J Syst Sci Complex (2014) 27: 1247–1262

CONTINUOUS AUXILIARY COVARIATE IN ADDITIVE HAZARDS REGRESSION FOR SURVIVAL DATA*

SHI Xiaoping · LIU Yanyan · WU Yuanshan

DOI: 10.1007/s11424-014-3010-3
Received: 16 January 2013 / Revised: 15 April 2013

Abstract This paper considers the additive hazards regression analysis by utilizing continuous auxiliary covariate information to improve the efficiency of the statistical inference when the primary covariate is ascertained only for a randomly selected subsample. The authors construct a martingale-based estimating equation for the regression parameter and establish the asymptotic consistency and normality of the resultant estimators. Simulation study shows that the proposed method can greatly improve the efficiency compared with the estimator which discards the auxiliary covariate information in a variety of settings. A real example is also provided as an illustration.

Keywords Additive hazards regression, continuous auxiliary covariate, estimating equation, kernel smoothing, survival analysis.

1 Introduction

Due to financial limitation or technical difficulty, it is often expensive to measure the primary exposure variable in many biomedical studies. In other words, the primary exposure variable may only be measured precisely in a subset of study cohort. This subset is often referred to as the validation set. Discarding the information of the subjects with missing values would result in efficiency loss for regression parameters. One useful accommodating approach is to measure some auxiliary covariate for primary exposure variable on all subjects, while conducting ascertainments on the primary exposure variable only for a randomly selected subsample. Usually,

SHI Xiaoping
School of Mathematics and System Sciences, Xinjiang University, Urumqi 830046*, China; School of Mathematics and Statistics, Wuhan University, Wuhan* 430072*, China.* Email: shixiaopingcn@gmail.com.
LIU Yanyan · WU Yuanshan
School of Mathematics and Statistics, Wuhan University, Wuhan 430072, *China.*
Email: liuyy@whu.edu.cn; shan@whu.edu.cn.
*This research was supported by the National Natural Science Foundation of China under Grant Nos. 11171263, 41261087, and the Doctoral Fund of Ministry of Education of China under Grant Nos. 20110141110004, 20110141120004.
◇ *This paper was recommended for publication by Editor SUN Liuquan.*

the auxiliary information is cheap and easy to obtain. Consequently, a natural and important question is that how to take use of the auxiliary information to improve the statistical inference. Some methods have been developed for this issue in the Cox proportional hazards model[1]. For example, the seminal work of Prentice[2] introduced a partial likelihood estimator based on the induced relative hazards function under "rare disease" assumption about censored failure time data. Pepe and Fleming[3] proposed an implemented method which is nonparametric with respect to the mismeasurement process. Lin and Ying[4] provided a general solution to survival data with missing cavariate. Zhou and Pepe[5] proposed an estimated partial likelihood method for discrete auxiliary covariate by using both validation and non-validation observations to enhance efficiency. Zhou and Wang[6] extended the nonparametric inference procedure of Zhou and Pepe[5] to handle a continuous auxiliary covariate. Greene and Cai[7] proposed using the SIMEX method to handle measurement errors for multivariate failure time data when a validation set is not available. Liu, Wu, and Zhou[8] and Liu, Zhou, and Cai[9] proposed the statistical inference procedures for multivariate survival data by utilizing auxiliary information for the discrete and continuous auxiliary covariates, respectively.

It can be seen that there is extensive literature on auxiliary covariate information under the framework of the Cox proportional hazards model. As an important complement for Cox proportional hazards model, the additive hazards model is also widely used in practice. See for example Breslow and Day[10, 11], Cox and Oakes[12], Thomas[13], and Lin and Ying[14]. Furthermore, O'Neill[15] has shown that use of the Cox proportional hazards model can result in serious bias when the additive hazards model is correct. Some researches on how to further utilize auxiliary information has been conducted under the framework of the additive hazards model. For example, Kulich and Lin[16] proposed a method based on correcting the pseudo-score function for the additive hazards model, which produces asymptotically unbiased estimation. However, the corrected pseudo-score method requires the conditional moments of surrogate covariate given the true covariate to be correctly specified. Jiang and Zhou[17] proposed an updated pseudo-score method, which relaxed the moment conditions and thus avoided the possibility of modelling miss-specifications. However, their updated pseudo-score method involves extensive computation.

In this paper, we consider the additive hazards model with continuous auxiliary covariate. We propose a method by implementing the missing items in estimating equations with its kernel smoothing estimators based on auxiliary information and then obtained an estimated estimating equation for the regression parameter. Our method does not need to specify the form of baseline hazard function. The auxiliary covariate could be mismeasured surrogate to the true covariate, or any covariate that is informative about the true covariate. Moreover, it is known that surrogate variable could be considered as the auxiliary covariate but the visa versa is not true. The proposed method is nonparametric with respect to the condition distribution of the primary covariate given the auxiliary covariate and can be applied to allow the rescue of the incomplete data and implemented easily in practice.

The rest of this paper is organized as follows. In Section 2, we propose an estimated estimating equation method for the additive hazards model by using auxiliary information. In Section 3, we establish the large sample properties of the resultant estimators. We conduct simulation studies to evaluate the finite sample performance of the proposed method in Section 4. The proposed method is illustrated with application to a real data set in Section 5. Some concluding remarks are provided in Section 6.

2 Inference Procedure

Suppose that there is a random sample of n independent subjects from an underlying population. Let $\widetilde{T}_i$ and C_i denote the potential failure time and censoring time for the ith subject, respectively. Due to censoring, we always observe $T_i = \min\{\widetilde{T}_i, C_i\}$. Let $\Delta_i = I(\widetilde{T}_i \le C_i)$ be the failure indicator. Let $W_i = (X_i^{\mathrm{T}}, Z_i^{\mathrm{T}})^{\mathrm{T}}$ denote a set of covariates which could be time-dependent, where X_i is the primary exposure subject to missing and $Z_i = (Z_{i1}, Z_{i2}, \cdots, Z_{ip})^{\mathrm{T}}$ is the remaining covariate vector which is observed completely. Assume that the hazard function of $\widetilde{T}_i$ associated with W_i takes the additive form:

$$\lambda(t|W_i) = \lambda_0(t) + \beta^{\mathrm{T}} X_i(t) + \gamma^{\mathrm{T}} Z_i(t), \tag{1}$$

where $\theta = (\beta^{\mathrm{T}}, \gamma^{\mathrm{T}})^{\mathrm{T}}$ is the regression parameter to be estimated and $\lambda_0(t)$ is an unspecified baseline hazard function.

We use the indicator variable η_i to indicate whether or not the ith subject has the primary covariate X_i precisely ascertained and assume that the missing mechanism for X_i is referred as missing completely at random in the sense of Rubin[18]. Denote $V = \{i\colon \eta_i = 1\}$ and $\overline{V} = \{i\colon \eta_i = 0\}$ as the validation and nonvalidation sets, respectively. Note that the primary covariate X_i is only observed in the validation set. Thus, the conventional method can be conducted based on the validation set (see [14]). However, this method could be suffered from the loss of efficiency. Let A_i denote the auxiliary covariate that is related and surrogate to the primary covariate X_i. The auxiliary covariate is ascertained for all subjects under study. Assume that, given X_i, A_i provides no additional information to regression model in the sense that $\lambda(t|W_i, A_i) = \lambda(t|W_i)$ for all $t \ge 0$. Then, the observed data structure is $\{T_i, \Delta_i, Z_i, X_i, A_i\}$ if $i \in V$ and otherwise $\{T_i, \Delta_i, Z_i, A_i\}$. We aim to provide the inference procedure by utilizing the auxiliary covariate information to improve the study efficiency.

If the ith subject belongs to the validation set V, then Z_i and X_i are observed and the hazards function of $\widetilde{T}_i$ takes the form as (1). Otherwise, using the argument of Prentice[2], Zhou and Pepe[5], and Zhou and Wang[6], it can be verified that the hazard function for $\lambda(t|Z_i, A_i)$ satisfies the induced hazards regression model as follows:

$$\begin{aligned}\lambda(t|Z_i, A_i) &\equiv \lim_{\Delta t \downarrow 0}\left[\frac{1}{\Delta t}\Pr\{t \le \widetilde{T}_i < t + \Delta t | \widetilde{T}_i \ge t, Z_i(t), A_i(t)\}\right]\\ &= \lambda_0(t) + \gamma^{\mathrm{T}} Z_i(t) + \beta^{\mathrm{T}} E\{X_i(t)|\widetilde{T}_i \ge t, Z_i(t), A_i(t)\}.\end{aligned}$$

Under the independent censoring assumption, conditioning on W_i, $\widetilde{T}_i$ and C_i are independent. Furthermore, we can rewrite the induced model as

$$\lambda(t|Z_i, A_i) = \lambda_0(t) + \gamma^{\mathrm{T}} Z_i(t) + \beta^{\mathrm{T}} E\{X_i(t)|Y_i(t) = 1, Z_i(t), A_i(t)\},$$

where $Y_i(t) = I(T_i \geq t)$ is the at-risk process. For notational simplicity, there exists A_i^* such that $P\{X_i(t) \leq x|T_i \geq t, Z_i(t), A_i(t)\} = P\{X_i(t) \leq x|T_i \geq t, A_i^*(t)\}$ for any $x \in R$. Thus, we further have

$$\lambda(t|Z_i, A_i) = \lambda_0(t) + \gamma^{\mathrm{T}} Z_i(t) + \beta^{\mathrm{T}} E\{X_i(t)|Y_i(t) = 1, A_i^*(t)\}. \tag{2}$$

Obviously, the induced hazard model (2) still maintains the structure of the additive hazards regression. Denote $W_i^*(t) = W_i(t)\eta_i + [E\{X_i(t)|Y_i(t) = 1, A_i^*(t)\}^{\mathrm{T}}, Z_i(t)^{\mathrm{T}}]^{\mathrm{T}}(1 - \eta_i)$ and the counting process for the ith subject by $N_i(t) = \Delta_i I(T_i \leq t)$. Based on (1) and (2), the process

$$M_i(t) \equiv N_i(t) - \int_0^t Y_i(u)\{d\Lambda_0(u) + \theta^{\mathrm{T}} W_i^*(u)du\}$$

is a local square integrable zero-mean martingale at the true parameter $\theta_0 = (\beta_0^{\mathrm{T}}, \gamma_0^{\mathrm{T}})^{\mathrm{T}}$(see [19]), where $\Lambda_0(t) = \int_0^t \lambda(u)du$ is the unknown cumulative baseline hazards function. Consequently, it is natural to obtain the Breslow-Aalen type estimator (see [20, 21]) for $\Lambda_0(t)$ with given θ:

$$\widetilde{\Lambda}_0(t;\theta) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u)\theta^{\mathrm{T}} W_i^*(u)du\}}{\sum_{j=1}^n Y_j(u)}.$$

The parameter θ can be estimated from the following estimating equation

$$U(\theta) \equiv \sum_{i=1}^n \int_0^\tau W_i^*(t)\{dN_i(t) - Y_i(t)d\widetilde{\Lambda}_0(t;\theta) - Y_i(t)\theta^{\mathrm{T}} W_i^*(t)dt\} = 0,$$

or equivalently,

$$U(\theta) = \sum_{i=1}^n \int_0^\tau \{W_i^*(t) - E(t)\}\{dN_i(t) - Y_i(t)\theta^{\mathrm{T}} W_i^*(t)dt\} = 0,$$

where $E(t) = \sum_{i=1}^n Y_i(t)W_i^*(t)/\{\sum_{i=1}^n Y_i(t)\}$ and τ is the end time of study.

Since $U(\theta)$ involves the unknown conditional expectation except the regression parameter, in what follows we first seek an estimate for the conditional expectation and then construct an estimated estimating function for $U(\theta)$. Assume that A_i^* is a d-vector continuous auxiliary covariate. If the ith subject lies in the nonvalidation set $\overline{V}$, we can estimate the conditional expectation $E\{X_i(t)|Y_i(t) = 1, A_i^*(t)\}$ by using the method of Nadaraya[22] and Watson[23] as follows:

$$\widehat{X}_i(t) \equiv \frac{\sum_{j\in V} Y_j(t)K\{B^{-1}(A_j^* - A_i^*)\}X_j(t)}{\sum_{j\in V} Y_j(t)K\{B^{-1}(A_j^* - A_i^*)\}},$$

where $K(\cdot)$ is a kernel function with bandwidth matrix B, which is $d \times d$ positive-definite, with its element possibly depending on n. For simplicity, we only consider the situation in which B is a diagonal matrix with element at (l,l) denoted by b_l. Then, replacing the unknown conditional expectation by its estimated counterpart $\widehat{X}_i(t)$ in $U(\theta)$, we obtained an estimated estimating equation, which is given by

$$\widehat{U}(\theta) \equiv \sum_{i=1}^{n} \int_0^\tau \{\widehat{W}_i(t) - \widehat{E}(t)\}\{dN_i(t) - Y_i(t)\theta^{\mathrm{T}}\widehat{W}_i(t)dt\} = 0,$$

where $\widehat{W}_i(t) = W_i(t)\eta_i + [\widehat{X}_i(t)^{\mathrm{T}}, Z_i(t)^{\mathrm{T}}]^{\mathrm{T}}(1-\eta_i)$ and $\widehat{E}(t) = \frac{\sum_{i=1}^n Y_i(t)\widehat{W}_i(t)}{\sum_{i=1}^n Y_i(t)}$. The proposed estimator, denoted by $\widehat{\theta}_E$, which solves $\widehat{U}(\theta) = 0$, can be explicitly obtained as follows,

$$\widehat{\theta}_E = \left[\sum_{i=1}^{n} \int_0^\tau Y_i(t)\{\widehat{W}_i(t) - \widehat{E}(t)\}^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^{n} \int_0^\tau \{\widehat{W}_i(t) - \widehat{E}(t)\} dN_i(t)\right],$$

where $a^{\otimes 2} = aa^{\mathrm{T}}$ for a column vector a. Furthermore, $\Lambda_0(t)$ can be estimated by $\widehat{\Lambda}_0(t;\widehat{\theta}_E)$, where

$$\widehat{\Lambda}_0(t;\theta) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u)\theta^{\mathrm{T}}\widehat{W}_i(u)du\}}{\sum_{j=1}^n Y_j(u)}.$$

Note that $\widehat{\Lambda}_0(t;\widehat{\theta}_E)$ may not be monotonicity in t. Define $\widehat{\Lambda}_0^*(t,\widehat{\theta}_E) = \max_{s\le t} \widehat{\Lambda}_0(s,\widehat{\theta}_E)$ to ensure the monotonicity. Using the similar arguments in [14], we have that $\widehat{\Lambda}_0^*(t,\widehat{\theta}_E) - \widehat{\Lambda}_0(t;\widehat{\theta}_E) = o_p(n^{-1/2})$ uniformly over t.

3 Asymptotic Properties

For simplicity, we introduce some notation. For a vector a, define $a^{\otimes 0} = 1, a^{\otimes 1} = a, ||a|| = \sup_i |a_i|$. For a matrix $D = (d_{ij})$, define $||D|| = \sup_{i,j} |d_{ij}|$. For $k = 0, 1$, define $\widehat{S}^{(k)}(t) = n^{-1}\sum_{i=1}^n Y_i(t)\widehat{W}_i^{\otimes k}(t)$, $S^{(k)}(t) = n^{-1}\sum_{i=1}^n Y_i(t)W_i^{*\otimes k}(t)$, $s^{(k)}(t) = E\{Y(t)W^{*\otimes k}(t)\}$, $e(t) = s^{(1)}(t)/s^{(0)}(t)$. For given t, let $F(y,z)$ be the joint distribution of $(Y(t), A^*(t))$ and $\mathcal{A}$ be the domain of $A^*(t)$. Denote $f_{A^*}(z) = \partial F(1,z)/\partial z$ and $\rho = \lim_{n\to\infty} v/n$, where v is the cardinality of the validation set V. For $j \in V$, let

$$Q_j(\theta) = \int_0^\tau \left[E\{W_j(t)|Y_j(t)=1, A_j^*\} - e(t)\right] Y_j(t)[W_j(t) - E\{W_j(t)|Y_j(t)=1, A_j^*\}]^{\mathrm{T}}\theta dt,$$

$$Q_j^{\overline{v}}(\theta) = \frac{1}{\overline{v}} \sum_{i\in\overline{V}} \int_0^\tau |B|^{-1}\{W_i^*(t) - e(t)\} \frac{Y_i(t)Y_j(t)K\{B^{-1}(A_j^* - A_i^*)\}}{f_{A^*}(a)} \{W_j(t) - W_i^*(t)\}^{\mathrm{T}}\theta dt$$

for any $a \in \mathcal{A}$, where $\overline{v} = n - v$.

We impose the following conditions through our derivations:

C1. $\int_0^\tau \lambda_0(t)dt < \infty$.

C2. $P\{Y(t)=1|A^* = a\} > 0$ for any $a \in \mathcal{A}$.

C3. $E\left\{\sup_{t\in[0,\tau]} ||Y(t)W^{*\otimes k}(t)||\right\} < \infty$ for $k = 0, 1$.

C4. Let $K(\cdot)$ be a multivariate symmetric kernel function, which is non-negative and uniformly bounded with a compact bounded support satisfying that

$$\int K(u)du = 1 \qquad \text{and} \qquad \int K^2(u)du < \infty.$$

Furthermore, the kernel function $K(\cdot)$ has order α_0 in the sense that

$$\alpha_0 \equiv \inf\left\{\alpha > d\colon \int_{R^d} u^\alpha K(u)du \neq 0\right\},$$

where $u^\alpha = u_1^{\alpha_1}u_2^{\alpha_2}\cdots u_d^{\alpha_d}$, $|\alpha| = \alpha_1+\alpha_2+\cdots+\alpha_d$, $u = (u_1, u_2, \cdots, u_d)$, $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_d)$, and α_l is non-negative integer. The bandwidth B satisfies that $||B|| \to 0$, $\sqrt{n}||B||^{\alpha_0} \to 0$, $\frac{\log n}{\sqrt{n}||B||^d} \to 0$.

C5. For given t, let $F(y, z)$ be the joint distribution of $(Y(t), A^*(t))$, $G(y, z, x)$ be the joint distribution of $(Y(t), A^*(t), X(t))$ and $H(y, z, w)$ be the joint distribution of $(Y(t), A^*(t), W(t))$. Suppose that $f_{A^*}(z) = \frac{\partial F(1,z)}{\partial z}$, $g(z, x) = \frac{\partial^2 G(1,z,x)}{\partial z\partial x}$ and $h(z, w) = \frac{\partial^2 H(1,z,w)}{\partial z\partial w}$ have the α_0th continuous derivation with respect to every component of z.

C6. $\sup_{t\in[0,\tau]} ||W^v(t)|| = O_P(1)$, where

$$W^v(t) = \sqrt{v}\left\{|B|^{-1}\frac{1}{v}\sum_{j\in V} Y_j(t)K\{B^{-1}(A_j^* - a)\}W_j(t) - f_{A^*}(a)E[W(t)|Y(t), A^* = a]\right\}$$

for any $a \in \mathcal{A}$.

Remark 3.1 Conditions C1–C3 are standard assumptions in survival analysis. Condition C4 is on the kernel function. Specifically, we can choose the bandwidth $B = 2\widehat{\sigma}_A n^{-1/3}$ for $d = 1$ and $\alpha_0 = 2$, where $\widehat{\sigma}_A$ is the sample standard deviation of A. Condition C5 is a technical assumption for proving. Condition C6 is imposed to simplify the derivations of the asymptotic properties, which can be satisfied if the class of functions $\{|B|^{-1}Y_j(t)K\{B^{-1}(A_j^* - a)\}W_j(t)\colon a \in \mathcal{A}, j = 1, 2, \cdots, n; t \in [0, \tau]\}$ is Donsker.

In the following, we use notation $\to_p$, $\to_{\text{a.s.}}$, and $\to_d$ to denote the convergence in probability, convergence in probability 1 and convergence in distribution, respectively. Denote the determinant of matrix B by $|B|$. Let $b = (b_1, b_2, \cdots, b_d)'$ and $\alpha_0^* = (\alpha_{01}^*, \alpha_{02}^*, \cdots, \alpha_{0d}^*)'$ such that each α_{0l}^* is non-negative integer and $|\alpha_0^*| \equiv \sum_{l=1}^d \alpha_{0l}^* = \alpha_0$. For d-vector $u = (u_1, u_2, \cdots, u_d)$ and $v = (v_1, v_2, \cdots, v_d)$, define $u^v = \prod_{l=1}^d u_l^{v_l}$ and $uv = (u_1v_1, u_2v_2, \cdots, u_dv_d)$. To derive our theorems, we need the following lemma.

Lemma 3.2 *Under conditions* C4–C5*,*

$$\sup_{t\in[0,\tau]} ||X_i(t) - \widehat{X}_i(t)|| \longrightarrow_p 0 \qquad for \qquad i \in \overline{V}.$$

Proof Let $\mathcal{A}$ be the domain of the process $A^*(t)$ for $t \in [0, \tau]$. Let P_n be the empirical measure from the n i.i.d. observations and P be the corresponding probability measure. For fixed $a \in \mathcal{A}$, let

$$L_n(t, a) = P_n\{|B|^{-1}K\{B^{-1}(A^* - a)\}Y(t)X(t)\}.$$

Note that the stochastic process $A^*(t), Y(t), X(t)$ have bounded total variation over $t \in [0, \tau]$. It follows from Lemma 9.10 in [24] that they are VC-subgraph with finite VC-index. Using the similar argument used in [25], thus, we obtain that

$$\sup_{t\in[0,\tau],a\in\mathcal{A}} ||L_n(t,a) - P\{|B|^{-1}K\{B^{-1}(A^* - a)\}Y(t)X(t)\}|| = O_p\left(\frac{\log n}{\sqrt{n}||B||^d}\right).$$

Furthermore, under condition C5 and using the Taylor expansion, we have

$$\begin{aligned}
& P\{|B|^{-1}K\{B^{-1}(A^* - a)\}Y(t)X(t)\} \\
&= \int (b_1 b_2 \cdots b_d)^{-1} K\{B^{-1}(z-a)\} g(z,x) x dz dx \\
&= \int K(u) g(bu + a, x) x du dx \\
&= \int K(u) \left[\sum_{|\alpha|=0}^{\alpha_0 - 1} \left\{ \frac{\partial^\alpha g(a,x)}{\partial u^\alpha} \frac{1}{\alpha!} b^\alpha u^\alpha \right\} + \frac{\partial^{\alpha_0^*} g(a^*,x)}{\partial u^{\alpha_0^*}} \frac{1}{\alpha_0 !} b^{\alpha_0^*} u^{\alpha_0^*} \right] x du dx \\
&= \int g(a,x) x dx + O(b^{\alpha_0^*}) \\
&= f_{A^*}(a) \int \frac{g(a,x)}{f_{A^*}(a)} x dx + O(||B||^{\alpha_0}) \\
&= f_{A^*}(a) E\left\{X(t)|Y(t)=1, A^* = a\right\} + O(||B||^{\alpha_0}),
\end{aligned}$$

where a^* is the line segment of a and bu.

Therefore, we can conclude that by condition C4,

$$\sup_{t\in[0,\tau],a\in\mathcal{A}} |L_n(t,a) - f_{A^*}(a)E\{X(t)|Y(t)=1, A^*=a\}| \longrightarrow_p 0.$$

Denote $Q_n(t,a) = P_n\{|B|^{-1}K\{B^{-1}(A^* - a)\}Y(t)\}$. Similarly, we can conclude

$$\sup_{t\in[0,\tau],a\in\mathcal{A}} |Q_n(t,a) - f_{A^*}(a)| \longrightarrow_p 0.$$

Consequently, it follows straightforwardly that

$$\sup_{t\in[0,\tau],a\in\mathcal{A}} \left| \frac{L_n(t,a)}{Q_n(t,a)} - E\{X(t)|Y(t)=1, A^*=a\} \right| \longrightarrow_p 0.$$

Furthermore, note that

$$\widehat{X}_i(t) = \frac{|B|^{-1}v^{-1}\sum_{j\in V} Y_j(t)K\{B^{-1}(A_j^* - A_i^*)\}X_j(t)}{|B|^{-1}v^{-1}\sum_{j\in V} Y_j(t)K\{B^{-1}(A_j^* - A_i^*)\}},$$

then we can conclude

$$\sup_{t\in[0,\tau]} ||X_i(t) - \widehat{X}_i(t)|| \to_p 0, \quad \text{for} \quad i \in \overline{V}.$$

Thus, the proof of Lemma 3.2 has been done. ∎

Furthermore, we can conclude that

$$\sup_{t\in[0,\tau]} ||\widehat{W}_i(t) - W_i^*(t)|| \to_p 0, \quad \text{for} \quad i \in \overline{V}. \tag{3}$$

According to the definition of $\widehat{S}^{(k)}(t)$, $S^{(k)}(t)$, and $s^{(k)}$, we also have that

$$\sup_{t\in[0,\tau]} ||\widehat{S}^{(k)}(t) - S^{(k)}(t)|| \to_p 0, \quad \text{for} \quad k = 0, 1.$$

On the other hand, it follows from the uniformly strong law of large numbers that

$$\sup_{t\in[0,\tau]} ||S^{(k)}(t) - s^{(k)}(t)|| \to_{\text{a.s.}} 0, \quad \text{for} \quad k = 0, 1.$$

Immediately, we have that

$$\sup_{t\in[0,\tau]} ||\widehat{S}^{(k)}(t) - s^{(k)}(t)|| \to_p 0, \quad \text{for} \quad k = 0, 1,$$

and by Slutsky Theorem, we obtain

$$\sup_{t\in[0,\tau]} ||\widehat{E}(t) - e(t)|| \to_p 0. \tag{4}$$

Analogously, using the method of Taylor expansion to $f_{A^*}(z)$ and $h(z, w)$, under condition C4 and definition of $W_i^*(t)$ for $i \in \overline{V}$, we have that

$$v^{-1} \sum_{j\in V} |B|^{-1} Y_j(t) K\{B^{-1}(A_j^* - a)\} \to_p f_{A^*}(a) \tag{5}$$

for all $a \in \mathcal{A}$ and

$$v^{-1} \sum_{j\in V} |B|^{-1} Y_j(t) K\{B^{-1}(A_j^* - a)\} W_i^*(t) \to_p f_{A^*}(a) E\{W(t)|Y(t) = 1, A^* = a\} \tag{6}$$

for any $a \in \mathcal{A}$ and $i \in \overline{V}$.

Furthermore, under the definition of $\widehat{W}_i(t)$ and conditions C3 and C6, by (5) and (6), we obtain

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^n \int_0^\tau \{W_i^*(t) - e(t)\} Y_i(t) \{W_i^*(t) - \widehat{W}_i(t)\}^{\mathrm{T}} \theta_0 dt \\ &= -n^{-1/2} \sum_{i\in\overline{V}} \Bigg(\int_0^\tau \frac{\{W_i^*(t) - e(t)\}}{f_{A^*}(a)} Y_i(t) \\ &\qquad \times \frac{1}{v} \sum_{j\in V} |B|^{-1} K\{B^{-1}(A_j^* - A_i^*)\} Y_j(t) \{W_j(t) - W_i^*(t)\}^{\mathrm{T}} \theta_0 dt \Bigg) \end{aligned} \tag{7}$$

for any $a \in \mathcal{A}$.

Lemma 3.3 *Under conditions* C1–C6*,*

$$n^{-1/2} \widehat{U}(\theta_0) \longrightarrow_d N\{0, (1-\rho)\Sigma_1(\theta_0) + \rho\Sigma_2(\theta_0)\}.$$

Proof The main idea is that we decompose $n^{-1/2}\widehat{U}(\theta_0)$ as two independent parts and use the martingale central limit theorem. Rewrite

$$\begin{aligned} n^{-1/2}\widehat{U}(\theta_0) &= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\widehat{W}_i(t)-\widehat{E}(t)\right\}dM_i(t) \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\widehat{W}_i(t)-\widehat{E}(t)\right\}Y_i(t)d\Lambda_0(t) \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\widehat{W}_i(t)-\widehat{E}(t)\right\}Y_i(t)\{W_i^*(t)-\widehat{W}_i(t)\}^{\mathrm{T}}\theta_0 dt \\ &\equiv B_{1n}+B_{2n}+B_{3n}. \end{aligned} \tag{8}$$

Furthermore, we rewrite

$$B_{1n}=n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\widehat{W}_i(t)-W_i^*(t)+W_i^*(t)-e(t)+e(t)-\widehat{E}(t)\right\}dM_i(t).$$

Note that $n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\widehat{W}_i(t)-W_i^*(t)\}dM_i(t)$ is a square integrable martingale, which converges in probability to zero by the Lenglart inequality (see [26]). Analogously,

$$n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{e(t)-\widehat{E}(t)\}dM_i(t)$$

also converges in probability to zero. Hence,

$$B_{1n}=n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{W_i^*(t)-e(t)\}\,dM_i(t)+o_P(1). \tag{9}$$

We also rewrite

$$B_{3n}=n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\widehat{W}_i(t)-W_i^*(t)+W_i^*(t)+e(t)-\widehat{E}(t)-e(t)\right\}Y_i(t)\{W_i^*(t)-\widehat{W}_i(t)\}^{\mathrm{T}}\theta_0 dt.$$

By (3), (5), (6), and condition C6, we have that

$$n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\widehat{W}_i(t)-W_i^*(t)\}Y_i(t)\{W_i^*(t)-\widehat{W}_i(t)\}^{\mathrm{T}}\theta_0 dt=o_P(1).$$

Then, using (4), (5), (6), and condition C6, we have also that

$$n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{e(t)-\widehat{E}(t)\}Y_i(t)\{W_i^*(t)-\widehat{W}_i(t)\}^{\mathrm{T}}\theta_0 dt=o_P(1).$$

Consequently, by (7) and the definition of $Q_j^{\overline{v}}(\theta)$, we obtain

$$B_{3n}=-n^{-1/2}\frac{\overline{v}}{v}\sum_{j\in V}Q_j^{\overline{v}}(\theta_0)+o_P(1).$$

Finally, according to the definition of $Q_j(\theta)$, we conclude that

$$n^{-1/2}\frac{\overline{v}}{v}\sum_{j\in V}\{Q_j^{\overline{v}}(\theta_0)-Q_j(\theta_0)\}\longrightarrow_p 0.$$

Thus,

$$B_{3n}=-n^{-1/2}\frac{\overline{v}}{v}\sum_{j\in V}Q_j(\theta_0)+o_P(1). \tag{10}$$

Obviously, noting $B_{2n}=0$ and combining (9) and (10), we derive (8) as follows:

$$\begin{aligned} n^{-1/2}\widehat{U}(\theta_0) &= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \{W_i^*(t)-e(t)\}\,dM_i(t)-n^{-1/2}\frac{\overline{v}}{v}\sum_{j\in V}Q_j(\theta_0)+o_P(1)\\ &= n^{-1/2}\sum_{i\in\overline{V}}\int_0^\tau \{W_i^*(t)-e(t)\}\,dM_i(t)\\ &\quad +n^{-1/2}\sum_{j\in V}\left[\int_0^\tau \left\{W_j^*(t)-e(t)\right\}dM_j(t)-\frac{\overline{v}}{v}Q_j(\theta_0)\right]+o_P(1). \end{aligned}$$

Note that the first term is a martingale, which converges to a mean-zero normal distribution with covariance $(1-\rho)\Sigma_1(\theta_0)$. Similarly, the second one also converges to a mean-zero normal distribution with covariance $\rho\Sigma_2(\theta_0)$ by noting that $E\{\frac{\overline{v}}{v}Q_j(\theta_0)\}=0$. The lemma follows immediately from that two terms are independent because they are summations over the nonvalidation and validation sets, respectively. ∎

Theorem 3.4 *Under conditions* C1–C6, $\widehat{\theta}_E$ *converges to* θ_0 *in probability.*

Proof We use Theorem 1 in [27] to prove the consistency of $\widehat{\theta}_E$ by verifying the following conditions.

(i) $n^{-1}\partial\widehat{U}(\theta)/\partial\theta$ exists and is continuous in an open neighborhood of θ_0;

(ii) $n^{-1}\partial\widehat{U}(\theta)/\partial\theta$ converges in probability to $\Sigma(\theta)$, uniformly in an open neighborhood of θ_0; Furthermore, every element of $\Sigma(\theta)$ is a continuous function of θ in the neighborhood of θ_0 and $\Sigma^{-1}(\theta_0)$ exists;

(iii) $n^{-1}\partial\widehat{U}(\theta_0)/\partial\theta$ is negative-definite with probability going to one;

(iv) $n^{-1}\widehat{U}(\theta_0)\longrightarrow_p 0$.

First, we have

$$-n^{-1}\frac{\partial\widehat{U}(\theta)}{\partial\theta}=n^{-1}\sum_{i=1}^{n}\int_0^\tau\left\{\widehat{W}_i(t)-\widehat{E}(t)\right\}\widehat{W}_i(t)^{\mathrm{T}}Y_i(t)dt.$$

Thus, (i) is satisfied. Second, rewrite

$$-n^{-1}\frac{\partial\widehat{U}(\theta)}{\partial\theta}=n^{-1}\sum_{i=1}^{n}\int_0^\tau\left\{\widehat{W}_i(t)-W_i^*(t)+e(t)-\widehat{E}(t)+W_i^*(t)-e(t)\right\}^{\otimes 2}Y_i(t)dt.$$

Combining (3), (4), and condition C3, after some algebraic manipulations, we can further conclude that

$$-n^{-1}\frac{\partial\widehat{U}(\theta)}{\partial\theta}=n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\{W_i^*(t)-e(t)\}^{\otimes 2}Y_i(t)dt+o_P(1).$$

It follows from the strong law of large numbers that

$$-n^{-1}\frac{\partial\widehat{U}(\theta)}{\partial\theta}\longrightarrow_p E\left[\int_0^{\tau}\{W^*(t)-e(t)\}^{\otimes 2}Y(t)dt\right]=\varSigma(\theta). \tag{11}$$

Thus, (ii) and (iii) are verified. Finally, (iv) also holds by Lemma 3.3. Hence, $\widehat{\theta}_E$ converges in probability to θ_0. ∎

Theorem 3.5 *Under conditions* C1–C6, $\sqrt{n}(\widehat{\theta}_E-\theta_0)$ *is asymptotically normal with mean zero and covariance matrix* $\varSigma(\theta_0)^{-1}\{(1-\rho)\varSigma_1(\theta_0)+\rho\varSigma_2(\theta_0)\}\{\varSigma(\theta_0)^{-1}\}^{\mathrm{T}}$, *where*

$$\varSigma(\theta)=E\left[\int_0^{\tau}\{W^*(t)-e(t)\}^{\otimes 2}Y(t)dt\right],$$

$$\varSigma_1(\theta)=E\left[\int_0^{\tau}\{W^*(t)-e(t)\}^{\otimes 2}Y(t)d\Lambda_0(t)\right]+E\left[\int_0^{\tau}\{W^*(t)-e(t)\}^{\otimes 2}Y(t)\theta^{\mathrm{T}}W^*(t)dt\right],$$

$$\varSigma_2(\theta)=E\left[\int_0^{\tau}\{W(t)-e(t)\}dM(t)-\frac{1-\rho}{\rho}Q(\theta)\right]^{\otimes 2}.$$

Proof Using Taylor expansion, we have

$$n^{-1/2}\widehat{U}(\theta_0)=\left\{-n^{-1}\frac{\partial\widehat{U}(\theta)}{\partial\theta}|_{\theta=\theta_0}\right\}n^{1/2}(\widehat{\theta}_E-\theta_0)+o_P(1).$$

It follows from Lemma 3.3, (11), and the consistency of $\widehat{\theta}_E$ that $\sqrt{n}(\widehat{\theta}_E-\theta_0)$ is asymptotically normal with mean zero and covariance matrix $\varSigma(\theta_0)^{-1}\{(1-\rho)\varSigma_1(\theta_0)+\rho\varSigma_2(\theta_0)\}\{\varSigma(\theta_0)^{-1}\}^{\mathrm{T}}$. Thus, we complete the proof of Theorem 3.5. ∎

4 Simulation Studies

In this section, we examined the finite sample properties of $\widehat{\theta}_E$ via simulation studies. We compared $\widehat{\theta}_E$ with two estimators. The first one is the validation set estimator, denoted by $\widehat{\theta}_V$, which is obtained by using method (see [14]) based only on the validation data. The other one is the naive estimator, denoted by $\widehat{\theta}_N$, which is the estimator by using the auxiliary covariate to replace the true primary covariate which is subject to missing. We compared these estimators under different levels of censoring proportions, validation fractions, and correlations between auxiliary and primary.

We generated the survival times $\widetilde{T}$ from the hazard model $\lambda(t|Z,X)=2+\beta_0X+\gamma_0Z$, where both X and Z were independently simulated from $\mathrm{Unif}(0,2)$. We construct the auxiliary covariate through $A=X+e$, where $e\sim N(0,\sigma^2)$. Here σ^2 is the parameter which controls the strength of the association between X and A. We generated η from the Bernoulli distribution with success probability ρ and the censoring time from Unif$(0, c)$. We set $\theta_0^{\mathrm{T}} = (\beta_0^{\mathrm{T}}, \gamma_0^{\mathrm{T}}) = (2, 2)$ and considered two different strength of association $\sigma = 0.2$ or 1, coupled with $\rho = 0.8$ and 0.5. The constant c was chosen to yield a censoring rate of 30% or 60%. Each configuration was replicated 1000 times under the sample size $n = 200$. We use the Epanechnikov kernel function[28] with bandwidth $b = 2\widehat{\sigma}_A n^{-1/3}$ and $\widehat{\sigma}_A$ is the sample standard deviation of A, which satisfies the bandwidth conditions. The corresponding results were summarized in Table 1. The column "Cen." is the censoring rate. "ρ" is the validation fraction. "σ" is the strength of association between primary and auxiliary. The column "Est" is the average value of the estimates. The sample standard derivation of the estimates is given in the column "SD". The column "SE" gives the average of the estimated standard errors and the column "95% CP" is the nominal 95% confidence interval coverage of the true parameter using the estimated standard errors.

Table 1 Simulation results based on the hazard model $\lambda(t|Z, X) = 2 + \beta_0^{\mathrm{T}} X + \gamma_0^{\mathrm{T}} Z$

Cen.				$\beta_0 = 2$				$\gamma_0 = 2$			
rate	ρ	σ	Method	Est	SD	SE	95%CP	Est	SD	SE	95%CP
30%	0.8		Validation	2.083	0.991	0.972	0.945	2.053	0.967	0.971	0.953
		0.2	Naive	2.057	0.866	0.865	0.949	2.043	0.852	0.865	0.961
			Proposed	2.065	0.869	0.865	0.945	2.043	0.852	0.854	0.954
		1	Naive	1.305	0.697	0.683	0.791	2.043	0.850	0.865	0.959
			Proposed	2.030	0.940	0.922	0.944	2.047	0.853	0.855	0.955
	0.5		Validation	2.090	1.278	1.239	0.946	2.021	1.232	1.235	0.956
		0.2	Naive	2.054	0.866	0.864	0.950	2.042	0.852	0.865	0.961
			Proposed	2.074	0.876	0.898	0.948	2.042	0.852	0.876	0.955
		1	Naive	0.830	0.558	0.546	0.420	2.046	0.852	0.865	0.960
			Proposed	1.857	1.062	1.066	0.949	2.049	0.851	0.878	0.962
60%	0.8		Validation	2.164	1.316	1.273	0.945	2.111	1.244	1.266	0.957
		0.2	Naive	2.133	1.136	1.132	0.942	2.115	1.117	1.130	0.955
			Proposed	2.142	1.141	1.132	0.948	2.115	1.116	1.124	0.958
		1	Naive	1.348	0.900	0.893	0.853	2.113	1.111	1.129	0.955
			Proposed	2.120	1.243	1.213	0.940	2.120	1.116	1.124	0.960
	0.5		Validation	2.220	1.690	1.616	0.954	2.082	1.583	1.611	0.962
		0.2	Naive	2.129	1.133	1.131	0.944	2.114	1.117	1.130	0.957
			Proposed	2.152	1.147	1.180	0.950	2.114	1.116	1.170	0.959
		1	Naive	0.847	0.709	0.716	0.603	2.113	1.113	1.129	0.952
			Proposed	1.991	1.409	1.442	0.948	2.118	1.116	1.175	0.970

From Table 1, we make the following observations: (i) All the estimates for γ_0 are essentially unbiased. For β_0, both $\widehat{\beta}_V$ and $\widehat{\beta}_E$ are virtually unbiased. However, $\widehat{\beta}_N$ is biased; (ii) It is natural that $\widehat{\theta}_E$ is more efficient than $\widehat{\theta}_V$, since $\widehat{\theta}_E$ use more information; (iii) When σ is large which means that A is less informative about X, $\widehat{\beta}_E$ is less accurate in estimating β_0. This bias, however, decreases as we increase the sample size to $n = 500$ (results not shown); (iv) The coverage rates are around the nominal level of 95%.

Table 2 compares the relative efficiency of $\widehat{\theta}_E$ versus $\widehat{\theta}_V$ under different censoring proportions, which are calculated through $\{\mathrm{SD}(\widehat{\theta}_V)/\mathrm{SD}(\widehat{\theta}_E)\}^2$, where $\mathrm{SD}(\widehat{\theta})$ is the sample standard derivation of estimate $\widehat{\theta}$. The column "σ" is the strength of association between primary and auxiliary covariates. "Cen." is the censoring rate. "ρ" is the validation fraction. "SD" is the sample standard derivation of the estimate. "RE" is relative efficiency of the proposed method over the validation set method. We observed that based on Table 2 when the validation fraction decreases, the efficiency gain of $\widehat{\theta}_E$ relative to $\widehat{\theta}_V$ increases. This suggests that when the validation fraction is small, using our proposed method is even more beneficial compared to the estimator based on the validation set only.

Table 2 The efficiency comparison between the proposed estimator $\widehat{\theta}_E$ and the validation set estimator $\widehat{\theta}_V$ with $\sigma = 0.2$

Cen. rate	ρ	SD ($\beta_0 = 2$) Validation	SD ($\beta_0 = 2$) Proposed	RE	SD ($\gamma_0 = 2$) Validation	SD ($\gamma_0 = 2$) Proposed	RE
30%	0.8	0.991	0.869	1.301	0.967	0.852	1.288
	0.5	1.278	0.876	2.128	1.232	0.852	2.091
	0.2	2.111	0.878	5.780	2.048	0.852	5.780
60%	0.8	1.316	1.141	1.330	1.244	1.116	1.243
	0.5	1.690	1.147	2.171	1.583	1.116	2.012
	0.2	2.767	1.146	5.828	2.673	1.116	5.738

5 The Primary Biliary Cirrhosis Data

We apply the proposed method to the data from the Mayo Clinic trial in the primary biliary cirrhosis (PBC) of the liver. The PBC is a chronic and fatal liver disease characterised by inflammatory destruction of the small bile ducts within the liver, which finally leads to cirrhosis of the liver. The cause of PBC is unknown, but it is generally thought to be an autoimmune disease because of the presence of autoantibodies. About 90% of patients with PBC are women. Patients often present abnormalities in their blood tests, such as elevated and gradually increasing serum bilirubin. In this randomised clinical trial, a total of 312 PBC patients met the eligibility criteria. The days from registration to the earlier of death, transplantation, or study analysis time were recorded. The covariates of interest include serum cholesterol level (Chol), treatment (Trt), and patients' sex (Sex). We fit the data using the following model:

$$\lambda(t|\boldsymbol{Z}, X) = \lambda_0(t) + \beta X_{\log(\mathrm{Chol})} + \gamma_{01} Z_{\mathrm{Trt}} + \gamma_{02} Z_{\mathrm{Sex}},$$

where β, γ_{01}, and γ_{02} are the unknown parameters.

A clinical background description and a more extend discussion for the trial and the covariates recorded can be found in [29] and [30].

About 9% outcomes of cholesterol were missing in this data set. Removing those observations could lead to efficiency loss. Our exploratory data analysis shows strong correlation between cholesterol and bilirubin (Bili), which is observed completely. Therefore, we use the serum bilirubin as the auxiliary covariate for cholesterol. We follow the literature clinical study and take the logarithmic transformation of cholesterol and bilirubin, respectively.

Table 3 displays the analysis results from the proposed method and the validation set method. We did not find any significant difference across the treatment group. In addition, it can be seen that patients associated with lower serum cholesterol level or the female could be expected to live longer. These findings coincide with previous analysis in the literature. On the other hand, the proposed method produced more precise assessment of the covariate effect of the serum cholesterol level, compared with the validation set method, which discarding the incomplete data could lead to the loss of efficiency.

Table 3 Analysis results for the PBC data

Method	Covariate	Est	SE	95%CI	p-value
Validation	Trt	0.001	0.013	$(-0.024, 0.025)$	0.955
	Sex	-0.052	0.025	$(-0.102, -0.004)$	0.036
	log(Chol)	0.049	0.021	$(0.008, 0.090)$	0.019
Proposed	Trt	-0.003	0.012	$(-0.026, 0.020)$	0.803
	Sex	-0.055	0.030	$(-0.113, 0.003)$	0.063
	log(Chol)	0.061	0.019	$(0.023, 0.099)$	0.002

6 Concluding Remarks

In this article, under the framework of the additive hazards model, we proposed an estimated estimating equation method for the survival data with continuous auxiliary information to further improve study efficiency. A key feature of this method is that it does not require to specify the association between the missing covariate and the auxiliary covariate. The resultant estimates were shown to be consistent and asymptotically normal. Simulation studies demonstrated that the proposed method works well with moderate sample size and that the resulting estimator outperforms the validation set estimator. The proposed variance estimator is also viable in practice. When the auxiliary covariate is more informative about the primary exposure, the proposed estimator is more efficient.

Our additional simulation study in shows that the proposed method is feasible when the dimension of A^* is two. However, the nonparametric kernel smoothing would suffer from the curse of dimensionality when the dimension is larger. The unstability of the nonparametric kernel smoothing estimation may further deteriorate the behaviors of the proposed method.

For higher dimension d, the techniques such as dimension reduction or parametric modeling could be considered (see [31]).

References

[1] Cox D R, Regression models and life-tables, *Journal of the Royal Statistical Society, Series B*, 1972, **34**: 187–202.

[2] Prentice R L, Covariate measurement errors and parameter estimation in a failure time regression model, *Biometrika*, 1982, **69**: 331–342.

[3] Pepe M S and Fleming T R, A nonparametric method for dealing with mismeasured covariate data, *Journal of the American Statistical Association*, 1991, **86**: 108–113.

[4] Lin D Y and Ying Z, Cox regression with incomplete covariate measurements, *Journal of the American Statistical Association*, 1993, **88**: 1341–1349.

[5] Zhou H and Pepe M S, Auxiliary covariate data in failure time regression analysis, *Biometrika*, 1995, **82**: 139–149.

[6] Zhou H and Wang C, Failure time regression with continuous covariates measured with error, *Journal of the Royal Statistical Society, Series B*, 2000, **62**: 657–665.

[7] Greene W F and Cai J, Measurement error in covariate in the marginal hazards model for multivariate failure time data, *Biometrics*, 2004, **60**: 987–996.

[8] Liu Y, Wu Y, and Zhou H, Multivariate failure times regression with a continuous auxiliary covariate, *Journal of Multivariate Analysis*, 2010, **101**: 679–691.

[9] Liu Y, Zhou H, and Cai J, Estimated pseudopartial-likelihood method for correlated failure time data with auxiliary covariates, *Biometrics*, 2009, **65**: 1184–1193.

[10] Breslow N E and Day N E, *Statistical Methods in Cancer Research, The Design and Analysis of Case-Control Studies*, Lyon, IARC, 1980.

[11] Breslow N E and Day N E, *Statistical Methods in Cancer Research, The Design and Analysis of Cohort Studies*, Lyon, IARC, 1987.

[12] Cox D R and Oakes D, *Analysis of Survival Data*, Chapman and Hall, London, 1984.

[13] Thomas D C, Use of auxiliary information in fitting nonproportional hazards models, in *Modern Statistical Methods in Chronic Disease Epidemiology* (ed. by Moolgavkar S H and Prentice R L), Wiley, New York, 1986, 197–210.

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

[15] O'Neill T J, Inconsistency of the misspecified proportional hazards model, *Statistics and Probability Letters*, 1986, **4**: 219–222.

[16] Kulich M and Lin D Y, Additive hazards regression with covariate measurement error, *Journal of the American Statistical Association*, 2000, **95**: 238–248.

[17] Jiang J and Zhou H, Additive hazard regression with auxiliary covariates, *Biometrika*, 2007, **92**: 359–369.

[18] Rubin D B, Inference and missing data, *Biometrika*, 1976, **63**: 581–592.

[19] Fleming T R and Harrinton D P, *Counting Processes and Survival Analysis*, Wiley, New York, 1991.

[20] Breslow N E, Discussion of paper of D. R. Cox, *Journal of the Royal Statistical Society, Series B*, 1972, **34**: 216–217.

[21] Aalen O O, A linear regression model for the analysis of life times, *Statistics in Medcine*, 1989, **8**: 907–925.

[22] Nadaraya E A, On estimating regression, *Theory of Probability and Its Applications*, 1964, **10**: 186–190.

[23] Watson G S, Smooth regression analysis, *Sankya, Series A*, 1964, **26**: 359–372.

[24] Kosorok M R, *Introduction to Empirical Processes and Semiparametric Inference*, Springer, New York, 2008.

[25] Yin G, Li H, and Zeng D, Partially linear additive hazards regression with varying coefficients, *Journal of the American Statistical Association*, 2008, **103**: 1200–1213.

[26] Andersen P K and Gill R D, Cox's regression model for counting processes: A large sample study, *The Annals of Statistics*, 1982, **10**: 1100–1120.

[27] Foutz R V, On the unique consistent solution to likelihood equations, *Journal of the American Statistical Association*, 1977, **72**: 147–148.

[28] Eubank R L, *Smoothing Spline and Nonparametric Regression*, Dekker, New York, 1988.

[29] Dickson E R, Grambsch P M, Fleming T R, Fisher L D, and Langworthy A, Prognosis in primary biliary cirrhosis: Model for decision making, *Hepatology*, 1989, **10**: 1–7.

[30] Markus B H, Dickson E R, Grambsch P M, Fleming T R, Mazzaferro V, Klintmalm G B, Wiesner R H, Van Thiel D H, and Starzl T E, Efficiency of liver transplantation in patients with primary biliary cirrhosis, *The New England Journal of Medicine*, 1989, **320**: 1709–1713.

[31] Wang J H, Zhou J, and Li Y, Variable selection for censored quantile regresion, *Statistica Sinica*, 2013, **23**: 145–167.