Statistica Sinica **26** (2016), 1087-1115
doi:http://dx.doi.org/10.5705/ss.2014.094

SEMIPARAMETRIC MAXIMUM LIKELIHOOD INFERENCE FOR TRUNCATED OR BIASED-SAMPLING DATA

Hao Liu[1], Jing Ning[2], Jing Qin[3] and Yu Shen[2]

[1]*Baylor College of Medicine,*
[2]*The University of Texas MD Anderson Cancer Center*
and [3]*National Institutes of Health*

Abstract: Sample selection bias has long been recognized in many fields including clinical trials, epidemiology studies, genome-wide association studies, and wildlife management. This paper investigates the maximum likelihood estimation for censored survival data with selection bias under the Cox regression models where the selection process is modeled parametrically. A novel expectation-maximization algorithm is proposed and shown to have considerable computational advantages. Rigorous asymptotic properties of the estimator are established. Extensive simulation studies and a data analysis are conducted to investigate the performance of the proposed estimation procedure.

Key words and phrases: Biased sampling, length bias data, truncated and right-censored survival data.

1. Introduction

Biased sampling is frequently encountered in the study of biology (Terwilliger et al. (1997)), economics (Wooldridge (2010, Chap. 17)), sociology (Vella (1998)), public health (Brookmeyer and Gail (1987)) and industrial engineering (Kvam (2008)). Researchers have long recognized the efficiency and convenience of biased sampling, but have also noted that the observed data do not represent the distribution of the target population (Heckman (1976, 1990); Smith (1993)). Since Heckman's seminal work (Heckman (1979)), many techniques for correcting sample selection bias under either parametric or semi-parametric models have been proposed, mainly for the outcome variable that has a normal distribution (Wooldridge (2010, Chap. 17)). When the outcome is event times subject to right censoring, sample selection bias can also occur in a wide range of applications including astronomical surveys with truncated data (Woodroofe (1985)), gene mapping studies (Terwilliger et al. (1997)), financial performance analyses with survivorship bias (Carpenter and Lynch (1999)), labor economy studies (Lancaster (1990); Vella (1998)), RNA sequencing studies with transcript

length bias (Oshlack and Wakefield (2009)), epidemiological studies with prevalent cohort sampling (Brookmeyer and Gail (1987)), and wildlife studies with area-biased sampling (Patil and Rao (1978); Horne, Garton, and Sager-Fradkin (2007)), among many others.

Consider a univariate continuous outcome denoted by $\tilde{T}$, for example, the unemployment duration of a subject in a target population. Given covariates $X = x$, assume that $\tilde{T}$ has the population density function denoted by $f(t|x)$. Suppose that a subject from the target population is selected to a study with a probability proportional to a weight function $w(t, x)$ such that the density function of the observed outcome is

$$\frac{w(t,x)f(t|x)}{\int w(u,x)f(u|x)du}. \tag{1.1}$$

Clearly, the density function of the outcome variable for the sampled sujects is a biased or weighted version of the density function $f(\cdot|x)$ for the targeted population. The sampling mechanism (1.1) is very general; with various forms of the weight function $w(t, x)$, it can describe situations as diverse as the truncated survival data, size-biased data, and missing data. If the sampling weight function $w(t, x)$ is known completely, then this is a form of biased sampling carried intentionally by design. If $w(t, x)$ is not known completely, then this is a form of biased sampling that occurs accidentally by the nature of the study. This includes the selection bias or missing data problem (Chen (2001)), and the propensity score methodology for reducing selection bias in estimating the treatment effect for observational studies (Rosenbaum and Rubin (1984)).

In this paper, we focus on a class of the general biased-sampling mechanism (1.1) when the sampling weight function is independent of the covariates, $w(t, x) = H(t)$, where $H(t)$ is a positive increasing function for $t > 0$. If $H(t)$ is proportional to the length of the failure time, $H(t) = t$, this is the length-biased sampling problem. The weight function $H(t)$ may be interpreted as proportional to the cumulative distribution function of the underlying truncation time. Therefore, our model generalizes the setup for left-truncation survival data.

For left-truncation survival data, extensive efforts have been made for the nonparametric estimation of the distribution of $\tilde{T}$ (Turnbull (1976); Vardi (1985); Tsui, Jewell, and Wu (1988); Lagakos, Barraj, and Gruttola (1988); Kalbfleisch and Lawless (1989)), and for the semiparametric estimation of the regression models for $f(t|x)$ (Lai and Ying (1991); Wang, Brookmeyer, and Jewell (1993); Gross and Lai (1996)). The existing estimation methods for the regression models are largely based on the approach conditional on the observed truncation time without modeling its distribution specifically (Keiding (1992); Andersen et al. (1992); Klein and Moeschberger (2003)). As acknowledged in the literature,

such conditional methods can lead to a loss of efficiency (Asgharian, M'Lan, and Wolfson (2002)). The analysis of length-biased data has recently attracted a considerable amount of work on nonparametric and semiparametric estimation of the distribution of unbiased time $\tilde{T}$, including the construction of consistent estimating equations and maximum likelihood estimation (Wang (1996); Asgharian and Wolfson (2005); Tsai (2009); Qin et al. (2011); Carone, Asgharian, and Jewell (2014)). A major challenge when analyzing length-biased data is to verify the stringent stationarity assumption (Asgharian, Wolfson, and Zhang (2006)), $w(t, x) = t$, which is equivalent to check the sampling weight function as a uniform distribution.

We study maximum likelihood estimation (MLE) under a general biased-sampling mechanism when $w(t, x) = H(t)$ in (1.1), for a positive increasing function $H(t)$ for $t > 0$ that is specified parametrically. Qin et al. (2011) studied maximum likelihood estimation for a semiparametric model of $f(t|x)$ under length-biased data that must satisfy a stationarity assumption. We generalize their model by considering a flexible class of parametric models for $H(t)$. The main focus is a semiparametric model for $f(t|x)$ for the failure time data subject to a general biased-sampling, often the main interest of the practitioners. By linking the general biased-sampling problem with the truncated survival data problem, we make two major contributions to the literature: we provide a new approach to the maximum likelihood estimation with a general left-truncation model; specifying $H(t)$ parametrically allows us to alleviate the stationarity assumption for length-biased data.

Research on general biased-sampling mechanisms is relatively limited, especially when the weight function is of a general form in (1.1). When the density function $f(t|x)$ is left unspecified or specified semiparametrically, the weight function $w(t, x)$ have to be modeled parametrically in order for model (1.1) to be identifiable. In the case of a left-truncated survival time, Wang (1989) showed that a full-likelihood approach is not possible due to identifiability issues, if the distributions of the survival time and the truncation time are both left completely unspecified. When the weight function $w(t, x)$ is known up to a single parameter, Gilbert, Lele, and Vardi (1999) showed that the biased-sampling model is identifiable for a nonparametric estimation of the distribution of the failure time. Kim et al. (2013) considered a general biased-sampling problem and proposed an estimating-equation-based approach. They acknowledged that their approach is less efficient than the likelihood-based inference that is still lacking in the literature.

Directly maximizing the likelihood function is computationally prohibitive. We devise an expectation-maximization (EM) algorithm that incorporates the biased-sampling mechanism into a missing-data framework. Compared to the

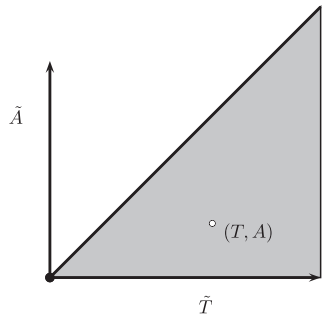

Figure 1. Sampling schema for time-to-event data (T, A).

EM method in Qin et al. (2011), the proposed EM algorithm does not need to impute the censored observations for the sampled individuals. As a result, the proposed procedure is computationally more efficient and can be implemented easily. The maximum likelihood estimator is shown to be asymptotically most efficient under the semiparametric Cox regression model. As a by-product, our unified approach can lead to the development of tools for checking the stationarity assumption underpinning the analysis of length-biased data.

This paper is organized as follows. In Section 2, we describe the intrinsic connection between biased-sampled data and truncation data, derive the full likelihood for the observed data, and present the key computational tool based on the EM algorithm. In Section 3, we establish large sample properties for the estimators. In Section 4, we present simulation studies and the analysis of a data example. We make concluding remarks in Section 5. Proofs are in the Appendix.

2. Methods

2.1. Model

Let $\tilde{T}$ and $\tilde{A}$ be positive random variables representing the unbiased event time measured from an initial event to an endpoint event, and the event time measured from the initial event to the sampling time, respectively. We model the association between a covariate vector X and the distribution of $\tilde{T}$, with observed biased-sampling data (T, A). Under this sampling schema, the data (T, A) can be only observed conditional on $\tilde{T} > \tilde{A}$ as depicted by Figure 1.

Let $f(\cdot|x)$ and $S(\cdot|x)$ be the respective density and survival function of $\tilde{T}$ given $X = x$. The biased-sampling data are the pair (T, A) with the joint density function

$$\frac{h_\theta(a)1(t > a)}{\int S(u|x)h_\theta(u)du} f(t|x) = \frac{h_\theta(a)1(t > a)}{\int H_\theta(u)f(u|x)du} f(t|x), \tag{2.1}$$

where $h_\theta(\cdot)$ and $H_\theta(\cdot)$ are proportional to the density and cumulative distribution function of $\tilde{A}$, respectively, with their forms known up to some parameters θ of finite dimension. Here, $1(\mathcal{C})$ is the indicator function taking the value 1 when $\mathcal{C}$ is true and 0 otherwise. Given $X = x$, the density function of T is a weighted function of $f(t|x)$ in the form (1.1)

$$\frac{H_\theta(t)f(t|x)}{\int H_\theta(u)f(u|x)du}.$$

When there is no censoring, the likelihood function based on (2.1) as a function of $(T, A) = (t, a)$ can be written as, for $t > a$,

$$\frac{f(t|x)}{S(a|x)}\left\{\frac{S(a|x)h_\theta(a)}{\int S(u|x)h_\theta(u)du}\right\}. \tag{2.2}$$

The first term of (2.2) is the conditional density function of $\tilde{T}$ given $A = a$, while the second term is the marginal density function of A, all for given $X = x$ and conditional on $\tilde{T} > \tilde{A}$. As the second term in (2.2) involves the distribution function $S(\cdot|x)$, any inference procedure based only on the conditional density function of $\tilde{T}$ given $A = a$ (the first term) loses information for the estimation of the distribution on $\tilde{T}$, even if the density function $h_\theta(\cdot)$ in (2.2) is completely known or known up to some parameters θ. On the other hand, if h_θ depends on x but is completely unspecified, then even with a parametric assumption on $S(\cdot|x)$, the second term in the likelihood cannot contribute additional information to the estimation of the distribution on $\tilde{T}$, as the term $S(\cdot|x)$ is absorbed into h_θ.

It is natural to base any statistical inference on the full likelihood approach as it is the most efficient. We thus illustrate the full likelihood approach for a general biased-sampling function $h_\theta(\cdot)$ specified parametrically in (2.2), jointly with a commonly used semiparametric Cox model for the survival function of $\tilde{T}$: conditional on covariates $X = x$,

$$S(t|x) = \exp\left\{-\int_0^t e^{\beta^{\mathrm{T}}x}d\Lambda(u)\right\}, \tag{2.3}$$

where β is the regression parameter, and the baseline cumulative hazard function $\Lambda(t)$ is not specified.

2.2. Likelihood

Denote the independently and identically distributed observed data from n subjects as $Y_i = \min\{T_i, (A_i + C_i)\}$ and $\delta_i = 1(T_i \le A_i + C_i)$, where C_i is the censoring time measured from the sampling time. As illustrated in Figure 2, the failure time T_i is subject to dependent right censoring by the time $A_i + C_i$.

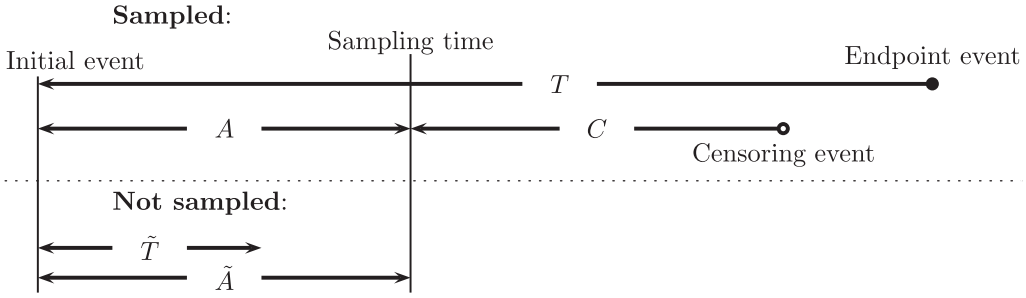

Figure 2. Dependent right censoring of bias-sampled data for a single subject.

Conditional on X_i, assume that the unbiased times $\tilde{T}_i$ and $\tilde{A}_i$ are independent. We also make the commonly used assumption that C_i is independent of (A_i, T_i) conditional on X_i. Under these assumptions, the full likelihood can be written as

$$\prod_{i=1}^{n} \left\{ \frac{f(Y_i|X_i)h_\theta(A_i)}{\int S(u|X_i)h_\theta(u)du} \bar{G}_C(Y_i - A_i|X_i) \right\}^{\delta_i} \left\{ \frac{S(Y_i|X_i)h_\theta(A_i)}{\int S(u|X_i)h_\theta(u)du} g_C(C_i|X_i) \right\}^{1-\delta_i},$$

where $g_C(\cdot|X_i)$ and $\bar{G}_C(\cdot|X_i)$ are the respective density function and survival function of the censoring time C_i for given X_i, and the parameters of interest are $\psi = (\theta, \beta, \Lambda(\cdot))$. It is evident that the terms on the censoring distributions can be factored out from the likelihood as they do not involve the parameters of interest on the distribution of $\tilde{T}$ and $\tilde{A}$. The likelihood function is thus proportional to

$$L_n(\theta, \beta, \Lambda) = \prod_{i=1}^{n} \frac{f^{\delta_i}(Y_i|X_i)S^{1-\delta_i}(Y_i|X_i)}{\int S(u|X_i)h_\theta(u)du} h_\theta(A_i). \tag{2.4}$$

Let $0 = t_0 < t_1 < t_2 < \cdots < t_K < \infty$ denote distinct observed time points, both censored and uncensored. Following an argument similar to that of Vardi, Y. (1989) and Qin et al. (2011), the nonparametric maximum likelihood estimator (NPMLE) for a discrete Λ has positive masses at $\{t_1, \ldots, t_K\}$ for any given (θ, β), in contrast to the Nelson-Aalen estimator in the traditional survival analysis. The NPMLE for the baseline function Λ is thus defined in the sense as described in Gill (1989).

2.3. EM algorithm

In this section, we exploit the underlying feature of the biased-sampling mechanism and devise an EM algorithm to find the MLE for ψ. Let $\lambda_k \equiv d\Lambda(t_k)$ be the positive masses of the discrete baseline function Λ at the times $t_1, \ldots, t_K$,

respectively, where $\Lambda(u) = \sum_{k=1}^{K} \lambda_k 1(t_k \le u)$. Let $\lambda = (\lambda_1, \ldots, \lambda_K)^{\mathrm{T}}$. The log-likelihood function can be expressed as

$$\sum_{i=1}^{n} \Bigg[\sum_{k=1}^{K} \Big\{ \delta_i \log f(t_k|X_i) - (1-\delta_i)\Lambda(t_k) e^{\beta^{\mathrm{T}} X_i} \Big\} 1(Y_i = t_k) - \log \int S(u|X_i) h_\theta(u) du + \log h_\theta(A_i) \Bigg].$$

With biased sampling, the data generating mechanism for each subject can be considered as sampling the unbiased times $(\tilde{T}, \tilde{A})$ for a random m_i times until $\tilde{T} > \tilde{A}$. For $i = 1, \ldots, n$, let the unobservable failure times be denoted by T_{ij}^* and A_{ij}^*, where $A_{ij}^* > T_{ij}^*$ with the corresponding covariates X_i for $j = 1, 2, \ldots, m_i$. In the presence of right censoring, the complete data for the ith subject include the observed data $(Y_i, A_i, \delta_i, X_i)$ and the latent unobservable data (T_{ij}^*, A_{ij}^*). The log-likelihood for the complete data can be written as

$$\sum_{i=1}^{n} \Bigg[\log h_\theta(A_i) + \sum_{j=1}^{m_i} \log h_\theta(A_{ij}^*) + \sum_{k=1}^{K} \sum_{j=1}^{m_i} 1(T_{ij}^* = t_k) \big\{ \log \lambda_k + \beta^{\mathrm{T}} X_i - \Lambda(t_k) e^{\beta^{\mathrm{T}} X_i} \big\} + \sum_{k=1}^{K} 1(Y_i = t_k) \big\{ \delta_i (\log \lambda_k + \beta^{\mathrm{T}} X_i) - \Lambda(t_k) e^{\beta^{\mathrm{T}} X_i} \big\} \Bigg],$$

where $\Lambda(t_k) = \sum_{k'=1}^{k} \lambda_{k'}$.

Denote the observed data for the ith subject by $\mathcal{O}_i \equiv (A_i, Y_i, \delta_i, X_i)$ for $i = 1, \ldots, n$. By the biased-data generating mechanism for the ith subject, the random integer m_i follows a geometric distribution with the success probability $P(\tilde{A}_i > \tilde{T}_i)$. Denote the current parameter value in the EM step by $\tilde{\psi} = \{\tilde{\theta}, \tilde{\beta}, \tilde{\lambda}\}$. Then, conditional on the observed data $\mathcal{O}_i$ the expectation of m_i is

$$E(m_i|\mathcal{O}_i) = \frac{1 - P(A_i \le T_i|X_i)}{P(A_i \le T_i|X_i)} = \frac{1 - \int f_{\tilde{\beta},\tilde{\lambda}}(u|X_i) H_{\tilde{\theta}}(u) du}{\int f_{\tilde{\beta},\tilde{\lambda}}(u|X_i) H_{\tilde{\theta}}(u) du},$$

where $f_{\tilde{\beta},\tilde{\lambda}}(t_k|X_i) = \tilde{\lambda}_k \exp(\tilde{\beta}^{\mathrm{T}} X_i) \exp\{-\tilde{\Lambda}(t_k) \exp(\tilde{\beta}^{\mathrm{T}} X_i)\}$, and $H_{\tilde{\theta}}(u) = \int_0^u h_{\tilde{\theta}}(v)\, dv$. The expected number of truncated latent subjects who would have the event time t_k is

$$w_{ik} = E\Bigg[\sum_{j=1}^{m_i} 1(T_{ij}^* = t_k) \Big| \mathcal{O}_i \Bigg] = E(m_i|\mathcal{O}_i) E\Big[1(T_{ij}^* = t_k) \Big| \mathcal{O}_i \Big] = \frac{f_{\tilde{\beta},\tilde{\lambda}}(t_k|X_i) \bar{H}_{\tilde{\theta}}(t_k)}{\int f_{\tilde{\beta},\tilde{\lambda}}(u|X_i) H_{\tilde{\theta}}(u) du},$$

where $\bar{H}_{\tilde{\theta}}(u) = 1 - H_{\tilde{\theta}}(u)$. The expectation of $\log h_\theta(A_{ij}^*)$ given the observed data under the biased-sampling constraint $A_{ij}^* > T_{ij}^*$ is

$$E\big\{\log h_\theta(A_{ij}^*)|A_{ij}^* > T_{ij}^*, \mathcal{O}_i\big\} = \frac{\int F_{\tilde{\beta},\tilde{\lambda}}(u|X_i)h_{\tilde{\theta}}(u)\log h_\theta(u)du}{\int F_{\tilde{\beta},\tilde{\lambda}}(u|X_i)h_{\tilde{\theta}}(u)du},$$

where $F_{\tilde{\beta},\tilde{\lambda}}(u|X_i) = \int_0^u f_{\tilde{\beta},\tilde{\lambda}}(v|X_i)dv$. It follows that the expected complete-data log-likelihood function given the current parameter estimate $\tilde{\psi} = \{\tilde{\theta}, \tilde{\beta}, \tilde{\lambda}\}$ is

$$\begin{aligned}\ell_E(\theta,\beta,\lambda) = \sum_{i=1}^n \Bigg[&\log h_\theta(A_i) + E(m_i|\mathcal{O}_i)\frac{\int F_{\tilde{\beta},\tilde{\lambda}}(u|X_i)h_{\tilde{\theta}}(u)\log h_\theta(u)du}{\int F_{\tilde{\beta},\tilde{\lambda}}(u|X_i)h_{\tilde{\theta}}(u)da} \\ &+ \sum_{k=1}^K w_{ik}\big\{\log\lambda_k + \beta^{\mathrm{T}}X_i - \Lambda(t_k)e^{\beta^{\mathrm{T}}X_i}\big\} \\ &+ \sum_{k=1}^K 1(Y_i = t_k)\big\{\delta_i(\log\lambda_k + \beta^{\mathrm{T}}X_i) - \Lambda(t_k)e^{\beta^{\mathrm{T}}X_i}\big\}\Bigg]. \end{aligned} \tag{2.5}$$

The M-step is to maximize the expected complete-data log-likelihood function conditional on the observed data with respect to θ, β, and λ. It turns out that the maximizer for λ_k has a closed form that depends only on β,

$$\lambda_k(\beta) = \frac{\sum_{i=1}^n\{w_{ik} + 1(Y_i = t_k)\delta_i\}}{\sum_{i=1}^n\sum_{k'=k}^K\{w_{ik'} + 1(Y_i = t_{k'})\}e^{\beta^{\mathrm{T}}X_i}}. \tag{2.6}$$

We notice the resemblance between (2.6) and the Breslow-type estimator in the traditional survival analysis, but they are different as w_{ik} in (2.6) is a function of both truncation and survival time distributions.

Maximizing the expected complete-data log-likelihood function with respect to β is equivalent to solving the equation

$$\frac{\partial\ell_E}{\partial\beta} = \sum_{i=1}^n\Bigg[\sum_{k=1}^K\Big\{w_{ik} + 1(Y_i = t_k)\delta_i - \{w_{ik} + 1(Y_i = t_k)\}e^{\beta^{\mathrm{T}}X_i}\Lambda(t_k)\Big\}X_i\Bigg] = 0. \tag{2.7}$$

Plugging $\lambda_j(\beta)$ for $j = 1, \ldots, K$ into (2.7), β can be solved from

$$\begin{aligned}\sum_{i=1}^n\Bigg[\sum_{k=1}^K\Bigg\{&w_{ik} + 1(Y_i = t_k)\delta_i \\ &- \{w_{ik} + 1(Y_i = t_k)\}\sum_{l=1}^k\frac{\sum_{j=1}^n\{w_{jl} + 1(Y_j = t_l)\delta_j\}e^{\beta^{\mathrm{T}}X_i}}{\sum_{j=1}^n\sum_{k'=l}^K\{w_{jk'} + 1(Y_j = t_{k'})\}e^{\beta^{\mathrm{T}}X_j}}\Bigg\}X_i\Bigg] = 0. \end{aligned} \tag{2.8}$$

In summary, the maximization in the M-step thus cycles through θ, β, and λ: given θ and β, the maximization with respect to λ is calculated explicitly using (2.6); given λ and β, the maximization with respect to θ is maximizing (2.5) with respect to θ, which only involves the first two terms; given θ and λ, the maximization with respect to regression β can be achieved using the existing software for the standard Cox regression model as follows.

To use the existing program for the Cox model with right-censored data, we note that β can be estimated by (2.8). First, create a data set for the unobserved truncated subjects, where the failure times are constructed by repeating the observed distinct times n times as $T_{nK} = (t_1, \ldots, t_K, \ldots, t_1, \ldots, t_K)$. The corresponding censoring indicator of T_{nK} is an identity vector of length nK, $\Delta_{nK} = (1, \ldots, 1)^{\mathrm{T}}$. Each vector of the covariate matrix is also repeated K times to match the truncated latent failure times, $X_{nK} = (X_1, \ldots, X_1, \ldots, X_n, \ldots, X_n)^{\mathrm{T}}$. The complete data are the combination of the observed data and the created data for unobserved truncated subjects. By using the function `coxph()` in S-PLUS (or R) with the option of `weight`, we can find the estimator of β at the M-step,

$$> \texttt{coxph}(\texttt{Surv}(\texttt{T}_\texttt{c}, \Delta_\texttt{c}) \sim \texttt{X}_\texttt{c}, \texttt{weight} = \texttt{W}_\texttt{c}),$$

where $T_c = (y_1, \cdots, y_n, T_{nK})$, $\Delta_c = (\delta_1, \cdots, \delta_n, \Delta_{nK})$, $X_c = (X_1, \cdots, X_n, X_{nK}^{\mathrm{T}})^{\mathrm{T}}$, and $W_c = (1, \cdots, 1, w_{11}, \cdots, w_{1K}, \cdots, w_{n1}, \cdots, w_{nK})$. The first n elements in the weight vector are associated with the observed data, so they have weight 1.

3. Asymptotic Properties

We establish the asymptotic properties of the maximum likelihood estimator (MLE), denoted by $\hat{\psi}_n \equiv (\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n(\cdot))$, where we use subscript n to emphasize its dependence on the sample size n. Using the counting processes formulations, the log-likelihood function has the form

$$\ell_n(\psi) = \sum_{i=1}^{n} \Bigg[\int_0^\tau \Big(\beta^{\mathrm{T}} X_i + \log d\Lambda(u)\Big) dN_i(u) - \int_0^\tau M_i(u) \exp(\beta^{\mathrm{T}} X_i) d\Lambda(u) + \log h_\theta(A_i) - \log \int_0^\tau S(u|X_i) h_\theta(u) du \Bigg], \tag{3.1}$$

where τ is the upper bound of the support for $\tilde{T}$, $N_i(t) = 1(A_i < Y_i \le t)\delta_i$, and $M_i(t) = 1(Y_i \ge t)1(Y_i > A_i)$. Under some mild regularity conditions listed in the Appendix, we first establish the strong consistency of the MLE using the classical Kullback-Leibler information approach (Gill (1989); Parner (1998)). We then apply the Z-theorem for the infinite-dimensional estimating equations to

prove the weak convergence of the estimators (van der Vaart and Wellner (1996, Thm. 3.3.1)). Proofs are in the Appendix.

3.1. Strong consistency

Let the true value be $\psi_0 = (\theta_0, \beta_0, \Lambda_0)$. As $\hat{\psi}_n$ maximizes the log-likelihood function, the empirical Kullback-Leibler information $\ell_n(\hat{\psi}_n) - \ell_n(\psi_0)$ must be always nonnegative. If $\hat{\psi}_n$ converges to ψ^*, say, then we can show that $\ell_n(\hat{\psi}_n) - \ell_n(\psi_0)$ must converge to the negative Kullback-Leibler distance between P_{ψ^*} and P_{ψ_0}. As the Kullback-Leibler information is always nonnegative, it implies that $P_{\psi^*} = P_{\psi_0}$ almost surely. Under condition (1) in Appendix A, the parametric family $\{h_\theta(a)\}$ is identifiable, while the Cox model for $\{f_{\beta,\Lambda}(t|x)\}$ is also identifiable. It follows from (2.1) that the model $\{P_\psi\}$ is identifiable, and $\psi^* = \psi_0$.

Let $\|\cdot\|_2$ be Euclidean distance. Suppose that τ is finite with $\Lambda(\tau) < \infty$.

Theorem 1. *Under the regularity conditions listed in Appendix A, the maximum likelihood estimators* $(\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n(\cdot))$ *are consistent:* $\|\hat{\theta}_n - \theta_0\|_2$, $\|\hat{\beta}_n - \beta_0\|_2$, *and* $\sup_{0\le u<\tau} |\hat{\Lambda}_n(u) - \Lambda_0(u)|$ *converge almost surely to* 0 *as* $n \to \infty$.

3.2. Weak convergence

Using the EM algorithm, we find the MLE estimators $(\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n)$ for the full likelihood function (3.1). The MLE hence must satisfy the infinite-dimensional score equations of (θ, β, Λ), as the baseline function $\Lambda(\cdot)$ is estimated nonparametrically. We establish weak convergence by applying the Z-theorem for infinite-dimensional estimating equations (van der Vaart and Wellner (1996, Thm. 3.3.1)). This approach has been applied to the semiparametric frailty models by Murphy (1995, Thm. 1) and Parner (1998, Thm. 2), among many others.

The maximizer of the likelihood function, the MLE estimators $(\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n)$, satisfy jointly the infinite-dimensional score equations of (θ, β, Λ),

$$U_{1n}(\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n) = 0, \quad U_{2n}(\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n) = 0, \quad \text{and} \quad U_{3n}(t, \hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n) = 0,$$

where the score functions $\{U_{1n}, U_{2n}, U_{3n}\}$ are calculated using the von Mises method for semiparametric MLE (Gill (1989)), by differentiating $\ell_n(\psi)$ with respect to θ, β, and a sub-model $d\Lambda_\eta(\cdot) = (1 + \eta\phi(\cdot))d\Lambda(\cdot)$, for a bounded and integrable function $\phi(\cdot)$, and a constant $\eta > 0$. The infinite-dimensional score

functions have the form

$$U_{1n}(\psi)=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\dot{h}_\theta(A_i)}{h_\theta(A_i)}1(Y_i>A_i)-\frac{\int_0^\tau S(u|X_i)\dot{h}_\theta(u)du}{\int_0^\tau S(u|X_i)h_\theta(u)du}\right\}, \tag{3.2}$$

$$U_{2n}(\psi)=\frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau X_i\left\{dN_i(u)-\left(M_i(u)-\frac{\int_u^\tau S(v|X_i)h_\theta(v)dv}{\int_0^\tau S(v|X_i)h_\theta(v)dv}\right)e^{\beta^{\mathrm{T}}X_i}d\Lambda(u)\right\}\right], \tag{3.3}$$

$$U_{3n}(t,\psi)=\frac{1}{n}\sum_{i=1}^{n}\left[\int_0^t\left\{dN_i(u)-\left(M_i(u)-\frac{\int_u^\tau S(v|X_i)h_\theta(v)dv}{\int_0^\tau S(v|X_i)h_\theta(v)dv}\right)e^{\beta^{\mathrm{T}}X_i}d\Lambda(u)\right\}\right], \tag{3.4}$$

where $\dot{h}_\theta(\cdot)$ is the first partial derivative of $h_\theta(\cdot)$ with respect to θ. Under (2.1), it can be easily confirmed that, conditional on $\tilde{T}>\tilde{A}$, the estimating equations are of mean zero, $E\{U_{1n}(\psi)\}=0$, $E\{U_{2n}(\psi)\}=0$ and $E\{U_{3n}(t,\psi)\}=0$.

Let $U_n(\cdot,\psi)\equiv\{U_{1n}(\psi),U_{2n}(\psi),U_{3n}(\cdot,\psi)\}$, and denote its expectation under the true values $\psi_0=(\theta_0,\beta_0,\Lambda_0)$ by

$$\begin{aligned}U_0(\cdot,\psi)&\equiv\{U_{10}(\psi),U_{20}(\psi),U_{30}(\cdot,\psi)\}\\&=\{E_0\{U_{1n}(\psi)\},E_0\{U_{2n}(\psi)\},E_0\{U_{3n}(\cdot,\psi)\}\}.\end{aligned}$$

It can be confirmed that the true value ψ_0 satisfies the population score equations $U_0(t,\psi_0)=0$. Evaluated at the true value ψ_0 the estimating functions can be written as an empirical process $\sqrt{n}U_n(t,\psi_0)=\sqrt{n}\{U_n(t,\psi_0)-U_0(t,\psi_0)\}$ indexed by t.

By the uniform central limit theorem, $\sqrt{n}U_n(\cdot,\psi_0)$ converges weakly to $\mathbb{W}(\cdot)=\{\mathbb{W}_1,\mathbb{W}_2(\cdot)\}$, where $\mathbb{W}_1$ is a Gaussian random vector and $\mathbb{W}_2(\cdot)$ is a tight Gaussian process. Letting $x^{\otimes2}=xx^{\mathrm{T}}$, the marginal covariance function for $\mathbb{W}$ has the form

$$\Sigma=\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix},$$

where

$$\Sigma_{11}=E_0\begin{pmatrix}U_{11}(\psi_0)\\U_{21}(\psi_0)\end{pmatrix}^{\otimes2},\quad \Sigma_{12}(t)=\Sigma_{21}(t)^{\mathrm{T}}=E_0\left\{\begin{pmatrix}U_{11}(\psi_0)\\U_{21}(\psi_0)\end{pmatrix}U_{31}(t,\psi_0)\right\},$$

$\Sigma_{22}(t_1,t_2)=E_0\{U_{31}(t_1,\psi_0)U_{31}(t_2,\psi_0)\}$.

Denote the Fréchet derivative of $U_0(\psi)$ evaluated at $\psi=\psi_0$ by $\dot{U}_0$. We confirm that the operator $\dot{U}_0$ is continuously invertible using the classical Fredholm theorem for the integral equations (Tricomi (1985)). To apply the Z-theorem for the infinite-dimensional estimating equations (van der Vaart and Wellner

(1996, Thm. 3.3.1)), we outline our proof in the Appendix by confirming the three main conditions of the Z-theorem: Fréchet differentiability and invertibility, weak convergence of $\sqrt{n}U_n(\psi_0)$, and a stochastic approximation condition of the estimating equations. Theorem 2 summarizes the results.

Theorem 2. *Under the regularity conditions listed in Appendix A,* $\sqrt{n}(\hat{\psi}_n - \psi_0)$ *converges weakly to a tight mean zero Gaussian process* $-\dot{U}_0^{-1}(\mathbb{W})$.

3.3. Asymptotic normality

We characterize the asymptotic distribution of the sequence $\sqrt{n}(\hat{\psi}_n - \psi_0)$ that is completely determined by the tightness of $\dot{U}_0^{-1}(\mathbb{W})$ and its marginal covariance function. Let $\ddot{h}_\theta(t) = d^2h_\theta(t)/d\theta^2$, $h_0(\cdot) = h_{\theta_0}(\cdot)$, $\dot{h}_0(\cdot) = \dot{h}_{\theta_0}(\cdot)$, and $\ddot{h}_0(\cdot) = \ddot{h}_{\theta_0}(\cdot)$. Denote for $l = 0, 1, 2$,

$$
\begin{aligned}
Q_0(\cdot, X) &= \frac{\exp\big\{-\int_0^{\cdot} e^{\beta_0^{\mathrm{T}} X} d\Lambda_0(u)\big\}}{\int_0^{\tau} \exp\big\{-\int_0^{u} e^{\beta_0^{\mathrm{T}} X} d\Lambda_0(u)\big\} h_0(u)du},\\
\kappa_0 &= E_0\bigg[\int_0^{\tau} Q(u, X)\ddot{h}_0(u)du + \bigg(\int_0^{\tau} Q(u, X)\dot{h}_0(u)du\bigg)^2\\
&\quad - \bigg\{\frac{\ddot{h}_0(A)}{h_0(A)} - \bigg(\frac{\dot{h}_0(A)}{h_0(A)}\bigg)^2\bigg\} 1(Y > A)\bigg],\\
\kappa_1^{(l)}(u) &= \int_u^{\tau} E_0\bigg[\bigg\{h_0(v)\int_0^{\tau} Q_0(z, X)\dot{h}_0(z)dz - \dot{h}_0(v)\bigg\} Q_0(v, X) e^{\beta_0^{\mathrm{T}} X} X^{\otimes l}\bigg] dv,\\
\kappa_2^{(l)}(u) &= \int_u^{\tau} E_0\bigg[\bigg\{\Lambda_0(v) - \int_0^{\tau} Q_0(z, X)\Lambda_0(z)h_0(z)dz\bigg\} Q_0(v, X) e^{2\beta_0{}^{\mathrm{T}} X} X^{\otimes l}\bigg]\\
&\qquad h_0(v)dv,\\
\kappa_3^{(l)}(u) &= E_0\bigg[\bigg\{M(u) - \int_u^{\tau} Q_0(z, X)h_0(z)dz\bigg\} e^{\beta_0{}^{\mathrm{T}} X} X^{\otimes l}\bigg],\\
\kappa_4(u, v) &= \int_u^{\tau} E_0\bigg\{\bigg(1(z \geq v) - \int_u^{\tau} Q_0(s, X)h_0(s)ds\bigg) Q(z, X) e^{2\beta_0{}^{\mathrm{T}} X}\bigg\} h_0(z)dz,\\
J_{11} &= \begin{pmatrix} \kappa_0 & \{\int_0^{\tau} \kappa_1^{(1)}(u)d\Lambda_0(u)\}^{\mathrm{T}} \\ \int_0^{\tau} \kappa_1^{(1)}(u)d\Lambda_0(u) & \int_0^{\tau} \{\kappa_2^{(2)}(u) + \kappa_3^{(2)}(u)\}d\Lambda_0(u) \end{pmatrix},\\
J_{21}(u) &= J_{12}(u)^{\mathrm{T}} = \Big(\int_0^{u} \kappa_1^{(0)}(z)d\Lambda_0(z) \int_0^{u} \{\kappa_2^{(1)}(z) + \kappa_3^{(1)}(z)\}^{\mathrm{T}} d\Lambda_0(z)\Big).
\end{aligned}
$$

By straightforward calculation, the Fréchet derivative of $U_0(\psi)$ is

$$
\dot{U}_0(\psi) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} \xi \\ \Lambda \end{pmatrix} = \begin{pmatrix} \sigma_{11}(\xi) + \sigma_{12}(\Lambda) \\ \sigma_{21}(\xi) + \sigma_{22}(\Lambda) \end{pmatrix},
$$

where $\xi = (\theta^{\mathrm{T}}, \beta^{\mathrm{T}})^{\mathrm{T}}$ are the finite dimensional parameters, and

$$\sigma_{11}(\xi) = J_{11}\xi, \quad \sigma_{12}(\Lambda) = \int_0^\tau J_{12}(u)d\Lambda(u), \quad \sigma_{21}(\xi)(t) = J_{21}(t)\xi,$$
$$\sigma_{22}(\Lambda)(t) = \int_0^t \left\{ \int_0^\tau \kappa_4(u,v)d\Lambda(v) \right\} d\Lambda_0(u) + \int_0^t \kappa_3^{(0)}(u)d\Lambda(u).$$

Here, J_{11} is the Fisher information for the finite diminutional parameters ξ when the true value Λ_0 is known. The derivative $\dot{U}_{\psi_0}$ is continuously invertible with this form:

$$\dot{U}_{\psi_0}^{-1}(\psi) \equiv \begin{pmatrix} \sigma_{11}^{-1} + \sigma_{11}^{-1}\sigma_{12}\Phi^{-1}\sigma_{21}\sigma_{11}^{-1} & -\sigma_{11}^{-1}\sigma_{12}\Phi^{-1} \\ \Phi^{-1}\sigma_{21}\sigma_{11}^{-1} & \Phi^{-1} \end{pmatrix} \begin{pmatrix} \xi \\ \Lambda \end{pmatrix},$$

where $\sigma_{11}^{-1}(\xi) = J_{11}^{-1}\xi$. Here, $\Phi = \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12}$ is continuously invertible with the inverse as a function of t that has the form

$$\Phi^{-1}(\Lambda)(t) = \int_0^t \frac{d\Lambda(u)}{\kappa_3^{(0)}(u)} + \int_0^\tau \left(\int_0^t R(v,u)dv \right) \frac{d\Lambda(u)}{\kappa_3^{(0)}(u)}, \tag{3.5}$$

where $R(t,u)$ satisfies the equation

$$R(t,u) = K(t,u) + \int K(t,v)R(v,t)dv,$$
$$K(t,u) = \frac{\lambda_0(t)}{\kappa_3^{(0)}(t)} \left\{ \begin{pmatrix} \kappa_1^{(0)}(t) \\ \kappa_2^{(1)}(t) + \kappa_3^{(1)}(t) \end{pmatrix}^{\mathrm{T}} J_{11}^{-1}J_{12}(u) - \kappa_4(t,u) \right\}. \tag{3.6}$$

It follows that the process $\sqrt{n}(\hat{\Lambda}_n - \Lambda_0)$ converges weakly to a tight Gaussian process $\Phi^{-1}\sigma_{21}\sigma_{11}^{-1}(\mathbb{W}_1) + \Phi^{-1}(\mathbb{W}_2)$, where $\Phi^{-1}(\mathbb{W}_2)$ is a Gaussian process given by

$$\Phi^{-1}(\mathbb{W}_2)(t) = \int_0^t \frac{d\mathbb{W}_2(u)}{\kappa_3^{(0)}(u)} + \int_0^\tau \left(\int_0^t R(v,u)dv \right) \frac{d\mathbb{W}_2(u)}{\kappa_3^{(0)}(u)}.$$

The stochastic integral is defined via integration by parts. Furthermore, for the estimator $\hat{\xi}_n = \{\hat{\theta}_n^{\mathrm{T}}, \hat{\beta}_n^{\mathrm{T}}\}^{\mathrm{T}}$ of the finite dimensional parameters, the random vector $\sqrt{n}(\hat{\xi}_n - \xi_0)$ converges in distribution to a mean zero normal random vector

$$\sigma_{11}^{-1}(\mathbb{W}_1) + \sigma_{11}^{-1}\sigma_{12}\Phi^{-1}\sigma_{21}\sigma_{11}^{-1}(\mathbb{W}_1) - \sigma_{11}^{-1}\sigma_{12}\Phi^{-1}(\mathbb{W}_2). \tag{3.7}$$

The first term $\sigma_{11}^{-1}(\mathbb{W}_1)$ in (3.7) has mean zero and the sandwich covariance matrix $J_{11}^{-1}\Sigma_{11}J_{11}^{-1}$. The extra terms $\sigma_{11}^{-1}\sigma_{12}\Phi^{-1}\sigma_{21}\sigma_{11}^{-1}(\mathbb{W}_1) - \sigma_{11}^{-1}\sigma_{12}\Psi^{-1}(\mathbb{W}_2)$ in (3.7) indicate the extra variability due to $\hat{\Lambda}_n$. The asymptotic variance of the finite dimensional estimator $\hat{\xi}_n$ is thus quite complicated, as it involves both $\mathbb{W}_1$

and $\mathbb{W}_2$, which depend on the randomness of the estimator $\hat{\Lambda}_n$ of the infinite-dimensional parameter.

While the asymptotic variance of the estimator for the finite-dimensional parameters can be calculated as the inverse of the empirical information matrix of the likelihood function profiled over $(\theta, \beta^{\mathrm{T}})$, direct evaluation of the variance-covariance matrix is extremely difficult as there is no explicit form for the information matrix. Since the rate of convergence for $\hat{\psi}_n$ is shown to be $n^{-1/2}$, we use an EM-aided computational differentiation approach (Chen and Little (1999)) to approximate the information matrix. For notation clarity, we drop the subscript n in the description of computational algorithm:

(i) Perturb each component of $\hat{\xi} = (\hat{\theta}_1, \ldots, \hat{\theta}_q, \hat{\beta}_1, \ldots, \hat{\beta}_p)$ by a small value $\epsilon = 1/n$ in its neighborhood in both directions to obtain the perturbed estimators $\hat{\xi}_l^{\epsilon+} = \hat{\xi} + (0, \ldots, \epsilon, \ldots, 0)$ and $\hat{\xi}_l^{\epsilon-} = \hat{\xi} - (0, \ldots, \epsilon, \ldots, 0)$, respectively, for $l = 1, \ldots, q + p$.

(ii) With $\xi = \hat{\xi}_l^{\epsilon+}$ or $\xi = \hat{\xi}_l^{\epsilon-}$, we calculate the conditional expectations required in the E-step, and update the respective estimate $\hat{\lambda}_{\hat{\xi}_l^{\epsilon+}}$ or $\hat{\lambda}_{\hat{\xi}_l^{\epsilon-}}$ by maximizing the expected log-likelihood in the M-step. We repeat the steps (i)−(ii) until convergence is achieved.

(iii) Calculate the expected complete-data scores with respect to (θ, β), evaluated at $(\hat{\xi}_l^{\epsilon+}, \hat{\lambda}_{\hat{\xi}_l^{\epsilon+}})$ and $(\hat{\xi}_l^{\epsilon-}, \hat{\lambda}_{\hat{\xi}_l^{\epsilon-}})$ and denoted by $\nu_E(\hat{\xi}_l^{\epsilon+}, \hat{\lambda}_{\hat{\xi}_l^{\epsilon+}})$ and $\nu_E(\hat{\xi}_l^{\epsilon-}, \hat{\lambda}_{\hat{\xi}_l^{\epsilon-}})$, respectively.

(iv) Approximate the lth row of the information matrix of $\hat{\xi}$ by

$$\frac{1}{2\epsilon}\left\{\nu_E(\hat{\xi}_l^{\epsilon-}, \hat{\lambda}_{\hat{\xi}_l^{\epsilon-}}) - \nu_E(\hat{\xi}_l^{\epsilon+}, \hat{\lambda}_{\hat{\xi}_l^{\epsilon+}})\right\}.$$

4. Numerical Results

4.1. Simulation

We conducted simulation studies to evaluate the finite sample performance of the proposed estimators and the variance estimation procedure. We also compared the efficiency of the proposed method with that of an existing MLE method for length-biased data (Qin et al. (2011)) and an conditional method for left-truncated data (Wang, Brookmeyer, and Jewell (1993)).

We generated the survival time $\tilde{T}$ from a proportional hazards model with baseline function $\Lambda(t) = t^2$ and two covariates (X_1, X_2), with X_1 a binary covariate following a Bernoulli distribution with parameter 0.5, and X_2 a continuous covariate generated from a uniform distribution on $(-0.5, 0.5)$. The regression coefficients $\beta = (\beta_1, \beta_2)$ was set to be $(0.5, 1)$. The underlying truncation time $\tilde{A}$

was independently generated from an exponential distribution of a single parameter θ. For length-biased data, the truncation time was generated independently from the uniform distribution. To form a prevalent cohort, we kept only the subjects satisfing the sampling constraint $\tilde{T} \geq \tilde{A}$. The censoring time C in the prevalent cohort was generated from a uniform distribution on the interval $[0, \tau_c]$, where τ_c was chosen so that the overall censoring rate was approximately 20%, 30%, and 50%, respectively. Sample sizes of 100, 200, and 400 were used, and each scenario used 500 simulation replicates.

We first compare the performance of three methods in analyzing length-biased data: the proposed method, the method described in Qin et al. (2011), and the conditional method of Wang, Brookmeyer, and Jewell (1993). The simulation results are summarized in Table 1, including the empirical means, the average of the asymptotic standard error estimators and the mean squared errors based on 500 replicates. The two maximum likelihood estimators have lower MSEs than the conditional method. As the data were generated under the length-biased sampling, the method of Qin et al. (2011) has a slightly lower MSE than the proposed method that is developed for general left-truncation data. This is expected because the proposed method has more parameters to estimate than the method by Qin et al. (2011). The proposed method has relatively smaller bias than the method of Qin et al. (2011) under high censoring rate (50%). In the simulation studies, the proposed EM algorithm was computationally more efficient than that of Qin et al. (2011) in terms of the CPU times under the two computational algorithm.

Table 2 summarizes the simulation results with left truncation data. We report the empirical means, the average of the asymptotic standard error estimators, the empirical standard deviations, and the mean squared errors based on 500 replicates. With a light (20%) or moderate (30%) censoring percentage, all three model parameters θ, β_1, and β_2 were well-estimated by the proposed method in that the biases of the estimates were small and the estimated standards errors were close to the empirical standard deviations, even with a small sample size (100). With heavy censoring (50%) and a small sample size (100), the biases of the estimated parameters were around 10%, but decreased when the sample size increased or the censoring rates were reduced. In additional simulation studies (results not shown), the bias vanished when sample size is very large (n =1,000). In Table 2, we present the simulation results using the conditional method for left-truncated data (Wang, Brookmeyer, and Jewell (1993)). As expected, the conditional method is less efficient than the proposed MLE method with larger mean squared errors.

Table 1. Summary of simulation studies with length-biased data. "Mean" is the empirical mean; "ASE" is the average of asymptotic standard error estimates; "MSE" is mean squared error.

Sample	Censoring	Proposed method			Qin et al. (2011)			Conditional method		
		Mean	ASE	MSE	Mean	ASE	MSE	Mean	ASE	MSE
200	20%	(0.529, 1.052)	(0.116, 0.215)	(0.015, 0.049)	(0.490, 0.975)	(0.106, 0.199)	(0.011, 0.040)	(0.506, 1.022)	(0.157, 0.305)	(0.025, 0.094)
	30%	(0.518, 1.029)	(0.123, 0.219)	(0.015, 0.049)	(0.479, 0.952)	(0.112, 0.200)	(0.013, 0.042)	(0.503, 1.019)	(0.179, 0.329)	(0.032, 0.109)
	50%	(0.489, 0.978)	(0.131, 0.237)	(0.017, 0.056)	(0.466, 0.930)	(0.120, 0.208)	(0.016, 0.048)	(0.494, 1.027)	(0.218, 0.394)	(0.048, 0.156)
400	20%	(0.521, 1.042)	(0.087, 0.142)	(0.008, 0.022)	(0.490, 0.979)	(0.081, 0.133)	(0.007, 0.018)	(0.499, 1.009)	(0.119, 0.196)	(0.014, 0.038)
	30%	(0.516, 1.031)	(0.090, 0.147)	(0.008, 0.023)	(0.480, 0.959)	(0.082, 0.135)	(0.007, 0.020)	(0.496, 1.009)	(0.130, 0.213)	(0.017, 0.045)
	50%	(0.505, 1.011)	(0.096, 0.159)	(0.009, 0.026)	(0.469, 0.940)	(0.086, 0.139)	(0.008, 0.023)	(0.491, 1.013)	(0.153, 0.261)	(0.024, 0.068)

Table 2. Summary of Simulation Studies with left-truncation data: Mean and Standard Errors. "Mean" is the empirical mean; "ASE" is the average of asymptotic standard error estimates; "ESE" is the empirical standard deviation; "MSE" is mean squared error.

Sample size	Censoring	$\theta = 1$ Proposed method				$\beta = (0.5, 1)$ Proposed method				Conditional method		
		mean	ASE	ESE	MSE	mean	ASE	ESE	MSE	mean	ASE	MSE
100	20%	1.023	0.226	0.219	0.052	(0.508, 1.011)	(0.211,0.364)	(0.206,0.365)	(0.045,0.132)	(0.505, 1.016)	(0.238,0.413)	(0.057,0.171)
	30%	0.971	0.237	0.221	0.057	(0.487,0.970)	(0.215,0.370)	(0.208,0.369)	(0.046. .138)	(0.505, 1.020)	(0.251,0.441)	(0.063,0.195)
	50%	0.901	0.247	0.206	0.071	(0.429,0.889)	(0.230,0.390)	(0.224,0.397)	(0.058,0.164)	(0.490, 1.051)	(0.312,0.513)	(0.097,0.266)
200	20%	1.005	0.156	0.154	0.024	(0.505,1.002)	(0.151,0.236)	(0.144,0.253)	(0.023,0.056)	(0.505,1.007)	(0.169,0.270)	(0.029,0.073)
	30%	0.974	0.167	0.156	0.029	(0.490,0.973)	(0.155,0.240)	(0.146,0.258)	(0.024,0.058)	(0.504, 1.008)	(0.181,0.291)	(0.033,0.085)
	50%	0.911	0.195	0.147	0.046	(0.440,0.900)	(0.166,0.269)	(0.159,0.282)	(0.031,0.082)	(0.488, 1.019)	(0.219,0.352)	(0.048,0.124)
400	20%	0.999	0.113	0.109	0.013	(0.505,0.996)	(0.106,0.178)	(0.101,0.177)	(0.011,0.031)	(0.505, 1.009)	(0.115,0.200)	(0.013,0.040)
	30%	0.976	0.124	0.110	.016	(0.495,0.978)	(0.109,0.181)	(0.103,0.182)	(0.012,0.033)	(0.508, 1.013)	(0.120,0.210)	(0.014,0.044)
	50%	0.927	0.146	0.105	0.027	(0.461,0.925)	(0.125,0.205)	(0.113,0.202)	(0.017,0.048)	(0.502, 1.029)	(0.147,0.250)	(0.022,0.063)

Table 3. Estimates (Standard Errors) of Regression Coefficients for Dementia Data.

	Proposed MLE	Qin et al. (2011)	Conditional method
Probable Alzheimer's	0.130 (0.064)	0.125 (0.062)	0.037 (0.093)
Vascular dementia	0.193 (0.078)	0.185 (0.077)	0.124 (0.113)
θ	0.0002(0.013)	NA	NA

4.2. Application

We applied the proposed method to a cohort of prevalent cases in one of the largest epidemiological studies of dementia, the Canadian Study of Health and Aging (The Canadian Study of Health and Aging Working Group (1994)), which has been analyzed previously (Asgharian, M'Lan, and Wolfson (2002); Qin et al. (2011); Carone, Asgharian, and Jewell (2014)). In the study, a total of 10,263 people agreed to participate, and were then screened for dementia. Among them, 1,132 individuals were identified as having the disease and their dates of dementia onset were ascertained from their medical records. These individuals identified as having dementia were followed until death or last follow-up in 1996. Excluding the participants with missing dates of disease onsets or classification of dementia subtype, the data set included 818 participants: 393 with probable Alzheimer's disease, 252 with possible Alzheimer's disease and 173 with vascular dementia.

We used the semiparametric maximum likelihood estimation as described in Section 2.3 to analyze the survival differences among the three dementia subtypes. We used the subgroup with possible Alzheimer's disease as the baseline group and included two binary indicators for the other two subtypes of dementia in the Cox proportional hazards model for the overall survival time from dementia onset. We assumed the truncation time followed truncated exponential distribution on $(0, \tau_{max})$, where τ_{max} is the maximum observed time from dementia onset.

The estimated covariate effects of the two subtypes of dementia from the two estimation methods and the estimated parameter of the exponential distribution for the truncation time by our method are summarized in Table 3. The proposed semiparametric MLE suggested a statistically significant survival difference between the group with possible Alzheimer's disease and the group with vascular dementia (0.193, SE=0.078), and a marginally significant difference between the groups with possible versus probable Alzheimer's disease (0.130, SE=0.064). The results are comparable with those reported in Qin et al. (2011), where the estimate was 0.185 (SE=0.077) for comparing possible Alzheimer's disease with vascular dementia, and was 0.125 (SE=0.062) for comparing probable Alzheimer's disease with possible Alzheimer's disease. The estimated effects using the conditional method do not show any statistically significant survival differences among the three subtypes of dementia. The estimated parameter for

the truncated exponential distribution was very small (0.0002) compared with its standard error (0.013), suggesting that the truncation time could be approximately uniformly distributed. This result is consistent with a previous finding by Asgharian, Wolfson, and Zhang (2006), who showed that the dementia data satisfied the stationarity assumption.

5. Discussion

Biased sampling is well recognized in a wide range of applications. Jointly modeling the bias selection process with a parametric specification and the survival time with a semiparametric specification, we propose a full maximum likelihood approach that works well for the problem of general left-truncation data, a special case of biased sampling. The proposed MLE is shown to be asymptotically consistent and efficient. A novel EM algorithm with desirable computational advantages is developed to find the semiparametric MLE with infinite-dimensional parameters. The M-step for the regression coefficient can be easily implemented by the existing software for the Cox regression model with right-censored survival data. As demonstrated in the numerical studies, the proposed MLEs are more efficient than the estimating equations approach based on the conditional likelihood. The proposed EM algorithm does not need to impute the censored observation. In spite of the additional step to estimate the parameters for the truncation time distribution, the proposed EM algorithm is computationally more efficient than the EM algorithm developed by Qin et al. (2011).

From scientific and statistical points of view, it is of great interest to estimate the truncation time distribution. In the setting of biased sampling, the truncation time is only partially observed because only subjects who had not have the failure event before the recruitment time are recruited. The proposed method has the advantage of evaluating the distribution of the underlying selection process in addition to efficiency improvements in estimating the survival functions of the failure events.

For the purpose of model identifiability, a parametric model $h_\theta(\cdot)$ without covariate is assumed for the distribution of the truncation time. It is critical to develop model checking techniques to test the goodness-of-fit of the parametric model assumption. In practice, one approach is to use a rich parameter family and then examine the model fitting. Additionally, the proposed approach can be generalized to include X in the weight function with a properly assumed parametric structure for $h_\theta(a|x)$.

Although statistical methodology has been proposed to verify the stationarity assumption for length-biased data (Asgharian, Wolfson, and Zhang (2006)), the proposed modeling framework can be used to develop a formal test of whether $h_\theta(\cdot)$ is uniformly distributed. Indeed, verifying whether the underlying disease

process is stationary can be of research interest in itself. In our numerical study, we considered a single parameter $\theta = 0$ for the left truncation distribution. However, testing $\theta = 0$ is on the boundary of the parameter space that requires special considerations. While asymptotic theory with boundary problems has been well developed for parametric models (Self and Liang (1987)), yo the best of our knowledge, little has been done in semiparametric modeling with boundary problems. The situation is more complicated in the setting of the biased sampling problem, which remains a future research topic.

We have focused on the Cox model for the survival data because of its popularity and availability in standard statistical software. More general survival models such as the transformation models can also be investigated. However, the extension of the EM algorithm for these models is not straightforward. Further investigation is warranted.

Acknowledgement

We thank Professor Masoud Asgharian and the investigators of the Canadian Study of Health and Aging (CHSA) for providing us with the dementia data, which were collected as part of the CHSA, funded by the Seniors' Independence Research Program, through the National Health Research and Development Program of Health Canada (Project no.6606-3954-MC(S)). Additional funding was provided by Pfizer Canada Incorporated through the Medical Research Council/Pharmaceutical Manufacturers Association of Canada Health Activity Program, NHRDP Project 6603-1417-302(R), Bayer Incorporated, and the British Columbia Health Research Foundation Projects 38 (93-2) and 34 (96-1). The study was coordinated through the University of Ottawa and the Division of Aging and Seniors, Health Canada. This work in this paper was supported in part by the National Institutes of Health.

Appendix: Proofs of Asymptotic Properties

Appendix A. Regularity conditions

We summarize the regularity conditions:

1. The parameter $\{\theta, \beta\}$ is in a compact set $\mathcal{B}$ in the real space $\mathbb{R}^q \times \mathbb{R}^p$. The parametric model $\{h_\theta(\cdot) : \theta \in \mathcal{B}\}$ is identifiable. The parameter Λ is in the set $\mathcal{A}$ of nondecreasing functions with $\Lambda(0) = 0$ and $\Lambda(\tau-) < \infty$. The true parameter $\Lambda_0(\cdot)$ is continuous and differentiable, and $P(\tilde{A} + C > \tau) > 0$.
2. The function $h_\theta(a) \geq \delta_1$ for some constant $\delta_1 > 0$ for every a.
3. The function $\dot{h}_\theta(\cdot)$ and $h_\theta(\cdot)$ satisfy a Lipschitz condition in θ. More specifically, there exist functions, F_1 and F_2 such that

$$|\dot{h}_{\theta_1}(a) - \dot{h}_{\theta_2}(a)| \leq \|\theta_1 - \theta_2\|_2 F_1(a), \quad \text{and} \quad |h_{\theta_1}(a) - h_{\theta_2}(a)| \leq \|\theta_1 - \theta_2\|_2 F_2(a),$$

where $E_0\{F_1(\tilde{A})^2\} < \infty$ and $E_0\{F_2(\tilde{A})^2\} < \infty$.

4. The covariate X is bounded, and $E_0\|X\|_2^2$, $E_0\{e^{|\beta^{\mathrm{T}}X|}\}$ are all bounded for every β in $\mathcal{B}$.
5. The matrix J_{11} is positive definite and the bivariate function $K(t,u)$ in (3.6) is square integrable.

Appendix B. Proof of consistency

To avoid the measurability issues, the probability measure is understood as the outer probability (van der Vaart and Wellner (1996)). As technical details are similar to those in the literature for semiparametric maximum likelihood estimation, see for example, Parner (1998), we provide only a sketch of the proof. The first step is to show that $\hat{\psi}_n = (\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n)$ stays bounded, in particular, $\overline{\lim}_n \hat{\Lambda}_n < \infty$. As $\{\hat{\theta}_n, \hat{\beta}_n\}$ are found in a bounded compact set of finite dimensions, we can find a convergence subsequence of $\{\hat{\theta}_n, \hat{\beta}_n\}$. To show that $\hat{\Lambda}_n$ stays bounded, we use proof by contradiction as follows. Suppose that $\hat{\Lambda}_n(\cdot)$ diverges. We can then construct some sequence $(\bar{\theta}_n, \bar{\beta}_n, \bar{\Lambda}_n)$ such that the empirical Kullback-Leibler distance $\ell_n(\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n) - \ell_n(\bar{\theta}_n, \bar{\beta}_n, \bar{\Lambda}_n)$ goes to negative infinity. This is a contradiction because $\hat{\psi}_n$ maximizes the log-likelihood function so that $\ell_n(\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n) - \ell_n(\bar{\theta}_n, \bar{\beta}_n, \bar{\Lambda}) \geq 0$ for every $(\bar{\theta}_n, \bar{\beta}_n, \bar{\Lambda})$ in the parameter set. The construction of the contradiction is along the lines of that in Parner (1998) for the gamma frailty model. Briefly, choose $\{\bar{\theta}_n, \bar{\beta}_n\} = \{\theta_0, \beta_0\}$, and choose $\bar{\Lambda}_n$ to be

$$\bar{\Lambda}_n(t) = \int_0^t \Bigg\{ \sum_{i=1}^n \Bigg(M_i(u) - \frac{\int_u^\tau S_0(v|X_i) h_0(v) dv}{\int_0^\tau S_0(v|X_i) h_0(v) dv} \Bigg) e^{\beta_0^{\mathrm{T}} X_i} \Bigg\}^{-1} d\Bigg\{ \sum_{j=1}^n N_j(u) \Bigg\},$$

where $S_0(\cdot|X_i) = \exp\{-\int_0^{\cdot} e^{\beta_0^{\mathrm{T}} X} d\Lambda_0(u)\}$ and $h_0(\cdot) = h_{\theta_0}(\cdot)$. It can be easily shown that using the mean-zero estimating equation (3.4), $\bar{\Lambda}_n(t)$ converges to Λ_0 almost surely and uniformly in t by an application of the Glivenko-Cantelli theorem. Using a technical argument similar to that in Parner (1998), we can show that $\ell_n(\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n) - \ell_n(\bar{\theta}_n, \bar{\beta}_n, \bar{\Lambda}_n) \to -\infty$ as $n \to \infty$, but this is impossible so $\hat{\Lambda}_n$ must stay bounded.

As $\hat{\Lambda}_n$ stays bounded, we can apply Helly's selection principle to find a convergent subsequence of $(\hat{\theta}_n, \hat{\beta}_{n_k}, \hat{\Lambda}_{n_k})$ for an arbitrary subsequence from the sequence indexed by $\{1, \ldots, n\}$. We can then show that by the strong law of large numbers, such convergent subsequence must converge to $(\theta_0, \beta_0, \Lambda_0)$, using the classical Kullback-Leibler information approach. For any given subsequence $\{n_k\}$, we can identify a further subsequence of $(\hat{\theta}_{n_k}, \hat{\beta}_{n_k}, \hat{\Lambda}_{n_k})$ that converges to $(\theta_0, \beta_0, \Lambda_0)$. Helly's selection theorem implies that the entire sequence $(\hat{\theta}_n, \hat{\beta}_n, \hat{\Lambda}_n(t))$ must converge to $(\theta_0, \beta_0, \Lambda_0(t))$ for each t. By the assumption that

$\Lambda_0(\cdot)$ is continuous, the convergence of $\hat{\Lambda}_n(t)$ at each t is also uniform in t. The convergence can be made almost-surely convergence by carrying out the proof for a fixed ω in the underlying probability space Ω, and applying the law of large numbers only countable many times.

Appendix C. Proof of asymptotic normality

C.1. Score equations

We first calculate the score equations for a single subject with the log-likelihood function

$$\ell(\psi) = \int_0^\tau \Big(\beta^{\mathrm{T}} X + \log d\Lambda(u)\Big) dN(u) - \int_0^\tau M(u) e^{\beta^{\mathrm{T}} X} d\Lambda(u) + \log h_\theta(A) - \log \int_0^\tau S(u|X) h_\theta(u) du, \tag{C.1}$$

where $N(t) = 1(A < Y \le t)\delta$ and $M(t) = 1(Y \ge t)1(Y > A)$. The score function for the infinite dimensional parameter $\Lambda(\cdot)$ is calculated through a submodel $d\Lambda_\eta(\cdot) = (1 + \eta\phi(\cdot))d\Lambda(\cdot)$, where $\phi(\cdot)$ is a bounded and integrable function, and $\eta > 0$ is a constant. By taking the derivative $\ell(\theta, \beta, \Lambda_\eta)$ with respect to η and evaluating at $\eta = 0$, the score operator for Λ has the form

$$\dot{\ell}_3(\psi, \mathcal{O})(\phi) = \int_0^\tau \phi(u) \left[dN(u) - \left\{ M(u) - \frac{\int_u^\tau S(v|X) h_\theta(v) dv}{\int_0^\tau S(v|X) h_\theta(v) dv} \right\} e^{\beta^{\mathrm{T}} X} d\Lambda(u) \right]. \tag{C.2}$$

Taking $\phi(\cdot) = 1(\cdot \le t)$ in (C.2), we have the equivalent score function for Λ

$$\dot{\ell}_3(t, \psi, \mathcal{O}) = \int_0^t \left[dN(u) - \left\{ M(u) - \frac{\int_u^\tau S(v|X) h_\theta(v) dv}{\int_0^\tau S(v|X) h_\theta(v) dv} \right\} e^{\beta^{\mathrm{T}} X} d\Lambda(u) \right]. \tag{C.3}$$

The score function for β has the form

$$\begin{aligned} \dot{\ell}_2(\psi, \mathcal{O}) &= \int_0^\tau X \left[dN(u) - \left\{ M(u) - \frac{\int_u^\tau S(v|X) h_\theta(v) dv}{\int_0^\tau S(v|X) h_\theta(v) dv} \right\} e^{\beta^{\mathrm{T}} X} d\Lambda(u) \right] \\ &= \int_0^\tau X d\dot{\ell}_3(t, \mathcal{O}, \psi). \end{aligned} \tag{C.4}$$

The score function for θ has the form

$$\dot{\ell}_1(\psi, \mathcal{O}) = \frac{\dot{h}_\theta(A)}{h_\theta(A)} - \frac{\int S(u|X) \dot{h}_\theta(u) du}{\int S(u|X) h_\theta(u) du}, \tag{C.5}$$

where $\dot{h}_\theta(\cdot)$ is the first partial derivative of $h_\theta(\cdot)$ with respect to θ.

C.2. Fréchet derivative

We first confirm that the expectation of the estimating equations is Fréchet differentiable and its Fréchet derivative is continuously invertible. Denote the expectation of the estimating equations by $U_0(\psi) = \{U_{10}(\psi), U_{20}(\psi), U_{30}(\cdot, \psi)\}$, where

$$U_{10}(\psi) = E_0\left[\left\{\frac{\dot{h}_\theta(A)}{h_\theta(A)}1(Y > A) - \frac{\int_0^\tau S(u|X)\dot{h}_\theta(u)du}{\int_0^\tau S(u|X)h_\theta(u)du}\right\}\right], \tag{C.6}$$

$$U_{20}(\psi) = \int_0^\tau E_0\left[X\left\{dN(u) - \left(M(u) - \frac{\int_u^\tau S(v|X)h_\theta(v)dv}{\int_0^\tau S(v|X)h_\theta(v)dv}\right)e^{\beta^{\mathrm{T}}X}d\Lambda(u)\right\}\right], \tag{C.7}$$

$$U_{30}(t, \psi) = \int_0^t E_0\left[\left\{dN(u) - \left(M(u) - \frac{\int_u^\tau S(v|X)h_\theta(v)dv}{\int_0^\tau S(v|X)h_\theta(v)dv}\right)e^{\beta^{\mathrm{T}}X}d\Lambda(u)\right\}\right]. \tag{C.8}$$

The Fréchet differentiability of $U_0(\psi)$ at $\psi = \psi_0$ can be verified by the definition, and the derivation can be calculated using the Gâteaux variations of $U_0(\psi)$. This is done by the differentiation of $U_0(\psi_\eta)$ with respect to η and evaluated at $\eta = 0$, where $\psi_\eta = (\theta_\eta, \beta_\eta, \Lambda_\eta) = (\beta_0, \theta_0, \Lambda_0(\cdot)) + \eta(\beta, \theta, \Lambda(\cdot))$.

The Gâteaux derivative of $U_{10}(\psi)$ evaluated at ψ_0 is $-\{s_{11}(\theta) + s_{12}(\beta) + s_{13}(\Lambda)\}$, where

$$\begin{aligned}
s_{11}(\theta) &\equiv \frac{\partial}{\partial\eta}U_{10}(\theta_\eta, \beta_0, \Lambda_0)\Big|_{\eta=0} = \theta\kappa_0,\\
s_{12}(\beta) &\equiv \frac{\partial}{\partial\eta}U_{10}(\theta_0, \beta_\eta, \Lambda_0)\Big|_{\eta=0} = \beta^{\mathrm{T}}\int_0^\tau \kappa_1^{(1)}(u)d\Lambda_0(u),\\
s_{13}(\Lambda) &\equiv \frac{\partial}{\partial\eta}U_{10}(\theta_0, \beta_0, \Lambda_\eta)\Big|_{\eta=0} = \int_0^\tau\left(\int_u^\tau \kappa_1^{(0)}(v)dv\right)d\Lambda(u).
\end{aligned}$$

The Gâteaux derivative of $U_{20}(\psi)$ evaluated at ψ_0 is $-\{s_{21}(\theta) + s_{22}(\beta) + s_{23}(\Lambda)\}$, where

$$\begin{aligned}
s_{21}(\theta) &\equiv \frac{\partial}{\partial\eta}U_{10}(\theta_\eta, \beta_0, \Lambda_0)\Big|_{\eta=0} = \theta\int_0^\tau \kappa_1^{(1)}(u)d\Lambda_0(u),\\
s_{22}(\beta) &\equiv \frac{\partial}{\partial\eta}U_{10}(\theta_0, \beta_\eta, \Lambda_0)\Big|_{\eta=0} = \beta^{\mathrm{T}}\left[\int_0^\tau\{\kappa_2^{(2)}(u) + \kappa_3^{(2)}(u)\}d\Lambda_0(u)\right],\\
s_{23}(\Lambda) &\equiv \frac{\partial}{\partial\eta}U_{10}(\theta_0, \beta_0, \Lambda_\eta)\Big|_{\eta=0} = \int_0^\tau\{\kappa_2^{(1)}(u) + \kappa_3^{(1)}(u)\}d\Lambda(u).
\end{aligned}$$

The Gâteaux derivative of $U_{30}(t, \psi)$ evaluated at ψ_0 is $-\{s_{31}(\theta)(t) + s_{32}(\beta)(t) +$

$s_{33}(\Lambda)(t)\}$, where

$$
\begin{aligned}
s_{31}(\theta)(t) &\equiv \frac{\partial}{\partial\eta} U_{30}(t,\theta_\eta,\beta_0,\Lambda_0)\Big|_{\eta=0} = \theta\int_0^t \kappa_1^{(0)}(u)d\Lambda_0(u),\\
s_{32}(\beta)(t) &\equiv \frac{\partial}{\partial\eta} U_{30}(t,\theta_0,\beta_\eta,\Lambda_0)\Big|_{\eta=0} = \beta^{\mathrm{T}}\bigg[\int_0^t \{\kappa_2^{(1)}(u)+\kappa_3^{(1)}(u)\}d\Lambda_0(u)\bigg],\\
s_{33}(\Lambda)(t) &\equiv \frac{\partial}{\partial\eta} U_{30}(t,\theta_0,\beta_0,\Lambda_\eta)\Big|_{\eta=0}\\
&= \int_0^t \bigg\{\int_0^\tau \kappa_4(u,v)d\Lambda(v)\bigg\}d\Lambda_0(u) + \int_0^t \kappa_3^{(0)}(u)d\Lambda(u).
\end{aligned}
$$

It follows by straightforward calculation that the Fréchet derivative $\dot{U}_0(\psi_0)$ has the form

$$
\dot{U}_{\psi_0}(\psi) \equiv \begin{pmatrix} s_{11}\ s_{12}\ s_{13}\\ s_{21}\ s_{22}\ s_{23}\\ s_{31}\ s_{32}\ s_{33}\end{pmatrix}\begin{pmatrix}\theta\\ \beta\\ \Lambda\end{pmatrix} \equiv \begin{pmatrix} s_{11}(\theta)+s_{12}(\beta)+s_{13}(\Lambda)\\ s_{21}(\theta)+s_{22}(\beta)+s_{23}(\Lambda)\\ s_{31}(\theta)+s_{32}(\beta)+s_{33}(\Lambda)\end{pmatrix}. \tag{C.9}
$$

Denote the finite dimensional parameters by $\xi = (\theta,\beta)^{\mathrm{T}}$. Take

$$
\begin{aligned}
J_{11} &= \begin{pmatrix} \kappa_0 & \{\int_0^\tau \kappa_1^{(1)}(u)d\Lambda_0(u)\}^{\mathrm{T}}\\ \int_0^\tau \kappa_1^{(1)}(u)d\Lambda_0(u) & \int_0^\tau\{\kappa_2^{(2)}(u)+\kappa_3^{(2)}(u)\}d\Lambda_0(u)\end{pmatrix},\\
J_{21}(t) &= J_{12}(u)^{\mathrm{T}} = \bigg(\int_0^t \kappa_1^{(0)}(u)d\Lambda_0(u)\ \int_0^t\{\kappa_2^{(1)}(u)+\kappa_3^{(1)}(u)\}^{\mathrm{T}}d\Lambda_0(u)\bigg).
\end{aligned}
$$

Then the Fréchet derivative $U_0(\psi_0)$ can be written as

$$
\dot{U}_0(\psi) = \begin{pmatrix}\sigma_{11}\ \sigma_{12}\\ \sigma_{21}\ \sigma_{22}\end{pmatrix}\begin{pmatrix}\xi\\ \Lambda\end{pmatrix} = \begin{pmatrix}\sigma_{11}(\xi)+\sigma_{12}(\Lambda)\\ \sigma_{21}(\xi)+\sigma_{22}(\Lambda)\end{pmatrix},
$$

where $\sigma_{22}(\Lambda) = s_{33}(\Lambda)$,

$$
\begin{aligned}
\sigma_{11}(\xi) &= -\begin{pmatrix} s_{11}(\theta)+s_{12}(\beta)\\ s_{21}(\theta)+s_{22}(\beta)\end{pmatrix} = J_{11}\xi,\\
\sigma_{21}(\xi)(t) &= -\{s_{31}(\theta)(t)+s_{32}(\beta)(t)\} = J_{21}(t)\xi,\\
\sigma_{12}(\Lambda) &= \begin{pmatrix} s_{13}(\Lambda)\\ s_{23}(\Lambda)\end{pmatrix} = \int_0^\tau J_{12}(u)d\Lambda(u).
\end{aligned}
$$

If the inverse of $\dot{U}_{\psi_0}$ exists, then its inverse must have the form:

$$
\dot{U}_{\psi_0}^{-1}(\psi) \equiv \begin{pmatrix} \Sigma_{11}^{-1}+\Sigma_{11}^{-1}\sigma_{12}\Phi^{-1}\sigma_{21}\sigma_{11}^{-1} & -\sigma_{11}^{-1}\sigma_{12}\Phi^{-1}\\ \Phi^{-1}\sigma_{21}\sigma_{11}^{-1} & \Phi^{-1}\end{pmatrix}\begin{pmatrix}\xi\\ \Lambda\end{pmatrix}, \tag{C.10}
$$

where $\Phi = \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12}$. Thus, to show that $\dot{U}_{\psi_0}$ is continuously invertible, we need only show that σ_{11} and Φ are continuously invertible.

First, the operator σ_{11} is identical to the symmetric matrix J_{11}. This matrix is the Fisher information on the estimation of (β_0, θ_0) when the baseline function Λ_0 is known. It is reasonable to assume that the information matrix J_{11} is positive definite so that it has inverse J_{11}^{-1}. It follows that the operator σ_{11} is continuously invertible with the inverse $\sigma_{11}^{-1}(\xi) = J_{11}^{-1}\xi$.

To determine whether the operator Φ is invertible, note that it has the form

$$\begin{aligned}\Phi(\Lambda) =&\sigma_{22}(\Lambda) - \sigma_{21}\sigma_{11}^{-1}\sigma_{12}(\Lambda)\\ =&\int_0^t \left\{ \int_0^\tau \kappa_4(u,v)d\Lambda(v)\right\}d\Lambda_0(u) + \int_0^t \kappa_3^{(0)}(u)d\Lambda(u)\\ &- \int_0^\tau J_{21}(t)J_{11}^{-1}J_{12}(u)d\Lambda(u).\end{aligned}$$

Showing that the operator Φ is continuously invertible is equivalent to showing that there exists a unique solution to the operator equation $\Phi(\Lambda) = \check{\Lambda}$ for each bounded function $\check{\Lambda}$. This can be written as an integral equation

$$\int_0^t \left\{ \int_0^\tau \kappa_4(u,v)d\Lambda(v)\right\}d\Lambda_0(u) + \int_0^t \kappa_3^{(0)}(u)d\Lambda(u) - \int_0^\tau J_{21}(t)J_{11}^{-1}J_{12}(u)d\Lambda(u) = \check{\Lambda}(t).$$

Taking the derivative with respect to t on both sides,

$$\lambda_0(t)\int_0^\tau \kappa_4(t,v)d\Lambda(v)+\kappa_3^{(0)}(t)d\Lambda(t)-\int_0^\tau \dot{J}_{21}(t)J_{11}^{-1}J_{12}(u)d\Lambda(u)=d\check{\Lambda}(t), \quad \text{(C.11)}$$

where

$$\dot{J}_{21}(t) = \left(\kappa_1^{(0)}(t)\lambda_0(t)\left\{\kappa_2^{(1)}(t) + \kappa_3^{(1)}(t)\right\}^{\mathrm{T}}\lambda_0(t).\right)$$

This can be written as a Fredholm integral equation of the second type:

$$\frac{d\check{\Lambda}(t)}{\kappa_3^{(0)}(t)} = d\Lambda(t) - \int_0^\tau K(t,u)d\Lambda(u), \quad \text{(C.12)}$$

$$K(t,u) = \frac{\lambda_0(t)}{\kappa_3^{(0)}(t)}\left\{ \begin{pmatrix} \kappa_1^{(0)}(t) \\ \kappa_2^{(1)}(t) + \kappa_3^{(1)}(t)\end{pmatrix}^{\mathrm{T}} J_{11}^{-1}J_{12}(u) - \kappa_4(t,u)\right\}. \quad \text{(C.13)}$$

The existence and uniqueness of its solution for given $\check{\Lambda}(\cdot)$ are well known (Tricomi (1985)). Under the regularity conditions, we have

$$\int \left|\frac{\check{\lambda}(t)}{\kappa_3^{(0)}(t)}\right|^2 dt < \infty, \qquad \int\int |K(t,u)|^2 dtdu < \infty.$$

By the classical existence and uniqueness theorems for Fredholm integral equations, there exists one and only one solution to (C.11). It follows that the inverse operator Φ^{-1} exists and has the form

$$\Phi^{-1}(\check{\Lambda}) = \int_0^t \frac{d\check{\Lambda}(u)}{\kappa_3^{(0)}(u)} + \int_0^\tau \int_0^t \frac{R(v,u)}{\kappa_3^{(0)}(u)} dv d\check{\Lambda}(u),$$

where $R(t,u)$ is independent of $\check{\Lambda}(\cdot)$ and satisfies the equation

$$R(t,u) = K(t,u) + \int K(t,v)R(v,t)dv.$$

C.3. Weak convergence

Both the score function U_n and its expectation U_0 are defined on the parameter set $\mathcal{B} \times \mathcal{A}$, where $\mathcal{B}$ is an open convex set in $\mathbb{R}^{p+q}$, and $\mathcal{A}$ is in the space of functions of bounded variation. As the true value ψ_0 satisfies the population estimating equations $U_0(\psi_0) = 0$, we have

$$\sqrt{n}U_n(\psi_0) = \sqrt{n}\Big\{U_{1n}(\psi_0) - U_{10}(\psi_0), U_{2n}(\psi_0) - U_{20}(\psi_0), U_{3n}(t,\psi_0) - U_{30}(t,\psi_0)\Big\}.$$

As the summation of independently and identically distributed (i.i.d.) random vectors, $\sqrt{n}\{U_{1n}(\psi_0) - U_{10}(\psi_0), U_{2n}(\psi_0) - U_{20}(\psi_0)\}$ converges in law to $\mathbb{W}_1$ by the multivariate central limit theorem. The process $\sqrt{n}\{U_{3n}(t,\psi_0) - U_{30}(t,\psi_0)\}$ is a sum of i.i.d. processes of bounded variation. By a lemma for the central limit theorem for processes of bounded variation (van der Vaart and Wellner (1996, Example 2.11.16)), $\sqrt{n}\{U_{3n}(t,\psi_0) - U_{30}(t,\psi_0)\}$ converges to a tight Gaussian process, denoted by $\mathbb{W}_2(\cdot)$. The weak convergence of $\sqrt{n}U_n(\psi_0)$ to $\mathbb{W} = \{\mathbb{W}_1, \mathbb{W}_2(\cdot)\}$ follows by the continuous mapping theorem (van der Vaart and Wellner (1996, Thm. 1.3.6)).

C.4. Stochastic approximation

To confirm the conditions for the Z-theorem (van der Vaart and Wellner (1996, Thm. 3.3.1)), we prove the stochastic approximation

$$\begin{aligned}&\sqrt{n}\big\{(U_n - U_0)(\psi_n) - (U_n - U_0)(\psi_0)\big\}\\ &\quad \equiv \sqrt{n}\big[\{U_n(\cdot,\psi_n) - U_0(\cdot,\psi_n)\} - \{U_n(\cdot,\psi_0) - U_0(\cdot,\psi_0)\}\big] = o_{P^*}(1).\end{aligned}$$

Let $\mathbb{P}_n$ be the empirical measure and let $\dot{\ell}(t,\psi,\mathcal{O}) = \{\dot{\ell}_1(\psi,\mathcal{O}), \dot{\ell}_2(\psi,\mathcal{O}), \dot{\ell}_3(t,\psi,\mathcal{O})\}$. Write $U_n(\psi) = \mathbb{P}_n\dot{\ell}(\cdot,\psi,\mathcal{O}) = n^{-1}\sum_{i=1}^n \dot{\ell}(\cdot,\psi,\mathcal{O}_i)$. Denote the empirical process by $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - P_0 f)$, where P_0 denote the expectation under the true value ψ_0. Then $\sqrt{n}(U_n - U_0)(\psi) = \mathbb{G}_n\dot{\ell}(\cdot,\psi,\mathcal{O})$ is the empirical

process indexed by the class of functions $\{\dot{\ell}(t,\psi,\mathcal{O}), \psi \in \mathcal{B}\times\mathcal{A}, t\in[0,\tau)\}$. Let the norm $\|\cdot\|_{\mathcal{H}}$ on $\mathcal{H}=\mathcal{B}\times\mathcal{A}$ be defined as $\|\psi\|_{\mathcal{H}} = |\beta| + \|\Lambda\|_\infty$. Then the stochastic condition is

$$\|\mathbb{G}_n\dot{\ell}(t,\hat{\psi}_n,\mathcal{O}) - \mathbb{G}_n\dot{\ell}(t,\psi_0,\mathcal{O})\|_{\mathcal{H}} = o_{P^*}(1).$$

We follow van der Vaart and Wellner (1996, Lemma 3.3.5) in proving the stochastic approximation. First, we show that the difference of the score functions $\{\dot{\ell}(t,\psi,\mathcal{O}) - \dot{\ell}(t,\psi_0,\mathcal{O}) : t\in[0,\tau), \|\psi-\psi_0\|_{\mathcal{H}}<\delta\}$ is P_0-Donsker. This is done by showing that the functions are sum, products and continuous transformations of P_0-Donsker classes.

Under regularity condition 3, the functions $\dot{h}_\theta$ and h_θ satisfy a Lipschitz condition. Thus, there exist functions, F_1 and F_2 such that

$$|\dot{h}_{\theta_1}(a) - \dot{h}_{\theta_2}(a)| \le \|\theta_1-\theta_2\|_2 F_1(a), \quad \text{and} \quad |h_{\theta_1}(a) - h_{\theta_2}(a)| \le \|\theta_1-\theta_2\|_2 F_2(a),$$

where $E_0\{F_1(\tilde{A})^2\} < \infty$ and $E_0\{F_2(\tilde{A})^2\} < \infty$. It follows that the classes of functions $\{\dot{h}_\theta : \|\psi-\psi_0\|_{\mathcal{H}}<\delta\}$ and $\{h_\theta : \|\psi-\psi_0\|_{\mathcal{H}}<\delta\}$ are P_0-Donsker (van der Vaart and Wellner (1996, Thm. 2.7.11)). We have assumed that $h_\theta \ge \delta_1$ for some constant $\delta_1>0$ for every θ satisfying $\|\psi-\psi_0\|_{\mathcal{H}}<\delta$. Then by the permanence of the Donsker property, the class of functions $\{\dot{h}_\theta/h_\theta : \|\psi-\psi_0\|_{\mathcal{H}}<\delta\}$ is also P_0-Donsker (van der Vaart and Wellner (1996, Example 2.10.9)).

Now

$$\begin{aligned}
&\dot{\ell}_3(t,\psi,\mathcal{O}) - \dot{\ell}_3(t,\psi_0,\mathcal{O})\\
&\quad= \int_0^t M(u)e^{\beta_0^{\mathrm{T}}X}d\Lambda_0(u) - \int_0^t Y(u)e^{\beta^{\mathrm{T}}X}d\Lambda(u)\\
&\qquad+\int_0^t \frac{S(v|X)h_\theta(v)dv}{\int_0^\tau S(v|X)h_\theta(v)dv}e^{\beta^{\mathrm{T}}X}d\Lambda(u) - \int_0^t \frac{S_0(v|X)h_0(v)dv}{\int_0^\tau S_0(v|X)h_0(v)dv}e^{\beta_0^{\mathrm{T}}X}d\Lambda_0(u).
\end{aligned}$$

Under the regularity conditions, β is in the compact set $\mathcal{B}$ so that $|\beta|$ is bounded. We assume that $E_0\{\exp(2|\beta^{\mathrm{T}}X|)\} < \infty$ for every β in $\mathcal{B}$. It follows that the class of functions $\{\exp(\beta^{\mathrm{T}}X) : \|\psi-\psi_0\|_{\mathcal{H}} < \delta\}$ is P_0-Donsker. The class of all monotone functions on a bound set $[0,\tau)$ is P_0-Donsker (van der Vaart and Wellner (1996, Thm. 2.7.5)). It then follows that the class of functions $\{\int_0^t Y(u)d\Lambda(u) : t\in[0,\tau), \|\psi-\psi_0\|_{\mathcal{H}}<\delta\}$ is P_0-Donsker, and so is the class of functions $\{\int_0^t Y(u)e^{\beta^{\mathrm{T}}X}d\Lambda(u) : t\in[0,\tau), \|\psi-\psi_0\|_{\mathcal{H}}<\delta\}$. Furthermore, the following classes of functions are also P_0-Donsker:

$$\begin{aligned}
&\{\exp\{-\exp(\beta^{\mathrm{T}}X)\Lambda(t)h_\theta(t)\} : t\in[0,\tau), \|\psi-\psi_0\|_{\mathcal{H}}<\delta\},\\
&\{\exp\{-\exp(\beta^{\mathrm{T}}X)\Lambda(t)\dot{h}_\theta(t)\} : t\in[0,\tau), \|\psi-\psi_0\|_{\mathcal{H}}<\delta\}.
\end{aligned}$$

The map $\varphi : f \mapsto \int_0^\tau f(u)du$ is a continuous map for f in the set of bounded variations. Hence, it follows that by the continuous mapping theorem, the following classes of functions are P_0-Donsker: $\{\int_t^\tau S(u|X)h_\theta(u)du : t \in [0,\tau), \|\psi - \psi_0\|_{\mathcal{H}} < \delta\} = \{\int_0^\tau 1(u \geq t)\exp\{-\exp(\beta^{\mathrm{T}}X)\Lambda(u)h_\theta(u)du\} : \|\psi - \psi_0\|_{\mathcal{H}} < \delta\}$, and $\{\int_t^\tau S(u|X)\dot{h}_\theta(u)du : t \in [0,\tau), \|\psi - \psi_0\|_{\mathcal{H}} < \delta\} = \{\int_0^\tau 1(u \geq t)\exp\{-\exp(\beta^{\mathrm{T}}X)\Lambda(u)\dot{h}_\theta(u)du\} : \|\psi - \psi_0\|_{\mathcal{H}} < \delta\}$. Finally, by the permanent property of Donsker classes and assuming that $\beta^{\mathrm{T}}X$ is bounded for every β in $\mathcal{B}$, we have confirmed the following classes of functions are P_0-Donsker:

$$\{\dot{\ell}_1(\psi,\mathcal{O}) - \dot{\ell}_1(\psi_0,\mathcal{O}) : \|\psi - \psi_0\|_{\mathcal{H}} < \delta\},$$
$$\{\dot{\ell}_2(\psi,\mathcal{O}) - \dot{\ell}_2(\psi_0,\mathcal{O}) : \|\psi - \psi_0\|_{\mathcal{H}} < \delta\},$$
$$\{\dot{\ell}_3(t,\psi,\mathcal{O}) - \dot{\ell}_3(t,\psi_0,\mathcal{O}) : t \in [0,\tau), \|\psi - \psi_0\|_{\mathcal{H}} < \delta\}.$$

When $\|\psi - \psi_0\|_{\mathcal{H}} \to 0$, $\dot{\ell}(t,\psi,\mathcal{O})$ converges to $\dot{\ell}(t,\psi_0,\mathcal{O})$ for each t. The convergence also holds in the square moment by the dominated convergence theorem. Hence,

$$\sup_{t\in[0,\tau)} E_0\|\dot{\ell}(t,\psi,\mathcal{O}) - \dot{\ell}(t,\psi_0,\mathcal{O})\|_{\mathcal{H}}^2 \to 0.$$

The stochastic approximation of $(U_n - U_0)(\hat{\psi}_n)$ to $(U_n - U_0)(\psi_0)$ follows by applying a result of van der Vaart and Wellner (1996, Lemma 3.3.5).

References

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1992). *Statistical Models Based on Counting Processes*. Springer, New York.

Asgharian, M., M'Lan, C. E. and Wolfson, D. B. (2002). Length-biased sampling with right censoring: an unconditional approach. *J. Amer. Statist. Assoc.* **97**, 201-209.

Asgharian, M. and Wolfson, D. B. (2005). Asymptotic behavior of the unconditional NPMLE of the lengthbiased survivor function from right censored prevalent cohort data. *Ann. Statist.* **33**, 2109-2131.

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

Brookmeyer, R. and Gail, M. (1987). Biases in prevalent cohorts. *Biometrics* **43**, 739-749.

Carone, M., Asgharian, M. and Jewell, N. P. (2014). Estimating the lifetime risk of dementia in the Canadian elderly population using cross-sectional cohort survival data. *J. Amer. Statist. Assoc.* **109**, 24-25.

Carpenter, J. and Lynch, A. (1999). Survivorship bias and attrition effects in measures of performance persistence. *J. Financial Econom.* **54**, 337-374.

Chen, H. Y. and Little, R. J. A. (1999). Proportional hazards regression with missing covariates. *J. Amer. Statist. Assoc.* **94**, 896-908.

Chen, K. (2001). Parametric models for response-biased sampling. *J. Roy. Statist. Soc. Ser. B* **63**, 775-789.

Gilbert, P. B., Lele, S. R. and Vardi, Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika* **86**, 27-43.

Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I (with discussion). *Scand. J. Statist.* **16**, 97-128.

Gross, S. T. and Lai, T. L. (1996). Nonparametric estimation and regression analysis with left-truncated and right-censored data. *J. Amer. Statist. Assoc.* **91**, 1166-1180.

Heckman, J. (1990). Varieties of selection bias. *Amer. Econom. Rev.* **80**, 313-318.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such model. *Ann. Econom. Social Measurement* **5**, 475-492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47**, 153-161.

Horne, J. S., Garton, E. O. and Sager-Fradkin, K. A. (2007). Correcting home-range models for observation bias. *J. Wildlife Management* **71**, 996-1001.

Kalbfleisch, J. D. and Lawless, J. F. (1989). Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS. *J. Amer. Statist. Assoc.* **84**, 360-372.

Keiding, N. (1992). Independent delayed entry. In *Survival Analysis: State of the Art*, (Edited by J. P. Klein and P. Goel), 309-326. Kluwer Academic Publishers, Boston.

Kim, J. K., Lu, W. B., Sit, T. and Ying, Z. L. (2013). A unified approach to semiparametric transformation models under general biased sampling schemes. *J. Amer. Statist. Assoc.* **108**, 217-227.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd edition. Springer, New York.

Kvam, P. (2008). Length bias in the measurements of carbon nanotubes. *Technometrics* **50**, 462-467.

Lagakos, S., Barraj, L. and Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with application to aids. *Biometrika* **75**, 515-523.

Lai, T. L. and Ying, Z. (1991). Estimating a distribution function with truncated and censored data. *Ann. Statist.* **19**, 417-442.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge University Press.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182-198.

Oshlack, A. and Wakefield, M. J. (2009). Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct* **4**.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty Model. *Ann. Statist.* **26**, 183-214.

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., Ning, J., Liu, H. and Shen, Y. (2011). Maximum likelihood estimations and EM algorithms with length-biased data. *J. Amer. Statist. Assoc.* **106**, 1434-1449.

Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *J. Amer. Statist. Assoc.* **79**, 516-524.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82**, 605-610.

Smith, T. M. F. (1993). Populations and selection: limitations of statistics. *J. Roy. Statist. Soc. Ser. A* **156**(2), 144-166.

Terwilliger, J., Shannon, W., Lathrop, G., Nolan, J., Goldin, L., Chase, G. and Weeks, D. (1997). True and false positive peaks in genomewide scans: applications of length-biased sampling to linkage mapping. *Amer. J. Human Genetics* **61**, 430-438.

The Canadian Study of Health and Aging Working Group (1994). Canadian study of health and aging: study methods and prevalence of dementia. *Canad. Medical Assoc. J.* **150**, 899-912.

Tricomi, F. G. (1985). *Integral Equations*. Dover Publications, New York.

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

Tsui, K. L., Jewell, N. P. and Wu, C. F. J. (1988). A nonparametric approach to the truncated regression problem. *J. Amer. Statist. Assoc.* **83**, 785-792.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **38**, 290-295.

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.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* **76**, 751-761.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *J. Human Resources* **33**, 127-169.

Wang, M. C. (1989). A semiparametric model for randomly truncated data. *J. Amer. Statist. Assoc.* **84**, 742-748.

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

Wang, M.-C., Brookmeyer, R. and Jewell, N. (1993). Statistical models for prevalent cohort data. *Biometrics* **49**, 1-11.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163-177.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd edition. MIT Press, Cambridge, Mass.

Division of Biostatistics, Baylor College of Medicine Dan L. Duncan Cancer Center, Houston, Texas 77030, USA.

E-mail: haol@bcm.edu

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA.

E-mail: jning@mdanderson.org

Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

E-mail: jingqin@niaid.nih.gov

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA.

E-mail: yshen@mdanderson.org

(Received March 2014; accepted August 2015)