Statistica Sinica **27** (2017), 229-242
doi:http://dx.doi.org/10.5705/ss.202015.0089

SMOOTH COMPOSITE LIKELIHOOD ANALYSIS OF LENGTH-BIASED AND RIGHT-CENSORED DATA WITH AFT MODEL

Xuerong Chen[1], Na Hu[2] and Jianguo Sun[3,4]

[1]*Southwestern University of Finance and Economics,*
[2]*Boehringer Ingelheim (China) Investment Co.,*
[3]*University of Missouri and* [4]*Jilin University*

Abstract: This article discusses regression analysis of length-biased and right-censored failure time data arising from the accelerated failure time model. A key feature of such data is the informative censoring induced by the length-biased sampling, and several methods have been proposed in the literature for their analysis. However, these may be less efficient or apply only to limited situations. We propose a kernel-smoothed composite likelihood method for estimation of covariate effects. The proposed estimators are proved to be consistent and asymptotically normal. Simulation studies conducted to assess the finite sample performance of the method suggest that it works well for practical situations. An illustrative example is provided.

Key words and phrases: AFT model, composite likelihood, length-biased, right-censored.

1. Introduction

Length-biased and right-censored failure time data arise naturally in the studies in which samples are drawn only from the individuals with a condition or disease at the time of enrollment Vardi (1989). It is well known that such designs are commonly used due to their relative efficiency in large epidemiological studies among others, especially when the disease is rare (Keiding (1991); Sansgiry and Akman (2000)). On the other hand, in these situations, one has to deal with selection bias as the observed failure time tends to be longer than the actual one from the target population. It is easy to see that length-biased data are a special case of left-truncated data when disease onset follows a stationary Poisson distribution.

A number of methods have been developed in the literature for regression analysis of length-biased and right-censored failure time data under various situations. For example, Wang (1996) considered the problem without censoring under the proportional hazards model and developed a pseudo-likelihood approach. Tsai (2009) and Qin and Shen (2010) studied the same problem in the presence of right censoring, and generalized the partial likelihood approach by using some weighted and bias-adjusted risk sets. Huang and Qin (2012) investigated the same problem and developed a composite partial likelihood method, motivated by the exchangeability of the truncation time and the forward recurrence time. One advantage of their method is that it can be easily carried out by using standard statistical software. Chen and Zhou (2012) discussed the quantile regression of length-biased and right-censored data.

The accelerated failure time (AFT) model is commonly used in failure time data analysis, and many inference procedures have been proposed for it when only right censoring exists (Cox and Oakes (1984); Jin, Lin and Ying (2003); Kalbfleisch and Prentice (2002); Lai and Ying (1991); Zeng and Lin (2007)). In the presence of length bias, a few methods have also been developed for estimation of covariate effects in Shen, Ning and Qin (2009), Ning, Qin and Shen (2011), Ning, Qin and Shen (2014a), and Ning, Qin and Shen (2014b). In particular, Shen, Ning and Qin (2009) gave some inverse weighted estimating equation approaches with the advantage that the resulting estimator has a closed form and thus is easy to determine. However, it may be less efficient and could be biased. Ning, Qin and Shen (2011) proposed a Buckley-James-type estimator that can be more efficient than the former approach, but the estimator does not have a closed form expression and its determination is not easy. Corresponding to these, Ning, Qin and Shen (2014a) presented two rank-based estimating equation methods, and Ning, Qin and Shen (2014b) derived some estimation equations based on the score equations derived from the likelihood functions under the embedded models for transformed failure data. A shortcoming of these estimating equation-based methods is that they are non-smooth to the parameter, which can make numerical computation and variance estimation difficult.

A common feature of all existing methods for fitting the AFT model to length-biased data is that they are estimating equation-based and thus may not be efficient. Corresponding to this, we develop a composite conditional likelihood approach that allows both length bias and right censoring and is expected to improve the efficiency. For this, we describe the AFT model in terms of the hazard function of the related random error, and then construct a composite conditional likelihood by following Huang and Qin (2012). For computational simplicity, we approximate the hazard function of the error by a piecewise constant function, treated as a nuisance function. After profiling out the nuisance function, we can show that the resulting log composite likelihood function converges to a function that involves several unknown distribution functions. To further optimize it, a kernel-smoothed function is employed to approximate them. We use the composite likelihood method and the resulting approach is more efficient and has computational simplicity because the optimization of the log composite likelihood function can be easily implemented by existing procedures.

The remainder of the paper is organized as follows. We begin in Section 2 by introducing some notation and assumptions, and discussing some likelihood functions. A kernel smoothing estimation procedure is then presented in Section 3, and the asymptotic properties of the resulting estimators are established. Section 4 gives some results obtained from a simulation study for evaluating the performance of the proposed method; they indicate that it works well for practical situations. An illustrative example is provided in Section 5, and Section 6 contains some discussion and concluding remarks.

2. Notation, Assumptions and Likelihood Functions

2.1. Notation and assumptions

Consider a study in which a failure time of interest is the time between an initial event and an end or failure event. Let $\tilde{T}$ denote the failure time of interest and $\tilde{A}$ the time from the initiating event to an examination before the failure event, and suppose that there exists a vector of covariates $\tilde{\mathbf{X}}$. The study consists only of the subjects with $\tilde{T} \geq \tilde{A} > 0$ and it is assumed that $\tilde{A}$ follows the uniform distribution. Thus, we have length-biased sampling. We use T, A and $\mathbf{X}$ to denote the same as $\tilde{T}$, $\tilde{A}$ and $\tilde{\mathbf{X}}$, but for the subjects included in the study. Let V denote the time from the examination to the failure event. Then we have $T = A + V$ and the joint distribution of (T, A), given $\mathbf{X}$, is the same as that of $(\tilde{T}, \tilde{A}) \,|\, \tilde{T} \geq \tilde{A}, \tilde{\mathbf{X}}$.

We assume that the covariate $\tilde{\mathbf{X}}$ is time-independent and, given $\tilde{\mathbf{X}}$, $\tilde{T}$ can be expressed as

$$\log \tilde{T} = -\boldsymbol{\beta}^T \tilde{\mathbf{X}} + \epsilon, \tag{2.1}$$

where ϵ is a measurement error whose distribution is assumed to be unknown and independent of $\tilde{\mathbf{X}}$. Thus, $\tilde{T}$ follows the accelerated failure time (AFT) model (Kalbfleisch and Prentice (2002)). Let $\lambda(t)$ and $\Lambda(t)$ denote the hazard and cumulative hazard functions of e^{ϵ}, respectively. Then the hazard and cumulative hazard functions of $\tilde{T}$, given $\tilde{\mathbf{X}}$, have the forms

$$\lambda_{\tilde{T}|\tilde{\mathbf{X}}}(t) = \lambda\Big(te^{\boldsymbol{\beta}^T\tilde{\mathbf{X}}}\Big)e^{\boldsymbol{\beta}^T\tilde{\mathbf{X}}} \;\;,\;\; \Lambda_{\tilde{T}|\tilde{\mathbf{X}}}(t) = \Lambda\Big(te^{\boldsymbol{\beta}^T\tilde{\mathbf{X}}}\Big),$$

respectively.

In practice, one may not observe T as there may exist a right censoring time. Then we may only observe the censored failure time $Y = \min(T, A + C)$ and the censoring indicator $\delta = I(V \leq C)$, where C denotes the residual censoring time measured from the examination. In the following, we assume that (A, V) is independent of C given $\mathbf{X}$. Note that for the situation here, however, we do not have independent right censoring on T since T and $A+C$ share the variable A. Thus, T and $A+C$ are related, or T is subject to informative censoring.

2.2. Likelihood functions

Suppose that the study consists of n independent subjects and the observed data are $\{(Y_i, A_i, \delta_i, \mathbf{X}_i);\ i=1,\ldots,n\}$, where Y_i, A_i, δ_i, and $\mathbf{X}_i$ are Y, A, δ, and $\mathbf{X}$, but associated with subject i. Let $f_{\tilde{T}|\tilde{\mathbf{X}}}(t|\mathbf{x})$ and $S_{\tilde{T}|\tilde{\mathbf{X}}}(t|\mathbf{x})$ denote the density and survival functions of $\tilde{T}$ given $\tilde{\mathbf{X}}$, respectively. As the joint distribution of (T, A), given $\mathbf{X}$, is

$$\frac{f(t|\mathbf{x})}{\mu(\mathbf{x})} I(t \geq a), \quad \text{where} \quad \mu(\mathbf{x}) = \int t f(t|\mathbf{x}) dt, \tag{2.2}$$

the full likelihood function of $\boldsymbol{\beta}$ and Λ has the form

$$\begin{aligned} L_F &= \frac{1}{n}\prod_{i=1}^{n} \frac{f_{\tilde{T}|\tilde{\mathbf{X}}}(Y_i|\mathbf{X}_i)^{\delta_i} S_{\tilde{T}|\tilde{\mathbf{X}}}(Y_i|\mathbf{X}_i)^{1-\delta_i}}{\mu(\mathbf{X}_i)} \\ &= \frac{1}{n}\prod_{i=1}^{n} \frac{\Big[\lambda\big(Y_i e^{\boldsymbol{\beta}^T\mathbf{X}_i}\big)e^{\boldsymbol{\beta}^T\mathbf{X}_i}\Big]^{\delta_i} \exp\Big(-\Lambda\big(Y_i e^{\boldsymbol{\beta}^T\mathbf{X}_i}\big)\Big)}{\int_0^\infty t\lambda\big(te^{\boldsymbol{\beta}^T\mathbf{X}_i}\big)e^{\boldsymbol{\beta}^T\mathbf{X}_i}\exp\Big(-\Lambda\big(te^{\boldsymbol{\beta}^T\mathbf{X}_i}\big)\Big)dt}. \end{aligned}$$

For estimation of $\boldsymbol{\beta}$ and Λ, it is apparent that one can directly maximize the full likelihood function. On the other hand, it is easy to see that the maximization process is computationally cumbersome because of the complex integration involved in the denominator, and the way that Λ is involved. To deal with this, we consider the composite conditional likelihood method (Arnold and Strauss (1988)).

To construct a composite conditional likelihood, note that $V = T - A$ represents the residual survival time from the examination and, in the absence of right censoring, it follows from (2.2) that given $\mathbf{X}$, the pair (A, V) has an exchangable joint density function

$$f_{A,V}(a, v|\mathbf{x}) = \frac{f_{\tilde{T}|\tilde{\mathbf{X}}}(a+v|\mathbf{x})}{\mu(\mathbf{x})}, \quad a \geq 0, v \geq 0.$$

Furthermore, it is easy to show that the conditional distribution of T given A is the same as that of T given V. This motivates the composite conditional likelihood

$$\mathcal{L}_C^* = \frac{1}{n}\prod_{i=1}^{n} \frac{f_{\tilde{T}|\tilde{\mathbf{X}}}(T_i|\mathbf{X}_i)}{S_{\tilde{T}|\tilde{\mathbf{X}}}(A_i|\mathbf{X}_i)} \times \frac{f_{\tilde{T}|\tilde{\mathbf{X}}}(T_i|\mathbf{X}_i)}{S_{\tilde{T}|\tilde{\mathbf{X}}}(V_i|\mathbf{X}_i)} \tag{2.3}$$

if there is no right censoring.

In the presence of right censoring, the residual failure time V is unobservable for censored subjects and thus the composite conditional likelihood function $\mathcal{L}_C^*$ is not available. For this, define $V^0 = \min\{T - A, C\}$, the observed residual failure time. It can be shown that, although the joint distribution of V^0 and A are not exchangeable, the conditional density function of A given V^0 is still identical to that of V^0 given A for uncensored failure times:

$$f_{A|V^0,\delta=1,\mathbf{X}}(a,v) = f_{V^0|A,\delta=1,\mathbf{X}}(a,v|\mathbf{x}) = \frac{f_{\tilde{T}|\tilde{\mathbf{X}}}(a+v|\mathbf{x})}{S_{\tilde{T}|\tilde{\mathbf{X}}}(v|\mathbf{x})}, \quad a \geq 0, v \geq 0.$$

Then by following Huang and Qin (2012), we can construct the composite conditional likelihood as

$$\begin{aligned}\mathcal{L}_C &= \frac{1}{n}\prod_{i=1}^{n}\Big[\frac{f_{\tilde{T}|\tilde{\mathbf{X}}}(Y_i|\mathbf{X}_i)}{S_{\tilde{T}|\tilde{\mathbf{X}}}(A_i|\mathbf{X}_i)} \times \frac{f_{\tilde{T}|\tilde{\mathbf{X}}}(Y_i|\mathbf{X}_i)}{S_{\tilde{T}|\tilde{\mathbf{X}}}(V_i^0|\mathbf{X}_i)}\Big]^{\delta_i}\Big[\frac{S_{\tilde{T}|\tilde{\mathbf{X}}}(Y_i|\mathbf{X}_i)}{S_{\tilde{T}|\tilde{\mathbf{X}}}(A_i|\mathbf{X}_i)}\Big]^{(1-\delta_i)} \\ &= \frac{1}{n}\prod_{i=1}^{n}\frac{\Big[\lambda\big(Y_i e^{\boldsymbol{\beta}^T\mathbf{X}_i}\big)e^{\boldsymbol{\beta}^T\mathbf{X}_i}\Big]^{2\delta_i}\exp\Big(-(1+\delta_i)\Lambda\big(Y_i e^{\boldsymbol{\beta}^T\mathbf{X}_i}\big)\Big)}{\exp\Big(-\Lambda\big(A_i e^{\boldsymbol{\beta}^T\mathbf{X}_i}\big) - \delta_i\Lambda\big(V_i^0 e^{\boldsymbol{\beta}^T\mathbf{X}_i}\big)\Big)}.\end{aligned}$$

The resulting log composite conditional likelihood function of $\boldsymbol{\beta}$ and Λ is then proportional to

$$\begin{aligned}\ell_C = \frac{1}{n}\sum_{i=1}^{n}\Big[&2\delta_i\boldsymbol{\beta}^T\mathbf{X}_i + 2\delta_i\log\lambda(e^{R_i(\boldsymbol{\beta})}) - (1+\delta_i)\Lambda(e^{R_i(\boldsymbol{\beta})}) \\ &+\Lambda(e^{H_i(\boldsymbol{\beta})}) + \delta_i\Lambda(e^{I_i(\boldsymbol{\beta})})\Big], \end{aligned} \tag{2.4}$$

where $R_i(\boldsymbol{\beta}) = \log(Y_i e^{\boldsymbol{\beta}^T\mathbf{X}_i})$, $H_i(\boldsymbol{\beta}) = \log(A_i e^{\boldsymbol{\beta}^T\mathbf{X}_i})$, and $I_i(\boldsymbol{\beta}) = \log(V_i^0 e^{\boldsymbol{\beta}^T\mathbf{X}_i})$. For estimation of $\boldsymbol{\beta}$ and Λ, although it seems natural to maximize (2.4), the process may not be easy, or yield efficient estimators, due to the nonsmoothness; One may want to approximate Λ by some smooth functions first.

Although the key components of our method are the smoothed likelihood approach in Zeng and Lin (2007) and the composite likelihood in Huang and Qin (2012), their combination process is not straightforward. The extension of the composite likelihood approach from Cox model to the AFT model with length-biased data is not easy. The baseline hazard function under Cox model, is simply a function of time t, the baseline hazard function under the AFT model also involves the parameter of interest. Further compared to right-censored data, the structure of length-biased and right-censored data is more complicated, it yields informative censoring. That makes kernel smoothing process of the resulting likelihood function more difficult. In all, the derivation of the kernel-smoothed log composite likelihood function and the asymptotic properties are more difficult and complicated.

3. Kernel-Smoothed Composite Likelihood Estimation

3.1. Estimation procedure

To approximate the cumulative hazard function Λ by some smooth functions, we partition the interval $[0,\tau]$ into J_n equally spaced intervals $0 = t_0 < t_1 < \cdots < t_{J_n} = \tau$, where $\tau = \sup_{\boldsymbol{\beta}} \big\{ \big(R_i(\boldsymbol{\beta}), H_i(\boldsymbol{\beta}), I_i(\boldsymbol{\beta})\big), i = 1, \ldots, n \big\}$. We consider the piecewise constant hazard function

$$\lambda(t) = \sum_{k=1}^{J_n} c_k I(t \in [t_{k-1}, t_k),$$

which gives

$$\Lambda(t) = \sum_{k=1}^{J_n} c_k (t - t_{k-1}) I(t_{k-1} \le t < t_k) + \frac{\tau}{J_n} \sum_{k=1}^{J_n} c_k I(t \ge t_k),$$

where the c_k's are some unknown constants. By plugging this into ℓ_C, we have

$$\begin{aligned}
\ell_n(\boldsymbol{\beta}, c_k' s) = {} & n^{-1} \sum_{i=1}^{n} 2\delta_i \boldsymbol{\beta}^T \mathbf{X}_i + n^{-1} \sum_{k=1}^{J_n} \log c_k \Big\{ \sum_{i=1}^{n} 2\delta_i I(e^{R_i(\boldsymbol{\beta})} \in [t_{k-1}, t_k) \Big\} \\
& - n^{-1} \sum_{k=1}^{J_n} c_k \Big\{ \sum_{i=1}^{n} (1+\delta_i)(e^{R_i(\boldsymbol{\beta})} - t_{k-1}) I(t_{k-1} \le e^{R_i(\boldsymbol{\beta})} < t_k) \\
& + \frac{\tau}{J_n} \sum_{i=1}^{n} (1+\delta_i) I(e^{R_i(\boldsymbol{\beta})} \ge t_k) \Big\} \\
& + n^{-1} \sum_{k=1}^{J_n} c_k \Big\{ \sum_{i=1}^{n} (e^{H_i(\boldsymbol{\beta})} - t_{k-1}) I(t_{k-1} \le e^{H_i(\boldsymbol{\beta})} < t_k) \\
& + \frac{\tau}{J_n} \sum_{i=1}^{n} I(e^{H_i(\boldsymbol{\beta})} \ge t_k) \Big\} \\
& + n^{-1} \sum_{k=1}^{J_n} c_k \Big\{ \sum_{i=1}^{n} \delta_i (e^{I_i(\boldsymbol{\beta})} - t_{k-1}) I(t_{k-1} \le e^{I_i(\boldsymbol{\beta})} < t_k) \\
& + \frac{\tau}{J_n} \sum_{i=1}^{n} \delta_i I(e^{I_i(\boldsymbol{\beta})} \ge t_k) \Big\}.
\end{aligned}$$

For given $\boldsymbol{\beta}$, one can easily obtain the maximum likelihood estimators of the c_k's as $\hat{c}_k = M_{0k}/(M_{1k} - M_{2k} - M_{3k})$, where

$$M_{0k} = \sum_{i=1}^{n} 2\delta_i I(e^{R_i(\boldsymbol{\beta})} \in [t_{k-1}, t_k)),$$

$$M_{1k} = \sum_{i=1}^{n}(1+\delta_i)(e^{R_i(\boldsymbol{\beta})} - t_{k-1})I\Big(e^{R_i(\boldsymbol{\beta})} \in [t_{k-1}, t_k)\Big) + \sum_{i=1}^{n}(1+\delta_i)I(e^{R_i(\boldsymbol{\beta})} \geq t_k)\frac{\tau}{J_n},$$

$$M_{2k} = \sum_{i=1}^{n}(e^{H_i(\boldsymbol{\beta})} - t_{k-1})I\Big(e^{H_i(\boldsymbol{\beta})} \in [t_{k-1}, t_k)\Big) + \sum_{i=1}^{n} I(e^{H_i(\boldsymbol{\beta})} \geq t_k)\frac{\tau}{J_n},$$

$$M_{3k} = \sum_{i=1}^{n}\delta_i(e^{I_i(\boldsymbol{\beta})} - t_{k-1})I\Big(e^{I_i(\boldsymbol{\beta})} \in [t_{k-1}, t_k)\Big) + \sum_{i=1}^{n}\delta_i I(e^{I_i(\boldsymbol{\beta})} \geq t_k)\frac{\tau}{J_n}.$$

It follows that the profile, approximate log composite conditional likelihood function has the form

$$\ell_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} 2\delta_i\boldsymbol{\beta}^T\mathbf{X}_i + \sum_{k=1}^{J_n}\frac{M_{0k}}{n}\times\log\frac{M_{0k}}{n} - \sum_{k=1}^{J_n}\frac{M_{0k}}{n}\times\log\frac{M_{1k}-M_{2k}-M_{3k}}{n}$$

by plugging the $\hat{c}_k$'s into $\ell_n(\boldsymbol{\beta}, c_k's)$.

To further smooth this function, by following Zeng and Lin (2007), we note that as $n \to \infty$, $J_n \to \infty$ and $n^{-1}J_n \to 0$, it can be shown that $\ell_n(\boldsymbol{\beta})$ converges uniformly in a compact set of $\boldsymbol{\beta}$ to

$$\ell(\boldsymbol{\beta}) = E\Big[\delta\boldsymbol{\beta}^T\mathbf{X} + \delta\log\Big(\frac{dP(\delta=1, Ye^{\boldsymbol{\beta}^T\mathbf{X}} \leq t)/dt}{N_1 - N_2 + N_3 - N_4}\Big)\Big|_{t=Ye^{\boldsymbol{\beta}^T\mathbf{X}}}\Big],$$

where

$$N_1 = P\big(Ye^{\boldsymbol{\beta}^T\mathbf{X}} \geq t\big)\ , \qquad N_2 = P\big(Ae^{\boldsymbol{\beta}^T\mathbf{X}} \geq t\big),$$
$$N_3 = P\big(\delta=1, Ye^{\boldsymbol{\beta}^T\mathbf{X}} \geq t\big)\ ,\ \ N_4 = P\big(\delta=1, V^0e^{\boldsymbol{\beta}^T\mathbf{X}} \geq t\big).$$

The proof of this is sketched in the Supplementary Material. In addition, for a given kernel function K with bandwidth b_n, and under some regularity conditions, one can prove that

$$\frac{1}{nb_n}\sum_{i=1}^{n}\delta_i K\Big(\frac{R_i(\boldsymbol{\beta})-\log t}{b_n}\Big) \to \frac{dP(\delta=1, R(\boldsymbol{\beta}) \leq s)}{ds}\Big|_{s=\log t} = \frac{dP(\delta=1, e^{R(\boldsymbol{\beta})} \leq t)}{dt}t,$$

$$\frac{1}{nb_n}\sum_{i=1}^{n}\int_{\log t}^{\infty} K\Big(\frac{R_i(\boldsymbol{\beta})-s}{b_n}\Big)ds \to P(R(\boldsymbol{\beta}) \geq \log t) = N_1,$$

$$\frac{1}{nb_n}\sum_{i=1}^{n}\int_{\log t}^{\infty} K\Big(\frac{H_i(\boldsymbol{\beta})-s}{b_n}\Big)ds \to P(H(\boldsymbol{\beta}) \geq \log t) = N_2,$$

$$\frac{1}{nb_n}\sum_{i=1}^{n}\int_{\log t}^{\infty}\delta_i K\Big(\frac{R_i(\boldsymbol{\beta})-s}{b_n}\Big)ds \to P(\delta=1, R(\boldsymbol{\beta})\geq \log t)=N_3,$$
$$\frac{1}{nb_n}\sum_{i=1}^{n}\int_{\log t}^{\infty}\delta_i K\Big(\frac{I_i(\boldsymbol{\beta})-s}{b_n}\Big)ds \to P(\delta=1, I(\boldsymbol{\beta})\geq \log t)=N_4.$$

It is then natural to approximate

$$\frac{dP(\delta=1, Ye^{\boldsymbol{\beta}^T\mathbf{X}}\leq t)/dt}{N_1-N_2+N_3-N_4}$$

by

$$\frac{1}{t}\frac{(nb_n)^{-1}\sum_{i=1}^{n}\delta_i K((R_i(\boldsymbol{\beta})-\log t)/b_n)}{\int_{\log t}^{\infty}(nb_n)^{-1}\sum_{i=1}^{n}\{K(\frac{R_i(\boldsymbol{\beta})-s}{b_n})-K(\frac{H_i(\boldsymbol{\beta})-s}{b_n})+\delta_i K(\frac{R_i(\boldsymbol{\beta})-s}{b_n})-\delta_i K(\frac{I_i(\boldsymbol{\beta})-s}{b_n})\}ds}$$

and $\ell(\boldsymbol{\beta})$ by

$$\begin{aligned}\ell_n^s(\boldsymbol{\beta}) = {} & \frac{1}{n}\sum_{i=1}^{n}\delta_i\boldsymbol{\beta}^T\mathbf{X}_i-\frac{1}{n}\sum_{i=1}^{n}\delta_i R_i(\boldsymbol{\beta})+\frac{1}{n}\sum_{i=1}^{n}\delta_i\log\Big\{\frac{1}{nb_n}\sum_{j=1}^{n}\delta_j K\Big(\frac{R_j(\boldsymbol{\beta})-R_i(\boldsymbol{\beta})}{b_n}\Big)\Big\}\\ & -\frac{1}{n}\sum_{i=1}^{n}\delta_i\log\Big\{\frac{1}{n}\sum_{j=1}^{n}\int_{(H_j(\boldsymbol{\beta})-R_i(\boldsymbol{\beta}))/b_n}^{(R_j(\boldsymbol{\beta})-R_i(\boldsymbol{\beta}))/b_n}K(s)ds\\ & +\frac{1}{n}\sum_{j=1}^{n}\delta_j\int_{(I_j(\boldsymbol{\beta})-R_i(\boldsymbol{\beta}))/b_n}^{(R_j(\boldsymbol{\beta})-R_i(\boldsymbol{\beta}))/b_n}K(s)ds\Big\},\end{aligned}$$

its kernel-smoothed empirical version.

For estimation of $\boldsymbol{\beta}$, we take $\widehat{\boldsymbol{\beta}}_n$ as the value of $\boldsymbol{\beta}$ that maximizes $\ell_n^s(\boldsymbol{\beta})$. Given $\widehat{\boldsymbol{\beta}}_n$, it is natural to estimate $\lambda(t)$ and $\Lambda(t)$ by

$$\widehat{\lambda}_n(t)=\frac{2(nb_nt)^{-1}\sum_{i=1}^{n}\delta_i K((R_i(\widehat{\boldsymbol{\beta}}_n)-\log t)/b_n)}{(1/n)\sum_{i=1}^{n}\int_{(H_i(\widehat{\boldsymbol{\beta}}_n)-\log t)/b_n}^{(R_i(\widehat{\boldsymbol{\beta}}_n)-\log t)/b_n}K(s)ds+\delta_i\int_{(I_i(\widehat{\boldsymbol{\beta}}_n)-\log t)/b_n}^{(R_i(\widehat{\boldsymbol{\beta}}_n)-\log t)/b_n}K(s)ds},$$
$$\widehat{\Lambda}_n(t)=\int_{-\infty}^{\log t}\frac{2(nb_nt)^{-1}\sum_{i=1}^{n}\delta_i K((R_i(\widehat{\boldsymbol{\beta}}_n)-u)/b_n)}{(1/n)\sum_{i=1}^{n}\int_{(H_i(\widehat{\boldsymbol{\beta}}_n)-u)/b_n}^{(R_i(\widehat{\boldsymbol{\beta}}_n)-u)/b_n}K(s)ds+\delta_i\int_{(I_i(\widehat{\boldsymbol{\beta}}_n)-u)/b_n}^{(R_i(\widehat{\boldsymbol{\beta}}_n)-u)/b_n}K(s)ds}du,$$

respectively. In the next subsection, we establish the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\Lambda}_n(t)$.

It is apparent that to implement this estimation procedure, one needs to choose a bandwidth b_n, for simplicity, one use a single value for b_n in all quantities. Among the quantities involving b_n, there are two kinds of density functions. One is for all subjects, such as N_1 and N_2, and the other is only for uncensored subjects such as $P(\delta=1, Ye^{\boldsymbol{\beta}^T\mathbf{X}}\leq t)$, N_3, and N_4 with $\delta=1$. In these situations, as suggested by Zeng and Lin (2007) one may want to employ two different bandwidths. In the numerical studies below, we use the optimal bandwidth $b_n = Cn^{-1/3}$ suggested in Zeng and Lin (2007), with different C for the two types of the quantities. Alternative is to employ a data-driven method such as the cross-validation procedure.

3.2. Asymptotic properties

Let $\boldsymbol{\beta}_0$ and Λ_0 denote the true values of $\boldsymbol{\beta}$ and Λ, respectively. To establish the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\Lambda}_n(t)$, we need some regularity conditions.

(C1). The parameter space $\mathcal{B}$ for $\boldsymbol{\beta}$ is compact.

(C2). If there exist a constant vector $\boldsymbol{\eta}$ and a deterministic function $g(.)$ such that $\boldsymbol{\eta}^T\mathbf{X} = g(\epsilon)$ with probability 1, then $\boldsymbol{\eta} = 0$ and $g = 0$.

(C3). The hazard function $\lambda_0(t)$ is three times continuously differentiable with $\lambda_0^{'}(0) > 0$ for all $t \geq 0$.

(C4). The matrix $\nabla^2_{\boldsymbol{\beta}} E\{\delta \log((dP(\delta=1, Ye^{\boldsymbol{\beta}^T\mathbf{X}} \leq t)/dt)/(N_1 - N2 + N_3 - N_4))\}$ is nonsingular in a neighborhood of $\boldsymbol{\beta}_0$.

(C5). The kernel function $K(.)$ is three times continuously differentiable with the derivatives having bounded variations. For the bandwidth b_n, $nb_n^2 \to \infty$, $nb_n^4 \to 0$, and $\log n/nb_n \to 0$ as $n \to \infty$.

(C6). The censoring time C has a positive and twice-continuously differentiable density function in $[0, \tau)$ and there exists a positive constant c_0 such that $P(C \geq \tau | \mathbf{X}, s \leq \tau) > c_0$ with probability 1.

The latter part of condition (C6) ensures that some subjects are censored at the end of study to avoid some technical complications regarding the tail behavior of the related limiting distribution. Many kernel functions, such as the Gaussian kernel and smooth kernels with a bounded support, satisfy.

Theorem 1. *If conditions* (C1)−(C6) *hold, as* $n \to \infty$, $\widehat{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}_0$ *and* $\sup_{t\in[0,\tau]} |\widehat{\Lambda}_n(t) - \Lambda_0(t)| \xrightarrow{P} 0$.

Theorem 2. *If conditions* (C1)−(C6) *hold, as* $n \to \infty$, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ *converges in distribution to normal with mean 0 and the variance-covariance matrix that can be consistently estimated by* $\hat{\mathbf{A}}^{-1}\hat{\mathbf{V}}\hat{\mathbf{A}}^{-1}$, *where*

$$\widehat{\mathbf{A}} = \frac{\partial^2 \ell_n^s(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^2} \quad , \quad \widehat{V} = \frac{1}{n}\sum_{i=1}^{n} \Big(\frac{\partial l_i^s(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}\Big)\Big(\frac{\partial l_i^s(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}\Big)^T,$$

with

$$l_i^s(\boldsymbol{\beta}) = \delta_i \boldsymbol{\beta}^T \mathbf{X}_i - \delta_i R_i(\boldsymbol{\beta}) + \delta_i \log \Big\{ \frac{1}{nb_n} \sum_{j=1}^n \delta_j K\Big(\frac{R_j(\boldsymbol{\beta}) - R_i(\boldsymbol{\beta})}{b_n}\Big) \Big\}$$
$$-\delta_i \log \Big\{ \frac{1}{n} \sum_{j=1}^n \int_{(H_j(\boldsymbol{\beta})-R_i(\boldsymbol{\beta}))/b_n}^{(R_j(\boldsymbol{\beta})-R_i(\boldsymbol{\beta}))/b_n} K(s)ds$$
$$+\frac{1}{n} \sum_{j=1}^n \delta_j \int_{(I_j(\boldsymbol{\beta})-R_i(\boldsymbol{\beta}))/b_n}^{(R_j(\boldsymbol{\beta})-R_i(\boldsymbol{\beta}))/b_n} K(s)ds \Big\}.$$

The proofs of these theorems are sketched in the Supplementary Material.

4. A Simulation Study

In this section, we present some results obtained from a simulation study conducted to assess the finite sample performance of the proposed estimation procedure. In the study, we assumed that $\tilde{T}$ follows the log-linear model $\log \tilde{T} = 2 + X_1 + X_2 + \epsilon$ with X_1 generated from the uniform distribution over $(0, 1)$, and X_2 from the Bernoulli distribution with the success probability of 0.5. For the error term distribution, we considered several choices: S_1: $N(0, 0.5)$; S_2: $extreme(0, 1)$ (the standard extreme distribution); S_3: $0.5N(0, 0.25)+0.5N(0, 1)$; and S_4: $exp(1)$ (the exponential distribution). Furthermore, the examination time $\tilde{A}$ was assumed to follow the uniform distribution over $(0, \omega)$ with ω being larger than the upper bound of $\tilde{T}$ to ensure the stationary assumption. The censoring time C was also generated from the uniform over $(0, a)$ with a chosen to give proper censoring percentages. The results given below are based on $n = 200$, the Gaussian kernel function, and 1,000 replications.

Table 1 presents the results on estimation of $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ given by the proposed estimation procedure with the percentage of right-censored observations (PCR) being 15%, 30%, or 50%, respectively. They include the estimated bias, the sample standard deviation of the estimators (SE), the average of the estimated standard errors (SEE), and the 95% empirical coverage probabilities (CP). For the bandwidths, we used $2.6\hat{\sigma}_1 n^{-1/3}$ and $1.6\hat{\sigma}_2 n^{-1/3}$ for the quantities involving uncensored subjects and all subjects, respectively, where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are the sample standard deviations of ϵ among the uncensored subjects and all subjects, respectively. For comparison, we considered the estimation procedure given in Shen, Ning and Qin (2009) referred to as SNQ, and obtained and included in the table the estimated bias and the sample standard deviation.

The results in Table 1 suggest that the proposed estimator is unbiased and the variance estimation is reasonable. As expected, the parameters can be estimated more accurately when the percentage of right-censored observations decreases. There seems to exist some biases for the estimators in

Table 1. Simulation results on estimation of $\boldsymbol{\beta}$.

			Proposed				SNQ	
Model	PCR		Bias	SE	SEE	CP	Bias	SE
S_1	15%	β_1	-0.027	0.148	0.171	0.962	-0.027	0.114
		β_2	0.024	0.089	0.095	0.928	0.020	0.079
	30%	β_1	-0.036	0.155	0.182	0.960	-0.015	0.119
		β_2	0.020	0.097	0.102	0.912	0.041	0.083
	50%	β_1	-0.038	0.184	0.211	0.958	0.018	0.127
		β_2	0.025	0.122	0.118	0.932	0.082	0.092
S_2	15%	β_1	-0.053	0.303	0.467	0.966	-0.157	0.597
		β_2	-0.028	0.311	0.279	0.972	0.012	0.521
	30%	β_1	-0.073	0.343	0.364	0.958	-0.110	0.607
		β_2	-0.080	0.405	0.390	0.952	0.047	0.521
	50%	β_1	-0.062	0.391	0.404	0.940	-0.007	0.635
		β_2	-0.067	0.479	0.416	0.948	-0.122	0.554
S_3	15%	β_1	-0.015	0.262	0.283	0.932	-0.003	0.267
		β_2	-0.003	0.170	0.183	0.930	0.018	0.194
	30%	β_1	-0.030	0.276	0.307	0.928	0.024	0.278
		β_2	0.010	0.190	0.189	0.938	0.056	0.206
	50%	β_1	-0.053	0.295	0.327	0.930	0.090	0.303
		β_2	0.010	0.212	0.199	0.924	0.110	0.220
S_4	15%	β_1	-0.051	0.301	0.444	0.964	-0.164	0.598
		β_2	-0.015	0.317	0.272	0.974	0.014	0.519
	30%	β_1	-0.068	0.345	0.334	0.944	-0.121	0.600
		β_2	-0.080	0.504	0.492	0.958	0.049	0.518
	50%	β_1	-0.062	0.391	0.404	0.940	-0.007	0.635
		β_2	-0.067	0.479	0.435	0.948	0.122	0.554

Table 2. Estimated mean square errors of the proposed estimators.

	bw		bw_1		bw_2		bw_3		bw_4	
PCR	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
15%	0.024	0.008	0.068	0.046	0.028	0.012	0.031	0.011	0.059	0.029
30%	0.025	0.010	0.089	0.057	0.043	0.014	0.027	0.011	0.034	0.014
50%	0.035	0.016	0.093	0.062	0.052	0.018	0.036	0.014	0.045	0.018

Shen, Ning and Qin (2009) when the data were generated from models S_2 and S_4. Under the scenarios S_1 and S_3, SNQ seems to have comparable or even better efficiency than the proposed method; it was derived based on the least square principle and tends to be more efficient when the distribution of $\log \tilde{T}$ is close to normal. For the non-normal S_2 and S_4, the proposed method is more efficient than SNQ. For simplicity, we used a fixed bandwidth, performance can be improved if a dynamic optimal bandwidth selection method such as cross-validation is used.

For the proposed estimation procedure, a question of interest concerns the effect of the bandwidth on the resulting estimators. To investigate this, we repeated the simulation study using different bandwidths and compared the mean square errors (MSE) of the resulting estimators. Table 2 gives the MSE when the data were generated from model S_1. Here, in addition to the bandwidth $bw = \{2.6\hat{\sigma}_1 n^{-1/3}, 1.6\hat{\sigma}_1 n^{-1/3}\}$ used above, we also considered $bw_1 = 0.3bw$, $bw_2 = 0.5bw$, $bw_3 = 1.5bw$, and $bw_4 = 2bw$. The results suggest that the proposed estimator is robust with respect to the bandwidth selection. We also investigated some other kernel functions and obtained similar results.

5. An Illustration

To illustrate our estimation procedure proposed above, we apply it to the set of length-biased and right-censored data discussed in Shen, Ning and Qin (2009), among others. It arose from the Canadian Study of Health and Aging on the patients diagnosed with dementia. After excluding the patients with the missing date of disease onset or classification of dementia subtype, it consists of 818 patients with probable Alzheimer's disease, possible Alzheimer, or vascular dementia. Among them, 638 subjects died during follow-up and the others gave right-censored death times. The main objective was to evaluate and compare the impact of different subtypes of dementia on the overall survival time. For the data here, the stationarity assumption has been validated by Addona and Wolfson (2006).

For the analysis, take $X_1 = 1$ if the patient had Vascular dementia and 0 otherwise and $X_2 = 1$ if the patient had possible Alzheimer's disease and 0 otherwise. Using the Gaussian kernel function we obtained $\hat{\beta}_{n1} = -0.065$ and $\hat{\beta}_{n2} = 0.184$, with estimated standard errors 0.073 and 0.095, respectively. For the bandwidths, we used $3.2\hat{\sigma}_1 n^{-1/3}$ and $3.9\hat{\sigma}_2 n^{-1/3}$ for the kernel densities of uncensored subjects and the cumulative kernel densities of all subjects, respectively. The results suggest that the patients with possible Alzheimer's disease have significantly longer lifetime than those with either probable Alzheimer's disease or Vascular dementia. However, there is no significant difference between the lifetimes of the patients in the two latter groups. For comparison, we also performed the analysis by using the method given in Shen, Ning and Qin (2009) and obtained $\hat{\beta}_{n1} = -0.075$ and $\hat{\beta}_{n2} = 0.129$, with estimated standard errors of 0.143 and 0.152, respectively. They indicate no significant differences among the lifetimes of all three types of patients.

6. Discussion and Concluding Remarks

Our focus has been on the failure time data arising from the AFT model. Sometimes the data arise from such other regression models as the additive hazards model or the linear transformation model, and it would be useful to develop estimation procedures for these models too. Although the idea discussed above can still apply to these situations, the development of specific estimation procedures is not straightforward due to the different structures of the models. In approximating the hazard or cumulative hazard function in Section 3, we used the piecewise constant functions. There are other approximations, such as B-spline functions, and it is straightforward to generalize the proposed estimation procedure to these situations.

Supplementary Materials

The proofs of our main results are given in the supplementary material available online.

Acknowledgement

We wish to thank the Editor, and two reviewers for their many helpful and insightful comments and suggestions that greatly improved the paper. Chen's work was supported by National Natural Science Foundation of China (NSFC) (11501461), the Fundamental Research Funds for the Central Universities (JBK140507, JBK120509).

References

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

Arnold, B. C. and Strauss, D. (1988). Bivariate distributions with exponential conditionals. *J. Amer. Statist. Assoc.* **83**, 522-527.

Chen, X. and Zhou, Y. (2012). Quantile regression for right-censored and length-biased data. *Acta Math. Appl. Sin-E* **28**, 443-462.

Cox, D. R. and Oakes, D. (1984).*Analysis of Survival Data.* Chapman and Hall, New York.

Huang, C. Y. and Qin, J. (2012). Composite partial likelihood estimation under lengthbiased sampling, with application to a prevalent cohort study of dementia. *J. Amer. Statist. Assoc.* **107**, 946-957.

Jin, Z., Lin, D. Y. and Ying, Z. (2003). Rank-based inference for the accelerated failure time models. *Biometrika* **90**, 341-353.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data.* John Wiely and Sons, New Jersey.

Keiding, N. (1991). Age- specified incidence and prevalence: a statistical perspective. *J. Roy. Statist. Soc. Ser. A* **154**, 371-412.

Lai, T. L. and Ying, Z. (1991). Large sample theory of a modified buckley-james estimator for regression analysis with censored data. *Ann. Statist.* **19**, 531-556.

Ning, J., Qin, J. and Shen, Y. (2011). Buckley-James-Type estimator with right-censored and length-biased data. *Biometrics* **67**, 1369-1378.

Ning, J., Qin, J. and Shen, Y. (2014a). Score estimating equation for embedded likelihood functions under accelerated failure time model. *J. Amer. Statist. Assoc.* **109**, 1625-1635.

Ning, J., Qin, J. and Shen, Y. (2014b). Semiparametric accelerated failure time model for length-biased data with application to dementia study. *Statist. Sinica* **24**, 313-333.

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

Sansgiry, P. and Akman, O. (2000). Transformations of the lognormal distribution as a selection model.*Am. Stat.* **54**, 307-309.

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.

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

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

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

Zeng, D. and Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *J. Amer. Statist. Assoc.* **102**, 1387-1396.

Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, 611130, China.

E-mail: chenxr522@foxmail.com

Boehringer Ingelheim (China) Investment Co., Ltd., Shanghai, 200040, China.

E-mail: nh2hd@mail.missouri.edu

Department of Statistics, University of Missouri, Columbia, Missouri, 65201, U.S.A.

Institute of Mathematics, Jilin University, Changchun, 130012, China.

E-mail: sunj@missouri.edu

(Received March 2015; accepted December 2015)