Published in final edited form as:

Scand Stat Theory Appl. 2016 June ; 43(2): 558–572. doi:10.1111/sjos.12191.

Locally Efficient Semiparametric Estimators for Proportional Hazards Models with Measurement Error

Yuhang Xu,

Department of Statistics, Iowa State University

Yehua Li, and

Department of Statistics, Iowa State University

Xiao Song

Department of Epidemiology and Biostatistics, University of Georgia

Yehua Li: yehuali@gmail.com

Abstract

We propose a new class of semiparametric estimators for proportional hazards models in the presence of measurement error in the covariates, where the baseline hazard function, the hazard function for the censoring time, and the distribution of the true covariates are considered as unknown infinite dimensional parameters. We estimate the model components by solving estimating equations based on the semiparametric efficient scores under a sequence of restricted models where the logarithm of the hazard functions are approximated by reduced rank regression splines. The proposed estimators are locally efficient in the sense that the estimators are semiparametrically efficient if the distribution of the error-prone covariates is specified correctly, and are still consistent and asymptotically normal if the distribution is misspecified. Our simulation studies show that the proposed estimators have smaller biases and variances than competing methods. We further illustrate the new method with a real application in an HIV clinical trial.

Keywords

Cox model; errors-in-variables; semiparametric efficiency; spline; survival analysis

1. Introduction

The proportional hazards model or the Cox model (Cox, 1972) is the most widely used model in survival analysis. When all or a subset of the covariates are measured with error, it is well known that the naive method which substitutes the mismeasured values for the true covariates leads to biased estimation (Prentice, 1982). There are a large amount of methods in the literature on measurement error problems for the Cox model, including approximation

Address: Yehua Li, Department of Statistics, Iowa State University, Ames, IA 50011, USA

Supporting information: Additional information for this article is available online including proofs for the theorems and recursive formulas for computing B-splines.

methods such as regression calibration (Prentice, 1982; Wang *et al.*, 1997) and SIMEX (Li & Lin, 2003), likelihood-based methods (Hu *et al.*, 1998; Su & Wang, 2012), Bayesian methods (Cheng & Crainiceanu, 2009) and score estimating equation methods. Typical score methods include the conditional score method (Tsiatis & Davidian, 2001; Song *et al.*, 2002) and the corrected score method (Nakamura, 1992; Huang & Wang, 2000; Song & Huang, 2005). The readers are referred to Carroll *et al.* (2006) for a comprehensive account of these methods.

The method we propose in this paper is most closely related to score estimating equation methods. Unlike likelihood-based methods, score methods do not assume a parametric model for the distribution of the true covariate. However there are still some drawbacks for existing score methods. For example, the conditional score approach relies on the existence of a complete and sufficient statistic and therefore is not applicable to situations where such a statistic cannot be easily obtained. One simple example where a complete and sufficient statistic does not exist is the quadratic logistic regression model described in Tsiatis & Ma (2004). More importantly, semiparametric efficiency has not yet been established for any existing score estimating equation methods in survival analysis.

Tsiatis & Ma (2004) proposed locally efficient semiparametric estimators in the general setting of functional measurement error models. Their method is based on semiparametric efficient scores rather than complete and sufficient statistics and therefore is more general. Their method was first proposed for parametric regression models, where the response depends on a parametric function of the true covariates, and was further extended by Ma & Carroll (2006) to a class of semiparametric regression models, where the response variable Y depends on a parametric form of an error-prone covariate $\mathbf{X}$ and a nonparametric function of an error-free univariate covariate $\mathbf{Z}$. Ma & Carroll (2006) proposed a backfitting algorithm, where the nonparametric function was estimated by a local estimating equation. In particular, they applied their method to the generalized partially linear models as a special case.

Unlike the models considered in Ma & Carroll (2006), the likelihood of the Cox model depends on an unknown baseline hazard function, the hazard function for the censoring time, and their integrals, i.e., the corresponding cumulative hazard functions. Constructing local estimating equations for the Cox model is difficult especially when some of the covariates are measured with error. Our strategy is to apply the method of Tsiatis & Ma (2004) to a restricted model where the logarithms of the baseline hazard function and the censoring hazard function are both approximated by regression splines (Zhou *et al.*, 1998; Zhu *et al.*, 2008). By slowly increasing the rank of the spline approximation with the sample size, the efficient score under the restricted model becomes an estimating equation with diverging number of covariates (Wang, 2011). There has also been a large amount of literature on estimating the baseline hazard function using splines, including Kooperberg *et al.* (1995), and Huang (1996). However, none of them has dealt with efficient estimation in measurement error problems.

The rest of the paper is organized as follows. We introduce the model framework in Section 2.1 and propose a class of locally efficient estimators under a class of broadly defined sub-

models in Section 2.2. We discuss the asymptotic properties of the resulting estimators while letting the rank of the spline approximation increase with the sample size in Section 2.3 and provide a specific implementation strategy based on spline approximations in Section 2.4. We provide two simulation studies in Section 3 and a real data application in AIDS clinical trials in Section 4 to illustrate the proposed method. Some concluding remarks are provided in Section 5, and all technical proofs are presented in the online supporting material.

2 Locally efficient semiparametric estimation

2.1 Model specification and assumptions

We consider right censored failure time data collected from n independent subjects. For subject i , let T_i and C_i be the failure and censoring times, $\mathbf{Z}_i$ be a p_1 -dimensional error-free covariate and $\mathbf{X}_i$ be a p_2 -dimensional error-prone covariate. Let $V_i = \min(T_i, C_i)$ be the observed event time and $\delta_i = \mathbb{I}(T_i \leq C_i)$ be the failure indicator. We assume the censoring time is independent of the failure time and the covariates. However, this assumption can be relaxed (see Section 5). The hazard of the failure time T_i is related to $\mathbf{X}_i$ and $\mathbf{Z}_i$ through a proportional hazards regression model

$$\lambda_i(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T_i < t+dt | T_i \geq t, \mathbf{X}_i, \mathbf{Z}_i)}{dt} = \lambda(t) \exp\{g(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\theta}_1)\}, \quad (1)$$

where $\lambda(\cdot)$ is an unspecified baseline hazard function and $g(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\theta}_1)$ is a known function of $\mathbf{X}_i$ and $\mathbf{Z}_i$ up to a p -dimensional parameter $\boldsymbol{\theta}_1$. The most commonly used parametric structure for $g(\cdot)$ is the linear structure, i.e. $g(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\theta}_1) = \mathbf{X}_i^T \boldsymbol{\beta} + \mathbf{Z}_i^T \boldsymbol{\alpha}$ with $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ and $p = p_1 + p_2$. However, we also allow more general parametric structures for $g(\cdot)$, such as a quadratic regression model $g(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\theta}_1) = X_{i1}\beta_1 + X_{i1}^2\beta_2 + \mathbf{Z}_i^T \boldsymbol{\alpha}$ with $\boldsymbol{\theta}_1 = (\beta_1, \beta_2, \boldsymbol{\alpha}^T)^T$ and $p = p_1 + p_2 + 1$. Instead of observing $\mathbf{X}_i$, we only observe $\mathbf{W}_i$, a surrogate for $\mathbf{X}_i$. We assume that (V_i, δ_i) and $\mathbf{W}$ are conditionally independent given the covariates $(\mathbf{X}, \mathbf{Z})$ and the conditional density of $\mathbf{W}$ given $\mathbf{X}$ and $\mathbf{Z}$, $p(\mathbf{w} | \mathbf{x}, \mathbf{z})$, is known. For example, under a Gaussian classical measurement error model $\mathbf{W} = \mathbf{X} + \mathbf{U}$, where $\mathbf{U} \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}_U)$ is a measurement error independent of $\mathbf{X}$, the assumption requires the covariance matrix $\boldsymbol{\Sigma}_U$ to be known. This assumption can be later relaxed to allow for unknown parameters in $p(\mathbf{w} | \mathbf{x}, \mathbf{z})$, see remark 3 in Section 2.2.

Denote the observed data as $\mathbf{O} = (V_i, \delta_i, \mathbf{W}_i^T, \mathbf{Z}_i^T)^T$ and the data comprising the true covariate as $\mathbf{D} = (V_i, \delta_i, \mathbf{W}_i^T, \mathbf{X}_i^T, \mathbf{Z}_i^T)^T$. Following (1), the probability density of $\mathbf{D}$ is

$$p(\mathbf{d}) = [\lambda(v) \exp\{g(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_1)\}]^\delta \exp[-\Lambda(v) \exp\{g(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_1)\}] \times \{\lambda_c(v)\}^{1-\delta} \exp[-\Lambda_c(v)] p(\mathbf{w} | \mathbf{x}, \mathbf{z}) p(\mathbf{z}), \quad (2)$$

where $\Lambda(t) = \int_0^t \lambda(u) du$ is the baseline cumulative hazard function, and $\lambda_C(\cdot)$ and $\Lambda_C(\cdot)$ are the hazard and cumulative hazard functions for the censoring time. We collect the four infinite dimensional nuisance parameters in the model into $\eta = \{p(\mathbf{x} | \mathbf{z}), p(\mathbf{z}), \lambda(\cdot), \lambda_C(\cdot)\}$.

The model specified in (2) is a semiparametric model (Tsiatis, 2006) with a parametric component θ_1 and a nonparametric component η . Define the Hilbert space H_D of all L_2 measurable functions $h(\mathbf{D})$ of $\mathbf{D}$ with finite variance, and equip H_D with the covariance inner product, $\langle h_1(\mathbf{D}), h_2(\mathbf{D}) \rangle = E\{h_1(\mathbf{D})h_2(\mathbf{D})\}$ for any $h_1(\mathbf{D}), h_2(\mathbf{D}) \in H_D$. The nuisance tangent space, denoted as Λ_F is the linear subspace in H_D spanned by the nuisance scores of all parametric sub-models (Tsiatis & Ma, 2004). Two elements $h_1, h_2 \in H_D$ are orthogonal to each other, or $h_1 \perp h_2$, if $\langle h_1, h_2 \rangle = 0$. For any $h \in H_D$ and any subspace $A \subset H_D$, define the projection of h on A , denoted by $\Pi(h|A)$, to be the unique element $h_a \in A$ such that $h - h_a$ is orthogonal to all elements in A , or $(h - h_a) \perp A$. If $\mathbf{h} = (h_1, \dots, h_r)$ is a vector of random elements in H_D , define $\Pi(\mathbf{h}|A)$ to be the vector resulted from projecting each entry of $\mathbf{h}$ on A . Also define the orthogonal complement of a subspace A as $A^\perp = \{g \in H_D : g \perp A\}$.

In our problem, the nuisance tangent space can be further decomposed into four subspaces associated with the four components in η . Let $N_C(u) = I(V - u, = 0)$ be the censoring indicator by time u , $N(u) = I(V - u, = 1)$ be the observed failure indicator, and $Y(u) = I(V - u, > 0)$ be the at-risk indicator. Define martingale increments $dM_C(u) = dN_C(u) - \lambda_C(u)Y(u)du$ and $dM(u, \mathbf{X}, \mathbf{Z}) = dN(u) - \lambda(u) \exp(\mathbf{X}^T \boldsymbol{\beta} + \mathbf{Z}^T \boldsymbol{\alpha}) Y(u)du$. As derived in Tsiatis & Ma (2004), the nuisance tangent subspaces associated with $p(\mathbf{x} | \mathbf{z})$ and $p(\mathbf{z})$ are

$$\begin{aligned}\Lambda_{1D} &= [h(\mathbf{X}, \mathbf{Z}) \in H_D : E\{h(\mathbf{X}, \mathbf{Z}) | \mathbf{Z}\} = 0], \\ \Lambda_{2D} &= [h(\mathbf{Z}) \in H_D : E\{h(\mathbf{Z})\} = 0].\end{aligned}$$

Tsiatis (2006) shows that the nuisance tangent subspaces associated with $\lambda(\cdot)$ and $\lambda_C(\cdot)$ are

$$\begin{aligned}\Lambda_{3D} &= \{ \int a(u) dM(u, \mathbf{X}, \mathbf{Z}) \}, \\ \Lambda_{4D} &= \{ \int a(u) dM_C(u) \},\end{aligned}$$

where $a(u)$ is any integrable function of u . It is easy to verify that the four nuisance tangent subspaces above are orthogonal to each other under the covariance inner product, and hence Λ_D is a direct sum of the four subspaces, i.e. $\Lambda_D = \Lambda_{1D} \oplus \Lambda_{2D} \oplus \Lambda_{3D} \oplus \Lambda_{4D}$.

Similarly, we define the Hilbert space H of functions on the observed data. Using arguments similar to those in Tsiatis & Ma (2004), we can see that $H = E(H_D | \mathbf{O}) = \{E(h | \mathbf{O}) : h \in H_D\}$ and the nuisance tangent space based on the observed data is

$$\Lambda = E(\Lambda_D | \mathbf{O}) = \Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4,$$

where $\Lambda_i = E(\Lambda_{iD} | \mathbf{O})$, $i = 1, \dots, 4$. Since Λ_{2D} and Λ_{4D} only consist of functions of the observed data, it is easy to see that $\Lambda_2 = \Lambda_{2D}$ and $\Lambda_4 = \Lambda_{4D}$, and they are both orthogonal to Λ_1 and Λ_3 . However, Λ_1 and Λ_3 are no longer orthogonal to each other.

The observed-data score is a p -dimensional vector $E[\partial / \partial \theta_1 \{\log p(\mathbf{d})\} | \mathbf{O}]$. The semiparametric efficient score is the projection of the observed-data score onto $\Lambda^\perp$, the orthogonal complement of the nuisance tangent space (Tsiatis & Ma, 2004). In theory, the root of the efficient score is the most efficient estimator for θ_1 . However, except for a few special cases, the semiparametric efficient score usually involves the unknown nuisance parameter η and does not have a close form.

2.2 Locally efficient estimators under a restricted sub-model

To motivate our locally efficient semiparametric estimator, we consider a sub-model of (2) where both $\nu(t; \gamma_1) = \log\{\lambda(t)\}$ and $\nu_C(t; \gamma_2) = \log\{\lambda_C(t)\}$ are modeled parametrically. Suppose the parameters γ_1 and γ_2 are of dimensions K_1 and K_2 respectively, and we will refer to this sub-model as the restricted model.

Define $\theta_2 = (\gamma_1^T, \gamma_2^T)^T$, $\theta = (\theta_1^T, \theta_2^T)^T$, $K = K_1 + K_2$ and $q = p + K$. Then the restricted model is a semiparametric sub-model with a q -dimensional parametric component θ and a nonparametric component $\eta^R = \{p(\mathbf{x} | \mathbf{z}), p(\mathbf{z})\}$. Under this restricted model, the nuisance tangent spaces for $\mathbf{D}$ and $\mathbf{O}$, denoted by Λ_D^R and Λ^R , have simpler structures

$$\Lambda_D^R = \Lambda_{1D} \oplus \Lambda_{2D}, \quad \Lambda^R = \Lambda_1 \oplus \Lambda_2,$$

where Λ_{iD}^R and Λ_i^R , $i = 1, 2$, are defined as in Section 2.1.

In locally efficient semiparametric estimation, the underlying conditional density $p(\mathbf{x} | \mathbf{z})$ is allowed to be misspecified. Denote the possibly incorrect conditional density by $p^*(\mathbf{x} | \mathbf{z})$, expectations or conditional expectations calculated under the misspecified distribution by $E^*(\cdot)$, and expectations calculated under the true distribution by $E(\cdot)$. Define the Hilbert spaces, H_D^* and H^* , for $\mathbf{D}$ and $\mathbf{O}$ under the misspecification, where the inner product is defined on $E^*(\cdot)$. Similarly, $\Lambda_D^{R*} = \Lambda_{1D}^* \oplus \Lambda_{2D}^*$ and $\Lambda^{R*} = \Lambda_1^* \oplus \Lambda_2^*$ are the nuisance tangent spaces of the restricted model under $p^*(\mathbf{x} | \mathbf{z})$. Denote Π^* as the projection operator under $E^*(\cdot)$.

The score vector based on $\mathbf{D}$ under the restricted model is

$$\mathbf{S}_{D,\theta}(V, \Delta, \mathbf{X}, \mathbf{Z}) = \partial / \partial \theta \{\log p(V, \Delta | \mathbf{X}, \mathbf{Z}; \theta)\},$$

and the observed-data score vector under $p^*(\mathbf{x} | \mathbf{z})$ is

$$S_{\theta}^*(O) = E^*\{S_{D,\theta}(V, \Delta, \mathbf{X}, \mathbf{Z}) | O\} = \frac{\int S_{D,\theta}(V, \Delta, \mathbf{x}, \mathbf{Z}) p(V, \Delta | \mathbf{x}, \mathbf{Z}; \theta_0) p(\mathbf{W} | \mathbf{x}, \mathbf{Z}) p^*(\mathbf{x} | \mathbf{Z}) d\mu(\mathbf{x})}{\int p(V, \Delta | \mathbf{x}, \mathbf{Z}; \theta_0) p(\mathbf{W} | \mathbf{x}, \mathbf{Z}) p^*(\mathbf{x} | \mathbf{Z}) d\mu(\mathbf{x})},$$

(3)

where $d\mu(\cdot)$ denotes the dominating measure in the domain of $\mathbf{X}$. Based on the observed score (3), the efficient score for the restricted model under $p^*(\mathbf{x} | \mathbf{Z})$ is

$$S_{\text{eff}}^*(O, \theta) = S_{\theta}^*(O) - \Pi^*\{S_{\theta}^*(O) | \Lambda^{R*}\}. \quad (4)$$

The following Lemma follows directly from Theorem 1 of Tsiatis & Ma (2004) and provides clues to find $\Pi^*\{S_{\theta}^*(O) | \Lambda^{R*}\}$.

Lemma 1 Suppose the restricted model is true, and $p^*(\mathbf{x} | \mathbf{Z})$ and $p(\mathbf{x} | \mathbf{Z})$ have the same support, then almost surely, the space orthogonal to $\Lambda_1^{R*} \oplus \Lambda_2^{R*}$ is

$$[h(O) : E\{h(O) | \mathbf{X}, \mathbf{Z}\} = 0].$$

By the definition of Λ^{R*} , $\Pi^*\{S_{\theta}^*(O) | \Lambda^{R*}\}$ is of the form $E^*\{h(\mathbf{X}, \mathbf{Z}) | O\}$. By Lemma 1, almost surely, $h(\mathbf{X}, \mathbf{Z})$ needs to satisfy the integral equation

$$E^*[S_{\theta}^*(O) - E^*\{h(\mathbf{X}, \mathbf{Z}) | O\} | \mathbf{X}, \mathbf{Z}] = 0. \quad (5)$$

Details of how to solve the integral equation above can be found in Section 2.4.

Remark 1 Under the restricted model, the efficient score in (4) becomes

$$S_{\text{eff}}^*(O, \theta) = S_{\theta}^*(O) - E^*\{h(\mathbf{X}, \mathbf{Z}) | O\}, \quad (6)$$

where $h(\mathbf{X}, \mathbf{Z})$ is the solution of (5). Lemma 1 implies that $E\{S_{\text{eff}}^*(O, \theta) | \mathbf{X}, \mathbf{Z}\} = 0$, which in turn implies $E\{S_{\text{eff}}^*(O, \theta)\} = 0$ and $S_{\text{eff}}^*(O, \theta)$ as an element in H is orthogonal to $\Lambda_1 + \Lambda_2$. In other words, even if $p^*(\mathbf{x} | \mathbf{Z})$ is misspecified, the estimating equation based on the score function in (6) is still unbiased.

Based on the efficient score (6), we construct the estimating equation

$$\mathbf{S}_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \mathbf{S}_{\text{eff}}^*(\mathbf{O}_i, \boldsymbol{\theta}) = \mathbf{0}, \quad (7)$$

where $\mathbf{O}_i = (V_i, \mathbf{W}_i, \mathbf{Z}_i)$, $i = 1, \dots, n$. Denote the solution of (7) by $\hat{\boldsymbol{\theta}}_n$ which yields a partition $(\hat{\boldsymbol{\theta}}_{1n}^T, \hat{\boldsymbol{\gamma}}_{1n}^T, \hat{\boldsymbol{\gamma}}_{2n}^T)^T$ the same way as for $\boldsymbol{\theta}$. Then $\hat{\boldsymbol{\theta}}_{1n}$ contains the estimators of the Cox regression coefficients that are of main interests.

Remark 2 The possibly misspecified conditional density $p^*(\mathbf{x} | \mathbf{z})$ is often assumed to have a parametric form $p^*(\mathbf{x} | \mathbf{z}; \boldsymbol{\tau})$, where $\boldsymbol{\tau}$ is an unknown finite-dimensional parameter of which a root- n consistent estimator $\hat{\boldsymbol{\tau}}$ exists. Tsiatis & Ma (2004) shows that replacing $p^*(\mathbf{x} | \mathbf{z}; \boldsymbol{\tau})$ with $p^*(\mathbf{x} | \mathbf{z}; \hat{\boldsymbol{\tau}})$ does not cause any additional variation in the final estimator $\hat{\boldsymbol{\theta}}_n$.

Remark 3 So far, we have assumed $p(\mathbf{w} | \mathbf{x}, \mathbf{z})$ to be known. In reality, this distribution usually depends on an unknown measurement error model parameter $\boldsymbol{\gamma}_{\text{mem}}$. For example, under the classical measurement error model $\mathbf{W} = \mathbf{X} + \mathbf{U}$, $\boldsymbol{\gamma}_{\text{mem}} = \boldsymbol{\Sigma}_{\mathbf{U}}$ is the covariance matrix of the measurement error. One can estimate this parameter through replicates of $\mathbf{W}$ and plug the estimate into the estimating equation (7). Replicates of the surrogate exist in many applications, including the AIDS clinical trial data in Section 4. However, as pointed out in Ma & Carroll (2006), though this plug-in method is robust, it may cause loss of efficiency. A more efficient method is to construct an additional estimating equation for $\boldsymbol{\gamma}_{\text{mem}}$ and solve this equation with (7) simultaneously. See Section 3.6 of Ma & Carroll (2006) for the construction of this additional equation.

2.3 Locally efficient estimators based on regression splines

We now provide a more specific strategy to parameterize $\nu(t)$ and $\nu_C(t)$, where we will use regression splines (Zhou *et al.*, 1998; Zhu *et al.*, 2008). For any interval $[a, b]$, let $a = \kappa_{1-r} = \dots = \kappa_0 < \kappa_1 < \dots < \kappa_{J+1} = \dots = \kappa_{J+r} = b$ be a sequence of knots, normalized B-spline functions (Schumaker, 1981; Zhou *et al.*, 1998) of order r are defined as

$$B_j(t) = (\kappa_j - \kappa_{j-r})[\kappa_{j-r}, \dots, \kappa_j](\kappa - t)_+^{r-1}, \quad j=1, \dots, K,$$

where $K = J+r$, $[\kappa_{j-r}, \dots, \kappa_j]g(\kappa)$ is the r th order divided difference of the function $g(\kappa)$ on $\kappa_{j-r}, \dots, \kappa_j$ and $(x)_+ = \max(x, 0)$. We collect B-spline basis functions in a K -dimensional vector as $\mathbf{B}(t) = \{B_1(t), \dots, B_K(t)\}^T$. The B-spline functions can also be evaluated using the recursive formula provided in Appendix D in the online supporting material.

Suppose the study is conducted within a compact time interval $\mathcal{T}$ so that any subject, who has not failed or dropped out by the end of $\mathcal{T}$, is automatically right censored. Define the class of Holder continuous functions on $\mathcal{T}$ as

$$C^{r,a}(\mathcal{T}) = \{f: \sup_{t_1, t_2 \in \mathcal{T}} |f^{(r)}(t_1) - f^{(r)}(t_2)| / |t_1 - t_2|^a < \infty\} \quad (8)$$

for some nonnegative integer r and some $a > 0$, where $f^{(r)}$ is the r th derivative of f and a is called the Hölder exponent. Hölder continuity is a strong form of continuity for functions. If $f \in C^{r,a}(\mathcal{T})$, not only is $f^{(r)}$ continuous over $\mathcal{T}$, but $f^{(r)}(t_1) \rightarrow f^{(r)}(t_2)$ no slower than the rate of $|t_1 - t_2|^a$ as $t_1 \rightarrow t_2$. Define the L_∞ norm of a function f on $\mathcal{T}$ as $\|f\|_\infty = \sup_{t \in \mathcal{T}} |f(t)|$.

Recall that $\nu(t) = \log\{\lambda(t)\}$ and $\nu_C(t) = \log\{\lambda_C(t)\}$ are the log baseline hazard functions for the failure time and censoring time respectively, we model $\nu(t)$ and $\nu_C(t)$ as splines to guarantee that estimators of $\lambda(t)$ and $\lambda_C(t)$ are non-negative. We assume that the true log hazard functions are $\nu_0 \in C_1^{r_1, a_1}$ and $\nu_{C,0} \in C_2^{r_2, a_2}(\mathcal{T})$ for some $r_1, r_2, a_1, a_2 > 0$. Let $\mathbf{B}_1(t)$ and $\mathbf{B}_2(t)$ be K_1 - and K_2 -dimensional vectors of normalized B-spline basis functions of orders r_1 and r_2 with interior knots equally placed in $\mathcal{T}$. We allow K_1 and K_2 to increase slowly with the sample size, and define $\tilde{\mathbf{B}}_1(t) = K_1^{-1/2} \mathbf{B}_1(t)$ and $\tilde{\mathbf{B}}_2(t) = K_2^{-1/2} \mathbf{B}_2(t)$ for the convenience of asymptotic analysis. Under the Hölder continuity assumption above, there exist the best spline approximations to the true log hazard functions denoted by

$$\nu^*(t) = \tilde{\mathbf{B}}_1^T(t) \gamma_1^* \text{ and } \nu_C^*(t) = \tilde{\mathbf{B}}_2^T(t) \gamma_2^* \text{ for some spline coefficient vectors } \gamma_1^* \text{ and } \gamma_2^* \text{ such that } \|\nu^* - \nu_0\|_\infty = O(K_1^{-r_1}) \text{ and } \|\nu_C^* - \nu_{C,0}\|_\infty = O(K_2^{-r_2}) \text{ (Schumaker, 1981).}$$

Now we turn to the restricted model with ν and ν_C parameterized as $\nu(t) = \tilde{\mathbf{B}}_1^T(t) \gamma_1$ and $\nu_C(t) = \tilde{\mathbf{B}}_2^T(t) \gamma_2$. The “true” parameters under this model are collected in $\theta^* = (\theta_{10}^T, \theta_2^{*T})^T$, where θ_{10} is the true value of θ_1 and $\theta_2^* = (\gamma_1^{*T}, \gamma_2^{*T})^T$ contains the coefficients of best possible spline approximations. When ν and ν_C are nonparametric and belong to the Hölder class defined above, the estimation equations in (7) are slightly biased with

$$E\{\mathbf{S}_n(\theta^*)\} = \tilde{O}(\max_{i=1,2} K_i^{-r_i}), \quad (9)$$

where $\tilde{O}(\cdot)$ denotes the order uniformly for all entries of a vector or a matrix.

To develop the asymptotic theory for the proposed estimator bases on regression spline approximation, we first introduce some notations and conditions. We use $\|\cdot\|$ to denote the L^2 norm of either a vector or a function: $\|\mathbf{a}\| = (\mathbf{a}^T \mathbf{a})^{1/2}$ for a vector $\mathbf{a}$ and $\|f\| = \{\int_{\mathcal{T}} f^2(t) dt\}^{1/2}$ for a function f in the Hölder class. Denote the parameter in (2) with nonparametric loghazard functions by $\Theta = (\theta_1, \nu, \nu_C)$, which is an element in a Banach space with the norm $\|\Theta\| = (\|\theta_1\|^2 + \|\nu\|^2 + \|\nu_C\|^2)^{1/2}$. Let $\mathcal{L}(\Theta) = \log\{p(\mathbf{D} | \Theta)\}$ be log-likelihood of data $\mathbf{D}$, define its Gâteaux derivative along the direction $\Theta^\dagger = (\theta_1^\dagger, \nu^\dagger, \nu_C^\dagger)$ as

$$d\ell(\Theta; \Theta^\dagger) = \lim_{t \rightarrow 0} \{\ell(\Theta + t\Theta^\dagger) - \ell(\Theta)\} / t.$$

Denote the projection into the subspace orthogonal to Λ^{R*} by $\Pi_\perp^*(\cdot)$ and define

$$\mathbb{S}^*(\Theta; \Theta^\dagger) = \Pi_\perp^* [E^* \{dl(\Theta; \Theta^\dagger) | \mathbf{O}\}]. \quad (10)$$

Let $\mathcal{S}(\Theta; \Theta^\dagger)$ be a special case of $\mathcal{S}^*(\Theta; \Theta^\dagger)$ when $p(\mathbf{x} | \mathbf{z})$ is correctly specified.

We will make the following assumptions:

Assumptions—(1) Denote the true parameter by $\Theta_0 = (\theta_{10}, v_0, v_{C,0})$ and assume that there exist some positive constants C_0, C_1 and C_2 such that

$$C_1 \leq E\{\mathbb{S}^*(\Theta; \Theta^\dagger)\mathbb{S}(\Theta; \Theta^\dagger)\} / \|\Theta^\dagger\|^2 \leq C_2,$$

for all $\Theta^\dagger$ and Θ with $\|\Theta - \Theta_0\| \leq C_0$.

(2) Assume $K_i = O(n^{l_i})$ for $i = 1, 2$, with $\max_{i=1,2}(l_i) < 1/3$ and $\min_{i=1,2}(l_i r_i) < 1/2$.

Remark 4 Assumption (1) imposes a nondegeneracy and boundedness condition for the semiparametric efficient score (Tsiatis & Ma, 2004; Shen, 1997). Assumption (2) specifies the rate of growth of the dimension of the spline spaces relative to the sample size. It means that if higher order B-spline basis functions are used, fewer functions are required. It also implies that $r_i \geq 2$, $i = 1, 2$, so linear or higher order spline functions should be used.

Under these assumptions, we present the theoretical properties of the proposed estimators based on spline approximations. Proofs of the following theorems are provided in the online supporting material.

Theorem 1 The estimating equations in (7) yield a sequence of consistent solutions with $\|\hat{\theta}_n - \theta^*\| = O_p(\delta_n)$, where $\delta_n = (K/n)^{1/2} + \max_{i=1,2} K_i^{-r_i+1/2}$.

Theorem 2 Suppose Γ and Σ defined in (A.7) in the online supporting material exist, $n^{1/2}(\hat{\theta}_{1n} - \theta_{10}) \rightarrow \text{Normal}(0, \Gamma^{-1} \Sigma \Gamma^{-1})$ in distribution. If the conjectured model $p^*(\mathbf{x} | \mathbf{z})$ is specified correctly, the asymptotic covariance matrix can be further simplified as Σ^{-1} .

Theorem 3 The proposed semiparametric estimator $\hat{\theta}_n$ is locally efficient. That is, the estimator is semiparametrically efficient if $p^*(\mathbf{x} | \mathbf{z})$ is specified correctly, and is still consistent if $p^*(\mathbf{x} | \mathbf{z})$ is specified incorrectly.

Theorem 1 suggests that there exist a sequence of solutions to the estimating equation that converge to the true parameter. However, as in all estimating equation literature, the uniqueness of the solution is hard to show and often not guaranteed. The result in Theorem 1 does not guarantee uniqueness of the solution to our estimating equation either in finite sample or in large sample limit. Carroll *et al.* (2006) recommend to solve the estimating equation with multiple initial values and, if distinct solutions are found, choose the one that

is the closest to the naive estimator. In our simulation studies, we have never encountered multiple root problems with our estimator.

Recall that the parameter vector θ contains both the parametric component θ_1 and the spline coefficients θ_2 . For both theoretical derivation (see our proof of Theorem 2) and practical calculation, we need to calculate the asymptotic covariance of $\hat{\theta}_n$. As the dimension of spline basis functions increases, this covariance matrix also has an increasing dimension. In Theorem 2, we focus on the asymptotic distribution of the parametric part, the dimension of which is finite, and thus can establish a root- n convergence rate. The asymptotic covariance of $\hat{\theta}_{1n}$ is the upper left sub-matrix of a much bigger covariance matrix with an increasing dimension. In calculating this sub-matrix, we define Σ and Γ as limits of various matrices. They have complicated forms that are hard to use directly. In practice, we estimate the covariance $\hat{\theta}_n$ directly using a sandwich formula (see Section 2.4 for more details), and the covariance of $\hat{\theta}_{1n}$ is extracted as a sub-matrix.

2.4 Implementation details

For given K_i and r_i , $i = 1, 2$, we use the following algorithm to solve $\hat{\theta}_n$ from the estimating equation (7) and obtain the estimate of its covariance matrix.

Algorithm—(1) Choose an appropriate initial value for $\theta = (\theta_1^T, \theta_2^T)^T$. Specifically, an initial value for θ_1 is obtained using the conditional score method (Tsiatis & Davidian, 2001; Song *et al.*, 2002). The initial value for $\theta_2 = (\gamma_1^T, \gamma_2^T)^T$ is obtained through a naive maximum likelihood method (Kooperberg *et al.*, 1995) where we model $v(t)$ and $v_C(t)$ as splines and replace X_i with W_i (or W_i if replicates of W_i are available).

(2) Solve $\hat{\theta}_n$ from (7) using a derivative-free optimization algorithm. For example, we use the function *lsqnonlin* in *MATLAB*. In order to solve (7), we first need to solve $h(\mathbf{X}, \mathbf{Z})$ from the integral equation (5), where we use a simple approximation technique proposed by Tsiatis & Ma (2004). We discretize the value of $\mathbf{X}$ on m points, $\mathbf{x}_1, \dots, \mathbf{x}_m$, spread across the support of $\mathbf{X}$, so that the function $h(\mathbf{X}, \mathbf{Z})$ becomes a q by m matrix $h(\mathbf{Z}) = \{h(\mathbf{x}_1, \mathbf{Z}), \dots, h(\mathbf{x}_m, \mathbf{Z})\}$ for any $\mathbf{Z}$. Then, the integral equation (5) is reduced to linear equations $\mathbf{A}(\mathbf{Z})\mathbf{h}^T(\mathbf{Z}) = \mathbf{b}^T(\mathbf{Z})$, where $\mathbf{A}(\mathbf{Z})$ and $\mathbf{b}(\mathbf{Z})$ are $m \times m$ and $q \times m$ matrices defined similarly as in equation (20) in Tsiatis & Ma (2004) but adapted to fit our setting.

(3) We estimate the covariance matrix of $\hat{\theta}_n$ by a sandwich formula similar to equation (18) in Tsiatis & Ma (2004)

$$\widehat{\text{cov}}(\hat{\theta}_n) = \{\mathbf{J}_n(\hat{\theta}_n)\}^{-1} \mathbf{G}_n(\hat{\theta}_n) \{\mathbf{J}_n^T(\hat{\theta}_n)\}^{-1},$$

where $\mathbf{J}_n(\theta) = -\partial \theta / \partial S_n(\theta)$ and $\mathbf{G}_n(\theta) = n^{-1} \sum_{i=1}^n \mathbf{S}_{\text{eff}}^*(O_i, \theta) \mathbf{S}_{\text{eff}}^{*T}(O_i, \theta)$. The derivative $\mathbf{J}_n(\theta)$ does not have an explicit form and is computed using numerical differentiation. The covariance of $\hat{\theta}_{1n}$ is obtained as the corresponding sub-matrix of $\widehat{\text{cov}}(\hat{\theta}_n)$.

In reality, the choice of spline basis functions, including the orders, numbers of knots and knots placement, is part of the model tuning. The optimal orders r_1 and r_2 depend on the smoothness of $\nu(t)$ and $\nu_C(t)$, but in practice the results are least sensitive to these choices and cubic splines (twice continuously differentiable, piecewise cubic polynomial) are often employed (Kooperberg *et al.*, 1995). Suppose the study is conducted within a compact time interval $\mathcal{T} = [t_{\min}, t_{\max}]$, we place the interior knots with equal distance in $\mathcal{T}$. This choice is equivalent to using a global bandwidth in kernel regression and works well in our simulation studies and data analysis. Placing the knots adaptively is possible (Stone *et al.*, 1997) but computationally much more intensive. With the orders and knots placement pre-determined, the performance of the method is mainly governed by the numbers of spline basis K_1 and K_2 . We choose these tuning parameters by minimizing the following Akaike information criteria (AIC)

$$\text{AIC}(K_1, K_2) = -2\ell_O^*(\hat{\theta}_n) + 2(p + K_1 + K_2),$$

where $\ell_O^*(\cdot)$ denotes the log-likelihood for the observed data by integrating out $p^*(\mathbf{x} | \mathbf{z})$ and $\hat{\theta}_n$ is the estimator in the restricted model. Under the assumption that the censoring time is independent of the failure time and the covariates, it can be shown that $\ell_O^*(\theta)$ can be separated into two parts, $\ell_O^*(\theta) = \ell_{O,1}^*(\theta_1, \gamma_1) + \ell_{O,2}^*(\gamma_2)$, so the AIC can also be separated into two parts. Thus, K_1 and K_2 can be chosen separately. AIC works well in our numerical studies although other information criteria such as BIC (Kooperberg *et al.*, 1995) can also be used.

3 Simulation studies

3.1 Simulation 1: a linear Cox model

In the first simulation study, we considered a Cox model where the log hazard is linearly related to a covariate X , i.e. $\lambda(t | X) = \lambda_0(t) \exp(\beta X)$. The covariate X was generated from Normal (μ_X, σ_X^2) with $\mu_X = 1$ and $\sigma_X = 1$. Instead of observing X , we observed two independent copies of the surrogate $W^{[j]} = X + U^{[j]}$, $j = 1, 2$, where $U^{[j]}$ was generated independently from Normal $(0, \sigma_U^2)$ with $\sigma_U = 1$. We considered the baseline hazard functions $\lambda_0(t) = 1$, corresponding to an exponential distribution, and $\lambda_0(t) = t$, corresponding to a Weibull distribution. The true value of β was taken to be one. We considered both independent and dependent censoring mechanisms. Let b be a general constant which was used to achieve censoring rates of 30% and 60% in different scenarios. Under independent censoring, C was generated independently from a uniform distribution on $[0, b]$. Under the dependent censoring mechanism, C and Z are dependent, which is a departure from our assumption. Specifically, if $X > \mu_X$, C follows a uniform distribution on $[0, b]$; otherwise, C follows a uniform distribution on $[0, 2b]$. We set the sample size to be $n = 500$ and repeated the simulation 1,000 times.

We used cubic splines to approximate the logarithms of $\lambda_0(t)$ and $\lambda_0^C(t)$. The numbers of spline basis functions K_1 and K_2 were selected by AIC as described in Section 2.4. The

proposed estimator involves solving the integral equation (5), which requires specification of the density $p^*(x)$. We considered two choices of $p^*(x)$: a normal density with mean μ_X and variance σ_X^2 which corresponded to the truth and a uniform density on $[\mu_X - 3\sigma_X, \mu_X + 3\sigma_X]$ to assess the effect of misspecification. We used the method-of-moment estimators to estimate μ_X , σ_U^2 , and σ_X^2 , that is to say, $\hat{\mu}_X = \sum_{i=1}^n \bar{W}_i / n$, $\hat{\sigma}_U^2 = 2S_{\bar{V}}^2$, and $\hat{\sigma}_X^2 = 2S_{\bar{W}}^2 - \hat{\sigma}_U^2 / 2$, where $\bar{W}_i = (W_i^{[1]} + W_i^{[2]}) / 2$, $\bar{V}_i = (W_i^{[1]} - W_i^{[2]}) / 2$, and $S_{\bar{W}}^2$ and $S_{\bar{V}}^2$ are sample variances for $\bar{\mathbf{W}} = (W_1, \dots, W_n)^T$ and $\bar{\mathbf{V}} = (V_1, \dots, V_n)^T$ respectively. To solve equation (5), we adopted a simple discrete approximation for $p^*(x)$ following Tsiatis & Ma (2004). Specifically, we took X to be discrete with masses on 12 equally spaced points in $[\mu_X - 3\sigma_X, \mu_X + 3\sigma_X]$ with probabilities proportional to the density $p^*(x)$. All integrals involved were calculated by Gaussian quadratures using 100 nodes.

For comparison, we also considered various existing methods. The naive estimator maximizes the partial likelihood using W as the covariate. Two approximation methods were considered: the regression calibration (RC) estimator (Chapter 4 of Carroll *et al.*, 2006), where X is replaced by $E(X|W)$ in the partial likelihood, and the simulation and extrapolation (SIMEX) estimator (Chapter 5 of Carroll *et al.*, 2006) with a quadratic extrapolation function. A joint Gaussian distribution was assumed for X and W in regression calibration. The nonparametric correction (NPC) estimator of Huang & Wang (2000) and the conditional score (CS) estimator of Tsiatis & Davidian (2001) were also included for comparison.

The results under independent censoring are summarized in Table 1, where we report the bias, the empirical standard deviation, the mean estimated standard error and the coverage rate of a 95% confidence interval for various estimators of β . As we can see, the naive estimator is severely biased; both approximation methods, i.e. RC and SIMEX, are still considerably biased; the NPC and CS estimators are consistent estimators and yield smaller biases than the approximation methods. In all cases shown in the table, our proposed estimators yield the smallest biases among all and have much smaller standard deviations than other consistent estimators (i.e. the NPC and CS) which illustrates the efficiency of our methods. For example, our methods are 77% to 89% more efficient than the CS estimator. The NPC method occasionally yields strange solutions and these outliers inflate the empirical bias and standard deviation. To be generous, we also report a cleaned version of NPC (labeled as NPC* in the table), where we remove 0.5% to 2% of outliers in different cases. Such failure to converge has never occurred to our estimators and even after removing these outliers the NPC still has much larger standard deviations than ours. In addition, both the NPC and CS methods significantly underestimate the standard errors, and our further simulations reveal that this underestimation of standard error can be severe when either the measurement error is large or the censoring rate is high. In contrast, the estimated standard errors of our methods do not show obvious underestimation. One striking finding of particular interest is that misspecification of $p(x)$ in our methods does not cause obvious loss of efficiency.

The results under dependent censoring are reported in Table 2. These results tell a rather similar story as Table 1. Our proposed method still vastly outperforms the competing

methods, which also suggests that the proposed estimator is robust to mild violation of the independent censoring assumption.

3.2 Simulation 2: a quadratic Cox regression model

To assess the performance of the proposed estimator when $g(\cdot)$ in (1) is nonlinear, we considered the model $\lambda(t|X) = \lambda_0(t) \exp(\beta_1 X + \beta_2 X^2)$, where $\beta_1 = \beta_2 = -1$. We generated X from $\text{Normal}(-1, 1)$ and $W = X + U$ where $U \sim \text{Normal}(0, \sigma_U^2)$ with $\sigma_U = 1/5$. The rest of the simulation setting is similar to that in Simulation 1.

The NPC estimator was proposed for the case where $g(\cdot)$ is linear and therefore was not included for comparison. Since no complete and sufficient statistic exists under this model, the CS approach cannot be applied directly. Instead, we considered the approximated conditional score method of Song *et al.* (2002), which is based on a linear approximation using the delta method. The results for $\lambda_0(t) = t$ under independent censoring are summarized in Table 3. The Naive estimator is serenely biased. The proposed estimator is robust against misspecification of $p(x)$ and has smaller bias than RC and approximated CS. The SIMEX estimator works surprisingly well under this particular setting and is comparable to our method.

4 Analysis of AIDS clinical trial data

We applied the proposed method to data from AIDS Clinical Trials Group (ACTG) 175 (Hammer *et al.*, 1996) to assess the effects of antiretroviral therapies and baseline CD4 count on the time to AIDS or death in antiretroviral-naïve patients. Four therapies were investigated in the ACTG 175 clinical trial, and previous studies (Hammer *et al.*, 1996; Huang & Wang, 2000) found that the therapy using zidovudine alone is inferior to the other three therapies while the effects of the other three therapies are rather similar. Following Song & Huang (2005), we considered two treatment groups in our analysis, zidovudine alone and the combination of the other three therapies. It is well known that CD4 measurements were subject to substantial measurement errors. Among the 1067 antiretroviral-naïve patients in the study, 1036 had two CD4 measurements within 3 weeks of randomization and 31 had only one CD4 measurement. The censoring rate for the time to AIDS or death was 91%.

Following the literature, we used the log transformed CD4 count as a covariate in the proportional hazards model. It is well-known that the CD4 count measures were subject to a significant amount of measurement error. We used the graphical tools described in Carroll *et al.* (2006) to check various assumptions on the measurement error. In the left panel of Figure 1, we show the normal Q-Q plot of the differences between replicates of $\log(\text{CD4})$ within the same subject. The plot indicates that the measurement error exhibits slightly heavy tails on both sides, which is a mild deviation from the Gaussian assumption. In the right panel of Figure 1, we also plot the standard deviation of $\log(\text{CD4})$ within a subject against the mean to check on the constant variance assumption. The regression line in this plot was fitted using the robust regression function *rlm* in the MASS package in R. The estimated slope is -0.0184 with a *p*-value of 0.067, and therefore there is no clear violation of the assumption

that the variance of measurement error is a constant. The estimated standard deviation is 0.182 for the measurement error and 0.276 for the true underlying $\log(\text{CD4})$.

We considered a proportional hazards model with two covariates, the error-prone $\log(\text{CD4})$ and an error-free treatment indicator which is zero for the zidovudine alone therapy and one for the other therapies. For the proposed locally efficient estimator, we considered two distributions for $\log(\text{CD4})$, a normal distribution and a uniform distribution. In Table 4, we compare the estimation results based on various estimators considered in the simulation studies. In addition to the parameter estimates, we provide two versions of standard errors estimated by either a sandwich formula or a bootstrap procedure with 1,000 bootstrap samples. As suggested by our simulation results, the sandwich formula can sometimes underestimate the true standard deviation, especially for the NPC and CS estimators. The bootstrap standard error, even though is computationally more expensive, is considered to be more reliable for a fair comparison among the different methods. Both the NPC and CS estimators are based on estimating equations, and the consistency of bootstrap for these estimators follows from the general theory by Chatterjee & Bose (2005) for bootstrapping estimating equations. Our estimator is based on a semiparametric estimating equation as we allow the number of splines and dimension of the estimating equation to diverge to infinity. We expect the consistency of bootstrap for our estimator to follow from similar arguments as in Chatterjee & Bose (2005), but a rigorous justification is out of the scope of this paper.

As we can see from Table 4, the treatment effect is not affected much by measurement error in CD4, since the naive estimate of the treatment effect is rather similar to the CS and the proposed estimators. The RC, SIMEX and NPC methods, however, seem to have overdone with bias-correction for the treatment effect. The naive estimator for the coefficient of $\log(\text{CD4})$, on the other hand, is significantly attenuated compared with other estimators. All methods that take into account the measurement error provide similar estimates for $\log(\text{CD4})$ except for the SIMEX which still shows a small degree of attenuation. The proposed estimator under either Gaussian or uniform assumption for $\log(\text{CD4})$ has a smaller bootstrap standard error than other consistent estimators (i.e. the CS and NPC), indicating some efficiency gain using our method.

5 Discussion

We propose a class of locally efficient semiparametric estimators for the proportional hazards models with covariates contaminated with measurement error. Compared with competing methods, the proposed estimator is robust against misspecification of the distribution of the true covariate and is semiparametrically efficient if this underlying distribution is correctly specified. We allow the effect of the error-prone variable on the failure time to have a very general parametric form, under which a sufficient statistic for the true covariate usually does not exist. The likelihood function of the proportion hazards model involves the integral of the baseline hazard function, which makes a local estimating equation approach like the one proposed by Ma & Carroll (2006) difficult to implement. We circumvent this difficulty using spline approximations. Our numerical studies show that our method vastly outperforms competing methods.

The efficiency loss for the proposed estimators under misspecified $p^*(\mathbf{x} | \mathbf{z})$ is unknown. However, the simulation results in this paper, as well as those in Tsiatis & Ma (2004) and Ma & Carroll (2006), show that the loss is virtually negligible even if the misspecification of $p^*(\mathbf{x} | \mathbf{z})$ is severe. In practice, it is still advisable to propose a proper specification of $p^*(\mathbf{x} | \mathbf{z})$ using some graphical tools. When the measurement error is very large, a nonparametric estimate of $p(\mathbf{x} | \mathbf{z})$ using deconvolution methods can also be employed (Delaigle *et al.*, 2008).

Our method is developed under the independent censoring assumption, but can be extended easily to situations where censoring time also depends on the covariates. In those cases, another proportional hazards model can be used to characterize the relationship between the censoring time and the covariates, $\lambda_{iC}(t) = \lambda_C(t) \exp\{g_C(\mathbf{X}_i, \mathbf{Z}_i, \theta_{1,C})\}$, where $g_C(\cdot)$ is a known function of $\mathbf{X}_i$ and $\mathbf{Z}_i$ with a parameter $\theta_{1,C}$. We can redefine the parameter of interest to be a combination of θ_1 and $\theta_{1,C}$. Then a semiparametric estimator can be constructed similarly and the locally efficient property of the estimator still holds. Under this setting, we can estimate not only the relationship between the failure time and the covariates but also the censoring mechanism.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research of Li and Xu was partially supported by the US National Science Foundation (DMS-1314118). The research of Song was partially supported by US National Science Foundation (DMS-1106816) and US National Institute of Health (ES17030, HL121347). The authors thank the associate editor and an anonymous referee for their constructive comments which lead to significant improvement in the paper.

References

- Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, CM. Measurement error in nonlinear models: a modern perspective. 2nd. Chapman and Hall CRC Press; London: 2006.
- Chatterjee S, Bose A. Generalized bootstrap for estimating equations. *Ann Statist.* 2005; 33:414–436.
- Cheng YJ, Crainiceanu CM. Cox models with smooth functional effects of covariates measured with error. *J Am Statist Assoc.* 2009; 104:1144–1154.
- Cox DR. Regression models and life tables (with discussion). *J R Statist Soc B.* 1972; 34:187–220.
- Delaigle A, Hall P, Meister A. On deconvolution with repeated measurements. *Ann Statist.* 2008; 36:665–685.
- Hammer SM, Katzstein DA, Hughes MD, Gundaker H, Schooley RT, Haubrich RH, Henry WK, Lederman MM, Phair JP, Niu M, Hirsch MS, Merigan TC. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *The New England Journal of Medicine.* 1996; 335:1081–1090. [PubMed: 8813038]
- Horn, R.; Johnson, CR. Matrix analysis. Cambridge University Press; New York: 1985.
- Hu P, Tsiatis AA, Davidian M. Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics.* 1998; 54:1407–1419. [PubMed: 9883541]
- Huang J. Efficient estimation for the proportional hazards model with interval censoring. *Ann Statist.* 1996; 24:540–568.
- Huang Y, Wang CY. Cox regression with accurate covariates unascertainable: a nonparametric correction approach. *J Am Statist Assoc.* 2000; 95:1209–1219.

- Kooperberg C, Stone CJ, Truong YK. Hazard regression. *J Am Statist Assoc.* 1995; 90:78–94.
- Li Y, Lin XH. Functional inference in frailty measurement error models for clustered survival data using the SIMEX approach. *J Am Statist Assoc.* 2003; 98:191–203.
- Ma Y, Carroll RJ. Locally efficient estimators for semiparametric models with measurement error. *J Am Statist Assoc.* 2006; 101:1465–1474.
- Nakamura T. Proportional hazards model with covariates subject to measurement error. *Biometrics.* 1992; 48:829–838. [PubMed: 1420844]
- Ortega, JM.; Rheinboldt, WC. Iterative solution of nonlinear equations in several variables. Academic Press; San Diego: 1970.
- Prentice R. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika.* 1982; 69:331–342.
- Schumaker, LL. Spline functions. Wiley; New York: 1981.
- Shen X. On methods of sieves and penalization. *Ann Statist.* 1997; 25:2555–2591.
- Song X, Davidian M, Tsiatia AA. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics.* 2002; 3:511–528. [PubMed: 12933595]
- Song X, Huang Y. On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics.* 2005; 61:702–714. [PubMed: 16135021]
- Stone CJ, Hansen M, Kooperberg C, Truong YK. Polynomial splines and their tensor products in extended linear modeling (with discussion). *Ann Statist.* 1997; 25:1371–1470.
- Su YR, Wang JL. Modeling left-truncated and right-censored survival data with longitudinal covariates. *Ann Statist.* 2012; 40:1465–1488.
- Tsiatis, AA. Semiparametric theory and missing data. New York: Springer-Verlag; 2006.
- Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika.* 2001; 88:447–458.
- Tsiatis AA, Ma Y. Locally efficient semiparametric estimators for functional measurement error models. *Biometrika.* 2004; 91:835–848.
- Wang CY, Hsu L, Feng ZD, Prentice RL. Regression calibration in failure time regression. *Biometrics.* 1997; 53:131–145. [PubMed: 9147589]
- Wang L. GEE analysis of clustered binary data with diverging number of covariates. *Ann Statist.* 2011; 39:389–417.
- Zhou S, Shen X, Wolfe DA. Local asymptotics for regression splines and confidence regions. *Ann Statist.* 1998; 26:1760–1782.
- Zhu Z, Fung WK, He X. On the asymptotics of marginal regression splines with longitudinal data. *Biometrika.* 2008; 95:907–917.

Figure 1. Model checking for the AIDS clinical trial data. The left panel is the normal QQ plot for the differences between replicates of $\log(\text{CD4})$; the right panel is the scatter plot of the standard deviation of $\log(\text{CD4})$ within a subject versus the mean $\log(\text{CD4})$ within the subject.

Table 1

Estimation results of β under Simulation 1 and independent censoring. The results are based on 1,000 replications.

Censoring	Estimator	Exponential				Weibull			
		Bias	ESD	ESE	Cov	Bias	ESD	ESE	Cov
30%	Naive	-428	51	49	0	-434	53	50	0
	RC	-139	88	74	518	-148	89	75	475
	SIMEX	-186	84	60	213	-192	87	61	183
	NPC	67	271	168	913	74	277	210	908
	NPC*	48	200	166	924	56	216	209	919
	CS	28	176	151	924	29	183	159	926
60%	Semi mis	12	128	121	932	14	140	127	929
	Semi true	11	128	119	930	16	134	128	939
	Naive	-396	65	63	0	-405	67	64	0
	RC	-90	108	95	786	-103	108	97	757
	SIMEX	-154	103	77	468	-162	106	78	450
	NPC	72	248	172	897	105	439	186	900
	NPC*	59	211	171	906	66	224	184	914
	CS	41	198	167	922	43	207	179	928
	Semi mis	21	149	141	957	25	154	148	953
	Semi true	21	148	139	954	23	152	145	949

Note: Naive, the naive estimator; RC, regression calibration; SIMEX, the simulation extrapolation estimator; NPC, the nonparametric correction estimator of Huang & Wang (2000); NPC*, the NPC estimator after removing some outliers; CS, the conditional score estimator; Semi mis and Semi true, the locally efficient semiparametric estimators under the misspecified and true distribution of X , respectively; Bias, the empirical bias ($\times 10^3$); ESD, the empirical standard deviation ($\times 10^3$); ESE, the mean estimated standard error ($\times 10^3$); Cov, the empirical coverage probability of a 95% Wald confidence interval ($\times 10^3$).

Table 2

Estimation results of β under Simulation 1 and dependent censoring. The results are based on 1,000 replications.

Censoring	Estimator	Exponential				Weibull			
		Bias	ESD	ESE	Cov	Bias	ESD	ESE	Cov
30%	Naïve	-432	52	50	0	-440	53	50	0
	RC	-143	88	75	492	-154	89	76	441
	SIMEX	-187	85	60	206	-195	87	61	190
	NPC	72	264	166	918	88	287	178	907
	NPC*	60	206	166	924	66	216	177	921
	CS	35	178	153	938	36	190	163	937
60%	Semi mis	6	130	119	934	8	137	128	927
	Semi true	7	131	119	935	28	149	141	944
	Naïve	-403	68	64	0	-419	69	66	0
	RC	-101	111	96	760	-126	112	99	696
	SIMEX	-159	107	78	470	-173	110	81	423
	NPC	73	337	168	887	95	418	185	880
	NPC*	53	216	167	896	61	231	183	893
	CS	37	200	163	924	39	210	178	935
	Semi mis	4	147	132	929	3	152	139	938
	Semi true	3	149	134	924	-9	155	143	933

Note: The layout of the table is similar to Table 1.

Table 3

Results of the Simulation 2 based on 1,000 replications

Censoring	Estimator	β_1				β_2			
		Bias	ESD	ESE	Cov	Bias	ESD	ESE	Cov
30%	Naïve	181	124	124	662	142	79	76	507
	RC	76	136	134	890	71	86	82	832
	SIMEX	12	154	144	926	5	103	91	906
	CS*	-53	175	176	962	-46	120	123	974
	Semi mis	33	151	149	934	22	98	95	927
60%	Semi true	26	152	149	934	17	99	94	930
	Naïve	144	189	190	846	117	119	116	781
	RC	35	206	206	941	44	129	126	908
	SIMEX	-15	225	218	950	-8	146	137	933
	CS*	-82	253	254	964	-59	168	168	968
	Semi mis	-9	225	221	955	-5	144	140	945
	Semi true	-15	227	223	952	-9	146	141	943

Note: The layout of the table is similar to Table 1. CS* is the approximated conditional score estimator of Song *et al.* (2002).

Table 4

Results for the ACTG 175 data

Estimator	log(CD4)			Treatment		
	Est	SW SE	Boot SE	Est	SW SE	Boot SE
Naive	-1.879	0.327	0.366	-0.569	0.220	0.233
RC	-2.286	0.434	0.414	-0.675	0.216	0.226
SIMEX	-2.198	0.421	0.389	-0.681	0.219	0.227
NPC	-2.261	0.446	0.436	-0.670	0.222	0.232
CS	-2.262	0.392	0.445	-0.575	0.218	0.235
Semi uniform	-2.286	0.417	0.428	-0.585	0.225	0.237
Semi normal	-2.253	0.411	0.423	-0.580	0.224	0.236

Note: Naive, the naive estimator; RC, the regression calibration method; SIMEX, the simulation extrapolation estimator; NPC, the nonparametric correction estimator of Huang & Wang (2000); CS, the conditional score estimator; Semi uniform and Semi normal, the proposed semiparametric estimators when log(CD4) is assumed to be uniformly and normally distributed, respectively; Est, estimate; SW SE, standard error estimated using a sandwich formula; Boot SE, the bootstrap standard error based on 1,000 bootstrap replications.