東海大學統計研究所

碩士論文

Cox-Aalen模型下左截右設限資料之分析

The Cox-Aalen Model for Left-truncated and Right-censored Data

指導教授:沈葆聖　　博士

研究生:翁梨寧

中華民國一百零六年六月

The Cox-Aalen Model for Left-truncated and Right-censored Data

Advisor: Pao-sheng Shen

Student: Li Ning Weng

Department of Statistics

Tunghai University

Taichung, Taiwan 40704

June 2017

東海大學碩士班研究生

論文口試委員審定書

統計學系碩士班翁梨寧君所提之論文

The Cox-Aalen Model for Left-truncated and Right-censored Data

---

經本委員會審議，認為符合碩士資格標準。

論文口試委員召集人 戴政（簽章）

委員 [illegible]

林正祥

中華民國 106 年 06 月 21 日

謝誌

首先我要感謝我的父母，辛苦工作將我們拉拔長大，讓我們在求學過程無後顧之憂，並接受我的負面情緒，一路上陪伴、支持著我;沒有你們，就沒今日的我，謝謝你們。

再來，我要感謝我的指導教授–沈葆聖老師，感謝你這一年來的指導，在這短短一年，從老師身上學到不止有專業知識，還有老師對做學問的熱忱與處事的態度，在往後的人生裡，我會謹記老師那份克服萬難的精神，在面臨困境時絕不輕言放棄。

此外，我要感謝 戴政博士及 林正祥博士在百忙之中抽空來擔任我的口試委員，口試對談過程給了我許多啟發讓我獲益良多，並給予我論文上的指導與建議，讓我的論文能夠更加完善。

感謝系上每一位老師與助教，這些年來的教學與幫助，讓我得以在這五年內有所成長。也感謝研究室的同學、學長，讓我研究所生涯多彩多姿，也謝謝你們的包容及照顧。

最後我要感謝這一路上曾經幫助過我、包容我的人，沒有你們我沒有辦法走到這裡，謝謝你們。

學生 翁梨寧 謹識于

東海大學統計研究所

中華民國 106 年 7 月

Abstract

We analyze left-truncated and right-censored (LTRC) data using an additive-multiplicative Cox-Aalen model proposed by Scheike and Zhang (2002), that extends the Cox regression model as well as the additive Aalen model. Based on the conditional likelihood function, we derive the weighted least squared (WLS) estimators for the regression parameters and cumulative intensity functions of the model. The estimators are shown to be consistent and asymptotically normal. A simulation study is conducted to investigate the performance of the proposed estimators.

Key Words: Left truncation, Aalen model, Cox regression, survival analysis, time-varying effect.

Contents

1 Introduction

Left-truncated and right-censored (LTRC) data often arise in epidemiology and individual follow-up studies (see Wang, 1991). Left truncation is a biased sampling plan as subjects with shorter survival times tend to be excluded from the sample such that individuals with larger survival times are oversampled. An example is the common use of prevalent cohort study, where a group of diseased individuals are recruited for a prospective study. The main target of a research project is to study the natural history of the disease for individuals who developed the disease during the calendar time period (τ_0, τ), $\tau_0 < \tau$. Under a prevalent cohort design, also refereed as cross-sectional sampling, individuals who have experienced a first event (e.g. diagnosed as having chronic diseases or HIV infection) between τ_0 and τ and have not experienced a second event (e.g. death or AIDS) are recruited at the time τ for a prospective follow-up study. Suppose that the initial time of the first event, denoted by T_s, can be quite accurately determined, such as HIV infection resulting from blood transfusion. Let T denote the time from T_s to the second event's endpoint. Let V denote the time from T_s to τ. Thus, left truncation occurs since those individuals who have experienced the second event, i.e., $T \leq V$, are not included in the study. Suppose that the follow-up study is terminated at τ^* ($\tau^* > \tau$). Let $C_1 = V + \tau^* - \tau$ denote the time from the first event to the end of study and C_2 denote the time from the first event to drop-out. Thus, T is left-truncated by V and right-censored by $C = \min(C_1, C_2)$. Figure 1 highlights all the different times for LTRC data as described above.

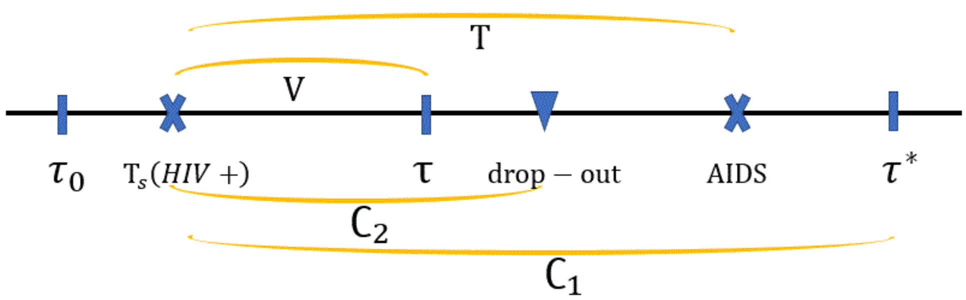

Figure 1: Schematic depiction of LTRC data

The proportional hazards regression model (Cox (1972)) specifies that the hazard function takes the form

$$\lambda(t|Z(t)) = \lambda(t)\exp(Z(t)^T\beta)$$

, where $Z(t)$ is a $p \times 1$ vector of covariates, β is a $p \times 1$ vector of unknown regression parameters and $\lambda(t)$ as an arbitrary baseline hazard function. The model may be extended by allowing the regression coefficients to be time-varying:

$$\lambda(t|Z(t)) = \lambda(t)\exp(Z(t)^T\beta(t))$$

(see, e.g. Murphy and Sen (1991), Huang (1999), Martinussen et al. (2001)). One problem with the extended Cox models is that the choice of a smoothing-parameter is needed for estimating the non-parametric terms. An alternative to proportional hazard models is the Aalen additive hazard model (Aalen (1980,1989); McKeague (1988); Huffer and McKeague (1991)), in which it is assumed that

$$\lambda(t|W(t)) = W(t)^T\alpha(t)$$

, where $W(t)$ is a $q\times1$ vector of covariates and $\alpha(t)$ is a $q\times1$ vector of time-varying regression parameters. One major advantage of the Aalen additive models is that time-varying effects are easy to estimate, and that no smoothing parameter needs to be chosen. Proportional and Additive hazard models postulate a different relationship between the hazard function and covariates and the two models can be used to complement each other. Scheike and Zhang (2002) proposed a new model, called the Cox-Aalen model, that combines the multiplicative and additive model. The hazard density function of the Cox-Aalen model is given as

$$\lambda(t|W(t), Z(t)) = W(t)^T\alpha(t)\exp(Z(t)^T\beta). \tag{1.1}$$

Under model (1.1), covariates $W(t)$ work additively on the risk and have non-parametric time-varying effects while covariates $Z(t)$ have multiplicative effect. The Cox-Aalen model provides a flexible class of models, which extends the Cox model by allowing the baseline intensity to depend on covariates through the additive Aalen model. For right-censored data, Scheike and Zhang (2002) proposed approximate maximum likelihood estimators of the baseline intensity functions and the relative risk parameters of the Cox model and established the large sample properties of the estimators. Scheike and Zhang (2003) showed how the Cox-Aalen model can lead to simple formulae for predicted probabilities and their standard errors. Kraus (2004) studied goodness-of-fit tests for the Cox-Aalen model based on the stratified martingale residual process.

When truncation is present, Pan and Chappell (2002) considered the non-parametric maximum likelihood estimate (NPMLE) of the regression coefficient for the Cox proportional hazards model with LTRC data. Shen (2014) analyzed LTRC data using Aalen's additive hazard models. Shen (2016) demonstrated Gandy and Jensen (2005)'s goodness-of-fit tests for Aalen's model can be extended to LTRC data and doubly censored data. In Section 2, based on the conditional likelihood function, we derive the weighted least squared (WLS) estimators for the regression parameters and cumulative intensity functions of model (1.1) with LTRC data. The proposed estimators are shown to be consistent and asymptotically normal. In Section 3, a simulation study is conducted to investigate the performance of the proposed estimators.

2 The Proposed Estimators

We assume that given $W(t)$ and $Z(t)$, T and (V, C) are independent of each other but V and C are dependent with $P(C \geq V) = 1$. For LTRC data, one can observe nothing if $T < V$ and observe $(X, V, \delta, W(t), Z(t))$, with $\delta = I_{[T<C]}$ and $X = \min(T, C)$, if $T \geq V$. Suppose that the left and right endpoints of T are independent of $W(t)$ and $Z(t)$. Let F, Q and G denote the cumulative distribution functions of T, C and V, respectively. Let a_F and b_F denote the left and right endpoints of T, and similarly, define (a_Q, b_Q) and (a_G, b_G) as the left and right endpoint of C, and V, respectively. Throughout this article, for identifiabilities of F, we assume that $a_G = a_F = a_Q = 0, b_G \leq \min(b_F, b_Q)$.

Let $(X_i, V_i, \delta_i, W_i(t), Z_i(t))$ $(i = 1, \ldots, n)$ be the observed truncated sample.

Let $Y_i(t) = I_{[V_i \leq t \leq X_i]}$ and $N_i(t) = I_{[X_i \leq t, \delta_i = 1]}$.

Let $\mathcal{F}(t)$ denote the complete σ-field generated by

$$\{V_i, W_i(x), Z_i(x), Y_i(x), I_{[V_i \leq X_i]}, \delta_i I_{[V_i < X_i \leq t]}, I_{[V_i < X_i \leq x]}, x \leq t; i = 1, \ldots, n\}.$$

Let

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) W_i(s)^T \alpha(s) \exp(Z_i(s)^T \beta) ds.$$

Since $E[dN_i(t)|\mathcal{F}(t-)] = Y_i(t) W_i(t)^T \alpha(t) \exp(Z_i(t)^T \beta)$, $M_i(t)$ is a martingale process with respect to $\mathcal{F}(t)$.

Let τ_c be some constant such that $\tau_c < b_F$. In practice, the value of τ_c is set at the largest values of X_i's with $\delta = 1$. Consider the counting process on $[0, \tau_c]$. This gives the following likelihood:

$$L = \prod_{i=1}^{n} \left\{ dF(X_i|W_i, Z_i) dG(V_i)[1 - Q(X_i|V_i)]/p_i \right\}^{\delta_i} \times \prod_{i=1}^{n} \left\{ dQ(X_i|V_i) dG(V_i)[1 - F(X_i|W_i, Z_i)]/p_i \right\}^{1-\delta_i}.$$

We decompose L into three factors yielding

$$L = \prod_{i=1}^{n} \left\{ \frac{[dF(X_i|W_i, Z_i)]^{\delta_i}[1 - F(X_i|W_i, Z_i)]^{1-\delta_i}}{1 - F(V_i - |W_i, Z_i)} \right\} \prod_{i=1}^{n} \left\{ \frac{dG(V_i)[1 - F(V_i - |W_i, Z_i)]}{p_i} \right\}$$

$$\times \prod_{i=1}^{n} \left\{ [1 - Q(X_i|V_i)]^{\delta_i}[dQ(X_i|V_i)]^{1-\delta_i} \right\} = L_1 L_2 L_3,$$

where L_1, L_2 and L_3 represent the likelihoods in the first, second and third braces. We do not consider the maximization of L_3 since the estimator of L_3 does not involve $F(\cdot|Z_i)$. Thus, the likelihood is proportional to

$$L(\beta, A) = \prod_{i=1}^{n} p_i^{-1} \prod_{t \le \tau_c} Y_i(t) \big[\big(W_i(t)^T dA(t) \exp(Z_i(t)^T \beta) \big)^{dN_i(t)} \big]$$

$$\times \prod_{i=1}^{n} p_i^{-1} \exp\left\{ - \int_0^{\tau_c} I_{[X_i \ge s]} W_i(s)^T \alpha(s) \exp(Z_i(s)^T \beta) ds \right\},$$

where $A(t) = \int_0^t \alpha(s)ds$, $p_i = \int_{a_G}^{b_G} \exp\{-\Lambda(v|W_i(\cdot), Z_i(\cdot))\} dG(v)$, $\Lambda(v|W_i(\cdot), Z_i(\cdot)) = \int_0^v I_{[X_i \ge s]} W_i(s)^T \alpha(s) \exp(Z_i(s)^T \beta) ds$.

We wish to estimate the cumulative intensity function $A(t)$ as well as the true value of the relative risk parameter β. Notice that $L(\beta, A)$ can be factorized as $L(\beta, A) = L_m(\beta, A) \times L_c(\beta, A)$, where

$$L_m(\beta, A) = \prod_{i=1}^{n} p_i^{-1} \exp\left\{ - \int_0^{V_i} I_{[X_i \ge s]} W_i(s)^T \alpha(s) \exp(Z_i(s)^T \beta) ds \right\}$$

and

$$L_c(\beta, A) = \prod_{i=1}^{n} \prod_{t \le \tau_c} Y_i(t) \big[\big(W_i(t)^T dA(t) \exp(Z_i(t)^T \beta) \big)^{dN_i(t)} \big]$$

$$\times \exp\left\{ - \int_0^{\tau_c} Y_i(t) \exp(Z_i(t)^T \beta) W_i(t)^T dA(t) \right\},$$

where $L_m(\beta, A)$ and $L_c(\beta, A)$ are the marginal likelihood for V_i and conditional likelihood for X_i given V_i.

The following Lemma shows that when $W_i(t)$ and $Z_i(t)$ are discrete and independent of time it suffices to maximize L_c.

Lemma 1. When $W_i(t)$ and $Z_i(t)$ are discrete and independent of time, maximizing $L_c(\beta, A)$ is equivalent to maximizing the full likelihood $L(\beta, A)$.

Remark 1: *Notice that more efficient estimators can be obtained by maximizing the full likelihood L. However, one difficulty arises in this approach since the p_i in L_m involves the unknown distribution function $G(x)$. Further research is required in this issue. In this article, we consider the conditional maximum likelihood estimators based on L_c although Lemma 1 holds only for the special case when covariates are discrete and independent of time.*

The logarithm of $L_c(\beta, A)$ can be written as

$$l_c(\beta, A) = \sum_i \int_0^{\tau_c} \log\big(Y_i(t) W_i(t)^T dA(t) \exp(Z_i(t)^T \beta)\big) dN_i(t)$$

$$- \sum_i \int_0^{\tau_c} Y_i(t) \exp(Z_i(t)^T \beta) W_i(t)^T dA(t),$$

Let $N(t) = (N_1(t), \ldots, N_n(t))^T$ be an n-dimensional counting process,

$M(t) = (M_1(t), \ldots, M_n(t))^T$ be an n-dimensional martingale.

Define a matrice $Y(\beta, t) = (Y_1(t) \exp(Z_1(t)^T \beta) W_1(t) , \ldots, Y_n(t) \exp(Z_n(t)^T \beta) W_n(t))^T$ and let $\tilde{Z}(t) = (Z_1(t), \ldots, Z_n(t))^T$

Now, we solve the joint score equations for β and α. Given $A(t)$, taking derivatives of the log-likelihood $l_c(\beta, A)$ with respect to β gives the score equation for β

$$U(\beta|A(t)) = \sum_i \int_0^{\tau_c} Y_i(t) Z_i(t) dN_i(t) - \sum_i \int_0^{\tau_c} Y_i(t) Z_i(t) \exp(Z_i(t)^T \beta) W_i(t) dA(t).$$

The score equation for α is given by

$$U_\alpha(A(t)) = Y^T(\beta, t) \text{diag}(1/\lambda_i(t))(dN(t) - Y(\beta, t) dA(t)),$$

where $\lambda_i(t; \beta) = W_i(t)^T \alpha(t) \exp(Z_i(t)^T \beta)$.

Given β, solving for $U_\alpha(A(t)) = 0$ yields the Aalen-Huffer-McKeague estimator $\hat{A}(t,\beta)$

$$\hat{A}(t,\beta) = \int_0^t Y^-(\beta, s; D(s;\beta))dN(s),$$

where

$$Y^-(\beta, s; D(s;\beta)) = [Y(\beta,s)^T D(s;\beta)Y(\beta,s)]^{-1}Y(\beta,s)^T D(s;\beta),$$

$D(s;\beta) = \mathrm{diag}(d_i(s;\beta))$ is a diagonal matrix with elements $d_i(s;\beta) = Y_i(s)/\lambda_i(s;\beta)$. Similar to the argument of Sasieni (1992), we may interpret the estimator $\hat{A}(t,\beta)$ as a conditional maximum likelihood estimator for $A(t)$, which is equivalent to WLS.

To solve the score equations simultaneously we insert the Aalen-Huffer-McKeague estimator $\hat{A}(t,\beta)$ into the score for β and get

$U(\beta, \tau_c|D)$

$= \sum_i \int_0^{\tau_c} Y_i(t)Z_i(t)dN_i(t) - \sum_i \int_0^{\tau_c} Y_i(t)Z_i(t)\exp(Z_i(t)^T\beta)W_i(t)Y^-(\beta, t; D(t))dN(t)$

$= \int_0^{\tau_c}(\tilde{Z}^T(t) - S^{(1)}(\beta,t)Y^-(\beta,t;D(t)))dN(t)$

$= \int_0^{\tau_c}(\tilde{Z}^T(t) - \tilde{Z}^T(t)Y(\beta,t)Y^-(\beta,t;D(t)))dN(t),$

where $S^{(k)}(\beta,t) = \sum_i Z_i^{\otimes k}(t)Y_i(t)\exp(Z_i(t)^T\beta)W_i(t)^T$, $z^{\otimes 0} = 1$, $z^{\otimes 1} = z$ and $z^{\otimes 2} = zz^T$. The process $U(\beta_0, t|D)$ is a martingale in t since the compensator of $U(\beta_0, t|D)$ is 0 for any weight matrix $D(t)$. Thus, we can consider the estimating equation for all the choices of $D(t)$. Given D, we define $\hat{\beta}$ as the solution to the score equation $U(\beta, \tau_c|D) = 0$. Following Scheike and Zhang (2002), we consider weights of the form $d_i(s) = d_i(s;\beta) = Y_i(t)\exp(-Z_i(t)^T\beta)/h_i(t)$. Based on the estimator $\hat{\beta}$ we can estimate the cumulative intensity function $A(t)$ by solving

$$\hat{A}(t,\hat{\beta}) = \int_0^t Y^-(\hat{\beta}, s; D(s;\hat{\beta}))dN(s), \tag{2.1}$$

where $D(s;\hat{\beta}) = \mathrm{diag}(d_i(s;\hat{\beta}))$ is a diagonal matrix with elements

$$d_i(s;\hat{\beta}(D)) = Y_i(s)\exp(-Z_i(s)^T\hat{\beta})/h_i(s). \text{ where } h_i(s) = W_i(s)^T\alpha(s).$$

One simple choice of $h_i(t)$ is $h_i(t) = 1$. Using $h_i(t) = 1$, we can solve $U(\beta, \tau_c|D) = 0$ for obtaining an initial estimator $\tilde{\beta}_n$. Based on $\tilde{\beta}_n$, we obtain an estimator for A, denoted by $\tilde{A}_n(t, \tilde{\beta}_n)$. Notice that no iteration is needed when $h_i(t) = 1$. We call $\tilde{\beta}_n$ and $\tilde{A}_n(t, \tilde{\beta}_n)$ the ordinary least squared (OLS) estimators. The other choice is to consider the conditional maximum likelihood weights $h_i(t) = W_i(t)^T\alpha(t)$. Under this weight, the score for β reduces to the score for partial likelihood in the case of Cox model with LTRC data, which is asymptotically efficient. Using $h_i(t) = 1$, we can obtain an initial estimator for $A(t)$, denoted by $\hat{A}_n^{(0)}(t) = \tilde{A}_n(t, \tilde{\beta}_n)$. Based on $\hat{A}^{(0)}(t)$, we obtain a kernel-smoothed estimator of $\alpha_j(t)$ by

$$\hat{\alpha}_j^{(0)}(t) = \int_{a_F}^{b_F} \frac{1}{h_n} K\left(\frac{t-u}{h_n}\right) d\hat{A}_j^{(0)}(u, \tilde{\beta}_n), \; j = 1, \ldots, q \tag{2.2}$$

where $K(\cdot)$ is a left-continuous function on $(0, 1]$ such that $\int_{(0,1]} K(u)du = 1$,

h_n is a positive bandwidth parameter that tends to 0 as $n \to \infty$.

In the second step, the estimators $\hat{\alpha}_j^{(0)}(t)$ $(j = 1, \ldots, q)$ are used to estimate the weight function $\hat{h}_i^{(0)}(t) = W_i(t)^T\hat{\alpha}^{(0)}(t)$ to obtain a update estimator for β, $\hat{\beta}_n^{(1)}$. Based on (2.1), we obtain a update estimator for $A(t)$, $\hat{A}_n(t, \hat{\beta}_n^{(1)})$. Iterate between solving score function $U(\beta, \tau_c|D) = 0$, and (2.1), (2.2) until convergence. Let $\hat{\beta}_n$ and $\hat{A}_n(t, \hat{\beta}_n)$ denote the converged estimators. We call them the weighted least squared (WLS) estimators with the conditional maximum likelihood (cMLE) weights.

Next, we derive the asymptotic properties of the OLS and WLS estimators. We denote the true value of β and $A(t)$ as β_0 and $A_0(t)$, respectively.

Let $\tilde{W}_i(\beta, t) = Y_i(t)\exp(Z_i(t)^T\beta)W_i(t)$ and

$$S^{(k)(j)}(\beta, t) = \sum_i Z_i^{\otimes k}(t)Y_i(t)d_i^j(t)\tilde{W}_i^{\otimes j}(\beta, t)\exp(Z_i(t)^T\beta)W_i(t)^T.$$

for $k+j \leq 2$, and defined as for $k+j = 2$ with the convention that $S^{(k)} = S^{(k)(0)}$, and with an additional transpose for $S^{(1)(1)}$ such that the dimensions match. We need the following conditions:

(C1) $\int_0^{\tau_c} \alpha(s)ds < \infty$.

(C2) There exists a compact neighbourhood $\mathcal{B}$ of β_0, and functions $s^{(k)(j)}$, for $k + j \leq 2$ defined on $\mathcal{B} \times [0, \tau_c]$ such that

$$\sup_{\beta \in \mathcal{B}, t \in [0, \tau_c]} ||n^{-1} S^{(k)(j)}(\beta, t) - s^{(k)(j)}(\beta, t)|| \xrightarrow{p} 0$$

(C3) $s^{(k)(j)}(\beta, t)$, $k + j \leq 2$ are uniformly continuous functions of $\beta \in \mathcal{B}$ and $t \in [0.\tau_c]$ and bounded on $\mathcal{B} \times [0, \tau_c]$. Let

$$\phi(\beta) = \int_0^{\tau_c} \{s^{(1)}(\beta_0, u) - s^{(1)}(\beta, u)(s^{(0)(1)}(\beta, u))^{-1} q^{(0)}(\beta, u)\}\alpha(u)du,$$

where $q^{(0)}(\beta, u)$ is the limiting distribution of $n^{-1}[Y(\beta, u)^T D(u) Y(\beta, u)]$. Assume that there exists a root of $\phi(\beta) = 0$ on $\mathcal{B}$, and β_0 is the only root of $\phi(\beta) = 0$. Assume that the following matrix is positive definite:

$$\Sigma_I = \int_0^{\tau_c} \{s^{(2)}(\beta_0, u) - s^{(1)}(\beta_0, u)(s^{(0)(1)}(\beta_0, u))^{-1}(s^{(1)(1)}(\beta_0, u)\alpha(u))\}du$$

(C4) With probability one, both $W(\cdot)$ and $Z(\cdot)$ have bounded total variation in $[0, \tau_c]$

We denote $\hat{\beta}$ and $\hat{A}(\hat{\beta}, t)$ as the solutions based on some weight matrix D.

Theorem 1. Under conditions (C1)-(C4), it follows that $n^{1/2}(\hat{\beta} - \beta_0)$ converges towards a normally distributed variable with mean 0 and a variance that may be estimated consistently by $\hat{\Sigma}_\beta = n\mathcal{I}^{-1}(\hat{\beta}, \tau_c)[\hat{U}(\hat{\beta}_n, \cdot)](\tau_c)\mathcal{I}^{-1}(\hat{\beta}, \tau_c)$, where

$$[\hat{U}(\hat{\beta}, \cdot)](\tau_c) = \int_0^{\tau_c} (Z^T(t) - S^{(1)}(\hat{\beta}, t)Y^-(\beta_0, t)) \times \mathrm{diag}(dN(t))(Z^T(t) - S^{(1)}(\hat{\beta}, t)Y^-(\hat{\beta}, t))^T$$

is the optional variation process of $U(\beta_0, \tau_c)$ with β_0 replaced by $\hat{\beta}$.

Remark 2: *For the cMLE weights the variance simplifies and $\mathcal{I}^{-1}(\hat{\beta}_n, \tau_c)$ estimate the variance of $(\hat{\beta}_n - \beta_0)$, where $\mathcal{I}(\beta, t)$ is the derivative of $U(\beta, t)$ and given as*

$$\begin{aligned}
\mathcal{I}(\beta, t) &= -\tfrac{\partial}{\partial \beta} U(\beta, t) \\
&= \textstyle\int_0^{\tau_c} S^{(2)}(\beta, t) Y^{-}(\beta, t; D(t)) dN(t) + \int_0^{\tau_c} S^{(1)}(\beta, t) \tfrac{\partial}{\partial \beta} Y^{-}(\beta, t; D(t)) dN(t) \\
&= \textstyle\int_0^{\tau_c} \tilde{Z}^T(t) \mathrm{diag}(Y(\beta, t) Y^{-}(\beta, t; D(t)) dN(t) \tilde{Z}(t) \\
&\quad - \textstyle\int_0^{\tau_c} \tilde{Z}^T(t) Y(\beta, t) Y^{-}(\beta, t; D(t)) \mathrm{diag}(Y(\beta, t) Y^{-}(\beta, t; D(t)) dN(t)) \tilde{Z}(t).
\end{aligned}$$

Next, we investigate the asymptotic properties of the estimator $\hat{A}(\hat{\beta}, t)$.

Theorem 2. Under conditions (C1)-(C4), it follows that $n^{1/2}(\hat{A}(\hat{\beta}, t) - A(t))$ converges in distribution towards a Gaussian process with variance that be estimated consistently by

$$\hat{\Sigma}_{\hat{A}} = n(n^{-1} H(\hat{\beta}, t)^T \hat{\Sigma}_{\hat{\beta}} H(\hat{\beta}, t) + [M_A(\cdot)](t) + H(\hat{\beta}, t)^T \mathcal{I}^{-1}(\hat{\beta}, t)[U(\cdot, M_A(\cdot)](t)$$

$$+ [U(\cdot, M_A(\cdot)](t) \mathcal{I}^{-1}(\hat{\beta}, \tau_c) H(\hat{\beta}, t),$$

where $H(\hat{\beta}, t) = \frac{\partial}{\partial \hat{\beta}} \int_0^t Y^{-}(\beta, s) dN(s)$ and $M_A(t) = \int_0^t Y^{-}(\hat{\beta}, s) dM(s)$.

3 Simulation Studies

We generate T based on the hazard density function

$$\lambda(t|W(t), Z(t)) = (0.5 + W_1 t)\exp(Z_1 0.3 + Z_2(-0.3))$$

,where W_1 is is a discrete uniform random variable on the integers $1, 2, \dots, 10$ and independent of Z_1 and Z_2 that are independent standard normals. The left-truncation variable V is generated from exponential distribution with mean μ_v equal to 0.17, 0.25 and 0.33 such that the proportion of truncation rate is equal to 0.25, 0.45 and 0.65, respectively. Right censoring variable C is generated from $V + d_0$, where d_0 is chosen as 0.7 and 0.45 such that the proportion of censoring $P(\delta_i = 0)$ is equal to 0.2 and 0.4, respectively. Sample size is $n = 200, 400$ and the replication time is 1000. The value of τ_c is set at the largest values of X_i's with $\delta_i = 1$. For each simulated dataset, we obtain the ordinary least squared (OLS) estimators $\tilde{\beta}_n = (\tilde{\beta}_{1n}, \tilde{\beta}_{2n})^T$, $\tilde{A}_n(t, \tilde{\beta}_n) = (\tilde{A}_{1n}(t, \tilde{\beta}_n), \tilde{A}_{2n}(t, \tilde{\beta}_n))^T$ and the weighted least squared (WLS) estimators with the cMLE weights $\hat{\beta}_n = (\hat{\beta}_{1n}, \hat{\beta}_{2n})^T$, $\hat{A}_n(t, \hat{\beta}_n) = (\hat{A}_{1n}(t, \hat{\beta}_n), \hat{A}_{2n}(t, \hat{\beta}_n))^T$. The weights are obtained based on Epanechnikov kernel with bandwidth $h = 0.3, 0.35, 0.4, 0.45, 0.5$ and $h = 0.2, 0.2, 0.25, 0.3, 0.35$ for $n = 200$ and $n = 400$, respectively. Using $\Sigma_{\tilde{\beta}_n}$ and $\Sigma_{\tilde{A}_n}$, we calculated the estimated standard deviations of $\tilde{\beta}_n$ and $\tilde{A}_n$. Similarly, we calculated the estimated standard deviations of $\hat{\beta}_n$ and $\hat{A}_n$ using $\Sigma_{\hat{\beta}_n}$ and $\Sigma_{\hat{A}_n}$. Approximate 0.95 confidence intervals for β and $A(t)$ are constructed using the normal approximation. Table 1 shows the simulated biases, simulated standard deviations (std), estimated standard deviations (estd), empirical coverage (cov) of $\tilde{\beta}_n$ and $\hat{\beta}_n$ and the ratio (denoted by ratio) of the root mean squared error (rmse) of $\tilde{\beta}_{in}$ to that of $\hat{\beta}_{in}$. Table 1 also shows the proportion of left-truncation (denoted by q) and right-censoring (denoted by $p_c = P(\delta_i = 0)$). Table 2 shows the simulation results for $\tilde{A}_1 n(t, \tilde{\beta}_n)$ and $\hat{A}_2 n(t, \hat{\beta}_n)$ at some selected points.

Table 1 Simulated biases and std. of $\hat{\beta}_n$ and $\tilde{\beta}_n$

					$\tilde{\beta}_{1n}$				$\hat{\beta}_{1n}$		
p_c	q	n	bias	std	estd	cov	bias	std	estd	cov	ratio
0.2	0.25	200	-0.001	0.085	0.086	0.95	-0.001	0.084	0.086	0.95	1.02
0.2	0.25	400	0.006	0.055	0.059	0.95	0.006	0.055	0.059	0.95	1.00
0.2	0.45	200	0.007	0.087	0.086	0.93	0.006	0.086	0.086	0.93	1.02
0.2	0.45	400	0.004	0.065	0.060	0.94	0.003	0.064	0.059	0.94	1.01
0.2	0.65	200	0.017	0.085	0.087	0.95	0.015	0.084	0.087	0.95	1.02
0.2	0.65	400	0.008	0.062	0.061	0.95	0.007	0.062	0.061	0.95	1.01
0.4	0.45	200	0.017	0.093	0.099	0.97	0.017	0.094	0.099	0.97	1.00
0.4	0.45	400	0.012	0.070	0.067	0.94	0.012	0.070	0.067	0.94	1.00
0.4	0.65	200	0.012	0.103	0.098	0.94	0.012	0.102	0.097	0.94	1.00
0.4	0.65	400	0.011	0.069	0.067	0.95	0.011	0.069	0.067	0.95	1.01

					$\tilde{\beta}_{2n}$				$\hat{\beta}_{2n}$		
p_c	q	n	bias	std	estd	cov	bias	std	estd	cov	ratio
0.2	0.25	200	0.003	0.082	0.085	0.95	0.003	0.081	0.085	0.95	1.01
0.2	0.25	400	-0.001	0.067	0.059	0.93	-0.001	0.067	0.059	0.93	1.00
0.2	0.45	200	-0.014	0.088	0.087	0.94	-0.013	0.087	0.087	0.94	1.01
0.2	0.45	400	-0.005	0.062	0.060	0.93	-0.005	0.062	0.059	0.93	1.00
0.2	0.65	200	-0.015	0.083	0.088	0.95	-0.014	0.082	0.087	0.95	1.01
0.2	0.65	400	-0.001	0.058	0.060	0.98	-0.001	0.058	0.060	0.98	1.00
0.4	0.45	200	0.007	0.109	0.099	0.92	0.007	0.108	0.100	0.92	1.00
0.4	0.45	400	-0.003	0.070	0.068	0.96	-0.003	0.070	0.068	0.96	1.01
0.4	0.65	200	-0.006	0.100	0.100	0.95	-0.005	0.099	0.099	0.95	1.02
0.4	0.65	400	-0.007	0.070	0.068	0.93	-0.007	0.070	0.068	0.93	1.00

Table 2. Simulated biases and std. of $\tilde{A}_1n(t,\tilde{\beta}_n)$ and $\hat{A}_2n(t,\hat{\beta}_n)$

					$\tilde{A}_{1n}(t,\tilde{\beta}_n)$				$\hat{A}_{1n}(t,\hat{\beta}_n)$			
p_c	q	t	n	bias	std	estd	cov	bias	std	estd	cov	ratio
0.2	0.25	0.3	200	-0.004	0.129	0.142	0.95	-0.006	0.118	0.127	0.90	1.09
0.2	0.25	0.6	200	-0.019	0.208	0.204	0.94	-0.021	0.190	0.181	0.91	1.09
0.2	0.25	0.9	200	-0.031	0.364	0.358	0.96	-0.019	0.314	0.309	0.93	1.16
0.2	0.25	0.3	400	0.008	0.139	0.146	0.97	0.007	0.108	0.149	0.93	1.28
0.2	0.25	0.6	400	0.004	0.178	0.180	0.97	0.008	0.140	0.176	0.97	1.27
0.2	0.25	0.9	400	-0.005	0.269	0.276	0.95	0.010	0.232	0.253	0.94	1.16
0.2	0.45	0.3	200	0.006	0.143	0.137	0.92	0.002	0.138	0.129	0.85	1.04
0.2	0.45	0.6	200	-0.003	0.202	0.202	0.94	-0.007	0.189	0.185	0.93	1.07
0.2	0.45	0.9	200	-0.005	0.336	0.323	0.96	-0.006	0.309	0.309	0.92	1.09
0.2	0.45	0.3	400	-0.008	0.102	0.099	0.93	-0.009	0.093	0.092	0.88	1.10
0.2	0.45	0.6	400	-0.009	0.146	0.144	0.95	-0.012	0.134	0.130	0.93	1.09
0.2	0.45	0.9	400	-0.011	0.240	0.230	0.93	0.001	0.211	0.202	0.93	1.14
0.2	0.65	0.3	200	-0.020	0.147	0.149	0.94	-0.018	0.132	0.142	0.83	1.11
0.2	0.65	0.6	200	-0.011	0.215	0.218	0.95	-0.012	0.197	0.207	0.91	1.09
0.2	0.65	0.9	200	0.000	0.339	0.324	0.96	0.005	0.289	0.292	0.94	1.17
0.2	0.65	0.3	400	0.004	0.127	0.126	0.92	-0.001	0.118	0.116	0.87	1.07
0.2	0.65	0.6	400	-0.014	0.169	0.168	0.94	-0.012	0.150	0.152	0.90	1.12
0.2	0.65	0.9	400	-0.010	0.250	0.234	0.94	-0.008	0.211	0.207	0.94	1.19
0.4	0.45	0.3	200	0.002	0.131	0.131	0.93	-0.001	0.122	0.123	0.91	1.08
0.4	0.45	0.6	200	-0.011	0.214	0.214	0.94	-0.003	0.199	0.330	0.92	1.08
0.4	0.45	0.9	200	0.036	0.472	0.430	0.95	0.024	0.448	0.477	0.91	1.05
0.4	0.45	0.3	400	-0.004	0.104	0.099	0.92	-0.008	0.097	0.093	0.88	1.07
0.4	0.45	0.6	400	-0.003	0.157	0.152	0.94	-0.008	0.145	0.138	0.94	1.08
0.4	0.45	0.9	400	0.017	0.282	0.292	0.96	0.009	0.270	0.260	0.93	1.05
0.4	0.65	0.3	200	-0.011	0.163	0.160	0.91	-0.015	0.153	0.146	0.86	1.06
0.4	0.65	0.6	200	-0.012	0.228	0.230	0.96	-0.022	0.210	0.207	0.93	1.08
0.4	0.65	0.9	200	-0.058	0.379	0.375	0.96	-0.043	0.320	0.327	0.93	1.19
0.4	0.65	0.3	400	-0.009	0.102	0.110	0.94	-0.013	0.098	0.102	0.92	1.04
0.4	0.65	0.6	400	-0.016	0.157	0.161	0.95	-0.017	0.145	0.146	0.93	1.09
0.4	0.65	0.9	400	-0.019	0.276	0.269	0.95	-0.033	0.255	0.234	0.89	1.08

Table 2. Simulated biases and std. of $\tilde{A}_1 n(t, \tilde{\beta}_n)$ and $\hat{A}_2 n(t, \hat{\beta}_n)$ (Continued)

				$\tilde{A}_{2n}(t, \tilde{\beta}_n)$				$\hat{A}_{2n}(t, \hat{\beta}_n)$				
p_c	q	t	n	bias	std	estd	cov	bias	std	estd	cov	ratio
0.2	0.25	0.3	200	0.002	0.026	0.029	0.99	0.002	0.023	0.026	0.97	1.13
0.2	0.25	0.6	200	0.008	0.050	0.047	0.96	0.008	0.045	0.042	0.93	1.10
0.2	0.25	0.9	200	0.011	0.109	0.110	0.95	0.008	0.103	0.098	0.95	1.06
0.2	0.25	0.3	400	-0.002	0.022	0.023	0.96	-0.002	0.017	0.024	0.94	1.29
0.2	0.25	0.6	400	0.000	0.033	0.035	0.97	-0.001	0.027	0.034	0.99	1.24
0.2	0.25	0.9	400	0.007	0.073	0.080	0.96	0.003	0.066	0.073	0.96	1.11
0.2	0.45	0.3	200	-0.002	0.024	0.025	0.96	-0.001	0.024	0.024	0.92	1.02
0.2	0.45	0.6	200	-0.002	0.046	0.044	0.94	-0.001	0.042	0.041	0.95	1.08
0.2	0.45	0.9	200	0.004	0.105	0.097	0.92	0.003	0.094	0.092	0.93	1.11
0.2	0.45	0.3	400	0.002	0.020	0.019	0.95	0.002	0.018	0.017	0.93	1.09
0.2	0.45	0.6	400	0.003	0.035	0.032	0.94	0.003	0.032	0.029	0.92	1.07
0.2	0.45	0.9	400	0.006	0.077	0.069	0.93	0.002	0.069	0.061	0.92	1.12
0.2	0.65	0.3	200	0.001	0.029	0.028	0.96	0.001	0.026	0.027	0.93	1.09
0.2	0.65	0.6	200	-0.001	0.044	0.048	0.98	-0.001	0.042	0.046	0.96	1.05
0.2	0.65	0.9	200	-0.001	0.100	0.094	0.94	-0.002	0.087	0.085	0.96	1.15
0.2	0.65	0.3	400	0.000	0.021	0.022	0.97	0.001	0.020	0.021	0.96	1.07
0.2	0.65	0.6	400	0.003	0.035	0.036	0.94	0.002	0.031	0.033	0.97	1.15
0.2	0.65	0.9	400	0.003	0.071	0.065	0.96	0.002	0.061	0.059	0.96	1.17
0.4	0.45	0.3	200	0.000	0.024	0.025	0.93	0.001	0.024	0.023	0.94	1.02
0.4	0.45	0.6	200	0.007	0.051	0.050	0.94	0.005	0.049	0.080	0.94	1.04
0.4	0.45	0.9	200	-0.004	0.150	0.138	0.91	0.001	0.147	0.151	0.93	1.02
0.4	0.45	0.3	400	-0.001	0.019	0.019	0.94	0.000	0.018	0.017	0.95	1.05
0.4	0.45	0.6	400	-0.002	0.035	0.034	0.95	-0.001	0.032	0.031	0.96	1.10
0.4	0.45	0.9	400	-0.013	0.082	0.090	0.94	-0.011	0.079	0.082	0.95	1.04
0.4	0.65	0.3	200	-0.001	0.029	0.030	0.96	0.000	0.028	0.027	0.93	1.04
0.4	0.65	0.6	200	-0.003	0.049	0.051	0.95	-0.001	0.046	0.046	0.94	1.06
0.4	0.65	0.9	200	0.008	0.126	0.114	0.92	0.002	0.108	0.101	0.94	1.17
0.4	0.65	0.3	400	0.000	0.019	0.020	0.95	0.001	0.019	0.019	0.95	1.00
0.4	0.65	0.6	400	0.001	0.036	0.036	0.94	0.001	0.033	0.032	0.95	1.07
0.4	0.65	0.9	400	0.003	0.082	0.081	0.94	0.008	0.078	0.073	0.93	1.05

Based on the results of Table 1 and 2, we have the following conclusions:

(i) The standard deviations of all the estimators increase as the proportion of left-truncation q or right censoring (p_c) increase. The standard deviations of $\hat{\beta}_n$ and $\hat{A}_n(t, \hat{\beta}_n)$ are smaller than that of $\tilde{\beta}_n$ and $\tilde{A}_n(t)$ for all the cases considered. In term of rmse, $\hat{\beta}_n$ and $\hat{A}_n$ outperform $\tilde{\beta}_n$ and $\tilde{A}_n$. The ratio of root mean squared error of $\tilde{\beta}_{in}$ to that of $\hat{\beta}_{in}$ ranges from 1.00 to 1.02. The ratio of root mean squared error of $\tilde{A}_{in}$ to that of $\hat{A}_{in}$ ranges from 1.00 to 1.29.

(ii) When $n = 200$, the estimated standard deviation underestimates the empirical standard deviation, resulting in less-than-nominal coverage of confidence intervals. However, when $n = 400$, the estimated standard deviation is close to the empirical standard deviation and the coverages of 95% confidence intervals based on the estimated standard deviations are close to nominal levels.

4 Application

To illustrate the proposed method, we consider the data of 103 heart transplant patients taken from Kalbfleisch and Prentice (2002, pages 387-389). According to the description of Crowley and Hu (1977), the patients agreed to participate in the Stanford program after a medical conference where it was decided that they were unlikely to respond to the other therapies. This data consist of 103 observations, 69 of whom received a transplant and from them 24 were still alive at the end of study. Although survival times were recorded for all the patients, the other covariates except age were not recorded for those who did not receive a transplant. Thus, to explore the relationship between survival time and the other covariates, such as mismatch scores, we can only use the truncated data consisting of 69 patients who received a transplant to fit the Cox-Aalen model. The proportional part of the model contains number of mismatches (Z_1), HLA-A2 antigen indicator variable (Z_2, presence=1) and mismatch scores (Z_3). The age of patients (W_1) may be seen as an additional cause of death at early stages and therefore seemed natural to include in the additive part of the model. Due to small sample size, the estimates are computed using the weights with $h_i(t) = 1$, i.e. the OLS estimator. Table 3 lists the estimated parameters $\tilde{\beta}_n$ for Z_1, Z_2 and Z_3 and the estimated parameters $\tilde{A}_n(t, \tilde{\beta}_n)$ for some selected quartile points. For the proportional part, the mismatch score is significant (p-value=0.043) and give a log-relative-risk increase at 0.684 per one unit increase in mismatch score. Figure 2 shows the cumulative additive effects of baseline and age estimate with 95% pointwise confidence bands. Both effects are clearly insignificant.

Table 3. The estimated parameters $\tilde{\beta}_n$ and $\tilde{A}_n$

the estimated parameters		OLS (p-value)	
$\tilde{\beta}_{1n}$(number of mismatches)		-0.134 (0.247)	
$\tilde{\beta}_{2n}$(HLA-A2 antigen)		0.047 (0.457)	
$\tilde{\beta}_{3n}$(mismatch scores)		0.684 (0.043)	
$\tilde{A}_{1n}(38, \tilde{\beta}_n)$	$\tilde{A}_{2n}(38, \tilde{\beta}_n)$	0.212 (0.300)	-0.002 (0.386)
$\tilde{A}_{1n}(65, \tilde{\beta}_n)$	$\tilde{A}_{2n}(65, \tilde{\beta}_n)$	0.202 (0.306)	0.000 (0.485)
$\tilde{A}_{1n}(77, \tilde{\beta}_n)$	$\tilde{A}_{2n}(77, \tilde{\beta}_n)$	0.026 (0.474)	0.005 (0.258)
$\tilde{A}_{1n}(109, \tilde{\beta}_n)$	$\tilde{A}_{2n}(109, \tilde{\beta}_n)$	-0.266 (0.241)	0.014 (0.033)
$\tilde{A}_{1n}(206, \tilde{\beta}_n)$	$\tilde{A}_{2n}(206, \tilde{\beta}_n)$	-0.377 (0.156)	0.018 (0.010)
$\tilde{A}_{1n}(339, \tilde{\beta}_n)$	$\tilde{A}_{2n}(339, \tilde{\beta}_n)$	-0.090 (0.435)	0.014 (0.104)
$\tilde{A}_{1n}(514, \tilde{\beta}_n)$	$\tilde{A}_{2n}(514, \tilde{\beta}_n)$	-0.117 (0.415)	0.015 (0.085)
$\tilde{A}_{1n}(732, \tilde{\beta}_n)$	$\tilde{A}_{2n}(732, \tilde{\beta}_n)$	-0.966 (0.077)	0.038 (0.007)
$\tilde{A}_{1n}(1031, \tilde{\beta}_n)$	$\tilde{A}_{2n}(1031, \tilde{\beta}_n)$	-1.201 (0.055)	0.049 (0.002)
$\tilde{A}_{1n}(1799, \tilde{\beta}_n)$	$\tilde{A}_{2n}(1799, \tilde{\beta}_n)$	-2.067 (0.047)	0.071 (0.013)

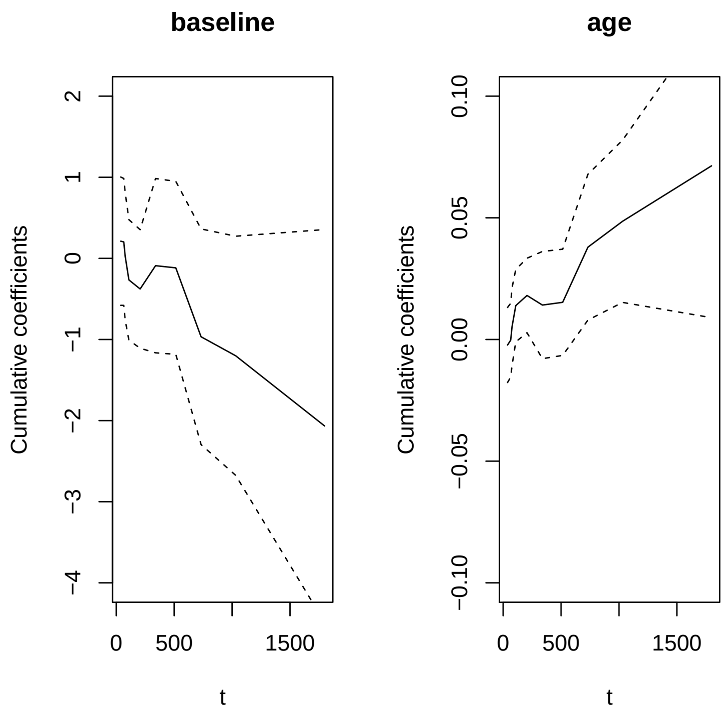

Figure 2: Cumulative risk for additive part of model

5 Conclusion

Under the Cox-Aalen model with LTRC data, we have derived the estimators of regression coefficients and cumulative intensities using the conditional likelihood approach. Simulation results indicate that although the WLS estimator is superior to the OLS estimator, we encountered only moderate gain in efficiency. Bandwidth selection appears to have some room for improvement. Further research is required in the area. To check the validity of the model assumption, one can use the procedure based on the the stratified martingale residual process proposed by Kraus (2004). In some situation, the distribution of truncation variables $G(x)$ can be parameterized as $G(x; \theta)$. A more efficient estimator can be developed by incorporating the available information on the distribution function of left truncation variable V.

References

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting process. In:Klonecki, W., Kozek, A., Rosiski, J. (Eds). Lecture Notes on Mathematical Statistics and Probability, Vol. 2. Springer, New York, pp. 1-25.

Aalen, O. O. (1989). A linear regression model for analysis of life times. *Statistics in Medicine*, **8**, 907-925.

Cox, D. (1972). Regression models and life tables (with Discussion). *Journal of Royal Statistical Society Series* **B**, **34**, 187-220.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *The Annals of Statistics*, **27**, 1536-1563.

Huffer, F. W. and Mckeague, F. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, **86**, 114-129.

Kraus, D. (2004). Goodness-of-fit inference for the Cox-Aalen additive-multiplicative regression model. *Statistics and Probability Letters*, **70**, 285-298.

Martinussen, T., Scheike, T. H. and Skovgaard, I. (2001). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics*, **29**, 59-76.

McKeagues, I. W. (1988). Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. *in Statistical Inference from Stochastic Processes*, 139-152.

Murphy, S. A. and Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Process and their Applications*, **39**, 153-180.

Pan, W. and Chappell, R. (2002) Estimation in the Cox proportional hazards model with left-truncated and interval-censored data. *Biometrics*, **58**, 64-70.

Sasieni, P. D. (1992). Information bounds for the additive and multiplicative intensity models. In: *Survival Analysis of the art*, (J.P. Klein and P.K. Goel, eds.), 249-265, Kluwer.

Scheike, T. H. and Zhang, M.-J. (2002). An additive-multiplicative Cox–Aalen regression model. *Scandinavian Journal of Statistics*, **29**, 75-88.

Scheike, T. H. and Zhang, M.-J. (2003). Extensions and applications of the Cox-Aalen survival model. *Biometrics*, **59**, 1036-1045.

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.

Shen, P.-S. (2016). On goodness-of-fit tests for Aalen's additive model based on left-truncated right-censored or doubly censored data. *Communications in Statistics: Theory and Methods*, **45**, 4812-4823.

Wang, M.-C.(1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*, *86*, 130-143.