Journal of Nonparametric Statistics, 2014
Vol. 26, No. 1, 101–113, http://dx.doi.org/10.1080/10485252.2013.834337

Taylor & Francis
Taylor & Francis Group

Auxiliary covariate in additive hazards regression for survival data

Xiaoping Shi[a,b], Yanyan Liu[b] and Yuanshan Wu[b*]

[a]*School of Mathematics and System Sciences, Xinjiang University, Urumqi, Xinjiang, People's Republic of China;* [b]*School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, People's Republic of China*

(*Received 3 December 2012; accepted 7 August 2013*)

We consider the additive hazards regression analysis by utilising auxiliary covariate information to improve the efficiency of the statistical inference when the primary covariate is ascertained only for a randomly selected subsample. We construct a martingale-based estimating equation for the regression parameter and establish the asymptotic consistency and normality of the resultant estimator. Simulation study shows that our proposed method can improve the efficiency compared with the estimator which discards the auxiliary covariate information. A real example is also analysed as an illustration.

Keywords: additive hazards regression; auxiliary covariate; estimating equation; survival analysis; validation set

2000 Mathematics Subject Classifications: 62N01; 62N02; 62F12

1. Introduction

The Cox proportional hazards model (Cox 1972) has been widely used for the analysis of survival data. In the case of violating the proportional hazards assumption, the additive hazards model is a useful alternative (see, e.g. Breslow and Day 1980, 1987; Cox and Oakes 1984; Thomas 1986; and Lin and Ying 1994). Both of these two models can assess the covariate effect on the hazard function and their statistical interpretations can complement with each other in practice.

In many biomedical studies, it is often expensive to ascertain the primary exposure, due to the reason of technical difficulty or budget limitations. One useful accommodating approach is to measure some auxiliary covariate for primary exposure on all subjects, while conducting ascertainments on the primary exposure only for a randomly selected subsample. The auxiliary information can be obtained more easily or less expensively. Consequently, a natural and important question is how to make use of the auxiliary information to improve the statistical inference. Some proposed methods have been developed for this issue. For example, Pepe and Fleming (1991) proposed an implemented method which is nonparametric with respect to the mismeasurement process. Lin and Ying (1993) provided a general solution to survival data with missing covariate. Zhou and Pepe (1995) proposed an estimated partial likelihood by using both validation and nonvalidation observations to enhance efficiency. Greene and Cai (2004) proposed using the

*Corresponding author. Email: shan@whu.edu.cn

simulation-extrapolation method to handle measurement errors for multivariate failure time data when a validation set is not available. Liu, Zhou, and Cai (2009) and Liu, Wu, and Zhou (2010) proposed the statistical inference procedures for multivariate survival data by utilising auxiliary information for the case of discrete and continuous auxiliary covariates, respectively. Fan and Wang (2009) also considered this issue under the framework of the common baseline hazard function by adopting the kernel smoothing technique.

The aforementioned studies are all based on the framework of the proportional hazards model. For additive hazards model, some researches on how to further utilise auxiliary information have been conducted. For example, Kulich and Lin (2000) proposed a method based on correcting of the pseudo-score function for 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 (2007) 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 proposed a method by implementing the missing items in estimating equations with their empirical estimators based on auxiliary information and then obtaining 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. The proposed method is nonparametric with respect to the conditional distribution of the primary covariate given auxiliary. Our proposed method can be used as a remedy for this kind of incomplete data and can also be implemented easily in practice.

The rest of this paper is organised as follows. In Section 2, we propose an estimated 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. We delineate the proofs of the asymptotic results in the appendix.

2. Inference procedure

Suppose that there is a random sample of n independent subjects from an underlying population. For the ith subject ($i = 1, \ldots, n$), let $\tilde{T}_i$ and C_i denote the failure time and the censoring time, respectively. Due to censoring, we always observe $T_i = \min\{\tilde{T}_i, C_i\}$. Let $\Delta_i = I(\tilde{T}_i \leq C_i)$ 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 $Z_i = (Z_{i1}, \ldots, Z_{ip})^{\mathrm{T}}$ is the remaining covariate vector which is observed completely. Assume that the hazard function of $\tilde{T}_i$ 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. Denote $V = \{i : \eta_i = 1\}$ and $\bar{V} = \{i : \eta_i = 0\}$ 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 (Lin and Ying 1994). However, appropriately utilising auxiliary information could lead to remarkable

efficiency enhancement. 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 conditional on 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 $t \geq 0$. 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 utilising 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 $\tilde{T}_i$ takes the form as Equation (1). Otherwise, using the argument of Prentice (1982) and Zhou and Pepe (1995), it can be verified that the hazard function for $\lambda(t|Z_i, A_i)$ satisfies the induced hazards regression model as follows:

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

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

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

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

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

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

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

is a local square integrable zero-mean martingale at the true parameter $\theta_0 = (\beta_0^{\mathrm{T}}, \gamma_0^{\mathrm{T}})^{\mathrm{T}}$ (Fleming and Harrington 1991), where $\Lambda_0(t) = \int_0^t \lambda(u)\,\mathrm{d}u$ is the unknown cumulative baseline hazards function. Consequently, it is natural to obtain the Breslow (1972) and Aalen (1989) type estimator for $\Lambda_0(t)$ with given θ:

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

The parameter θ can be estimated from the following estimating equation:

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

or equivalently,

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

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 categorical variable with finitely possible values a_m $(m = 1, \ldots, q)$ and has the identical distribution $P(A_i^* = a_m) = p_m, m = 1, \ldots, q$, such that $\sum_{m=1}^{q} p_m = 1$. If the ith subject lies in the nonvalidation set $\bar{V}$, we can estimate the conditional expectation $\mathrm{E}\{X_i(t)|Y_i(t) = 1, A_i^*(t)\}$ by using the method of Nadaraya (1964) and Watson (1964),

$$\hat{X}_i(t) \equiv \frac{\sum_{j\in V} Y_j(t) I\{A_j^* = A_i^*\} X_j(t)}{\sum_{j\in V} Y_j(t) I\{A_j^* = A_i^*\}}.$$

Then, replacing the unknown conditional expectation by its estimated counterpart $\hat{X}_i(t)$ in $U(\theta)$, we obtained an estimated estimating equation, which is given by

$$\hat{U}(\theta) \equiv \sum_{i=1}^{n} \int_0^\tau \{\hat{W}_i(t) - \hat{E}(t)\}\{\mathrm{d}N_i(t) - Y_i(t)\theta^{\mathrm{T}} \hat{W}_i(t)\,\mathrm{d}t\} = 0.$$

The proposed estimator, denoted by $\hat{\theta}_E$, which solves $\hat{U}(\theta) = 0$, can be explicitly obtained as follows,

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

where $\hat{W}_i(t) = W_i(t)\eta_i + [\hat{X}_i(t)^{\mathrm{T}}, Z_i(t)^{\mathrm{T}}]^{\mathrm{T}}(1 - \eta_i)$, $\hat{E}(t) = \sum_{i=1}^{n} Y_i(t)\hat{W}_i(t)/\sum_{i=1}^{n} Y_i(t)$, and $a^{\otimes 2} = aa^{\mathrm{T}}$ for a column vector a. Furthermore, $\Lambda_0(t)$ can be estimated by $\hat{\Lambda}_0(t; \hat{\theta}_E)$, where

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

3. Asymptotic properties

For simplicity, we introduce some notations. For a vector a, define $a^{\otimes 0} = 1, a^{\otimes 1} = a, \|a\| = \sup_i |a_i|$. For a matrix $A = (a_{ij})$, define $\|A\| = \sup_{i,j} |a_{ij}|$. For $k = 0, 1$, define $\hat{S}^{(k)}(t) = n^{-1}\sum_{i=1}^{n} Y_i(t)\hat{W}_i^{\otimes k}(t)$, $S^{(k)}(t) = n^{-1}\sum_{i=1}^{n} Y_i(t)W_i^{*\otimes k}(t)$, $s^{(k)}(t) = \mathrm{E}\{Y(t)W^{*\otimes k}(t)\}$. Denote $H_{a_m}(t) = P(Y(t) = 1|A^* = a_m)$ and $\rho = \lim_{n\to\infty} v/n$, where v is the cardinality of the validation set V. For $j \in V$, let

$$Q_j(\theta) = \int_0^\tau \left[\mathrm{E}\{W_j(t)|Y_j(t) = 1, A_j^*\} - \frac{s^{(1)}(t)}{s^{(0)}(t)}\right] Y_j(t)[W_j(t) - \mathrm{E}\{W_j(t)|Y_j(t) = 1, A_j^*\}]^{\mathrm{T}}\theta\,\mathrm{d}t,$$

$$Q_j^{\bar{v}}(\theta) = \frac{1}{\bar{v}} \sum_{i\in\bar{V}} \int_0^\tau \left\{W_i^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)}\right\} \frac{Y_i(t)Y_j(t)I\{A_i^* = A_j^*\}}{p_{A_i^*} H_{A_i^*}(t)}\{W_j(t) - W_i^*(t)\}^{\mathrm{T}}\theta\,\mathrm{d}t,$$

where $\bar{v} = n - v$.

We impose the following conditions through our derivations:

C1. $\int_0^\tau \lambda_0(t)\,\mathrm{d}t < \infty$.
C2. $P\{Y(t) = 1|A^* = a_m\} > 0$ for $m = 1, \ldots, q$.
C3. $\mathrm{E}\{\sup_{t\in[0,\tau]} \|Y(t)W^{*\otimes k}(t)\|\} < \infty$ for $k = 0, 1$.

C4. $\sup_{t\in[0,\tau]} \|W^v(t)\| = O_P(1)$, where

$$W^v(t) = \sqrt{v}\left\{\frac{1}{v}\sum_{i\in V} I\{Y_i(t)=1, A_i^* = a_m\}W_i(t) - \mathrm{E}[I\{Y(t)=1, A^* = a_m\}W(t)]\right\}$$

for $m = 1, \ldots, q$.

Conditions C1–C3 are standard assumptions in survival analysis. Condition C4 is imposed to simplify the derivations of the asymptotic properties, which can be satisfied if the class of functions $\{I\{Y_i(t)=1, A_i^* = a_m\}W_i(t) : i = 1, \ldots, n;\ t \in [0,\tau]\}$ is Donsker. We summarise the asymptotic results of the proposed estimator in the following theorems.

THEOREM 3.1 *Under conditions* C1–C4, $\hat{\theta}_E$ *converges to* θ_0 *in probability.*

THEOREM 3.2 *Under conditions* C1–C4, $\sqrt{n}(\hat{\theta}_E - \theta_0)$ *is asymptotically normal with mean zero and covariance matrix* $\Sigma(\theta_0)^{-1}\{(1-\rho)\Sigma_1(\theta_0) + \rho\Sigma_2(\theta_0)\}\{\Sigma(\theta_0)^{-1}\}^{\mathrm{T}}$, *where*

$$\Sigma(\theta) = \mathrm{E}\left[\int_0^\tau \left\{W^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)}\right\}^{\otimes 2} Y(t)\,\mathrm{d}t\right],$$

$$\Sigma_1(\theta) = \mathrm{E}\left[\int_0^\tau \left\{W^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)}\right\}^{\otimes 2} Y(t)\,\mathrm{d}\Lambda_0(t)\right] + \mathrm{E}\left[\int_0^\tau \left\{W^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)}\right\}^{\otimes 2} Y(t)\theta^{\mathrm{T}}W^*(t)\,\mathrm{d}t\right],$$

$$\Sigma_2(\theta) = \mathrm{E}\left[\int_0^\tau \left\{W(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)}\right\}\mathrm{d}M(t) - \frac{1-\rho}{\rho}Q(\theta)\right]^{\otimes 2}.$$

The consistency of $\hat{\theta}_E$ follows by verifying conditions in Theorem 1 in Foutz (1977). Using the Taylor expansion and martingale representation theory, we can show that the estimated estimating equation is asymptotically equivalent to a sum of two independent terms. Each of the items is also shown to be asymptotically equivalent to a sum of independent vectors. Thus, the central limit theorem can be employed to obtain the asymptotic normality of $\hat{\theta}_E$. The outline of the proofs of theorems and some related lemma are provided in the appendix.

4. Simulation studies

In this section, we examined the finite sample properties of $\hat{\theta}_E$ via simulation studies. We compared $\hat{\theta}_E$ with two estimators. The first one is the validation set estimator, denoted by $\hat{\theta}_V$, which is obtained by using Lin and Ying (1994)'s method based only on the validation data. The other one is the naive estimator, denoted by $\hat{\theta}_N$, 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 $\tilde{T}$ from the hazard model $\lambda(t|Z,X) = 2 + \beta_0 X + \gamma_0 Z$, where both X and Z were independently simulated from Unif$(0,2)$. We constructed the auxiliary covariate through $A^* = I(X + e > Q_{0.5})$, where $e \sim \mathrm{N}(0,\sigma^2)$ and $Q_{0.5}$ denotes the sample median of the variable $X + e$. Here σ^2 is the parameter which controls the strength of the association between X

Table 1. Simulation results based on the hazard model $\lambda(t|Z,X) = 2 + \beta_0 X + \gamma_0 Z$ with the censoring rate of 30%.

				$\beta_0 = 2$				$\gamma_0 = 2$			
n	ρ	σ	Method	Est	SD	SE	95%CP	Est	SD	SE	95%CP
100	0.8		Validation	2.042	1.453	1.403	0.957	2.034	1.425	1.403	0.951
		0.2	Naive	1.807	1.212	1.196	0.944	2.050	1.267	1.249	0.957
			Proposed	2.057	1.316	1.273	0.951	2.048	1.260	1.204	0.939
		1	Naive	1.636	1.175	1.170	0.944	2.056	1.269	1.249	0.955
			Proposed	2.042	1.387	1.328	0.949	2.052	1.265	1.206	0.943
	0.5		Validation	2.013	1.784	1.823	0.961	2.047	1.845	1.801	0.960
		0.2	Naive	1.665	1.223	1.200	0.943	2.055	1.283	1.249	0.954
			Proposed	2.075	1.396	1.418	0.966	2.047	1.268	1.245	0.956
		1	Naive	1.261	1.201	1.159	0.887	2.062	1.283	1.250	0.955
			Proposed	2.020	1.659	1.692	0.959	2.053	1.280	1.260	0.948
200	0.8		Validation	2.084	0.991	0.972	0.945	2.053	0.967	0.971	0.953
		0.2	Naive	1.791	0.818	0.823	0.944	2.043	0.855	0.865	0.958
			Proposed	2.069	0.892	0.893	0.953	2.043	0.852	0.848	0.952
		1	Naive	1.628	0.813	0.808	0.918	2.046	0.856	0.865	0.958
			Proposed	2.075	0.958	0.938	0.948	2.047	0.854	0.849	0.953
	0.5		Validation	2.090	1.278	1.239	0.946	2.021	1.232	1.235	0.956
		0.2	Naive	1.662	0.819	0.822	0.936	2.043	0.858	0.865	0.959
			Proposed	2.085	0.946	0.981	0.961	2.043	0.852	0.873	0.958
		1	Naive	1.260	0.818	0.793	0.838	2.049	0.859	0.865	0.955
			Proposed	2.060	1.164	1.190	0.955	2.046	0.853	0.884	0.957
500	0.8		Validation	2.001	0.614	0.601	0.950	1.989	0.596	0.606	0.953
		0.2	Naive	1.750	0.531	0.517	0.915	1.986	0.530	0.542	0.961
			Proposed	1.995	0.570	0.562	0.951	1.985	0.532	0.535	0.952
		1	Naive	1.588	0.525	0.507	0.867	1.987	0.531	0.542	0.962
			Proposed	1.998	0.605	0.589	0.936	1.986	0.533	0.536	0.951
	0.5		Validation	2.029	0.759	0.769	0.953	1.946	0.760	0.767	0.950
		0.2	Naive	1.622	0.519	0.515	0.880	1.986	0.531	0.542	0.960
			Proposed	2.004	0.593	0.617	0.958	1.985	0.533	0.548	0.959
		1	Naive	1.239	0.507	0.497	0.667	1.988	0.530	0.542	0.962
			Proposed	2.005	0.707	0.746	0.949	1.986	0.531	0.557	0.963

Notes: The column '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.

and A^*. We generated η from the Bernoulli distribution with success probability ρ and the censoring times from Unif$(0, c)$, where ρ is the fraction of the size of validation set over all the samples and c was chosen to yield a censoring rate of 30%. 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. Each configuration was replicated 1000 times under the sample sizes $n = 100$, 200, and 500, respectively. The corresponding results are summarised in Table 1. 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.

From Table 1, we made the following observations: (i) All the estimates for γ_0 are essentially unbiased. For β_0, both $\hat{\beta}_V$ and $\hat{\beta}_E$ are virtually unbiased with reasonable coverage probabilities. However, $\hat{\beta}_N$ is biased; (ii) The proposed estimator $\hat{\theta}_E$ is more efficient than $\hat{\theta}_V$; (iii) $\hat{\beta}_E$ works well even σ is large ($\sigma = 1$), in other words, A^* is less informative about X as the sample size n is increased to 500.

To investigate the effect of the censoring rate on the performance of the proposed method, we chose c to yield a censoring rate of 60% while keeping the remaining set-ups the same as

Table 2. Simulation results based on the hazard model $\lambda(t|Z,X) = 2 + \beta_0 X + \gamma_0 Z$ with the censoring rate of 60%.

				$\beta_0 = 2$				$\gamma_0 = 2$			
n	ρ	σ	Method	Est	SD	SE	95%CP	Est	SD	SE	95%CP
100	0.8		Validation	2.162	1.871	1.844	0.949	2.262	1.826	1.836	0.958
		0.2	Naive	1.917	1.585	1.568	0.954	2.263	1.592	1.633	0.963
			Proposed	2.169	1.688	1.662	0.948	2.260	1.584	1.601	0.958
		1	Naive	1.767	1.561	1.546	0.946	2.266	1.589	1.633	0.963
			Proposed	2.161	1.796	1.757	0.947	2.262	1.584	1.603	0.956
	0.5		Validation	2.140	2.335	2.375	0.955	2.315	2.363	2.353	0.964
		0.2	Naive	1.780	1.578	1.570	0.950	2.269	1.613	1.634	0.961
			Proposed	2.192	1.785	1.813	0.951	2.261	1.599	1.640	0.962
		1	Naive	1.394	1.557	1.532	0.926	2.270	1.595	1.634	0.960
			Proposed	2.160	2.118	2.217	0.956	2.261	1.594	1.657	0.964
200	0.8		Validation	2.164	1.316	1.273	0.945	2.111	1.244	1.266	0.957
		0.2	Naive	1.861	1.096	1.079	0.940	2.115	1.123	1.130	0.955
			Proposed	2.134	1.172	1.159	0.945	2.115	1.120	1.118	0.956
		1	Naive	1.714	1.086	1.065	0.936	2.120	1.123	1.130	0.952
			Proposed	2.150	1.271	1.228	0.945	2.122	1.119	1.120	0.954
	0.5		Validation	2.220	1.690	1.616	0.954	2.082	1.583	1.611	0.962
		0.2	Naive	1.722	1.101	1.075	0.941	2.113	1.122	1.130	0.950
			Proposed	2.146	1.229	1.246	0.955	2.114	1.117	1.139	0.961
		1	Naive	1.319	1.072	1.047	0.893	2.119	1.120	1.130	0.952
			Proposed	2.141	1.529	1.524	0.951	2.119	1.116	1.149	0.962
500	0.8		Validation	2.058	0.813	0.812	0.952	2.069	0.800	0.811	0.951
		0.2	Naive	1.819	0.710	0.693	0.937	2.062	0.706	0.723	0.955
			Proposed	2.055	0.757	0.747	0.950	2.063	0.707	0.721	0.956
		1	Naive	1.662	0.687	0.683	0.917	2.063	0.705	0.723	0.955
			Proposed	2.056	0.793	0.792	0.955	2.063	0.708	0.722	0.955
	0.5		Validation	2.075	1.002	1.028	0.960	2.041	1.002	1.027	0.950
		0.2	Naive	1.696	0.702	0.692	0.927	2.061	0.707	0.723	0.955
			Proposed	2.076	0.780	0.800	0.953	2.062	0.709	0.732	0.960
		1	Naive	1.294	0.666	0.674	0.817	2.062	0.705	0.723	0.956
			Proposed	2.056	0.918	0.981	0.957	2.061	0.706	0.737	0.966

Notes: The column '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.

before. The simulation results are presented in Table 2, from which we can conclude the similar conclusions as that in Table 1.

The relative efficiency of $\hat{\theta}_E$ versus $\hat{\theta}_V$ along the validation fraction ρ under different censoring rates is summarised in Table 3. We define the relative efficiency of $\hat{\theta}_E$ versus $\hat{\theta}_V$ as $\{\mathrm{SD}(\hat{\theta}_V)/\mathrm{SD}(\hat{\theta}_E)\}^2$, where $\mathrm{SD}(\hat{\theta})$ is the sample standard derivation of estimate $\hat{\theta}$. Based on Table 3, we observed that the efficiency gain of the proposed method is becoming more significant as the validation fraction ρ is decreasing. 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.

Furthermore, we considered the situation where it is more informative discretisation of the auxiliary covariate. The auxiliary covariate A^* is assigned the values of 1,…, or 8 if the observation $X + e$ lies in the interval $(-\infty, Q_1]$, $(Q_i, Q_{i+1}]$, and $(Q_7, +\infty)$, respectively, where Q_i are the sample $i/8$ $(i = 1, \ldots, 7)$ quantiles of $X + e$. The remaining simulation set-ups were kept the same as before. The corresponding simulation results for β_0 are summarised in Table 4. It can be seen that the proposed method results in smaller MSEs with more informative discretisation. However, if it discretises the auxiliary as too many stratums, a larger sample size is needed to

Table 3. The efficiency comparison between the proposed estimator $\hat{\theta}_E$ and the validation set estimator $\hat{\theta}_V$ with $\sigma = 0.2$ and $n = 200$.

		SD ($\beta_0 = 2$)			SD ($\gamma_0 = 2$)		
Cen. rate	ρ	Validation	Proposed	RE	Validation	Proposed	RE
30%	0.8	0.991	0.913	1.178	0.967	0.854	1.282
	0.5	1.278	0.983	1.690	1.232	0.853	2.086
	0.2	2.111	1.087	3.781	2.048	0.861	5.659
60%	0.8	1.316	1.199	1.205	1.244	1.120	1.238
	0.5	1.690	1.277	1.751	1.583	1.116	2.012
	0.2	2.767	1.371	4.070	2.673	1.122	5.675

Notes: 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, which is calculated by $\{\mathrm{SD}(\hat{\theta}_V)/\mathrm{SD}(\hat{\theta}_E)\}^2$.

Table 4. Simulation results for $\beta_0 = 2$ based on the hazard model $\lambda(t|Z, X) = 2 + \beta_0 X + 2Z$ with $n = 200$ and A^* taking two or eight possible values.

			Two categories					Eight categories				
Cen. rate	ρ	σ	Est	SD	SE	MSE	95%CP	Est	SD	SE	MSE	95%CP
30%	0.8	0.2	2.069	0.892	0.893	0.800	0.953	2.059	0.865	0.866	0.752	0.948
		1	2.075	0.958	0.938	0.923	0.948	2.027	0.936	0.918	0.877	0.944
	0.5	0.2	2.085	0.946	0.981	0.902	0.961	2.058	0.871	0.893	0.762	0.951
		1	2.060	1.164	1.190	1.358	0.955	1.830	1.050	1.060	1.131	0.945
60%	0.8	0.2	2.134	1.172	1.159	1.392	0.945	2.131	1.141	1.131	1.319	0.944
		1	2.150	1.271	1.228	1.638	0.945	2.107	1.267	1.210	1.617	0.940
	0.5	0.2	2.146	1.229	1.246	1.532	0.955	2.127	1.143	1.155	1.323	0.946
		1	2.141	1.529	1.524	2.358	0.951	1.935	1.412	1.395	1.998	0.945

Note: 'MSE' is the mean square error.

ensure that we can utilise the observations in the validation set to 'replace' the counterpart with missing in the nonvalidation set.

5. The primary biliary cirrhosis data

We applied 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 the treatment (Trt), patients' sex (Sex), and serum cholesterol level (Chol). A clinical background description and a more extend discussion for the trial and the covariates recorded can be found in Dickson, Grambsch, Fleming, Fisher, and Langworthy (1989) and Markus et al. (1989).

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

Table 5. Basic descriptions of the covariates in PBC data.

Covariate	Type	Missing rate
Trt	categorical	0
Sex	categorical	0
Chol	continuous	9%

Note: Trt = 1 or 0, whether the patient was treated with D-penicillamine or placebo; Sex = 1 or 0, whether the patient is female or male.

Table 6. Analysis results for the real example with the missing rate of 9%.

Method	Covariate	Est	SE	SE95%CI	*p*-Value
Validation	Trt	0.0007	0.0125	(−0.0238, 0.0252)	0.9551
	Sex	−0.0525	0.0250	(−0.1015, −0.0035)	0.0355
	log(Chol)	0.0490	0.0208	(0.0082, 0.0898)	0.0185
Proposed	Trt	−0.0031	0.0117	(−0.0260, 0.0198)	0.7936
	Sex	−0.0555	0.0295	(−0.1133, 0.0023)	0.0597
	log(Chol)	0.0543	0.0193	(0.0165 , 0.0921)	0.0049

Table 7. Analysis results for the real example with the artificially constructed missing rate of 50%.

Method	Covariate	Est	SE	95%CI	*p*-Value
Validation	Trt	0.0089	0.0154	(−0.0214, 0.0390)	0.5691
	Sex	−0.0482	0.0307	(−0.1083, 0.0119)	0.1159
	log(Chol)	0.0390	0.0237	(−0.0074, 0.0854)	0.1002
Proposed	Trt	−0.0022	0.0126	(−0.0269, 0.0226)	0.8645
	Sex	−0.0523	0.0284	(−0.1080, 0.0034)	0.0657
	log(Chol)	0.0906	0.0261	(0.0394, 0.1418)	0.0005

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. The auxiliary covariate is then assigned the value 0 or 1 based on whether or not the logarithm of bilirubin is less than its empirical median 0.2994. The basic descriptions of the covariates are presented in Table 5.

Table 6 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.

To demonstrate the effectiveness of the proposed method for moderate missing rate (50%), we artificially constructed the missing serum cholesterol levels through simple random sampling approach. The analysis results are summarised in Table 7. Under the moderate missing rate of 50%, the covariate log(Chol) is still significant at the level of 5% in the proposed method while it is insignificant in the validation set method.

6. Concluding remarks

In this article, we proposed an estimated estimating equation method for the survival data with 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 different sample sizes and that the resulting estimator outperforms the validation set estimator. The proposed variance estimator also performs well. When the auxiliary covariate A^* is more informative about the primary exposure X, the proposed estimator is more efficient.

When the auxiliary covariate is continuous, one way is to discretise it into categories and then apply the proposed method, as we did in real data analysis. A more efficient method is to develop a nonparametric kernel smoothing version for unspecified conditional expectation in estimation equation. The related work is currently underway.

We consider that only one missing covariate in our proposal, nevertheless, the proposed method could be extended to the case where multiple covariates are subject to missing. In application, much larger sample size is needed in such scenario to ensure that we can utilise the observations in the validation set to 'replace' the counterpart with missing in the nonvalidation set. When the censoring mechanism is not independent, that is, the censoring mechanism is informative, the statistical inference for the additive hazards regression with auxiliary information becomes a challenge because the censoring variable also implies the information of the regression parameter of interested. New statistical method is needed to be developed along this research direction.

Acknowledgements

We thank the Associate Editor and two referees for their helpful comments which improved the paper substantially. This research is supported in part by the National Science Foundation of China (11171263, 41261087, 11201350), the Doctoral Fund of Ministry of Education of China (201110141110004, 20110141120004), and the Fundamental Research Funds for the Central Universities.

References

Aalen, O.O. (1989), 'A Linear Regression Model for the Analysis of Life Times', *Statistics in Medicine*, 8, 907–925.

Andersen, P.K., and Gill, R.D. (1982), 'Cox's Regression Model for Counting Processes: A Large Sample Study', *The Annals of Statistics*, 10, 1100–1120.

Breslow, N.E. (1972), 'Discussion of Paper of D. R. Cox', *Journal of the Royal Statistical Society, Series B*, 34, 216–217.

Breslow, N.E., and Day, N.E. (1980), *Statistical Methods in Cancer Research* (Vol. 1), The Design and Analysis of Case-Control Studies, Lyon: IARC.

Breslow, N.E., and Day, N.E. (1987), *Statistical Methods in Cancer Research* (Vol. 2), The Design and Analysis of Cohort Studies, Lyon: IARC.

Cox, D.R. (1972), 'Regression Models and Life-Tables', *Journal of the Royal Statistical Society, Series B*, 34, 187–202.

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

Dickson, E.R., Grambsch, P.M., Fleming, T.R., Fisher, L.D., and Langworthy, A. (1989), 'Prognosis in Primary Biliary Cirrhosis: Model for Decision Making', *Hepatology*, 10, 1–7.

Fan, Z., and Wang, X.-F. (2009), 'Marginal Hazards Model for Multivariate Failure Time Data with Auxiliary Covariates', *Journal of Nonparametric Statistics*, 21, 771–786.

Fleming, T.R., and Harrington, D.P. (1991), *Counting Processes and Survival Analysis*, New York: Wiley.

Foutz, R.V. (1977), 'On the Unique Consistent Solution to Likelihood Equations', *Journal of the American Statistical Association*, 72, 147–148.

Greene, W.F., and Cai, J. (2004), 'Measurement Error in Covariates in the Marginal Hazards Model for Multivariate Failure Time Data', *Biometrics*, 60, 987–996.

Jiang, J., and Zhou, H. (2007), 'Additive Hazard Regression with Auxiliary Covariates', *Biometrika*, 92, 359–369.

Kulich, M., and Lin, D.Y. (2000), 'Additive Hazards Regression with Covariate Measurement Error', *Journal of the American Statistical Association*, 95, 238–248.

Lin, D.Y., and Ying, Z. (1993), 'Cox Regression with Incomplete Covariate Measurements', *Journal of the American Statistical Association*, 88, 1341–1349.

Lin, D.Y., and Ying, Z. (1994), 'Semiparametric Analysis of the Additive Risk Model', *Biometrika*, 81, 61–71.
Liu, Y., Wu, Y., and Zhou, H. (2010), 'Multivariate Failure Times Regression with a Continuous Auxiliary Covariate', *Journal of Multivariate Analysis*, 101, 679–691.
Liu, Y., Zhou, H., and Cai, J. (2009), 'Estimated Pseudopartial-Likelihood Method for Correlated Failuretime Data with Auxiliary Covariates', *Biometrics*, 65, 1184–1193.
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. (1989), 'Efficiency of Liver Transplantation in Patients with Primary Biliary Cirrhosis', *The New England Journal of Medicine*, 320, 1709–1713.
Nadaraya, E.A. (1964), 'On Estimating Regression', *Theory of Probability and Its Applications*, 10, 186–190.
Pepe, M.S., and Fleming, T.R. (1991), 'A Nonparametric Method for Dealing with Mismeasured Covariate Data', *Journal of the American Statistical Association*, 86, 108–113.
Prentice, R.L. (1982), 'Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model', *Biometrika*, 69, 331–342.
Thomas, D.C. (1986), 'Use of Auxiliary Information in Fitting Nonproportional Hazards Models', in *Modern Statistical Methods in Chronic Disease Epidemiology*, eds. S.H. Moolgavkar and R.L. Prentice, New York: Wiley, pp. 197–210.
Watson, G.S. (1964), 'Smooth Regression Analysis', *Sankya, Series A*, 26, 359–372.
Zhou, H., and Pepe, M.S. (1995), 'Auxiliary Covariate Data in Failure Time Regression Analysis', *Biometrika*, 82, 139–149.

Appendix

In what follows, 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.

Since $Y(t)I\{A^* = a_m\}X(t)$ and $Y(t)I\{A^* = a_m\}$ as functions of $t \in [0, \tau]$ are left continuous with right-hand limit. According to the arguments in Andersen and Gill (1982, Theorem III.1) and noting that condition C3 implies that $\mathrm{E}[\sup_{t\in[0,\tau]} Y(t)I\{A^* = a_m\}X(t)] < \infty$, we have $\sup_{t\in[0,\tau]} \|X_i(t) - \hat{X}_i(t)\| \to_{a.s.} 0$ for $i \in \bar{V}$. Furthermore, we can conclude that

$$\sup_{t\in[0,\tau]} \|\hat{W}_i(t) - W_i^*(t)\| \to_{a.s.} 0, \quad i = 1, \ldots, n. \tag{A1}$$

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

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

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

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

Immediately, we have that

$$\sup_{t\in[0,\tau]} \|\hat{S}^{(k)}(t) - s^{(k)}(t)\| \to_{a.s.} 0 \quad \text{for } k = 0, 1. \tag{A2}$$

LEMMA A.1 *Under conditions* C1–C4,

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

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

$$\begin{aligned} n^{-1/2}\hat{U}(\theta_0) &= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \left\{\hat{W}_i(t) - \frac{\hat{S}^{(1)}(t)}{\hat{S}^{(0)}(t)}\right\} \mathrm{d}M_i(t) \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \left\{\hat{W}_i(t) - \frac{\hat{S}^{(1)}(t)}{\hat{S}^{(0)}(t)}\right\} Y_i(t)\,\mathrm{d}\Lambda_0(t) \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \left\{\hat{W}_i(t) - \frac{\hat{S}^{(1)}(t)}{\hat{S}^{(0)}(t)}\right\} Y_i(t)\{W_i^*(t) - \hat{W}_i(t)\}^{\mathrm{T}}\theta_0\,\mathrm{d}t \\ &\equiv B_{1n} + B_{2n} + B_{3n}. \end{aligned} \tag{A3}$$

Furthermore, we rewrite

$$B_{1n} = n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \left\{ \hat{W}_i(t) - W_i^*(t) + W_i^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)} + \frac{s^{(1)}(t)}{s^{(0)}(t)} - \frac{\hat{S}^{(1)}(t)}{\hat{S}^{(0)}(t)} \right\} \mathrm{d}M_i(t).$$

Note that $n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \{\hat{W}_i(t) - W_i^*(t)\} \mathrm{d}M_i(t)$ is a square integrable martingale, which converges in probability to zero by the Lenglart inequality (Andersen and Gill, 1982). Analogously, $n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \{\frac{s^{(1)}(t)}{s^{(0)}(t)} - \frac{\hat{S}^{(1)}(t)}{\hat{S}^{(0)}(t)}\} \mathrm{d}M_i(t)$ also converges in probability to zero. Hence,

$$B_{1n} = n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \left\{ W_i^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\} \mathrm{d}M_i(t) + o_P(1). \tag{A4}$$

We also rewrite

$$B_{3n} = n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \left\{ \hat{W}_i(t) - W_i^*(t) + W_i^*(t) - \frac{\hat{S}^{(1)}(t)}{\hat{S}^{(0)}(t)} \right\} Y_i(t) \{W_i^*(t) - \hat{W}_i(t)\}^{\mathrm{T}} \theta_0 \, \mathrm{d}t.$$

Using Equation (A1) and condition C4, we have that

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

Combining Equation (A2) and the definitions of $\hat{W}_i(t)$ and $Q_j^{\bar{v}}(\theta)$, we have that

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

Similarly, using the definition of $Q_j(\theta_0)$ and condition C4, we also have that

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

Thus,

$$B_{3n} = -n^{-1/2} \frac{\bar{v}}{v} \sum_{j \in V} Q_j(\theta_0) + o_P(1). \tag{A5}$$

Obviously, noting $B_{2n} = 0$ and combining Equations (A4) and (A5), we derive Equation (A3) as follows:

$$\begin{aligned} n^{-1/2} \hat{U}(\theta_0) &= n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \left\{ W_i^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\} \mathrm{d}M_i(t) - n^{-1/2} \frac{\bar{v}}{v} \sum_{j \in V} Q_j(\theta_0) + o_P(1) \\ &= n^{-1/2} \sum_{i \in \bar{V}} \int_0^{\tau} \left\{ W_i^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\} \mathrm{d}M_i(t) \\ &\quad + n^{-1/2} \sum_{j \in V} \left[\int_0^{\tau} \left\{ W_j^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\} \mathrm{d}M_j(t) - \frac{\bar{v}}{v} Q_j(\theta_0) \right] + o_P(1). \end{aligned}$$

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 $\mathrm{E}\{(\bar{v}/v)Q_j(\theta_0)\} = 0$. The lemma follows immediately from that two terms are independent because they are summations over the nonvalidation and validation sets, respectively. ■

Proof of Theorem 3.1 We use Theorem 1 in Foutz (1977) to prove the consistency of $\hat{\theta}_E$ by verifying the following conditions.

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

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

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

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

First, we have

$$-n^{-1}\frac{\partial \hat{U}(\theta)}{\partial \theta} = n^{-1}\sum_{i=1}^{n}\int_0^\tau \left\{\hat{W}_i(t) - \frac{\hat{S}^{(1)}(t)}{\hat{S}^{(0)}(t)}\right\}\hat{W}_i(t)^{\mathrm{T}} Y_i(t)\,\mathrm{d}t$$

Thus, (i) is satisfied. Second, rewrite

$$-n^{-1}\frac{\partial \hat{U}(\theta)}{\partial \theta} = n^{-1}\sum_{i=1}^{n}\int_0^\tau \left\{\hat{W}_i(t) - W_i^*(t) + \frac{s^{(1)}(t)}{s^{(0)}(t)} - \frac{\hat{S}^{(1)}(t)}{\hat{S}^{(0)}(t)} + W_i^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)}\right\}^{\otimes 2} Y_i(t)\,\mathrm{d}t.$$

Combining Equations (A1), (A2), and condition C3, after some algebraic manipulations, we can further conclude that

$$-n^{-1}\frac{\partial \hat{U}(\theta)}{\partial \theta} = n^{-1}\sum_{i=1}^{n}\int_0^\tau \left\{W_i^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)}\right\}^{\otimes 2} Y_i(t)\,\mathrm{d}t + o_P(1).$$

It follows from the strong law of large numbers that

$$-n^{-1}\frac{\partial \hat{U}(\theta)}{\partial \theta} \longrightarrow_p \mathrm{E}\left[\int_0^\tau \left\{W^*(t) - \frac{s^{(1)}(t)}{s^{(0)}(t)}\right\}^{\otimes 2} Y(t)\,\mathrm{d}t\right] = \Sigma(\theta). \tag{A6}$$

Thus, (ii) and (iii) are verified. Finally, (iv) also holds by Lemma A.1. Hence, $\hat{\theta}_E$ converges in probability to θ_0. ■

Proof of Theorem 3.2 Using the Taylor expansion, we have

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

It follows from Lemma A.1, Equation (A6), and the consistency of $\hat{\theta}_E$ that $\sqrt{n}(\hat{\theta}_E - \theta_0)$ is asymptotically normal with mean zero and covariance matrix $\Sigma(\theta_0)^{-1}\{(1-\rho)\Sigma_1(\theta_0) + \rho\Sigma_2(\theta_0)\}\{\Sigma(\theta_0)^{-1}\}^{\mathrm{T}}$. Thus, we complete the proof of Theorem 3.2. ■