time model for length-biased data with application to dementia study. *Statistica Sinica* **24** 313–333. MR3183686

[32] Ning, J., Qin, J., and Shen, Y. (2014b). Score estimating equations from embedded likelihood functions under accelerated failure time model. *Journal of the American Statistical Association* **109** 1625–1635. MR3293615

[33] Pollard, D. (1990). *Empirical Processes: Theory and Applications. NSF-CBMS Regional Conference Series in Probability and Statistics (Vol. 2)*, Hayward, CA: IMS. MR1089429

[34] Qin, J. and Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics* **66** 382–392. MR2758818

[35] Qin, J., Ning, J., Liu, H., and Shen, Y. (2011). Maximum likelihood estimations and EM algorithms with length-biased data. *Journal of the American Statistical Association* **106** 1434–1449. MR2896847

[36] Reinhard, H. (1986). *Differential Equations: Foundations and Applications*. North Oxford, London.

[37] Shen, P. S. (2011). Semiparametric analysis of transformation models with left-truncated and right-censored data. *Computational Statistics* **26** 521–537. MR2833146

[38] Shen, Y., Ning, J., and Qin, J. (2009). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association* **104** 1192–1202. MR2750244

[39] Shen, Y., Ning, J., and Qin, J. (2012). Likelihood approaches for the invariant density ratio model with biased-sampling data. *Biometrika* **99** 363–378. MR2931259

[40] Tsai, W. Y., Jewell, N. P., and Wang, M. C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74** 883–886.

[41] Tsai, W. Y. (2009). Pseudo-partial likelihood for proportional hazards models with biased-sampling data. *Biometrika* **96** 601–615. MR2538760

[42] Turnbull, B. U. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B* **38** 290–295. MR0652727

[43] Wang, M. C. (1996). Hazards regression analysis for length-biased data. *Biometrika* **83** 343–354. MR1439788

[44] Wang, X. and Wang, Q. H. (2015). Estimation for semiparametric transformation models with length-biased sampling. *Journal of Statistical Planning and Inference* **156** 80–89. MR3267514

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

Wenhua Wei
School of Statistics and Management
Shanghai University of Finance and Economics
Shanghai
China
E-mail address: wei.wenhua@mail.shufe.edu.cn

Yong Zhou
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
Beijing
China
E-mail address: yzhou@amss.ac.cn

Alan T. K. Wan
Department of Management Sciences
City University of Hong Kong
Kowloon
Hong Kong
E-mail address: Alan.Wan@cityu.edu.hk