 DOI: 10.1111/biom.12119

Semiparametric Analysis of Linear Transformation Models with Covariate Measurement Errors

Samiran Sinha,* Yanyuan Ma**

Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, U.S.A.

**email:* sinha@stat.tamu.edu

***email:* ma@stat.tamu.edu

Summary. We take a semiparametric approach in fitting a linear transformation model to a right censored data when predictive variables are subject to measurement errors. We construct consistent estimating equations when repeated measurements of a surrogate of the unobserved true predictor are available. The proposed approach applies under minimal assumptions on the distributions of the true covariate or the measurement errors. We derive the asymptotic properties of the estimator and illustrate the characteristics of the estimator in finite sample performance via simulation studies. We apply the method to analyze an AIDS clinical trial data set that motivated the work.

Key words: Counting process; Estimating equation; Induced hazard; Kernel density; Non-differential measurement errors; U-statistics.

1. Introduction

Errors in covariates are common in clinical studies, and standard data analysis tools ignoring the fact that the covariates are measured with errors may result in biased and inconsistent parameter estimation. Traditionally two approaches exist in handling errors in covariates: the functional approach where unobserved true covariates are assumed to be unknown parameters, and structural approach where unobserved true covariates are treated as a random variable with a probability distribution. The modern perspective further makes a difference between whether or not a method imposes a parametric model for the distribution of the unobserved covariates, and considers the parametric modeling as structural approach, while classifies the nonparametric modeling as a kind of functional approach. From this perspective, our work belongs to the modern structural approach framework.

Errors in covariates in the Cox proportional hazard (CPH) model has been studied extensively by Prentice (1982), Nakamura (1992), Hu, Tsiatis, and Davidian (1998), Huang and Wang (2000), Hu and Lin (2004) and Zucker (2005). In the extended cure rate model context, Ma and Yin (2008) proposed a corrected score estimator to handle errors in covariates. However, relatively less attention has been given for developing general methodology for handling errors in covariates in the proportional odds (PO) model and in the linear transformation model which contains the CPH and PO models as two special cases. Some recent works in the area of linear transformation models include Fine, Ying, and Wei (1998), Gao and Tsiatis (2005), Lu and Zhang (2010). Wen and Chen (2012) developed a conditional score approach for handling errors in covariates in the context of current status data using the PO model. Cheng and Wang (2001) considered measurement errors in covariates in analyzing right censored data using the linear transformation model. In their set up, they assumed parametric models for both the paired differences of the measurement errors and the paired differences of the unobserved true covariates.

In this article we develop a semiparametric method for analyzing right censored failure time data using the linear transformation model while a covariate is measured with error. We consider the scenario where replicated measurements of a surrogate variable are available. We build our estimator through forming an estimating equation which inherits the structure of the estimating equation in Chen, Jin, and Ying (2002), where estimation and inference of linear transformation model without covariate measurement errors is considered. The nice feature of such estimating equation approach is that there exists a closed form expression for the standard errors for the finite dimensional regression parameter estimator. An additional advantage of our approach is that we can accommodate any error distributions. Furthermore, unlike Cheng and Wang (2001), our approach is applicable as long as the time-to-failure and the censoring time are independent conditional on the covariates.

This work is motivated by an AIDs clinical trial data Hammer et al. (1996). The goal of this trial was to compare the efficacy of several therapies on the HIV-infected adults. One of the critical parameter for HIV-infection is the CD4 cell counts, the building blocks of body's immune system. In this trial, CD4 cell counts were first measured for screening purpose, then repeated measurements were taken at the baseline, and finally they were measured on a regular basis after the treatment began. Since the true CD4 counts is impossible to obtain, we treated the two baseline measurements as the erroneous values of the actual CD4 counts. Our interest is in finding the effect of baseline CD4 cell counts and the four therapies on the time to death or AIDS using a linear transformation model.

The rest of the article is organized as the following. In Section 2, we describe the model and its background information.

We then explain how to estimate the distribution of the unobservable covariate conditional on the observable ones and the surrogates in Section 3. The main estimation methodology is given in Section 4, and we derive the corresponding large sample properties of the estimator in Section 5. The method and its finite sample performance are investigated through simulation studies in Section 6. We analyze the AIDs trial data using our method in Section 7 and conclude the article in Section 8. All the technical details are collected in an Appendix and in the supplementary materials.

2. Basic Model and Background

Suppose that the observed data are n iid copies of $(T^*, \Delta, \mathbf{Z}, \boldsymbol{W}^*)$, where $\mathbf{Z}$ is a $p \times 1$ vector of observed covariates, $T^* = \min(T, C)$, where T is the time-to-failure, C is the censoring time, and $\Delta = I(T \leq C)$. We assume that conditional on the complete covariates $(X, \mathbf{Z})$, where X is unobservable, T and C are independent. Here $\boldsymbol{W}^* = (W_1^*, \ldots, W_m^*)^{\mathrm{T}}$ is a surrogate measurement for the scalar covariate X, which implies that conditional on X, $\boldsymbol{W}^*$ is independent of T and $\mathbf{Z}$. Furthermore, under the additive measurement errors

$$W_{ij}^* = X_i + U_{ij}^*,\ i = 1, \ldots, n,\ j = 1, \ldots, m,$$

where the measurement errors U_{ij}^* are assumed to be iid copies of a random variable U^* which follows a symmetric distribution centered at 0. We assume that U^* is independent of $(T, X, \mathbf{Z}, C)$.

The linear transformation model is

$$H(T) = -\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{Z} - \beta_2 X + e, \tag{1}$$

where H is an unknown monotone transformation function, e is a random variable with a known distribution and is independent of $\mathbf{Z}$ and X, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\mathrm{T}}, \beta_2)^{\mathrm{T}}$ is an unknown regression parameter of interest. The proportional hazards model and the proportional odds model are two special cases of (1) with e following the extreme-value distribution and the standard logistic distribution, respectively. Let $\lambda(\bullet)$ and $\Lambda(\bullet)$ be the corresponding hazard and cumulative hazard functions of e, respectively. In the following paragraph we describe the estimation procedure given in Chen et al. (2002) when X is observed without any measurement errors.

Define $Y_i(t) = I(T_i^* \geq t)$, $N_i(t) = I(T_i^* \leq t, \Delta_i = 1)$ and $M^*(t) = N(t) - \int_0^t Y(u)\mathrm{d}\Lambda\{\boldsymbol{\beta}_{10}^{\mathrm{T}}\mathbf{Z} + \beta_{20}X + H_0(u)\}$, where $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{10}^{\mathrm{T}}, \beta_{20})^{\mathrm{T}}$ and $H_0(\bullet)$ are the true values of $\boldsymbol{\beta}$ and H, respectively. Note that $M^*(t)$ is a martingale process with respect to filtration $\sigma\{Y(u), N(u), \mathbf{Z}, X, 0 \leq u < t\}$. The nonparametric estimate of H will be derived at the observed failure times. Define $\mathcal{H}$ as the collection of non-decreasing step functions defined on $[0, \infty)$. Also, for any $H \in \mathcal{H}$, set $H(0) = -\infty$. Chen et al. (2002) recommended to estimate $\boldsymbol{\beta}$ and H by solving the following estimating equations:

$$U_\beta(\beta, H) = \sum_{i=1}^n \int_0^\infty \begin{pmatrix} \mathbf{Z}_i \\ X_i \end{pmatrix} [\mathrm{d}N_i(u) - Y_i(u)\mathrm{d}\Lambda\{\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{Z}_i + \beta_2 X_i + H(u)\}] = 0,$$

$$U_H(\beta, H) = \sum_{i=1}^n [\mathrm{d}N_i(u) - Y_i(u)\mathrm{d}\Lambda\{\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{Z}_i + \beta_2 X_i + H(u)\}] = 0,\ \text{for all } u \geq 0.$$

Clearly the $\hat{H}$ that solves $U_H(\beta, H) = 0$ belongs to $\mathcal{H}$.

3. Estimation of $f_{X|W,\mathbf{Z}}(x|w, z)$

To handle the measurement errors, we need to understand the relation between the unobservable true covariates and the observed ones. To this end, we first point out that under the independent symmetric error assumption, the measurement error distribution can be easily estimated. For example, we can form $W_i = \sum_{j=1}^m a_j W_{ij}^* = X_i + \sum_{j=1}^m a_j U_{ij}^*$ and $V_i = \sum_{j=1}^m b_j W_{ij}^* = \sum_{j=1}^m b_j U_{ij}^*$, where $\sum_{j=1}^m a_j = 1$, and $\sum_{j=1}^m b_j = 0$. Write $U_i = \sum_{j=1}^m a_j U_{ij}^* = W_i - X_i$. As long as U_i's and V_i's have the same distribution, we can use the observed V_i's to estimate $f_U(\cdot)$. In terms of minimizing the error variance, $\{a_1, \cdots, a_m\}$ should be chosen so that $\sum_{j=1}^m a_j U_{ij}^*$ has a minimum variance. Let $\eta = [m/2]$ be the largest integer smaller or equal to $m/2$, then the optimal choice of a_i's and b_i's yield

$$W_i = \sum_{j=1}^{\eta} \frac{W_{ij}^*}{2\eta} + \sum_{j=\eta+1}^{m} \frac{W_{ij}^*}{2m - 2\eta},\ V_i = \sum_{j=1}^{\eta} \frac{W_{ij}^*}{2\eta} - \sum_{j=\eta+1}^{m} \frac{W_{ij}^*}{2m - 2\eta}.$$

We can estimate $f_U(\cdot)$ using the above formed $V_1, \ldots, V_n$. For example, a nonparametric estimator $\widehat{f}_U(u) = (nh)^{-1}\sum_{i=1}^n K\left\{(V_i - u)/h\right\}$ is given in Hall and Ma (2007), where $K(\cdot)$ is a symmetric kernel function and $h > 0$ is a bandwidth.

The availability of W_i's and $\mathbf{Z}_i$'s easily allow us to obtain an assessment of $f_{W|\mathbf{Z}}(w \mid z)$. Because $f_U(u)$ can also be estimated, this allows us to recover the distribution of $f_{X|\mathbf{Z}}(x \mid z)$, or at least provides us sufficient information to propose a suitable model for $f_{X|\mathbf{Z}}(x \mid z)$. See also Li and Vuong (1998) for more in depth study on how to assess $f_{X|\mathbf{Z}}(x \mid z)$ from multiple observations. Thus, we use a parametric model $f_{X|\mathbf{Z}}(x \mid \mathbf{Z}, \boldsymbol{\theta})$ to denote the conditional density of X given $\mathbf{Z}$, where $\boldsymbol{\theta}$ is a finite dimensional parameter. Therefore we work in a partly structural model framework by assuming the functional form of $f_{X|\mathbf{Z}}$ to be known, while leaving the error distribution unspecified. Although the functional form of $f_{X|\mathbf{Z}}$ is assumed to be known, it does not have to belong to any specific family such as the normal distribution family. Observe that $f_{X|W,\mathbf{Z}}(x|w, z; \boldsymbol{\theta}, f_U) = f_{X|\mathbf{Z}}(x \mid z, \boldsymbol{\theta})f_U(w - x)/\int f_{X|\mathbf{Z}}(x \mid z, \boldsymbol{\theta})f_U(w - x)\mathrm{d}x$. Here and in the following text, we use $f_{X|W,\mathbf{Z}}(x \mid w, z; \boldsymbol{\theta}, f_U)$ to emphasize that the conditional probability density function of X on $W, \mathbf{Z}$ relies on f_U. In the case when $f_U(u)$ is estimated nonparametrically, we write $f_{X|W,\mathbf{Z}}(x \mid w, z; \boldsymbol{\theta}, \widehat{f}_U)$. Note that $\widehat{f}_U$ can be obtained based on V_i's directly. Subsequently, $\widehat{\boldsymbol{\theta}}$ can be obtained through maximizing

$$\sum_{i=1}^n \log\left\{\int f_{X|\mathbf{Z}}(x \mid z_i; \boldsymbol{\theta})\widehat{f}_U(w_i - x)\mathrm{d}x\right\}. \tag{2}$$

While allowing the error distribution f_U to be unspecified is very flexible, we retained the parametric assumption on $f_{X|\mathbf{Z}}$

in our model. Obviously, this implies certain restriction. For example, because X is unobservable, $f_{X|\mathbf{Z}}$ is harder to model than if X had been available. However, once f_U is estimated, $f_{X|\mathbf{Z}}$ is identifiable through a deconvolution step, hence its modeling is not completely unfounded. In addition, in our numerical experiments (See Section 6), we have found that slight or even moderate model mis-specification of $f_{X|\mathbf{Z}}$ does not cause severe estimation bias, while a parametric approach of $f_{X|\mathbf{Z}}$ brings dramatic simplification both theoretically and numerically. Hence we adopted this modeling approach. We would like to point out that Zucker (2005) assumed that $f_{X|\mathbf{Z},W}$ is known up to a finite dimensional parameter, which is a much easier model to handle than the one considered here.

4. Estimation of $\boldsymbol{\beta}$ With Estimated $f_{X|W,\mathbf{Z}}(x \mid w, z)$

Having obtained $\widehat{\boldsymbol{\theta}}$ and $\widehat{f}_U$, we can plug in the resulting $f_{X|W,\mathbf{Z}}(x \mid w, \mathbf{z}; \widehat{\boldsymbol{\theta}}, \widehat{f}_U)$ and carry out the estimation procedure described. Specifically, the induced model of T given $(W, \mathbf{Z})$ is

$$\mathrm{pr}(T \geq t|W, \mathbf{Z}; \widehat{\boldsymbol{\theta}}, \widehat{f}_U) = \int \exp[-\Lambda\{\boldsymbol{\beta}_1^T\mathbf{Z} + \beta_2 x + H(t)\}] f_{X|W,\mathbf{Z}}(x \mid W, \mathbf{Z}; \widehat{\boldsymbol{\theta}}, \widehat{f}_U)\mathrm{d}x.$$

Therefore, the induced cumulative hazard of T given W and $\mathbf{Z}$ is

$$\Lambda_T(t|W, \mathbf{Z}; H, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U) = -\log \int \exp[-\Lambda\{\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{Z} + \beta_2 X + H(t)\}] f_{X|W,\mathbf{Z}}(x \mid W, \mathbf{Z}; \widehat{\boldsymbol{\theta}}, \widehat{f}_U)\mathrm{d}x,$$

and the induced hazard for T given W and $\mathbf{Z}$ is $\lambda_T(t|W, \mathbf{Z}; H, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U) = J(t|W, \mathbf{Z}; H, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U)\dot{H}(t)$, where $\dot{H}(t) = \partial H(t)/\partial t$, and $J(t|W, \mathbf{Z}; H, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U) = \int \lambda\{\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{Z} + \beta_2 x + H(t)\}G(x|t, W, \mathbf{Z}; H, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U)\mathrm{d}x$ with $G(x|t, W, Z; H, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U) = \exp[-\Lambda\{\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{Z} + \beta_2 x + H(t)\}]\, f_{X|W,\mathbf{Z}}(x \mid W, \mathbf{Z}; \widehat{\boldsymbol{\theta}}, \widehat{f}_U) \,/\, \int \exp[-\Lambda\{\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{Z} + \beta_2 x + H(t)\}] f_{X|W,\mathbf{Z}}(x \mid W, \mathbf{Z}; \widehat{\boldsymbol{\theta}}, \widehat{f}_U)\mathrm{d}x$.

Let $\boldsymbol{\beta}_0$, H_0, $\boldsymbol{\theta}_0$ be the true values of $\boldsymbol{\beta}$, H and $\boldsymbol{\theta}$ respectively. Then, $M(t) = N(t) - \int_0^t Y(u)\lambda_T(u \mid W, \mathbf{Z}; H_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U)\mathrm{d}u$ is a martingale process with respect to filtration $\sigma\{Y(u), N(u), \mathbf{Z}, W, 0 \leq u < t\}$. Now define

$$U_\beta(\boldsymbol{\beta}, H, \widehat{\boldsymbol{\theta}}, \widehat{f}_U) = \sum_{i=1}^n \int_0^\infty \begin{pmatrix} \mathbf{Z}_i \\ W_i \end{pmatrix} [\mathrm{d}N_i(u) - Y_i(u)\lambda_T(u|W_i, \mathbf{Z}_i; H, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U)\mathrm{d}u] = 0, \quad (3)$$

$$U_H(\boldsymbol{\beta}, H, \widehat{\boldsymbol{\theta}}, \widehat{f}_U) = \sum_{i=1}^n [\mathrm{d}N_i(u) - Y_i(u)\lambda_T \times (u|W_i, Z_i; H, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U)\mathrm{d}u] = 0 \text{ for all } u \geq 0. \quad (4)$$

Our proposal is to estimate $\boldsymbol{\beta}$ and H by solving estimating equations (3) and (4). Let $\widehat{H}(u, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U)$ be the solution of Equation (4).

5. Asymptotic Theory

To facilitate the derivation and statement of the asymptotic theories, we define $\dot{\lambda}(u) = \mathrm{d}\lambda(u)/\mathrm{d}u$, $C_D(u) = E[Y(u)J\{u|W, \mathbf{Z}, H_0(u), \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U\}]$, $C_N(u) = E[Y(u)\partial J\{u|W, \mathbf{Z}, H_0(u), \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U\}/\partial H_0(u)] = E\{Y(u)\int I_2(u, x, W, \mathbf{Z}, H_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U)\mathrm{d}x\}$, $\lambda^*\{H_0(t)\} = \exp[\int_a^t \{C_N(s)/C_D(s)\}\mathrm{d}H_0(s)]$, for $0 < a \leq t \leq \tau$, and $I_1(u, x, W, \mathbf{Z}, H_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U) = [\lambda\{\boldsymbol{\beta}_{10}^T\mathbf{Z} + \beta_{20}x + H_0(u)\} - J\{u|W, \mathbf{Z}, H_0(u), \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U\}] \times G(x|u, W, \mathbf{Z}, H_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U)$, $I_2(u, x, W, \mathbf{Z}, H_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U) = [\dot{\lambda}\{\boldsymbol{\beta}_{10}^T\mathbf{Z} + \beta_{20}x + H_0(u)\} - \lambda^2\{\boldsymbol{\beta}_{10}^T\mathbf{Z} + \beta_{20}x + H_0(u)\} + J\{u|W, \mathbf{Z}, H_0(u), \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U\}\lambda\{\boldsymbol{\beta}_{10}^T\mathbf{Z} + \beta_{20}x + H_0(u)\}]G(x|u, W, \mathbf{Z}, H_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U)$ for $0 < a \leq t \leq \tau$, where $\tau = \inf\{t : \mathrm{pr}(T^* > t) = 0\}$. Here $\partial J/\partial H_0(u)$ is obtained by replacing the $H_0(u)$ function in J with a variable, taking partial derivative of J with respect to this variable, and then replacing this variable by $H_0(u)$ in the result. For a generic function $g(u, \boldsymbol{\beta}, \cdot)$, write $g_u(u, \boldsymbol{\beta}, \cdot) = \partial g(u, \boldsymbol{\beta}, \cdot)/\partial u$, $g_\beta(u, \boldsymbol{\beta}, \cdot) = \partial g(u, \boldsymbol{\beta}, \cdot)/\partial \boldsymbol{\beta}$. Also assume that $0 < t_1 < \cdots < t_{K(n)}$ are the distinct failure times in the observed data. Before we describe the asymptotic properties of the estimator, we require some lemmas.

LEMMA 1. *For any $t \in (0, \tau]$*

$$(i) \lim_{n\to\infty} \hat{H}_\beta(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U) = \lim_{n\to\infty} \frac{\partial \hat{H}(t, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U)}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \boldsymbol{\gamma}_1(t, \boldsymbol{\beta}_0, H_0, \boldsymbol{\theta}_0, f_U) + o_p(1),$$

$$(ii) \lim_{n\to\infty} \hat{H}_{\beta t}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U) = \lim_{n\to\infty} \frac{\partial^2 \hat{H}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U)}{\partial \boldsymbol{\beta}\partial t} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \boldsymbol{\gamma}_2(t, \boldsymbol{\beta}_0, H_0, \boldsymbol{\theta}_0, f_U) + o_p(1),$$

where $\boldsymbol{\gamma}_1(t) = \boldsymbol{\gamma}_1(t, \boldsymbol{\beta}_0, H_0, \boldsymbol{\theta}_0, f_U)$, $\boldsymbol{\gamma}_2(t) = \boldsymbol{\gamma}_2(t, \boldsymbol{\beta}_0, H_0, \boldsymbol{\theta}_0, f_U)$ and

$$\boldsymbol{\gamma}_1(t) = -\frac{1}{\lambda^*\{H_0(t)\}} \int_0^t \frac{\lambda^*\{H_0(s)\}}{C_D(s)} E[Y(s) \times \frac{\partial}{\partial \boldsymbol{\beta}_0} J\{s|W, Z, H_0(s), \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U\}]\mathrm{d}H_0(s), \quad (5)$$

$$\boldsymbol{\gamma}_2(t) = -\frac{E[Y(t)\partial J\{t|W_i, \mathbf{Z}_i, H_0(t), \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U\}/\partial \boldsymbol{\beta}_0] + C_N(t)\boldsymbol{\gamma}_1(t)}{C_D(t)} \times \dot{H}_0(t). \quad (6)$$

The proof of Lemma 1 and the remaining lemmas are given in the supplementary material. We now summarize the asymptotic properties of the estimator $\widehat{\boldsymbol{\beta}}$.

THEOREM 1. *Under the regularity conditions listed in the Appendix,*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to \mathrm{Normal}(0, \Sigma_1^{-1}\Sigma_*\Sigma_1^{-1})$$

in distribution when $n \to \infty$, where the expressions for Σ_* and Σ_1 are given in (A1) and (A2) respectively in the Appendix.

Despite of the relatively complex form of Σ_1 and Σ_*, we are able to derive their consistent estimators which are essential for inference purpose. Specifically,

$$\widehat{\Sigma}_1 = n^{-1}\sum_{i=1}^{n}\Bigg(\int_0^{T_i^*}\begin{pmatrix}\mathbf{Z}_i\\ W_i\end{pmatrix}\int\left\{\begin{pmatrix}\mathbf{Z}_i\\ x\end{pmatrix}+\widehat{\boldsymbol{\gamma}}_1(u)\right\}^T \times I_2(u,x,W_i,\mathbf{Z}_i,\widehat{H},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U)\mathrm{d}x\widehat{H}_u(u)\mathrm{d}u\Bigg) + n^{-1}\sum_{i=1}^{n}\left[\int_0^{T_i^*}\begin{pmatrix}\mathbf{Z}_i\\ W_i\end{pmatrix}J\{u|W_i,\mathbf{Z}_i,\widehat{H},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U\}\widehat{\boldsymbol{\gamma}}_2^{\mathrm{T}}(u)\mathrm{d}u\right],$$

where $\hat{H}(u)=\widehat{H}(u,\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U)$ is the solution of (4), $\widehat{\boldsymbol{\gamma}}_1(u)=\widehat{\boldsymbol{\gamma}}_1(u,\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U)$, $\widehat{\boldsymbol{\gamma}}_2(u)=\widehat{\boldsymbol{\gamma}}_2(u,\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U)$, and

$$\widehat{\Sigma}_* = \frac{1}{n}\sum_{i=1}^{n}\sum_{k:t_k\le T_i^*}\widehat{\Phi}_i(t_k)\widehat{\Phi}_i^T(t_k)J(t_k|W_i,\mathbf{Z}_i,\widehat{H},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U) \times\{\widehat{H}(t_k)-\widehat{H}(t_{k-1})\}+\frac{1}{n}\sum_{i=1}^{n}\left[\sum_k\widehat{\Upsilon}_i(t_k)\{\widehat{H}(t_k)-\widehat{H}(t_{k-1})\}\right]^{\otimes 2},$$

with $a^{\otimes 2}\equiv aa^T$ for any matrix or vector a and

$$\widehat{\Phi}_i(u) = \begin{pmatrix}\mathbf{Z}_i\\ W_i\end{pmatrix} - \frac{\widehat{\lambda}^*\{\widehat{H}(u)\}}{n\widehat{C}_D(u)}\sum_{j=1}^{n}\left[\int_u^{\tau}Y_j(s)\begin{pmatrix}\mathbf{Z}_j\\ W_j\end{pmatrix} \times\left\{\int I_2(s,x,W_j,\mathbf{Z}_j,\widehat{H},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U)\mathrm{d}x\right\}\frac{\mathrm{d}\widehat{H}(s)}{\widehat{\lambda}^*\{\widehat{H}(s)\}}\right] - \frac{1}{n\widehat{C}_D(u)}\sum_{j=1}^{n}\left[\begin{pmatrix}\mathbf{Z}_j\\ W_j\end{pmatrix}Y_j(u)J\{u|W_j,\mathbf{Z}_j;\widehat{H}(u),\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U\}\right] + \frac{\widehat{\lambda}^*\{\widehat{H}(u)\}}{n\widehat{C}_D(u)}\int_u^{\tau}\frac{\mathrm{d}\widehat{H}(s)\widehat{C}_N(s)}{\widehat{\lambda}^*\{\widehat{H}(s)\}\widehat{C}_D(s)} \times\sum_{j=1}^{n}\left[\begin{pmatrix}\mathbf{Z}_j\\ W_j\end{pmatrix}Y_j(s)J\{s|W_j,\mathbf{Z}_j;\widehat{H}(s),\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U\}\right],$$

and $\widehat{\lambda}^*\{\widehat{H}(u)\}=\exp[\int_a^u\{\widehat{C}_N(s)/\widehat{C}_D(s)\}\mathrm{d}\widehat{H}(s)]$, where $\widehat{C}_D(u)$ and $\widehat{C}_N(u)$ are obtained by replacing the true parameters by their consistent estimators and expectations by their empirical averages in $C_D(u)$ and $C_N(u)$. Although $\widehat{\Sigma}_1$ contains terms like $\partial\widehat{H}(\cdot,\boldsymbol{\beta},\widehat{\boldsymbol{\theta}},\widehat{f}_U)/\partial\boldsymbol{\beta}$, it does not involve $\partial\widehat{\boldsymbol{\theta}}/\partial\boldsymbol{\beta}$ or $\partial\widehat{f}_U/\partial\boldsymbol{\beta}$ as $\boldsymbol{\theta}$ and f_U were estimated independent of $\boldsymbol{\beta}$. Note that in $\widehat{\Sigma}_*$, the first term is due to the linear transformation model whereas the second term involving $\widehat{\Upsilon}_i(u)$ is due to estimated $\boldsymbol{\theta}$ and $f_U(\cdot)$, and

$$\widehat{\Upsilon}_i(u) = \Bigg[-\widehat{\mathbf{Q}}_i(\mathbf{Z}^*,u)+\widehat{\mathbf{D}}_1(\mathbf{Z}^*,u)\exp\left\{\int_0^u\frac{-\widehat{C}_N(s)\mathrm{d}\widehat{H}(s)}{\widehat{C}_D(s)}\right\} \times\int_0^u\exp\left\{\int_0^s\frac{\widehat{C}_N(l)\mathrm{d}\widehat{H}(l)}{\widehat{C}_D(l)}\right\}\frac{\widehat{\mathbf{Q}}_i(1,s)\mathrm{d}\widehat{H}(s)}{\widehat{C}_D(s)}\Bigg] + \frac{1}{n}\sum_{i=1}^{n}\left[\begin{pmatrix}\mathbf{Z}_i\\ W_i\end{pmatrix}Y_i(u)J\{u|W_i,\mathbf{Z}_i;\widehat{H}(u),\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U\}\right] \times\exp\left\{\int_0^u\frac{-\widehat{C}_N(s)\mathrm{d}\widehat{H}(s)}{\widehat{C}_D(s)}\right\} \times\Bigg[\exp\left\{\int_0^u\frac{\widehat{C}_N(s)\mathrm{d}\widehat{H}(s)}{\widehat{C}_D(s)}\right\}\frac{\widehat{\mathbf{Q}}_i(1,u)}{\widehat{C}_D(u)} - \frac{\widehat{C}_N(u)}{\widehat{C}_D(u)}\int_0^u\exp\left\{\int_0^s\frac{\widehat{C}_N(l)\mathrm{d}\widehat{H}(l)}{\widehat{C}_D(l)}\right\}\frac{\widehat{\mathbf{Q}}_i(1,s)\mathrm{d}\widehat{H}(s)}{\widehat{C}_D(s)}\Bigg].$$

Due to limited space, here we provide consistent estimators for $\mathbf{D}_1$ and $\mathbf{Q}_i$, and the actual expressions are given in the supplementary materials. Specifically, $\widehat{C}_N(t)=n^{-1}\sum_{i=1}^{n}Y_i(t)\int I_2(t,x,W_i,\mathbf{Z}_i,\widehat{H},\widehat{\boldsymbol{\theta}},\widehat{f}_U)\,\mathrm{d}x$, $\widehat{\mathbf{D}}_1(\mathbf{Z}^*,t) = n^{-1}\sum_{i=1}^{n}(\mathbf{Z}_i^T,W_i)^T\,Y_i(t)\int I_2(t,x,W_i,\mathbf{Z}_i,\widehat{H},\ \widehat{\boldsymbol{\theta}},\widehat{f}_U)\mathrm{d}x$. Let $\widehat{\mathbf{A}}_{W,\mathbf{Z}}=n^{-1}\sum_{j=1}^{n}\partial[\{\int f'_{X|\mathbf{Z}}\ (x|\mathbf{Z}_j,\boldsymbol{\theta})\widehat{f}_U(W_j-x)\mathrm{d}x\}/\{\int f_{X|\mathbf{Z}}(x|\mathbf{Z}_j,\boldsymbol{\theta})\widehat{f}_U(W_j-x)\mathrm{d}x\}]/\partial\boldsymbol{\theta}^{\mathrm{T}}$, then

$$\widehat{\mathbf{D}}_2(\mathbf{Z}^*,t) = \frac{1}{n}\sum_{j=1}^{n}\Bigg\{\int\widehat{E}\{Y(t)\mid x,\mathbf{Z}_j\}\begin{pmatrix}\mathbf{Z}_j\\ W_j\end{pmatrix} \times I_1(t,x,W_j,\mathbf{Z}_j,\widehat{H},\widehat{\boldsymbol{\theta}},\widehat{f}_U)\mathrm{d}x\mathbf{S}^{\mathrm{T}}_{W,\mathbf{Z},\boldsymbol{\theta}}(W_j,\mathbf{Z}_j,\widehat{\boldsymbol{\theta}}) - \int\widehat{E}\{Y(t)\mid x,\mathbf{Z}_j\}\begin{pmatrix}\mathbf{Z}_j\\ W_j\end{pmatrix}\mathbf{S}^{\mathrm{T}}_{X,\mathbf{Z},\boldsymbol{\theta}}(x,\mathbf{Z}_j,\widehat{\boldsymbol{\theta}}) \times I_1(t,x,W_j,\mathbf{Z}_j,\widehat{H},\widehat{\boldsymbol{\theta}},\widehat{f}_U)\mathrm{d}x\Bigg\}\widehat{\mathbf{A}}^{-1}_{W,\mathbf{Z}},$$

$$\widehat{\mathbf{Q}}_i(\mathbf{Z}^*,t) = \widehat{\mathbf{D}}_2(\mathbf{Z}^*,t)\Bigg\{\mathbf{S}_{W,\mathbf{Z},\boldsymbol{\theta}}(W_i,\mathbf{Z}_i,\widehat{\boldsymbol{\theta}}) - \frac{1}{n}\sum_{j=1}^{n}\int\mathbf{S}_{W,\mathbf{Z},\boldsymbol{\theta}}(W,\mathbf{Z}_j,\widehat{\boldsymbol{\theta}})f_{X|\mathbf{Z}}(W-v_i|\mathbf{Z}_j)\mathrm{d}W\Bigg\} + \frac{1}{n}\sum_{j=1}^{n}\int\Bigg(\widehat{E}\{Y(t)\mid W,\mathbf{Z}_j\}\begin{pmatrix}\mathbf{Z}_j\\ W\end{pmatrix} \times[\lambda\{\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}}\mathbf{Z}_j+\widehat{\beta}_2(W-v_i)+\widehat{H}(t)\} - J(t\mid W,\mathbf{Z}_j,\widehat{H},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{f}_U)]\exp [-\Lambda\{\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}}\mathbf{Z}_j+\widehat{\beta}_2(W-v_i)+\widehat{H}(t)\}] \Big/\int\exp[-\Lambda\{\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}}\mathbf{Z}_j+\widehat{\beta}_2x+\widehat{H}(t)\}]$$

Table 1

Results of the simulation study where $log(T) = -Z - X + \epsilon$. The number of replications is 500. NV, CW, and SP stand for the naive, Cheng and Wang's method, and the proposed semiparametric approach. Here SD, MSE, ESE, and CP denote the standard deviation of the estimates, mean squared error, estimated standard error based on the formula, and 95% coverage probability. The sample size was $n = 400$ and $U^* \sim \mathrm{Normal}(0, \sigma_U^2)$

		10% Censoring						50% Censoring					
		NV		CW		SP		NV		CW		SP	
r		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
0	Bias	−0.081	−0.021	0.001	0.015	0.014	0.037	−0.053	−0.159	0.089	0.394	0.019	0.088
	SD	0.187	0.063	0.233	0.103	0.215	0.105	0.241	0.069	0.354	0.218	0.262	0.113
	MSE	0.042	0.004	0.054	0.011	0.046	0.012	0.061	0.030	0.133	0.203	0.069	0.021
	ESE	0.187	0.058			0.217	0.105	0.242	0.067			0.298	0.127
	CP	0.924	0.082			0.958	0.954	0.942	0.366			0.968	0.972
0.5	Bias	−0.037	−0.173	0.028	0.019	0.001	0.033	−0.025	−0.138	0.055	0.313	0.022	0.084
	SD	0.256	0.082	0.265	0.108	0.274	0.116	0.304	0.084	0.375	0.181	0.321	0.124
	MSE	0.067	0.036	0.071	0.012	0.075	0.015	0.093	0.026	0.144	0.131	0.103	0.022
	ESE	0.265	0.081			0.279	0.117	0.309	0.089			0.341	0.126
	CP	0.96	0.42			0.952	0.964	0.954	0.652			0.968	0.96
1	Bias	−0.014	−0.161	0.029	0.013	0.014	0.028	−0.014	−0.139	0.065	0.252	0.023	0.065
	SD	0.331	0.103	0.309	0.121	0.346	0.137	0.353	0.099	0.432	0.199	0.368	0.137
	MSE	0.109	0.036	0.096	0.014	0.119	0.019	0.124	0.029	0.191	0.103	0.136	0.023
	ESE	0.342	0.101			0.352	0.133	0.356	0.106			0.370	0.142
	CP	0.958	0.614			0.956	0.956	0.954	0.714			0.948	0.954
1.5	Bias	−0.011	−0.157	0.029	0.011	0.007	0.022	−0.005	−0.140	0.062	0.177	0.025	0.054
	SD	0.406	0.125	0.378	0.145	0.417	0.161	0.403	0.117	0.502	0.230	0.416	0.155
	MSE	0.164	0.040	0.144	0.0211	0.174	0.026	0.162	0.033	0.255	0.084	0.174	0.026
	ESE	0.419	0.120			0.427	0.154	0.405	0.119			0.417	0.156
	CP	0.962	0.714			0.958	0.946	0.954	0.76			0.952	0.946
2	Bias	−0.008	−0.157	0.032	0.008	0.005	0.018	0.000	−0.141	0.100	0.151	0.026	0.047
	SD	0.479	0.148	0.406	0.152	0.491	0.186	0.448	0.131	0.567	0.269	0.460	0.169
	MSE	0.229	0.046	0.165	0.023	0.241	0.034	0.201	0.037	0.331	0.095	0.212	0.031
	ESE	0.498	0.141			0.505	0.178	0.452	0.133			0.463	0.172
	CP	0.964	0.792			0.966	0.942	0.948	0.80			0.95	0.958

$$
\times f_{X|W,\mathbf{Z}}(x \mid W, \mathbf{Z}_j)\mathrm{d}x
- \int E\{Y(t) \mid x, \mathbf{Z}_j\} \begin{pmatrix} \mathbf{Z}_j \\ W \end{pmatrix} I_1(t, x, W, \mathbf{Z}_j, \widehat{H}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U)\mathrm{d}x \Bigg)
\times f_{X|\mathbf{Z}}(W - v_i|\mathbf{Z}_j)\mathrm{d}W,
$$

where $\widehat{E}\{Y(t) \mid W_j, \mathbf{Z}_j\}$ is an estimate of $E\{Y(t) \mid W_j, \mathbf{Z}_j\} = \mathrm{pr}(Y(t) = 1 \mid W_j, \mathbf{Z}_j)$. To estimate $E\{Y(t) \mid x, \mathbf{Z}_j\} = \mathrm{pr}(Y(t) = 1 \mid x, \mathbf{Z}_j)$, we make use of $\mathrm{pr}\{Y(t) = 1 \mid w, \mathbf{z}\} = \int \mathrm{pr}\{Y(t) = 1 \mid x, \mathbf{z}\} f_{X|W,\mathbf{Z}}(x \mid w, \mathbf{z})\mathrm{d}x$. We have used a linear logistic model to estimate $\mathrm{pr}\{Y(t) = 1 \mid w, \mathbf{z}\}$ and $\mathrm{pr}\{Y(t) = 1 \mid x, \mathbf{z}\}$. Alternatively, $\mathrm{pr}\{Y(t) = 1 \mid w, \mathbf{z}\}$ can be estimated through the nonparametric kernel method and $\mathrm{pr}\{Y(t) = 1 \mid x, \mathbf{z}\}$ can be estimated via the deconvolution approach (Carroll and Hall, 1989). The estimator of $\boldsymbol{\theta}$ and f_U are plugged into the estimating equations (3) and (4), and then $\widehat{\boldsymbol{\beta}}$ is obtained. Because of this profiling procedure, analyzing the asymptotic behavior $\widehat{\boldsymbol{\beta}}$ requires us to study the asymptotic behavior of $\widehat{\boldsymbol{\theta}}$ and $\widehat{f}_U$ as well. A sketch proof of the theorem is given in the appendix whereas the detailed derivation is collected in the online supplementary materials.

6. Simulation Study

In order to investigate the performance of the proposed approach in finite sample we carried out simulation studies. We generated X and Z from a Normal$(0, 1)$ and a Uniform$(0, 1)$ distribution, respectively. We further generated the time to event, T, from the model $\log(T) = -Z - X + e$, where e is generated from the distribution with its hazard function $\lambda(u) = \exp(u)/\{1 + r\exp(u)\}$. We considered $r = 0, 0.5, 1, 1.5$, and 2. Note that $r = 0$ and 1 correspond to the proportional hazard model and the proportional odds model, respectively. We set the censoring variable $C = X^2 + \mathrm{Uniform}(0, K)$, and choose K to yield 10% and 50% right censored data. The erroneous measurement W^* is created by adding noise U^* to X, and we generated U^* from two different distributions, Normal$(0, \sigma_U^2)$ with $\sigma_U^2 = 0.5$ and Uniform$(-1.75, 1.75)$. We generated three replicates of W^* for each subject, and here we present the results for $n = 400$. The supplementary material contains the results for $n = 200$.

We analyzed each data set by using the naive (NV) approach, the method proposed in Cheng and Wang (2001) (hereafter referred to as CW), and the proposed semiparametric (SP) approach. In the naive approach, we used the method proposed in Chen et al. (2002) and used $\overline{W}_i^* = \sum_{j=1}^m W_{ij}^*/m$ in

Table 2

Results of the simulation study where log(T) = −Z − X + ϵ. The number of replications is 500. NV, CW, and SP stand for the naive, Cheng and Wang's method, and the proposed semiparametric approach. Here SD, MSE, ESE, and CP denote the standard deviation of the estimates, mean squared error, estimated standard error based on the formula, and 95% coverage probability. The sample size was $n = 400$ and $U^* \sim \sigma_U$Uniform(−1.75, 1.75)

		10% Censoring						50% Censoring					
		NV		CW		SP		NV		CW		SP	
r		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
0	Bias	−0.078	−0.212	0.013	0.020	0.018	0.037	−0.054	−0.162	0.102	0.394	0.019	0.086
	SD	0.181	0.061	0.234	0.105	0.209	0.102	0.241	0.072	0.359	0.224	0.263	0.115
	MSE	0.038	0.046	0.055	0.011	0.044	0.012	0.061	0.031	0.139	0.205	0.069	0.021
	ESE	0.187	0.058			0.216	0.101	0.242	0.067			0.290	0.118
	CP	0.946	0.072			0.952	0.952	0.944	0.36			0.972	0.936
0.5	Bias	−0.032	−0.176	0.023	0.018	0.014	0.032	−0.024	−0.139	0.072	0.311	0.024	0.085
	SD	0.253	0.079	0.267	0.107	0.272	0.114	0.304	0.086	0.376	0.191	0.320	0.124
	MSE	0.065	0.037	0.072	0.012	0.074	0.014	0.093	0.026	0.146	0.133	0.103	0.023
	ESE	0.265	0.081			0.279	0.109	0.309	0.089			0.323	0.120
	CP	0.964	0.406			0.96	0.946	0.962	0.626			0.952	0.949
1	Bias	−0.006	−0.165	0.026	0.016	0.020	0.026	−0.012	−0.141	0.089	0.253	0.026	0.067
	SD	0.331	0.099	0.312	0.118	0.347	0.134	0.354	0.102	0.433	0.198	0.369	0.141
	MSE	0.109	0.037	0.098	0.014	0.121	0.018	0.125	0.030	0.195	0.103	0.137	0.024
	ESE	0.343	0.101			0.352	0.133	0.356	0.105			0.371	0.143
	CP	0.96	0.594			0.96	0.954	0.946	0.706			0.956	0.942
1.5	Bias	−0.003	−0.161	0.018	0.015	0.015	0.021	−0.003	−0.141	0.076	0.174	0.029	0.058
	SD	0.405	0.119	0.375	0.133	0.418	0.155	0.404	0.120	0.503	0.226	0.418	0.161
	MSE	0.164	0.040	0.141	0.018	0.175	0.024	0.163	0.034	0.259	0.081	0.176	0.029
	ESE	0.419	0.120			0.427	0.155	0.404	0.119			0.417	0.156
	CP	0.964	0.708			0.962	0.958	0.958	0.76			0.956	0.948
2	Bias	0.000	−0.161	0.032	0.011	0.014	0.017	0.003	−0.142	0.098	0.138	0.031	0.051
	SD	0.480	0.141	0.409	0.149	0.493	0.179	0.449	0.135	0.570	0.265	0.462	0.176
	MSE	0.230	0.046	0.168	0.022	0.243	0.032	0.202	0.038	0.335	0.089	0.214	0.034
	ESE	0.498	0.141			0.505	0.171	0.452	0.132			0.464	0.172
	CP	0.96	0.766			0.96	0.950	0.954	0.788			0.948	0.952

place of X_i, where $m = 3$. The standard error of the naive approach was calculated based on the formula given in (8) of Chen et al. (2002). In the CW method we estimated the parameters by solving Equation (7) of CW and we assumed that censoring mechanism is independent of T, X, Z and W^*. Furthermore, we assumed that for $i \neq j$, $X_{ij} \equiv (X_i - X_j)|W_{ij}^+ \sim \text{Normal}\{(1-\rho^2)W_{ij}^+, \rho^2\sigma_W^2\}$, where $W_{ij}^+ \equiv \overline{W}_i^* - \overline{W}_j^*$, $\rho^2 = \sigma_e^2/(m\sigma_W^2 + \sigma_e^2)$, $\sigma_e^2 = \text{var}(W_{il}^* - W_{jl'}^*|X_{ij})$, $\sigma_W^2 = \text{var}(X_{ij})$. For $m = 3$, we used $\widehat{\sigma_W^2} = (3n^2)^{-1}\sum_{i=1}^n\sum_{j=1}^n(W_{ij1}^*W_{ij2}^* + W_{ij1}^*W_{ij3}^* + W_{ij2}^*W_{ij3}^*)$, and $\widehat{\sigma_e^2} = (6n^2)^{-1}\sum_{i=1}^n\sum_{j=1}^n\{(W_{ij1}^*)^2 + (W_{ij2}^*)^2 + (W_{ij3}^*)^2\}$, where $W_{ijl}^* = W_{il}^* - W_{jl}^*$. The survival probability of the censoring variable was estimated by the Kaplan-Meier method using the data $\{T_i^*, (1-\delta_i)\}$. In the SP method we assumed that X given Z follows a normal distribution with mean $\gamma_0 + \gamma_1 Z$ and variance σ_x^2. To estimate $f_U(\cdot)$ following the process described in Section 3, we selected the bandwidth h via the plug-in bandwidth selection method from Sheather and Jones (1991). The standard error of the SP method was calculated based on the formula given in Section 5.

For the NV and SP methods we present the bias, the standard deviation of the estimates, the mean squared error (MSE), the estimated standard error, and the coverage rate of the 95% confidence intervals. For the CW method, we present only the bias, standard deviation, and the MSE of the estimates. The results shown in Tables 1 and 2 (and Tables 1 and 2 in the supplementary materials) clearly indicate that the estimates of β_2, the coefficient for X, are highly biased in the NV method and the corresponding coverage probabilities are strikingly lower than the nominal level 0.95. The CW method has large biases when the censoring rate is high, caused by the violation of the censoring mechanism assumption. In contrast, the proposed SP method has less bias compared with the naive approach and the CW approach when 50% data are censored, the estimated standard errors are closer to the sample standard deviation of the estimates and the coverage rates are close to 95%. In addition, using the proposed SP method, the amount of estimation bias decreases as the sample sizes increase, and the standard errors are decreasing with sample size. These features reflect the established asymptotic properties of the proposed estimator.

In order to assess the robustness of our approach we simulated Z from the Normal(0, 1) distribution and X from a two component mixture of normals distribution RNormal$(-0.6, 0.5^2)$ $+(1-R)$Normal$(1.25, 0.5^2)$, where $R \sim$ Bernoulli(0.33) distribution. We generated U^* from the Normal$(0, 0.71^2)$ distribution and an asymmetric distribution R^*Uniform$(-1, 0) + (1-R^*)$Normal$(0.499, 0.65^2)$, where R^*

Table 3

Results of the simulation study where $log(T) = -Z - X + \epsilon$. The number of replications is 500. NV, CW, and SP stand for the naive, Cheng and Wang's method, and the proposed semiparametric approach. Here SD, MSE, ESE, and CP denote the standard deviation of the estimates, mean squared error, estimated standard error based on the formula, and 95% coverage probability. Here X followed a mixture of normals and the sample size was $n = 400$

		10% Censoring						80% Censoring					
		NV		CW		SP		NV		CW		SP	
r		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$U^* \sim \text{Normal}(0, 0.71^2)$													
0	Bias	−0.093	−0.216	0.004	0.015	0.001	0.020	−0.022	−0.228	−0.336	−0.431	0.017	−0.051
	SD	0.193	0.059	0.236	0.105	0.219	0.104	0.356	0.111	0.309	0.111	0.378	0.149
	MSE	0.046	0.050	0.056	0.011	0.048	0.011	0.127	0.064	0.208	0.198	0.143	0.025
	ESE	0.187	0.059			0.349	0.112	0.215	0.103			0.365	0.150
	CP	0.928	0.072			0.956	0.956	0.952	0.472			0.954	0.918
1	Bias	−0.023	−0.156	0.007	0.010	0.008	0.012	0.008	−0.197	−0.354	−0.409	0.024	−0.033
	SD	0.332	0.099	0.313	0.117	0.354	0.134	0.462	0.139	0.401	0.134	0.481	0.185
	MSE	0.111	0.034	0.098	0.014	0.125	0.018	0.214	0.058	0.286	0.185	0.232	0.035
	ESE	0.343	0.100			0.353	0.128	0.465	0.145			0.477	0.187
	CP	0.956	0.626			0.952	0.946	0.952	0.68			0.956	0.944
2	Bias	−0.005	−0.154	0.019	0.008	0.012	0.005	−0.009	−0.178	−0.244	−0.301	0.015	−0.018
	SD	0.483	0.144	0.409	0.147	0.496	0.183	0.524	0.156	0.542	0.166	0.537	0.202
	MSE	0.233	0.044	0.168	0.022	0.246	0.034	0.275	0.056	0.353	0.118	0.289	0.041
	ESE	0.502	0.140			0.509	0.175	0.528	0.160			0.542	0.209
	CP	0.966	0.794		0.96	0.946	0.954	0.766				0.96	0.956
$U^* \sim r\text{Uniform}(-1, 0) + (1 - r)\text{Normal}(0.499, 0.65^2)$, $r \sim \text{Bernoulli}(0.5)$													
0	Bias	−0.097	−0.225	0.004	0.009	0.009	0.020	−0.027	−0.238	−0.335	−0.432	0.019	−0.065
	SD	0.197	0.063	0.239	0.107	0.223	0.109	0.353	0.108	0.309	0.114	0.379	0.152
	MSE	0.048	0.055	0.057	0.012	0.050	0.012	0.125	0.068	0.208	0.200	0.144	0.027
	ESE	0.186	0.059			0.216	0.104					0.366	0.150
	CP	0.906	0.076			0.952	0.956	0.954	0.438			0.946	0.90
1	Bias	−0.02	−0.154	0.010	0.026	0.018	0.022	−0.011	−0.204	−0.353	−0.418	0.18	−0.060
	SD	0.336	0.101	0.313	0.118	0.357	0.134	0.474	0.144	0.403	0.135	0.484	0.186
	MSE	0.113	0.034	0.098	0.015	0.128	0.018	0.225	0.062	0.287	0.193	0.267	0.038
	ESE	0.344	0.101			0.354	0.129	0.466	0.146			0.479	0.188
	CP	0.946	0.61			0.948	0.950	0.952	0.66			0.954	0.93
2	Bias	−0.003	−0.142	0.027	0.017	0.019	0.015	0.006	−0.186	−0.255	−0.303	0.029	−0.035
	SD	0.486	0.143	0.408	0.146	0.500	0.179	0.527	0.162	0.541	0.184	0.539	0.204
	MSE	0.236	0.041	0.167	0.022	0.250	0.032	0.278	0.061	0.358	0.126	0.291	0.043
	ESE	0.502	0.140			0.509	0.175	0.527	0.160			0.541	0.204
	CP	0.96	0.792			0.946	0.960	0.956	0.744			0.952	0.958

has a Bernoulli(0.5) distribution. This is a clear violation of our assumption. The censoring variable C was generated from a Uniform$(0, K)$ distribution independently of any other variables, and we choose K to yield 10% and 80% right censored data, the latter resembles the very high percentage of censoring in the real data example. In the SP method we assumed that conditional on Z, X follows a normal distribution. Table 3 contains the results for $r = 0, 1$ and 2. The results indicate that SP method works quite well even when (1) the assumed model for X given Z is quite wrong and (2) measurement errors follow an asymmetric distribution. This suggests that SP seems to have certain robustness property, and moderate model violations may not have much visible impact on the results of the analysis. However, we would like to caution that in our experience, when the distribution of X conditional on Z is very different from the truth, estimation bias does occur.

To avoid possible convergence issues, we recentered W^*_{ij} by subtracting the mean of all W^*_{ij}. In the numerical calculations, we used the Hermite quadrature formula to compute integrals with respect to x. Since $\hat{H}(\cdot)$ is a step function, any Stieltjes integral with respect to $\widehat{H}(\cdot)$ is a sum of the product of the integrand and $\mathrm{d}\widehat{H}(\cdot)$, that is, $\int f \mathrm{d}\widehat{H} = \sum_{j:T^*_j,\Delta_j=1} f(T^*_j)\mathrm{d}\widehat{H}(T^*_j)$. The naive estimates were used as the initial values for the computation of the SP method. We did not encounter any convergence issue in using the Newton–Raphson procedure to estimate the parameters. We point out that the computation time of these methods changes with the sample sizes and the percentage of censoring. For example, for $n = 400$, 10% censored data, $r > 0$, and 500 replications, computation of the estimates takes about 3.15 hours for the SP method and 16.3 hours for the CW method. All simulations were done on a 2.8GHz Intel Xeon X5560 processor.

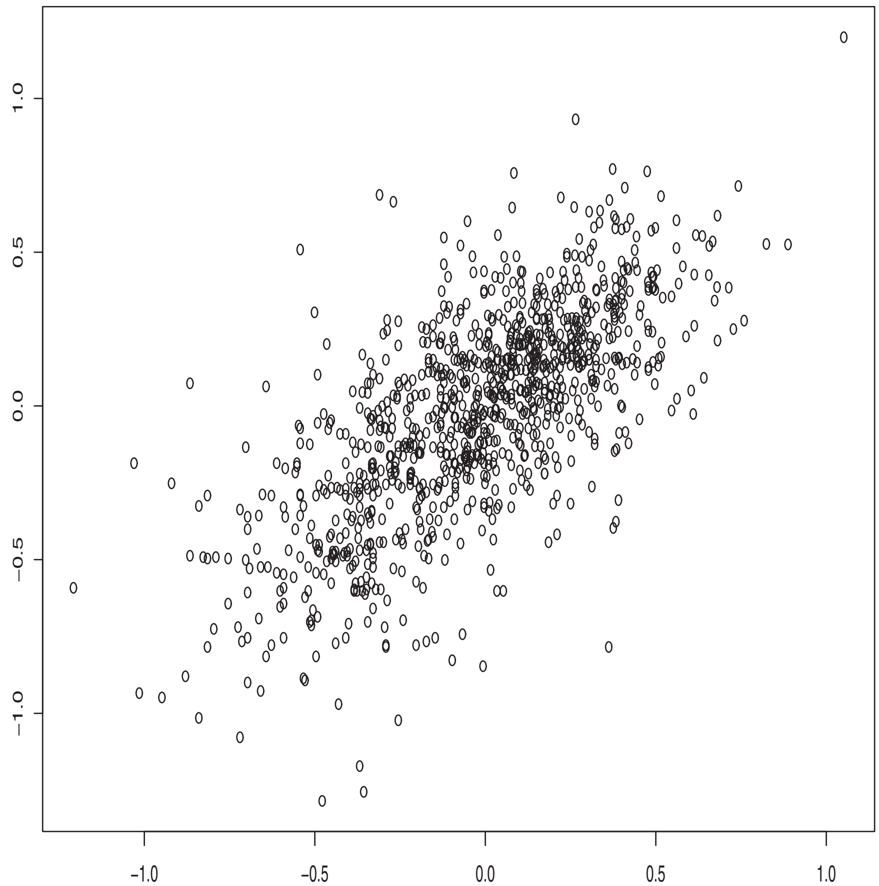

Figure 1. Plot of two log(CD4) measurements of 1036 subjects in the ACTG 175 data set.

7. Analysis of the AIDS Clinical Trial Data

We now apply the proposed method to analyze a subset of the data from the ACTG 175 study. This was a randomized, double-blind, placebo-controlled trial to understand the effect of the four therapies, 600 mg of zidovudine, 600 mg of zidovudine plus 400 mg of didanosine, 600 mg of zidovudine plus 2.25 mg of zalcitabine and 400 mg of didanosine on HIV-1 infected patients, see Hammer et al. (1996) for details. For our analysis we considered only $n = 1036$ subjects who did not have antiretroviral treatment before this trial, and 262, 257, 260, 257 subjects received the above 4 treatments, respectively. These subjects had two (i.e., $m = 2$) baseline measurements of CD4 counts prior to the start of their treatment, see Figure 1 for a scatter plot of these measurements. We considered time to AIDs or death from the date treatment started as the response variable T. Among the 1036 subjects, 85 subjects experienced the above event, and the median and average follow-up time were approximately 27 and 32 months, respectively.

We fit model (1) to this data set, where the logarithm of the actual CD4 count at the baseline minus 5.89 is considered as X. The choice of 5.89 is to make the distribution centered around 0. The two baseline measurements are considered to be two erroneous measurements for X. The three dummy variables corresponding to the four treatments are considered to be error free covariates $\mathbf{Z}$ where 600 mg of zidovudine was considered as the reference category. We modeled the hazard function of e in model (1) through $h_e(u) = \exp(u)/\{1 + r\exp(u)\}$, and choose $r = 0, 0.5, 1, 1.5, 2$. Table 4 contains the estimate and the estimated standard error based on NV, CW, and SP methods. Figure 2 shows the deconvoluted density of $X = \log(\text{CD4})$ under Normal and Laplacian errors for each treatment category obtained from $\overline{W}_1^*, \cdots, \overline{W}_n^*$ where $\overline{W}_i^* = \sum_{j=1}^{2} W_{ij}^*/2$. The bandwidth was selected using 1000 bootstrap samples. Since none of the deconvoluted densities deviates much from a unimodal bell-shaped curve, in the SP method we assume that X given $\mathbf{Z}$ follows a normal distribution.

The results of both the naive and the proposed methods indicate that log(CD4) has a statistically significant effect on the time to event. More importantly, after adjusting for the measurement errors, the estimate of the coefficient for CD4 counts, β_2, is very different from the naive estimate. Overall, the naive estimate tends to underestimate the effect of CD4 counts. We find that compared to the monotherapy with zidovudine, other three therapies have statistically significant effect on delaying the time to event.

8. Conclusions

We have proposed a semiparametric method for handling mismeasured covariates in the linear transformation model. This approach resolves the errors-in-covariates issue in many

Table 4

Analysis of the ACTG 175 aids clinical trial data. Here Est. and SE stand for estimate and standard error, respectively. Also, Z, Z+D, Z+Z, and D stand for zidovudine, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine, respectively

		$r = 0$		$r = 0.5$		$r = 1$		$r = 1.5$		$r = 2$	
Method		Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
NV	Z+D (Ref: Z)	−0.706	0.292	−0.736	0.304	−0.765	0.317	−0.795	0.330	−0.824	0.343
	Z+Z (Ref: Z)	−0.919	0.310	−0.956	0.322	−0.993	0.334	−1.030	0.345	−1.067	0.358
	D (Ref: Z)	−0.703	0.282	−0.733	0.295	−0.763	0.308	−0.793	0.321	−0.882	0.334
	log(CD4)	−2.033	0.348	−2.114	0.368	−2.195	0.388	−2.275	0.406	−2.355	0.425
CW	Z+D (Ref: Z)	−0.109	0.082	−0.137	0.111	−0.164	0.135	−0.191	0.159	−0.219	0.178
	Z+Z (Ref: Z)	−0.178	0.066	−0.222	0.089	−0.266	0.107	−0.311	0.127	−0.356	0.143
	D (Ref: Z)	−0.145	0.076	−0.181	0.099	−0.216	0.119	−0.253	0.139	−0.289	0.154
	log(CD4)	−0.569	0.129	−0.711	0.159	−0.853	0.188	−0.996	0.217	−1.139	0.246
SP	Z+D (Ref: Z)	−0.736	0.295	−0.770	0.309	−0.804	0.324	−0.837	0.338	−0.871	0.352
	Z+Z (Ref: Z)	−0.943	0.313	−0.986	0.326	−1.029	0.339	−1.071	0.353	−1.113	0.367
	D (Ref: Z)	−0.697	0.286	−0.730	0.300	−0.762	0.315	−0.794	0.329	−0.826	0.342
	log(CD4)	−2.602	0.456	−2.714	0.485	−2.826	0.513	−2.937	0.540	−3.049	0.569

models including the proportional hazard model and proportional odds model. We make minimal assumptions on the errors associated with the covariate, and resort to a nonparametric estimation of this error density using repeated measurements. We work in the structural model framework and make a parametric model assumption regarding the

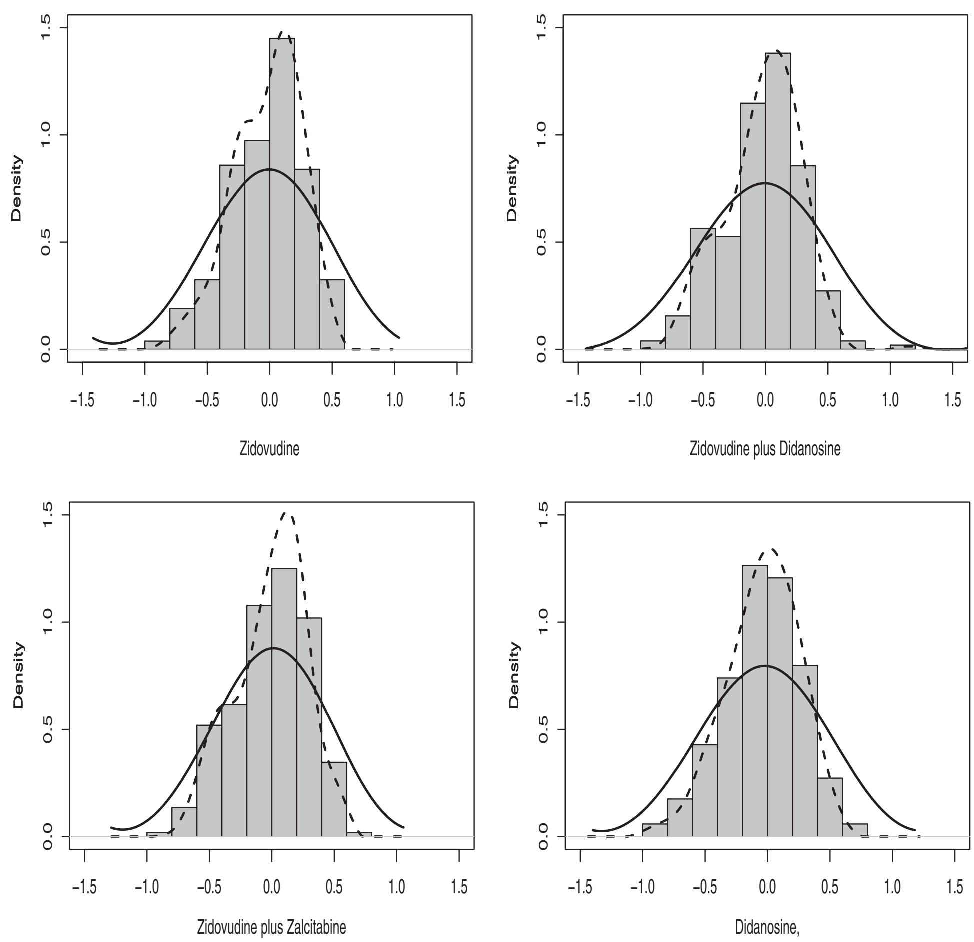

Figure 2. Deconvoluted density plot of log(CD4) − 5.89 based on laplacian (dotted) and normal errors (solid) along with the histogram of the average of the erroneous measurements of log(CD4) − 5.89.

distribution of X given Z. Although a misspecification of this assumption will lead to a biased estimator in principle, in practice, one can make this bias negligible by taking a flexible parametric model, such as a mixture of normals. Simulation study indicated that a moderate model violation may not have much impact on the results, while severe model violation certainly will bring estimation bias.

The proposed method has two steps. In the first step, we estimate the model parameters involved in the distribution of X given Z. In the second step, we use these estimated parameters to estimate the main regression parameters. The two step procedure is easy to implement and is tractable analytically. Alternatively, a maximum likelihood based estimator is also possible, where one would maximize the joint density of (T_i^*, Δ_i, W_i^*) given $\mathbf{Z}_i$ with $f_U(\cdot)$ being estimated separately and plugged in. This will require separate analysis both theoretically and numerically. The difference between this approach and our approach is that of the difference between the Martingale based approach and the NPMLE based approach in general survival analysis, and further investigation will be interesting and fruitful.

Results of the simulation study clearly indicate the usefulness of the proposed approach. In principle, the proposed method can be extended to handle the scenario where the main data contains only a single and possibly a biased measurement of the true covariate, where an external calibration study contains this biased surrogate variable and some gold standard measurements of the true covariate to assess the error distribution.

9. Supplementary Materials

Proofs and tables referenced in Sections 5 and 6, and the source code for computation are available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

This research was partially supported by NSF grant SES-0961618, 1206693, 1000354 and NIH grants NS073671-01 and 1R03CA176760-01 from the NCI.

Appendix: A sketch proof of Theorem 1.

A Taylor expansion of the estimating equation yields

$$\begin{aligned}
0 &= n^{-1/2}\mathbf{U}_\beta\{\widehat{\boldsymbol\beta}, \widehat{H}(\cdot, \widehat{\boldsymbol\beta}, \widehat{\boldsymbol\theta}, \widehat{f}_U), \widehat{\boldsymbol\theta}, \widehat{f}_U\} \\
&= E\left[\frac{\partial}{n\partial\boldsymbol\beta_0^{\mathrm{T}}}\mathbf{U}_\beta\{\boldsymbol\beta_0, \hat{H}(\cdot, \boldsymbol\beta_0, \boldsymbol\theta_0, f_U), \boldsymbol\theta_0, f_U\}\right]\sqrt{n}(\hat{\boldsymbol\beta} - \boldsymbol\beta_0) \\
&\quad + \frac{1}{\sqrt{n}}\mathbf{U}_\beta\{\boldsymbol\beta_0, \widehat{H}(\cdot, \boldsymbol\beta_0, \widehat{\boldsymbol\theta}, \widehat{f}_U), \widehat{\boldsymbol\theta}, \widehat{f}_U\} \\
&\quad + o_p(1) = -\Sigma_1\sqrt{n}(\hat{\boldsymbol\beta} - \boldsymbol\beta_0) \\
&\quad + \frac{1}{\sqrt{n}}\mathbf{U}_\beta\{\boldsymbol\beta_0, \widehat{H}(\cdot, \boldsymbol\beta_0, \widehat{\boldsymbol\theta}, \widehat{f}_U), \widehat{\boldsymbol\theta}, \widehat{f}_U\} + o_p(1).
\end{aligned}$$

Thus, we first consider the asymptotic expansion of $n^{-1/2}\mathbf{U}_\beta\{\boldsymbol\beta_0, \hat{H}(\cdot, \boldsymbol\beta_0, \widehat{\boldsymbol\theta}, \widehat{f}_U), \widehat{\boldsymbol\theta}, \widehat{f}_U\}$. In the supplementary materials we show that

$$\begin{aligned}
&n^{-1/2}U_\beta\{\boldsymbol\beta_0, \hat{H}(\cdot, \boldsymbol\beta_0, \widehat{\boldsymbol\theta}, \widehat{f}_U), \widehat{\boldsymbol\theta}, \widehat{f}_U\} \\
&\quad = n^{-1/2}\sum_{i=1}^n \int_0^\tau \{\Phi_i(u)\mathrm{d}M_i(u) + \Upsilon_i(u)\mathrm{d}H_0(u)\} + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\Phi_i(u) = {}& \begin{pmatrix}\mathbf{Z}_i \\ W_i\end{pmatrix} - \frac{\lambda^*\{H_0(u)\}}{C_D(u)} E\Bigg(\int_u^\tau Y(s)\begin{pmatrix}\mathbf{Z} \\ W\end{pmatrix} \\
&\times \Bigg[\int \dot\lambda\{\boldsymbol\beta_{10}^T\mathbf{Z} + \beta_{20}x + H_0(s)\} \\
&\times G(x|s, W, \mathbf{Z}; H_0, \boldsymbol\beta_0, \boldsymbol\theta_0, f_U)\mathrm{d}x \\
&- \int \lambda^2\{\boldsymbol\beta_{10}^T\mathbf{Z} + \beta_{20}x + H_0(s)\} \\
&\times G(x|s, W, \mathbf{Z}; H_0, \boldsymbol\beta_0, \boldsymbol\theta_0, f_U)\mathrm{d}x \\
&+ J^2(s|W, \mathbf{Z}; H_0, \boldsymbol\beta_0, \boldsymbol\theta_0, f_U)\Bigg]\frac{\mathrm{d}H_0(s)}{\lambda^*\{H_0(s)\}}\Bigg) \\
&- \frac{1}{C_D(u)}E\left[\begin{pmatrix}\mathbf{Z} \\ W\end{pmatrix}Y(u)J\{u|W, \mathbf{Z}; H_0(u), \boldsymbol\beta_0, \boldsymbol\theta_0, f_U\}\right] \\
&+ \frac{\lambda^*\{H_0(u)\}}{C_D(u)}\int_u^\tau \frac{\mathrm{d}H_0(s)C_N(s)}{\lambda^*\{H_0(s)\}C_D(s)} \\
&\times E\left[\begin{pmatrix}\mathbf{Z} \\ W\end{pmatrix}Y(s)J\{s|W, \mathbf{Z}; H_0(s), \boldsymbol\beta_0, \boldsymbol\theta_0, f_U\}\right],
\end{aligned}$$

$$\begin{aligned}
\Upsilon_i(u) = {}& \Bigg[-\mathbf{Q}_i(\mathbf{Z}^*, u) + \mathbf{D}_1(\mathbf{Z}^*, u)\exp\left\{\int_0^u \frac{-C_N(s)\mathrm{d}H_0(s)}{C_D(s)}\right\} \\
&\times \int_0^u \exp\left\{\int_0^s \frac{C_N(l)\mathrm{d}H_0(l)}{C_D(l)}\right\}\frac{\mathbf{Q}_i(1, s)\mathrm{d}H_0(s)}{C_D(s)}\Bigg] \\
&+ E\left[\begin{pmatrix}\mathbf{Z} \\ W\end{pmatrix}Y(u)J\{u|W, \mathbf{Z}; H_0(u), \boldsymbol\beta_0, \boldsymbol\theta_0, f_U\}\right] \\
&\times \exp\left\{\int_0^u \frac{-C_N(s)\mathrm{d}H_0(s)}{C_D(s)}\right\} \\
&\times \Bigg[\exp\left\{\int_0^u \frac{C_N(s)\mathrm{d}H_0(s)}{C_D(s)}\right\}\frac{\mathbf{Q}_i(1, u)}{C_D(u)} - \frac{C_N(u)}{C_D(u)} \\
&\times \int_0^u \exp\left\{\int_0^s \frac{C_N(l)\mathrm{d}H_0(l)}{C_D(l)}\right\}\frac{\mathbf{Q}_i(1, s)\mathrm{d}H_0(s)}{C_D(s)}\Bigg].
\end{aligned}$$

Observe that $\Phi_i(u)$ is a predictable and bounded process for $u \in (0, \tau]$ with respect to the filtration $\mathcal{F}_{u-} = \sigma\{Y(s), N(s), \mathbf{Z}, W, 0 \le s < u\}$. Due to the martingale property $E\{\int_0^\tau \Phi_i(u)\mathrm{d}M_i(u)\} = 0$. On the other hand, $\Upsilon_i(u)$ belongs to a Hilbert space of square integrable random variable with zero mean, that is, $E\{\Upsilon_i^2(u)\} < \infty$, $E\{\Upsilon_i(u)\} = 0$ for all $u \in (0, \tau]$. Now, using the Martingale central limit theorem we can write

$n^{-1/2}U_\beta\{\boldsymbol{\beta}_0, \hat{H}(\cdot, \boldsymbol{\beta}_0, \widehat{\boldsymbol{\theta}}, \widehat{f}_U), \widehat{\boldsymbol{\theta}}, \widehat{f}_U\}$ asymptotically follows a normal distribution with mean 0 and variance

$$\Sigma_* = E\left[\left\{\int_0^\tau \Phi_i(u)\Phi_i^T(u)Y_i(u)\lambda_T(u|W_i, \mathbf{Z}_i, H_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U)\right\}\mathrm{d}u + \left\{\int_0^\tau \Upsilon_i(u)\mathrm{d}H_0(u)\right\}^{\otimes 2}\right]. \quad \text{(A1)}$$

The above equality used the fact that $\mathrm{cov}\{\int_0^\tau \Phi_i(t)\mathrm{d}M_i(t), \int_0^\tau \Upsilon_i(t)\mathrm{d}H_0(t)\} = 0$. Observe that the randomness of $\Upsilon_i(u)$ comes only from its random covariates $W_i, \mathbf{Z}_i$, and consequently for any $u, u' \in (0, \tau]$, $\mathrm{cov}\{\mathrm{d}M_i(u), \Upsilon_i(u')\} = E[E\{\mathrm{d}M_i(u)\Upsilon_i(u')|\mathcal{F}_{u-}\}] = E[\Upsilon_i(u')E\{\mathrm{d}M_i(u)|\mathcal{F}_{u-}\}] = 0$. Hence,

$$\begin{aligned}
&\mathrm{cov}\{\int_0^\tau \Phi_i(t)\mathrm{d}M_i(t), \int_0^\tau \Upsilon_i(t)\mathrm{d}H_0(t)\} \\
&= E\left\{\int_0^\tau\int_0^\tau \Phi_i(u)\mathrm{d}M_i(u)\Upsilon_i(u')\mathrm{d}H_0(u')\right\} \\
&= E\left\{\int_0^\tau\int_0^\tau \Upsilon_i(u')\Phi_i(u)E(\mathrm{d}M_i(u)|\mathcal{F}_{u-})\mathrm{d}H_0(u')\right\} = 0.
\end{aligned}$$

We now consider the calculation of Σ_1. Observe that

$$\begin{aligned}
&\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\mathbf{U}_\beta\{\boldsymbol{\beta}, \widehat{H}(\cdot, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U), \boldsymbol{\theta}_0, f_U\} \\
&= \frac{1}{n}\sum_{i=1}^n \begin{pmatrix}\mathbf{Z}_i \\ W_i\end{pmatrix} \frac{\partial}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\int_0^\tau \{\mathrm{d}N_i(u) - Y_i(u) \\
&\quad J(u|W_i, \mathbf{Z}_i, \widehat{H}, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U)\hat{H}_u(u, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U)\}\mathrm{d}u \\
&= -\frac{1}{n}\sum_{i=1}^n \begin{pmatrix}\mathbf{Z}_i \\ W_i\end{pmatrix}\int_0^\tau Y_i(u)\big[J_\beta\{u|W_i, \mathbf{Z}_i, \hat{H} \\
&\quad \times (u, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U), \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U\}\hat{H}_u(u, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U) \\
&\quad + J(u|W_i, \mathbf{Z}_i, \hat{H}, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U)\hat{H}_{\beta u}(u, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U)\big]^{\mathrm{T}}\mathrm{d}u,
\end{aligned}$$

where $J_\beta\{u|W_i, \mathbf{Z}_i, \hat{H}(u, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U), \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U\} = \int[\dot{\lambda}\{\boldsymbol{\beta}_1^T\mathbf{Z}_i + \beta_2 x + \hat{H}_0(u, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U)\} - \lambda^2\{\boldsymbol{\beta}_1^T\mathbf{Z}_i + \beta_2 x + \hat{H}(u, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U)\} + J\{u|W_i, \mathbf{Z}_i, \hat{H}(\cdot, \boldsymbol{\beta}), \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U\}\lambda\{\boldsymbol{\beta}_1^T\mathbf{Z}_i + \beta_2 x + \hat{H}(u, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U)\}](\mathbf{Z}_i^T, x)^T + \hat{H}_\beta(u, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U)]G\{x|u, W_i, \mathbf{Z}_i, \hat{H}(\cdot, \boldsymbol{\beta}), \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U\}\mathrm{d}x$. After setting $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ we obtain

$$\begin{aligned}
&-\frac{1}{n}E\frac{\partial}{\partial\boldsymbol{\beta}^{\mathrm{T}}}U_\beta\{\boldsymbol{\beta}, \hat{H}(\cdot, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}, \widehat{f}_U), \widehat{\boldsymbol{\theta}}, \widehat{f}_U\}\mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \stackrel{a.s}{\rightarrow} \\
&-\frac{1}{n}E\frac{\partial}{\partial\boldsymbol{\beta}^{\mathrm{T}}}U_\beta\{\boldsymbol{\beta}, \hat{H}(\cdot, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U), \boldsymbol{\theta}_0, f_U\}\mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \stackrel{a.s}{\rightarrow} \Sigma_1,
\end{aligned}$$

where

$$\begin{aligned}
\Sigma_1 = E\Bigg(&\int_0^\tau Y(u)\begin{pmatrix}\mathbf{Z} \\ W\end{pmatrix}\int\Big[\dot{\lambda}\{\boldsymbol{\beta}_{10}^T Z + \beta_{20}x + H_0(u)\} \\
&- \lambda^2\{\boldsymbol{\beta}_{10}^T\mathbf{Z} + \beta_{20}x + H_0(u)\} \\
&+ J\{u|W, \mathbf{Z}, H_0(u), \beta_0, \boldsymbol{\theta}_0, f_U\}\lambda\{\boldsymbol{\beta}_{10}^T\mathbf{Z} + \beta_{20}x + H_0(u)\}\Big] \\
&\times\left\{\begin{pmatrix}\mathbf{Z} \\ x\end{pmatrix} + \boldsymbol{\gamma}_1(u)\right\}^T G(x|u, W_i, \mathbf{Z}_i, H_0, \boldsymbol{\beta}, \boldsymbol{\theta}_0, f_U) \\
&\times \mathrm{d}x\mathrm{d}H_0(u)\Bigg) + E\Bigg[\int_0^\tau Y(u)\begin{pmatrix}\mathbf{Z} \\ W\end{pmatrix} \\
&\times J\{u|W, \mathbf{Z}, H_0(u), \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, f_U\}\boldsymbol{\gamma}_2^{\mathrm{T}}(u)\mathrm{d}u\Bigg]. \quad \text{(A2)}
\end{aligned}$$

References

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* **83**, 1184–1186.

Chen, K., Jin, Z., and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659–668.

Cheng, S.-C. and Wang, N. (2001). Linear transformation models for failure time data with covariate measurement error. *Journal of the American Statistical Association* **96**, 706–716.

Fine, J. P., Ying, Z., and Wei, L. G. (1998). On the linear transformation model for censored data. *Biometrika* **85**, 980–986.

Gao, G. and Tsiatis, A. A. (2005). Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika* **92**, 875–891.

Hall, P. and Ma, Y. (2007). Measurement error models with unknown error structure. *Journal of Royal Statistical Society, Series B* **69**, 429–446.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R., Henry, W. K., Lederman, M, M., Phair, J. P., Niu, M., Hirsch, M. S., and Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic milimeter. *The New England Journal of Medicine* **335**, 1081–1090.

Hu, C. and Lin, D. Y. (2004). Semiparametric failure time regression with replicates of mismeasured covariates. *Journal of the American Statistical Association* **99**, 105–118.

Hu, P., Tsiatis, A. A. and Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **54**, 1407–1419.

Huang, Y. and Wang, C. Y. (2000). Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *Journal of the American Statistical Association* **95**, 1209–1219.

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis* **65**, 139–165.

Lu, W., and Zhang, H. H. (2010). On estimation of partially linear transformation models. *Journal of the American Statistical Association* **105**, 683–691.

Ma, Y. and Yin, G. (2008). Cure rate model with mismeasured covariates under transformation. *Journal of the American Statistical Association* **103**, 743–756.

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics* **48**, 829–838.

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

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B* **53**, 683–690.

Wen, C. C. and Chen Y-H. (2012). Conditional score approach to errors in-variable current status data under the proportional odds model. *Scandinavian Journal of Statistics* **39**, 635–644.

Zucker, D. M. (2005). A pseudo partial likelihood method for semiparametric survival regression with covariate errors. *Journal of the American Statistical Association* **100**, 1264–1277.

Received March 2013. *Revised September* 2013. *Accepted October* 2013.