Statistics and Its Interface Volume 12 (2019) 135–148

Pseudo likelihood estimation for the additive hazards model with data subject to left-truncation and right-censoring

Li-Pang Chen

Analysis of left-truncated and right-censored (LTRC) survival data has received extensive interest. Many inference methods have been developed for the various survival models, including the Cox proportional hazards model and the transformation model. The additive hazards model is also concerned in survival analysis, and several methods have also been developed without left-truncation. However, little work has been available in the literature for the additive hazards model with left-truncation and right-censoring. In this paper, we explore this important problem under the additive hazards model. We develop the pseudo-likelihood inference for the estimation of the survival model parameters, which yields a more efficient estimator. Besides, we assess the performance of our proposed methods using simulation studies. Through the conducted simulations, the proposed estimator is further found to outperform the existing competitors in the literature.

Keywords and phrases: Kernel estimator, Left-truncation, Misspecification, Prevalent sampling, Pseudo likelihood.

1. INTRODUCTION

In the study of a disease history, the time from the onset of an initiating event to the focused disease event (or failure) is usually the interest in the epidemiological and biomedical researches. One of the most attractive data comes from the prevalent sampling design, in which individuals only experience the initiating event but not the failure event before the recruiting time. Under this sampling scheme, individuals might not be observed because they experience the failure event before the recruiting time. Such a phenomenon caused by the delayed entry is called *left-truncation* and tends to produce a biased sample. Meanwhile, individuals who are recruited in the study may drop out or may not experience the failure event at the end of the study. It is called *right-censoring* in the dataset. In this study, we focus on investigating the statistical inference procedures with left-truncated and right-censored (LTRC) data. Here we give two specific examples.

Example 1: The Channing house data.
The Channing house is a retirement center in Palo Alto, California. These data were collected between the opening of the house in 1964 until July 1, 1975. Suppose that individuals who survive to an age of 60 or higher are allowed to enter the retirement center. Residents in the population were born before 1916. In addition, let u and v be the calendar time of birth and the calendar time of death, respectively. Let ξ be the calendar time of entry. Therefore, we can define $T^* = v - u$ as the length from birth to death and denote $A^* = \xi - u$ the length from birth to the time of entry. It is obvious that an individual becomes the element of a sample if and only if $T^* \geq A^*$. For those individuals in the house, their ages on entry and also on leaving or death were recorded. Besides, in this dataset, a large number of individuals were right-censored because they left the house prior to July 1, 1975, or that they were still alive and living in the center on that date.

Example 2: The Worcester Heart Attack Study (WHAS500) data.
The main goal of this study is to determine the factors associated with trends over time in the incidence and survival rates following hospital admission for acute myocardial infarction (MI). The data were collected over thirteen 1-year periods beginning in 1975 and extending through 2001 on all MI patients admitted to the hospitals in Worcester, Massachusetts. There are 500 observations and 22 variables in this dataset. Specifically, as discussed in Hosmer, Lemeshow, and May [9], the beginning of survival time was defined as the time the subject was admitted to the hospital. The main interest is the survival time of a patient who was discharged and still alive. Hence, an inclusion criterion is that only those subjects who are discharged and still alive are eligible to be included in the analysis. That is, the minimum survival time would be the length of the time a patient stayed in the hospital; individuals whose observation times are shorter than the minimum survival time are not included in this analysis.

Basically, the data are pertinent to three important events in calendar time: time of hospital admission, time of hospital discharge, and time of last follow-up (which is either failure or censoring). The total length of follow-up is defined as the length of time between hospital admission and the last follow-up, and the length of hospital stay is defined as the time length between hospital admission and hospital discharge. Data can only be collected for those individuals whose total length of follow-up is longer than the length of hospital stay, which is the so-called left-truncation (e.g., [17, Section 1.3]; [19, Section 2.4]).

In the case of stable disease, however, the occurrence of disease onset follows the stationary Poisson process. It implies that the truncation time follows the uniform distribution, and the survival time in the prevalent cohort has a length-biased sampling distribution since the probability of the survival time is proportional to the length of survival time (e.g., [32, 10, 11]). Hence, the length-biased sampling can be regarded as a special case of LTRC data (e.g., [2, 25]).

In this article, we mainly focus on the model development with covariates under LTRC data. Among all models in survival analysis, the Cox proportional hazards (PH) model has highly attracted the most research attention. Briefly speaking, given covariate Z, the Cox PH model is given by

$$\lambda(t|z) = \lambda_0(t) \exp\left(\beta^\top z\right),$$

where $\lambda_0(\cdot)$ is the unspecified baseline hazard function and β is the unknown parameter.

For the Cox PH model, several estimation methods have been developed based on LTRC data. For instance, suppose that for sample $i = 1, \cdots, n$, T_i, A_i and C_i are the failure time, the truncation time, and censoring time, respectively. Denote $Y_i = \min\{T_i, C_i\}$ the lifetime and define $\delta_i = I(T_i \leq C_i)$ as an indicator of the failure time. Wang, Brookmeyer, and Jewell [34] expressed the conditional likelihood as the following form:

$$L_C(\beta, \lambda_0) \propto \prod_{i=1}^{n} \frac{\left\{\lambda_0(y_i)\exp\left(\beta^\top z_i\right)\right\}^{\delta_i} S\left(y_i|z_i\right)}{S\left(a_i|z_i\right)}, \tag{1}$$

where $S(t|z) = \exp\left\{-\Lambda_0(t)\exp(\beta^\top z)\right\}$. Moreover, $L_C(\beta, \lambda_0)$ can be decomposed as the product of the partial likelihood (Kalbfleisch and Lawless [16])

$$L_P(\beta, \lambda_0) \propto \prod_{i=1}^{n} \left\{ \frac{\exp(\beta^\top z_i)}{\sum\limits_{i=1}^{n} \exp(\beta^\top z_j) I(a_j \leq y_i \leq y_j)} \right\}^{\delta_i} \tag{2}$$

and the residual likelihood $L_R(\beta, \lambda_0)$. Wang, Brookmeyer, and Jewell [34] showed that $L_R(\beta, \lambda_0)$ is ancillary and (2) is fully efficient with respect to (1). In both LTRC data and the length-biased sampling, however, directly maximizing (2) with respect to β is expected to yield the inefficient estimator. Hence, in order to improve the efficiency, several non-parametric or semi-parametric approaches have been developed. Specifically, the past literature mainly focused on the length-biased sampling and the Cox PH model. For example, Tsai [29] proposed the pseudo-partial likelihood method for the Cox PH model based on the length-biased sampling. Qin and Shen [25] proposed two different methods of the estimating equations to estimate β. Huang, Follman, and Qin [11] proposed the semiparametric likelihood inference for the Cox PH model based on the length-biased sampling. Su and Wang [28] developed the semi-parametric approach for the joint modelling between the LTRC survival outcomes and the longitudinal covariates. In addition to the Cox PH model, different types of model were also concerned. For example, Cheng and Wang [5] discussed the transformation model for causal inference. Chen and Shen [3] developed the conditional maximum likelihood approach in the transformation model.

Not only the models mentioned above, different type of models are also discussed in the developments of survival analysis based on specific purposes. For example, different from the investigation of the hazard ratio based on the Cox PH model, sometimes researchers may be more interested in the risk difference attributed to the risk factors. Based on this purpose, the additive hazards model is considered, and the formulation is given by

$$\lambda(t|z) = \lambda_0(t) + \beta^\top z, \tag{3}$$

where $\lambda(t|z)$ is the conditional hazard function of the survival time given the covariates Z, $\lambda_0(t)$ is the unspecified baseline hazard function and β is the $p \times 1$ vector of parameters.

In the absence of the left truncation, many methods were developed for the additive hazards model in the past literature. For example, Lin and Ying [22, 23] developed the estimating equations approach to derive the estimator of β and the cumulative baseline hazard function $\Lambda_0(\cdot)$. When the left-truncation occurs, however, little work has been available in the developments of the additive hazards model. Similar to the Cox model, the estimator would be inefficient if we only considered the conditional part. To overcome this problem and improve efficiency, Huang and Qin [12] proposed the modified conditional estimating equations under the martingale theory (e.g., [22]) and applied the pairwise pseudo-marginal likelihood (e.g., [20]) to deal with the remaining part. Finally combining the two methods yields the estimators, and the efficiency of the estimators is largely improved. However, the model misspecification was not concerned in the development of Huang and Qin [12]. That is, the properties of the martingale theory only hold when the model is true (e.g., [21]). Besides, the estimating equation approach was frequently studied in the development of the additive hazards model. To the best of our knowledge, however, there is no existing method to estimate β by the maximization of the likelihood function based on the additive hazards model. Hence, in this article, the main goal is to develop the pseudo-likelihood method to improve efficiency.

The rest of this article is organized as follows. We first introduce the structure of LTRC data and the corresponding likelihood functions in Section 2. We next review the existing estimation methods which were proposed by Huang and Qin [12] in Section 3. After that, we present our methods in Section 4. Basically, we first estimate the distribution function of the truncation time, and then use the smoothing technique to estimate the baseline hazard function. In the last step, we propose the pseudo-likelihood estimation to derive the estimator of the main parameter. We give some model settings to examine the numerical performances of the estimator and compare with methods proposed by Huang and Qin [12] in Section 5. Finally, the real data analysis is given in Section 6.

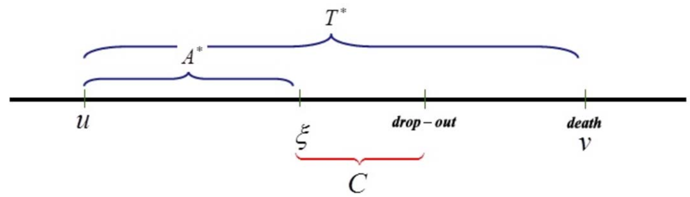

Figure 1. Schematic depiction of LTRC data for $T^* \geq A^*$.

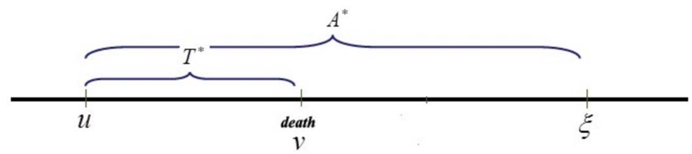

Figure 2. Schematic depiction of LTRC data. Truncation occurs when $T^* < A^*$.

2. LTRC DATA AND LIKELIHOOD CONSTRUCTION

2.1 Data introduction

For an individual in the target disease population, let ξ be the calendar time of the recruitment (e.g., the recruitment starts right at the hospital discharge) and let u and v denote the calendar time of the initiating event (e.g., hospital admission) and the failure event (e.g., death), respectively, where $u < v$ and $u < \xi < v$. Let $T^* = v - u$ be the lifetime (e.g., the time length between the hospital admission and the failure) and $A^* = \xi - u$ be the truncation time (e.g., the time length between the hospital admission and the hospital discharge). Let Z^* be the associated covariates of dimension $p \times 1$. Let $h(a)$ be the unspecified probability density function of A^*, and let $H(a) = \int_0^a h(\zeta)d\zeta$ denote the distribution function of A^*. Let $f(t)$ and $S(t)$ be the density function and the survivor function of failure time T^*, respectively. Define

$$(A, T, Z) = \begin{cases} (A^*, T^*, Z^*), & \text{if } T^* \geq A^*, \\ 0, & \text{otherwise.} \end{cases}$$

That is, A, T, and Z represent the truncation time, the survival time and the covariates for those subjects who are recruited in the study, respectively. Hence, (A, T, Z) has the same joint distribution as (A^*, T^*, Z^*) given $T^* \geq A^*$. In addition, we let C denote the censoring time for a recruited subject. Let $Y = \min\{T, A + C\}$ be the observed survival time and $\Delta = I(T \leq A + C)$ be the indicator of a failure event. Figure 1 gives an illustration of the relationship among those defined variables. However, if $T^* < A^*$, as shown in Figure 2, the individual is not included in the study so that the researcher cannot obtain any information of such individual.

For the following development, we make standard assumptions which are commonly considered for survival data analysis and related frameworks (e.g., [11, 12]):

(A1) Conditional on Z^*, T^* are independent of A^*;
(A2) Censoring time is non-informative.

2.2 Construction of the likelihood function

Suppose that we have a sample of n subjects and that for $i = 1, \cdots, n$, $(Y_i, A_i, \Delta_i, Z_i)$ has the same distribution as (Y, A, Δ, Z) and $(y_i, a_i, \delta_i, z_i)$ represents realizations of $(Y_i, A_i, \Delta_i, Z_i)$.

Under assumption (A1), the joint density function of (A, T) given Z is proportional to

$$\frac{f(t|z)h(a)}{\int_0^\infty S(u|z)h(u)du} = \frac{f(t|z)}{S(a|z)} \times \frac{S(a|z)h(a)}{\int_0^\infty S(u|z)h(u)du}, \tag{4}$$

where $\frac{f(t|z)}{S(a|z)}$ is the density function of T given A and Z, and $\frac{S(a|z)h(a)}{\int_0^\infty S(u|z)h(u)du}$ is the density function of A given Z. The detailed derivation of (4) is placed in Appendix A. In addition to the failure event, right-censoring may occur for those recruited subjects in the study. Under assumption (A2), we can derive the full likelihood function based on the independent and identically distributed data $(Y_i, A_i, \Delta_i, Z_i)$, which is proportional to

$$L_F \propto \prod_{i=1}^n \frac{f(y_i|z_i)^{\delta_i} S(y_i|z_i)^{1-\delta_i} h(a_i)}{\int S(u|z_i)h(u)du}. \tag{5}$$

Specifically, we focus on the additive hazards model in this article. Thus, under (5) and model (3), we can obtain

$$L_F(\beta, \lambda_0, h) = \prod_{i=1}^n \frac{\left\{\lambda_0(y_i) + \beta^\top z_i\right\}^{\delta_i} S(y_i|z_i)h(a_i)}{\int S(u|z_i)h(u)du}, \tag{6}$$

where $S(t|z_i) = \exp\left\{-\Lambda_0(t) - \beta^\top z_i t\right\}$ is the survival function under model (3). Moreover, we can decompose (6) into $L_C \times L_M$, where

$$L_C(\beta, \lambda_0) = \prod_{i=1}^n \frac{\left(\lambda_0(y_i) + \beta^\top z_i\right)^{\delta_i} S(y_i|z_i)}{S(a_i|z_i)} \tag{7}$$

is the likelihood of (Y, Δ) given A, Z; and

$$(8) \qquad L_M(\beta, \lambda_0, h) = \prod_{i=1}^{n} \frac{S(a_i|z_i)h(a_i)}{\int S(u|z_i)h(u)du}$$

is the likelihood of A given Z.

One can easily observe that it is difficult to directly maximize (6) since it involves unspecified functions $\lambda_0(\cdot)$ and $h(\cdot)$, and we cannot imitate Wang, Brookmeyer, and Jewell [34] to decompose (7) into the partial likelihood and the residual likelihood based on the additive hazards model. The existing methods, which were proposed by Huang and Qin [12], to estimate β and $\lambda_0(\cdot)$ under model (3) are the conditional estimating equations, the pairwise pseudo-marginal likelihood and the combination of those two methods. The spirit of those methods is to imitate the roles of likelihood functions (7) and (8). We will review those existing methods in the next section.

Remark 1. As described in Section 1, if the disease incidence occurs over calendar time at a constant rate, then it implies that the distribution of the truncation time follows the uniform distribution (e.g., [18]). Based on this situation, the likelihood function can be expressed as

$$(9) \qquad L = \prod_{i=1}^{n} \frac{f(y_i|z_i)^{\delta_i} S(y_i|z_i)^{1-\delta_i}}{\int S(u|z_i)du},$$

which can be viewed as a special case of the likelihood function (5) (e.g., [25]). For the statistical inference, it becomes simple since the density function of the truncation time $h(a)$ is no need to be estimated, and we only need to estimate $\Lambda_0(\cdot)$ and β.

3. REVIEW OF EXISTING METHODS

3.1 The conditional estimating equation (CEE)

As the absence of the left truncation, Lin and Ying [22] proposed the estimating equations method to obtain the estimator of β and the cumulative baseline hazard function $\Lambda_0(t)$ based on the additive hazards model. If truncation time occurs, Huang and Qin [12] modified the estimating equations proposed by Lin and Ying [22] by the following procedures. First of all, define $N_i(t) = \Delta_i I(Y_i \leq t)$ as the counting process for the observed failure events, and the modified at-risk process is denoted by $R_i(t) = I(A_i \leq t \leq Y_i)$ for the adjustment of the truncation time.

Next, define $M_i(t) = N_i(t) - \int R_i(u)\{d\Lambda_0(u) + \beta^\top Z_i du\}$, which can be verified that it is a local square-integrable martingale. Hence, we can estimate β and $\Lambda_0(\cdot)$ by solving the two estimating equations $\sum_{i=1}^{n} \int Z_i dM_i(u) = 0$ and $\sum_{i=1}^{n} \int dM_i(u) = 0$. From $\sum_{i=1}^{n} \int dM_i(u) = 0$, we can obtain

$$(10) \qquad \widehat{\Lambda}_0(t;\beta) = \int_0^t \frac{\sum_{i=1}^{n} \{dN_i(u) - R_i(u)\beta^\top Z_i du\}}{\sum_{i=1}^{n} R_i(u)}.$$

Substituting (10) into $\sum_{i=1}^{n} \int Z_i dM_i(u) = 0$ yields the estimating equation

$$(11) \qquad \sum_{i=1}^{n} \phi_i(\beta) = 0,$$

where $\phi_i(\beta) = \int \{Z_i - \bar{Z}(t)\} dM_i(t)$ and $\bar{Z}(t) = \frac{\sum Z_i R_i(t)}{\sum R_i(t)}$. Finally, the estimator of β is obtained by solving (11), and the solution is denoted by

$$(12) \qquad \begin{aligned} \widehat{\beta}_\phi = & \left(\sum \int \{Z_i - \bar{Z}(t)\}^{\otimes 2} R_i(t) dt \right)^{-1} \\ & \times \left(\sum \int \{Z_i - \bar{Z}(t) dN_i(t)\} \right), \end{aligned}$$

where $a^{\otimes 2} = aa^\top$ for any vector a.

Moreover, it can be shown that $\sqrt{n}\left\{\widehat{\Lambda}_0(t;\beta) - \Lambda_0(t)\right\}$ converges weakly to a zero-mean Gaussian process by the counting process theory and $\sqrt{n}\left(\widehat{\beta}_\phi - \beta_0\right)$ converges weakly to a zero-mean multivariate normal distribution, where β_0 is the true parameter.

3.2 The pairwise pseudo likelihood method (PPL)

Although the estimator of β is obtained by the conditional estimating equation, it is expected that the estimator from the estimating equation approach is not efficient since the information of β in the marginal likelihood (8) is not used in the estimating procedure. Hence, in order to analyze the marginal likelihood and improve the efficiency of the estimator of β, the pairwise likelihood method was proposed.

The idea is as follows. Under the additive hazards model, the marginal density function of A given Z is

$$(13) \qquad \frac{S(a|z)h(a)}{\int S(u|z)h(u)du} = \frac{S_0(a)\exp(-\beta^\top za)h(a)}{\int S(u|z)h(u)du},$$

where $S_0(a) = \exp\{-\Lambda_0(a)\}$.

From (13), we observe that obtaining the estimator of β by maximizing the marginal likelihood is difficult since it involves unspecified functions $h(\cdot)$ and $S_0(\cdot)$. Hence, to estimate the parameter β in (13) and eliminate the unknown functions $S_0(\cdot)$ and $h(\cdot)$ simultaneously, the pairwise pseudo likelihood (e.g., [15, 20]) is modified to handle (13). For the observed unordered pair (A_i, A_j), the pairwise pseudo marginal likelihood of (A_i, A_j) conditional on (Z_i, Z_j) with $i < j$ is given by

$$\frac{\frac{S(a_i|z_i)h(a_i)}{\int S(u|z_i)h(u)du} \times \frac{S(a_j|z_j)h(a_j)}{\int S(u|z_j)h(u)du}}{\frac{S(a_i|z_i)h(a_i)}{\int S(u|z_i)h(u)du} \times \frac{S(a_j|z_j)h(a_j)}{\int S(u|z_j)h(u)du} + \frac{S(a_i|z_j)h(a_i)}{\int S(u|z_j)h(u)du} \times \frac{S(a_j|z_i)h(a_j)}{\int S(u|z_i)h(u)du}},$$

which can also be re-written as

$$\begin{aligned} & \frac{\exp(-\beta^\top a_i z_i - \beta^\top a_j z_j)}{\exp(-\beta^\top a_i z_i - \beta^\top a_j z_j) + \exp(-\beta^\top a_i z_j - \beta^\top a_j z_i)} \\ = \quad & \frac{1}{1 + \exp\{\beta^\top (a_i - a_j)(z_i - z_j)\}}. \end{aligned}$$

Define $\rho_{ij} = (a_i - a_j)(z_i - z_j)$, and we can estimate β by maximizing the log pairwise pseudo marginal likelihood

$$\sum_{1 \le i < j \le n} - \log \left\{1 + \exp(\beta^\top \rho_{ij})\right\}.$$

It is equivalent to solve the estimating equation

$$\psi(\beta) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \psi_{ij}(\beta) = 0, \tag{14}$$

where

$$\psi_{ij}(\beta) = \frac{-\rho_{ij}}{1 + \exp(-\beta^\top \rho_{ij})}.$$

Here we denote $\widehat{\beta}_\psi$ the solution of (14). The asymptotic result of $\widehat{\beta}_\psi$ can be found in Huang and Qin [12].

3.3 The method of Huang and Qin [12]

So far, the consistent estimator of β can be obtained from both (11) and (14) separately. Hence, to obtain the estimator of β from both parts and to improve the efficiency, Huang and Qin [12] proposed to incorporate those two estimating equations (11) and (14), and the estimator of β can be obtained by solving the estimation equation

$$\eta(\beta) = \phi(\beta) + \psi(\beta) = 0, \tag{15}$$

where $\phi(\beta) = n^{-1} \sum_{i=1}^{n} \phi_i(\beta)$. Let $\widehat{\beta}$ denote the estimator satisfying $\eta(\widehat{\beta}) = 0$. The asymptotic result of $\widehat{\beta}$ is given in Theorem 1 of Huang and Qin [12].

3.4 Additional remarks

We can observe that the goal of the two estimating equations (11) and (14) is to deal with two parts (7) and (8), respectively. However, both methods are not equivalent to the maximization of (7) and (8). Besides, the conditional estimating equation in Section 3.1 is constructed under the martingale theory. However, the properties of the martingale method are only valid when the model is correct. It implies that model misspecification may not be involved in their method. As pointed out by Lin and Wei [21], if the model is incorrect or misspecified, then the estimating equation method does not work. Therefore, instead of using the estimating equation approach to derive the estimator (e.g., [1]), Lin and Wei [21] proposed the maximization of the likelihood function to derive the estimator based on the Cox PH model. Furthermore, as the best of our knowledge, there is no development for the additive hazard model based on the likelihood method. Hence, in the next section, we propose a new method to derive the estimator and our goal is to obtain the more precise estimator and improve the efficiency of the estimator.

4. PSEUDO LIKELIHOOD ESTIMATION

We now propose the pseudo likelihood method to estimate β for the additive hazards model under LTRC data in this section. From (6), there exists a unknown parameter β and two unspecified functions $\lambda_0(\cdot)$ and $h(\cdot)$. First of all, we estimate the distribution function of the truncation time in (8). We then propose the smoothing technique to estimate $\lambda_0(\cdot)$ and replace $\Lambda_0(\cdot)$ by (10). Finally, replacing $\lambda_0(\cdot)$ and $h(\cdot)$ by their estimators in the likelihood function (6) yields the pseudo likelihood function. The last step is to obtain estimator of β by maximizing the pseudo likelihood function.

4.1 Estimation of $H(a)$

From (6), we have a unknown parameter β and two unknown functions $\lambda_0(\cdot)$ and $h(\cdot)$. As discussed in Section 2.2, we can decompose (6) into (7) and (8). Moreover, we observe that $h(\cdot)$ only appears in (8). So, in this section, we mainly discuss the estimation procedure for $h(\cdot)$ in (8).

Since (8) contains unknown β, $\Lambda_0(\cdot)$, and $h(\cdot)$, we first replace $\Lambda_0(\cdot)$ by the consistent estimator (10). This approach is valid even our method is not based on the martingale method. The detailed derivation is available in Appendix B. After that, we obtain the pseudo marginal likelihood

$$PL_M(\beta, H) = \prod_{i=1}^{n} \frac{\widehat{S}(a_i|z_i)h(a_i)}{\int_0^\infty \widehat{S}(u|z_i)dH(u)}, \tag{16}$$

where $H(a)$ is the corresponding cumulative distribution function of $h(a)$, and $\widehat{S}(t|z_i) = \exp\left\{-\widehat{\Lambda}_0(t) - \beta^\top z_i t\right\}$. Hence, we can estimate $H(a)$ by

$$\widehat{H}(a) = \underset{H}{\operatorname{argmax}}\, PL_M(\beta, H).$$

In (16), we observe that $h(a)$ only appears at the truncation time A_j in both the numerator and the denominator. In addition, our goal is that the likelihood function can be as large as possible, so value in numerator must be large, and value in denominator has to be small. Hence, in the spirit of the empirical likelihood estimation, we can search the maximizer $\widehat{H} \in \{H : h(t) = h_l$ if $t = a_l, l = 1, \ldots, n$ and $h(t) = 0$ otherwise $\}$. As a result, we can discretize (16) and therefore the estimator of $H(a)$ is given by

$$\widehat{H}(a) = \underset{\Phi}{\operatorname{argmax}}\, PEL_M(\beta, H), \tag{17}$$

where $\Phi = \{h_l : 0 \leq h_l \leq 1 \text{ and } \sum\limits_l h_l = 1\}$ and

$$PEL_M(\beta, H) = \prod_{i=1}^{n} \frac{\widehat{S}(a_i|z_i) h_i}{\sum_{j=1}^{n} \widehat{S}(a_j|z_i) h_j}. \tag{18}$$

Next, since $\mathbf{h} = (h_1, \ldots, h_n)$ and β are unknown in (18), so we have to estimate $\mathbf{h}$ and β directly from (18). First of all, we take the partial derivative of $\log PEL_M(\beta, H)$ with respect to h_l, and solve $\frac{\partial}{\partial h_l} \log PEL_M(\beta, H) = 0$ for $l = 1, \ldots, n$. By simple computations, we can obtain the iterated form

$$h_l = \left(\sum_{i=1}^{n} \frac{\widehat{S}(a_l|z_i)}{\sum_{j=1}^{n} \widehat{S}(a_l|z_i) h_j} \right)^{-1} \tag{19}$$

for $l = 1, \ldots, n$. We can easily observe that (19) follows the form of the fixed point, say $x = g(x)$. Therefore, we can discuss its convergence and find its convergent value. On the other hand, we have to find the estimator of β. However, a crucial problem is that we cannot find the closed form for the estimator of β. In this case, we apply the numerical method to find the estimator of β by maximizing (18). The following is the algorithm to find the estimators of β and $\mathbf{h}$:

Step 1 : Given initial value $h_i^{(0)} = \frac{1}{n}$ $\forall i = 1, \ldots, n$, and $\beta^{(0)} = \widehat{\beta}$.

Step 2 : Update $\mathbf{h}$ by formula (19). That is, $h_i^{(k+1)} = g(h_i^{(k)}), i = 1, \ldots, n$.

Step 3 : Given the updated value $\mathbf{h}^{(k+1)}$ in Step 2, update β by maximizing $PEL_M(\beta, H)$, which is given by

$$\beta^{(k+1)} = \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{n} \left\{ \log \widehat{S}(a_i|z_i) + \log h_i^{(k+1)} - \log \left(\sum_{j=1}^{n} \widehat{S}(a_j|z_i) h_j^{(k+1)} \right) \right\}.$$

Step 4 : Continue Steps 2 and 3 until $\mathbf{h}^{(k+1)}$ converges and denote it as $\widehat{\mathbf{h}}$.

By (19) and the algorithm, we have the following proposition of $\widehat{\mathbf{h}}$. Its proof is placed in Appendix C.

Proposition 4.1. *Suppose that there is no tied in the truncation time, i.e.,* $a_1 < a_2 < \cdots < a_n$*, then the estimator* $\widehat{\mathbf{h}}$ *is convergent.*

Finally, we can obtain estimator of $H(a)$, which is given by

$$\widehat{H}(a) = \sum_{i=1}^{n} I(A_i \leq a) \widehat{h}_i, \tag{20}$$

where $\widehat{h}_i, i = 1, \ldots, n$ is obtained from (19) in the algorithm.

Actually, Huang and Qin [12] also presented the estimator of $H(\cdot)$, which is given by

$$\widetilde{H}(a) = \frac{\sum\limits_i \widehat{S}_0^{-1}(A_i) \exp(\widehat{\beta}^\top Z_i A_i) I(A_i \leq a)}{\sum\limits_i \widehat{S}_0^{-1}(A_i) \exp(\widehat{\beta}^\top Z_i A_i)}, \tag{21}$$

where $\widehat{S}_0(t) = \exp\left\{ -\widehat{\Lambda}_0 \left(t, \widehat{\beta}_\phi \right) \right\}$. We shall compare the performances of (20) and (21) in the simulation study.

4.2 Estimation of $\lambda_0(t)$

In addition to L_M, the likelihood function $L_C(\lambda_0, \beta)$ also contains information of β, so we have to incorporate $L_C(\lambda_0, \beta)$ in the analysis. We can observe that there exists not only $\Lambda_0(\cdot)$ and β, $\lambda_0(\cdot)$ is also involved in the likelihood function $L_C(\lambda_0, \beta)$. Besides, different from the Cox PH model, we can not derive (7) as the partial likelihood which contains β only (e.g., [34, 11]), so the unknown baseline hazard function $\lambda_0(\cdot)$ can not be directly eliminated when we estimate β. These difficulties make challenges, and it seems not possible to derive the estimator of β by maximizing the conditional likelihood directly. Therefore, we need to deal with the baseline hazard function $\lambda_0(\cdot)$ before deriving the estimator of β.

In order to drive the estimator of $\lambda_0(\cdot)$, we apply the method of the smoothing estimation on $\lambda_0(\cdot)$ [8, Section 6.2.4] and implement this estimator to the likelihood function (7). Let $K(\cdot)$ be the second order symmetric kernel function and let σ be the positive-value bandwidth. The kernel estimator of $\lambda_0(\cdot)$ is given by

$$\widehat{\lambda}_\sigma(y) = \frac{1}{\sigma} \int K\left(\frac{y - \zeta}{\sigma} \right) d\widehat{\Lambda}_0(\zeta, \widehat{\beta}_\phi), \tag{22}$$

where $\widehat{\Lambda}_0(\zeta, \widehat{\beta}_\phi)$ is the consistent estimator of the cumulative baseline hazard (10). Moreover, it can be verified that (22) is the consistent estimator and the asymptotic normality of $(n\sigma)^{1/2} \left\{ \widehat{\lambda}_0(y) - \lambda_0(y) \right\}$ can be established [8].

Specifically, we consider the kernel function $K(x) = \frac{3}{4}(1 - x^2) I(|x| \leq 1)$ in our numerical studies, and we can estimate $\lambda_0(y)$ by

$$\widehat{\lambda}_\sigma(y) = \frac{1}{\sigma} \sum_{j=1}^{n} \frac{3}{4} \left\{ 1 - \left(\frac{y - \zeta_j}{\sigma} \right)^2 \right\} I\left(\left| \frac{y - \zeta_j}{\sigma} \right| \leq 1 \right) \times \left\{ \widehat{\Lambda}_0(\zeta_j, \widehat{\beta}_\phi) - \widehat{\Lambda}_0(\zeta_{j-1}, \widehat{\beta}_\phi) \right\}.$$

4.3 Pseudo likelihood estimation for β

So far, we have estimators $\widehat{\Lambda}_0(\cdot)$, $\widehat{\lambda}_0(\cdot)$ and $\widehat{\mathbf{h}}$ derived in previous subsections. Finally, plugging in consistent estimators (10), (22), and the convergent value $\widehat{\mathbf{h}}$ to the full likelihood (6) yields the pseudo likelihood function of β, which is given by

$$(23) \qquad PL_\sigma(\beta) = \prod_{i=1}^{n} \frac{\left(\widehat{\lambda}_\sigma(y_i) + \beta^\top z_i\right)^{\delta_i} \widehat{S}(y_i|z_i)\widehat{h}_i}{\sum_{j=1}^{n} \widehat{S}(a_j|z_i)\widehat{h}_j}.$$

Furthermore, in the estimator (22), one of crucial issues is the bandwidth selection. Our approach to select bandwidth is the *leave one subject out* cross-validation. By deleting-one subject in $\widehat{\Lambda}_0(\cdot)$, we have

$$(24) \qquad \widehat{\Lambda}_0^{(-i)}(t, \widehat{\beta}_\phi) = \int_0^t \frac{\sum_{j\neq i} \left\{dN_j(u) - R_j(u)\widehat{\beta}_\phi Z_j du\right\}}{\sum_{j\neq i} R_j(u)}.$$

From (22) and (24), we have the deleted-one-subject smoothing estimator

$$(25) \qquad \widehat{\lambda}_\sigma^{(-i)}(y) = \frac{1}{\sigma}\int K\left(\frac{y-\zeta}{\sigma}\right) d\widehat{\Lambda}_0^{(-i)}(\zeta, \widehat{\beta}_\phi).$$

Plugging in (25) to (23) yields

$$(26) \qquad PL_\sigma(\beta) = \prod_{i=1}^{n} \frac{\left(\widehat{\lambda}_\sigma^{(-i)}(y_i) + \beta^\top z_i\right)^{\delta_i} \widehat{S}^{(-i)}(y_i|z_i)\widehat{h}_i}{\sum_{j=1}^{n} \widehat{S}^{(-i)}(a_j|z_i)\widehat{h}_j}.$$

The estimator $\widehat{\sigma}$ is derived by maximizing (26) for any fixed β.

Finally, replacing σ in (23) by $\widehat{\sigma}$, we can derive the estimator of β by maximizing (23), and denote the estimator by

$$(27) \qquad \widehat{\beta}_F = \underset{\beta}{\operatorname{argmax}}\, PL_{\widehat{\sigma}}(\beta).$$

5. SIMULATION

Several simulation scenarios are conducted to investigate the performance of the pseudo likelihood estimation under designed additive hazards models and different censoring rates (c.r.). Further, some existing methods described in Section 3, including the conditional estimating equations (CEE) approach, the pairwise pseudo likelihood estimation (PPL) approach and the combined estimating equation approach proposed by Huang and Qin [12], are conducted in this section. In the numerical experiments, each simulation settings are repeatedly generated 1000 times with sample sizes $n = 200, 300$ and 400.

5.1 Model settings

Four model formulations for (A^*, T^*) are considered in this simulation study as follows:
Model 1. $\lambda(t|z) = 0.5\sqrt{t} + z$, $A^* \sim U(0, 100)$.
Model 2. $\lambda(t|z) = 0.5\sqrt{t} + z$, $A^* \sim \exp(10)$.
Model 3. $\lambda(t|z) = 1.5\sqrt{t} + 0.5z$, $A^* \sim \exp(10)$.
Model 4. $\lambda(t|z) = \frac{4}{3}\sqrt[3]{t} + z$, $A^* \sim \exp(10)$.

It is noted that Models 1 to 2 are the same with the settings in Huang and Qin [12]. Apparently, the survival times are generated from the additive hazards model of the form $\lambda(t|z) = \lambda_0(t) + \beta z$. In addition, Z^* is generated from the uniform distribution $U(0, 1)$. As mentioned in Section 2, the collected data of the form $\{(A_i, T_i, Z_i) : i = 1, \ldots, n\}$ are obtained from (A_i^*, T_i^*, Z_i^*) given $T^* \geq A^*$. Independent of (A_i^*, T_i^*, Z_i^*), the censoring time C_i is generated from the uniform distribution $U(0, \tau_c)$ with τ_c being specified to produce the expected censoring rates of about 0%, 25%, and 50%, respectively.

5.2 Simulation results

Tables 1 to 4 summarize the empirical mean, the empirical standard deviation (S.E.), the mean squared error (MSE) and the coverage probability (CP) under Models 1 to 4 with the different censoring rates and sample sizes. In our four model settings, we observe that the variation increases as the censoring rate increases and variations in the small sample are obviously larger than variations in the larger sample for all estimators. For comparisons, the biases of the proposed method are comparable with other methods for Models 1 and 2, and the biases do not have significant differences under Models 3 and 4. For S.E., we first observe that both $\widehat{\beta}_\phi$ and $\widehat{\beta}_\psi$ produce noticeable S.E., which verifies that the estimator would be inefficient if either (7) or (8) is ignored. Second, both $\widehat{\beta}$ and $\widehat{\beta}_F$ are estimators obtained by the combinations of (7) and (8), so we can see that the efficiency of the estimator is improved. Finally, from the comparisons of S.E. between $\widehat{\beta}$ and $\widehat{\beta}_F$, we can see that the proposed estimator $\widehat{\beta}_F$ has the smaller S.E. than estimator $\widehat{\beta}$ proposed by Huang and Qin [12]. Besides, MSE of $\widehat{\beta}_F$ is the smallest among all estimators. From simulation studies, the numerical results indeed show that the proposed method improves the efficiency of the estimator and the performance is better than others. In addition, we can further observe that the coverage probability based on the CEE approach or PPL method are relatively large. The main reason is that both CEE and PPL methods produce the larger S.E., which yields the wider interval and therefore the coverage probability is over-estimated. On the contrary, the coverage probability of our proposed method is close to 95%, which means that our proposed method gives the more precise estimator.

5.3 Investigation of the distribution of truncation time

Huang and Qin [12] discussed the performance of the estimator when different distributions of the truncation time

Table 1. Simulation result for Model 1

n	c.r.	CEE($\widehat{\beta}_\phi$)				PPL($\widehat{\beta}_\psi$)				HuangQin($\widehat{\beta}$)				Proposed($\widehat{\beta}_F$)			
		bias	SE	MSE	CP	bias	SE	MSE	CP	bias	SE	MSE	CP	bias	SE	MSE	CP
200	0%	0.00	0.27	0.07	97	0.04	0.32	0.11	97	0.01	0.21	0.05	94	0.00	0.20	0.04	95
	25%	-0.01	0.31	0.10	96	0.03	032	0.11	96	0.00	0.23	0.05	95	0.00	0.21	0.04	95
	50%	-0.04	0.36	0.13	97	0.02	0.34	0.12	93	-0.01	0.26	0.07	94	-0.01	0.24	0.06	95
300	0%	-0.02	0.23	0.05	98	0.03	0.28	0.08	97	0.00	0.18	0.03	97	0.00	0.16	0.03	96
	25%	0.00	0.25	0.06	96	0.04	0.27	0.08	94	0.01	0.19	0.04	95	0.01	0.18	0.03	95
	50%	-0.01	0.30	0.09	94	0.04	0.27	0.08	93	0.02	0.21	0.04	96	0.01	0.19	0.04	95
400	0%	0.00	0.17	0.03	99	-0.01	0.23	0.05	97	-0.01	0.15	0.02	90	0.00	0.13	0.01	95
	25%	-0.06	0.38	0.15	82	0.01	0.22	0.47	97	-0.01	0.19	0.04	97	-0.02	0.17	0.03	94
	50%	-0.01	0.22	0.05	94	0.01	0.22	0.05	98	0.00	0.16	0.03	96	0.00	0.14	0.02	96

Note:
c.r.: Censoring rate.
bias: Difference between the empirical mean and the true value.
SE: Standard error. Square root of the empirical variance.
MSE: Mean square error.
CP: Coverage of probability (%).
CEE: The Conditional Estimation Equation in Section 3.1.
PPL: The Pairwise Pseudo Likelihood Method in Section 3.2.
HuangQin: The Method proposed by Huang and Qin [12] in Section 3.3.
Proposed: The Proposed method in Section 4.

Table 2. Simulation result for Model 2

n	c.r.	CEE($\widehat{\beta}_\phi$)				PPL($\widehat{\beta}_\psi$)				HuangQin($\widehat{\beta}$)				Proposed($\widehat{\beta}_F$)			
		bias	SE	MSE	CP	bias	SE	MSE	CP	bias	SE	MSE	CP	bias	SE	MSE	CP
200	0%	-0.02	0.28	0.08	93	0.02	0.35	0.12	97	-0.01	0.23	0.05	96	0.00	0.20	0.04	95
	25%	-0.01	0.31	0.10	96	0.04	0.35	0.13	98	0.01	0.24	0.06	95	-0.01	0.22	0.05	95
	50%	-0.04	0.36	0.13	96	0.04	0.36	0.13	97	0.00	0.27	0.07	95	-0.01	0.24	0.06	95
300	0%	-0.01	0.22	0.05	93	0.03	0.30	0.09	97	0.00	0.19	0.04	95	0.00	0.17	0.03	95
	25%	-0.01	0.25	0.06	93	0.03	0.29	0.08	96	0.00	0.20	0.04	95	0.00	0.18	0.03	95
	50%	-0.01	0.29	0.08	94	0.03	0.29	0.08	96	0.00	0.21	0.04	95	-0.01	0.19	0.04	95
400	0%	-0.02	0.17	0.03	97	0.02	0.25	0.06	96	-0.01	0.15	0.02	94	0.00	0.14	0.02	95
	25%	-0.03	0.35	0.12	94	0.05	0.26	0.07	98	0.03	0.21	0.05	95	0.01	0.19	0.04	95
	50%	-0.03	0.30	0.09	94	0.06	0.25	0.07	96	0.02	0.21	0.04	95	0.01	0.19	0.03	95

Table 3. Simulation result for Model 3

n	c.r.	CEE($\widehat{\beta}_\phi$)				PPL($\widehat{\beta}_\psi$)				HuangQin($\widehat{\beta}$)				Proposed($\widehat{\beta}_F$)			
		bias	SE	MSE	CP	bias	SE	MSE	CP	bias	SE	MSE	CP	bias	SE	MSE	CP
200	0%	-0.02	0.42	0.18	98	0.02	0.53	0.28	98	-0.01	0.33	0.11	95	0.00	0.28	0.08	95
	25%	-0.02	0.46	0.21	97	0.02	0.53	0.28	98	-0.01	0.35	0.12	96	0.01	0.31	0.09	95
	50%	-0.02	0.51	0.27	97	0.01	0.53	0.28	98	-0.01	0.37	0.14	96	0.01	0.34	0.12	95
300	0%	-0.01	0.35	0.12	92	-0.01	0.46	0.21	92	-0.01	0.28	0.08	94	0.00	0.24	0.06	95
	25%	-0.01	0.36	0.13	94	0.01	0.42	0.18	91	-0.01	0.27	0.07	95	0.02	0.24	0.06	96
	50%	0.00	0.42	0.18	93	0.01	0.43	0.19	89	0.00	0.31	0.09	95	0.01	0.27	0.07	95
400	0%	0.05	0.29	0.08	93	0.05	0.35	0.13	94	0.04	0.22	0.05	93	0.02	0.20	0.05	95
	25%	-0.09	0.46	0.22	90	0.05	0.36	0.13	91	0.00	0.28	0.08	95	0.00	0.26	0.07	95
	50%	-0.09	0.41	0.18	91	0.05	0.35	0.13	94	-0.01	0.27	0.07	95	0.00	0.26	0.07	96

Table 4. Simulation result for Model 4

n	c.r.	CEE($\widehat{\beta}_\phi$)				PPL($\widehat{\beta}_\psi$)				HuangQin($\widehat{\beta}$)				Proposed($\widehat{\beta}_F$)			
		bias	SE	MSE	CP	bias	SE	MSE	CP	bias	SE	MSE	CP	bias	SE	MSE	CP
200	0%	-0.01	0.43	0.19	97	0.02	0.52	0.27	98	0.01	0.34	0.12	95	0.00	0.30	0.10	95
	25%	-0.04	0.48	0.24	96	0.03	0.51	0.27	96	-0.01	0.36	0.13	94	0.01	0.33	0.11	94
	50%	-0.06	0.54	0.29	94	0.03	0.51	0.26	96	-0.01	0.39	0.15	95	0.01	0.35	0.13	95
300	0%	-0.01	0.36	0.13	97	0.01	0.43	0.19	98	-0.01	0.28	0.08	96	0.01	0.25	0.07	95
	25%	-0.02	0.40	0.16	96	0.01	0.43	0.19	97	-0.01	0.30	0.09	95	0.02	0.27	0.08	94
	50%	-0.02	0.46	0.21	97	0.00	0.43	0.19	96	-0.01	0.32	0.10	96	0.00	0.28	0.08	94
400	0%	-0.03	0.32	0.10	95	-0.03	0.35	0.12	98	-0.03	0.22	0.05	95	0.02	0.20	0.04	95
	25%	-0.06	0.51	0.26	95	-0.01	0.39	0.15	90	-0.02	0.32	0.10	95	0.02	0.29	0.09	95
	50%	0.07	0.41	0.17	93	0.06	0.40	0.16	92	0.06	0.28	0.08	94	0.04	0.27	0.08	96

Table 5. Different distributions of the truncation time and the corresponding truncation rates

Distribution	Truncation rate
$U(0,1)$	0.33
$U(0,10)$	0.89
$U(0,100)$	0.99
$4Exp(0.1)$	0.22
$Exp(10)$	0.89
$Bin(2,0.75)$	0.68
$Bin(4,0.75)$	0.90

Note:
$U(0,100)$ and $Exp(10)$ are simulation designs in Section 5.1; the others are additional designs to compare the truncation rate.

are concerned. In this article, we also investigate the performance of the estimator under different distributions of the truncation time and compare two different estimators of $H(\cdot)$ as described in (20) and (21).

We first introduce the truncation rate. The truncation rate is defined as $P(T^* < A^*)$, and it means the proportion that individuals who are not recruited in the collected dataset. Be more specific, in the simulation, we can repeatedly generate data for subjects, but we only recruit subject who satisfies $T^* \geq A^*$. Similar to the description in Section 5.1, the repetition of data generation procedure stops when the assigned sample size n is achieved. Here we let tr denote the total number of the repetition in generating data. Therefore, the truncation rate is determined by $1-n/tr$. The detailed data generation is deferred in Appendix D.

We now take the uniform distribution, the exponential distribution, and the binomial distribution with different rates as examples to examine the estimators in the simulation study. We first summarize their truncation rates in Table 5. Here we take the uniform distribution as an example to explain the effect of the truncation rate. If the truncation time follows $U(0,1)$, then it may be no need to generate much data to achieve the desired sample size n. Thus the truncation rate is lower since tr is small. On the other hand, if the truncation time follows $U(0,10)$, then $P(T^* \geq A^*)$ is lower. It means that those individuals have a higher probability not to be recruited in the study, and this case may yield a large truncation rate. For the estimation of $H(\cdot)$, it is expected that the estimator performs well if the truncation rate is low. The reason is that if the truncation rate is low, then most of the truncation times A^* satisfying $T^* \geq A^*$ are in the dataset, and therefore the estimator can perform well. On the contrary, if the truncation rate is high, then the truncation time A^* which satisfies $T^* \geq A^*$ in the dataset is totally different from those in the population. Hence, the performance of the estimator would be worse in this situation. To see this phenomenon, we compare the two estimators (20) and (21) with the true curve, and plot them in Figures 3 to 6. Figures 3 and 4 are $U(0,1)$ with sample sizes $n=200$ and $n=300$, respectively, we can see that the two estimators are close to the true curve.

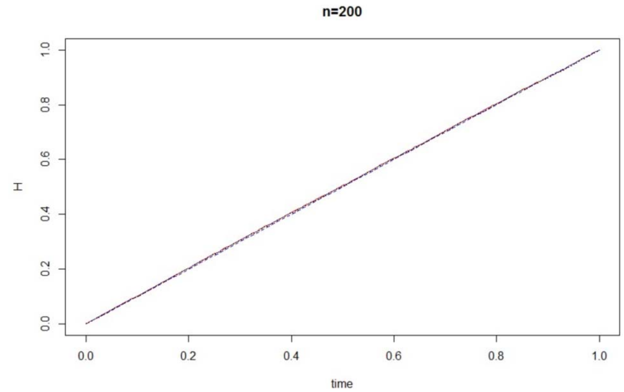

Figure 3. Distribution function of truncation time with $n=200$. Solid line is estimated curve $\widehat{H}(a)$, dot line is estimated curve $\widetilde{H}(a)$, and dash line is the true curve $H(a)=a$.

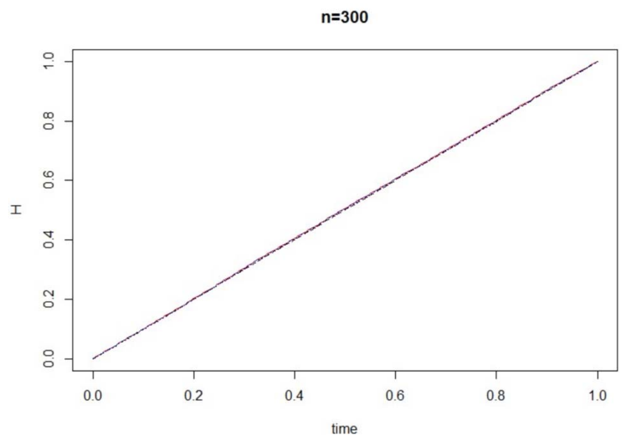

Figure 4. Distribution function of truncation time with $n=300$. Solid line is estimated curve $\widehat{H}(a)$, dot line is estimated curve $\widetilde{H}(a)$, and dash line is the true curve $H(a)=a$.

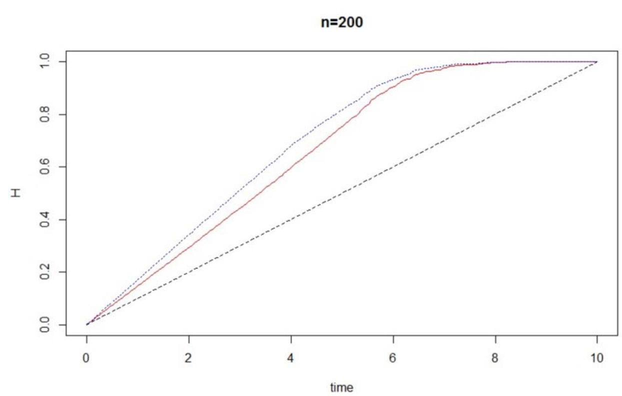

Figure 5. Distribution function of truncation time with $n=200$. Solid line is estimated curve $\widehat{H}(a)$, dot line is estimated curve $\widetilde{H}(a)$, and dash line is the true curve $H(a)=\frac{a}{10}$.

Table 6. Summary statistics for the estimated regression parameter under different truncation distributions. The censoring rate is set to 0%

n	dist.	CEE($\widehat{\beta}_\phi$)			PPL($\widehat{\beta}_\psi$)			HuangQin($\widehat{\beta}$)			Proposed($\widehat{\beta}_F$)		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
200	$U(0,1)$	-0.01	0.25	0.06	-0.03	0.79	0.63	-0.01	0.24	0.06	-0.01	0.22	0.05
	$U(0,10)$	-0.03	0.27	0.08	0.02	0.34	0.12	-0.01	0.22	0.05	0.00	0.20	0.04
	$4Exp(0.1)$	-0.01	0.25	0.06	-0.02	0.75	0.56	-0.02	0.25	0.06	-0.01	0.20	0.04
	$Bin(4,0.75)$	-0.09	0.32	0.11	0.01	0.32	0.10	-0.03	0.23	0.06	-0.02	0.23	0.05
	$Bin(2,0.75)$	-0.03	0.29	0.09	0.00	0.39	0.15	-0.03	0.24	0.06	-0.02	0.21	0.05
300	$U(0,1)$	-0.02	0.20	0.04	-0.06	0.69	0.48	-0.02	0.20	0.04	-0.03	0.18	0.03
	$U(0,10)$	0.01	0.22	0.05	0.00	0.26	0.07	0.00	0.18	0.03	0.00	0.16	0.03
	$4Exp(0.1)$	0.00	0.20	0.04	-0.02	0.66	0.43	0.00	0.18	0.03	0.00	0.16	0.03
	$Bin(4,0.75)$	-0.06	0.27	0.07	0.01	0.25	0.06	-0.03	0.19	0.04	-0.01	0.18	0.03
	$Bin(2,0.75)$	-0.01	0.24	0.06	0.02	0.32	0.10	0.00	0.19	0.04	-0.01	0.18	0.03

Note:
dist. - Distribution of truncation time;
$U(0,a)$ - Uniform distribution with support $[0,a]$;
$Exp(\mu)$ - Exponential distribution with mean μ;
$Bin(n,0.75)$ - Binomial distribution with size n and success probability 0.75.

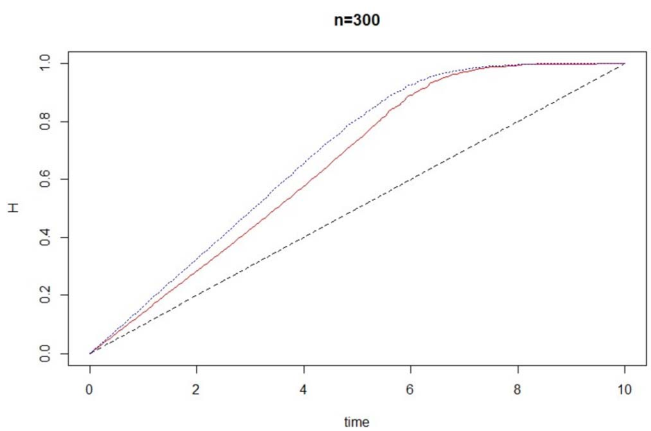

Figure 6. Distribution function of truncation time with $n = 300$. Solid line is estimated curve $\widehat{H}(a)$, dot line is estimated curve $\widetilde{H}(a)$, and dash line is the true curve $H(a) = \frac{a}{10}$.

On the other hand, Figures 5 and 6 are $U(0,10)$ with sample sizes $n = 200$ and $n = 300$, respectively, and they show that the both estimators (20) and (21) have larger pointwise biases. For the comparisons between (20) and (21), as shown in Figures 3 and 4, if the truncation rate is low, then both estimated curves have the similar performances. On the contrary, if the truncation rate is high, then Figures 5 and 6 illustrate that our proposed curves are much closer to the true curves. It implies that the estimated curves of our proposed method have the smaller pointwise bias.

Finally, we study the performance of the estimator of β. Here we consider the additive hazards model $\lambda(t|z) = 0.5\sqrt{t} + z$ with $Z \sim U(0,1)$. Different distributions of the truncation time, including the exponential distribution, the binomial distribution, and the uniform distribution, are examined. The censoring rate is set to be 0%. We again calculate the estimators with sample sizes $n = 200$ and 300. We summarize the mean, S.E., and MSE in Table 6. The numerical results show that the proposed method has no significant difference when the truncation rates change. Besides, we can see that the biases, S.E., and MSE of the proposed method are smaller than the other three estimators. It indicates that the performances of our method are always better and more robust than others regardless of the change of the distributions of the truncation time and the truncation rates.

6. ANALYSIS OF WORCESTER HEART ATTACK STUDY DATA

In this section, we apply our proposed methods to analyze the data arising from the Worcester Heart Attack Study (WHAS500), which is described in Section 1. Specifically, as discussed by Hosmer, Lemeshow, and May [9], the beginning of a survival time was defined as the time that subject was admitted to a hospital. The main interest is in the survival times of patients who were discharged alive from hospitals. Hence, a selection criterion was imposed that only those subjects who were discharged alive were eligible to be included in the analysis. That is, their minimum survival time would be the length of their hospital stay; individuals whose failure times did not exceed the minimum survival time were not enrolled in this analysis, and hence the left-truncation happens. With such a criterion, a sample of size 461 was selected and the truncation rate was approximately 7.8%. Be more specifically, total length of follow-up (lenfol) is the last event time (i.e., $Y_i = \min\{T_i, A_i + C_i\}$), length of hospital stay (los) is the truncation time (i.e., A_i), and vital status at last follow-up (fstat) is δ_i. These 461 patients contribute the measurements which satisfy the constraint $T_i \geq A_i$. In this dataset, the censoring rate is 61.8%. To be consistent with the simulation settings and to emphasize the estimation methods, here we employ the body mass index (BMI) as the only covariate and denote it as Z.

To investigate the risk difference attributed to the risk factors, here we use the additive hazards model to fit this data. We use the proposed method to obtain the estimator, $\widehat{\beta}_F = 0.0112$, which means that for the given time $t = t_0$, the risk difference, $\lambda(t_0|z+1) - \lambda(t_0|z)$, is approximately 0.0112. In order to do the inference and the hypothesis test, we need to construct the 95% confidence interval. Here we adopt the non-parametric bootstrap method by sampling 461 subjects with replacement from the dataset with 2000 times repetition.

From the bootstrap method, the 95% confidence interval is $(0.0006, 0.0217)$ and the p-value is roughly 0.037. Since the p-value is smaller than the significant level $\alpha = 0.05$, then we conclude that BMI may be significant in the risk difference. For the comparisons, we also illustrate the estimators discussed in Section 3. For those estimators, we have $\widehat{\beta}_\phi = 0.0217$, $\widehat{\beta}_\psi = 0.0435$ and $\widehat{\beta} = 0.0147$ with the 95% confidence intervals $(0.0108, 0.0542)$, $(0.0020, 0.08898)$ and $(0.0098, 0.0392)$, respectively. Moreover, the p-values of $\widehat{\beta}_\phi$, $\widehat{\beta}_\psi$ and $\widehat{\beta}$ are 0.059, 0.0163, 0.047, respectively. Therefore, we conclude that two methods in Sections 3.2 and 3.3 give the similar results with our proposed method, while the estimator based on the condition estimating equation approach illustrates that BMI is not significant on the risk difference.

7. CONCLUSION

Analysis of left-truncated and right-censored data is an important problem and a challenging topic in survival analysis. Different from the usual data structure, biased and incomplete data is involved. In the past literature, the inferences of the Cox PH model and the transformation model have been developed. On the contrary, little work has been available on the analysis of the additive hazards model with survival data subject to left-truncation and right-censoring. The existing method to deal with the additive hazards model is the estimating equations approach. In order to improve the efficiency of the estimator, we develop the pseudo-likelihood method in this article.

There are some key strengths in this article. First of all, the model misspecification is considered. As mentioned previously, the martingale approach meets its properties, such as zero expectation, only when the assumed model is correct. However, we never know what the true model is in the practical situations. Instead, our proposed method has a valid and robust performance without the model assumption. Secondly, different from the existing methods, the proposed method implements the consistent estimators to the likelihood function (6), and the estimator of β is derived from the likelihood approach. Hence, it is expected that the proposed method can yield the consistent and efficient estimator. Moreover, the simulation results show that the proposed method outperforms other existing methods with relative robustness and easiness of handling different types of dataset when fitting the additive hazards model. Even though the theoretical properties are not fully developed, we can also conclude that the proposed method is doable and the efficiency is improved from the comprehensively numerical studies. The theoretical development will be our future investigation.

Although we only present the results for the time-independent covariates, the proposed method is also expected to handle the time-dependent covariates. In the future research, we will develop the rigorous method for the case of the time-dependent covariate in the future.

APPENDIX A. DERIVATION OF EQUATION (4)

Since the joint distribution of (T, A) is equivalent to $(T^*, A^*)|T^* \geq A^*$, then

$$\begin{aligned} P(T<t, A<a) &= P(T^*<t, A^*<a|T^* \geq A^*) \\ &= \frac{P(T^*<t, A^*<a, T^* \geq A^*)}{P(T^* \geq A^*)} \\ &= \frac{P(T^*<t)\,P(A^*<a)}{P(T^* \geq A^*)}, \end{aligned} \tag{A.1}$$

where the last identity is due to the assumption (A1). On the other hand, let $F(\cdot)$ be the cumulative density function (CDF) of T^*, then

$$\begin{aligned} P(T^* \geq A^*) &= \int_0^\infty \int_0^a f(t)h(a)dtda \\ &= \int_0^\infty F(t)h(a)\Big|_a^\infty da \\ &= \int_0^\infty \{1-F(a)\}\,h(a)da \\ &= \int_0^\infty S(a)h(a)da, \end{aligned} \tag{A.2}$$

where the last identity is due to the definition of the survivor function. Therefore, together with (A.1) and (A.2) and conditional on the covariate Z, the joint probability density function of (T, A) is

$$\frac{f(t|z)h(a)}{\int_0^\infty S(a|z)h(a)da}, \tag{A.3}$$

which is the left hand side of equation (4). Furthermore, we decompose (A.3) by multiplying/dividing $S(a|z)$, so that we have

$$\frac{f(t|z)}{S(a|z)} \times \frac{S(a|z)h(a)}{\int_0^\infty S(u|z)h(u)du}.$$

Therefore, we complete the derivation. □

APPENDIX B. THE ESTIMATOR OF $\Lambda_0(\cdot)$ BASED ON LIKELIHOOD APPROACH

In this section, we discuss the derivation for the estimator of $\Lambda_0(\cdot)$ based on likelihood function approach.

In the conditional likelihood function (7), we let the baseline hazard function $\lambda_0(\cdot)$ jump only at y_i. That is, $\lambda_j = \lambda_0(y_j)$ for $j = 1, \cdots, n$, and zero otherwise. Therefore, the cumulative baseline function can be expressed as $\Lambda_0(t) = \sum_{j=1}^{n} I(y_j < t)\lambda_j$. Taking log on likelihood function (7) yields

$$\begin{aligned}\ell &= \sum_{i=1}^{n} \Big\{ \delta_i \log\left(\lambda_i + \beta^\top z_i\right) \\ &\quad - \sum_{j=1}^{n} \lambda_j I(a_i < y_j < y_i) - \beta^\top z_i (y_i - a_i) \Big\}. \end{aligned} \tag{B.1}$$

Given k with $k = 1, \cdots, n$, taking partial derivative on (B.1) with respective to λ_k gives

$$\ell_k = \frac{\partial \ell}{\partial \lambda_k} = \frac{\delta_k}{\lambda_k + \beta^\top z_k} - \sum_{i=1}^{n} I(a_i < y_k < y_i).$$

Solving $\ell_k = 0$ yields

$$\widetilde{\lambda}_k = \frac{\delta_k - \beta^\top z_k \sum_{i=1}^{n} I(a_i < y_k < y_i)}{\sum_{i=1}^{n} I(a_i < y_k < y_i)}.$$

Hence, consistent with notation in Section 3.1, the estimator of $\Lambda_0(\cdot)$ is

$$\begin{aligned}\widetilde{\Lambda}_0(t) &= \sum_{k=1}^{n} I(y_k < t)\widetilde{\lambda}_k \\ &= \frac{\sum_{k=1}^{n} I(y_k < t)\left\{\delta_k - \beta^\top z_k \sum_{i=1}^{n} I(a_i < y_k < y_i)\right\}}{\sum_{i=1}^{n} I(a_i < y_k < y_i)} \\ &= \int_0^t \frac{\sum_{i=1}^{n}\{dN_i(u) - R_i(u)\beta^\top Z_i du\}}{\sum_{i=1}^{n} R_i(u)}. \end{aligned}$$

We can observe that $\widetilde{\Lambda}_0(t)$ is exactly same as (10) based on the martingale approach. Therefore, we can directly implement (10) in our proposed method. □

APPENDIX C. THE PROOF OF PROPOSITION 4.1

Lemma C.1. *Let $g : \mathbb{R} \to \mathbb{R}$ be a differentiable function. If $x^* = g(x^*)$ and $|g'(x^*)| < 1$, then the iterative scheme is locally convergent.*

Proof of Lemma C.1. If x^* is a fixed point, then for the error at the kth iteration, we have

$$e_{k+1} = x_{k+1} - x^* = g(x_k) - g(x^*) \tag{C.1}$$

By the Mean Value Theorem, there is a point θ_k between x_k and x^*, such that

$$g(x_k) - g(x^*) = g'(\theta_k)(x_k - x^*) \tag{C.2}$$

From (C.2), we can obtain

$$e_{k+1} = g'(\theta_k) e_k \tag{C.3}$$

We do not know the value of θ_k, but if $|g'(x^*)| < 1$, then by starting the iterations close enough to x^*, we can be assured that there is a constant C such that $|g'(\theta_k)| \leq C < 1$, for $k = 0, 1, \ldots$. Thus, by (2), we have

$$|e_{k+1}| \leq C|e_k| \leq \cdots \leq C^k |e_0|$$

Since $C < 1$ implies $C^k \to 0$, so $|e_k| \to 0$ and the sequence is converge. □

After stating Lemma C.1, then we now prove Proposition 4.1.

Proof of Proposition 4.1. Without loss of generality, we consider iterated form $\mathbf{h}$ in arbitrary l, say

$$h_l = \left(\sum_{i=1}^{n} \frac{\widehat{S}(a_l|z_i)}{\sum_{j=1}^{n} \widehat{S}(a_j|z_i) h_j} \right)^{-1}, \quad l = 1, \cdots, n.$$

One can observe that it is the form of fixed point, say $h = g(h)$, where

$$g(h_l) = \left(\sum_{i=1}^{n} \frac{\widehat{S}(a_l|z_i)}{\sum_{j=1}^{n} \widehat{S}(a_j|z_i) h_j} \right)^{-1}.$$

To show h_l^{k+1} is convergent, by Lemma C.1, it is sufficient to prove $\left|\frac{\partial g}{\partial h_l}\right| < 1$ for all $l = 1, \ldots, n$. Define $b_i = \frac{\widehat{S}(a_l|z_i)}{\sum_{j=1}^{n} \widehat{S}(a_j|z_i) h_j}$, and a simple algebraic computation yields $\left|\frac{\partial g}{\partial h_l}\right| = \frac{\sum_{i=1}^{n} b_i^2}{\left(\sum_{i=1}^{n} b_i\right)^2}$. Since $b_i > 0$ for all $i = 1, 2, \ldots, n$, so $\sum_{i=1}^{n} b_i^2 \leq \left(\sum_{i=1}^{n} b_i\right)^2$ always holds. Hence, we obtain $\left|\frac{\partial g}{\partial h_l}\right| = \frac{\sum_{i=1}^{n} b_i^2}{\left(\sum_{i=1}^{n} b_i\right)^2} \leq 1$ for all $l = 1, \ldots, n$. To show the strictly inequality, we suppose that there is no tied in the truncation times, which means that

$a_1 < a_2 < \cdots, a_n$. It is well-known that the survivor function is the decreasing function. Therefore, given z_i, we have $\widehat{S}(a_1|z_i) > \widehat{S}(a_2|z_i) > \cdots > \widehat{S}(a_n|z_i)$. From the expression by b_i, we also have $1 > b_1 > b_2 > \cdots > b_n > 0$. Thus, we derive that $\sum_{i=1}^{n} b_i^2 < \left(\sum_{i=1}^{n} b_i\right)^2$. □

APPENDIX D. DATA GENERATION AND CALCULATION OF TRUNCATION RATE

```
library(rootSolve)
s=1
n=200
tr = 0
data0 = NULL
while(s<=n)
{
 Z <-runif(1,0,1)
 A <-4*rexp(1,10)
 U <-runif(1,0,1)
 failure=function(T) {
      (1/3) * T^(3/2) + Z * T
    + log(1-U, base=exp(1)) }
 T=uniroot.all(failure,c(0,100))
 if(T>A)
 {
  data0=rbind(data0,c(Z,T,A))
  s=s+1
  }
 tr = tr+1
}
1 - n/tr  # truncation rate.
```

APPENDIX E. CODE FOR FITTING THE MODELS

```
# ----------Conditional Part----------
phi = function(b) {

R=outer(data[,3],data[,1],"<=")
      *outer(data[,1],data[,1],">=")
riskset=colSums(R)
riskset[riskset==0]<-10^10
k=colSums(matrix(data[,4],n,n,F)*R)
zbar=k/riskset


Lambday = Y1 - b*Y2
lambday = Lambday - c(0,Lambday[1:n-1])

N = outer(data[,1],data[,1],"<=")
    *matrix(data[,2],n,n,F)
dN = N - cbind(rep(0,n),N[,1:n-1])
dt = data[,1] - c(0,data[1:n-1,1])
dM = dN - R*matrix(lambday,n,n,T)
    - b*R*matrix(data[,4],n,n,F)
      *matrix(dt,n,n,T)


ans1 = rowSums(matrix(data[,4],n,n,F)*dM)
ans2 = rowSums(matrix(zbar,n,n,T)*dM)

result = sum(ans1 - ans2)
return(result)
}



 test1 = NULL
 for(i in 1:200) {
 test1 = rbind(test1,c(time[i],
                abs(phi(time[i]))))
 }
 betac = test1[which(test1[,2]
                ==min(test1[,2])),1]
 Beta_coll1 = c(Beta_coll1,betac)

# ----------Marginal Part----------

rho = outer(data[,4],data[,4],"-")
     *outer(data[,3],data[,3],"-")
shi = function(b) {
ans = -sum(rho/(1+exp(-b*rho)))
      * (2/(n*(n-1)))
return(ans)
}


 test2 = NULL
 for(i in 1:200) {
 test2 = rbind(test2,c(time[i],
                abs(shi(time[i]))))
 }
 betam = test2[which(test2[,2]
                ==min(test2[,2])),1]
 Beta_coll2 = c(Beta_coll2,betam)
# ----------Composite Part----------

 test3 = NULL
 for(i in 1:200) {
 test3 = rbind(test3,c(time[i],
      abs(phi(time[i])/n+shi(time[i]))))
 }
 betaH = test3[which(test3[,2]
                ==min(test3[,2])),1]
 Beta_coll3 = c(Beta_coll3,betaH)
```

ACKNOWLEDGEMENTS

The author would like to extend great gratitude to an Editor, an Associate Editor and two referees for their valuable suggestions and useful comments to make this paper better.

Received 25 September 2017

REFERENCES

[1] Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993) *Statistical Models Based on Counting Processes*. Springer. MR1198884

[2] Asgharian, M. and Wolfson D. B. (1966). (2005) Asymptotic behavior of the unconditional NPMLE of the length-biased survival function from right censored prevalent cohort data. *The Annals of Statistics*, 33, 2109–2131. MR2211081

[3] Chen, C. M. and Shen, P. S. (2017). Conditional maximum likelihood estimation in semiparametric transformation model with LTRC data. *Lifetime Data Analysis*, 24, 250–272. MR3773859

[4] Cheng, Y. J. and Wang, M. C. (2012) Estimating propensity scores and causal survival functions using prevalent survival data. *Biometrics*, 68, 707–716. MR3055175

[5] Cheng, Y. J. and Wang, M. C. (2015) Causal estimation using semiparametric transformation models under prevalent sampling. *Biometrics*, 71, 302–312. MR3366235

[6] Cortese, G., Holmboe, S. A and Scheike, T. H. (2017) Regression models for the restricted residual mean life for right-censored and left-truncated data. *Statist. Med.*, 36, 1803–1822. MR3648623

[7] Heath, M. T. (2005) *Scientific Computing: An Introductory Survey*. McGraw-Hill.

[8] Horowitz, J. L. (2009) *Semiparametric and Nonparametric Methods in Econometrics*. Springer. MR2535631

[9] Hosmer, D. W., Lemeshow, S. and May, S. (2008) *Applied Survival Analysis: Regression Modeling of Time to Event Data*. John Wiley and Sons Inc. MR1674644

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

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

[12] Huang, C. Y. and Qin, J. (2013) Semiparametric estimation for the additive hazards model with left-truncated and right-censored data. *Biometrika*, 100, 877–888. MR3142338

[13] Huang, C. Y., Ning, J. and Qin, J. (2015) Semiparametric likelihood inference for left-truncated and right-censored data. *Biostatistics*, 16, 785–798. MR3449843

[14] Jiang, J. and Zhou, H. (2007) Additive hazard regression with auxiliary covariates. *Biometrika*, 94, 359–369. MR2380565

[15] Kalbfleisch, J. D. (1978) Likelihood methods and nonparametric tests. *J. Am. Statist. Assoc*, 73, 167–170. MR0518600

[16] Kalbfleisch, J. D. and Lawless, J. F. (1991) Regression models for right truncated data with applications to AIDS incubation times and reporting lags. *Statist. Sinica*, 1, 19–32.

[17] Kalbfleisch, J. D. and Prentice, R. L. (2002) *The Statistical Analysis of Failure Time Data*. Wiley. MR0570114

[18] Lancaster, T. (1990) *The Econometric Analysis of Transition Data*. Cambridge. MR1167199

[19] Lawless, J. F. (2003) *Statistical Models and Methods for Lifetime Data*. Wiley. MR1940115

[20] Liang, K. Y. and Qin, J. (2000) Regression analysis under non-standard situations: a pairwise pseudolikelihood approach. *J. R. Statist. Soc., Series B*, 62, 773–786. MR1796291

[21] Lin, D. Y. and Wei, L. J. (1989) The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, 84, 1074–1078. MR1134495

[22] Lin, D. Y. and Ying, Z. (1994) Semiparametric analysis of the additive risk model. *Biometrika*, 81, 61–71. MR1279656

[23] Lin, D. Y. and Ying, Z. (1995) Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The Annals of Statistics*, 23, 1712–1734. MR1370304

[24] Owen, A. B. (2001) *Empirical Likelihood*. CRC Press.

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

[26] Shen, P. S. (2009) Semiparametric analysis of survival data with left-truncation and right-censoring. *Computational Statistics and Data Analysis*, 53, 4417–4432. MR2744335

[27] Shen, P. S. (2014) Aalen's additive risk model for left-truncated and right-censored data. *Communications in Statistics: Simulation and Computation*, 43, 1006–1019. MR3215763

[28] Su, Y. R. and Wang, J. L. (2012) Modeling left-truncated and right-censored survival data with longitudinal covariate. *The Annals of Statistics*, 40, 1465–1488. MR3015032

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

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

[31] Wang, M. C. (1989) A semiparametric model for randomly truncated data. *J. Am. Statist. Assoc*, 84, 742–748. MR1132590

[32] Wang, M. C. (1991a) Nonparametric estimation from cross-sectional survival data. *J. Am. Statist. Assoc*, 86, 130–43. MR1137104

[33] Wang, M. C. (1991b) Hazards regression analysis for length-biased data. *Biometrika*, 83, 343–354. MR1439788

[34] Wang, M. C., Brookmeyer, R. and Jewell, N.P. (1993) Statistical models for prevalent cohort data. *Biometrics*, 49, 1–11. MR1221402

Li-Pang Chen
Department of Statistics and Actuarial Science
University of Waterloo,
200 University Ave W, Waterloo, ON N2L 3G1
E-mail address: L358CHEN@uwaterloo.ca