

SEMPARAMETRIC ADDITIVE RISKS REGRESSION FOR TWO-STAGE DESIGN SURVIVAL STUDIES

Gang Li and Tong Tong Wu

University of California, Los Angeles and University of Maryland, College Park

Abstract: In this article we study a semiparametric additive risks model (McKeague and Sasieni (1994)) for two-stage design survival data where accurate information is available only on second stage subjects, a subset of the first stage study. We derive two-stage estimators by combining data from both stages. Large sample inferences are developed. As a by-product, we also obtain asymptotic properties of the single stage estimators of McKeague and Sasieni (1994) when the semiparametric additive risks model is misspecified. The proposed two-stage estimators are shown to be asymptotically more efficient than the second stage estimators. They also demonstrate smaller bias and variance for finite samples. The developed methods are illustrated using small intestine cancer data from the SEER (Surveillance, Epidemiology, and End Results) Program.

Key words and phrases: Censored data, correlation, efficiency, measurement errors, missing covariates.

1. Introduction

Two-stage designs are useful in medical studies and other fields of research. The first stage sample of a two-stage design consists of a set of subjects under study with surrogate, inaccurate, or missing information. The second stage sample is a subset of individuals from the first stage with accurate and complete data. Typical scenarios include measurement error and missing covariates problems. For example, when a complete survey is complicated, expensive, and time consuming, researchers often use a simplified version for all study subjects in the first stage. The complete version is taken only by a small subset of study subjects. Two-stage data also arise in applications where certain information from more recently available technology, such as a genome-wide scan, is collected only for newly-diagnosed patients. A medical device postmarketing surveillance example was given by Li and Tseng (2008): St Jude Medical conducted a postmarket surveillance study to evaluate the safety and efficacy of five pacing electrodes by collecting information of the adverse events or failures. The database maintained by the medical device company, which contains all the market devices, has serious under-reporting problems that might underestimate the failure and

adverse event rates. To offset the under-reporting bias, St Jude Medical drew an active follow-up sample and collected accurate and complete information on this sample. This typical two-stage survival study consists of the company administrative data (first stage data) and the active follow-up data (second stage data). In general, analysis based on the first stage data alone could be biased. On the other hand, analysis based on the second stage data alone would not be the most efficient since it does not utilize information from the first stage. It would be desirable to combine data from both stages to increase efficiency of the second stage data analysis.

The two-stage design has been studied extensively for complete data; see, e.g., White (1982), Schill et al. (1993), Breslow and Holubkov (1997), and references therein. However, there are relatively few methods available for analysis of two-stage censored survival data, especially when both the outcome variable and covariates are subject to error in the first stage sample. Among others, Zhou and Pepe (1995) and Wang et al. (1997) studied the surrogate covariates problem for a multiplicative semiparametric hazard model using regression calibration techniques. Kulich and Lin (2000) proposed a corrected pseudo-score estimator for the additive risks model of Lin and Ying (1994) with measurement errors on covariates. Based on the work of Chen and Chen (2000) on regression models in two-stage designs, Chen (2002) and Tseng (2004) studied the Cox model for two-stage survival data, where both survival time and covariates are subject to measurement error. Li and Tseng (2008) studied nonparametric estimation of survival functions for two-stage survival data. Jiang and Zhou (2007) studied a two-stage design problem for Lin and Ying's (1994) model.

In this paper we study the semiparametric additive risks model of McKeague and Sasieni (1994) (referred to as MS hereafter) for analysis of two-stage survival data where both the survival time and covariates are subject to measurement errors. Let $h(t|x, z)$ denote the conditional hazard function of a survival time given x and z . The MS model postulates that

$$h(t|x, z) = \alpha(t)'x + \beta'z.$$

The MS model provides a useful alternative to the Cox (1972) model when the proportional hazards assumption is violated. Including Lin and Ying's (1994) model as a special case, the MS is more parsimonious than Aalen's (1978) additive risks model.

We derive two-stage estimates for the parametric and nonparametric regression coefficients by bridging the first stage and second stage estimates through their asymptotic joint distribution. The estimators introduced in this paper take the form

$$\hat{\theta}_2 = \hat{\theta}_2^V - \Sigma_{21}\Sigma_{11}^{-1}(\hat{\theta}_1^V - \hat{\theta}_1),$$

where $\hat{\theta}_2^V$ is the second stage estimator, $\hat{\theta}_1^V$ and $\hat{\theta}_1$ are first stage estimators for different sample sizes, Σ_{21} is the covariance matrix between $\hat{\theta}_2^V$ and $\hat{\theta}_1^V$, and Σ_{11} is the variance of $\hat{\theta}_1^V$. The second stage estimator $\hat{\theta}_2^V$ can be improved by incorporating the information from the first stage, i.e. $\hat{\theta}_1^V - \hat{\theta}_1$. The use of information from both stages allows us to fit a model with full information. A major challenge in establishing the asymptotic joint distribution of the first-stage and second-stage estimates in our model is the loss of the martingale property, which is the key to the theoretical development of the MS model for the single stage estimates. Moreover, we need to derive the properties of the MS estimators under a misspecified model. We use a different approach to study the asymptotic joint distribution by deriving i.i.d. representations. The same approach is then used to establish large sample properties of the proposed two-stage estimates and to develop large sample inferences.

Our methods are developed under a very general setting that incorporates measurement errors on both covariates and the survival outcome without requiring specific model specifications for the errors. No assumption is needed for the relationship between surrogate variables and target variables. We allow misspecified models for the first stage data and derive a robust sandwich variance estimate for the MS model.

The paper is organized as follows. In Section 2, we study the properties of the single stage MS estimators under misspecified models, and propose two-stage estimators for the regression coefficients and the conditional survival function. Large sample properties of our proposed estimators are given in Section 3. Pointwise and simultaneous confidence intervals for the conditional survival function are derived. Section 4 presents a simulation study to evaluate the performance of our methods. In Section 5, we illustrate our method using small intestine cancer data from the Surveillance, Epidemiology, and End Results (SEER) Program supported by the National Cancer Institute (NCI). Section 6 provides some concluding remarks. The proofs are provided in the appendix.

2. Two-Stage Estimators

2.1. Notation and assumptions

Supposed there are N subjects in the first stage and only coarse measurements, denoted as $(x_{1i}, z_{1i}, T_{1i}, \delta_{1i})$, $i = 1, \dots, N$, are available. Here $x_{1i} \in R^{p_1}$, $z_{1i} \in R^{q_1}$ are the observed surrogate covariates that might depend on time, $T_{1i} = \min\{\tilde{T}_{1i}, C_{1i}\}$, \tilde{T}_{1i} is a survival time, C_{1i} is a censoring time conditionally independent of \tilde{T}_{1i} given the covariates, and $\delta_{1i} = I(\tilde{T}_{1i} \leq C_{1i})$ is the censoring indicator. In the second stage, accurate data $(x_{2i}, z_{2i}, T_{2i}, \delta_{2i})$, $i \in V(n)$, are collected for a random validation subsample $V(n)$ of n subjects from the first stage, where $x_{2i} \in R^{p_2}$, $z_{2i} \in R^{q_2}$, $T_{2i} = \min\{\tilde{T}_{2i}, C_{2i}\}$, \tilde{T}_{2i} is the true survival time, C_{2i}

is a censoring time conditionally independent of \tilde{T}_{2i} given the covariates x_{2i} and z_{2i} , and $\delta_{2i} = I(\tilde{T}_{2i} \leq C_{2i})$ is the censoring indicator.

Assume the following MS model for the survival time \tilde{T}_{2i} :

$$\begin{aligned} h_2(t|x_{2i}, z_{2i}) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq \tilde{T}_{2i} < t + \Delta t | \tilde{T}_{2i} \geq t, x_{2i}, z_{2i})}{\Delta t} \\ &= \alpha_2(t)'x_{2i} + \beta_2'z_{2i}, \end{aligned}$$

for $i \in V(n)$. For the second stage sample, let

$$\begin{aligned} \hat{\beta}_2^V &= \left[\int_0^\tau (Z_2^V)' \hat{H}_2^V Z_2^V dt \right]^{-1} \left[\int_0^\tau (Z_2^V)' \hat{H}_2^V dN_2^V(t) \right], \\ \hat{A}_2^V(t) &= \int_0^t [(X_2^V)' \hat{W}_2^V X_2^V]^{-1} (X_2^V)' \hat{W}_2^V [dN_2^V(t) - Z_2^V \hat{\beta}_2^V ds], \end{aligned}$$

be the weighted least squares estimators of β_2 and $A_2(t) = \int_0^t \alpha_2(u)du$, respectively (McKeague and Sasieni (1994)), where $Z_2^V = Z_2^V(t) = [z_{21}Y_{21}(t), \dots, z_{2n}Y_{2n}(t)]'$, $X_2^V = X_2^V(t) = [x_{21}Y_{21}(t), \dots, x_{2n}Y_{2n}(t)]'$, $Y_{2i}(t) = I(T_{2i} \geq t)$ is the at-risk process, $N_2^V(t) = [N_{21}(t), \dots, N_{2n}(t)]'$, $N_{2i}(t) = I(T_{2i} \leq t, \delta_{2i} = 1)$ is the counting process,

$$\hat{H}_2^V = \hat{W}_2^V - \hat{W}_2^V X_2^V [(X_2^V)' \hat{W}_2^V X_2^V]^{-1} (X_2^V)' \hat{W}_2^V,$$

with $\hat{W}_2^V = \hat{W}_2^V(t) = \text{diag}[1/\hat{h}_{21}(t), \dots, 1/\hat{h}_{2n}(t)]$, $\hat{h}_{2i}(t)$ is a uniformly consistent estimate of the weight function $h_{2i}(t) \equiv h_2(t|x_{2i}, z_{2i})$ for subject i , and τ is the last time point in the study (see a more rigorous definition in the appendix). Similarly, we define the first stage estimators $[\hat{\beta}_1, \hat{A}_1(t)]$ and $[\hat{\beta}_1^V, \hat{A}_1^V(t)]$ using the first stage data based on all N subjects and the n subjects in the validation sample, respectively.

2.2. Asymptotic properties of the MS estimators under misspecified models

Our theorem gives large sample properties of $\hat{\beta}_1^V$ and $\hat{A}_1^V(t)$ without making any model assumption for the first stage data. It is a nontrivial generalization of the result of MS by allowing the model to be misspecified. Scheike (2002) considered a particular misspecification of the MS model, where the form holds only for the rate function and not the intensity, while our results work for general misspecification.

It is shown in the appendix that $\sqrt{n}(\hat{\beta}_1^V - \beta_1)$ is equivalent to a sum of independent and identically distributed random variables with mean zero,

$$\sqrt{n}(\hat{\beta}_1^V - \beta_1) = \Phi_1^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n w_{1i} \right) + o_p(1),$$

where $\beta_1 = \Phi_1^{-1} E \left(\int_0^\tau u_{1i}(t) dN_{1i}(t) \right)$ is the unknown working parameter, $\Phi_1 = \int_0^\tau [R_1(t) - V_1(t)U_1^{-1}(t)V_1(t)]dt$, and

$$w_{1i} = \left\{ \int u_{1i}(t) dN_{1i}(t) - E \left(\int u_{1i}(t) dN_{1i}(t) \right) \right\} - \int \left[Y_{1i}(t) \frac{z_{1i}x'_{1i}}{h_{1i}(t)} - V_1'(t) \right] U_1^{-1}(t) d\tilde{F}^x(t) + \int V_1'(t) U_1^{-2}(t) \left[Y_{1i}(t) \frac{x_{1i}x'_{1i}}{h_{1i}(t)} - U_1(t) \right] d\tilde{F}^x(t),$$

with

$$u_{1i}(t) = Y_{1i}(t) \frac{[z_{1i} - V_1'(t)U_1^{-1}(t)x_{1i}]}{h_{1i}(t)},$$

$$R_1(t) = \lim_{n \rightarrow \infty} \tilde{R}_1^V(t), \quad \tilde{R}_1^V(t) = \frac{1}{n} Z_1^V(t)' W_1^V(t) Z_1^V(t),$$

$$V_1(t) = \lim_{n \rightarrow \infty} \tilde{V}_1^V(t), \quad \tilde{V}_1^V(t) = \frac{1}{n} X_1^V(t)' W_1^V(t) Z_1^V(t),$$

$$U_1(t) = \lim_{n \rightarrow \infty} \tilde{U}_1^V(t), \quad \tilde{U}_1^V(t) = \frac{1}{n} X_1^V(t)' W_1^V(t) X_1^V(t),$$

$$\tilde{F}^x(t) \equiv \lim_{n \rightarrow \infty} \tilde{F}_n^x(t), \quad \tilde{F}_n^x(t) = \frac{1}{n} \int_0^t X_1^V(s)' \hat{W}_1 dN_1(s).$$

The variance of $\sqrt{n}(\hat{\beta}_1^V - \beta_1)$ is therefore $\Phi_1^{-1} E (w_{1i}^{\otimes 2}) \Phi_1^{-1}$ and can be consistently estimated by $(\hat{\Phi}_1^V)^{-1} [(1/n) \sum_{i=1}^n (\hat{w}_{1i}^V)^{\otimes 2}] (\hat{\Phi}_1^V)^{-1}$ with the unknown quantities replaced by their estimates.

Similarly, we prove in the appendix that

$$\sqrt{n}[\hat{A}_1(t) - A_1(t)] = \frac{1}{\sqrt{n}} \sum_{i=1}^n v_{1i}(t) + o_p(1),$$

where

$$A_1(t) = \int_0^t U_1^{-1}(s) [d\tilde{F}^x(s) - V_1(s)\beta_1 ds], \tag{2.1}$$

$$v_{1i}(t) = \int_0^t U_1^{-1}(s) \left[Y_{1i}(s) \frac{x_{1i}}{h_{1i}(s)} dN_{1i}(s) - d\tilde{F}^x(s) \right] - \int_0^t U_1^{-2}(s) \left[Y_{1i}(s) \frac{x_{1i}x'_{1i}}{h_{1i}(s)} - U_1(s) \right] d\tilde{F}^x(s) + \int_0^t U_1^{-2}(s) \left[Y_{1i}(s) \frac{x_{1i}x'_{1i}}{h_{1i}(s)} - U_1(s) \right] V_1(s) ds \cdot \beta_1$$

$$\begin{aligned}
 & - \int_0^t U_1^{-1}(s) \left[Y_{1i}(s) \frac{x_{1i} z'_{1i}}{h_{1i}(s)} - V_1(s) \right] ds \cdot \beta_1 \\
 & - \int_0^t U_1^{-1}(s) V_1(s) ds \cdot w_{1i}.
 \end{aligned} \tag{2.2}$$

The pointwise variance of the asymptotic Gaussian process can be estimated by $(1/n) \sum_{i=1}^n [\hat{v}_{1i}(t)]^{\otimes 2}$, with the unknown quantities replaced by the estimates.

The asymptotic results for a misspecified MS model is summarized in the theorem, and the proof is in the appendix.

Theorem 1. *Under the regularity conditions (C1)–(C3) stated in the appendix, in a misspecified model*

$$\sqrt{n}(\hat{\beta}_1^V - \beta_1) \longrightarrow_d \mathcal{N}(0, \Sigma_{\beta,11}) \quad \text{as } n \rightarrow \infty,$$

where $\Sigma_{\beta,11} = \Phi_1^{-1} E(w_{1i}^{\otimes 2}) \Phi_1^{-1}$ with $a^{\otimes 2} = aa'$. The variance $\Sigma_{\beta,11}$ can be consistently estimated by $\hat{\Sigma}_{\beta,11} = (\hat{\Phi}_1^V)^{-1} [(1/n) \sum_{i=1}^n (\hat{w}_{1i}^V)^{\otimes 2}] (\hat{\Phi}_1^V)^{-1}$, with $\hat{\Phi}_1^V = (1/n) \int (Z_1^V)' \hat{H}_1^V Z_1^V dt$ and \hat{w}_{1i}^V defined in (A.2). Moreover,

$$\sqrt{n}[\hat{A}_1^V(\cdot) - A_1(\cdot)] \longrightarrow_d \mathcal{W}_1^V(\cdot) \quad \text{in } \mathcal{D}[0, \tau] \text{ as } n \rightarrow \infty,$$

where $\mathcal{D}[0, \tau]$ is the standard Skorohod space on $[0, \tau]$, $\tau = \sup\{t : S_1(t|x, z) S_2(t|x, z) S_C(t|x, z) > 0 \text{ for all } x, z\}$ (see regularity assumption (C1) in Appendix), and $\mathcal{W}_1^V(\cdot)$ is a zero-mean Gaussian process with covariance function $\kappa_1(t_1, t_2) = E \left\{ \left[\int_0^{t_1} v_{1i}(s) ds \right] \cdot \left[\int_0^{t_2} v_{1i}(s) ds \right]' \right\}$. The variance function of $\sqrt{n}[\hat{A}_1^V(t) - A_1(t)]$ is given by $\Sigma_{A,11}(t) = \kappa_1(t, t)$, which can be consistently estimated by $\hat{\Sigma}_{A,11}(t) = (1/n) \sum_{i=1}^n [\hat{v}_{1i}(t)]^{\otimes 2}$ with the unknown quantities replaced by the estimates.

The estimators $\hat{\beta}_1$ and $\hat{A}_1(t)$, based on all the N first stage subjects, have the same asymptotic properties as stated in Theorem 1. Notice that no model is assumed for the first stage data in Theorem 1. If the MS model holds for the first stage data, then β_1 and $A_1(t)$ coincide with the regression parameters in the true MS model.

2.3. Two-stage estimators of β and $A(t)$

To develop the two-stage estimator for β_2 , we first give the joint distribution of $(\hat{\beta}_1^V, \hat{\beta}_2^V)$.

Lemma 1. *Assume the regularity conditions (C1)–(C3) given in the appendix, then*

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_1^V - \beta_1 \\ \hat{\beta}_2^V - \beta_2 \end{pmatrix} \longrightarrow_d \mathcal{N} \left(0, \begin{pmatrix} \Sigma_{\beta,11} & \Sigma_{\beta,12} \\ \Sigma_{\beta,21} & \Sigma_{\beta,22} \end{pmatrix} \right) \quad \text{as } n \rightarrow \infty, \tag{2.3}$$

where $\Sigma_{\beta,gg} = \Phi_g^{-1} E(w_{gi}^{\otimes 2})\Phi_g^{-1}$, $\Sigma_{\beta,12} = \Phi_1^{-1} E(w_{1i} \cdot w'_{2i})\Phi_2^{-1}$, and $g = 1, 2$ indicates the stage. The covariance matrix can be estimated by $\hat{\Sigma}_{\beta,gg} = (\hat{\Phi}_g^V)^{-1} \cdot [(1/n) \sum_{i \in V(n)} (\hat{w}_{gi}^V)^{\otimes 2}] (\hat{\Phi}_g^V)^{-1}$, and $\hat{\Sigma}_{\beta,12} = (\hat{\Phi}_1^V)^{-1} [(1/n) \sum_{i \in V(n)} \hat{w}_{1i}^V(t) \cdot \hat{w}'_{2i}(t)'] (\hat{\Phi}_2^V)^{-1}$. Here β_2 , Φ_2 , $w_{2i}(t)$, $\hat{\Phi}_2^V$, and $\hat{w}_{2i}^V(t)$ are defined similarly as β_1 , Φ_1 , $w_{1i}(t)$, $\hat{\Phi}_1^V$, and $\hat{w}_{1i}^V(t)$.

It follows from (2.3) that $E(\hat{\beta}_2^V - \beta_2 | \hat{\beta}_1^V - \beta_1) \approx \Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} (\hat{\beta}_1^V - \beta_1)$. This suggests that β_2 be estimated by

$$\hat{\beta}_2 = \hat{\beta}_2^V - \hat{\Sigma}_{\beta,21} \hat{\Sigma}_{\beta,11}^{-1} (\hat{\beta}_1^V - \hat{\beta}_1). \tag{2.4}$$

Next, we consider the joint distribution of $\hat{A}_1^V(t)$ and $\hat{A}_2^V(t)$.

Lemma 2. *Let $A_2(t)$ and $v_{2i}(t)$ be defined similarly as (2.1) and (2.2), respectively, based on the second stage sample. Under the regularity conditions (C1)–(C3), as $n, N \rightarrow \infty$, $n/N \rightarrow \rho$,*

$$\sqrt{n} \begin{pmatrix} \hat{A}_1^V(t) - A_1(t) \\ \hat{A}_2^V(t) - A_2(t) \end{pmatrix} \rightarrow_d \begin{pmatrix} \mathcal{W}_1^V(t) \\ \mathcal{W}_2^V(t) \end{pmatrix}$$

in $\mathcal{D}[0, \tau]^2 = \mathcal{D}[0, \tau] \times \mathcal{D}[0, \tau]$, where $[\mathcal{W}_1^V(t), \mathcal{W}_2^V(t)]$ is a zero-mean Gaussian random field, with variance-covariance function

$$\begin{pmatrix} \Sigma_{A,11}(t_1) & \Sigma_{A,12}(t_1, t_2) \\ \Sigma_{A,21}(t_1, t_2) & \Sigma_{A,22}(t_2) \end{pmatrix},$$

where $\Sigma_{A,kl}(t_1, t_2) = E[v_{ki}(t_1) \cdot v_{li}(t_2)']$ for $k, l \in \{1, 2\}$, and $\Sigma_{A,gg}(t) = \Sigma_{A,gg}(t, t)$. The variance and covariance functions can be consistently estimated by $\hat{\Sigma}_{A,kl}(t_1, t_2)$ as defined in (A.4) in the Appendix.

By Lemma 2 and the argument leading to (2.4), we take

$$\hat{A}_2(t) = \hat{A}_2^V(t) - \hat{\Sigma}_{A,21}(t) \hat{\Sigma}_{A,11}^{-1}(t) [\hat{A}_1^V(t) - \hat{A}_1(t)], \tag{2.5}$$

where $\hat{\Sigma}_{A,21}(t) = \hat{\Sigma}_{A,21}(t, t)$.

Our two-stage estimators possess some appealing properties. In particular, if the first stage data are barely correlated with the second stage data, then the proposed estimate $\hat{\beta}_2$ is close to the second stage estimate $\hat{\beta}_2^V$. This is a desirable property since the first stage data are not expected to contribute much useful information for estimating β_2 . The same comment applies to $\hat{A}_2(t)$. It can also be easily verified that when the first stage sample contains precise and complete information, the proposed estimates $\hat{\beta}_2$ and $\hat{A}_2(t)$ are identical to the estimates $\hat{\beta}_1$ and $\hat{A}_1(t)$. This means that we should use all the first stage data to estimate

the parameters and make statistical inference when no bias is present in the first stage sample.

This method is general enough to allow variables to have different types of coefficients in the two stages, as well as different sets of covariates for the first- and second-stage models. However, we recommend using the same type of coefficients for a variable in both stages whenever possible, based on the intuition that the effects of a variable are expected to be similar for both stages. For example, we use constant coefficients for age and gender for both the first- and second-stage models in our data example in Section 5.

3. Asymptotic Properties and Inferences

3.1. Asymptotic properties of $(\hat{\beta}_2, \hat{A}_2(t))$

The following result states the weak convergence property of the joint distribution of the proposed estimators $\hat{\beta}_2$ and $\hat{A}_2(t)$.

Theorem 2. *Under conditions (C1)–(C3), we have*

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_2 - \beta_2 \\ \hat{A}_2(t) - A_2(t) \end{pmatrix} \longrightarrow_d \begin{pmatrix} Z_2 \\ \mathcal{W}_2(t) \end{pmatrix}, \tag{3.1}$$

where $Z_2 \sim \mathcal{N}(0, \Sigma_{\beta_2}^*)$ with $\Sigma_{\beta_2}^* = \Sigma_{\beta,22} - (1 - \rho)\Sigma_{\beta,21}\Sigma_{\beta,11}^{-1}\Sigma_{\beta,12}$, and $\mathcal{W}_2(t)$ is a zero-mean Gaussian process with covariance function

$$\begin{aligned} \zeta(s, t) = & \Sigma_{A,22}(s, t) - [\Sigma_{A,21}(s, t) - \sqrt{\rho}\Sigma_{A,21}(s, t)] \Sigma_{A,11}^{-1}(t) \Sigma_{A,12}(t) \\ & - \Sigma_{A,21}(s) \Sigma_{A,11}^{-1}(s) [\Sigma_{A,12}(s, t) - \sqrt{\rho}\Sigma_{A,12}(s, t)] \\ & + \Sigma_{A,21}(s) \Sigma_{A,11}^{-1}(s) [\Sigma_{A,11}(s, t) - 2\sqrt{\rho}\Sigma_{A,11}(s \wedge t) \\ & + \rho\Sigma_{A,11}(s, t)] \Sigma_{A,11}^{-1}(t) \Sigma_{A,12}(t), \end{aligned}$$

with $\Sigma_{A,kl}(s, t)$ defined in Lemma 2. The variance of $\mathcal{W}_2(t)$ is $\Sigma_{A_2}^*(t) = \zeta(t, t) = \Sigma_{A,22}(t) - (1 - \rho)\Sigma_{A,21}(t)\Sigma_{A,11}^{-1}(t)\Sigma_{A,12}(t)$. The covariance between Z_2 and $\mathcal{W}_2(t)$ is

$$\begin{aligned} \Sigma_{\beta_2 A_2}^*(t) = & \Sigma_{\beta_2 A_2}(t) - (1 - \sqrt{\rho})\Sigma_{\beta_2 A_1}(t)\Sigma_{A,11}^{-1}(t)\Sigma_{A,12}(t) \\ & - (1 - \sqrt{\rho})\Sigma_{\beta,21}\Sigma_{\beta,11}^{-1}\Sigma_{\beta_1 A_2}(t) \\ & + (1 - 2\sqrt{\rho} + \rho)\Sigma_{\beta,21}\Sigma_{\beta,11}^{-1}\Sigma_{\beta_1 A_1}(t)\Sigma_{A,11}^{-1}(t)\Sigma_{A,12}(t), \end{aligned}$$

where $n/N \rightarrow \rho$ for some constant $0 < \rho < 1$ as $n, N \rightarrow \infty$, and $\Sigma_{\beta_k A_l}(t) = \Phi_k^{-1} \cdot \mathbb{E}\{w_{ki} \cdot v_{li}(t)'\}$ for $k, l \in \{1, 2\}$.

Furthermore, $\Sigma_{\beta_2}^*$ can be consistently estimated by $\hat{\Sigma}_{\beta_2}^* = \hat{\Sigma}_{\beta,22} - (1 - (n/N))\hat{\Sigma}_{\beta,21}$.

$\hat{\Sigma}_{\beta,11}^{-1} \hat{\Sigma}_{\beta,12}$, and $\Sigma_{A_2}^*(t)$ can be consistently estimated by $\hat{\Sigma}_{A_2}^*(t) = \hat{\Sigma}_{A,22}(t) - (1 - (n/N)) \hat{\Sigma}_{A,21}(t) \hat{\Sigma}_{A,11}^{-1}(t) \hat{\Sigma}_{A,12}(t)$ for any $t \in [0, \tau]$. The covariance function $\Sigma_{\beta_2 A_2}^*(t)$ can be estimated by

$$\begin{aligned} \hat{\Sigma}_{\beta_2 A_2}^*(t) &= \hat{\Sigma}_{\beta_2 A_2}(t) - (1 - \sqrt{\frac{n}{N}}) \hat{\Sigma}_{\beta_2 A_1}(t) \hat{\Sigma}_{A,11}^{-1}(t) \hat{\Sigma}_{A,12}(t) \\ &\quad - (1 - \sqrt{\frac{n}{N}}) \hat{\Sigma}_{\beta,21} \hat{\Sigma}_{\beta,11}^{-1} \hat{\Sigma}_{\beta_1 A_2}(t) \\ &\quad + (1 - 2\sqrt{\frac{n}{N}} + \frac{n}{N}) \hat{\Sigma}_{\beta,21} \hat{\Sigma}_{\beta,11}^{-1} \hat{\Sigma}_{\beta_1 A_1}(t) \hat{\Sigma}_{A,11}^{-1}(t) \Sigma_{A,12}(t) \end{aligned}$$

with $\hat{\Sigma}_{\beta_k A_l}(t) = \hat{\Phi}_k^{-1} \cdot \{ (1/n) \sum_{i=1}^n \hat{w}_{ki} \cdot [\int_0^t \hat{v}_{li}(s) ds]' \}$.

Obviously, $\Sigma_{\beta_2}^* \leq \Sigma_{\beta,22}$ (i.e. $\Sigma_{\beta,22} - \Sigma_{\beta_2}^*$ is nonnegative definite). Hence our proposed two-stage estimators are asymptotically more efficient than the estimators using the second stage data alone. We will compare their finite sample performance in Section 4.

3.2. Estimation of the conditional survival function

We consider the problem of estimating the conditional cumulative hazard function $H_2(t) = H_2(t|x_0, z_0)$ and the conditional survival function $S_2(t) = S_2(t|x_0, z_0)$, for some given covariates x_0, z_0 . Let $\hat{H}_2(t) = \hat{A}_2(t)'x_0 + \hat{\beta}'_2 z_0 \cdot t$ and $\hat{S}_2(t) = \prod_{s \leq t} [1 - \Delta \hat{H}_2(s)]$, where $\Delta \hat{H}_2(s) = \hat{H}_2(s) - \hat{H}_2(s-)$.

Theorem 3. Assume that $n/N \rightarrow \rho$ for some constant $0 < \rho < 1$ as $n, N \rightarrow \infty$. Under the regularity conditions (C1)–(C3),

$$\begin{aligned} E_n(t) &= \sqrt{n}[\hat{H}_2(\cdot) - H_2(\cdot)] \longrightarrow_d \mathcal{V}_2(\cdot) \quad \text{in } \mathcal{D}[0, \tau], \\ \sqrt{n}[\hat{S}_2(\cdot) - S_2(\cdot)] &\longrightarrow_d S_2(\cdot) \mathcal{V}_2(\cdot) \quad \text{in } \mathcal{D}[0, \tau], \end{aligned}$$

where $\mathcal{V}_2(\cdot)$ is a zero-mean Gaussian process with covariance function $\Sigma_{H_2}^*(t_1, t_2)$ equal to $x_0' \zeta(t_1, t_2) x_0 + z_0' \Sigma_{\beta_2}^* z_0 \cdot t_1 t_2 + z_0' \Sigma_{\beta_2 A_2}^*(t_1) x_0 \cdot t_1 + z_0' \Sigma_{\beta_2 A_2}^*(t_2) x_0 \cdot t_2$; this can be consistently estimated by replacing each term by its estimate.

Thus at any $t \in [0, \tau]$, $100(1 - \alpha)\%$ pointwise confidence intervals for $H_2(t)$ and $S_2(t)$ are given by $\hat{H}_2(t) \pm z_{1-\alpha/2} \sqrt{\hat{\Sigma}_{H_2}(t)}$, and $\hat{S}_2(t) \pm z_{1-\alpha/2} \hat{S}_2(t) \sqrt{\hat{\Sigma}_{H_2}(t)}$, where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ th percentile of the standard normal distribution.

Notice that the proposed estimator $\hat{S}_2(t)$ is not necessarily monotonically non-increasing in t . As mentioned by Li and Tseng (2008), this problem is local and minor, especially when the sample size is large. In practice, one can improve

the estimates by the ‘Poor-Adjacent Violator’ algorithm (Barlow et al. (1972)) or some simpler modification (c.f. Lin and Ying (1994)).

Theorem 3 cannot be readily applied to construct simultaneous confidence bands for $S_2(t)$ over a given interval $[\tau_1, \tau_2]$ since the distribution of $\sup |\hat{S}_2(t) - S_2(t)|$ is intractable. Using an idea similar to that in Lin and Ying (1994), we develop a Monte Carlo method for constructing simultaneous confidence bands for $H_2(t)$ and $S_2(t)$. It can be shown that the process $E_n(t)$ in Theorem 3 is asymptotically equivalent to a sum of i.i.d. random variables. Specifically,

$$\begin{aligned} E_n(t) &= z'_0 \cdot \sqrt{n}(\hat{\beta}_2 - \beta_2) \cdot t + x'_0 \cdot \sqrt{n}[\hat{A}_2(t) - A_2(t)] \\ &= z'_0 \cdot \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_2^{-1} w_{2i} - \Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_1^{-1} w_{1i} \right. \right. \\ &\quad \left. \left. - \frac{1}{\sqrt{N}} \sqrt{\frac{n}{N}} \sum_{i=1}^N \Phi_1^{-1} w_{1i} \right] \right\} \cdot t \\ &\quad + x'_0 \cdot \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n v_{2i}(t) - \Sigma_{A,21}(t) \Sigma_{A,11}^{-1}(t) \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n v_{1i}(t) \right. \right. \\ &\quad \left. \left. - \frac{1}{\sqrt{N}} \sqrt{\frac{n}{N}} \sum_{i=1}^N v_{1i}(t) \right] \right\} + o_p(1). \end{aligned}$$

To approximate the distribution of $E_n(t)$, we define another process $\hat{E}_n(t)$ as

$$\begin{aligned} \hat{E}_n(t) &= z'_0 \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\Phi}_2^V)^{-1} \hat{w}_{2i}^V G_i - \hat{\Sigma}_{\beta,21} \hat{\Sigma}_{\beta,11}^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\Phi}_1^V)^{-1} \hat{w}_{1i}^V G_i \right. \right. \\ &\quad \left. \left. - \frac{1}{\sqrt{N}} \sqrt{\frac{n}{N}} \sum_{i=1}^N \hat{\Phi}_1^{-1} \hat{w}_{1i} G_i \right] \right\} t \\ &\quad + x'_0 \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{v}_{2i}^V(t) G_i - \hat{\Sigma}_{A,21}(t) \hat{\Sigma}_{A,11}^{-1}(t) \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{v}_{1i}^V(t) G_i \right. \right. \\ &\quad \left. \left. - \frac{1}{\sqrt{N}} \sqrt{\frac{n}{N}} \sum_{i=1}^N \hat{v}_{1i}(t) G_i \right] \right\}, \end{aligned}$$

where G_i are i.i.d. $\mathcal{N}(0, 1)$. We prove in the appendix that $E_n(t)$ and $\hat{E}_n(t)$ have the same limiting distribution.

Theorem 4. *Conditioned on the data $(x_{1i}, z_{1i}, T_{1i}, \delta_{1i}), i = 1, \dots, N$, and $(x_{2j}, z_{2j}, T_{2j}, \delta_{2j}), j \in V(n)$, the random process $\hat{E}_n(t)$ converges weakly to $\mathcal{V}_2(t)$ in $\mathcal{D}[0, \tau]$.*

Theorem 4 suggests that the limiting distribution of $E_n(t)$ can be approximated by that of $\hat{E}_n(t)$. The latter can be obtained by generating a large number of independent Monte Carlo random samples G_1, \dots, G_N from the standard normal distribution. Similar to Lin and Ying (1994), the confidence bands for $H_2(t)$ and $S_2(t)$ can be obtained as $\phi^{-1}\{\phi[\hat{H}_2(t)] \mp n^{-1/2}q_\alpha/g(t)\}$ and $\exp\{-\phi^{-1}\{\phi[\hat{H}_2(t)] \pm n^{-1/2}q_\alpha/g(t)\}\}$, where q_α is the critical value of $\Pr(\sup_{t \in [\tau_1, \tau_2]} |\sqrt{n}g(t)\phi'[\hat{H}_2(t)]\hat{E}_n(t)| > q_\alpha) = \alpha$, g is a weight function, and ϕ is a known transformation function with non-zero and continuous first derivative ϕ' . Specifically, we consider $g(t) = \hat{H}_2(t)/\hat{\Sigma}_{\hat{H}_2}^*(t)$ to get an equal-precision band, and set $\phi(t) = \log(t)$ to obtain bands on meaningful ranges and to attain better coverage probabilities.

4. Simulation Studies

We present a small simulation study to illustrate and evaluate the finite sample performance of the proposed two-stage estimators. The two-stage estimators are compared with the second stage estimators in terms of bias, variance, mean square error (MSE), and achieved confidence interval coverage probabilities.

The weight matrices $W_g(t), g = 1, 2$, in Section 2.1 require consistent estimates of the conditional hazard functions $h_{gi}(t|x_{gi}, z_{gi})$; they are given by $\hat{h}_{gi}(t|x_{gi}, z_{gi}) = \hat{\alpha}_g(t)'x_{gi} + \hat{\beta}'_g z_{gi}$, where

$$\hat{\alpha}_g(t) = \frac{1}{b} \int_{-\infty}^{\infty} K\left(\frac{t-s}{b}\right) d\hat{A}_g(s),$$

$K(t)$ is a kernel function, and b is the bandwidth. In the simulation study, we used the Epanechnikov kernel, $K(x) = 3(1-x^2)/4$ for $|x| < 1$, with $b = 0.4\tau$, around a given point t . The boundary effects are corrected by the modified asymmetric kernel proposed by Gasser and Muller (1979). In a more comprehensive simulation, Wu (2006) observed that both second stage and two-stage estimators are not very sensitive to the choice of the smoothing parameter.

The second stage survival time was generated from $h_2(t|x_{2i}, z_{2i,1}, z_{2i,2}) = \alpha_{2,0}(t) + \alpha_{2,1}(t)x_{2i} + \beta_{2,1}z_{2i,1} + \beta_{2,2}z_{2i,2}$, where $\alpha_{2,0}(t) = 1$, $\alpha_{2,1}(t) = t$, $\beta_{2,1} = \beta_{2,2} = 1$ ($q_2 = 2$), and $x_{2i}, z_{2i,1}, z_{2i,2} \sim$ i.i.d. $\text{Unif}[0, 1]$. The random censoring times $C_{2i}, i = 1, \dots, n$, were generated from $h(c|x_{2i}, z_{2i}) = 0.1 + 0.1cx_{2i} + 0.5z_{2i,1} + 0.5z_{2i,2}$. About 20 percent of the subjects were censored under this model. In the first stage, we considered a general situation by incorporating both the measurement error problem and missing covariate problem. The working model for the stage data was $h_1(t|x_{1i}, z_{1i,1}) = \alpha_{1,0}(t) + \alpha_{1,1}(t)x_{1i} + \beta_{1,1}z_{1i,1}$, where $\tilde{T}_{1i} = \tilde{T}_{2i} + \text{Unif}[0, 0.1]$, $x_{1i} = x_{2i} + \text{Unif}[0, 0.1]$, $z_{1i,1} = z_{2i,1} + \text{Unif}[0, 0.5]$, and the covariate $z_{2i,2}$ is missing for all subjects. We generated 1,000 Monte Carlo samples for each size $(n, N) = (100, 1,000), (500, 1,000),$ and $(500, 2,000)$.

With different sample sizes at both the first and second stages, we can evaluate: (1) the performance of the variance estimator under a misspecified model at finite sample sizes; and (2) the improvement of the second stage estimator when incorporating different amount of information from the first stage.

Table 1 presents the bias, variance, estimated variance, and MSE of $\hat{\beta}_1^V$, $\hat{\beta}_1$, $\hat{\beta}_2^V$, and $\hat{\beta}_2$, together with the achieved coverage probabilities of the respective confidence intervals. When the second stage sample was small ($n = 100$), the variance of the second stage estimate $\hat{\beta}_{2,1}^V$ (0.69) was underestimated (0.58). As a comparison, we can see that the first stage estimates had large bias, which indicates that the information from the first stage was biased. Even so, there was gain from combining the first stage data of $z_{1i,1}$. The estimated variance of $\hat{\beta}_{2,1}$ (0.23) was close to the variance (0.24), which was much smaller than the variance of the second stage estimate. We also observed a better coverage probability of $\hat{\beta}_{2,1}$ than $\hat{\beta}_{2,1}^V$. On the other hand, $\hat{\beta}_{2,2}$ showed little improvement over $\hat{\beta}_{2,2}^V$. This can be explained by the fact that the first stage data contained no information on $z_{2i,2}$. As n increased to 500, the second stage estimates were improved with better variance estimates and higher coverage probabilities – the variance estimator for a misspecified model worked well when the sample size got bigger. Our proposed estimates $\hat{\beta}_{2,1}$ had smaller variances, better variance estimates, and better coverage probabilities. As N increased to 2,000 and n remained at 500, the variance of $\hat{\beta}_{2,1}$ got further reduced since more information from stage one was obtained. As for comparison, Tables 1 also shows the “ideal” estimates, denoted as $\tilde{\beta}_{21}$ and $\tilde{\beta}_{22}$, using the complete and accurate information for all N subjects.

Figure 1 depicts the mean and variance of two-stage and second stage estimates of $A_{2,0}(t)$, $A_{2,1}(t)$, and $S_2(t|z_0 = (0.5, 0, 5), x_0 = 0.5)$, respectively, for sample size $(n, N) = (100, 1, 000)$. The two-stage estimates (thick solid line) show less bias at the right tail than the second stage estimates (dashed line). Moreover, the two-stage estimates have much smaller variances throughout. Table 2 presents the simulated coverage probabilities of the pointwise 95% confidence intervals of $A(t)$ and $S(t|z_0, x_0)$ for $(n, N) = (100, 1, 000)$. The coverage probabilities are quite satisfactory for most time points. As sample size increases, the performance is improved in the far right tail. The results for $(n, N) = (500, 1, 000)$ and $(500, 2, 000)$ are similar and thus not reported here.

5. An Example

5.1. Data description

We illustrate our method using a data set on small intestine cancer from the SEER program supported by NCI. Surgery and radiation therapy are the most commonly used treatments for small intestine cancer. In this study we wanted to know how these treatments affect both survival time and the development of

Table 1. Simulated bias, variance, means square error (MSE), and 95% coverage probability (CP) of $\hat{\beta}_1^V$, $\hat{\beta}_1$, $\hat{\beta}_2^V = (\hat{\beta}_{21}^V, \hat{\beta}_{22}^V)$ and $\hat{\beta}_2 = (\hat{\beta}_{21}, \hat{\beta}_{22})$. The true parameter value was $\beta_2 = (\beta_{21}, \beta_{22}) = (1, 1)$.

(n, N)	Estimate	Bias	Var	Est. Var	MSE	95% CP
(100, 1, 000)	$\hat{\beta}_{11}^V$ (Stage 1)	-0.290	0.380	0.310	NA	NA
	$\hat{\beta}_{11}$ (Stage 1)	-0.320	0.037	0.033	NA	NA
	$\hat{\beta}_{21}^V$ (Stage 2)	0.040	0.690	0.580	0.690	0.910
	$\hat{\beta}_{21}$ (Proposed)	< 0.001	0.240	0.230	0.240	0.940
	$\tilde{\beta}_{21}$ (Ideal)	< 0.001	0.060	0.061	0.061	0.950
	$\hat{\beta}_{22}^V$ (Stage 2)	-0.060	0.650	0.580	0.660	0.920
	$\hat{\beta}_{22}$ (Proposed)	-0.040	0.670	0.570	0.680	0.910
	$\tilde{\beta}_{22}$ (Ideal)	-0.020	0.058	0.061	0.058	0.950
(500, 1, 000)	$\hat{\beta}_{11}^V$	-0.320	0.075	0.075	NA	NA
	$\hat{\beta}_{11}$	-0.310	0.038	0.039	NA	NA
	$\hat{\beta}_{21}^V$ (Stage 2)	-0.010	0.130	0.120	0.130	0.940
	$\hat{\beta}_{21}$ (Proposed)	-0.010	0.078	0.077	0.078	0.950
	$\tilde{\beta}_{21}$ (Ideal)	< 0.001	0.060	0.061	0.060	0.950
	$\hat{\beta}_{22}^V$ (Stage 2)	0.010	0.130	0.120	0.130	0.940
	$\hat{\beta}_{22}$ (Proposed)	0.010	0.130	0.120	0.130	0.940
	$\tilde{\beta}_{22}$ (Ideal)	-0.010	0.058	0.060	0.058	0.950
(500, 2, 000)	$\hat{\beta}_{11}^V$	-0.310	0.074	0.075	NA	NA
	$\hat{\beta}_{11}$	-0.320	0.019	0.020	NA	NA
	$\hat{\beta}_{21}^V$ (Stage 2)	< 0.001	0.120	0.120	0.120	0.950
	$\hat{\beta}_{21}$ (Proposed)	< 0.001	0.054	0.054	0.054	0.950
	$\tilde{\beta}_{21}$ (Ideal)	< 0.001	0.030	0.031	0.031	0.950
	$\hat{\beta}_{22}^V$ (Stage 2)	< 0.001	0.130	0.120	0.130	0.940
	$\hat{\beta}_{21}$ (Proposed)	< 0.001	0.130	0.120	0.130	0.940
	$\tilde{\beta}_{21}$ (Ideal)	< 0.001	0.030	0.030	0.030	0.950

Table 2. Simulated coverage probabilities of the nonparametric two-stage estimators at nominal level 95%.

Time	0.2	0.4	0.6	0.8	1.0	1.2
$\hat{A}_{2,0}(t)$	0.91	0.92	0.91	0.93	0.93	0.90
$\hat{A}_{2,1}(t)$	0.95	0.96	0.96	0.95	0.92	0.90
$\hat{S}_2(t z_0, x_0)$	0.95	0.95	0.95	0.94	0.93	0.92

subsequent tumors. Therefore, we defined the survival time as the time from the diagnosis of the first primary small intestine cancer to the diagnosis of the second primary cancer or death. We considered eleven covariates: *surgery status* (1 if

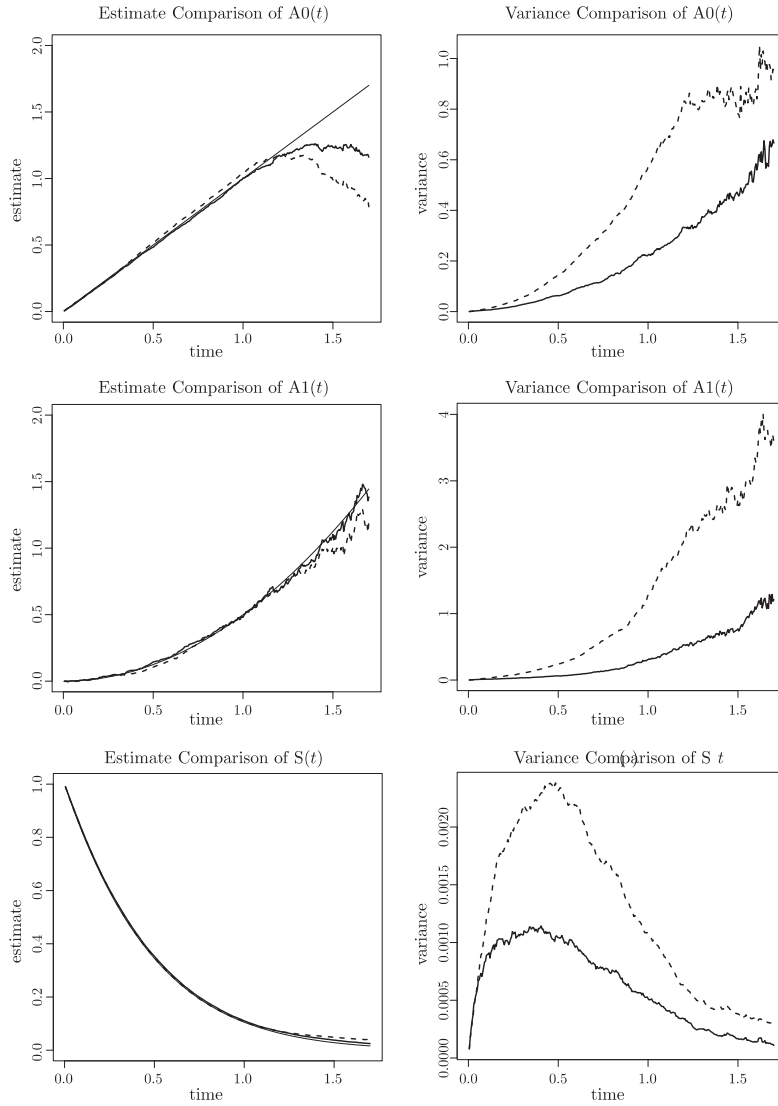


Figure 1. Comparison of the two-stage estimates (thick solid line) and second stage estimates (dash line) with the true coefficients (solid line) for sample size $(n, N) = (100, 1,000)$. The top panel gives the estimates and variances of $\hat{A}_{2,0}^V(t)$ and $\hat{A}_{2,0}(t)$; the middle panel shows the estimates and variances of $\hat{A}_{2,1}^V(t)$ and $\hat{A}_{2,1}(t)$; the bottom panel gives the estimates and variances of $\hat{S}_{2,0}^V(t|z_0, x_0)$ and $\hat{S}_{2,0}(t|z_0, x_0)$. The first column is the plot of point estimates compared to the true value (solid line), and the second column is the plot of variance estimates.

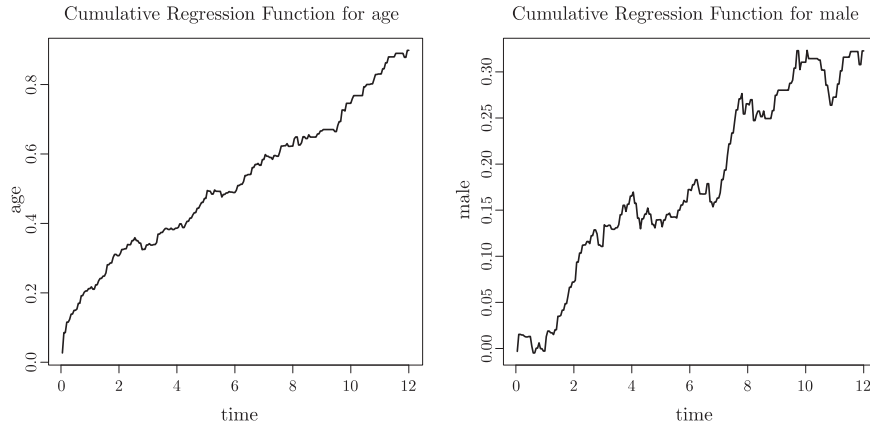


Figure 2. Plots of cumulative hazard function for *age* and *gender* show linear trends.

yes, or 0 if no), *radiation therapy* (1 if yes, or 0 if no), *age* at the first primary cancer diagnosis (1 if age < 60, or 0 if age \geq 60), *gender* (1 if male, or 0 if female), dummy variable *race* (*black*, *other races*, and the reference group *white*), dummy variable *stage* (*regional stage*, *distant stage*, and the reference group *local stage*), and dummy variable *tumor grade* (*grade II*, *grade III*, *grade IV*, and the reference group *grade I*).

To illustrate our method, we constructed a two-stage design data set as follows. The second stage sample consists of 300 patients (censoring rate = 33.7%) randomly chosen from the 2,669 patients (censoring rate = 26.5%) in the data set with all eleven covariates and survival information. The first stage data include all 2,669 patients, however the variable *tumor grade* is missing. The 2,669 patients with all variables were used as the reference population.

5.2. Analysis results

With all eleven covariates as time-dependent variables, we plotted the cumulative hazard function $\hat{A}(t) = \int_0^t \alpha(u) du$ for each variable based on the Aalen (1978) nonparametric additive risks model (Wu (2006)). The linear trends of *age* and *gender* (Figure 2) suggest that these two variables have time-independent effects and might be used as Z ($q_2 = 2$) in the MS model. The nonlinear trend of the other nine covariates (*radiation*, *race*, *tumor grade*, *stage*, and *surgery*) suggests that they be used as X ($p_2 = 9$). For more information on how to assign x and z , refer to Martinussen and Scheike (2006).

With *age* and *gender* as covariate Z ($q_2 = 2$) with time-independent effects, and the other nine covariates (*radiation therapy*, *black*, *other races*, *grade II*, *grade III*, *grade IV*, *regional stage*, *distant stage*, and *surgery status*) as X

Table 3. Comparison of the second stage and the proposed two-stage estimators for *age* and *gender*.

Covariate	$\hat{\beta}_2^Y$		$\hat{\beta}_2$	
	Point Est.	95% Conf. Interval	Point Est.	95% Conf. Interval
<i>age</i>	0.053	(0.017, 0.089)	0.062	(0.046, 0.078)
<i>gender</i>	0.036	(0.001, 0.072)	0.031	(0.016, 0.047)

($p_2 = 9$) with time-dependent effects, Table 3 compares the proposed and second stage estimates for *age* and *gender*. Both methods show significant higher risks associated with being male and older (≥ 60 years). We note that both confidence intervals cover the reference parameter values obtained from the complete data (“ideal” estimates), but the interval based on the two-stage estimates is much narrower than the second stage estimate.

Figure 3 displays the estimates and confidence bands of the cumulative regression functions for *surgery* and *radiotherapy*, which are of most interest in this study. It is seen from Figure (3b) that, after adjusting for other factors, surgery significantly reduces the risk of second primary cancer or death during the first 2.5 years; after this time period, the efficacy of surgery decreases. However, the effects of surgery are inconclusive based on the second stage estimate (Figure (3a)), since its confidence band is very wide and contains zero. Radiation therapy, as opposed to surgery, seems to have no significant impact on subsequent cancer development. We note that the proposed two-stage estimators are usually closer to the reference estimate from the complete data, with much narrower confidence bands than the second-stage estimator.

Figure 4 depicts a 95% simultaneous confidence band for the conditional survival function for a white male patient who is diagnosed as cancer grade IV, at distant stage, younger than 60, and treated by both surgery and radiation therapy. The proposed two-stage estimate is more accurate with a much narrower confidence band. For example, the survival probabilities at year 1 and year 3 are estimated to be 0.613 (with variance 0.0016) and 0.332 (0.0019), respectively, using the two-stage estimator, and 0.625 (0.0065) and 0.340 (0.0088), respectively, using the second stage data only.

6. Discussion

We propose two-stage estimators for the partial linear semiparametric hazard model introduced by McKeague and Sasieni (1994) in a two-stage design setting. We allow measurement errors for survival time, censoring time, and covariates in the first stage data. We also allow missing covariates in the first stage. The proposed estimators are consistent and asymptotically normal. Confidence

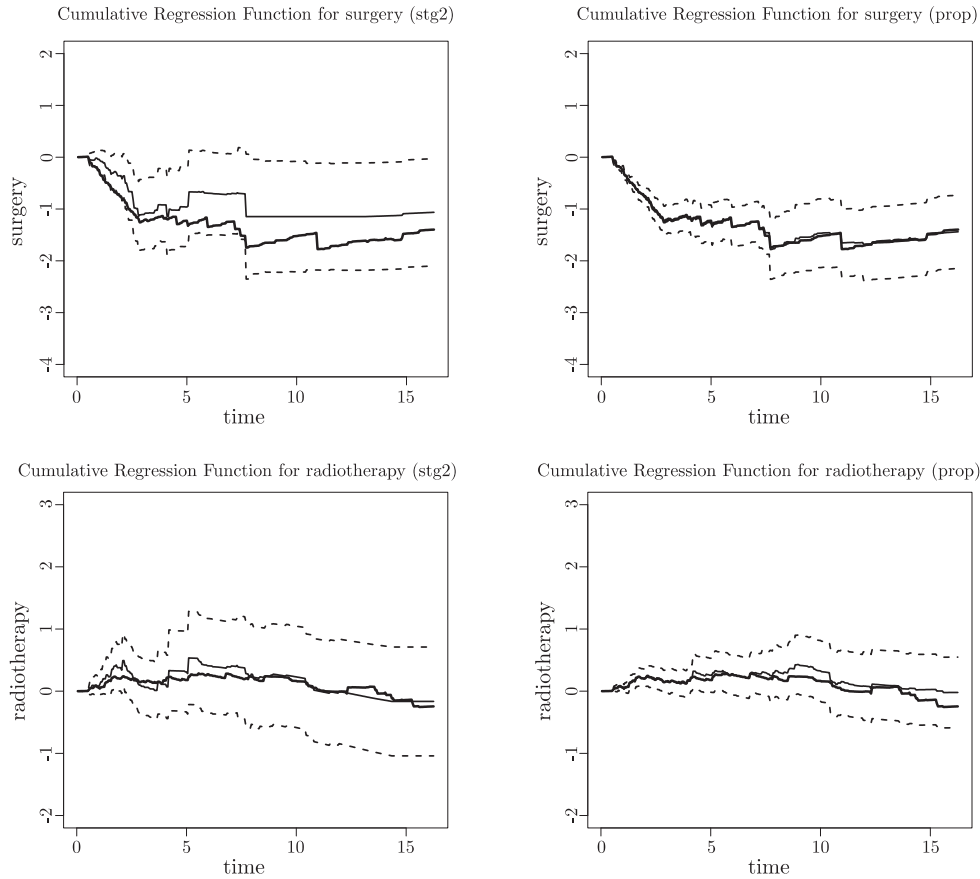


Figure 3. 95% confidence bands (dash line) of the second stage and two-stage estimators (solid line) of $A_j(t)$ for *surgery* and *radiation therapy*. The thick solid line is obtained from the complete data.

bands are developed to assess time-varying covariate effects, and to predict conditional survival probabilities. By utilizing information from the first stage, our estimators are more efficient than the second stage estimators for both large and small samples. Reduction in bias is also observed for small samples.

The estimators introduced in this paper take a form that is similar to that of a trick used for variance reduction in the theory of Monte Carlo methods (sometimes called “control variate”):

$$m^* = m - \frac{\sigma_m}{\sigma_t} \rho_{mt}(t - \tau),$$

where m^* and m are unbiased estimates for a parameter of interest, say μ , $E(t) = \tau$, σ_m and σ_t are variance of m and t , respectively, and $\rho_{mt} = corr(m, t)$. In the

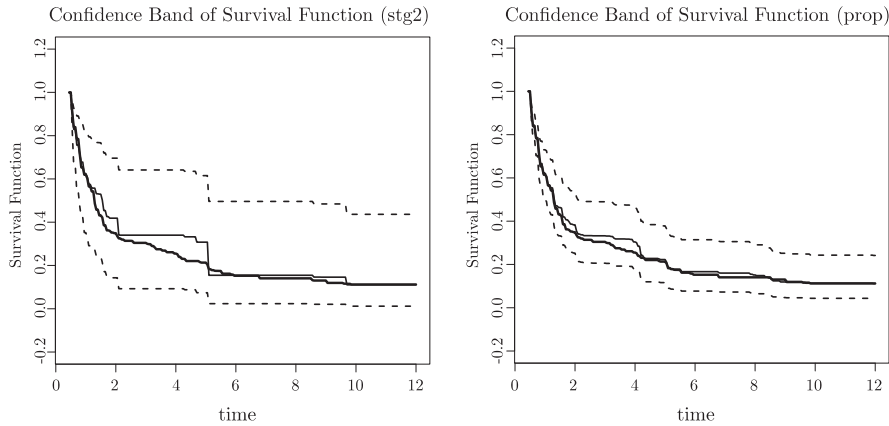


Figure 4. 95% simultaneous confidence bands (dash line) for the conditional survival function (solid line) for a white male patient who is diagnosed as grade IV, at distant stage, younger than 60, and has both surgery and radiation therapy. The thick solid line is obtained from the complete data.

control variate theory, σ_m , σ_t , and $\rho_{mt} = \text{corr}(m, t)$ can be estimated across the Monte Carlo replicates if they are known. The estimator m is improved by incorporating the control variate t . Since $E(t) = \tau$, the improved estimator m^* is unbiased and has the same expectation as m . In our method, since β_1 and $A_1(t)$ (τ) are unknown, we use the first stage estimates $\hat{\beta}_1$ and $\hat{A}_1(t)$ to replace their expected values. On the other hand, $\hat{\beta}_1$ and $\hat{A}_1(t)$ are more efficient than $\hat{\beta}_1^V$ and $\hat{A}_1^V(t)$ due to larger sample sizes, and therefore can serve as good estimates for β_1 and $A_1(t)$. Another difference between our method and the control variate method is that we use $\Sigma_{21} = \text{Cov}(\hat{\theta}_2^V, \hat{\theta}_1^V)$ instead of $\text{Cov}(\hat{\theta}_2^V, \hat{\theta}_1^V - \hat{\theta}_1)$, since the subjects outside the validation set are independent of the subjects in the validation set.

It is worth noting that our method also works for a working model other than the MS model. For any working model, one can define an estimator of the same form as our proposed estimator and obtain its asymptotic properties following similar steps, provided that the parameter estimator has a similar i.i.d. representation. An interesting question is then how to select the best possible working model, if exists, for the first stage data in order to maximize the benefit of combining information from the two stages. We do not have a definite answer to this question by far. The issue can be very complicated since the answer may depend on many factors, such as the form of the working model, the parameters to be estimated, the surrogate variables involved, and the criterion for optimality. Future research is warranted. However, as implied by the expression of the asymptotic variance of the proposed estimator, for a given second stage model the first stage working model should be chosen in a way such that the first

and second stage estimators have high correlation coefficients. In practice, it is convenient to use the same type of models in the two stages, a strategy that would work well when the first stage data are similar (or highly correlated) to the second stage data. Similarly, we suggest assigning the same type of coefficients to a variable in the two stages. For instance, age and gender have constant coefficients in both stages in our data example.

We only consider a simple design where the second stage is a simple random sample from the first stage sample. In many studies, the second-stage subjects are chosen with different selection probabilities depending on the outcomes of the first stage, the covariates, or both. The unequal-selection-probability sampling scheme is of special importance in medical and epidemiological studies, especially for rare diseases. In a sequel, we will extend our methods to incorporate biased-sampling problems.

Acknowledgement

Gang Li's research was supported in part by the U.S. National Institute of Health grants CA016042 and P01AT003960. The authors thank an associate editor and the referees for their helpful comments. Tong Tong Wu's research is partly supported by NSF CCF-0926194.

Appendix

A.1. Regularity assumptions

Let $g \in \{1, 2\}$ be the stage, and let

$$\begin{aligned}\tilde{U}_g^V(t) &= \frac{1}{n} X_g^V(t)' W_g^V(t) X_g^V(t), \\ \tilde{V}_g^V(t) &= \frac{1}{n} X_g^V(t)' W_g^V(t) Z_g^V(t), \\ \tilde{R}_g^V(t) &= \frac{1}{n} Z_g^V(t)' W_g^V(t) Z_g^V(t), \\ \tilde{U}_1(t) &= \frac{1}{N} X_1(t)' W_1(t) X_1(t), \\ \tilde{V}_1(t) &= \frac{1}{N} X_1(t)' W_1(t) Z_1(t), \\ \tilde{R}_1(t) &= \frac{1}{N} Z_1(t)' W_1(t) Z_1(t).\end{aligned}$$

The following assumptions are made for the theoretical development.

- (C1) **Finite intervals.** Let $\tau = \sup\{t : S_1(t|x, z)S_2(t|x, z)S_C(t|x, z) > 0 \text{ for all } x, z\}$ be a finite constant, where $S_g(t|x, z)$ is the survival function of stage

g , and $S_C(t|x, z)$ is the survival function for censoring time. The covariates x_{gi} and z_{gi} are restricted to a bounded set.

- (C2) **Limiting bounds.** The hazard functions $h_g(t|x, z)$, $g = 1, 2$, are bounded uniformly below and above in t, x , and z by some constants b and B , respectively.
- (C3) **Asymptotic limits.** For the subjects in $V(n)$, $\tilde{U}_g^V(t)$, $\tilde{V}_g^V(t)$, and $\tilde{R}_g^V(t)$ converge in L_∞ norm (defined as $\|M\|_\infty = \max_i \sum_{j=1}^n |m_{ij}|$ for a matrix M of size $m \times n$) uniformly in time $t \in [0, \tau]$, in probability, to some deterministic functions $U_g(t)$, $V_g(t)$, and $R_g(t)$, respectively. The functions $\tilde{U}_1(t)$, $\tilde{V}_1(t)$, and $\tilde{R}_1(t)$ have limiting functions $U_1(t)$, $V_1(t)$, and $R_1(t)$, respectively. These functions are uniformly continuous and bounded absolutely by the constant matrices K_U, K_V , and K_R , respectively, in the interval $[0, \tau]$. All the matrices are of full rank.

A.2. Proof of Theorem 1

To simplify the notation, we omit the superscript V for the validation set. Let $\hat{u}_{1i}(t) = Y_{1i}(t)[z_{1i} - \hat{V}'_1(t)\hat{U}_1^{-1}(t)x_{1i}]/\hat{h}_{1i}(t)$, $\hat{U}_1(t) = (1/n) \sum_{i=1}^n Y_{1i}(t)x_{1i}x'_{1i}/\hat{h}_{1i}(t)$, $\hat{\Phi}_1 = (1/n) \int_0^\tau Z'_1 \hat{H}_1 Z_1 dt$, $\hat{V}_1(t) = (1/n) \sum_{i=1}^n Y_{1i}(t)x_{1i}z_{1i}/\hat{h}_{1i}(t)$, and $\tilde{F}_n^x(t) = (1/n) \int_0^t X'_1 \hat{W}_1 dN_{1i}(s) = (1/n) \sum_{i=1}^n \int_0^t Y_{1i}(s)(x_{1i}/\hat{h}_{1i}(s))dN_{1i}(s)$. Let $\beta_1 = \Phi_1^{-1} E \left(\int_0^\tau u_{1i}(t)dN_{1i}(t) \right)$, where $\Phi_1 = \int_0^\tau [R_1 - V_1 U_1^{-1} V_1] dt$. We have

$$\begin{aligned} \sqrt{n}(\hat{\beta}_1 - \beta_1) &= \hat{\Phi}_1^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \int \hat{u}_{1i}(t)dN_{1i}(t) \right) - \sqrt{n}\Phi_1^{-1} E \left(\int u_{1i}(t)dN_{1i}(t) \right) \\ &= \Phi_1^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\int u_{1i}(t)dN_{1i}(t) - E \left(\int u_{1i}(t)dN_{1i}(t) \right) \right] \right\} \\ &\quad + \Phi_1^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \int [\hat{u}_{1i}(t) - u_{1i}(t)]dN_{1i}(t) \right) \\ &\quad + \left(\hat{\Phi}_1^{-1} - \Phi_1^{-1} \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \int u_{1i}(t)dN_{1i}(t) \right) \\ &\quad + \left(\hat{\Phi}_1^{-1} - \Phi_1^{-1} \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \int [\hat{u}_{1i}(t) - u_{1i}(t)]dN_{1i}(t) \right) \\ &\equiv I_1 + I_2 + I_3 + I_4. \end{aligned}$$

It can be shown that

$$I_2 = -\Phi_1^{-1} \left(\sqrt{n} \int [\hat{V}'_1(t) - V'_1(t)] U_1^{-1}(t) d\tilde{F}^x(t) \right)$$

$$+\Phi_1^{-1}\left(\sqrt{n}\int V_1'(t)U_1^{-2}(t)\left[\hat{U}_1(t)-U_1(t)\right]d\tilde{F}^x(t)\right)+o_p(1),$$

and $I_3 = o_p(1)$ and $I_4 = o_p(1)$ in L_∞ norm in probability; see Wu (2006) for more details. Thus

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = \Phi_1^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n w_{1i}\right) + o_p(1), \tag{A.1}$$

where

$$\begin{aligned} w_{1i} = & \int u_{1i}(t)dN_{1i}(t) - \mathbb{E}\left(\int u_{1i}(t)dN_{1i}(t)\right) \\ & - \int \left[Y_{1i}(t)\frac{z_{1i}x'_{1i}}{h_{1i}(t)} - V_1'(t)\right]U_1^{-1}(t)d\tilde{F}^x(t) \\ & + \int V_1'(t)U_1^{-2}(t)\left[Y_{1i}(t)\frac{x_{1i}x'_{1i}}{h_{1i}(t)} - U_1(t)\right]d\tilde{F}^x(t). \end{aligned}$$

It can be shown that $\mathbb{E}[w_{1i}] = 0$ and $\|w_{1i}\|_\infty \leq K$ for some constant K . By (A.1) and the Multivariate Central Limit Theorem, we have $\sqrt{n}(\hat{\beta}_1 - \beta_1) \rightarrow_d \mathcal{N}(0, \Sigma_{\beta,11})$, where $\Sigma_{\beta,11} = \Phi_1^{-1}\mathbb{E}(w_{1i}^{\otimes 2})\Phi_1^{-1}$.

Define

$$\hat{\Sigma}_{\beta,11} = \hat{\Phi}_1^{-1}\left(\frac{1}{n}\sum_{i=1}^n \hat{w}_{1i}^{\otimes 2}\right)\hat{\Phi}_1^{-1},$$

where

$$\begin{aligned} \hat{w}_{1i} = & \int \hat{u}_{1i}(t)dN_{1i}(t) - \left(\frac{1}{n}\sum_{i=1}^n \int \hat{u}_{1i}(t)dN_{1i}(t)\right) \\ & - \int \left[Y_{1i}(t)z_{1i}x'_{1i} - \hat{V}_1'(t)\right]\hat{U}_1^{-1}(t)d\tilde{F}_n^x(t) \\ & + \int \hat{V}_1'(t)\hat{U}_1^{-2}(t)\left[Y_{1i}(t)x_{1i}x'_{1i} - \hat{U}_1(t)\right]d\tilde{F}_n^x(t). \end{aligned} \tag{A.2}$$

The consistency of the variance estimator can be established by the consistency of $\hat{\Phi}_1$, $\hat{U}_1(t)$, $\hat{V}_1(t)$, $\tilde{F}_n^x(t)$, and $\hat{\beta}_1$.

Similarly, it can be shown that

$$\begin{aligned} \sqrt{n}[\hat{A}_1(t) - A_1(t)] = & \sqrt{n}\int_0^t \hat{U}_1^{-1}(s)[d\tilde{F}_n^x(s) - \hat{V}_1(s)\hat{\beta}_1 ds] \\ & - \sqrt{n}\int_0^t U_1^{-1}(s)[d\tilde{F}^x(s) - V_1(s)\beta_1 ds] \end{aligned}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n v_{1i}(t) + o_p(1). \tag{A.3}$$

It follows from (A.3) and the Multivariate Central Limit Theorem that the finite dimensional distributions of $\sqrt{n}(\hat{A}_1(t) - A_1(t))$ converge to those of a multivariate normal distribution with covariance $\kappa_1(t_1, t_2) = E\{v_{1i}(t_1) \cdot v_{1i}(t_2)'\}$. The ‘tightness’ can be checked using Theorem 13.5 in Billingsley (1999). Thus, $\sqrt{n}[\hat{A}_1(\cdot) - A_1(\cdot)]$ converges to a zero-mean Gaussian process.

The variance $\Sigma_{A,11}(t) = \kappa_1(t, t)$ can be estimated by

$$\hat{\Sigma}_{A,11}(t) = \frac{1}{n} \sum_{i=1}^n \left\{ \hat{v}_{1i}(t) \right\}^{\otimes 2},$$

where $\hat{v}_{1i}(t)$ is obtained by replacing $U_1(t), V_1(t), h_{1i}(t), \tilde{F}^x(s), \beta_1,$ and w_{1i} in (2.2) by their respective sample estimates $\hat{U}_1(t), \hat{V}_1(t), \hat{h}_{1i}(t), \tilde{F}_n^x(s), \hat{\beta}_1,$ and \hat{w}_{1i} . The consistency of the variance estimator follows immediately from the consistency of $\hat{U}_1(t), \hat{V}_1(t), \hat{w}_{1i}, \hat{\beta}_1,$ and $\hat{A}_1(t)$.

A.3. Proof of Lemma 1

Similar to (A.1), it can be shown that

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_1^V - \beta_1 \\ \hat{\beta}_2^V - \beta_2 \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} \Phi_1^{-1} w_{1i} \\ \Phi_2^{-1} w_{2i} \end{pmatrix} + o_p(1).$$

This, together with the Multivariate Central Limit Theorem, proves (2.3). The uniform convergence of the variance and covariance estimators can be easily verified using Theorem 1 of Rao (1963).

A.4. Proof of Lemma 2

Similar to the proof of Lemma 1, we can prove that the joint distributions of $\sqrt{n}[\hat{A}_1^V(t_1) - A_1(t_1), \hat{A}_2^V(t_2) - A_2(t_2)]$ converge to those of a zero-mean multivariate normal distribution. It can be verified that tightness holds. Therefore we have $\sqrt{n}[\hat{A}_1^V(t_1) - A_1(t_1), \hat{A}_2^V(t_2) - A_2(t_2)]$ converges to a zero-mean Gaussian random field $[\mathcal{W}_1^V(t_1), \mathcal{W}_2^V(t_2)]$ with the variance-covariance function

$$\begin{pmatrix} \Sigma_{A,11}(t_1) & \Sigma_{A,12}(t_1, t_2) \\ \Sigma_{A,21}(t_1, t_2) & \Sigma_{A,22}(t_2) \end{pmatrix}.$$

The covariance function $\hat{\Sigma}_{A,kl}(t_1, t_2)$ can be consistently estimated by

$$\hat{\Sigma}_{A,kl}(t_1, t_2) = \frac{1}{n} \sum_{i \in V(n)} \{ \hat{v}_{ki}(t_1) \cdot \hat{v}_{li}(t_2)' \}. \tag{A.4}$$

The uniform consistency of $\hat{\Sigma}_{A,kl}(t_1, t_2)$ can be verified using Theorem 1 of Rao (1963).

A.5. Proof of Theorem 2

As in Lemma 1, we can show that

$$\begin{aligned} \sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_1^V - \beta_1 \\ \hat{\beta}_2^V - \beta_2 \end{pmatrix} &= \begin{pmatrix} \frac{\sqrt{n}}{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^N \Phi_1^{-1} w_{1i} \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_1^{-1} w_{1i} \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_2^{-1} w_{2i} \end{pmatrix} + o_p(1) \\ &\longrightarrow_d \mathcal{N} \left(0, \begin{pmatrix} \rho \Sigma_{\beta,11} & \sqrt{\rho} \Sigma_{\beta,11} & \sqrt{\rho} \Sigma_{\beta,12} \\ \sqrt{\rho} \Sigma_{\beta,11} & \Sigma_{\beta,11} & \Sigma_{\beta,12} \\ \sqrt{\rho} \Sigma_{\beta,21} & \Sigma_{\beta,21} & \Sigma_{\beta,22} \end{pmatrix} \right), \end{aligned}$$

where $\rho = \lim n/N$ as $n, N \rightarrow \infty$. The proposed estimator then can be written as

$$\begin{aligned} &\sqrt{n}(\hat{\beta}_2 - \beta_2) \\ &= \sqrt{n}(\hat{\beta}_2^V - \beta_2) - \Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \left[\sqrt{n}(\hat{\beta}_1^V - \beta_1) - \sqrt{n}(\hat{\beta}_1 - \beta_1) \right] + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_2^{-1} w_{2i} - \Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_1^{-1} w_{1i} - \frac{\sqrt{n}}{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^N \Phi_1^{-1} w_{1i} \right) + o_p(1). \end{aligned}$$

We have

$$\begin{aligned} \sqrt{n} \begin{pmatrix} \hat{A}_1(t_1) - A_1(t_1) \\ \hat{A}_1^V(t_2) - A_1(t_2) \\ \hat{A}_2^V(t_3) - A_2(t_3) \end{pmatrix} &= \begin{pmatrix} \frac{\sqrt{n}}{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^N v_{1i}(t_1) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n v_{1i}(t_2) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n v_{2i}(t_3) \end{pmatrix} + o_p(1) \\ &\longrightarrow_d \begin{pmatrix} \mathcal{W}_1(t_1) \\ \mathcal{W}_1^V(t_2) \\ \mathcal{W}_2^V(t_3) \end{pmatrix} \equiv \mathcal{W}(t), \end{aligned}$$

with variance-covariance function among $\mathcal{W}_1(t_1), \mathcal{W}_1^V(t_2), \mathcal{W}_2^V(t_3)$ being

$$\begin{pmatrix} \rho \Sigma_{A,11}(t_1) & \sqrt{\rho} \Sigma_{A,11}(t_1 \wedge t_2) & \sqrt{\rho} \Sigma_{A,12}(t_1, t_3) \\ \sqrt{\rho} \Sigma_{A,11}(t_1 \wedge t_2) & \Sigma_{A,11}(t_2) & \Sigma_{A,12}(t_2, t_3) \\ \sqrt{\rho} \Sigma_{A,21}(t_1, t_3) & \Sigma_{A,21}(t_2, t_3) & \Sigma_{A,22}(t_3) \end{pmatrix}.$$

We can write

$$\begin{aligned} & \sqrt{n}[\hat{A}_2(t) - A_2(t)] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n v_{2i}(t) - \Sigma_{A,21}(t)\Sigma_{A,11}^{-1}(t) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n v_{1i}(t) - \frac{\sqrt{n}}{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^N v_{1i}(t) \right) + o_p(1). \end{aligned}$$

The joint distribution of the proposed estimators $\hat{\beta}_2$ and $\hat{A}_2(t)$ is

$$\begin{aligned} \sqrt{n} \begin{pmatrix} \hat{\beta}_2 - \beta_2 \\ \hat{A}_2(t) - A_2(t) \end{pmatrix} &= \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_2^{-1} w_{2i} - \Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_1^{-1} w_{1i} \right] \\ - \frac{\sqrt{n}}{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^N \Phi_1^{-1} w_{1i} \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n v_{2i}(t) - \Sigma_{A,21}(t)\Sigma_{A,11}^{-1}(t) \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n v_{1i}(t) \right] \\ - \frac{\sqrt{n}}{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^N v_{1i}(t) \end{pmatrix} \\ &+ o_p(1) \\ &\xrightarrow{d} \begin{pmatrix} Z_2 \\ \mathcal{W}_2(t) \end{pmatrix}, \end{aligned}$$

where $Z_2 \sim \mathcal{N}(0, \Sigma_{\beta_2}^*)$, the variance of $\sqrt{n}(\hat{\beta}_2 - \beta_2)$ can be derived easily by Slutsky's Theorem and the delta-method:

$$\begin{aligned} \text{Var}(\sqrt{n}(\hat{\beta}_2 - \beta_2)) &\rightarrow_p \Sigma_{\beta,22} + \Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \left(\Sigma_{\beta,11} + \rho \Sigma_{\beta,11} - 2\sqrt{\rho} \Sigma_{\beta,11} \right) \Sigma_{\beta,11}^{-1} \Sigma_{\beta,12} \\ &\quad - 2\Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \left(\Sigma_{\beta,12} - \sqrt{\rho} \Sigma_{\beta,12} \right) \\ &= \Sigma_{\beta,22} - (1 - \rho) \Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \Sigma_{\beta,12}, \end{aligned}$$

and $\mathcal{W}_2(t)$ is a zero-mean Gaussian process with covariance function given by

$$\begin{aligned} \zeta(s, t) &= \Sigma_{A,22}(s, t) - \left[\Sigma_{A,21}(s, t) - \sqrt{\rho} \Sigma_{A,21}(s, t) \right] \Sigma_{A,11}^{-1}(t) \Sigma_{A,12}(t) \\ &\quad - \Sigma_{A,21}(s) \Sigma_{A,11}^{-1}(s) \left[\Sigma_{A,12}(s, t) - \sqrt{\rho} \Sigma_{A,12}(s, t) \right] \\ &\quad + \Sigma_{A,21}(s) \Sigma_{A,11}^{-1}(s) \left[\Sigma_{A,11}(s, t) - 2\sqrt{\rho} \Sigma_{A,11}(s \wedge t) \right. \\ &\quad \left. + \rho \Sigma_{A,11}(s, t) \right] \Sigma_{A,11}^{-1}(t) \Sigma_{A,12}(t). \end{aligned}$$

Covariance of the two-stage estimators $\hat{\beta}_2$ and $\hat{A}_2(t)$ converges to

$$\begin{aligned} \Sigma_{\beta_2 A_2}^*(t) &= \text{E} \left\{ \left[\sqrt{n}(\hat{\beta}_2^V - \beta_2) - \hat{\Sigma}_{\beta,21} \hat{\Sigma}_{\beta,11}^{-1} \sqrt{n}(\hat{\beta}_1^V - \hat{\beta}_1) \right] \right. \\ &\quad \left. \cdot \left[\sqrt{n}(\hat{A}_2^V(t) - A_2(t)) - \hat{\Sigma}_{A,21}(t) \hat{\Sigma}_{A,11}^{-1}(t) \sqrt{n}(\hat{A}_1^V(t) - \hat{A}_1(t)) \right]' \right\} \end{aligned}$$

$$\begin{aligned} &\rightarrow \Sigma_{\beta_2 A_2} - (1 - \sqrt{\rho})\Sigma_{\beta_2 A_1} \Sigma_{A,11}^{-1} \Sigma_{A,12} \\ &\quad - (1 - \sqrt{\rho})\Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \Sigma_{\beta_1 A_2} \\ &\quad + (1 - 2\sqrt{\rho} + \rho)\Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \Sigma_{\beta_1 A_1} \Sigma_{A,11}^{-1} \Sigma_{A,12}. \end{aligned}$$

The variance and covariance matrices can be estimated by replacing each term by their respective sample estimates. The consistency can be established by Slutsky’s Theorem.

A.6. Proof of Theorem 4

To show that $E_n(t)$ and $\hat{E}_n(t)$ have the same limiting distribution $\mathcal{V}_2(t)$, we define an intermediate processes $E_n^*(t)$ as

$$\begin{aligned} E_n^*(t) = & z_0' \cdot \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_2^{-1} w_{2i} G_i \right. \\ & \left. - \Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_1^{-1} w_{1i} G_i - \frac{1}{\sqrt{N}} \sqrt{\frac{n}{N}} \sum_{i=1}^N \Phi_1^{-1} w_{1i} G_i \right) \right\} \cdot t \\ & + x_0' \cdot \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n v_{2i}(t) G_i \right. \\ & \left. - \Sigma_{A,21}(t) \Sigma_{A,11}^{-1}(t) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n v_{1i}(t) G_i - \frac{1}{\sqrt{N}} \sqrt{\frac{n}{N}} \sum_{i=1}^N v_{1i}(t) G_i \right) \right\}. \end{aligned}$$

It follows that, conditionally on the data, $E_n^*(t)$ converges weakly in probability to $\mathcal{V}_2(t)$, which is the limiting Gaussian distribution of $E_n(t)$, according to Theorem 2.9.6 in van der Vaart and Wellner (1996). To complete the proof, we need to show that $\|\hat{E}_n(t) - E_n^*(t)\| \rightarrow_d 0$. We write

$$\begin{aligned} &\hat{E}_n(t) - E_n^*(t) \\ &= z_0' \cdot \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left((\hat{\Phi}_2^V)^{-1} \hat{w}_{2i} - \Phi_2^{-1} w_{2i} \right) G_i \right. \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\hat{\Sigma}_{\beta,21} \hat{\Sigma}_{\beta,11}^{-1} (\hat{\Phi}_1^V)^{-1} \hat{w}_{1i} - \Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \Phi_1^{-1} w_{1i} \right) G_i \\ &\quad \left. + \frac{1}{\sqrt{N}} \sqrt{\frac{n}{N}} \sum_{i=1}^N \left(\hat{\Sigma}_{\beta,21} \hat{\Sigma}_{\beta,11}^{-1} \hat{\Phi}_1^{-1} \hat{w}_{1i} - \Sigma_{\beta,21} \Sigma_{\beta,11}^{-1} \Phi_1^{-1} w_{1i} \right) G_i \right\} \cdot t \\ &\quad + x_0' \cdot \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\hat{v}_{2i}^V(t) - v_{2i}(t) \right) G_i \right. \end{aligned}$$

$$\begin{aligned}
& -\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\hat{\Sigma}_{A,21}(t) \hat{\Sigma}_{A,11}^{-1}(t) \hat{v}_{1i}^V(t) - \Sigma_{A,21}(t) \Sigma_{A,11}^{-1}(t) v_{1i}(t) \right) G_i \\
& -\frac{1}{\sqrt{N}} \sqrt{\frac{n}{N}} \sum_{i=1}^N \left(\hat{\Sigma}_{A,21}(t) \hat{\Sigma}_{A,11}^{-1}(t) \hat{v}_{1i}(t) - \Sigma_{A,21}(t) \Sigma_{A,11}^{-1}(t) v_{1i}(t) \right) G_i \Big\} \\
& \equiv E_{n,1}^d(t) + E_{n,2}^d(t) + E_{n,3}^d(t) + E_{n,4}^d(t) + E_{n,5}^d(t) + E_{n,6}^d(t).
\end{aligned}$$

Then $\|\hat{E}_n(t) - \hat{E}_n^*(t)\| \leq \|E_{n,1}^d(t)\| + \|E_{n,2}^d(t)\| + \|E_{n,3}^d(t)\| + \|E_{n,4}^d(t)\| + \|E_{n,5}^d(t)\| + \|E_{n,6}^d(t)\|$. We first check that

$$\begin{aligned}
\|E_{n,1}^d(t)\| &= \left\| z'_0 \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left((\hat{\Phi}_2^V)^{-1} \hat{w}_{2i} - \Phi_2^{-1} w_{2i} \right) \right\} G_i \cdot t \right\| \\
&\leq \left\| z'_0 \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left((\hat{\Phi}_2^V)^{-1} - \Phi_2^{-1} \right) \hat{w}_{2i} \right\} G_i \cdot t \right\| \\
&\quad + \left\| z'_0 \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_2^{-1} (\hat{w}_{2i} - w_{2i}) \right\} G_i \cdot t \right\|,
\end{aligned}$$

which converges to zero in probability uniformly in t . It can be similarly shown that the other terms converge to zero in probability uniformly in t .

References

- Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701-726.
- Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York.
- Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley, New York.
- Breslow, N. E. and Holubkov, R. (1997). Weighted likelihood, pseudo-likelihood and maximum likelihood methods for logistic regression analysis of two-stage data. *Statist. Medicine* **16**, 103-116.
- Chen, Y.-H. (2002). Cox regression in cohort studies with validation sampling. *J. Roy. Statist. Soc. Ser. B* **64**, 51-62.
- Chen, Y.-H. and Chen, H. (2000). A unified approach to regression analysis under double sampling design. *J. Roy. Statist. Soc. Ser. B* **64**, 449-460.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.
- Gasser, T. and Muller, H. (1979). *Kernel Estimation of Regression Functions, Smoothing Techniques for Curve Estimation. Lecture Notes in Mathematics* 757. Springer-Verlag, Berlin.
- Jiang, J. and Zhou, H. (2007). Additive hazard regression with auxiliary covariates. *Biometrika* **94**, 359-369.
- Kulich, M. and Lin, D. Y. (2000). Additive hazards regression with covariate measurement error. *J. Amer. Statist. Assoc.* **95**, 238-248.

- Li, G. and Tseng, C. (2008). Non-parametric estimation of a survival function with two-stage design studies. *Scand. J. Statist.* **35**, 193-211.
- Lin, D. Y. and Ying, Z. L. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.
- Martinussen, T. and Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data*. Springer, New York.
- McKeague, I. and Sasieni, P. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501-514.
- Rao, R. (1963). The law of large numbers for $d[0,1]$ -valued random variables. *Theory Probab. Appl.* **8**, 70-74.
- Scheike, T. H. (2002). The additive nonparametric and semiparametric aalen model as the rate function for a counting process. *Lifetime Data Analysis* **8**, 247-262.
- Schill, W., Jockel, K.-H., Drescher, K. and Timm, J. (1993). Logistic analysis in case-control studies under validation sampling. *Biometrika* **80**, 339-352.
- Tseng, C. H. (2004). Survival analysis with two stage design studies. Ph. D. thesis, University of California, Los Angeles.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Verlag, New York.
- Wang, C.-Y., Hsu, L., Feng, Z. D. and Prentice, R. L. (1997). Regression calibration in failure time regression. *Biometrics* **53**, 131-145.
- White, J. E. (1982). A two stage design for the study of the relationship between a rare exposure and a rare disease. *Amer. J. Epidemiology* **115**, 119-128.
- Wu, T. T. (2006). A Partial Linear Semiparametric Additive Risks Model for Two-Stage Design Survival Studies. Ph. D. thesis, University of California, Los Angeles.
- Zhou, H. and Pepe, M. (1995). Auxiliary covariate data in failure time regression. *Biometrika* **82**, 139-149.

Department of Biostatistics, University of California, Los Angeles, CA 90095, USA.

E-mail: vli@ucla.edu

Departments of Epidemiology and Biostatistics, University of Maryland, College Park, MD 20742, USA.

E-mail: ttwu@umd.edu

(Received February 2009; accepted June 2009)