Statistica Sinica **26** (2016), 1129-1158
doi:http://dx.doi.org/10.5705/ss.2013.117

SEMIPARAMETRIC INFERENCE FOR THE PROPORTIONAL MEAN RESIDUAL LIFE MODEL WITH RIGHT-CENSORED LENGTH-BIASED DATA

Fangfang Bai[1], Jian Huang[2,3] and Yong Zhou[3,4]

[1] *University of International Business and Economics,* [2] *University of Iowa,*
[3] *Shanghai University of Finance and Economics*
and [4] *Chinese Academy of Sciences*

Abstract: We propose a semiparametric inference approach for proportional mean residual life model with right-censored length-biased data, that arise frequently in observational studies, especially in epidemiological cohort studies. A challenge in the analysis of such data is the presence of informative censoring. Another challenge is that the distribution of the observed data is different from that of the underlying model. We develop an inverse probability weighted approach to estimation based on estimating equations. We establish large sample properties and study the semiparametric efficiency and double robustness property of the proposed estimators. We also propose an improved estimator that chooses the most efficient one in the class of augmented inverse probability weighted estimators. We use simulation studies to evaluate the proposed method, and illustrate its application using a data analysis.

Key words and phrases: Dependent censoring, estimating equation, length-biased data, proportional mean residual model, semiparametric efficiency.

1. Introduction

Length-biased data arise frequently in such observational studies, as screening programs for chronic diseases (Zelen and Feinleib (1969)), aetiological studies (Simon (1980)), and wildlife studies (Patil and Rao (1978)). In such studies, for a fixed enrollment period, the initiation time of a subject is recorded and the subject is followed up for a fixed period of time. Only the subjects who have survived at or beyond the enrollment time can be observed, and so they are not a random sample from the underlying population. Generally, in this situation, the probability of observing a subject is proportional to the length of the period from the initiation time to the failure time. For this reason the data is said to be length-biased, observed individuals tending to live longer than those randomly selected from the underlying population. An example of right-censored length-biased data is the dementia data from the Canadian Study of Health and Aging

(CSHA), see Wolfson et al. (2001). In this data set, the longer the duration of an individual's dementia symptoms, the larger the probability of obtaining the observation from this individual, it is length-biased.

There are two main difficulties in the analysis of right-censored length-biased data: the distribution of the observed right-censored length-biased data is different from that of a random sample from the underlying population; with right-censored length-biased data, the censoring mechanism is informative since the observed failure time and the censoring time share the same initiation time. Several authors have considered the problem of regression analysis with right-censored length-biased data. For example, Wang (1996) considered the proportional hazards regression model using a bias-adjusted risk set method. Other work attempted to avoid the bias due to the informative censoring by means of not allowing it (Vardi (1985),Wang (1996)). More recently, Shen, Ning, and Qin (2009) proposed an inverse probability weighted approach to solve the problem. Using a similar approach, Qin and Shen (2010) studied the estimation in the proportional hazard model based on estimating equations.

The mean residual life function (MRLF) is an important characteristic of the remaining life of an individual who has survived up to a certain time. In many situations we are interested in the remaining life expectancy rather than the probability of immediate failure or the distribution of a failure time. Thus, a cancer patient may care most about how long he or she can expect to survive from the time of diagnosis, and insurance companies are interested in the mean residual life times of their customers. Mean residual life models have received much attention in the literature. Examples include Bickel et al. (1993), Rojo and Ghebremichael (2006), Oakes and Dasu (1990) and Maguluri and Zhang (1994), who considered estimation in the mean residual life model with complete data; and Chen and Cheng (2005), Chen et al. (2005), Chen and Cheng (2006) and Sun and Zhang (2009), who considered estimation in the same model with censored data.

Motivated by these considerations, we study the estimation in the proportional mean residual life model with right-censored length-biased data. Because of the sampling scheme of the length-biased data, the distribution of the observations have a unique structure that is different from that of traditional survival data. Ignoring this fact can lead to bias in estimation, so it is necessary to develop new methods to analyze the mean residual life data. We propose an estimating equation approach using the inverse probability weighted technique to correct for length bias and to adjust for informative censoring. Two estimating equations are constructed: the first is used to get an initial estimator, the second to obtain the final estimator by making use of the internal information. In this way, the efficiency of the estimator is improved. Moreover, our procedure can be easily generalized to the additive mean residual life model or the transformation model in Sun and Zhang (2009). We investigate the semiparametric efficiency of the proposed estimator for a given weight function and construct the most efficient estimator in a class of estimators. In addition, we study the double robustness property of the most efficient estimator.

The rest of this paper is organized as follows. In Section 2 we introduce the notation and assumptions used throughout the paper. We also describe the estimation approach and derive the asymptotic properties of the proposed estimator. In Section 3 we study the efficiency of the proposed estimator and provide a way to construct a class of more efficient estimators. The double robustness property is discussed there. In Section 4 we report on simulation studies to evaluate the finite sample properties of our estimators. In Section 5 a data example is used to illustrate the application of the proposed method. The proofs of theoretical results are given in the Appendix.

2. Estimation Method and Asymptotic Properties

2.1. Notation and assumptions

Let $\widetilde{T}$ denote the true failure time of the population of interest, that is, the time from the initiating event to the failure event, let A be the time from the initiation event to enrollment, let V be the time from enrollment to the failure event, and let C be the time from enrollment to censoring. Here $T = A + V$ is the observed failure time. Note that T can be observed only when $\widetilde{T} \geq A$. Let X be a p-vector baseline covariate, and assume that C is independent of both A and V given X.

Take $Y = \min\{T, A + C\}$ and $\delta = I\{A + V \leq A + C\}$, where $I\{\cdot\}$ is an indicator function. Then the observed data is $\{(Y_i, \delta_i, X_i, A_i), i = 1, \cdots, n\}$. As $A + V$ and $A + C$ are dependent, censoring is informative.

Let $f(\cdot|x)$ be the conditional density function of $\widetilde{T}$ given covariates $X = x$. The conditional density function of the observed length-biased T is

$$g(t|x) = \frac{tf(t|x)}{\mu(x)},$$

where $\mu(x) = \int tf(t|x)dt$.

Write $m(t|X) = E[\widetilde{T} - t|\widetilde{T} > t, X]$ as the mean residual life function associated with the covariate X. To assess the covariate effects, we consider the proportional mean residual life model

$$m(t|X) = m_0(t)\exp(\beta^T X), \tag{2.1}$$

where $m_0(t)$ is an unknown baseline mean residual life function and β is the parameter of interest that describes the effect of the covariate on the mean residual life.

We assume that the upper support of the censoring variable C is longer than that of V and take $0 < \tau = \sup\{t : S_{\widetilde{T}}(t) > 0\} < \infty$, where $S_{\widetilde{T}}(t)$ is the survival function of the true failure time $\widetilde{T}$. This assumption ensures that the survival function of $\widetilde{T}$ is estimable.

2.2. Estimation method when censoring variable C is independent of covariates

If the censoring variable C is independent of the covariates, the conditional probability of observing the data $(Y = y, \delta = 1, A = a)$ given X is

$$P(A = a, Y = y, \delta = 1|X = x) = P(A = a, V = y - a, C \geq y - a|X = x) = \frac{f(y|X = x)S_C(y - a)}{\mu(x)},$$

where $S_C(\cdot)$ is the survival function of censoring variable C.

Let

$$M_i(t) = \frac{\delta_i I(Y_i > t)}{Y_i S_C(Y_i - A_i)}\Big[(Y_i - t) - m_0(t)\exp(\beta^T X_i)\Big]. \tag{2.2}$$

Given the observed data $\{Y_i, \delta_i, X_i, A_i, i = 1, \cdots, n\}$, $\{M_i(t), 0 \leq t \leq \tau\}$ are zero-mean stochastic processes. Therefore we can construct the estimating equations for β and the baseline MRLF $m_0(\cdot)$ as

$$\sum_{i=1}^{n} \frac{\delta_i I(Y_i > t)}{Y_i S_C(Y_i - A_i)}\Big[(Y_i - t) - m_0(t)\exp(\beta^T X_i)\Big] = 0, \tag{2.3}$$

$$\sum_{i=1}^{n} \int_0^{\tau} \frac{\delta_i I(Y_i > t)}{Y_i S_C(Y_i - A_i)} X_i \Big[(Y_i - t) - m_0(t)\exp(\beta^T X_i)\Big] dH(t) = 0, \tag{2.4}$$

where $H(t)$ is an increasing known weight function and $0 \leq t \leq \tau$ in (2.3).

Here we use inverse probability of weighted (IPW) method in (2.3) and (2.4). The first estimating equation (2.3) is used to obtain an initial estimator of β and a point by point estimator of $m_0(t)$. Notice that for different t, the covariate effect β is the same. Making use of this information, estimating equation (2.4) is constructed, in this way, the efficiency of the parameter β is improved. This idea can be readily used in the additive mean residual life model, or the general transformation model of Sun and Zhang (2009).

Since the survival function $S_C(Y_i - A_i)$ can be estimated by its Kaplan-Meier estimator $\widehat{S}_C(Y_i - A_i)$, $m_0(t)$ can be estimated based on the estimating equation

$$\sum_{i=1}^{n} \frac{\delta_i I(Y_i > t)}{Y_i \widehat{S}_C(Y_i - A_i)}\Big[(Y_i - t) - m_0(t)\exp(\beta^T X_i)\Big] = 0, \quad 0 \leq t \leq \tau. \tag{2.5}$$

The resulting estimator has a closed form,

$$\widehat{m}_0(t,\beta) = \frac{\sum_{i=1}^n \delta_i I(Y_i > t)\{Y_i\widehat{S}_C(Y_i - A_i)\}^{-1}(Y_i - t)}{\sum_{i=1}^n \delta_i I(Y_i > t)\{Y_i\widehat{S}_C(Y_i - A_i)\}^{-1}\exp(\beta^T X_i)}. \tag{2.6}$$

To obtain an estimator of β, we replace $m_0(t)$ with $\widehat{m}_0(t,\beta)$ in (2.4) to get the estimating equation

$$U(\beta) = \sum_{i=1}^n \int_0^\tau \frac{\delta_i I(Y_i > t)}{Y_i\widehat{S}_C(Y_i - A_i)} X_i \Big[(Y_i - t) - \widehat{m}_0(t,\beta)\exp(\beta^T X_i)\Big] dH(t) = 0. \tag{2.7}$$

Let $\widehat{\beta}$ be the solution of (2.7), and write $\widehat{m}_0(t) = \widehat{m}_0(t,\widehat{\beta})$. We have the final estimators $\widehat{\beta}$ and $\widehat{m}_0(t)$ when C is independent of the covariates.

2.3. Estimation method when censoring variable C depends on covariates

Assume that C depends on Z, a part of the covariate X, or just X itself. As in Section 2.2, estimating equations can be constructed as

$$\sum_{i=1}^n \frac{\delta_i I(Y_i > t)}{Y_i S_C(Y_i - A_i|Z_i)} \Big[(Y_i - t) - m_0(t)\exp(\beta^T X_i)\Big] = 0, \quad 0 \le t \le \tau, \tag{2.8}$$

$$\sum_{i=1}^n \int_0^\tau \frac{\delta_i I(Y_i > t)}{Y_i S_C(Y_i - A_i|Z_i)} X_i \Big[(Y_i - t) - m_0(t)\exp(\beta^T X_i)\Big] dH(t) = 0, \tag{2.9}$$

where $S_C(\cdot|Z)$ is the survival function of C given Z. Thus, if $S_C(\cdot|Z)$ can be estimated properly, the final estimator can be derived as in the independent censoring case.

When censoring depends on the covariates, $S_C(\cdot|Z)$ can be estimated by semiparametric methods based on a regression model specified for the censoring time (e.g., the Cox proportional hazards model or transformation model), or nonparametric methods using the local Kaplan-Meier estimator of the survival function.

When Z is discrete taking a finite number of possible values and the sample size is sufficiently large, we can obtain the Kaplan-Meier estimators $\hat{S}_C(\cdot|Z_i)$ for each $Z = Z_i$. In general, we can model the dependence of C on Z and obtain an estimator of $S_C(\cdot|Z)$. Here we assume a Cox proportional hazards model for the conditional hazard function of C given Z, $\lambda_C(t|Z) = \lambda_0(t)\exp(\alpha^T Z)$, where $\lambda_0(t)$ is an unknown baseline hazard function and α is a vector of unknown parameters. We estimate α using the Cox partial likelihood estimator $\widehat{\alpha}$, the solution of

$$\sum_{i=1}^n \int_0^\tau (Z_i - \overline{Z}(t;\alpha))dN_i^C(t) = 0, \tag{2.10}$$

where $N_i^C(t) = I(Y_i - A_i \le t, \delta_i = 0)$, $Y_i(t) = I(Y_i - A_i \ge t)$, $\overline{Z}(t;\alpha) = S^{(1)}(t;\alpha)/S^{(0)}(t;\alpha)$, and for $k = 0, 1, 2$, $S^{(k)}(t;\alpha) = (1/n)\sum_{i=1}^n Y_i(t) Z_i^{\otimes k} \exp(\alpha^T Z_i)$. Here for a vector a, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, $a^{\otimes 2} = aa^T$. Let $\hat{\Lambda}_0(t)$ be the Breslow estimator of $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$,

$$\hat{\Lambda}_0(t) = \sum_{i=1}^n \int_0^t \frac{dN_i^C(s)}{\sum_{j=1}^n Y_j(s)\exp(\hat{\alpha}^T Z_j)}.$$

So $S_C(t|Z)$ can be estimated by $\hat{S}_C(t|Z) = \exp\{-\exp(\hat{\alpha}^T Z)\hat{\Lambda}_0(t)\}$. Plug $\hat{S}_C(t|Z)$ into (2.8) and (2.9), we obtain

$$\sum_{i=1}^n \frac{\delta_i I(Y_i > t)}{Y_i \hat{S}_C(Y_i - A_i|Z_i)}\Big[(Y_i - t) - m_0(t)\exp(\beta^T X_i)\Big] = 0, \quad 0 \le t \le \tau, \tag{2.11}$$

$$\sum_{i=1}^n \int_0^\tau \frac{\delta_i I(Y_i > t)}{Y_i \hat{S}_C(Y_i - A_i|Z_i)} X_i\Big[(Y_i - t) - m_0(t)\exp(\beta^T X_i)\Big] dH(t) = 0. \tag{2.12}$$

Let

$$\widehat{m}_{0d}(t,\beta) = \frac{\sum_{i=1}^n \delta_i I(Y_i > t)\{Y_i \widehat{S}_C(Y_i - A_i|Z_i)\}^{-1}(Y_i - t)}{\sum_{i=1}^n \delta_i I(Y_i > t)\{Y_i \widehat{S}_C(Y_i - A_i|Z_i)\}^{-1}\exp(\beta^T X_i)}.$$

As in Section 2.2, we can obtain the estimators $\hat{\beta}_d$ and $\hat{m}_{0d}(t)$ when C depends on the covariates, where $\hat{m}_{0d}(t) = \widehat{m}_{0d}(t, \hat{\beta}_d)$.

The Cox model may not capture the real relationship between C and the covariates. We propose a double robustness estimator in Section 3: if the Cox model is not satisfied, we can still obtain a consistent and asymptotically normal estimator.

2.4. Large sample properties

We first consider the large sample properties of the estimators when C is independent of the covariates. Let

$$\widehat{\pi}(t) = \frac{1}{n}\sum_{i=1}^n I(Y_i - A_i \ge t), N_i^C(t) = I(Y_i - A_i \le t, \delta_i = 0),$$

$$\widehat{\Lambda}_C(t) = \int_0^t \frac{\sum_{i=1}^n dN_i^C(s)}{\sum_{i=1}^n I(Y_i - A_i \ge s)},$$

$$\widehat{M}_i^C(t) = N_i^C(t) - \int_0^t I(Y_i - A_i \ge u) d\widehat{\Lambda}_C(u),$$

$$\widehat{M}_i(t) = \frac{\delta_i I(Y_i > t)}{Y_i \widehat{S}_C(Y_i - A_i)}\big[(Y_i - t) - \widehat{m}_0(t)\exp(\widehat{\beta}^T X_i)\big],$$

$$\widehat{Q}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \widehat{M}_i(s)\{X_i - \overline{X}(s,\widehat{\beta})\}dH(s)I(Y_i - A_i \geq t),$$

$$\overline{X}(t,\widehat{\beta}) = \frac{\sum_{i=1}^{n}\delta_i I(Y_i > t)\{Y_i\widehat{S}_C(Y_i - A_i)\}^{-1}\exp(\widehat{\beta}^T X_i)X_i}{\sum_{i=1}^{n}\delta_i I(Y_i > t)\{Y_i\widehat{S}_C(Y_i - A_i)\}^{-1}\exp(\widehat{\beta}^T X_i)}.$$

Denote the true value of the parameters by β_0 and $m_0(t)$, respectively.

Theorem 1. *Under assumptions* (A1)−(A4) *in the Appendix, $\widehat{\beta}$ and $\widehat{m}_0(t)$ exist and are unique. Moreover, $\widehat{\beta}$ is a strong consistent estimator of β_0, and $\widehat{m}_0(t)$ is a strong consistent estimator of $m_0(t)$ in $t \in [0,\tau]$.*

Theorem 2. *Under assumptions* (A1)−(A4) *in the Appendix, we have:*

(1) $\sqrt{n}(\widehat{\beta} - \beta_0)$ *is asymptotically normal with mean zero and a covariance matrix that can be consistently estimated by $\widehat{B}^{-1}\widehat{\Sigma}\widehat{B}^{-1}$, where $\widehat{\Sigma} = (1/n)\sum_{i=1}^{n}\widehat{\xi}_i^{\otimes 2}$,*

$$\widehat{B} = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \frac{\delta_i I(Y_i > t)}{Y_i\widehat{S}_C(Y_i - A_i)}\big(X_i - \overline{X}(t,\widehat{\beta})\big)^{\otimes 2}\widehat{m}_0(t)\exp(\widehat{\beta}^T X_i)dH(t),$$

$$\widehat{\xi}_i = \int_0^\tau \widehat{M}_i(t)\{X_i - \overline{X}(t,\widehat{\beta})\}dH(t) + \int_0^\tau \frac{\widehat{Q}(t)}{\widehat{\pi}(t)}d\widehat{M}_i^C(t);$$

(2) $\{\sqrt{n}(\widehat{m}_0(t) - m_0(t)),\ 0 \leq t \leq \tau\}$ *converges weakly to a zero-mean Gaussian process whose covariance function at (s,t) can be consistently estimated by*

$$\widehat{\Gamma}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\psi}_i(s)\widehat{\psi}_i(t),$$

where

$$\widehat{\psi}_i(t) = \widehat{\Phi}(t)^{-1}[\widehat{M}_i(t) + \int_0^\tau \frac{\widehat{R}(t,\mu)}{\widehat{\pi}(\mu)}d\widehat{M}_i^C(\mu)] - \overline{X}(t,\widehat{\beta})\widehat{m}_0(t)\widehat{B}^{-1}\widehat{\xi}_i,$$

$$\widehat{\Phi}(t) = \frac{1}{n}\sum_{i=1}^{n}\delta_i I(Y_i > t)\{Y_i\widehat{S}_C(Y_i - A_i)\}^{-1}\exp(\widehat{\beta}^T X_i),$$

and

$$\widehat{R}(t,\mu) = \frac{1}{n}\sum_{i=1}^{n}\widehat{M}_i(t)I(Y_i - A_i \geq \mu).$$

This result can be used as a basis for making statistical inference for β_0 and $m_0(t)$. While it is easy to obtain the pointwise confidence interval for $m_0(t)$ for any fixed t by the asymptotic normality of $\widehat{m}_0(t)$, simultaneous confidence bands are more difficult to construct, since for an interval $[t_1,\, t_2]$, the distribution of $\sup_{t\in[t_1,t_2]}(\widehat{m}_0(t) - m_0(t))$ has to be calculated. This is difficult because the limiting process of $n^{1/2}\{\widehat{m}_0(t) - m_0(t)\}$ does not have an independent increment structure. To get around this difficulty, we use a resampling scheme to approximate the distribution of $\sqrt{n}\{\widehat{m}_0(t) - m_0(t)\}$. Let $\widehat{W}(t) = n^{-1/2}\sum_{i=1}^n \widehat{\psi}_i(t)\Omega_i$, where $\Omega_i, i = 1, \ldots, n$ are independent standard normal random variables independent of $\{Y_i, \delta_i, X_i, A_i\}$. Then the distribution of the process $\sqrt{n}\{\widehat{m}_0(t) - m_0(t)\}$ can be approximated by that of a zero-mean Gaussian process $\widehat{W}(t)$ by the results in Lin et al. (2000). So an approximate $1-\alpha$ simultaneous confidence band for $m_0(t)$ can be obtained as follows. First, for fixed $\{Y_i, \delta_i, X_i, A_i\}$, generate repeatedly $(\Omega_1, \ldots, \Omega_n)$ to obtain many realizations of $\widehat{W}(t)$. Second, approximate the distribution of $\sqrt{n}\{\widehat{m}_0(t) - m_0(t)\}$ by the realizations of $\widehat{W}(t)$.

We now consider the large sample properties of the estimators when C depends on the covariates. Let

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big\{\frac{S^{(2)}(t;\hat{\alpha})}{S^{(0)}(t;\hat{\alpha})} - \overline{Z}(t;\hat{\alpha})^{\otimes 2}\Big\} dN_i^C(t),$$

$$\widehat{M}_i^d(t) = N_i^C(t) - \int_0^t I(Y_i - A_i \geq u)\exp(\hat{\alpha}^T Z_i)d\hat{\Lambda}_0(u),$$

$$\widehat{M}_i^*(t) = \frac{\delta_i I(Y_i > t)}{Y_i\widehat{S}_C(Y_i - A_i|Z_i)}\big[(Y_i - t) - \widehat{m}_0(t)\exp(\widehat{\beta}_d^T X_i)\big],$$

$$\overline{X}_d(t, \hat{\beta}_d) = \frac{\sum_{i=1}^n \delta_i I(Y_i > t)\{Y_i\widehat{S}_C(Y_i - A_i|Z_i)\}^{-1}\exp(\widehat{\beta}_d^T X_i)X_i}{\sum_{i=1}^n \delta_i I(Y_i > t)\{Y_i\widehat{S}_C(Y_i - A_i|Z_i)\}^{-1}\exp(\widehat{\beta}_d^T X_i)},$$

$$\hat{Q}_d(t) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \widehat{M}_i^*(u)(X_i - \overline{X}_d(u, \hat{\beta}_d))dH(u)\exp\{\hat{\alpha}^T Z_i\}I(Y_i - A_i \geq t),$$

$$\hat{D}_d = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \widehat{M}_i^*(u)(X_i - \overline{X}_d(u, \hat{\beta}_d))dH(u)\hat{\Lambda}_0(Y_i - A_i)\exp\{\hat{\alpha}^T Z_i\}Z_i^T,$$

$$\hat{\Phi}_d(t) = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i I(Y_i > t)\exp(\widehat{\beta}_d^T X_i)}{Y_i\widehat{S}_C(Y_i - A_i|Z_i)},$$

$$\hat{R}_1(t) = \frac{1}{n}\sum_{i=1}^n \widehat{M}_i^*(t)\hat{\Lambda}_0(Y_i - A_i)\exp\{\hat{\alpha}^T Z_i\}Z_i^T,$$

$$\hat{R}_d(t, u) = \frac{1}{n}\sum_{i=1}^n \widehat{M}_i^*(t)\exp\{\hat{\alpha}^T Z_i\}I(Y_i - A_i \geq u).$$

Theorem 3. *Under assumptions* (A1)−(A3) *and* (A5) *in the Appendix,* $\widehat{\beta}_d$ *and* $\widehat{m}_{0d}(t)$ *exist and are unique. Moreover,* $\widehat{\beta}_d$ *is a strong consistent estimator of* β_0 *and* $\widehat{m}_{0d}(t)$ *is a strong consistent estimator of* $m_0(t)$ *in* $t \in [0, \tau]$.

Theorem 4. *Under assumptions* (A1)−(A3) *and* (A5) *in the Appendix, we have:*

(1) $\sqrt{n}(\widehat{\beta}_d - \beta_0)$ *is asymptotically normal with mean zero and a covariance matrix that can be consistently estimated by* $\widehat{B}_d^{-1}\widehat{\Sigma}_d\widehat{B}_d^{-1}$, *where* $\widehat{\Sigma}_d = (1/n)\sum_{i=1}^n \widehat{\xi}_i^{d\otimes 2}$,

$$\widehat{B}_d = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{\delta_i I(Y_i > t)}{Y_i \widehat{S}_C(Y_i - A_i|Z_i)} \big(X_i - \overline{X}_d(t, \widehat{\beta}_d)\big)^{\otimes 2} \widehat{m}_{0d}(t) \exp(\widehat{\beta}_d^T X_i) dH(t),$$

$$\widehat{\xi}_i^d = \int_0^\tau \widehat{M}_i^*(t)\{X_i - \overline{X}_d(t, \widehat{\beta}_d)\} dH(t) + \int_0^\tau \frac{\widehat{Q}_d(t)}{S^{(0)}(t;\hat{\alpha})} d\widehat{M}_i^d(t) + \big[\hat{D}_d - \int_0^\tau \widehat{Q}_d(t)\overline{Z}^T(t;\hat{\alpha}) d\hat{\Lambda}_0(t)\big]\hat{\Omega}^{-1} \int_0^\tau (Z_i - \overline{Z}(t;\hat{\alpha})) d\widehat{M}_i^d(t).$$

(2) $\{\sqrt{n}(\widehat{m}_{0d}(t) - m_0(t)),\ 0 \le t \le \tau\}$ *converges weakly to a zero-mean gaussian process whose covariance function at* (s, t) *can be consistently estimated by*

$$\widehat{\Gamma}_d(s,t) = \frac{1}{n}\sum_{i=1}^n \widehat{\psi}_i^d(s)\widehat{\psi}_i^d(t),$$

where

$$\widehat{\psi}_i^d(t) = \widehat{\Phi}_d(t)^{-1}\Big\{\widehat{M}_i^*(t) + \int_0^\tau \frac{\hat{R}_d(t,u)}{S^{(0)}(u;\alpha)} d\widehat{M}_i^d(u) + \Big[\hat{R}_1(t) - \int_0^\tau \hat{R}_d(t,u)\overline{Z}^T(u;\hat{\alpha}) d\hat{\Lambda}_0(u)\Big]\hat{\Omega}^{-1}\int_0^\tau (Z_i - \overline{Z}(u;\hat{\alpha})) d\widehat{M}_i^d(u)\Big\} - \overline{X}_d(t, \widehat{\beta}_d)\hat{m}_{0d}(t)\hat{B}_d^{-1}\hat{\xi}_i^d.$$

Simultaneous confidence bands for $\hat{m}_{0d}(t)$ on some interval $[t_1, t_2]$ can be obtained as for $\hat{m}_0(t)$.

Proofs are deferred to the Appendix.

3. Efficiency Study

In this section, we discuss the efficiency and double robustness of the proposed estimators. For simplicity, we focus on the case when C is independent of the covariates. When C depends on the covariates, similar method can be used to study these issues. This is more complicated and we do not discuss it here.

3.1. An improved estimator and its efficiency

We focus on the issue of efficiency based on semiparametric efficiency theory (Bickel et al. (1993) and Robins, Rotnitzky, and Zhao (1994)). As stated in Tsiatis (2006), the most efficient estimator can be obtained in theory for regular and asymptotically linear estimators. However, in applications it is usually difficult to derive it. Fortunately, when data is monotone coarsened, there is an explicit form of the most efficient estimator. Here the most efficient estimator is the estimator that achieves the semiparametric efficiency bound.

With length-biased data, the true unbiased $\widetilde{T}$ cannot be observed, the data obtained is biased T. Then based on the biased data (A, T), the observation $(A, Y, \delta), Y = T \wedge (A + C), \delta = I(T \leq (A + C))$ can be handled as the usual right censored data since A is always observed. Following the theory in Tsiatis (2006), right censored data is monotone coarsening data. Hence, making use of the relationship between the biased T and the unbiased $\tilde{T}$, we can use the theory of monotone coarsening in Tsiatis (2006) (Theorems 10.1 and 10.4) to derive the most efficient estimator in our setting.

A coarsening variable $\tilde{C}$ can be defined as follows. For $0 < r < \infty$, $\tilde{C} = r$ is defined as $C = r, T - A > r$, and $G_r(A, T) = \{A, TI(T - A \leq r)\}$; $\tilde{C} = \infty$ is defined as $T - A \leq C$ and $G_\infty(A, T) = \{A, T\}$; for $r \geq T - A$, let $G_r(A, T) = G_\infty(A, T)$. In this way, the observed data is $(\tilde{C}, G_{\tilde{C}}(T))$, which is equivalent to

$$\{\tilde{C} = r, G_r(A, T)\} = \{A, C = r, T - A > r\},$$
$$\{\tilde{C} = \infty, G_\infty(A, T)\} = \{A, T, T - A \leq C\},$$

and so length-biased right-censored data is monotone coarsening.

After Theorem 2, our proposed estimator has an asymptotically linear representation,

$$\sqrt{n}(\widehat{\beta} - \beta_0) = B^{-1} \frac{1}{\sqrt{n}} U(\beta_0) + o_p(1). \tag{3.1}$$

In addition, it can be verified that the conditions in Theorem 2.2 of Newey (1990) on regular estimators are satisfied. Therefore, $\hat{\beta}$ is a regular estimator.

Let $D_i = (1/T_i) \int_0^\tau I(T_i > t)\big(X_i - \overline{x}(t)\big)\big[(T_i - t) - m_0(t) \exp(\beta_0^T X_i)\big] dh(t)$, $M_i^C(t) = I(Y_i - A_i \leq t, \delta_i = 0) - \int_0^t I(Y_i - A_i \geq s) d\Lambda_C(s)$, where $\Lambda_C(t)$ is the cumulative hazard function of the censoring variable C.

By the martingale integral representation of the Kaplan-Meier estimator (Gill (1980)), we can write

$$\frac{S_C(t) - \widehat{S}_C(t)}{S_C(t)} = \int_0^t \frac{\widehat{S}_C(u-)}{S_C(u)} \frac{\sum_{i=1}^n dM_i^C(u)}{n\widehat{\pi}(u)}. \tag{3.2}$$

Based on the results of Robins and Rotnitzky (1992), we have

$$\frac{\delta_i}{S_C(Y_i - A_i)} = 1 - \int_0^\tau \frac{dM_i^C(t)}{S_C(t)}. \tag{3.3}$$

Using (3.2) and (3.3), we can write

$$U(\beta_0) = \sum_{i=1}^n \frac{\delta_i D_i}{\widehat{S}_C(Y_i - A_i)} + o_p(n^{-1/2})$$

$$= \sum_{i=1}^{n} D_i - \sum_{i=1}^{n} \int_0^{\tau} \big[D_i - K(t,D)\big] \frac{dM_i^C(t)}{S_C(t)} + o_p(n^{-1/2}), \qquad (3.4)$$

where $K(t,D) = E[DI(T - A \geq t)]/S_V(t)$ and $S_V(t)$ is the survival function of V.

Combining (3.1) and (3.4), and based on the results of Robins and Rotnitzky (1992), the class of all the influence functions for the parameter β is

$$B^{-1}\Big\{D_i - \int_0^{\tau} \big[D_i - K(t,D)\big] \frac{dM_i^C(t)}{S_C(t)} + \int_0^{\tau} [e(u,X_i) - K(u,e)] \frac{dM_i^C(u)}{S_C(t)}\Big\},$$

where $e(u,X)$ is an arbitrary function of X and $K(u,e)$ is defined similarly as $K(t,D)$. Following the theory of semiparametric efficiency in Tsiatis (2006) (Theorems 10.1 and 10.4) involving censoring, when $e(u,X_i) = E\big[D_i|T_i - A_i \geq t, X_i, A_i\big]$, the corresponding estimator is the most efficient estimator in the sense of achieving the semiparametric efficiency bound.

Although this theoretically elegant, in general, it is virtually impossible to derive this conditional expectation through it, or even to get an estimator of it without additional assumptions on the covariates. There are two strategies to deal with this problem (Bang and Tsiatis (2000)): posit a model for the conditional expectation, or find an improved estimator, which means getting a more efficient estimator. In the first strategy it is difficult to select a correct model for the conditional expectation, and when it is incorrect, the estimator is not most efficient, although it is still consistent and asymptotically normal.

We now describe a way to obtain a more efficient estimator. Let $e(t,X)$ be a given q-dimensional function. Consider the class of influence functions

$$B^{-1}\Big\{D_i - \int_0^{\tau} \big[D_i - K(t,D)\big] \frac{dM_i^C(t)}{S_C(t)} + \gamma \int_0^{\tau} [e(u,X_i) - K(u,e)] \frac{dM_i^C(u)}{S_C(t)}\Big\}, \qquad (3.5)$$

where γ is a $p \times q$-dimensional constant matrix. Choose an optimal parameter γ that makes the corresponding estimator of β achieve the minimal variance in this class. Note that the first term is independent of the last two terms in (3.5). Thus finding the optimal γ turns out to be a regression problem.

Let

$$\Sigma_1 = E\Big[\int_0^{\tau} (D - K(t,D))(e(t,X) - K(t,e))^T I(Y - A \geq t) \frac{d\Lambda^C(t)}{S_C^2(t)}\Big],$$

$$\Sigma_2 = E\Big[\int_0^{\tau} (e(t,X) - K(t,e))^{\otimes 2} I(Y - A \geq t) \frac{d\Lambda^C(t)}{S_C^2(t)}\Big].$$

From least squares theory of linear regression, the optimal $\gamma_{opt} = \Sigma_1 \Sigma_2^{-1}$. The covariance of the corresponding optimal influence function is $B^{-1}\big\{Var(D) +$

$Var(\int_0^\tau [D-K(t,D)]dM^C(t)/S_C(t))-\Sigma_1\Sigma_2^{-1}\Sigma_1^T\}B^{-1}$, which is smaller than the covariance of the influence function of $\widehat{\beta}$, $B^{-1}\{Var(D)+Var(\int_0^\tau [D-K(t,D)]$ $dM^C(t)/S_C(t))\}B^{-1}$. Therefore, we obtain the most efficient estimator in this class of influence functions (3.5) for fixed q-dimensional $e(t,X)$. Here we used $e(t,X)=X$.

For general $e(t,X)$, we can estimate γ by $\widehat{\gamma}=\widehat{\Sigma}_1\widehat{\Sigma}_2^{-1}$, where

$$\widehat{\Sigma}_1=\frac{1}{n}\sum_{i=1}^n\int_0^\tau\frac{\delta_i}{\widehat{S}_C(Y_i-A_i)}(\widehat{D_i}(\widehat{\beta})-\widehat{K}(t,D))(X_i-\widehat{K}(t,X))^T I(Y_i-A_i\geq t)\frac{d\widehat{\Lambda}^C(t)}{\widehat{S}_C^2(t)},$$

$$\widehat{\Sigma}_2=\frac{1}{n}\sum_{i=1}^n\int_0^\tau(X_i-\widehat{K}(t,X))^{\otimes 2}I(Y_i-A_i\geq t)\frac{d\widehat{\Lambda}^C(t)}{\widehat{S}_C^2(t)},$$

$$\widehat{D_i}(\beta)=\frac{1}{T_i}\int_0^\tau I(T_i>t)\big(X_i-\overline{X}(t,\beta)\big)\big[(T_i-t)-\widehat{m}_0(t)\exp(\beta^TX_i)\big]dH(t),$$

and

$$\widehat{K}(t,e)=\frac{\sum e(t,X_i)I(Y_i-A_i\geq t)}{\sum I(Y_i-A_i\geq t)}.$$

In the case of $e(t,X)=X$, we denote the corresponding estimator by $\widehat{\beta}_{imp}$, and it is the solution of the estimating equation

$$U_{imp}(\beta)=\sum_{i=1}^n\Big[\frac{\delta_i\widehat{D_i}(\beta)}{\widehat{S}_C(Y_i-A_i)}+\widehat{\gamma}\int_0^\tau\big(X_i-\widehat{K}(t,X)\big)\frac{dN_i^C(t)}{\widehat{S}_C(t)}\Big]=0. \qquad (3.6)$$

The variance of $\sqrt{n}(\widehat{\beta}_{imp}-\beta_0)$ can be consistently estimated by $\widehat{B}_1^{-1}\widehat{\Sigma}_{imp}\widehat{B}_1^{-1}$, where

$$\widehat{B}_1=-\frac{1}{n}\frac{\partial U_{imp}(\widehat{\beta}_{imp})}{\partial\beta^T}=-\frac{1}{n}\frac{\partial U(\widehat{\beta}_{imp})}{\partial\beta^T},\quad \widetilde{K}(t,D)=\frac{1}{n}\sum_{i=1}^n\frac{\delta_i\widehat{D_i}(\beta)I(T_i-A_i\geq t)}{\widehat{S}_C(T_i-A_i)\widehat{S}_V(t)},$$

$$\widehat{\Sigma}_{imp}=\frac{1}{n}\sum_{i=1}^n\frac{\delta_i\widehat{D}_i(\widehat{\beta}_{imp})^{\otimes 2}}{\widehat{S}_C(Y_i-A_i)}+\frac{1}{n}\sum_{i=1}^n\int_0^\tau\big\{\widetilde{K}(t,\widehat{D}^{\otimes 2}(\widehat{\beta}_{imp}))-\widetilde{K}(t,\widehat{D}(\widehat{\beta}_{imp}))^{\otimes 2}\big\}\frac{dN_i^C(t)}{\widehat{S}_C^2(t)}-\widehat{\Sigma}_1\widehat{\Sigma}_2^{-1}\widehat{\Sigma}_1^T.$$

3.2. Double robustness property

The concept of double robust estimation was first introduced by Scharfstein, Rotnitzky, and Robins (1999) and was further studied by Robins (1999), Robins, Rotnitzky, and Van der Laan (2000), and Lunceford and Davidian (2004). In a censored data model, an estimator is double robust if the estimator is consistent when either the censoring distribution or the distribution of the failure time is correctly specified. Because with observational data, one can never be sure that the assumed censoring mechanism or the distribution of the failure time is correct, the best one can hope for is to find a double robust estimator. In this section, we show that the most efficient estimator given above is double robust.

As discussed, when $e(t, X_i) = E[D_i|T_i - A_i \geq t, X_i, A_i]$, the corresponding estimator is most efficient. The most efficiency estimator is based on the estimating function

$$\frac{\delta_i D_i}{S_C(Y_i - A_i)} + \int_0^\tau E[D_i|T_i - A_i \geq t, X_i, A_i] \frac{dM_i^C(t)}{S_C(t)}. \tag{3.7}$$

Take $Q_T(t, X, A) = -E[D|T - A \geq t, X, A] = (S_T(A+t|X, A))^{-1} \int_{A+t}^\tau D(u, X)\, dS_T(u|X, A)$, where $S_T(u|X, A)$ is the conditional survival function of T given X and A. In (3.7) both $S_C(t)$ and $S_T(t|X, A)$ are unknown. In order to obtain the estimator, we have to specify models for $S_C(t)$ and $S_T(t|X, A)$, respectively. If the models are not correctly specified, the estimator is biased.

Assume that a specified model for the survival function of the censoring variable C is $\widetilde{S}_C(t)$ (it may depend on the covariate, but, we use this notation here) and a specified model for the survival function of T given (X, A), is $\widetilde{S}_T(t|X, A)$. If the survival function of C is correctly specified, then $\widetilde{S}_C(t) = S_C(t)$; similarly if the survival function of T is correctly specified, then $\widetilde{S}_T(t|X, A) = S_T(t|X, A)$. We show in the Appendix that the most efficient estimator has the double robustness property. By double robustness with monotone coarsening theory in Tsiatis (2006, p.248), the most efficient estimator with estimating function (3.7), is double robust, that is, it is consistent if either $\widetilde{S}_T(t|X, A) = S_T(t|X, A)$ or the censoring mechanism assumption is true. In other words, in contrast to the estimator derived by simple inverse probability weighted method, i.e the solution to (2.7), the most efficient estimator can provide a double protection to obtain a consistent estimator. More importantly, the double robustness estimator with estimating function (3.7) does not lead to efficiency loss, since it belongs to the class of the augmented inverse probability complete weight estimators, see Van der Laan and Robins (2003), Rotnitzky and Robins (2005).

4. Simulation Studies

We conducted simulation studies to evaluate the proposed methods. We considered a proportional mean residual life model with two covariates $X = (X_1, X_2)$, where $X_1 \sim$ Bernoulli(0.5), X_2 was uniform on $[0, 1]$. Two simulation studies were carried out, the first focusing on the case in which the censoring variable C is independent of Z; the second on the case in which C depends on Z. We compared our procedures with that in Chan, Chen, and Di (2012).

Simulation study 1. The censoring variable measured from the enrollment time, C, was independently generated from a uniform on $[0, u_c]$, where u_c was specified to yield the desired censoring percentages, 10% and 30%. The initial time A was uniform on $[0, u_a]$. Here u_a was chosen to be larger than the upper bound of $\tilde{T}$ to ensure the stationarity assumption, as in Qin and Shen (2010).

The baseline mean residual life function $m_0(t)$ has a linear mean residual life of the Hall-Wellner family, $m_0(t) = (D_1 t + D_2)^+$, where $D_1 > -1$, $D_2 > 0$ and $d^+ = dI(d \geq 0)$ for any real number d. Here, we considered two cases: $D_1 = -0.5, D_2 = 0.5$ and $D_1 = -0.5, D_2 = 1$, corresponding to a uniform distribution on $[0, 1]$ and a random variable with survival function $S(t) = 1 - 0.5t,\ 0 \leq t \leq 2$, respectively, when $X = (0, 0)$. Let $m_0^{(1)}(t) = (0.5 - 0.5t)^+$ and $m_0^{(2)}(t) = (1 - 0.5t)^+$. Each study consisted of 1,000 replications. The sample sizes $n = 200$ and $n = 300$ were used.

We considered two choices for the weight function, $H_1(t) = t$, for which the integral in (2.7) is the area beneath the curve, and $H_2(t) = (1/n)\sum_{i=1}^n I(Y_i \leq t, \delta_i = 1)$, for which the integral in (2.7) is a weighted sum of integrand over all the jump points of the uncensored observations.

The true value of β_0 was taken to be (0, 0) and (0.2, 0.4). Tables 1-4 present the simulation results of $\widehat{\beta}$ and $\widehat{\beta}_{imp}$, where $\widehat{\beta}$ is the estimator obtained from (2.7), and $\widehat{\beta}_{imp}$ is derived from (3.6). In all these tables, Bias is the bias of the estimators; se is the standard error of the estimators; sd is the mean of the standard deviation of the estimators; CP% denotes the 95% empirical coverage probability.

Table 1 shows the results when the baseline function is $m_0^{(1)}(t)$ and the weight function is $H_1(t) = t$. Both estimators behave well. The confidence intervals based on these estimators have an empirical coverage probability close to the nominal 95% level. There is a good agreement between the estimated standard deviation and the empirical standard deviation. In addition, as expected, $\widehat{\beta}_{imp}$ performs better than $\widehat{\beta}$, in the sense that it has a smaller variance. Table 2 presents similar results for $m_0^{(1)}(t)$ and $H_2(t)$. In Table 3, the results for $m_0^{(2)}(t)$ and $H_1(t)$ are provided. For $m_0^{(2)}(t)$ and $H_2(t)$, see Table 4. Here the estimators perform well and have a reasonable 95% empirical coverage probability. As expected, the performance of the estimators improves as the sample size increases from 200 to 300. The results are similar for the two weight functions $H_1(t)$ and $H_2(t)$ in every case.

We have shown that the estimator obtained from the estimating function (3.7) is double robust. The estimating equation based on (3.7) is

$$U_{double}(\beta) = \sum_{i=1}^{n} \frac{\delta_i D(Y_i, X_i, \beta)}{\widehat{S}_C(Y_i - A_i)} - \int_0^{\tau} \widehat{Q}_T(t, X_i, A_i, \beta) \frac{dM_i^C(t)}{\widehat{S}_C(t)} = 0, \qquad (4.1)$$

Table 1. Simulations results based on 1,000 runs in the case of $H_1(t)$ and $m_0^{(1)}(t)$.

β_0	c%		Bias	se	sd	CP%
			$n = 200$			
$(0, 0)$	10	$\widehat{\beta}$	(0.0051, 0.0010)	(0.0799, 0.1325)	(0.0783, 0.1348)	(95.3, 95.3)
		$\widehat{\beta}_{imp}$	(0.0049, 0.0011)	(0.0789, 0.1316)	(0.0775, 0.1335)	(95.3, 95.4)
		$\widehat{\beta}_{double}$	(0.0048, 0.0012)	(0.0776, 0.1296)	(0.0765, 0.1321)	(95.4, 95.7)
		$\widehat{\beta}_{chan}$	(0.0045, 0.0019)	(0.0787, 0.1351)	(0.0787, 0.1357)	(95.3, 95.4)
	30	$\widehat{\beta}$	(0.0055, 0.0065)	(0.0961, 0.1616)	(0.0909, 0.1565)	(93.6, 93.9)
		$\widehat{\beta}_{imp}$	(0.0062, 0.0085)	(0.0915, 0.1519)	(0.0858, 0.1478)	(92.7, 94.6)
		$\widehat{\beta}_{double}$	(0.0070, 0.0073)	(0.0848, 0.1388)	(0.0825, 0.1423)	(94.7, 95.1)
		$\widehat{\beta}_{chan}$	(0.0065, 0.0042)	(0.0850, 0.1438)	(0.0840, 0.1449)	(94.6, 95.2)
$(0.2, 0.4)$	10	$\widehat{\beta}$	(0.0012, -0.0008)	(0.0660, 0.1082)	(0.0622, 0.1045)	(93.3, 93.3)
		$\widehat{\beta}_{imp}$	(0.0014, -0.0006)	(0.0656, 0.1070)	(0.0619, 0.1040)	(93.4, 93.6)
		$\widehat{\beta}_{double}$	(0.0013, -0.0020)	(0.0653, 0.1058)	(0.0614, 0.1032)	(93.7, 93.7)
		$\widehat{\beta}_{chan}$	(0.0018, -0.0009)	(0.0749, 0.1272)	(0.0714, 0.1220)	(93.0, 94.0)
	30	$\widehat{\beta}$	(0.0007, 0.0027)	(0.0735, 0.1220)	(0.0695, 0.1170)	(93.7, 94.1)
		$\widehat{\beta}_{imp}$	(0.0006, 0.0045)	(0.0717, 0.1181)	(0.0674, 0.1135)	(93.3, 93.8)
		$\widehat{\beta}_{double}$	(0.0001, 0.0002)	(0.0695, 0.1151)	(0.0655, 0.1101)	(93.8, 93.7)
		$\widehat{\beta}_{chan}$	(0.0015, 0.0008)	(0.0805, 0.1366)	(0.0759, 0.1301)	(93.2, 93.8)
			$n = 300$			
$(0, 0)$	10	$\widehat{\beta}$	(0.0022, -0.0050)	(0.0644, 0.1130)	(0.0639, 0.1103)	(94.9, 94.1)
		$\widehat{\beta}_{imp}$	(0.0020, -0.0049)	(0.0638, 0.1110)	(0.0633, 0.1093)	(94.8, 94.3)
		$\widehat{\beta}_{double}$	(0.0023, -0.0056)	(0.0631, 0.1097)	(0.0625, 0.1079)	(95.5, 94.0)
		$\widehat{\beta}_{chan}$	(0.0022, -0.0025)	(0.0647, 0.1132)	(0.0642, 0.1107)	(95.1, 93.7)
	30	$\widehat{\beta}$	(0.0019, -0.0079)	(0.0800, 0.1350)	(0.0746, 0.1286)	(92.6, 93.3)
		$\widehat{\beta}_{imp}$	(0.0014, -0.0067)	(0.0748, 0.1266)	(0.0703, 0.1213)	(93.7, 94.5)
		$\widehat{\beta}_{double}$	(0.0013, -0.0058)	(0.0698, 0.1190)	(0.0675, 0.1164)	(94.5, 94.2)
		$\widehat{\beta}_{chan}$	(0.0025, -0.0043)	(0.0699, 0.1208)	(0.0685, 0.1185)	(94.5, 95.4)
$(0.2, 0.4)$	10	$\widehat{\beta}$	(0.0015, 0.0045)	(0.0517, 0.0896)	(0.0508, 0.0859)	(95.0, 93.9)
		$\widehat{\beta}_{imp}$	(0.0017, 0.0043)	(0.0515, 0.0889)	(0.0506, 0.0854)	(94.7, 93.8)
		$\widehat{\beta}_{double}$	(0.0015, 0.0039)	(0.0509, 0.0879)	(0.0502, 0.0848)	(94.8, 94.1)
		$\widehat{\beta}_{chan}$	(0.0021, 0.0051)	(0.0581, 0.1057)	(0.0582, 0.1000)	(94.9, 94.1)
	30	$\widehat{\beta}$	(0.0018, 0.0048)	(0.0582, 0.1012)	(0.0568, 0.0962)	(94.5, 93.9)
		$\widehat{\beta}_{imp}$	(0.0023, 0.0048)	(0.0564, 0.0972)	(0.0551, 0.0934)	(95.2, 93.9)
		$\widehat{\beta}_{double}$	(0.0014, 0.0039)	(0.0547, 0.0937)	(0.0536, 0.0906)	(94.9, 93.8)
		$\widehat{\beta}_{chan}$	(0.0023, 0.0077)	(0.0625, 0.1126)	(0.0620, 0.1064)	(94.6, 94.3)

Note: c%: the censoring %; Bias: the bias of the estimators; se: the standard error of the estimators; sd: the mean of the standard deviation of the estimators; CP%: the empirical 95% covering probability.

where $\widehat{S}_C(\cdot)$ is the Kaplan-Meier estimator for censoring variable C, $\widehat{S}_{\tilde{T}}(t|X_i)$ is the survival function of the true failure time $\tilde{T}$ given X_i, which can be estimated

Table 2. Simulations results based on 1,000 runs in the case of $H_2(t)$ and $m_0^{(1)}(t)$.

β_0	c%		Bias	se	sd	CP%
			$n=200$			
$(0,0)$	10	$\widehat{\beta}$	(0.0031, 0.0026)	(0.0733, 0.1271)	(0.0797, 0.1371)	(96.4, 96.9)
		$\widehat{\beta}_{imp}$	(0.0030, 0.0026)	(0.0725, 0.1264)	(0.0728, 0.1253)	(94.8, 94.8)
	30	$\widehat{\beta}$	(0.0038, 0.0100)	(0.0940, 0.1633)	(0.1061, 0.1811)	(97.2, 96.4)
		$\widehat{\beta}_{imp}$	(0.0046, 0.0122)	(0.0905, 0.1558)	(0.0846, 0.1452)	(92.9, 93.4)
$(0.2, 0.4)$	10	$\widehat{\beta}$	(0.0018, 0.0004)	(0.0565, 0.0915)	(0.0561, 0.0936)	(94.9, 94.9)
		$\widehat{\beta}_{imp}$	(0.0019, 0.0006)	(0.0564, 0.0908)	(0.0534, 0.0893)	(93.0, 94.6)
	30	$\widehat{\beta}$	(0.0020, 0.0040)	(0.0667, 0.1076)	(0.0692, 0.1155)	(95.2, 95.4)
		$\widehat{\beta}_{imp}$	(0.0020, 0.0055)	(0.0657, 0.1052)	(0.0613, 0.1026)	(92.6, 94.4)
			$n=300$			
$(0,0)$	10	$\widehat{\beta}$	(0.0008,-0.0044)	(0.0597, 0.1055)	(0.0638, 0.1098)	(96.6, 96.5)
		$\widehat{\beta}_{imp}$	(0.0006,-0.0043)	(0.0593, 0.1040)	(0.0592, 0.1021)	(95.3, 95.2)
	30	$\widehat{\beta}$	(0.0003,-0.0090)	(0.0786, 0.1335)	(0.0862, 0.1469)	(96.5, 96.4)
		$\widehat{\beta}_{imp}$	(-0.0002,-0.0078)	(0.0744, 0.1270)	(0.0693, 0.1194)	(93.6, 93.1)
$(0.2, 0.4)$	10	$\widehat{\beta}$	(0.0007, 0.0045)	(0.0448, 0.0764)	(0.0450, 0.0757)	(95.1, 94.1)
		$\widehat{\beta}_{imp}$	(0.0008, 0.0043)	(0.0447, 0.0761)	(0.0433, 0.0729)	(94.3, 92.9)
	30	$\widehat{\beta}$	(0.0014, 0.0059)	(0.0529, 0.0913)	(0.0556, 0.0935)	(95.8, 94.8)
		$\widehat{\beta}_{imp}$	(0.0019, 0.0061)	(0.0517, 0.0886)	(0.0499, 0.0842)	(93.9, 93.5)

by

$$\widehat{S}_{\tilde{T}}(t|X_i,\beta) = \frac{\hat{m}_0(0,\beta)}{\hat{m}_0(t,\beta)} \exp\Big\{ -\int_0^t \frac{1}{\hat{m}_0(s,\beta)} ds \exp(-\beta^T X_i)\Big\},$$
$$\widehat{Q}_T(t,X_i,A_i,\beta) = \frac{1}{\widehat{S}_{\tilde{T}}(t|X_i,\beta)} \int_{a+t}^{\tau} D(u,X_i,\beta) d\widehat{S}_{\tilde{T}}(t|X_i,\beta). \qquad (4.2)$$

By making use the relationship between $S_T(t|X)$ and $S_{\tilde{T}}(t|X_i,\beta)$, $Q_T(t,X_i,A_i,\beta)$ can be estimated based on (4.2). The estimator of β is denoted as $\widehat{\beta}_{double}$. The results are shown in Table 1. The estimator is unbiased and has the smallest variance. When both the censoring model and the failure time model are correct, the double robustness estimator is semiparametric efficient, meaning it has the smallest asymptotic variance among all the estimators obtained with a given weight function $H(t)$.

Our method can also be used to estimate the baseline mean residual life function $m_0(t)$. Pointwise confidence intervals can be constructed and simultaneous confidence bands can also be obtained based on resampling. For simplicity, we only give the results for $m_0(t) = m_0^{(1)}(t)$, $H_1(t) = t$ and $n = 200$. Figure 1 shows the 95% pointwise confidence intervals and the 95% confidence bands. The solid line denotes the true curve of $m_0(t)$, the dotted dash line represents

Table 3. Simulations results based on 1,000 runs in the case of $H_1(t)$ and $m_0^{(2)}(t)$.

β_0	c%		Bias	se	sd	CP%
			$n = 200$			
$(0, 0)$	10	$\widehat{\beta}$	(0.0019, -0.0050)	(0.0789, 0.1371)	(0.0781, 0.1348)	(94.6, 94.2)
		$\widehat{\beta}_{imp}$	(0.0019, -0.0055)	(0.0775, 0.1360)	(0.0773, 0.1335)	(94.6, 94.4)
	30	$\widehat{\beta}$	(0.0026, -0.0018)	(0.0964, 0.1650)	(0.0913, 0.1573)	(93.9, 94.1)
		$\widehat{\beta}_{imp}$	(0.0036, -0.0028)	(0.0898, 0.1552)	(0.0860, 0.1482)	(93.7, 94.0)
$(0.2, 0.4)$	10	$\widehat{\beta}$	(0.0005, -0.0013)	(0.0630, 0.1068)	(0.0623, 0.1048)	(94.7, 94.1)
		$\widehat{\beta}_{imp}$	(0.0004, -0.0007)	(0.0626, 0.1060)	(0.0620, 0.1043)	(95.0, 93.5)
	30	$\widehat{\beta}$	(0.0015, 0.0024)	(0.0699, 0.1205)	(0.0697, 0.1170)	(95.1, 94.1)
		$\widehat{\beta}_{imp}$	(0.0004, 0.0044)	(0.0671, 0.1165)	(0.0676, 0.1135)	(94.8, 94.9)
			$n = 300$			
$(0, 0)$	10	$\widehat{\beta}$	(0.0049, -0.0046)	(0.0641, 0.1105)	(0.0640, 0.1105)	(95.0, 94.8)
		$\widehat{\beta}_{imp}$	(0.0050, -0.0046)	(0.0636, 0.1102)	(0.0634, 0.1095)	(94.7, 94.6)
	30	$\widehat{\beta}$	(0.0043, -0.0047)	(0.0779, 0.1341)	(0.0753, 0.1299)	(93.9, 93.8)
		$\widehat{\beta}_{imp}$	(0.0049, -0.0054)	(0.0729, 0.1275)	(0.0707, 0.1220)	(93.7, 94.0)
$(0.2, 0.4)$	10	$\widehat{\beta}$	(0.0005, 0.0039)	(0.0529, 0.0882)	(0.0509, 0.0860)	(94.6, 94.9)
		$\widehat{\beta}_{imp}$	(0.0004, 0.0039)	(0.0526, 0.0880)	(0.0506, 0.0856)	(94.2, 95.1)
	30	$\widehat{\beta}$	(0.0010, 0.0054)	(0.0591, 0.0986)	(0.0568, 0.0961)	(94.0, 94.9)
		$\widehat{\beta}_{imp}$	(0.0009, 0.0066)	(0.0573, 0.0968)	(0.0551, 0.0932)	(93.7, 94.7)

Note: c%: the censoring %; Bias: the bias of the estimators; se: the standard error of the estimators; sd: the mean of the standard deviation of the estimators; CP%: the empirical 95% covering probability.

the estimated curve by the proposed approach, the dashed lines represent the 95% pointwise confidence intervals, and the dotted lines represent the simultaneous 95% confidence bands. The plot in the top panel corresponds to the result with 30% censoring rate and the plot in the bottom panel corresponds to the result with 10% censoring rate. The simultaneous confidence band is wider than the pointwise confidence interval and $m_0(t)$ can be estimated very well by our approach.

Simulation study 2. Here, the censoring variable C followed a Cox model, depending on $Z = X_1$. The corresponding hazard function was $\lambda_C(t|Z) = \lambda_0(t)\exp(\alpha^T Z)$. We took $\alpha = 2$ and $\lambda_0(t)$ a constant, determined by the censoring percentages, 10% and 30%. The correlation between C and Z was about -0.52. We only considered the case $H_1(t) = t$ and $m_0^{(1)}(t) = 0.5(1-t)^+$. All the other settings were the same as those in Simulation study 1.

Results are given in Table 5. The values of the estimator $\hat{\beta}_d$ were calculated based on (2.11) and (2.12). We also give the value of the estimator $\hat{\beta}$, obtained under the assumption that C is independent of Z, ignoring the dependence of C

Table 4. Simulations results based on 1,000 runs in the case of $H_2(t)$ and $m_0^{(2)}(t)$.

β_0	c%		Bias	se	sd	CP%
				$n=200$		
$(0,0)$	10	$\widehat{\beta}$	(0.0014, -0.0019)	(0.0724, 0.1293)	(0.0794, 0.1368)	(96.3, 96.1)
		$\widehat{\beta}_{imp}$	(0.0015, -0.0024)	(0.0715, 0.1285)	(0.0726, 0.1253)	(94.2, 94.2)
	30	$\widehat{\beta}$	(0.0027, 0.0018)	(0.0950, 0.1660)	(0.1069, 0.1812)	(96.5, 96.3)
		$\widehat{\beta}_{imp}$	(0.0033, 0.0009)	(0.0901, 0.1577)	(0.0850, 0.1459)	(93.5, 93.3)
$(0.2,0.4)$	10	$\widehat{\beta}$	(0.0005,-0.0012)	(0.0530, 0.0935)	(0.0562, 0.0936)	(96.3, 95.0)
		$\widehat{\beta}_{imp}$	(0.0004,-0.0009)	(0.0529, 0.0928)	(0.0535, 0.0894)	(95.2, 94.1)
	30	$\widehat{\beta}$	(0.0021, 0.0015)	(0.0622, 0.1104)	(0.0695, 0.1142)	(96.5, 95.3)
		$\widehat{\beta}_{imp}$	(0.0012, 0.0029)	(0.0605, 0.1083)	(0.0615, 0.1023)	(95.2, 94.1)
				$n=300$		
$(0,0)$	10	$\widehat{\beta}$	(0.0050, -0.0042)	(0.0599, 0.1039)	(0.0637, 0.1101)	(96.6, 96.4)
		$\widehat{\beta}_{imp}$	(0.0051, -0.0042)	(0.0596, 0.1033)	(0.0592, 0.1022)	(95.4, 94.5)
	30	$\widehat{\beta}$	(0.0043, -0.0052)	(0.0785, 0.1320)	(0.0874, 0.1496)	(97.3, 96.1)
		$\widehat{\beta}_{imp}$	(0.0049, -0.0058)	(0.0741, 0.1274)	(0.0697, 0.1199)	(93.5, 93.4)
$(0.2,0.4)$	10	$\widehat{\beta}$	(-0.0012, 0.0027)	(0.0456, 0.0742)	(0.0450, 0.0758)	(95.0, 95.4)
		$\widehat{\beta}_{imp}$	(-0.0012, 0.0028)	(0.0454, 0.0741)	(0.0433, 0.0731)	(94.0, 94.8)
	30	$\widehat{\beta}$	(-0.0004, 0.0040)	(0.0543, 0.0894)	(0.0553, 0.0929)	(95.6, 95.1)
		$\widehat{\beta}_{imp}$	(-0.0004, 0.0051)	(0.0530, 0.0881)	(0.0498, 0.0839)	(93.3, 93.3)

Note: c%: the censoring %; Bias: the bias of the estimators; se: the standard error of the estimators; sd: the mean of the standard deviation of the estimators; CP%: the empirical 95% covering probability.

on Z.

In Table 5, $\widehat{\beta}_d$ performs very well under each censoring percentage. It is unbiased. The estimated standard error(se) is not far from the empirical standard deviation (sd). The empirical coverage probabilities of the confidence intervals are close to the nominal 95% level. The standard deviation of $\widehat{\beta}_d$ decreases when the sample size increases from 200 to 300.

If the dependence of C on Z is ignored, the estimator $\widehat{\beta}$, which does not have the double robustness property, is biased. On the other hand, when the model of C is misspecified, the double robustness estimator should be consistent since the failure time model is correct. To verify this, we calculated the double robustness estimator $\widehat{\beta}_{double}$ by solving the estimating equation (4.1). The results are given in Table 5, which support the theoretical result.

In our simulation studies, we also compared our method with that of Chan, Chen, and Di (2012). Let $\widehat{\beta}_{chan}$ denote their estimator. For simplicity, we only provide the results when the baseline function is $m_0^{(1)}(t)$ and the weight function is $H_1(t)=t$. The results with C independent of Z are presented in Table 1,

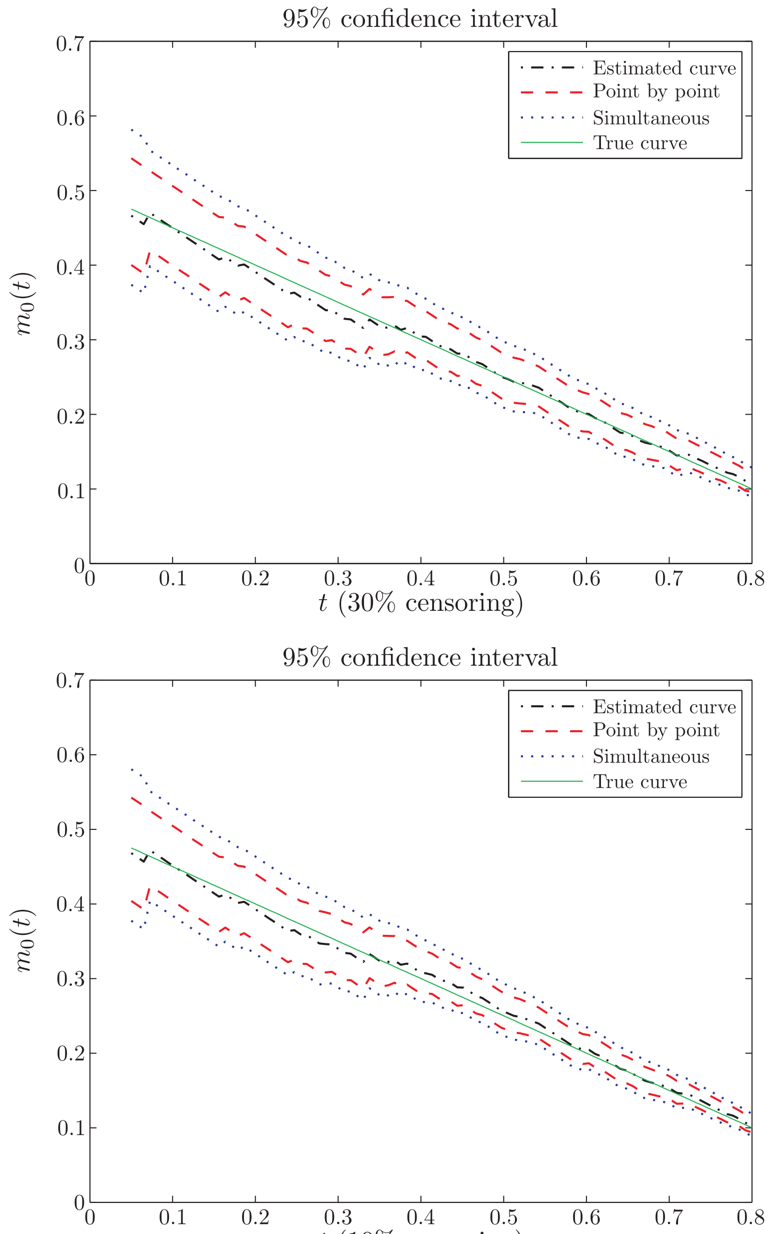

Figure 1. The estimation of $m_0(t)$ when C is independent of Z.

and with C depending on Z in Table 5. From Tables 1 and 5, we see that our estimators and $\widehat{\beta}_{chan}$ are unbiased, and they have similar performance.

When C depends on the covariate, $m_0(t)$ can also be estimated. Figure 2 shows the estimated curve, the corresponding 95% pointwise confidence intervals

Table 5. Simulations results based on 1,000 runs in the case of $H_1(t)$ and $m_0^{(1)}(t)$.

β_0	c%		Bias	se	sd	CP%
			$n = 200$			
(0,0)	10	$\widehat{\beta}_d$	(-0.0013, 0.0005)	(0.0796, 0.1397)	(0.0777, 0.1359)	(94.8, 94.5)
		$\widehat{\beta}$	(-0.0325, 0.0004)	(0.0809, 0.1387)	(0.0786, 0.1352)	(92.8, 95.0)
		$\widehat{\beta}_{double}$	(-0.0006, 0.0006)	(0.0786, 0.1361)	(0.0761, 0.1326)	(95.2, 94.7)
		$\widehat{\beta}_{chan}$	(0.0009, 0.0024)	(0.0812, 0.1410)	(0.0789, 0.1365)	(94.7, 93.5)
	30	$\widehat{\beta}_d$	(-0.0114, 0.0023)	(0.0993, 0.1776)	(0.0924, 0.1629)	(93.0, 93.2)
		$\widehat{\beta}$	(-0.1258, 0.0014)	(0.0999, 0.1582)	(0.0955, 0.1527)	(74.5, 94.4)
		$\widehat{\beta}_{double}$	(-0.0027, 0.0012)	(0.0870, 0.1486)	(0.0742, 0.1431)	(90.0, 94.1)
		$\widehat{\beta}_{chan}$	(0.0013, 0.0032)	(0.0884, 0.1538)	(0.0861, 0.1472)	(94.6, 93.8)
(0.2,0.4)	10	$\widehat{\beta}_d$	(0.0034, -0.001)	(0.0655, 0.1075)	(0.0612, 0.1045)	(93.3, 94.3)
		$\widehat{\beta}$	(-0.0130, 0.0004)	(0.0653, 0.1087)	(0.0610, 0.1055)	(91.9, 93.8)
		$\widehat{\beta}_{double}$	(0.0024, -0.0011)	(0.0646, 0.1068)	(0.0607, 0.1034)	(93.3, 94.0)
		$\widehat{\beta}_{chan}$	(0.0046, 0.0036)	(0.0741, 0.1230)	(0.0709, 0.1223)	(93.4, 94.9)
	30	$\widehat{\beta}_d$	(0.0000, 0.0051)	(0.0721, 0.1231)	(0.0675, 0.1196)	(93.5, 94.2)
		$\widehat{\beta}$	(-0.0595, 0.0060)	(0.0732, 0.1233)	(0.0686, 0.1201)	(84.5, 94.5)
		$\widehat{\beta}_{double}$	(-0.0030, 0.0009)	(0.0680, 0.1172)	(0.0595, 0.1115)	(91.7, 94.3)
		$\widehat{\beta}_{chan}$	(0.0034, 0.0014)	(0.0788, 0.1357)	(0.0763, 0.1322)	(93.5, 93.4)
			n=300			
(0,0)	10	$\widehat{\beta}_d$	(0.0038, -0.0009)	(0.0626, 0.1141)	(0.0635, 0.1110)	(95.8, 94.2)
		$\widehat{\beta}$	(-0.0279, -0.0008)	(0.0634, 0.1131)	(0.0643, 0.1105)	(93.1, 94.5)
		$\widehat{\beta}_{double}$	(0.0035, -0.0006)	(0.0619, 0.1106)	(0.0623, 0.1083)	(95.4, 94.4)
		$\widehat{\beta}_{chan}$	(0.0029, -0.0011)	(0.0629, 0.1142)	(0.0642, 0.1107)	(95.3, 94.1)
	30	$\widehat{\beta}_d$	(-0.0034, -0.0023)	(0.0785, 0.1418)	(0.0759, 0.1348)	(94.3, 92.7)
		$\widehat{\beta}$	(-0.1216, -0.0027)	(0.0785, 0.1251)	(0.0782, 0.1247)	(66.4, 94.0)
		$\widehat{\beta}_{double}$	(0.0008, -0.0014)	(0.0697, 0.1170)	(0.0607, 0.1168)	(91.8, 95.2)
		$\widehat{\beta}_{chan}$	(0.0021, -0.0002)	(0.0704, 0.1203)	(0.0700, 0.1195)	(95.4, 94.2)
(0.2,0.4)	10	$\widehat{\beta}_d$	(0.0009, -0.0038)	(0.0513, 0.0820)	(0.0503, 0.0857)	(94.1, 96.1)
		$\widehat{\beta}$	(-0.0156, -0.0028)	(0.0511, 0.0830)	(0.0502, 0.0865)	(92.8, 96.4)
		$\widehat{\beta}_{double}$	(-0.0002, -0.0039)	(0.0511, 0.0818)	(0.0499, 0.0849)	(94.6, 96.1)
		$\widehat{\beta}_{chan}$	(0.0025, -0.0021)	(0.0585, 0.1027)	(0.0579, 0.1000)	(93.6, 94.6)
	30	$\widehat{\beta}_d$	(-0.0006, -0.003)	(0.0576, 0.098)	(0.0553, 0.0982)	(94.5, 94.5)
		$\widehat{\beta}$	(-0.0620, -0.0005)	(0.0588, 0.0977)	(0.0563, 0.0983)	(79.7, 95.3)
		$\widehat{\beta}_{double}$	(-0.0047, -0.0049)	(0.0547, 0.0915)	(0.0488, 0.0916)	(92.5, 95.2)
		$\widehat{\beta}_{chan}$	(0.0023, -0.0043)	(0.0620, 0.1118)	(0.0623, 0.1082)	(94.5, 94.5)

Note: c%: the censoring %; Bias: the bias of the estimators; se: the standard error of the estimators; sd: the mean of the standard deviation of the estimators; CP: the empirical 95% covering probability.

and the 95% simultaneously confidence bands. The legends in the figure are the same as those in Figure 1. Our approach provides a good estimate of $m_0(t)$, and

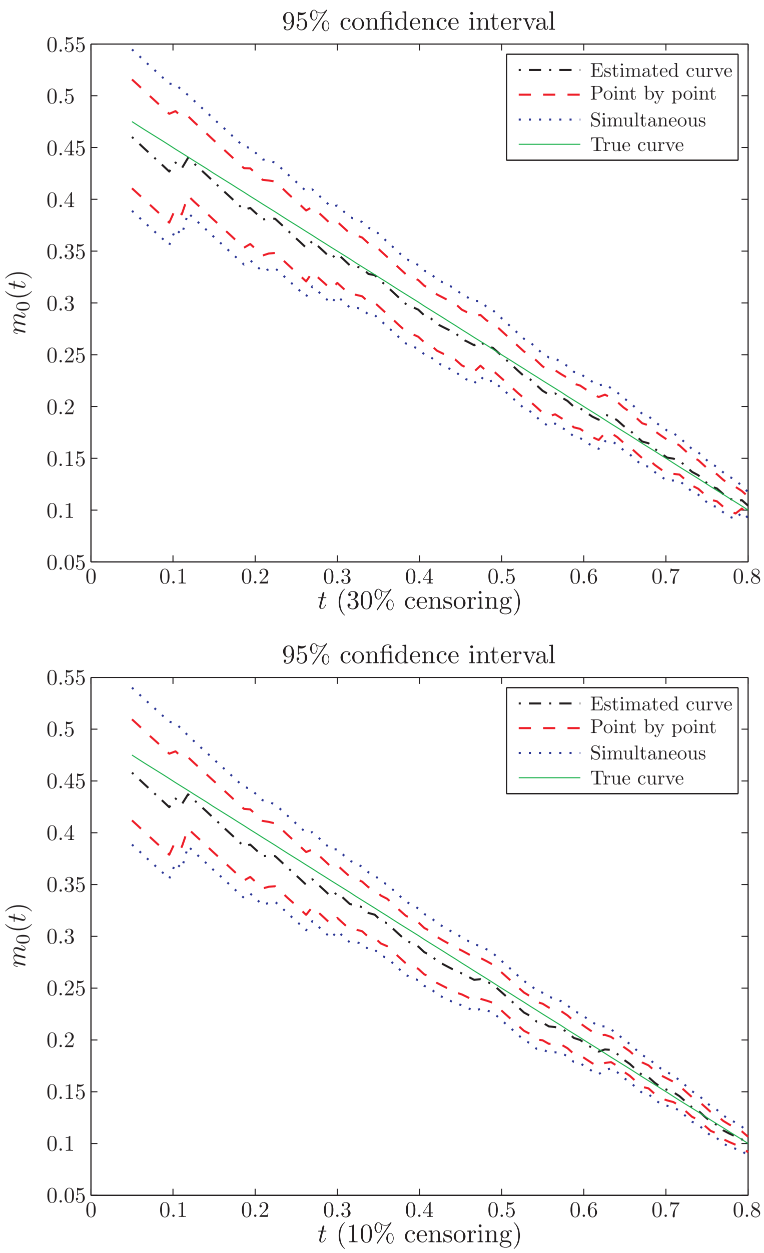

Figure 2. The estimation of $m_0(t)$ when C depends on the covariates.

as expected, the simultaneously confidence band is wider than the point by point confidence interval.

Table 6. Covariate effects on mean residual life.

	Adjusted method(se)	Naive method(se)	Improved estimator(se)
male	-0.0172 (0.0229)	-0.0455 (0.0294)	-0.0156 (0.0228)

Note: Adjusted method: adjust the length bias; Naive method: ignoring the length bias; Improved estimator: obtained by the improved method proposed in this paper.

5. Data Example

In this section, we illustrate our method using the Channing House data. A description of this data set can be found in Klein and Moeschberger (1997). Channing House is a retirement center located in Palo Alto, California. The data were collected from January 1964 to July 1975, with 97 male and 365 female individuals. For each individual, the age at entry and leaving or death were recorded. During the study, 46 men and 130 women died at Channning House. Since an individual must survive to a sufficient age to enter the retirement center, the data are left truncated and right censored. Here, truncation times were the ages in years when individuals entered the retirement community.

We used a subsample of this data set with individuals who lived longer than 786 months (65.5 years old), consisting of 448 individuals. This subsample is a length-biased data set satisfying the stationarity assumption. For this data set, the stationarity assumption was formally checked by the methods in Addona and Wolfson (2006) and Asgharian, Wolfson, and Zhang (2006). We took gender as the covariate X and evaluated its effect on mean residual life.

To gain some insights into the dependence of C and X, we assumed that C followed a Cox proportional hazard model. The covariate effect was not significant with p-value 0.898. Notice that there is a possibility that C does not follow a Cox model and C depends on X in some other model structures. For purposes of illustration, we assumed a Cox model with C correct for the time being.

The analysis results are given in Tables 6−7. Table 6 shows the estimated covariate effect based on the methods with and without the length-biased adjustment. The Adjusted method is the simple inverse probability weighted method described in Section 2 with $H(t) = t$, that adjusts the length bias, the Naive method ignores the length bias, and the Improved estimator is the estimator obtained by the improved method proposed in the efficiency study. From Table 6-7, the results from the Adjusted method and the Improved estimator are almost the same. It is clear from the results that the effect of gender on mean residual lifetime is not statistically significant.

Comparing the results from the naive method, we found a weaker effect between gender and residual life time, even though both methods yield statistically insignificant differences. Thus, the mean residual life times $m(T - t_0|T > t_0, X)$ at given years $t_0 = 70, 75, 80, 85, 90, 95$ are listed in Table 7. For example, for a

Table 7. Estimated mean residual life $m(T - t_0|T > t_0, X)$.

	t_0(years)					
	70	75	80	85	90	95
Adjusted method:						
female	13.6	9.4	6.2	4.4	3.4	3.0
male	13.3	9.3	6.1	4.4	3.3	3.0
Naive method:						
female	18.1	13.6	9.6	6.8	5.0	3.5
male	17.3	13.0	9.1	6.5	4.8	3.3
Improved estimator:						
female	13.5	9.4	6.2	4.4	3.4	3.0
male	13.3	9.3	6.1	4.4	3.3	3.0

Note: Adjusted method: adjust the length bias; Naive method: ignoring the length bias; Improved estimator: obtained by the improved method proposed in this paper.

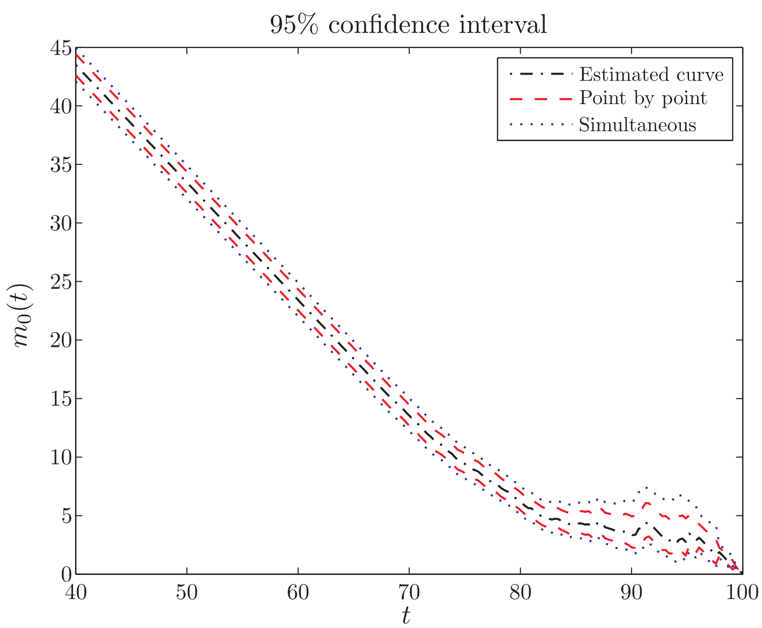

Figure 3. The estimation of $m_0(t)$ in data example.

female who has survived up to 70 years old, when ignoring the length bias, the estimated mean residual lifetime is 18.1 years, and when adjusting the length bias, the estimated mean residual lifetime is 13.6 years.

The baseline mean residual function $m_0(t)$ can also be estimated. The estimated curve, the 95% point by point confidence interval and 95% simultaneous confidence band are given in Figure 3.

6. Concluding Remarks

In this paper, we focus on proportional mean residual life model with length-biased data. Usually, the model structure for the observed data is different from that assumed for the true failure time and the censoring is informative. Making use of the inverse probability weighted approach, unbiased estimating equations are constructed to adjust the bias induced by the length bias and information censoring. In this way, the covariate effects on the mean residual life can be estimated based on the model. Moreover, we study the efficiency and double robustness of the proposed estimator. A new and more efficient estimator is derived.

There are many interesting problems about the mean residual life model. For example, with proportional mean residual life model, a natural restriction is that the derivative of $m(t|X) + t$ be monotone. It would be interesting to study how to incorporate this restriction into estimation. Another interesting issue is that when data are heavily skewed to the right or have heavy tails, the mean life may not exist or is not a reasonable summary of residual life. Here, the median residual life function may be more reasonable.

Supplementary Materials

Proofs of the main results in this paper are in the online Supplementary Material.

Acknowledgements

Bai's work was supported by National Natural Science Foundation of China (NSFC) (11501104), and Program for Innovative Research Team in UIBE. Zhou's work was supported by National Natural Science Foundation of China (NSFC) (71271128), the State Key Program of National Natural Science Foundation of China (71331006), NCMIS, Key Laboratory of RCSDS, CAS and Shanghai University of Finance and Economics through Project 211 Phase IV, Shanghai First-class Discipline A and IRTSHUFE, PCSIRT (IRT13077). We are very grateful to an associate editor and two anonymous referees for their valuable suggestions.

Appendix

We give an outline of the proofs of the main results. More detail is in the Supplementary Material. Denote the true values of β and $m(t)$ by β_0 and $m_0(t)$, respectively.

We require the following technical assumptions.

(A1) $m_0(t)$ is continuously differential in $[0, \tau]$.

(A2) $H(t)$ converges almost surely to a nonrandom and bounded function $h(t)$ uniformly in $[0, \tau]$.

(A3) X is bounded.

(A4) $B = E\Big[\int_0^\tau \mu^{-1}(X) S_{\widetilde{T}}(t|X)(X - \overline{x}(t))^{\otimes 2} m_0(t) \exp(\beta_0^T X) dh(t)\Big]$ is nonsingular where $\overline{x}(t) = \lim \overline{X}(t, \beta_0)$ is a nonrandom function.

(A5) $B_d = E\Big[\int_0^\tau \mu^{-1}(X) S_{\widetilde{T}}(t|X)(X - \overline{x}_d^*(t))^{\otimes 2} m_0(t) \exp(\beta_0^T X) dh(t)\Big]$ is nonsingular where $\overline{x}_d^*(t) = \lim \overline{X}_d^*(t, \beta_0)$ is a nonrandom function.

Proof of Theorem 1. It can be shown that $\widehat{m}_0(t, \beta)$ converges almost surely and uniformly to $m_0(t, \beta)$ in $t \in [0, \tau]$ and $\beta \in \mathcal{B} = \{\beta : ||\beta - \beta_0|| \leq \varepsilon\}$. Here, $m_0^*(t) = m_0(t, \beta_0)$. Therefore, in order to prove the existence and uniqueness of $\widehat{\beta}_1$ and $\widehat{m}_0(t)$, it suffices to show that there exists a unique solution to $U(\beta) = 0$.

Let $\widehat{A}(\beta) = -(1/n)(\partial U(\beta)/\partial \beta^T)$, which is always nonnegative definite. It can be easily shown that $\overline{X}(t, \beta)$ converges to some nonrandom function $\overline{x}(t, \beta)$ uniformly in $t \in [0, \tau]$. Together with the uniformly convergence of $\widehat{m}_0(t, \beta)$ to $m_0(t, \beta)$ in $t \in [0, \tau]$ and $\beta \in \mathcal{B} = \{\beta : ||\beta - \beta_0|| \leq \varepsilon\}$, we conclude that $\widehat{A}(\beta)$ converges uniformly to a nonrandom function $A(\beta)$ uniformly in $\beta \in \mathcal{B} = \{\beta : ||\beta - \beta_0|| \leq \varepsilon\}$. Let $B = A(\beta_0)$ and $S_{\widetilde{T}}(t|X)$ be the survival function of $\widetilde{T}$, where $A(\beta) = E\Big[\int_0^\tau \mu^{-1}(X) S_{\widetilde{T}}(t|X)(X - \overline{x}(t, \beta))^{\otimes 2} m_0(t, \beta) \exp(\beta^T X) dh(t)\Big]$.

It can be checked easily that $(1/n)U(\beta_0)$ converges to 0 almost surely. By (A4), A is nonsingular. On the other hand, $\widehat{A}(\beta)$ converges uniformly to a nonrandom function $A(\beta)$ uniformly in $\beta \in \mathcal{B} = \{\beta : ||\beta - \beta_0|| \leq \varepsilon\}$; thus there exists a small neighborhood of β_0 in which $\widehat{A}(\beta)$, especially $\widehat{A}(\beta_0)$, is nonsingular for sufficient large n. It follows from the Inverse Function Theorem(Rudin (1976))) that within a small neighborhood of β_0, there exists a unique solution $\widehat{\beta}$ to the equation $U(\beta) = 0$ when n is large enough. Furthermore by the nonnegative definiteness of $\widehat{A}(\beta)$ in the entire domain of β, the solution $\widehat{\beta}$ is global unique. Hence, there exists a unique estimator $\widehat{\beta}$ and $\widehat{m}_0(t)$, for $t \in [0, \tau]$. Following the proof of the estimator's uniqueness, we can see that $\widehat{\beta}$ is actually strongly consistent and then $\widehat{m}_0(t) = \widehat{m}_0(t, \widehat{\beta})$ converges uniformly to $m_0(t)$ almost surely in $t \in [0, \tau]$.

Proof of Theorem 2.

(1) Let $\pi(t) = P(Y - A \geq t)$ and

$$\overline{X}^*(t) = \frac{\sum \delta_i I(Y_i > t)\{Y_i \widehat{S}_C(Y_i - A_i)\}^{-1} \exp(\beta_0^T X_i) X_i}{\sum \delta_i I(Y_i > t)\{Y_i \widehat{S}_C(Y_i - A_i)\}^{-1} \exp(\beta_0^T X_i)}.$$

Let $\overline{x}^*(t)$ be the limit of $\overline{X}^*(t)$. After some calculation, we obtain

$$U(\beta_0) = \sum_{i=1}^n \int_0^\tau M_i(t)\{X_i - \overline{x}^*(t)\}dh(t) + \sum_{i=1}^n \int_0^\tau \frac{Q(t)}{\pi(t)} dM_i^C(t) + o_p(n^{1/2}),$$

where $Q(t) = (1/n)\sum_{i=1}^n \int_0^\tau M_i(s)\{X_i - \overline{x}^*(s)\}dh(s)I(Y_i - A_i \geq t)$ and $Q(t)$ converges to some nonrandom process $q(t)$. By Lemma 1 in Lin et al. (2000), we have $(1/\sqrt{n})U(\beta_0) = (1/\sqrt{n})\sum_{i=1}^n \xi_i + o_p(1)$, where $\xi_i = \int_0^\tau M_i(t)\{X_i - \overline{x}^*\}dh(t) + \int_0^\tau (q(t)/\pi(t))dM_i^C(t)$. It follows from the Multivariate Central Limit Theorem that $n^{-1/2}U(\beta_0)$ is asymptotically normal with mean 0 and covariance matrix $\Sigma = E[\xi_i^{\otimes 2}]$. By the Taylor expansion of $U(\beta)$ at β_0, we have $n^{1/2}(\widehat{\beta} - \beta_0) = B^{-1}n^{-1/2}U(\beta_0) + o_p(1)$. Therefore, $\sqrt{n}(\widehat{\beta} - \beta_0)$ is asymptotically zero-mean normal with covariance matrix $B^{-1}\Sigma B^{-1}$, which can be consistently estimated by $\widehat{B}^{-1}\widehat{\Sigma}\widehat{B}^{-1}$.

(2) We first show the weak convergence of $\widehat{m}_0(t)$. Note that

$$\sqrt{n}(\widehat{m}_0(t) - m_0(t)) = \sqrt{n}(\widehat{m}_0(t, \widehat{\beta}) - \widehat{m}_0(t, \beta_0)) + \sqrt{n}(\widehat{m}_0(t, \beta_0) - m_0(t)).$$

Let $\Phi(t) = (1/n)\sum_{i=1}^n \delta_i I(Y_i > t)\{Y_i \widehat{S}_C(Y_i - A_i)\}^{-1}\exp(\beta_0^T X_i)$ and $R(t, \mu) = (1/n)\sum_{i=1}^n M_i(t)I(Y_i - A_i \geq \mu)$. After some derivation, we obtain

$$\sqrt{n}(\widehat{m}_0(t, \widehat{\beta}) - \widehat{m}_0(t, \beta_0)) = -\overline{x}^*(t)m_0(t)\frac{1}{\sqrt{n}}\sum_{i=1}^n B^{-1}\xi_i + o_p(1),$$

$$\sqrt{n}(\widehat{m}_0(t, \beta_0) - m_0(t)) = \phi(t)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \left[M_i(t) + \int_0^\tau \frac{r(t,\mu)}{\pi(\mu)} dM_i^C(\mu)\right] + o_p(1),$$

where $\phi(t)$ and $r(t, \mu)$ is the corresponding nonrandom limit of $\Phi(t)$ and $R(t, \mu)$.

Therefore $\sqrt{n}(\widehat{m}_0(t) - m_0(t)) = (1/\sqrt{n})\sum_{i=1}^n \psi_i(t) + o_p(1)$, where

$$\psi_i(t) = \phi(t)^{-1}[M_i(t) + \int_0^\tau \frac{r(t,\mu)}{\pi(\mu)} dM_i^C(\mu)] - \overline{x}(t, \beta_0)m_0(t)B^{-1}\xi_i.$$

Since the terms in last equation are independent zero-mean random variables for every fixed t, the Multivariate Central Limit Theorem implies that the finite dimensional distributions of the process $\sqrt{n}\{\widehat{m}_0(t) - m_0(t)\}(0 \leq t \leq \tau)$ converge to those of a zero-mean Gaussian process. To prove weak convergence, it suffices to show the tightness. This reduces to the tightness of $n^{-1/2}\sum_{i=1}^n M_i(t)$. By (A1), $m_0(t)$ is of bounded variation. Since $M_i(t)$ can be written as the sum or product of monotone functions of t and is thus manageable, it follows from the functional central limit theorem that $n^{-1/2}\sum_{i=1}^n M_i(t)$ is tight. Therefore $n^{1/2}\{\widehat{m}_0(t) - m_0(t)\}(0 \leq t \leq \tau)$ is tight and converges weakly to a zero-mean Gaussian process with the covariance function $\Gamma(s,t) = E\{\psi_i(s)\psi_i(t)\}$ at (s,t), which can be consistently estimated by $\widehat{\Gamma}(s,t)$.

Proof of Theorem 3. By the maximum partial likelihood theory in Flemming and Harrington (1991), we can obtain the uniform consistency of $\hat{\alpha}$ and $\hat{\Lambda}_0(t)$ in $[0,\tau]$. Then similar arguments as that in the proof of Theorem 1 can be used to obtain the conclusion of Theorem 3. We omit the details.

Proof of Theorem 4. (1) Since C follows the Cox proportional hazards model, from Flemming and Harrington (1991), we can obtain

$$\hat{\alpha} - \alpha_0 = \Omega^{-1}\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}(Z_i - \overline{z}(t))dM_i^d(t) + o_p(n^{-1/2}),$$

$$\hat{\Lambda}_0(t) - \Lambda_0(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{t}\frac{M_i^d(u)}{s^{(0)}(u)} - \int_0^{t}\overline{z}(u)'d\Lambda_0(u)(\hat{\alpha}-\alpha_0) + o_p(n^{-1/2}), \quad (A.1)$$

where $M_i^d(t) = N_i^C(t) - \int_0^t Y_i(u)\exp(\alpha_0^T Z_i)d\Lambda_0(u)$, $s^{(k)}(t;\alpha) = \lim_{n\to\infty} S^{(k)}(t;\alpha)$, $s^{(0)}(t) = S^{(0)}(t;\alpha_0)$, $\overline{z}(t;\alpha) = \lim_{n\to\infty}\overline{Z}(t;\alpha)$, $\overline{z}(t) = \overline{z}(t;\alpha_0)$, $\Omega = \lim_{n\to\infty}\hat{\Omega}$.

Then combined with functional delta method in Van der Vaart and Wellner (1996), similar arguments as those in the proof of Theorem 2 can be used to complete the proof. We omit the details.

Proof of the double robustness of the most efficient estimator. To show double robustness, it suffices to show that

$$E\Big\{\frac{\delta_i D_i}{\widetilde{S}_C(T_i - A_i)} - \frac{(1-\delta_i)\widetilde{Q}_T(Y_i - A_i, X_i, A_i)}{\widetilde{S}_C(Y_i - A_i)} + \int_0^{Y_i - A_i}\widetilde{Q}_T(u, X_i, A_i)\frac{d\Lambda_i^C(u)}{\widetilde{S}_C(u)}\Big\} = 0,$$

if $S_C(t) = \widetilde{S}_C(t)$ or $S_T(t|X,A) = \widetilde{S}_T(t|X,A)$.

Let $\widetilde{Q}_T(t,X,A) = (1/\widetilde{S}_T(A+t|X,A))\int_{A+t}^{\tau} D(u,X)d\widetilde{S}_T(u|X,A)$. Under the assumptions on the two survival functions, (3.7) in the paper becomes

$$\frac{\delta_i D_i}{\widetilde{S}_C(T_i - A_i)} - \frac{(1-\delta_i)\widetilde{Q}_T(Y_i - A_i, X_i, A_i)}{\widetilde{S}_C(Y_i - A_i)} - \int_0^{Y_i - A_i}\widetilde{Q}_T(u, X_i, A_i)\frac{d\widetilde{S}_C(u)}{\widetilde{S}_C^2(u)}. \quad (A.2)$$

First of all,

$$E\big[\frac{\delta D}{\widetilde{S}_C(T-A)}\big|X,A\big] = -\int_a^{\tau}\frac{D(u,X)S_C(u-a)}{\widetilde{S}_C(u-A)}dS_T(u|X,A),$$

then,

$$E\Big[(1-\delta)\frac{\widetilde{Q}_T(Y-A,X,A)}{\widetilde{S}_C(Y-A)}\big|X,A\Big] = -\int\frac{\widetilde{Q}_T(c,X,A)}{\widetilde{S}_C(c)}S_T(a+c|X)dS_C(c)$$

$$= -\int_a^\tau \int_0^{u-a} \frac{S_T(a+c|X,A)}{\widetilde{S}_T(a+c|X,A)} \frac{dS_C(c)}{\widetilde{S}_C(c)} D(u,X) d\widetilde{S}_T(u|X,A),$$

and then,

$$E\Big[\int_0^{Y-A} \widetilde{Q}_T(u,X,A) \frac{d\widetilde{S}_C(u)}{\widetilde{S}_C^2(u)} \big|X,A\Big] = E\Big[\int_0^{Y-A} \frac{d\widetilde{S}_C(u)}{\widetilde{S}_C^2(u)} \widetilde{Q}_T(u,X,A)\big|X,A\Big]$$
$$= \int_a^\tau \int_0^{u-a} \frac{S_T(t+a|X,A)S_C(t)}{\widetilde{S}_C^2(t)\widetilde{S}_T(t+u|X,A)} d\widetilde{S}_C(t) D(u,X) d\widetilde{S}_T(u|X,A).$$

Since $d(S_C(t)\widetilde{S}_C(t)^{-1}) = dS_C(t)\widetilde{S}_C(t)^{-1} - S_C(t)\widetilde{S}_C^{-2}(t)d\widetilde{S}_C(t)$, then

$$E\Big\{\frac{\delta_i D_i}{\widetilde{S}_C(T_i - A_i)} - \frac{(1-\delta_i)\widetilde{Q}_T(Y_i - A_i, X_i, A_i)}{\widetilde{S}_C(Y_i - A_i)} + \int_0^{Y_i - A_i} \widetilde{Q}_T(u, X_i, A_i) \frac{d\Lambda_i^C(u)}{\widetilde{S}_C(u)} |X,A\Big\}$$
$$= -\int_a^\tau \frac{S_C(u-a)}{\widetilde{S}_C(u-a)} D(u,X) dS_T(u|X,A) + \int_a^\tau \int_0^{u-a} \frac{S_T(a+c|X,A)}{\widetilde{S}_T(a+c|X,A)} d\Big(\frac{S_C(c)}{\widetilde{S}_C(c)}\Big) D(u,X) d\widetilde{S}_T(u|X,A),$$

if $S_T(t|X,A) = \widetilde{S}_T(t|X,A)$, then

$$E\Big\{\frac{\delta_i D_i}{\widetilde{S}_C(T_i - A_i)} - \frac{(1-\delta_i)\widetilde{Q}_T(Y_i - A_i, X_i, A_i)}{\widetilde{S}_C(Y_i - A_i)} + \int_0^{Y_i - A_i} \widetilde{Q}_T(u, X_i, A_i) \frac{d\Lambda_i^C(u)}{\widetilde{S}_C(u)} |X\Big\} = 0,$$

and if $S_C(t) = \widetilde{S}_C(t)$, then

$$E\Big\{\frac{\delta_i D_i}{\widetilde{S}_C(T_i - A_i)} - \frac{(1-\delta_i)\widetilde{Q}_T(Y_i - A_i, X_i, A_i)}{\widetilde{S}_C(Y_i - A_i)} + \int_0^{Y_i - A_i} \widetilde{Q}_T(u, X_i, A_i) \frac{d\Lambda_i^C(u)}{\widetilde{S}_C(u)} |X\Big\} = 0,$$

since $d\Big(S_C(c)/\widetilde{S}_C(c)\Big) = 0$ and $E_A\Big[-\int_a^\tau [S_C(u-a)/\widetilde{S}_C(u-a)] D(u,X) dS_T(u|X, A)\Big] = E_A\big[DI(T > a)|X,A\big] = 0$ by the mean residual life assumption. Therefore, the most efficient estimator possesses double robustness.

References

Addona, V. and Wolfson, D. B. (2006). A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow up. *Lifetime Data Anal.* **12**, 264-274.

Asgharian, M., Wolfson, D. B. and Zhang, X. (2006). Checking staionarity of the incidence rate using prevalent cohort survival data. *Statist. Medicine* **25**, 1751-1767.

Bang, H. and Tsiatis, A. A. (2000). Estimating medical costs with censored data. *Biometrika* **87**, 329-343.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Inference in Semiparametric Models.* Johns Hopkins University Press, Baltimore, MD.

Chan, K. C., Chen, Y. Q. and Di, C. (2012). Proportional mean residual life model for right-censored length-biased data. *Biometrika* **99**, 995-1000.

Chen, Y. Q. and Cheng, S. (2005). Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika* **92**, 19-29.

Chen, Y. Q., Jewell, N. P., Lei, X. and Cheng, S. C. (2005). Semiparametric estimation of proportional mean residual life model in presence of censoring. *Biometrics* **61**, 170-178.

Chen, Y. Q. and Cheng, S. (2006). Linear life expectancy regression with censored data. *Biometrika* **93**, 303-313.

Flemming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

Gill, R. D. (1980). *Censoring and Stochastic Integrals.* Mathematisch Centrum, Amsterdam.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data.* New York, Springer.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. Ser. B* **62**, 711-730.

Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effcts: A comparative study. *Statist. Medicine* **23**, 2937-2960.

Maguluri, G. and Zhang, C. H. (1994). Estimation in the mean residual life regression model. *J. Roy. Statist. Soc. Ser. B* **56**, 477-489.

Newey, W. K. (1990). Semiparametric efficiency bounds. *J. Appl. Econom.* **5**, 99-135.

Oakes, D. and Dasu, T. (1990). A note on residual life. *Biometrika* **77**, 409-410.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* **34**, 179-189.

Qin, J. and Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under Cox Model. *Biometrics* **66**, 382-392.

Robins, J. M. (1999). Robust estimation in sequetially ignorable missing data and causal inference models. *Proceedings of the American Statistical Association Section on Bayesian Statistical Science,* 6-10. American Statistical Association, Alexandria, VA.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology-Methodological Issues* (Edited by N. Jewell, K. Dietz and V. Farewell), 297-331. Boston, Birkhauser.

Robins, J. M., Rotnitzky A. and Van der Laan, M. (2000). Comment on "On profile likelihood". *J. Amer. Statist. Assoc.* **95**, 477-482.

Robins, J. M., Rotnitzky A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 846-866.

Rojo, J. and Ghebremichael, M. (2006). Estimation of two ordered bivariate mean residual life functions. *J. Multivariate Anal.* **97**, 431-454.

Rotnitzky, A. and Robins, J. M. (2005). Inverse probability weighting in survival analysis. *Encylopedia of Biostatistics.* Wiley.

Rudin, W. B. (1976). *Principles of Mathematical Analysis.* 3rd Edition. McGraw-Hill.

Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models(with discussion). *J. Amer. Statist. Assoc.* **94**, 1096-1146.

Shen, Y., Ning, J. and Qin, J. (2009). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *J. Amer. Statist. Assoc.* **104**, 1192-1202.

Simon, R. (1980). Length-biased sampling in etiologic studies. *Amer. J. Epidemiology* **111**, 444-452.

Sun, L. and Zhang Z. (2009). A class of transformed mean residual life models with censored survival data. *J. Amer. Statist. Assoc.* **104**, 803-815.

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data.* Springer Verlag, New York.

Van der Laan, M. J. and Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality.* Springer Verlag, New York.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer Verlag, New York.

Vardi, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13**, 178-203.

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

Wolfson, C., Wolfson, D. B., Asgharian, M., M'Lan, C. E., Ostbye, T., Rockwood, K. and Hogan, D. B. for the Clinical Progression of Dementia Study Group. (2001). A reevaluation of the duration of survival after the onset of dementia. *New England J. Medicine* **344**, 1111-1116.

Zelen, M. and Feinleib M. (1969). On the theory of screening for chronic diseases. *Biometrika* **56**, 601-614.

School of Statistics, University of International Business and Economics, Chaoyang District, Beijing 100029, China.

E-mail: bff03021@163.com

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China.

Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242, U.S.A.

E-mail: jian-huang@uiowa.edu

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Haidian district, Beijing 100190, China.

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China.

E-mail: yzhou@amss.ac.cn

(Received May 2013; accepted August 2015)