Journal of Multivariate Analysis 101 (2010) 679–691

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Multivariate failure times regression with a continuous auxiliary covariate

Yanyan Liu [a,*], Yuanshan Wu [a,b], Haibo Zhou [b]

[a] School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, PR China

[b] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA

ARTICLE INFO

Article history:
Received 5 March 2008
Available online 30 September 2009

AMS subject classifications:
primary 62N02
secondary 62H12

Keywords:
Multivariate failure times
Auxiliary covariate
Pseudo-partial likelihood
Kernel smoothing
Validation sample

ABSTRACT

How to take advantage of the available auxiliary covariate information when the primary covariate of interest is not measured is a frequently encountered question in biomedical study. In this paper, we consider the multivariate failure times regression analysis in which the primary covariate is assessed only in a validation set, but a continuous auxiliary covariate for it is available for all subjects in the study cohort. Under the frame of marginal hazard model, we propose to estimate the induced relative risk function in the non-validation set through kernel smoothing method and then obtain an estimated pseudo-partial likelihood function. The proposed estimator which maximizes the estimated pseudo-partial likelihood is shown to be consistent and asymptotically normal. We also give an estimator of the marginal cumulative baseline hazard function. Simulation studies are conducted to evaluate the finite sample performance of our proposed estimator. The proposed method is illustrated by analyzing a heart disease data from the Study of Left Ventricular Dysfunction (SOLVD).

1. Introduction

In biomedical studies, it is common that the true exposure variable of interest, X, is only measured for a subset of the whole study cohort. This subset is often referred to as the validation set where it is usually just a simple random sample of the whole cohort. In many cases, there exists some easily available auxiliary covariate about X, denoted by W, that can be easily assessed for the full study cohort. Real examples for data with auxiliary covariate can be found in cancer research (e.g. [1,2]) and heart disease research (e.g. [3]).

It is well known that simply using W in place of X when X is not observed could lead to biased results. On the other hand, analysis based solely on the validation set would not be the most efficient one as information from the non-validation set, the set of individuals whose X is not assessed, is not utilized. It would be always desirable to incorporate the available auxiliary covariate information to improve the efficiency of inference. For the censored failure time data with auxiliary covariate, some methods have been proposed to improve the study efficiency. Among others, Prentice [4] proposed using induced partial likelihood for the Cox regression under the rare events assumption. Zhou and Pepe [5] removed the rare diseases assumption and proposed an estimated partial likelihood method for discrete auxiliary variable. Their work was extended by Zhou and Wang [6] to handle continuous auxiliary covariate based on a kernel smoother approach. Kulich and Lin [7] and Jiang and Zhou [8] studied failure time data with auxiliary covariate under additive hazards model of Lin and Ying [9].

All the aforementioned studies assumed that the failure time is univariate. In real studies, multivariate failure time data with auxiliary problem are just as frequently encountered. There are limited methods available for auxiliary covariate

* Corresponding author.
E-mail addresses: liuyy@whu.edu.cn (Y. Liu), wuyuanshan@gmail.com (Y. Wu), zhou@bios.unc.edu (H. Zhou).

0047-259X/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2009.09.008

problem when the failure times are correlated. For multiplicative hazards models, Hu and Lin [10] proposed a corrected estimating function methods for the marginal Cox proportional hazards model under the assumption that the measurement error is symmetrically distributed. Liu, Zhou and Cai [11] proposed an estimated pseudo-partial likelihood method for marginal Cox proportional hazards model with discrete auxiliary. However, the continuous auxiliary often arises in practical studies (e.g. SOLVD data we studied in Section 5). In this case, Liu, Zhou and Cai suggested to discretize the continuous auxiliary first and then use the available method. This simplification would entail a loss of information and other limitations [11]. In this paper, we develop a method to handle continuous auxiliary W under the marginal Cox model framework. The proposed method is based on nonparametric multivariate kernel smoothing procedure and makes no assumption about the relationship between the auxiliary variable and the true exposure variable. The proposed estimator for the regression parameter which maximizes the estimated pseudo-partial likelihood is shown to be consistent and asymptotically normal.

The rest of this paper is organized as follows. In Section 2, we propose to estimate the induced risk function through kernel method, derive the proposed estimator for the regression parameter from an estimated pseudo-partial likelihood function, and finally give a Breslow-type estimator for the baseline cumulative hazard function. Asymptotic properties of our proposed estimators are established in Section 3. Section 4 presents a simulation study to evaluate the performance of our proposed method. We apply the proposed method to study the effect of ejection fraction on the risk of heart failure and first myocardial infarction using the data from the SOLVD [3] in Section 5. Some concluding remarks are provided in Section 6. Sketched proofs of the asymptotic properties are presented in the Appendix.

2. Model and estimation

2.1. Notation and data structure

Suppose that there is a random sample of n independent groups from an underlying population and that there are K different correlated members in each group. Let (i, k) denote the kth member in the ith group. Let $\widetilde{T}_{ik}$ be the failure time for subject (i, k), C_{ik} the correspondingly potential censoring time, and $T_{ik} = \min(\widetilde{T}_{ik}, C_{ik})$ the observed time. The censoring indicator is denoted by $\Delta_{ik} = I(\widetilde{T}_{ik} \leq C_{ik})$, where $I(\cdot)$ is the indicator function. $Y_{ik}(t) = I(T_{ik} \geq t)$ denotes the at-risk indicator process. For simplicity, we assume that each cluster potentially has the same number of subjects, that is, K is fixed. However, we may allow the cluster size to change by setting appropriate $C_{ik} = 0$. Write the transpose of a matrix A as A' and let $(X'_{ik}(t), Z'_{ik}(t))'$ be a p-vector of possibly time-dependent covariate, where $X_{ik}(t)$ is the primary covariate subject to missing and $Z_{ik}(t)$ is the remaining covariate that is always observed. Let $W_{ik}(t)$ be the auxiliary covariate for $X_{ik}(t)$, and assume that conditional on $X_{ik}(t)$, $W_{ik}(t)$ provides no additional information to the regression model in the sense that

$$\begin{aligned}\lambda_{ik}(t; X_{ik}(t), Z_{ik}(t), W_{ik}(t)) &\equiv \lim_{\Delta t \downarrow 0}\left[\frac{1}{\Delta t}\Pr\left(t \leq \widetilde{T}_{ik} < t + \Delta t \mid \widetilde{T}_{ik} \geq t, X_{ik}(t), Z_{ik}(t), W_{ik}(t)\right)\right] \\ &= \lim_{\Delta t \downarrow 0}\left[\frac{1}{\Delta t}\Pr\left(t \leq \widetilde{T}_{ik} < t + \Delta t \mid \widetilde{T}_{ik} \geq t, X_{ik}(t), Z_{ik}(t)\right)\right] \\ &\equiv \lambda_{ik}(t; X_{ik}(t), Z_{ik}(t)).\end{aligned}$$

This is a kind of surrogate condition that asserts a conditional independence, given $X_{ik}(t)$ and $Z_{ik}(t)$, of failure rate at t and $W_{ik}(t)$ for subject (i, k) and requires the auxiliary covariate $W_{ik}(t)$ to have no predictive value given the covariates $X_{ik}(t)$ and $Z_{ik}(t)$. Let η_{ik} be an indicator variable with $\eta_{ik} = 1$ indicating that subject (i, k) is in the validation set and with $\eta_{ik} = 0$ in the non-validation set. Let $\bar{\eta}_{ik} = 1 - \eta_{ik}$. Denote the kth marginal validation set by $V_k = \{i : \eta_{ik} = 1\}$ and non-validation set by $\bar{V}_k = \{i : \bar{\eta}_{ik} = 1\}$, respectively. Let n_k denote the number of subjects in V_k and assume $\frac{n_k}{n} \to \rho_k$, as $n \to \infty$, where ρ_k is an unknown positive constant representing the fraction of kth marginal validation set. If $\eta_{ik} = 1$, then the observed data for subject (i, k) are $\{T_{ik}, \Delta_{ik}, Y_{ik}(t), X_{ik}(t), Z_{ik}(t), W_{ik}(t)\}$ and if $\eta_{ik} = 0$, then the observed data for subject (i, k) are $\{T_{ik}, \Delta_{ik}, Y_{ik}(t), Z_{ik}(t), W_{ik}(t)\}$. Suppose that the data are observed on the time interval $[0, \tau]$, where $0 < \tau < \infty$ is a fixed quantity. For fixed k, suppose $(\widetilde{T}_{ik}, C_{ik}, \eta_{ik}, X_{ik}(t), Z_{ik}(t), W_{ik}(t); t \in [0, \tau])(i = 1, \ldots, n)$ are independent and identically distributed.

2.2. Model and kernel-based estimation of pseudo-partial likelihood function

Assume that the marginal hazards function for the subject (i, k) takes the form:

$$\lambda_{ik}(t; X_{ik}(t), Z_{ik}(t)) = \lambda_{0k}(t) \exp\{\beta'_1 X_{ik}(t) + \beta'_2 Z_{ik}(t)\}, \tag{1}$$

where $\beta = (\beta'_1, \beta'_2)'$ is the relative risk parameter to be estimated, and $\lambda_{0k}(t)$ is an unspecified marginal baseline hazard function pertaining to the kth marginal subjects. If subject (i, k) belongs to the validation set V_k, then $X_{ik}(t)$ and $Z_{ik}(t)$ are observed and the marginal model takes the form as (1). If subject (i, k) belongs to the non-validation set $\bar{V}_k$, then $X_{ik}(t)$ is not observed. Under this situation, it can be verified, using the arguments of Prentice [4] and Zhou and Wang [6], that the hazard function $\lambda_{ik}(t; Z_{ik}(t), W_{ik}(t))$ satisfies the induced model

$$\lambda_{ik}(t; Z_{ik}(t), W_{ik}(t)) \equiv \lim_{\Delta t \downarrow 0}\left[\frac{1}{\Delta t}\Pr\left(t \le \widetilde{T}_{ik} < t + \Delta t \mid \widetilde{T}_{ik} \ge t, W_{ik}(t), Z_{ik}(t)\right)\right]$$
$$= \lambda_{0k}(t)E\{e^{\beta_1' X_{ik}(t)} | \widetilde{T}_{ik} \ge t, W_{ik}(t), Z_{ik}(t)\}\exp(\beta_2' Z_{ik}(t)).$$

Under the independent censoring assumption described by equation (5) in Prentice [4] that

$$\lambda_{ik}(t; Z_{ik}(t), W_{ik}(t), \text{no censorship in } [0, t)) = \lambda_{ik}(t; Z_{ik}(t), W_{ik}(t)),$$

we can rewrite the induced model as

$$\lambda_{ik}(t; Z_{ik}(t), W_{ik}(t)) = \lambda_{0k}(t)E\{e^{\beta_1' X_{ik}(t)} | Y_{ik}(t) = 1, W_{ik}(t), Z_{ik}(t)\}\exp(\beta_2' Z_{ik}(t)). \tag{2}$$

Note that this induced hazard model (2) is also a proportional hazard model with the relative risk function

$$\phi_{ik}(t; \beta) \equiv E\{e^{\beta_1' X_{ik}(t)} | Y_{ik}(t) = 1, W_{ik}(t), Z_{ik}(t)\}\exp(\beta_2' Z_{ik}(t)),$$

which is a weighted average of the relative risks in model (1), given $W_{ik}(t)$ and $Z_{ik}(t)$ at risk prior to time t.

Based on (1) and (2), the relative risk function can be concluded in general as

$$r_{ik}(t; \beta) \equiv \varphi_{ik}(t; \beta)\eta_{ik} + \phi_{ik}(t; \beta)(1 - \eta_{ik}), \tag{3}$$

where $\varphi_{ik}(t; \beta) = \exp(\beta_1' X_{ik}(t) + \beta_2' Z_{ik}(t))$.

If all the subjects under study are independent, we can write the partial likelihood as

$$PPL(\beta) \equiv \prod_{k=1}^{K}\prod_{i=1}^{n}\left[\frac{r_{ik}(T_{ik}; \beta)}{\sum_{j=1}^{n} Y_{jk}(T_{ik}) r_{jk}(T_{ik}; \beta)}\right]^{\Delta_{ik}}. \tag{4}$$

When the failure times within each group are correlated, the above partial likelihood function is referred to as the pseudo-partial likelihood [12]. Since the induced relative risk function $r_{ik}(t; \beta)$ includes unknown conditional expectation except the regression parameter, we should estimate it by using data from the validation set.

To estimate $r_{ik}(t; \beta)$, it suffices to estimate $\phi_{ik}(t; \beta)$. The conditional expectation in $\phi_{ik}(t; \beta)$ depends on the underlying distributions of $X_{ik}(t)$ and $Z_{ik}(t)$. If $f(X_{ik}(t) | T_{ik} \ge t, Z_{ik}(t), W_{ik}(t))$ is a known function up to a parameter θ, then the inference about β and θ can be derived from the usual pseudo-partial likelihood (4) based on the general relative function $r_{ik}(t; \beta, \theta)$ [13]. To avoid making undesirable parameter assumption, Liu, Zhou and Cai [11] used a marginal empirical relative risk estimator for those in the non-validation set when $W_{ik}(t)$ is discrete. For continuous $W_{ik}(t)$, we propose to estimate the relative risk in non-validation set through a kernel smoother method.

Let $Z_{ik}^* = (Z_{ik}', W_{ik}')'$ be a d-dimension vector. Using the method of Nadaraya [14] and Watson [15], we could estimate $\phi_{ik}(t; \beta)$ as:

$$\widehat{\phi}_{ik}(t; \beta) = \frac{\sum_{j=1}^{n} \eta_{jk} Y_{jk}(t) Q_k\left\{B_k^{-1}\{Z_{jk}^*(t) - Z_{ik}^*(t)\}\right\}\exp\{\beta_1' X_{jk}(t)\}}{\sum_{j=1}^{n} \eta_{jk} Y_{jk}(t) Q_k\left\{B_k^{-1}\{Z_{jk}^*(t) - Z_{ik}^*(t)\}\right\}}\exp\{\beta_2' Z_{ik}(t)\}, \tag{5}$$

where $Q_k(\cdot)$ is a kernel function with bandwidth matrix B_k, which is $d \times d$ positive-definite, with its elements possibly depending on n. For simplicity, we only consider the situation in which B_k is a diagonal matrix with element at (l, l) denoted by b_{lk}.

Replacing $\phi_{ik}(t; \beta)$ by $\widehat{\phi}_{ik}(t; \beta)$ in (3), we obtain an estimated induced relative risk function:

$$\widehat{r}_{ik}(t; \beta) = \varphi_{ik}(t; \beta)\eta_{ik} + \widehat{\phi}_{ik}(t; \beta)(1 - \eta_{ik}). \tag{6}$$

Note that the kernel-based relative risk estimator is not defined when the denominator in (5) is equal to zero, which occurs when the risk set at time t in the kth marginal validation set V_k is the null set. This could be treated by interpolating the neighboring points of continuous $Z_{ik}^*(t)$ as suggested by Zhou and Wang [6].

Substituting $\widehat{r}_{ik}(t; \beta)$ for $r_{ik}(t; \beta)$ in (4), we obtain an estimated pseudo-partial likelihood function:

$$EPPL(\beta) \equiv \prod_{k=1}^{K}\prod_{i=1}^{n}\left[\frac{\widehat{r}_{ik}(T_{ik}; \beta)}{\sum_{j=1}^{n} Y_{jk}(T_{ik}) \widehat{r}_{jk}(T_{ik}; \beta)}\right]^{\Delta_{ik}}. \tag{7}$$

The proposed estimator, denoted by $\widehat{\beta}_E$, which is defined as the maximizer of (7), is used to estimate β_0, the true value of β. $\widehat{\beta}_E$ can be obtained by solving the resultant estimated pseudo-partial likelihood score equation $\widehat{U}(\beta) = 0$, where

$$\widehat{U}(\beta) = \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^{\tau}\left[\frac{\widehat{r}_{ik}^{(1)}(u;\beta)}{\widehat{r}_{ik}(u;\beta)} - \frac{\sum_{j=1}^{n} Y_{jk}(u)\widehat{r}_{jk}^{(1)}(u;\beta)}{\sum_{j=1}^{n} Y_{jk}(u)\widehat{r}_{jk}(u;\beta)}\right]\mathrm{d}N_{ik}(u), \tag{8}$$

$N_{ik}(t) = I(T_{ik} \le t, \Delta_{ik} = 1)$ is the observed counting process corresponding to the subject (i, k), and $\widehat{r}_{ik}^{(j)}(t;\beta)$ denotes the jth derivative of $\widehat{r}_{ik}(t;\beta)$ with respect to β, especially, $\widehat{r}_{ik}^{(0)}(t;\beta) = \widehat{r}_{ik}(t;\beta)$. The same definitions go for $r_{ik}^{(j)}(t;\beta)$. We can use Newton–Raphson iterative procedure to get $\widehat{\beta}_E$.

Furthermore, a Breslow-type estimator of the marginal cumulative baseline hazard function $\Lambda_{0k}(t)$ can be consequently constructed as follows:

$$\widehat{\Lambda}_{0k}(t) = \int_0^t \frac{\sum_{i=1}^{n}\mathrm{d}N_{ik}(u)}{\sum_{i=1}^{n} Y_{ik}(u)\widehat{r}_{ik}(u;\widehat{\beta}_E)}.$$

3. Asymptotic properties

To investigate the asymptotic properties of $\widehat{\beta}_E$ and $\widehat{\Lambda}_{0k}(t)$, we define the following notation. For a vector a, define $a^{\bigotimes 2} = aa'$ and $\|a\| = \sup_i |a_i|$. For a matrix A, define $\|A\| = \sup_{i,j}|a_{ij}|$. Unless otherwise stated, all the limits are taken as $n \to \infty$.

For $k = 1, \ldots, K$, let

$$S_k^{(a)}(t;\beta) = \frac{1}{n}\sum_{i=1}^{n} Y_{ik}(t) r_{ik}^{(a)}(t;\beta), \quad a = 0, 1, 2,$$

$$S_k^{(3)}(t;\beta) = \frac{1}{n}\sum_{i=1}^{n} Y_{ik}(t)\frac{r_{ik}^{(2)}(t;\beta)}{r_{ik}(t;\beta)} r_{ik}(t;\beta_0),$$

$$S_k^{(4)}(t;\beta) = \frac{1}{n}\sum_{i=1}^{n} Y_{ik}(t)\left\{\frac{r_{ik}^{(1)}(t;\beta)}{r_{ik}(t;\beta)}\right\}^{\bigotimes 2} r_{ik}(t;\beta_0).$$

For $l = 0, \ldots, 4$, we define $\widehat{S}_k^{(l)}(t;\beta)$ by substituting $\widehat{r}_{ik}(t;\beta)$ for $r_{ik}(t;\beta)$ in $S_k^{(l)}(t;\beta)$ and denote $s_k^{(l)}(t;\beta) = ES_k^{(l)}(t;\beta)$. Except the aforementioned assumptions, we impose the following conditions throughout our discussions:

Conditions:

C1. $\Lambda_{0k}(\tau) < \infty$, for $k = 1, \ldots, K$.

C2. $P\{Y_{ik}(t) = 1\} > 0$, for all $t \in [0, \tau]$ and $i = 1, \ldots, n$ and $k = 1, \ldots, K$.

C3. For every i and k, each component of $(X'_{ik}, Z'_{ik}, W'_{ik})'$ has bounded total variation almost surely over $[0, \tau]$.

C4. There exists an open set $\mathcal{B}$, containing β_0 as its interior point, such that $\phi_{ik}(t;\beta)$ is bounded away from 0 on $[0, \tau] \times \mathcal{B}$. $A(\beta_0) \equiv \sum_{k=1}^{K} A_k(\beta_0)$ is positive-definite, where

$$A_k(\beta_0) = \int_0^{\tau}\left[\frac{s_k^{(4)}(t;\beta_0)}{s_k^{(0)}(t;\beta_0)} - \left\{\frac{s_k^{(1)}(t;\beta_0)}{s_k^{(0)}(t;\beta_0)}\right\}^{\bigotimes 2}\right] s_k^{(0)}(t;\beta_0)\lambda_{0k}(t)\mathrm{d}t.$$

C5. For $k = 1, \ldots, K$, multivariate kernel $Q_k(\cdot)$ is non-negative and uniformly bounded with finite support satisfying that $\int Q_k(\mathbf{u})\mathrm{d}\mathbf{u} = 1$ and $\int Q_k^2(\mathbf{u})\mathrm{d}\mathbf{u} < \infty$. Furthermore, the kernel $Q_k(\cdot)$ has order α_{0k} in the sense that $\alpha_{0k} \equiv \inf\{|\alpha| > d;\ \int_{R^d}\mathbf{u}^{\alpha}Q_k(\mathbf{u})\mathrm{d}\mathbf{u} \ne 0\}$, where $\mathbf{u}^{\alpha} = u_1^{\alpha_1}\cdots u_d^{\alpha_d}$, $|\alpha| = \alpha_1 + \cdots + \alpha_d$, $\mathbf{u} = (u_1, \ldots, u_d)$, $\alpha = (\alpha_1, \ldots, \alpha_d)$, and α_i's are non-negative integers. The bandwidth B_k satisfies that $\sqrt{n}\|B_k\|^{\alpha_{0k}} \to 0$ and $\frac{\log n}{\sqrt{n}\|B_k\|^d} \to 0$.

C6. For given t, let $H_k(u, \mathbf{v}, \mathbf{s})$ be the joint distribution of $(\eta_k Y_k(t), Z_k^*(t), X_k(t))$. Suppose that $h_k(\mathbf{v}, \mathbf{s}) = \frac{\partial^2 H_k(1,\mathbf{v},\mathbf{s})}{\partial\mathbf{v}\partial\mathbf{s}}$ has the α_{0k}th continuous derivatives with respect to every component of $\mathbf{v}$.

Conditions C1–C4 are regular conditions similar to those given in Spiekerman and Lin [16]; condition C5 is the usual condition for kernel smoothing; condition C6 is a technical assumption for proving.

To develop the asymptotic properties of the proposed estimators, we first claim that $s_k^{(l)}$ can be consistently estimated by $\widehat{S}_k^{(l)}$ for all $t \in [0, \tau]$ in the following lemma.

Lemma 1. *Under conditions* C1–C6, *for $k = 1, \ldots, K$, and $l = 0, \ldots, 4$, there exists a neighborhood $\mathcal{B}$ of β_0 such that $s_k^{(l)}(t;\beta)$ is continuous function of $\beta \in \mathcal{B}$ uniformly in $t \in [0, \tau]$ and is bounded on $[0, \tau] \times \mathcal{B}$. $s_k^{(0)}(t;\beta)$ is bounded away from zero on $[0, \tau] \times \mathcal{B}$. Furthermore, for $a = 0, 1, 2$,*

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}} |\widehat{\phi}_{ik}^{(a)}(t;\beta) - \phi_{ik}^{(a)}(t;\beta)| \to_p 0;$$

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}} \|\widehat{S}_k^{(l)}(t;\beta) - S_k^{(l)}(t;\beta)\| \to_p 0;$$

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}} \|\widehat{S}_k^{(l)}(t;\beta) - s_k^{(l)}(t;\beta)\| \to_p 0.$$

The proof of Lemma 1 is outlined in the Appendix. Furthermore, let the filtration $\{\mathcal{F}_t^k : t \in [0,\tau]\}$ be the kth marginal data history up to time t, that is,

$$\mathcal{F}_t^k = \sigma\{N_{ik}(s), Y_{ik}(s+), Z_{ik}(s+), X_{ik}(s+) : 0 \le s \le t, i = 1,\ldots,n\}.$$

Define $M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(u) r_{ik}(u;\beta_0)\lambda_{0k}(u)\mathrm{d}u$, then $M_{ik}(t)$ is a $\mathcal{F}_t^k$ martingale. Following closely the arguments of Foutz [17], we can show that $\widehat{\beta}_E$ is consistent for β_0. To show the asymptotic normality of $\widehat{\beta}_E$, we use the Taylor expansion of the estimated pseudo-partial likelihood score equation which, using martingale convergence theory and standard kernel estimate theory, can be shown to be asymptotically equivalent to a sum of two independent terms. Each of the terms can be shown to be a scaled sum of independent vectors. The multivariate central limit theorem is then applicable. We summarize the results in the following theorem and give the outline of the proofs in the Appendix.

Theorem 1. *Under conditions* C1–C6, $\widehat{\beta}_E$ *is a consistent estimator of* β_0*. Furthermore,* $\sqrt{n}(\widehat{\beta}_E - \beta_0)$ *is asymptotically normal distributed with mean zero and variance matrix*

$$\Sigma(\beta_0) = A^{-1}(\beta_0)[B_1(\beta_0) + B_2(\beta_0)]A^{-1}(\beta_0),$$

where $A(\beta_0)$ *is defined as in condition* C4*, and*

$$B_1(\beta) = \sum_{k=1}^K \sum_{l=1}^K \sqrt{\rho_k}\sqrt{\rho_l} E\left(\mu_{1k}(\beta)\mu'_{1l}(\beta)\right),$$

$$B_2(\beta) = \sum_{k=1}^K \sum_{l=1}^K \sqrt{1-\rho_k}\sqrt{1-\rho_l} E\left(\nu_{1k}(\beta)\nu'_{1l}(\beta)\right),$$

with

$$\mu_{jk}(\beta) = \int_0^\tau \left[\frac{\varphi_{jk}^{(1)}(t;\beta)}{\varphi_{jk}(t;\beta)} - \frac{s_k^{(1)}(t;\beta)}{s_k^{(0)}(t;\beta)}\right]\mathrm{d}M_{jk}(t) - \frac{1-\rho_k}{\rho_k} H_{jk}(\beta),$$

$$H_{jk}(\beta) = \int_0^\tau \left[\frac{\phi_{jk}^{(1)}(t;\beta)}{\phi_{jk}(t;\beta)} - \frac{s_k^{(1)}(t;\beta)}{s_k^{(0)}(t;\beta)}\right]\left(\varphi_{jk}(t;\beta) - \phi_{jk}(t;\beta)\right) Y_{jk}(t)\mathrm{d}\Lambda_{0k}(t),$$

$$\nu_{jk}(\beta) = \int_0^\tau \left[\frac{\phi_{jk}^{(1)}(t;\beta)}{\phi_{jk}(t;\beta)} - \frac{s_k^{(1)}(t;\beta)}{s_k^{(0)}(t;\beta)}\right]\mathrm{d}M_{jk}(t).$$

Remark 1. Observe that, when the validation fraction $\rho_k = 1$ for every k, the variance matrix is the same as that of the usual pseudo-partial likelihood estimator [12], as it should be.

It follows from Theorem 1 and Lemma 1 that the covariance matrix $\Sigma(\beta_0)$ can be consistently estimated by

$$\widehat{\Sigma}(\widehat{\beta}_E) = \widehat{A}^{-1}(\widehat{\beta}_E)[\widehat{B}_1(\widehat{\beta}_E) + \widehat{B}_2(\widehat{\beta}_E)]\widehat{A}^{-1}(\widehat{\beta}_E),$$

where

$$\widehat{A}(\widehat{\beta}_E) = \sum_{k=1}^K \int_0^\tau \left[\frac{\widehat{S}_k^{(4)}(t;\widehat{\beta}_E)}{\widehat{S}_k^{(0)}(t;\widehat{\beta}_E)} - \left\{\frac{\widehat{S}_k^{(1)}(t;\widehat{\beta}_E)}{\widehat{S}_k^{(0)}(t;\widehat{\beta}_E)}\right\}^{\otimes 2}\right]\widehat{S}_k^{(0)}(t;\widehat{\beta}_E)\mathrm{d}\widehat{\Lambda}_{0k}(t),$$

$$\widehat{B}_1(\widehat{\beta}_E) = \frac{1}{n}\sum_{k=1}^K \sum_{l=1}^K \sum_{i\in V_k}\sum_{j\in V_l} \widehat{\mu}_{ik}(\widehat{\beta}_E)\widehat{\mu}'_{jl}(\widehat{\beta}_E),$$

$$\widehat{B}_2(\widehat{\beta}_E) = \frac{1}{n}\sum_{k=1}^K \sum_{l=1}^K \sum_{i\in \bar{V}_k}\sum_{j\in \bar{V}_l} \widehat{\nu}_{ik}(\widehat{\beta}_E)\widehat{\nu}'_{jl}(\widehat{\beta}_E),$$

$$\widehat{\mu}_{jk}(\widehat{\beta}_E) = \int_0^\tau \left[\frac{\varphi_{jk}^{(1)}(t;\widehat{\beta}_E)}{\varphi_{jk}(t;\widehat{\beta}_E)} - \frac{\widehat{S}_k^{(1)}(t;\widehat{\beta}_E)}{\widehat{S}_k^{(0)}(t;\widehat{\beta}_E)}\right]\mathrm{d}\widehat{M}_{jk}(t;\widehat{\beta}_E) - \frac{n-n_k}{n_k}\widehat{H}_{jk}(\widehat{\beta}_E),$$

$$\widehat{\nu}_{jk}(\widehat{\beta}_E) = \int_0^\tau \left[\frac{\widehat{\phi}^{(1)}_{jk}(t;\widehat{\beta}_E)}{\widehat{\phi}_{jk}(t;\widehat{\beta}_E)} - \frac{\widehat{S}^{(1)}_k(t;\widehat{\beta}_E)}{\widehat{S}^{(0)}_k(t;\widehat{\beta}_E)} \right] \mathrm{d}\widehat{M}_{jk}(t;\widehat{\beta}_E),$$

$$\widehat{H}_{jk}(t;\widehat{\beta}_E) = \int_0^\tau \left[\frac{\widehat{\phi}^{(1)}_{jk}(t;\widehat{\beta}_E)}{\widehat{\phi}_{jk}(t;\widehat{\beta}_E)} - \frac{\widehat{S}^{(1)}_k(t;\widehat{\beta}_E)}{\widehat{S}^{(0)}_k(t;\widehat{\beta}_E)} \right] (\varphi_{jk}(t;\widehat{\beta}_E) - \widehat{\phi}_{jk}(t;\widehat{\beta}_E)) Y_{jk}(t) \mathrm{d}\widehat{\Lambda}_{0k}(t),$$

$$\mathrm{d}\widehat{M}_{jk}(t;\widehat{\beta}_E) = \mathrm{d}N_{jk}(t) - Y_{jk}(t)\widehat{r}_{jk}(t;\widehat{\beta}_E)\mathrm{d}\widehat{\Lambda}_{0k}(u).$$

Define the baseline stochastic processes $W_n(t) = \left(\widehat{\Lambda}_{01}(t) - \Lambda_{01}(t), \ldots, \widehat{\Lambda}_{0K}(t) - \Lambda_{0K}(t)\right)'$. Let $\mathcal{D}[0,\tau]^K$ be a space consisting of right-continuous functions $\{a_1(t), \ldots, a_K(t)\}^T$ with left limits, where $a_k(t) : [0,\tau] \to R$ for $k = 1, \ldots, K$. Make $\mathcal{D}[0,\tau]^K$ a metric space by equipping it with the metric $\rho_K(\mathbf{a}, \mathbf{b}) = \max_{t\in[0,\tau]}\{|a_k(t) - b_k(t)|, 1 \le k \le K\}$ for $\mathbf{a}, \mathbf{b} \in \mathcal{D}[0,\tau]^K$. The essential asymptotic results for the baseline cumulative hazard function estimator are summarized by the following theorem.

Theorem 2. *Under conditions* C1–C6, *$W_n(t)$ converges uniformly in $t \in [0,\tau]$ to zero in probability, and $\sqrt{n}W_n(t)$ converges weakly to a zero-mean Gaussian random field $\mathcal{W}(t)$ in $\mathcal{D}[0,\tau]^K$, where $\mathcal{W}(t) = (\mathcal{W}_1(t), \ldots, \mathcal{W}_K(t))'$. The covariance function between $\mathcal{W}_j(t)$ and $\mathcal{W}_k(t)$ is $\xi_{jk}(s,t) = E(\Phi_{1j}(s)\Phi_{1k}(t))$, where*

$$\Phi_{ik}(t) = \int_0^t \frac{\mathrm{d}M_{ik}(u)}{s_k(u;\beta_0)} - \left[\int_0^t \frac{s^{(1)}_k(u;\beta_0)}{s_k(u;\beta_0)} \mathrm{d}\Lambda_{0k}(u) \right]' A^{-1}(\beta_0) \sum_{l=1}^K (\eta_{il}\mu_{il}(\beta_0) + \bar{\eta}_{il}\nu_{il}(\beta_0)).$$

Furthermore, $\xi_{jk}(s,t)$ can be consistently estimated by $\widehat{\xi}_{jk}(s,t) = \frac{1}{n}\sum_{i=1}^n \widehat{\Phi}_{ij}(s)\widehat{\Phi}_{ik}(t)$, where

$$\widehat{\Phi}_{ik}(t) = \int_0^t \frac{\mathrm{d}\widehat{M}_{ik}(u;\widehat{\beta}_E)}{\widehat{S}_k(u;\widehat{\beta}_E)} - \left[\int_0^t \frac{\widehat{S}^{(1)}_k(u;\widehat{\beta}_E)}{\widehat{S}_k(u;\widehat{\beta}_E)} \mathrm{d}\widehat{\Lambda}_{0k}(u) \right]' \widehat{A}^{-1}(\widehat{\beta}_E) \sum_{l=1}^K \left(\eta_{il}\widehat{\mu}_{il}(\widehat{\beta}_E) + \bar{\eta}_{il}\widehat{\nu}_{il}(\widehat{\beta}_E)\right).$$

4. Simulation study

Simulation studies are conducted to evaluate the finite sample performance of the proposed estimator ($\widehat{\beta}_E$). The proposed estimator is compared with the following three estimators: (a) The full-data estimate ($\widehat{\beta}_F$) given by maximizing (4) with X observed for all the subjects under study, which is the well known pseudo-partial likelihood estimator [12]; (b) The complete case estimate ($\widehat{\beta}_C$), which is the pseudo-partial likelihood estimator based only on the validation set; (c) The naive estimate ($\widehat{\beta}_N$), which is the pseudo-partial likelihood estimator by substituting the auxiliary covariate W for the covariate X when X is unobserved. Obviously, $\widehat{\beta}_F$ can only be calculated in simulation studies because X is only observed for subjects in the validation set in real studies.

We simulate $K = 2$ failure types with the two baseline hazard functions being both 1. The partially observed covariates X_{i1} and X_{i2} are generated from the uniform distribution $U[0,1]$. The completely observed covariate (Z_{i1}, Z_{i2}) follows a bivariate normal distribution with marginal mean 0, standard deviation 1, and $Corr(Z_{i1}, Z_{i2}) = 0.8$. The multivariate failure times $(\tilde{T}_{i1}, \tilde{T}_{i2})$ are generated from the commonly used multivariate Clayton and Cuzick distribution [18], with the joint survival function as:

$$S(t_1, t_2; Z_1, Z_2; X_1, X_2) = \left\{ \sum_{k=1}^2 \exp(\theta^{-1} t_k \mathrm{e}^{\beta' D_k}) - 1 \right\}^{-\theta},$$

where $\beta = (\beta_1, \beta_2)'$ and $D_k = (X_k, Z_k)'$. The positive parameter θ controls the degree of dependence between $(\tilde{T}_1, \tilde{T}_2)$, with $\theta \to \infty$ corresponding to independence and $\theta \to 0$ to the increasing positive correlation. Simple calculations show that the failure times (t_1, t_2) are generated through:

$$t_1 = -\mathrm{e}^{-\beta' D_1} \times \ln(1 - u_1),$$

$$t_2 = \mathrm{e}^{-\beta' D_2} \ln[1 - a_1 + a_1(1 - u_2)^{-(\theta+1)^{-1}}],$$

where u_1 and u_2 are independently generated from $U(0,1)$ and $a_1 = \mathrm{e}^{-\theta^{-1} t_1 \mathrm{e}^{\beta' D_1}}$. We set $\theta = 0.25$ to represent a strong dependence between two type failure times. For $k = 1, 2$, the auxiliary covariate W_k is defined as

$$W_k = X_k + e_k,$$

where $e_k \sim N(0, \sigma^2)$ is a random error and σ is the parameter that controls the strength of the association between W_k and X_k. As σ increases, W_k becomes less informative about X_k. Censoring times are generated from $U(0, C)$, where C is chosen to yield a censoring rate, approximately being 80%.

Noting that the covariates $X = (X_1, X_2)$ and $Z = (Z_1, Z_2)$ are generated independently, we could let $Z^*_k = W_k$ in (5). We use the Epanechnikov kernel function [19] with bandwidth $B_k = (b_{1k})$, where $b_{1k} = 2\widehat{\sigma}_{W_k} n_k^{-1/3}$ and $\widehat{\sigma}_{W_k}$ is the sample standard deviation of W_k in the kth marginal validation set with sample size being n_k. Note that B_k satisfies the bandwidth conditions C5 in Section 3. We use the approach to interpolating the neighboring points of continuous Z_k^* as suggested by Zhou and Wang [6] for the small sample correction.

The true parameter values in our simulation studies are $\beta_1 = 0.693$ and $\beta_2 = -0.2$. We set $\sigma = 0.1$ and 0.6, which lead to the correlation coefficient between X and W approximately being 0.94 and 0.43, representing the high and low informative strength of W for X, respectively. The validation set proportion studied are ρ of 30% and 50%. The number of independent groups are $n = 300$ and $n = 600$. For each simulation configuration, 1000 replicated samples are generated. The sample standard deviations of the 1000 estimates are given in the corresponding SD columns. The SE columns give the average of the estimated standard errors. The coverage probabilities of the 95% confidence intervals for the true parameters using the estimated standard errors are listed in the CI columns.

Table 1 displays the simulation results. All four estimators for β_2 work well in all considered scenarios. The proposed estimator for β_2 is shown to be more efficiency than the complete case estimator $\widehat{\beta}_C$. For the estimators of β_1, we have the following observations: (i) The naive estimator $\widehat{\beta}_N$ is biased and the bias increases as the association between W and X becomes weaken (i.e. σ increases). The proposed estimator $\widehat{\beta}_E$ corrects this bias very well, especially when the sample size is not too small. For example, when $\sigma = 0.6$, which corresponds to the situation that W is less informative about X, $\widehat{\beta}_E$ is less accurate in estimating β_1 when group size $n = 300$ as indicated by the amount of biases shown in the β_1 estimates. However, this bias is deducted drastically when we increase n to 600. (ii) The standard deviation (SD) of the proposed estimator is always smaller than that of the complete case estimator, and is not much larger than that of the full-data estimator when $\sigma = 0.1$. Specifically, when $n = 300$ and $\sigma = 0.1$, the relative efficiency of the proposed estimator versus the full-data estimator, which is defined as the ratios of empirical variance of the full-data estimator to that of the proposed estimator, are 0.90 and 0.975 at 30% and 50% validation fraction, respectively. (iii) The relative efficiency of proposed estimator over the complete case estimator is higher when σ is small or the validation fraction is low. For example, the relative efficiency of $\widehat{\beta}_E$ over $\widehat{\beta}_C$ for validation fraction being 0.3 and 0.5 are 3.35 and 2.19 at $\sigma = 0.1$, 1.76 and 1.35 at $\sigma = 0.6$, respectively. This suggests that the proposed method is more beneficial compared with the complete case estimator when used in situations with a small validation fraction or a high informative auxiliary covariate. (iv) The estimated standard errors (SE) are very close to the true standard deviations (SD), and the coverage probabilities of the 95% confidence intervals also suggest that the asymptotic approximations in the sample sizes considered are of satisfactory.

5. Analysis of the SOLVD study

In this section, we illustrate the proposed method with a data set from the Study of Left Ventricular Dysfunction (SOLVD) [3]. This data has been previously analyzed by several authors, including Liu, Zhou and Cai [11]. The SOLVD study was a randomized, double-masked, and placebo-controlled trial between 1986 and 1991. The trial had a three-year recruitment and a two-year follow-up. The basic inclusion criteria for the prevention trial were: age between 21 and 80 years, inclusive, no overt symptoms of congestive heart failure, and left ventricular ejection fraction less than 35 percent. Ejection fraction is a number between 0 and 100 that measures the efficiency of the heart in ejecting blood. A total of 4228 patients with asymptomatic left ventricular dysfunction were randomly assigned to receive either enalapril or placebo at one of the 83 hospitals linked to 23 centers in the United States, Canada, and Belgium.

The primary clinical issues of interest are the effects of covariates on the risk of heart failure and on the first myocardial infarction (MI) after adjusting for the confounding variables. The covariates of interest are ejection fraction, patient's gender (SEX), which is coded 1 for male and 0 for female; treatment (TRT), which is coded as 1 for enalapril and 0 for placebo, and patient's age (AGE), which was measured in years. The average age of the patients is 59 years old with a standard deviation of 10 years. The covariates of SEX, AGE, and TRT were recorded for almost all of the subjects, but only 108 among the total of 4228 patients have their ejection fraction accurately measured using a standardized radionucleotide technique (LVEF). A related nonstandardized measure (EF) was, however, ascertained for all the patients. Therefore, the nonstandardized measure (EF) can be used as the auxiliary for the standardized measure for ejection fraction (LVEF) in this case. Both LVEF and EF are continuous quantities measured in percentage. One way is to discrete the EF variable through its quartiles as Liu, Zhou and Cai [11], however, discretizing the continuous variable may lead to loss of efficiency because a lower order scale is used in the analysis. Hence, our proposed method is required to handle the continuous auxiliary covariate in the multivariate failure times in the SOLVD study to investigate the effect of LVEF on both the risk of heart failure and the non-fatal MI adjusted by other confounding variables.

Let k denote failure type with $k = 1$ for heart failure and $k = 2$ for non-fatal MI and i denote the patient with $i = 1, \ldots, 4228$. Considering that the effects of the covariates on different failure times may be different, we set $X_{ik} = (LVEF_{ik}, LVEF_{ik} * I(k = 2))'$, $W_{ik} = EF_{ik}$, $Z_{ik} = (TRT_{ik}, SEX_{ik}, AGE_{ik}, TRT_{ik} * I(k = 2), SEX_{ik} * I(k = 2), AGE_{ik} * I(k = 2))'$ in terms of the notation introduced in the previous sections. Denote by $\alpha = (\alpha_1', \alpha_2')'$ the unknown regression coefficients, where $\alpha_1 = (\beta_1, \gamma_1)'$ and $\alpha_2 = (\beta_2, \beta_3, \beta_4, \gamma_2, \gamma_3, \gamma_4)'$.

We use the following marginal hazards model to fit the SOLVD data:

$$\lambda_{ik}(t|X_{ik}, Z_{ik}, W_{ik}) = \lambda_{0k}(t) r_{ik}(\alpha, t), \tag{9}$$

where

$$r_{ik}(\alpha, t) = \begin{cases} \exp(\alpha_1' X_{ik} + \alpha_2' Z_{ik}) & \text{when } X_{ik} \text{ is observed}, \\ \phi_{ik}(\alpha, t) & \text{when } X_{ik} \text{ is missing}, \end{cases}$$

Table 1
Simulation results comparing $\widehat{\beta}_E$, $\widehat{\beta}_N$, $\widehat{\beta}_C$ and $\widehat{\beta}_F$ with censoring rate 80% based on 1000 replications. ρ is the validation fraction[a].

			$\beta_1 = 0.693$				$\beta_2 = -0.2$			
ρ	σ	Method	Mean	SD	SE	CI	Mean	SD	SE	CI
			$n = 300$							
1	–	$\widehat{\beta}_F$	0.695	0.322	0.322	0.950	−0.197	0.091	0.093	0.950
0.3	–	$\widehat{\beta}_C$	0.722	0.622	0.601	0.939	−0.189	0.173	0.172	0.944
	0.1	$\widehat{\beta}_N$	0.642	0.302	0.310	0.958	−0.202	0.093	0.093	0.949
		$\widehat{\beta}_E$	0.692	0.340	0.348	0.960	−0.200	0.092	0.095	0.955
	0.6	$\widehat{\beta}_N$	0.165	0.163	0.158	0.084	−0.197	0.091	0.093	0.957
		$\widehat{\beta}_E$	0.562	0.469	0.462	0.936	−0.195	0.090	0.095	0.965
0.5	–	$\widehat{\beta}_C$	0.709	0.482	0.461	0.941	−0.194	0.132	0.132	0.943
	0.1	$\widehat{\beta}_N$	0.663	0.303	0.313	0.950	−0.201	0.093	0.093	0.950
		$\widehat{\beta}_E$	0.700	0.326	0.337	0.957	−0.199	0.092	0.094	0.955
	0.6	$\widehat{\beta}_N$	0.213	0.179	0.178	0.224	−0.197	0.091	0.093	0.956
		$\widehat{\beta}_E$	0.630	0.415	0.405	0.942	−0.196	0.090	0.094	0.958
			$n = 600$							
1	–	$\widehat{\beta}_F$	0.700	0.225	0.227	0.950	−0.198	0.064	0.065	0.956
0.3	–	$\widehat{\beta}_C$	0.707	0.416	0.418	0.951	−0.198	0.124	0.120	0.939
	0.6	$\widehat{\beta}_N$	0.172	0.113	0.112	0.005	−0.197	0.064	0.065	0.955
		$\widehat{\beta}_E$	0.673	0.412	0.417	0.956	−0.196	0.064	0.067	0.961
0.5	–	$\widehat{\beta}_C$	0.703	0.279	0.272	0.944	−0.199	0.076	0.078	0.949
	0.6	$\widehat{\beta}_N$	0.298	0.148	0.146	0.236	−0.197	0.064	0.065	0.957
		$\widehat{\beta}_E$	0.698	0.278	0.272	0.946	−0.195	0.064	0.066	0.957

[a] The marginal models $\lambda_k(t; X_k, Z_k) = \exp(\beta_1 X_k + \beta_2 Z_k)(k = 1, 2)$ with $\beta_1 = \log(2) \doteq 0.693$ and $\beta_2 = -0.2$, where X_1 and X_2 are generated independently from the $U(0, 1)$ and (Z_1, Z_2) is from a multi-normal distribution with mean zero, unit standard error and $Corr(Z_1, Z_2) = 0.8$. The auxiliary variable $W_k = X_k + e_k$, where $e_k \sim N(0, \sigma^2)$. $\widehat{\beta}_E$ = proposed estimator, $\widehat{\beta}_C$ = complete case estimator, $\widehat{\beta}_N$ = naive estimator, $\widehat{\beta}_F$ = full-data estimator. '–' represents the corresponding method does not depend on the σ. Mean is the sample mean of the estimator $\widehat{\beta}$, SD is the sampling standard deviation of $\widehat{\beta}$, SE is the sampling mean of the standard error estimator, and CI is the coverage probability of the 95% confidence interval.

Table 2
SOLVD data analysis results: The proposed method versus the complete case method.

	Proposed method				Complete case method			
Covariate	Est.	exp(Est.)	SE	p-value	Est.	exp(Est.)	SE	p-value
Heart failure								
LVEF (β_1)	−0.064	0.938	0.003	<0.001	−0.075	0.928	0.038	0.051
TRT (β_2)	−0.530	0.589	0.066	<0.001	−0.835	0.434	0.564	0.140
SEX (β_3)	−0.283	0.753	0.115	0.014	0.405	1.499	1.089	0.710
AGE (β_4)	0.026	1.026	0.002	<0.001	0.035	1.035	0.033	0.300
Non-fatal MI								
LVEF ($\beta_1 + \gamma_1$)	−0.121	0.886	0.058	0.038	−0.012	0.988	0.042	0.780
TRT ($\beta_2 + \gamma_2$)	−0.382	0.682	0.526	0.467	−0.703	0.495	0.875	0.420
SEX ($\beta_3 + \gamma_3$)	−0.043	0.958	1.769	0.981	−0.838	0.433	1.058	0.430
AGE ($\beta_4 + \gamma_4$)	0.012	1.012	0.153	0.938	0.003	1.003	0.032	0.920

with $\phi_{ik}(\alpha, t) = E\{e^{\alpha_1' X_{ik}(t)} | Y_{ik}(t) = 1, W_{ik}(t), Z_{ik}(t)\} \exp(\alpha_2' Z_{ik}(t))$. Liu, Zhou and Cai [11] checked the conditional dependence structure for SOLVD data and found that LVEF is conditional independent of TRT, SEX, and AGE, giving EF. Therefore, when X_{ik} is missing, $\phi_{ik}(\alpha, t)$ can be regarded as $E\{e^{\alpha_1' X_{ik}(t)} | Y_{ik}(t) = 1, W_{ik}(t)\} \exp(\alpha_2' Z_{ik}(t))$ and thus can be estimated by (5) with $Z_{ik}^* = W_{ik}$. We use Epanechnikov kernel with the bandwidth $b_{1k} = 2\widehat{\sigma}_k n_k^{-1/3}$ with $n_k = 108$ being the number of subjects in validation set and $\widehat{\sigma}_k$ being the sample standard deviation of W_{ik}'s in the kth marginal validation set.

Table 2 presents the results of data analysis. The proposed estimated pseudo-partial likelihood method is compared with the complete case pseudo-partial likelihood method. The proposed method utilizes the information contained in both the auxiliary (EF) for all subjects and data in the non-validation set while the complete case analysis relies only on validation set with the available true ejection fraction (LVEF).

The p-values in Table 2 indicate that at 0.05 significance level, LVEF, TRT, SEX, and AGE all have statistically significant effects on the heart failure under the proposed method, while only the effect of LVEF on the heart failure is approximately significant under the complete case analysis. The results from the proposed method also indicate that the effects of TRT, SEX, and AGE are different on the heart failure and on the non-fatal MI. Specifically, TRT, SEX, and AGE do not seem to affect the risk of non-fatal MI, but is related to the risk of heart failure.

The variance for coefficient α is estimated using the proposed method. For estimating the effect of LVEF on heart failure, there are substantial efficiency gains by using the proposed approach over the complete case analysis. The proposed method provides more precise 95% confidence intervals for the effect of covariates on heart failure. For instance, the 95% confidence interval for effect of LVEF on heart failure is (−0.070, −0.058) for the proposed method, and (−0.149, −0.001) for the complete case method.

Utilizing the auxiliary information in the proposed method, we had in effect regained the statistical power of the study that would have been lost had one conducted the complete case analysis. Therefore, our proposed method is more efficient and applicable than the complete case method.

6. Concluding remarks

In this paper, we studied an estimated pseudo-partial likelihood approach for multivariate survival data with continuous auxiliary. The proposed method is based on the kernel smoother technique and is nonparametric with respect to the association between the unobserved primary covariate and available auxiliary variable. Simulation studies demonstrate that the proposed estimator works well in the scenarios considered and outperforms the estimator which uses only data from the validation set. The real data example also indicates that a much more precise estimator can be obtained by incorporating the auxiliary covariate information into the statistical inference. Hence, the proposed method can achieve more statistical efficiency than what would be gained by using only the validation set.

We have a couple of cautionary notes on the limitations of the proposed method. First, for the reason of curse of dimensionality, the proposed method will not work well if the dimension of kernel smoother is high (e.g. $d > 3$). One possible way to avoid the potential issue due to the high dimensionality is to create the auxiliary using a predicted model with possible multiple predictors so that one-dimensional kernel smoother is applicable, or to employ some techniques (e.g. introducing some additive structure) in the dimension reduction. Second, when the auxiliary W is less informative about X, the proposed estimator tends to underestimate the true parameter. Increasing the fraction of the validation set sample or the total sample size can help to alleviate this problem as shown in the simulation study.

Although frequently used bandwidth is adopted in both our simulation study and SOLVD data analysis, it is valuable to consider some bandwidth selection criteria such as the generalized cross validation (GCV). On the other hand, Cai and Prentice [20] showed that more efficient estimator for the regression parameter could be obtained by introducing weights into the pseudo-partial likelihood score equations. In the sequel, we will introduce suitable weights to our proposed method to further improve the efficiency of estimator.

Finally, since the additive hazards model is an alternative for Cox's proportional hazards model when it is unsuitable to fit the data, therefore it is worthwhile to consider the nonparametric kernel smoothing method in the additive hazards model when there exists auxiliary covariate. The related investigations are currently under way.

Acknowledgments

The authors are very grateful for the valuable comments and suggestions from the Associate Editor and the referees. This research was supported by National Nature Science Fund of China (10771163) (Y. Liu) and U.S. NIH grants R01 CA 079949 (Y. Wu and H. Zhou).

Appendix. Proofs of the asymptotic properties

In the following, we use notation $\to_p$, $\to_{a.s.}$, and $\to_d$ to denote the convergence in probability, convergence in probability 1, and convergence in distribution, respectively.

Denote the determinant of matrix B_k by $|B_k|$. Let $\mathbf{b}_k = (b_{1k}, \ldots, b_{dk})'$ and $\alpha^*_{0k} = (\alpha^*_{0k1}, \ldots, \alpha^*_{0kd})$ such that each α^*_{0kj} is non-negative integer and $|\alpha^*_{0k}| \equiv \sum_{j=1}^d \alpha^*_{0kj} = \alpha_{0k}$. For l-vector $\mathbf{u} = (u_1, \ldots, u_l)$ and $\mathbf{v} = (v_1, \ldots, v_l)$, define $\mathbf{u}^{\mathbf{v}} = \prod_{j=1}^l {u_l}^{v_l}$ and $\mathbf{uv} = (u_1v_1, \ldots, u_lv_l)$.

Proof of Lemma 1. Let $\mathcal{Z}^*_k$ be the domain of the process $Z^*_k(t)$, $t \in [0, \tau]$. Let $\mathbf{P}_n$ be the empirical measure from the n i.i.d. observations and $\mathbf{P}$ be the corresponding probability measure. For fixed $z^*_{k;0} \in \mathcal{Z}^*_k$, let

$$L^k_n(t, \beta_1, z^*_{k;0}) = \mathbf{P}_n\left[|B_k|^{-1}\eta_k Y_k(t) Q_k\left\{B_k^{-1}\{Z^*_k(t) - z^*_{k;0}\}\exp\{\beta_1' X_k(t)\}\right\}\right].$$

Note that the stochastic processes $Z^*_k(t)$, $X_k(t)$, $Y_k(t)$, and $\exp\{\beta_1' X_k(t)\}$ have bounded total variation over $t \times \beta_1 \in [0, \tau] \times \mathcal{B}_1$. It follows from the Lemma 9.10 of Kosorok [21] that they are all VC-subgraph with finite VC-index. Using the similar arguments used in Yin, Li and Zeng [22], we can obtain that

$$\sup_{\Theta_k}\left|L^k_n(t, \beta_1, z^*_{k;0}) - \mathbf{P}\left[|B_k|^{-1}\eta_k Y_k(t) Q_k\left\{B_k^{-1}\{Z^*_k(t) - z^*_{k;0}\}\exp\{\beta_1' X_k(t)\}\right\}\right]\right|$$
$$= O_p\left(\frac{\log n}{\sqrt{n}\|B_k\|^d}\right),$$

where $\Theta_k = [0, \tau] \times \mathcal{B}_1 \times \mathcal{Z}^*_k$.

On the other hand, by the Taylor expansion,

$$\begin{aligned}
&\mathbf{P}\left[|B_k|^{-1}\eta_k Y_k(t) Q_k\left\{B_k^{-1}\{Z_k^*(t)-z_{k;0}^*\}\exp\{\beta_1' X_k(t)\}\right\}\right] = \int (b_{1k}\cdots b_{dk})^{-1} Q_k\{B_k^{-1}(\mathbf{v}-z_{k;0}^*)\} h_k(\mathbf{v},\mathbf{s})\exp(\beta_1'\mathbf{s})\mathrm{d}\mathbf{v}\mathrm{d}\mathbf{s}\\
&= \int Q_k(\mathbf{u}) h_k(\mathbf{b}_k\mathbf{u}+z_{k;0}^*,\mathbf{s})\exp(\beta_1'\mathbf{s})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{s}\\
&= \int Q_k(\mathbf{u})\sum_{|\alpha|=0}^{\alpha_{0k}-1}\frac{D_{\mathbf{v}}^{\alpha} h_k(z_{k;0}^*,\mathbf{s})}{\alpha!}\mathbf{b}_k^{\alpha}\mathbf{u}^{\alpha}\exp(\beta_1'\mathbf{s})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{s} + \left(\int \frac{D_{\mathbf{v}}^{\alpha_{0k}^*} h_k(z_{k;0}^{*\ddagger},\mathbf{s})}{\alpha_{0k}!}\mathbf{u}^{\alpha_{0k}^*} Q_k(\mathbf{u})\exp(\beta_1'\mathbf{s})\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{u}\right)\mathbf{b}_k^{\alpha_{0k}^*}\\
&= \int h_k(z_{k;0}^*,\mathbf{s})\exp(\beta_1'\mathbf{s})\mathrm{d}\mathbf{s} + O(\mathbf{b}_k^{\alpha_{0k}^*})\\
&= g_k(z_{k;0}^*)\int \frac{h_k(z_{k;0}^*,\mathbf{s})}{g_k(z_{k;0}^*)}\exp(\beta_1'\mathbf{s})\mathrm{d}\mathbf{s} + O(\|B_k\|^{\alpha_{0k}})\\
&= g_k(z_{k;0}^*) E\{\exp(\beta_1' X_k(t)) \mid Y_k(t)=1, Z_k^*(t)=z_{k;0}^*\} + O(\|B_k\|^{\alpha_{0k}}),
\end{aligned}$$

where $D_{\mathbf{v}}^{\alpha} h_k(\mathbf{v},\mathbf{s})$ is the αth order derivative of $h_k(\mathbf{v},\mathbf{s})$ respect to $\mathbf{v}$, $g_k(z_{k;0}^*)$ is the density function of $(Y_k(t)=1, Z_k^*(t)=z_{k;0}^*)$, and $z_{k;0}^{*\ddagger}$ is on the line segment of $z_{k;0}^*$ and $\mathbf{b}_k\mathbf{u}$.

Therefore, we can conclude that by condition C5,

$$\sup_{\Theta_k}\left|L_n^k(t,\beta_1,z_{k;0}^*) - g_k(z_{k;0}^*)E\{\exp(\beta_1' X_k(t)) \mid Y_k(t)=1, Z_k^*(t)=z_{k;0}^*\}\right| \to_p 0.$$

On the other hand, let $\beta_1=0$, then we can obtain that

$$\sup_{\Theta_k}\left|L_n^k(t,0,z_{k;0}^*) - g_k(z_{k;0}^*)\right| \to_p 0.$$

Consequently, it follows straightforwardly that

$$\sup_{\Theta_k}\left|\frac{L_n^k(t,\beta_1,z_{k;0}^*)}{L_n^k(t,0,z_{k;0}^*)} - E\{\exp(\beta_1' X_k(t)) \mid Y_k(t)=1, Z_k^*(t)=z_{k;0}^*\}\right| \to_p 0.$$

Furthermore, note that

$$\widehat{\phi}_{ik}(t;\beta) = \frac{L_n^k(t,\beta_1,Z_{ik}^*(t))}{L_n^k(t,0,Z_{ik}^*(t))}\exp(\beta_2' Z_{ik}(t)),$$

then

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}}\left|\widehat{\phi}_{ik}(t;\beta) - \phi_{ik}(t;\beta)\right| \to_p 0.$$

Similarly, for $a=1,2$, we can also prove that

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}}\|\widehat{\phi}_{ik}^{(a)}(t;\beta) - \phi_{ik}^{(a)}(t;\beta)\| \to_p 0. \tag{10}$$

Thus, we have the first half of the lemma proved. With respect to the second half part, we sketch the proof when $l=0$. The remaining conclusions can be proved in the similar way so they are omitted here.

First of all, by (10) and the definition of $\widehat{S}_k$ and S_k, we can prove

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}}\|\widehat{S}_k^{(0)}(t;\beta) - S_k^{(0)}(t;\beta)\| \to_p 0.$$

Second, by the uniform strong law of large numbers, we have

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}}\|S_k^{(0)}(t;\beta) - s_k^{(0)}(t;\beta)\| \to_{a.s.} 0.$$

It follows directly that

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}}\|\widehat{S}_k^{(0)}(t;\beta) - s_k^{(0)}(t;\beta)\| \to_p 0.$$

Thus, the proof of Lemma 1 has been done. □

Proof of Theorem 1. Consistency

Note that $\widehat{\beta}_E$ solves $n^{-1}\widehat{U}(\beta) = 0$. Following closely the arguments of Foutz [17], one can show that $\widehat{\beta}_E$ is consistent for β_0, provided that:

(I) $n^{-1}\partial\widehat{U}(\beta)/\partial\beta$ exists and is continuous in an open neighborhood B of β_0;
(II) $n^{-1}\partial\widehat{U}(\beta)/\partial\beta$ converges in probability to a fixed function, say, $H(\beta)$, uniformly in an open neighborhood of β_0; Furthermore, every element of $H(\beta)$ is a continuous function of β in the neighborhood of β_0 and $H^{-1}(\beta_0)$ exists;
(III) $n^{-1}\partial\widehat{U}(\beta_0)/\partial\beta$ is negative-definite with probability going to 1;
(IV) $n^{-1}\widehat{U}(\beta_0) \to_p 0$.

Obviously, (I) is satisfied. To verify (II), define

$$\widehat{U}_k(t;\beta) = \sum_{i=1}^{n}\int_0^t \left[\frac{\widehat{r}_{ik}^{(1)}(u;\beta)}{\widehat{r}_{ik}(u;\beta)} - \frac{\widehat{S}_k^{(1)}(u;\beta)}{\widehat{S}_k^{(0)}(u;\beta)}\right]\mathrm{d}N_{ik}(u),$$

then from (8), $\widehat{U}(\beta) = \sum_{k=1}^{K}\widehat{U}_k(\tau;\beta)$. After simple algebraic manipulations, we obtain that

$$\frac{\partial\widehat{U}_k(t;\beta)}{\partial\beta} = \sum_{i=1}^{n}\int_0^t \left[\frac{\widehat{r}_{ik}^{(2)}(u;\beta)}{\widehat{r}_{ik}(u;\beta)} - \left\{\frac{\widehat{r}_{ik}^{(1)}(u;\beta)}{\widehat{r}_{ik}(u;\beta)}\right\}^{\otimes 2} - \frac{\widehat{S}_k^{(2)}(u;\beta)}{\widehat{S}_k^{(0)}(u;\beta)} + \left\{\frac{\widehat{S}_k^{(1)}(u;\beta)}{\widehat{S}_k^{(0)}(u;\beta)}\right\}^{\otimes 2}\right] dN_{ik}(u).$$

Define

$$C_k(t;\beta) = \sum_{i=1}^{n}\int_0^t \left[\frac{\widehat{r}_{ik}^{(2)}(u;\beta)}{\widehat{r}_{ik}(u;\beta)} - \left\{\frac{\widehat{r}_{ik}^{(1)}(u;\beta)}{\widehat{r}_{ik}(u;\beta)}\right\}^{\otimes 2} - \frac{\widehat{S}_k^{(2)}(u;\beta)}{\widehat{S}_k^{(0)}(u;\beta)} + \left\{\frac{\widehat{S}_k^{(1)}(u;\beta)}{\widehat{S}_k^{(0)}(u;\beta)}\right\}^{\otimes 2}\right] \times Y_{ik}(u)r_{ik}(u;\beta_0)\lambda_{0k}(u)\mathrm{d}u,$$

then

$$\frac{1}{n}\frac{\partial\widehat{U}_k(t;\beta)}{\partial\beta} - \frac{1}{n}C_k(t;\beta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[\frac{\widehat{r}_{ik}^{(2)}(u;\beta)}{\widehat{r}_{ik}(u;\beta)} - \left\{\frac{\widehat{r}_{ik}^{(1)}(u;\beta)}{\widehat{r}_{ik}(u;\beta)}\right\}^{\otimes 2} - \frac{\widehat{S}_k^{(2)}(u;\beta)}{\widehat{S}_k^{(0)}(u;\beta)} + \left\{\frac{\widehat{S}_k^{(1)}(u;\beta)}{\widehat{S}_k^{(0)}(u;\beta)}\right\}^{\otimes 2}\right]\mathrm{d}M_{ik}(u)$$

is a square integrable martingale, which converges in probability to zero uniformly in $\beta \in \mathcal{B}$ by Lenglart's inequality [23]. Thus, $\frac{1}{n}\frac{\partial\widehat{U}_k(t;\beta)}{\partial\beta}$ converges in probability to the same limit as does $\frac{1}{n}C_k(t;\beta)$, uniformly in $\beta \in \mathcal{B}$. It follows that $\frac{1}{n}\frac{\partial\widehat{U}(t;\beta)}{\partial\beta}$ and $\frac{1}{n}\sum_{k=1}^{K}C_k(t;\beta)$ converges in probability to the same limit. Let $H(\beta)$ denote the uniformly convergence limit of $\frac{1}{n}\sum_{k=1}^{K}C_k(\tau;\beta)$. On the other hand, by Lemma 1, we can show that

$$H(\beta) = \sum_{k=1}^{K}\int_0^\tau \left\{s_k^{(3)}(t;\beta) - s_k^{(4)}(t;\beta) - \left[\frac{s_k^{(2)}(t;\beta)}{s_k^{(0)}(t;\beta)} - \left\{\frac{s_k^{(1)}(t;\beta)}{s_k^{(0)}(t;\beta)}\right\}^{\otimes 2}\right]s_k^{(0)}(t;\beta_0)\right\}\lambda_{0k}(t)\mathrm{d}t.$$

Hence,

$$\sup_{\beta\in\mathcal{B}}\left\|\frac{1}{n}\frac{\partial\widehat{U}(\beta)}{\partial\beta} - H(\beta)\right\| \to_p 0. \tag{11}$$

Note that $s_k^{(3)}(t;\beta_0) = s_k^{(2)}(t;\beta_0)$, then $H(\beta_0) = -A(\beta_0)$ is just as what we defined in condition C4 and is negative-definite. Thus, (II) is done. (III) is followed straightforwardly from (I), (II), and (11).

To prove (IV), we first rewrite

$$\widehat{U}_k(t;\beta) = \widehat{U}_{k:1}(t;\beta) + \widehat{U}_{k:2}(t;\beta),$$

where

$$\widehat{U}_{k:1}(t;\beta) = \sum_{i=1}^{n}\int_0^t \left[\frac{\widehat{r}_{ik}^{(1)}(u;\beta)}{\widehat{r}_{ik}(u;\beta)} - \frac{\widehat{S}_k^{(1)}(u;\beta)}{\widehat{S}_k^{(0)}(u;\beta)}\right]\mathrm{d}M_{ik}(u),$$

and

$$\widehat{U}_{k:2}(t;\beta) = \sum_{i=1}^{n}\int_0^t \left[\frac{\widehat{r}_{ik}^{(1)}(u;\beta)}{\widehat{r}_{ik}(u;\beta)} - \frac{\widehat{S}_k^{(1)}(u;\beta)}{\widehat{S}_k^{(0)}(u;\beta)}\right]Y_{ik}(u)r_{ik}(u;\beta)\lambda_{0k}(u)\mathrm{d}u.$$

By Lemma 1, the fact that $M_{ik}(t)$ is a square integrable martingale, and Lenglart's inequality [23], we can show that

$$\frac{1}{\sqrt{n}}\widehat{U}_{k:1}(\tau;\beta_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left[\frac{r_{ik}^{(1)}(u;\beta_0)}{r_{ik}(u;\beta_0)}-\frac{s_k^{(1)}(u;\beta_0)}{s_k^{(0)}(u;\beta_0)}\right]\mathrm{d}M_{ik}(u)+o_p(1).$$

By the arguments used in Zhou and Wang [6], it can be also shown that

$$\frac{1}{\sqrt{n}}\widehat{U}_{k:2}(\tau;\beta_0)=-\frac{1}{\sqrt{n}}\frac{n-n_k}{n_k}\sum_{j\in V_k}H_{jk}(\beta_0)+o_p(1).$$

Hence, we obtain that

$$\begin{aligned}\frac{1}{\sqrt{n}}\widehat{U}_k(\tau;\beta_0)&=\frac{1}{\sqrt{n}}\sum_{j\in V_k}\left\{\int_0^{\tau}\left[\frac{\varphi_{jk}^{(1)}(u;\beta_0)}{\varphi_{jk}(u;\beta_0)}-\frac{s_k^{(1)}(u;\beta_0)}{s_k^{(0)}(u;\beta_0)}\right]\mathrm{d}M_{jk}(u)-\frac{1-\rho_k}{\rho_k}H_{jk}(\beta_0)\right\}\\&\quad+\frac{1}{\sqrt{n}}\sum_{i\in\bar{V}_k}\int_0^{\tau}\left[\frac{\phi_{ik}^{(1)}(u;\beta_0)}{\phi_{ik}(u;\beta_0)}-\frac{s_k^{(1)}(u;\beta_0)}{s_k^{(0)}(u;\beta_0)}\right]\mathrm{d}M_{ik}(u)+o_p(1)\\&=\frac{1}{\sqrt{n}}\sum_{j\in V_k}\mu_{jk}(\beta_0)+\frac{1}{\sqrt{n}}\sum_{i\in\bar{V}_k}\nu_{ik}(\beta_0)+o_p(1).\end{aligned}\tag{12}$$

Note that $M_{ik}(t)$ is a martingale with mean zero and $E(H_{jk}(\beta_0))=0$, then by strong law of large numbers, we have $\frac{1}{n}\widehat{U}_k(\tau;\beta_0)\rightarrow_{a.s.}0$, so does $\frac{1}{n}\widehat{U}(\beta_0)\rightarrow_{a.s.}0$. Thus, (IV) is verified.

Hence, we have shown that $\widehat{\beta}_E$ converges in probability to β_0.

Asymptotic normality

By Taylor expansion of $\widehat{U}(\beta)$ around the true parameter β_0, we have

$$\sqrt{n}(\widehat{\beta}_E-\beta_0)=\left[-\frac{1}{n}\frac{\partial\widehat{U}(\beta^*)}{\partial\beta}\right]^{-1}\left[\frac{1}{\sqrt{n}}\widehat{U}(\beta_0)\right],\tag{13}$$

where β^* is on the line segment between $\widehat{\beta}_E$ and β_0.

By (11), the consistency of $\widehat{\beta}$, and the continuity of $\frac{1}{n}\frac{\partial\widehat{U}(\beta)}{\partial\beta}$, we have

$$\left[-\frac{1}{n}\frac{\partial\widehat{U}(\beta^*)}{\partial\beta}\right]^{-1}\rightarrow_p A^{-1}(\beta_0).\tag{14}$$

On the other hand, the two sums on the right-hand side of (12) are mutually independent since they are from validation set and non-validation set, respectively. Thus, by multivariate central limit theorem,

$$\frac{1}{\sqrt{n}}\widehat{U}(\beta_0)\rightarrow_d N(0,B_1(\beta_0)+B_2(\beta_0)).\tag{15}$$

Combining (13)–(15), we have that

$$\sqrt{n}(\widehat{\beta}_E-\beta_0)\rightarrow_d N(0,\Sigma(\beta_0)).\quad\square$$

Proof of Theorem 2. Note that

$$\sup_{t\in[0,\tau]}\left|\widehat{\Lambda}_{0k}(t)-\Lambda_{0k}(t)\right|\le\sup_{t\in[0,\tau]}\left|\int_0^t\frac{\frac{1}{n}\sum_{i=1}^{n}\mathrm{d}M_{ik}(u)}{\widehat{S}_k^{(0)}(u;\widehat{\beta}_W)}\right|+\sup_{t\in[0,\tau]}\left|\int_0^t\frac{S_k^{(0)}(u;\beta_0)-\widehat{S}_k^{(0)}(u;\widehat{\beta}_W)}{\widehat{S}_k^{(0)}(u;\widehat{\beta}_W)}\mathrm{d}\Lambda_{0k}(u)\right|.$$

The first term on the right-hand side of above inequality converges almost surely to zero by Lemma 1 and Lemma A.1 in the Appendix of Kulich and Lin [7]; Likewise, it can be shown that the second term is also asymptotically negligible. Therefore, $\widehat{\Lambda}_{0k}(t)$ converges almost surely to the cumulative hazard function $\Lambda_{0k}(t)$, uniformly in t.

By Lemma 1 and Theorem 1, it can be shown that

$$\sqrt{n}(\widehat{\Lambda}_{0k}(t)-\Lambda_{0k}(t))=\frac{1}{\sqrt{n}}\int_0^t\frac{\sum_{i=1}^{n}\mathrm{d}M_{ik}(u)}{\widehat{S}_k^{(0)}(u;\widehat{\beta}_W)}+\sqrt{n}\int_0^t\frac{S_k^{(0)}(u;\beta_0)-\widehat{S}_k^{(0)}(u;\widehat{\beta}_W)}{\widehat{S}_k^{(0)}(u;\widehat{\beta}_W)}\mathrm{d}\Lambda_{0k}(u)$$

$$= \frac{1}{\sqrt{n}} \int_0^t \frac{\sum_{i=1}^n \mathrm{d}M_{ik}(u)}{s_k^{(0)}(u;\beta_0)} - \left[\int_0^t \frac{s_k^{(1)}(u;\beta_0)}{s_k^{(0)}(u;\beta_0)} \mathrm{d}\Lambda_{0k}(u)\right]' \sqrt{n}(\widehat{\beta}_W - \beta_0) + o_p(1)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_{ik}(t) + o_p(1).$$

Thus

$$\sqrt{n}W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\Phi_{i1}(t), \ldots, \Phi_{iK}(t))' + o_p(1),$$

which converges in finite-dimensional distributions to a normal distribution with mean zero by the multivariate central limit theorem. If the tightness of $\sqrt{n}W_n(t)$ holds, we can conclude that $\sqrt{n}W_n(t)$ converges to a zero-mean Gaussian process $\mathcal{W}(t) = (\mathcal{W}_1(t), \ldots, \mathcal{W}_K(t))'$ with covariance function between $\mathcal{W}_j(t)$ and $\mathcal{W}_k(t)$ being $\xi_{jk}(s,t) = E(\Phi_{1j}(s)\Phi_{1k}(t))$.

Therefore, to complete the proof, we only need to prove the tightness of $\sqrt{n}W_n(t)$. Since the space of $\mathcal{D}[0,\tau]^K$ is equipped with uniform metric and thus marginal tightness implies joint tightness, it suffices to verify the tightness of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \frac{dM_{ik}(u)}{s_k^{(0)}(u;\beta_0)}$ and the tightness of $\left[\int_0^t \frac{s_k^{(1)}(u;\beta_0)}{s_k^{(0)}(u;\beta_0)} \mathrm{d}\Lambda_{0k}(u)\right]' \sqrt{n}(\widehat{\beta}_W - \beta_0)$. From the weak convergence of $\frac{1}{\sqrt{n}}\sum_{i=1}^n M_{ik}(t)$, $\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \frac{dM_{ik}(u)}{s_k^{(0)}(u;\beta_0)}$ converges weakly to a zero-mean Gaussian process. It then follows from Theorem 10.2 of Pollard [24] that $\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \frac{dM_{ik}(u)}{s_k^{(0)}(u;\beta_0)}$ is tight. The tightness of $\left[\int_0^t \frac{s_k^{(1)}(u;\beta_0)}{s_k^{(0)}(u;\beta_0)} \mathrm{d}\Lambda_{0k}(u)\right]' \sqrt{n}(\widehat{\beta}_W - \beta_0)$ follows from Theorem 1. Thus, we complete the proof. □

References

[1] G.J. D'Angio, N.E. Breslow, J.B. Beckwith, A. Evans, H. Baum, A. Delorimier, D. Fernbach, E. Hrabovsky, B. Jones, P. Kelalis, H.B. Othersen, M. Tefft, P.R.M. Thomas, Treatment of Wilms' tumor. Results of the third national Wilms' tumor study, Cancer 64 (1989) 349–360.
[2] D.M. Green, N.E. Breslow, J.B. Beckwith, J.Z. Finklestein, P.G. Grundy, P.R.M. Thomas, T. Kim, S. Shochat, G.M. Haasw, M.L. Ritchey, P.P. Kelalits, G.J. D'Angio, Comparison between single-dose and divided-dose administration of dactinomycin and doxorubicin for patients with Wilms' tumor: A report from the national Wilms' tumor study group, Journal of Clinical Oncology 16 (1998) 237–245.
[3] SOLVD Investigators, Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure, New England Journal of Medicine 325 (1991) 293–302.
[4] R.L. Prentice, Covariate measurement errors and parameter estimation in a failure time regression model, Biometrika 69 (1982) 331–342.
[5] H. Zhou, M.S. Pepe, Auxiliary covariate data in failure time regression analysis, Biometrika 82 (1995) 139–149.
[6] H. Zhou, C.Y. Wang, Failure time regression with continuous covariates measured with error, Journal of the Royal Statistical Society, Series B 62 (2000) 657–665.
[7] M. Kulich, D.Y. Lin, Additive hazards regression with covariate measurement error, Journal of the American Statistical Association 95 (2000) 238–248.
[8] J. Jiang, H. Zhou, Additive hazard regression with auxiliary covariates, Biometrika 94 (2007) 359–369.
[9] D.Y. Lin, Z. Ying, Semiparametric analysis of the additive risk model, Biometrika 81 (1994) 61–71.
[10] C. Hu, D.Y. Lin, Semiparametric failure time regression with replicates of mismeasured covariates, Journal of the American Statistical Association 99 (2004) 105–118.
[11] Y. Liu, H. Zhou, J. Cai, Estimated pseudo-likelihood method for correlated failure time data with auxiliary covariates, Biometrics (2009), in press (doi:10.1111/j.1541-0420.2009.01198.x).
[12] L.J. Wei, D.Y. Lin, L. Weissfeld, Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, Journal of the American Statistical Association 84 (1989) 1065–1073.
[13] R.L. Prentice, S.G. Self, Asymptotic distribution theory for Cox-type regression models with general relative risk form, Annals of Statistics 11 (1983) 804–813.
[14] E.A. Nadaraya, On estimating regression, Theory of Probability and Its Applications 10 (1964) 186–190.
[15] G.S. Watson, Smooth regression analysis, Sankya, Series A 26 (1964) 359–372.
[16] C.F. Spiekerman, D.Y. Lin, Marginal regression models for multivariate failure time data, Journal of the American Statistical Association 93 (1998) 1164–1175.
[17] R.V. Foutz, On the unique consistent solution to the likelihood equations, Journal of the American Statistical Association 72 (1977) 147–148.
[18] D. Clayton, J. Cuzick, Multivariate generalizations of the proportional hazards model, Journal of the Royal Statistical Society, Series A 148 (1985) 82–117.
[19] R.L. Eubank, Smoothing Spline and Nonparametric Regression, Dekker, New York, 1988.
[20] J. Cai, R.L. Prentice, Estimating equations for hazard ratio parameters based on correlated failure time data, Biometrika 82 (1995) 151–164.
[21] M.R. Kosorok, Introduction to Empirical Processes and Semiparametric Inference, Springer, New York, 2008.
[22] G. Yin, H. Li, D. Zeng, Partially linear additive hazards regression with varying coefficients, Journal of the American Statistical Association 103 (2008) 1200–1213.
[23] P.K. Andersen, R.D. Gill, Cox's regression model for counting processes: A large-sample study, Annals of Statistics 10 (1982) 1100–1120.
[24] D. Pollard, Empirical Processes: Theories and Applications, Institute of Mathematical Statistics, Hayward, 1990.