University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working Paper Series

Year 2013 *Paper* 102

A Frailty Approach for Survival Analysis with Error-prone Covariate

Sehee Kim* Yi Li†

Donna Spiegelman‡

*University of Michigan - Ann Arbor, seheek@umich.edu

†University of Michigan - Ann Arbor

‡Harvard School of Public Health

A Frailty Approach for Survival Analysis with Error-prone Covariate

Sehee Kim, Yi Li, and Donna Spiegelman

Abstract

This paper discovers an inherent relationship between the survival model with covariate measurement error and the frailty model. The discovery motivates our using a frailty-based estimating equation to draw inference for the proportional hazards model with error-prone covariates. Our established framework accommodates general distributional structures for the error-prone covariates, not restricted to a linear additive measurement error model or Gaussian measurement error. When the conditional distribution of the frailty given the surrogate is unknown, it is estimated through a semiparametric copula function. The proposed copula-based approach enables us to fit flexible measurement error models without the curse of dimensionality as in nonparametric approaches, and to be applicable with an external validation study. Large sample properties are derived and finite sample properties are investigated through extensive simulation studies. The methods are applied to a study of physical activity in relation to breast cancer mortality in the Nurses' Health Study.

A frailty approach for survival analysis with error-prone covariate

Sehee Kim[1,*], Yi Li[1,2], and Donna Spiegelman[3]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109

[2]Kidney Epidemiology and Cost Center, Ann Arbor, MI 48109

[3]Departments of Epidemiology and Biostatistics, Harvard University, MA 02115

[*]*email*: seheek@umich.edu

Abstract

This paper discovers an inherent relationship between the survival model with covariate measurement error and the frailty model. The discovery motivates our using a frailty-based estimating equation to draw inference for the proportional hazards model with error-prone covariates. Our established framework accommodates general distributional structures for the error-prone covariates, not restricted to a linear additive measurement error model or Gaussian measurement error. When the conditional distribution of the frailty given the surrogate is unknown, it is estimated through a semiparametric copula function. The proposed copula-based approach enables us to fit flexible measurement error models without the curse of dimensionality as in nonparametric approaches, and to be applicable with an external validation study. Large sample properties are derived and finite sample properties are investigated through extensive simulation studies. The methods are applied to a study of physical activity in relation to breast cancer mortality in the Nurses' Health Study.

Key words: *Bias correction; Copula; Error-prone covariate; Frailty; Measurement error; Survival analysis.*

1 Introduction

In epidemiologic studies, risk factors such as nutrient intake, physical activity or air pollutants are often subject to measurement error. When the error-prone risk factors are included in a Cox (1972) survival model, the estimated effects of these model covariates can be under- or over-estimated, even for covariates that are measured without error. To address the bias caused by the mis-measured exposure, we focus on modeling the association between the true exposure "X" and its surrogate measure "Z". By viewing either the true exposure X or the surrogate Z as a function of a frailty, and by noting the structural similarity between frailty and measurement error methodology, we develop a novel method for measurement error correction that draws on the many years of development of frailty models in survival data analysis. The measurement error model considered here allows the observed exposure distribution to differ from the distribution of the true in a way that is unknown and to be estimated from the data. Our research is motivated by a study of physical activity in relation to breast cancer mortality in the Nurses' Health Study (Holmes et al., 2005). Limited existing models are available to prescribe the relationship between the true and surrogate measurements of physical activity. For example, the commonly used linear additive measurement error model fails in this case, because the surrogate measurements have a much heavier density around zero than the true counterpart.

In the presence of covariate measurement error, the Cox regression model has been studied by many authors. Among them, a simple and intuitive way of handling measurement error is the regression calibration approach (Wang et al., 1997; Spiegelman et al., 1997; Xie et al., 2001), which replaces the unobserved true exposure X with an "estimate" of X given its surrogate, and then obtains the standard partial likelihood estimator. Since these regression calibration estimators are based on a linear approximation to the expectation of X given Z, they may result in an inconsistent regression coefficient estimator. Another approach which does not require any distributional assumption on the measurement error is the "estimated" partial likelihood method proposed by Zhou and Pepe (1995) and Zhou and Wang (2000).

Unlike the regression calibration estimators, they deal with the induced relative risk function directly in a nonparametric way, and then estimate the regression coefficients from the resulting estimated partial likelihood function. More recently, Zucker (2005) also proposed a method based on the partial likelihood approach, however, their partial likelihood function is induced by a parametric specification for the true and surrogate measures. Our approach is related to estimation problems considered by Zhou and Wang (2000) and Zucker (2005), but is quite distinct from them through an alternative simple form of estimating equations.

Most of work to date requires subsampling within a main cohort study, i.e. an internal validation study, (Zhou and Pepe, 1995; Wang et al., 1997; Zhou and Wang, 2000; Chen, 2002), or requires replicate measurements on at least a subset of the study population, thereby requiring the classical additive model (Xie et al., 2001). However, in our motivating data example, only external validation samples that are independent from the main cohort were available, which precludes the direct use of these existing methods. Recently, Zucker (2005) proposed a pseudo partial likelihood-based method that can be applied with an external validation study. However, their approach requires a known parametric measurement error model. Distinct from existing methods, the regression coefficient estimator proposed in this paper does not require either linear additivity or any parametric distributional assumption on the measurement error model, and yet is applicable to an external validation study. The key feature that makes this greater flexibility possible is the use of a semi-parametric copula-based procedure for estimating the joint distribution of the true exposure X and the surrogate Z, greatly reducing bias due to exposure measurement error.

2 Notation and Survival and Measurement Error Models

Let $\tilde{T} = \min(T, C)$ be the observed follow-up time, where T and C are the failure and censoring times, respectively. A natural model that links the outcome T to the covariates is

the following proportional hazards model

$$\lambda_c(t|X,W) = \lambda(t)\exp(\beta X + \gamma^T W), \tag{2.1}$$

where $\lambda(\cdot)$ is an unspecified baseline hazard function, β and γ are unknown regression parameters corresponding to X, an error-prone covariate (e.g., the detailed physical activity diary), and W, a vector of error-free covariates (e.g., age and gender), respectively. Usually, X is the covariate of main interest and is difficult or expensive to measure. In a typical epidemiological study, we observe Z (e.g., the self-administered physical activity measure) in lieu of X, and Z is often termed the surrogate for X.

Deviating from the common measurement error literature, we postulate a general framework that directly deals with $f_X(x|z) = f(X = x|Z = z)$, the conditional density of the true covariate given the observed surrogate. This is equal to a general measurement error model

$$X = \mu(Z) + \epsilon, \tag{2.2}$$

where $\mu(z) = E[X|Z = z]$, which represents a location shift of $f_X(x|z)$, and ϵ is a mean-zero random error that depends on $Z = z$ with density $f_\epsilon(\epsilon|z) = f_X(\mu(z) + \epsilon|z)$. This general formulation encompasses many well-known measurement error models including the Berkson model as special cases.

Indeed, when both X and Z are deemed as random variables, which is always the case in observational studies, it is a matter of mathematical convenience whether the measurement error model is specified conditional on X or on Z. Conditioning on X is more often adopted in classical measurement error settings for ease of interpretation. However, specifying the model in this way makes stronger transportability assumptions than the other way around. Specifically, for an external validation study, the estimated $f_Z(z|x)$ in the validation study may not be transportable to the main study. This happens because $f_Z(z|x)$ may not entirely be identifiable over the support of X as X may distribute differently across the main

and external validation samples. On the other hand, working directly with $f_X(x|z)$ may circumvent such a difficulty as Z's are fully observed in both main and external validation samples.

Suppose we have i.i.d. observations on n individuals in the main cohort study. Using the counting process notation, let $Y_i(t) = I(\tilde{T}_i \geq t)$ be the at-risk process, and $N_i(t) = I(\tilde{T}_i \leq t, T_i \leq C_i)$ be the counting process, where $I(\cdot)$ is the indicator function. The main cohort study consists of $\{\tilde{T}_i, Y_i(t), N_i(t), Z_i, W_i; 0 \leq t \leq \tau\}$ $(i = 1, \ldots, n)$, where τ is the duration of study. We assume that Z_i is independent of outcome T_i given X_i and W_i, and X_i is independent of W_i given Z_i. The former assumption representation corresponds to the non-differential measurement error assumption, while the latter is assumed for notational ease and can easily be relaxed without loss of generality of the methods proposed. Finally, we assume that C_i and T_i are conditionally independent given observed Z_i and W_i, that is a non-informative censoring condition commonly used in survival analysis.

Under these assumptions, the hazard function for T_i, conditional on the observed covariates $\{Z_i, W_i\}$ and a frailty term ϵ, takes the form

$$\lambda_c(t|Z_i, W_i, \epsilon) = \lambda(t) \exp\{\beta\mu(Z_i) + \gamma^T W_i + \beta\epsilon\},$$

under models (2.1) and (2.2).

Such a formulation has important implications. First, it clearly establishes the linkage between the frailty concept and the measurement error framework, facilitating cross-fertilization between the two well-studied fields for methodological innovation and applications. Second, this new formulation advances the standard frailty models because the frailty term is not restricted to follow a certain parametric distribution (e.g. Gamma, log-normal etc.) or to be independent of the surrogate Z, as commonly assumed.

Following Prentice (1982) and Zucker (2005), we specify $S_i^*(t)$ the expected survival

function given the observed data only, i.e.,

$$\begin{aligned} S_i^*(t) &= \Pr[T_i > t | Z_i, W_i] \\ &= \int \exp\left[-\Lambda(t)\exp\left\{\beta\mu(Z_i) + \gamma^T W_i + \beta\epsilon\right\}\right] f_\epsilon(\epsilon|Z_i)\, d\epsilon, \end{aligned}$$

where $\Lambda(t)$ is the cumulative baseline hazard function. The corresponding conditional hazard function is given by

$$\begin{aligned} d\Lambda_i^*(t) &= \lambda(t)\, \frac{\int \psi_i(\epsilon;\, \beta, \gamma) \exp\{-\Lambda(t)\psi_i(\epsilon;\, \beta, \gamma)\} f_\epsilon(\epsilon|Z_i)\, d\epsilon}{\int \exp\{-\Lambda(t)\psi_i(\epsilon;\, \beta, \gamma)\} f_\epsilon(\epsilon|Z_i)\, d\epsilon} \\ &= \lambda(t)\, E[\psi_i(\epsilon;\, \beta, \gamma) | Z_i, W_i, T_i > t] \\ &\equiv \lambda(t)\, \eta_i(\beta, \gamma, \Lambda(t)), \end{aligned} \tag{2.3}$$

where $\psi_i(\epsilon;\, \beta, \gamma) = \exp\{\beta\mu(Z_i) + \gamma^T W_i + \beta\epsilon\}$. The derivation holds because $f_\epsilon(\epsilon|Z_i, W_i) = f_\epsilon(\epsilon|Z_i)$ under the assumption that ϵ is independent of W_i given Z_i.

3 Frailty-copula Estimator

3.1 When $f_X(x|z)$ is known up to a parametric form

The treatment of the conditional density $f_\epsilon(\epsilon|z)$ is the key to this development. We approach it by estimating $f_X(x|z)$, which is equivalent to $f_\epsilon(\epsilon|z)$ in distribution since $\mu(z)$ is simply the location shift. We first develop the case where $f_X(x|z)$ is known in Section 3.1. When $f_X(x|z)$ is unknown, we propose to estimate it using a copula framework as discussed in Section 3.2, with the remaining inference procedures following Section 3.1 by replacing $f_X(x|z)$ with its estimated counterpart.

Suppose f_X belongs to a family of parametric models indexed by a vector parameter ξ in $\mathbb{R}^p$, which is denoted by $f_X(x|z;\, \xi)$. We denote the true value of β by β_0, the true vector of γ by γ_0, and the true $\Lambda(t)$ by $\Lambda_0(t)$. Let $M_i(t;\, \beta, \gamma, \Lambda) = N_i(t) - \int_0^t Y_i(s)\, d\Lambda_i^*(s) = N_i(t) -$

$\int_0^t Y_i(s)\eta_i(\beta,\gamma,\Lambda(s))\,d\Lambda(s)$. With non-informative censoring, $M_i(t) \equiv M_i(t;\beta_0,\gamma_0,\Lambda_0)$ $(i = 1,\ldots,n)$ is a martingale process with respect to the filtration $\sigma\{N_i(s-), Y_i(s), W_i, Z_i, 0 \le s \le t\}$. As such, we propose the following estimating equations:

$$0 = \sum_{i=1}^{n} \int_0^{\tau} \mu(Z_i)\, dM_i(t;\, \beta,\gamma,\Lambda), \tag{3.1}$$

$$0 = \sum_{i=1}^{n} \int_0^{\tau} W_i\, dM_i(t;\, \beta,\gamma,\Lambda), \tag{3.2}$$

$$0 = \sum_{i=1}^{n} \left[dN_i(t) - Y_i(t)\,\eta_i(\beta,\gamma,\Lambda(t))\,d\Lambda(t)\right]. \tag{3.3}$$

Similar estimating equations have been proposed by Pipper and Martinussen (2004) in a parametric frailty setting, in the absence of measurement problems. To our knowledge, this is the first attempt to adopt such estimating equations in a measurement error framework. The equation (6) yields a Breslow-type estimator for $\Lambda_0(t)$

$$\begin{aligned} \hat{\Lambda}(t) &= \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)}{\sum_{i=1}^n Y_i(u)\,\eta_i(\beta,\gamma,\hat{\Lambda}(u-))}, \quad \text{or} \\ \hat{\lambda}(u_m) &= \frac{d_m}{\sum_{i=1}^n Y_i(u_m)\,\eta_i(\beta,\gamma,\hat{\Lambda}(u_{m-1}))}, \end{aligned}$$

for $u_m \le t < u_{m+1}$, where $u_1, u_2, \ldots, u_M$ are the ordered observed event times, d_m is the number of events at u_m, and $\hat{\Lambda}(0) = 0$. A similar estimate $\hat{\Lambda}(t)$ was obtained by Zucker (2005) in a different context based on a pseudo-partial likelihood function. Define $Q_i = (\mu(Z_i),\, W_i^T)^T$ as a vector of the observed covariates and $\theta_0 = (\beta_0,\, \gamma_0^T)^T$ as the vector of corresponding true regression coefficients. By substituting $\hat{\Lambda}(t)$ for $\Lambda(t)$ in $M_i(t;\,\beta,\gamma,\Lambda)$, the estimating equations for θ_0 in (3.1) and (3.2) become

$$U(\theta,\,\hat{\Lambda}) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ Q_i - \frac{S^{(1)}(t;\,\theta,\hat{\Lambda})}{S^{(0)}(t;\,\theta,\hat{\Lambda})} \right\} dN_i(t), \tag{3.4}$$

where $S^{(k)}(t;\,\theta,\Lambda) = n^{-1}\sum_{i=1}^n Y_i(t)\,Q_i^{\otimes k}\eta_i(\theta,\Lambda(t))$ $(k = 0,\,1,\,2)$, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and

$a^{\otimes 2} = aa^T$. Then, $\hat{\theta}$ is the solution to $U(\theta, \hat{\Lambda}) = 0$, which can be found numerically using the Newton-Raphson algorithm, for example. Our estimating equations have a much simpler form than Zucker's (2005) score functions because Zucker (2005) had $\partial\{\log \eta_i\}/\partial\beta$ in the first term of (3.4), which can be very computationally involved.

To compute the covariance matrix of $\hat{\theta}$, we further define

$$\begin{aligned} S_\theta^{(k)}(t;\, \theta, \Lambda) &= n^{-1} \sum_{i=1}^{n} Y_i(t)\, Q_i^{\otimes k} \dot{\eta}_{\theta i}(\theta, \Lambda(t)), \\ S_\Lambda^{(k)}(t;\, \theta, \Lambda) &= n^{-1} \sum_{i=1}^{n} Y_i(t)\, Q_i^{\otimes k} \dot{\eta}_{\Lambda i}(\theta, \Lambda(t)), \end{aligned}$$

where

$$\begin{aligned} \dot{\eta}_{\theta i} &= \partial \eta_i(\theta, \Lambda(t))/\partial\theta \\ &= \frac{E_{X|Z}\Big[\tilde{Q}_i \psi_i(X_i;\theta) e^{-\Lambda(t)\psi_i(X_i;\theta)} - \psi_i^2(X_i;\theta) e^{-\Lambda(t)\psi_i(X_i;\theta)} \{\tilde{Q}_i \Lambda(t) + \partial_\theta \Lambda(t)\}\Big]}{E_{X|Z}\big[e^{-\Lambda(t)\psi_i(X_i;\theta)}\big]} \\ &\quad + \frac{E_{X|Z}\big[\psi_i(X_i;\theta) e^{-\Lambda(t)\psi_i(X_i;\theta)}\big]\, E_{X|Z}\Big[\psi_i(X_i;\theta) e^{-\Lambda(t)\psi_i(X_i;\theta)} \{\tilde{Q}_i \Lambda(t) + \partial_\theta \Lambda(t)\}\Big]}{\big\{E_{X|Z}\big[e^{-\Lambda(t)\psi_i(X_i;\theta)}\big]\big\}^2}, \end{aligned}$$

and

$$\begin{aligned} \dot{\eta}_{\Lambda i} &= \partial \eta_i(\theta, \Lambda(t))/\partial\Lambda(t) \\ &= -\frac{E_{X|Z}\big[\psi_i^2(X_i;\theta) e^{-\Lambda(t)\psi_i(X_i;\theta)}\big]}{E_{X|Z}\big[e^{-\Lambda(t)\psi_i(X_i;\theta)}\big]} + \left\{ \frac{E_{X|Z}\big[\psi_i(X_i;\theta) e^{-\Lambda(t)\psi_i(X_i;\theta)}\big]}{E_{X|Z}\big[e^{-\Lambda(t)\psi_i(X_i;\theta)}\big]} \right\}^2 \end{aligned}$$

at a fixed time t, where $\tilde{Q}_i = (X_i, W_i^T)^T$, $E_{X|Z}$ denotes the conditional expectation with respect to X given Z, and

$$\partial_\theta \Lambda(t) = -\sum_{i=1}^{n} \int_0^t \left\{ \sum_i Y_i(u) \dot{\eta}_{\theta i}(\theta, \Lambda(u-)) \right\} \left\{ \sum_i Y_i(u) \eta_i(\theta, \Lambda(u-)) \right\}^{-2} dN_i(u).$$

The theorem below stipulates that the covariance matrix of $n^{1/2}\,\hat{\theta}$ can be consistently

estimated by $\hat{D}^{-1} + \hat{D}^{-1}\hat{H}\hat{D}^{-1}$, where

$$\begin{aligned}
\hat{D} &= n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left\{S_\theta^{(1)}/S^{(0)}(t;\,\hat{\theta},\hat{\Lambda}) - S^{(1)}S_\theta^{(0)}/(S^{(0)})^2(t;\,\hat{\theta},\hat{\Lambda})\right\}dN_i(t),\\
\hat{H} &= n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\frac{G(t;\,\hat{\theta},\hat{\Lambda})^{\otimes 2}\,\hat{R}(t-)^2}{\{\sum_{i=1}^{n}Y_i(t)\,\eta_i(\hat{\theta},\hat{\Lambda}(t))\}^2}\,dN_i(t),\\
G(t;\,\hat{\theta},\hat{\Lambda}) &= \sum_{i=1}^{n}\int_t^{\tau}\{S^{(1)}S_\Lambda^{(0)}/(S^{(0)})^2(u;\,\hat{\theta},\hat{\Lambda}) - S_\Lambda^{(1)}/S^{(0)}(u;\,\hat{\theta},\hat{\Lambda})\}/\hat{R}(u)\,dN_i(u),\\
\hat{R}(t) &= \prod_{u\le t}\left\{1+\sum_{i=1}^{n}nS_\Lambda^{(0)}/(S^{(0)})^2(u;\,\hat{\theta},\hat{\Lambda})\,dN_i(u)\right\}.
\end{aligned}$$

Theorem 1 *Under regularity conditions (C1) - (C6), $\hat{\theta}$ is a consistent estimator of θ_0 and $n^{1/2}(\hat{\theta}-\theta_0)$ is asymptotically normally distributed with mean 0 and variance-covariance matrix $D^{-1}+D^{-1}HD^{-1}$.*

The asymptotic covariance is given by replacing the sample quantities in $\hat{D}^{-1}+\hat{D}^{-1}\hat{H}\hat{D}^{-1}$ with their corresponding population quantities. The regularity conditions and proofs of Theorems 1 - 3 (Theorems 2 and 3 appear below) are given in the Web Appendices.

When ξ is unknown, ξ can be estimated from an external validation study using a likelihood-based approach. The external data consist of $\{X_j, Z_j;\, j=1,\ldots,n_v\}$, where n_v is the sample size of the validation study. The following theorem establishes the asymptotic normality of the estimator $\hat{\theta}$ when ξ is unknown. Condition (C7) is added to guarantee identifiability of ξ.

Theorem 2 *Under regularity conditions (C1) - (C7), $\hat{\theta}$ is a consistent estimator of θ_0 and $n^{1/2}(\hat{\theta}-\theta_0)$ is asymptotically normally distributed with mean 0 and variance-covariance matrix $D(\theta_0)^{-1}+D(\theta_0)^{-1}[H+V(\xi_0)\,\Omega\,V(\xi_0)^T]D(\theta_0)^{-1}$, where Ω is the variance of the maximum likelihood estimator $\hat{\xi}$ and $V(\xi_0)$ is the limit of $n^{-1}\partial U(\theta_0,\,\Lambda_0,\,\xi_0)/\partial\xi$.*

The asymptotic variance-covariance matrix can be consistently estimated by replacing $f_X(x|z;\,\xi_0)$

with $f_X(x|z;\hat{\xi})$ in $\hat{D}$ and $\hat{H}$, along with the consistent estimator for $V(\xi_0)$

$$\begin{aligned}\hat{V} &= n^{-1}\sum_i \int_0^\tau \left[\dot{Q}_{\xi i}(\hat{\xi}) + \frac{\sum_i Y_i(t)\,\eta_i(\hat{\theta},\hat{\Lambda},\hat{\xi})\,Q_i(\hat{\xi})\,\sum_i Y_i(t)\,\dot{\eta}_{\xi i}(\hat{\theta},\hat{\Lambda},\hat{\xi})^T}{\{\sum_i Y_i(t)\,\eta_i(\hat{\theta},\hat{\Lambda},\hat{\xi})\}^2}\right] dN_i(t) \quad (3.5)\\ &\quad -n^{-1}\sum_{i=1}^n \int_0^\tau \left[\frac{\sum_i Y_i(t)\,Q_i(\hat{\xi})\,\dot{\eta}_{\xi i}(\hat{\theta},\hat{\Lambda},\hat{\xi})^T\,\sum_i Y_i(t)\,\eta_i(\hat{\theta},\hat{\Lambda},\hat{\xi})\,\dot{Q}_{\xi i}(\hat{\xi})}{\sum_i Y_i(t)\,\eta_i(\hat{\theta},\Lambda_0,\hat{\xi})}\right] dN_i(t),\end{aligned}$$

where $\dot{Q}_{\xi i}$ and $\dot{\eta}_{\xi i}$ are the partial derivatives of $Q_i(\xi)$ and $\eta_i(\theta,\Lambda,\xi)$ with respect to ξ.

3.2 When the parametric form of $f_X(x|z)$ is unknown

Given the availability of a validation study where both X and Z are observed, $f_X(x|z)$ can be, in theory, estimated non-parametrically. However, when the sample size in the validation data set is moderate or small and when X and Z are continuous, as in our motivating example, estimation of $f_X(x|z)$ non-parametrically would be unstable, which would further deteriorate the performance of the proposed method through the propagation of further error in the estimating process. As a remedy, we propose a semi-parametric method that utilizes a copula framework for estimating $f_X(x|z)$. The motivation stems from the fact that $f_X(x|z)$ can be viewed as a functional of F_X and F_Z, the marginal distributions of X and Z respectively, in a copula setting.

Specifically, by invoking the Sklar (1959) theorem, it follows that

$$f_X(x|z) = \frac{f_{XZ}(x,z)}{f_Z(z)} = C'_\xi(F_X(x), F_Z(z))\, f_X(x),$$

provided that $f_Z(z) > 0$, where C'_ξ is a copula density function with a dependence parameter ξ in $\mathbb{R}$. Furthermore, if F_X and F_Z are continuous as in our case, then the copula distribution function C_ξ is uniquely determined (Sklar, 1959).

We propose to estimate $f_X(x|z)$ as follows:

Step 1. Estimate $f_X(x)$ and $f_Z(z)$ separately in the validation study through kernel density estimation (Wand and Jones, 1995). Suppose $(X_j,\ Z_j)$ is the jth sample in the

validation study. The kernel density estimator of f_X at the point x is given by

$$\hat{f}_X(x) = n_v^{-1} \sum_{j=1}^{n_v} b^{-1} K\left(\frac{x - X_j}{b}\right),$$

where the kernel K satisfies $\int K(x)\,dx = 1$ and the bandwidth b controls the degree of smoothness of the density function.

Step 2. Compute $\hat{F}_X(x_j)$ and $\hat{F}_Z(z_j)$ at $\{(x_j, z_j);\, j = 1, \ldots, n_v\}$ from the validation study samples.

Step 3. Given a specific copula form C_ξ, estimate its dependence parameter ξ by maximizing the likelihood of the validation study, i.e., $\prod_{j=1}^{n_v} C_\xi(\hat{F}_X(x_j), \hat{F}_Z(z_j))$. The estimate of $f_X(x|z)$ is computed as $\hat{f}_X(x|z) = C'_{\hat{\xi}}(\hat{F}_X(x), \hat{F}_Z(z))\,\hat{f}_X(x)$.

In Step 1, the kernel K can be chosen to be a unimodal probability density function symmetric about zero, satisfying the conditions: $\int xK(x)\,dx = 0$ and $\int x^2K(x)\,dx > 0$. Our method can be applied with any choice of K. However, we consider particularly the two most popular choices for K: the Gaussian kernel (Eubank, 1988) $K_G(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ and the Epanechnikov kernel (Eubank, 1988) $K_E(u) = 3/4\,(1 - u^2)I(|u| \leq 1)$. In Step 2, a numerical integration algorithm such as Gaussian quadrature can be used. When the Gaussian kernel is specified in Step 1, Gaussian quadrature with the Hermite polynomials is used for approximating the integral over $(-\infty, \infty)$, while Legendre polynomials are used for approximating the integral over the finite support $[-1, 1]$ when the Epanechnikov kernel is specified. In Step 3, there are many parametric copula families available to control the structure and the strength of dependence. Hutchinson and Lai (1990) and Nelsen (2006, Chap 4.3) provide a thorough coverage of bivariate copulas and their properties. Among them, we investigated three types of copula families: Gaussian copulas, Clayton copulas, and Gumbel copulas. These copula families encompass a variety of dependence structures: the Gaussian copula describes symmetric dependence, while Clayton and Gumbel are suited

for stronger negative and positive tail correlations, respectively. The maximum likelihood estimator for the dependence parameter ξ of the copula is obtained, and then the best copula family for the data at hand is selected using a likelihood-based criterion such as the Akaike Information Criterion (Genest et al., 2009). More detail on other copula families and guidance on how to choose them can be found in Nelsen (2006).

Once $\hat{f}_X(x|z)$ is available, the frailty-copula estimator of θ can be obtained as described in Section 3.1. The next theorem summarizes the asymptotic properties of the resulting estimator $\hat{\theta}$.

Theorem 3 *Under regularity conditions (C1) - (C11), $\hat{\theta}$ is a consistent estimator of θ_0, and $n^{1/2}(\hat{\theta} - \theta_0)$ weakly converges to a distribution with mean 0 and asymptotic covariance $D(\theta_0)^{-1} + D(\theta_0)^{-1}[H + V_f]\, D(\theta_0)^{-1}$, where V_f is the variance of $U_f'(M)$, U_f' is the Hadamard derivative of $U(\theta, \Lambda, f_{X|Z})$ at $f_{X|Z}$, and M is a random variable following a mean-zero normal random variate with covariance matrix $\{C'_{\xi_0}(F_X, F_Z)\}^2 f_X \int K^2(u)\, du$.*

However, the variance estimator based on the formula given in Theorem 3 requires knowledge of the exact forms of the functional and the kernels, which is not realistic in practice. A more practical alternative approach to variance estimation is a non-parametric bootstrap, which we found to perform well as demonstrated in Section 4.

4 Numerical Results

4.1 Simulation Studies

Extensive simulations were conducted to evaluate the finite sample properties of the proposed estimator in various settings representative of what may occur in practice. We considered the proportional hazards model with two covariates, the error-prone covariate X and the error-free covariate $W \sim N(0, 1)$. To generate the error-prone covariate and surrogate (X, Z), we considered two models: (Model A) (X, Z) were from a normal density with mean

(0, 0) and variance (1, 1.5) and correlation 0.8, and (Model B) (X, Z) were generated from the same distribution as in Model A, but Z was truncated at -1. The former simulated the common classical measurement error model, while the latter simulated a nonlinear relationship between the true exposure and surrogate, mimicking the setting of our motivating data example.

To generate right-censored failure time data, we used an exponential baseline hazard with mean ν and fixed the censoring time to be 1. The constant hazard ν was varying according to the desired censoring rates of 40% and 90%, reflecting the relatively high censoring rates typically found in epidemiologic applications. We considered sample sizes $n = 500$ and $n_v = 100$ for the common disease setting, and $n = 3000$ and $n_v = 200$ for the rare disease setting as in our motivating example.

To estimate $f_X(x)$, we considered the Gaussian kernel and Epanechnikov kernel functions with a bandwidth $b = 0.9 \min(\hat{\sigma}_X, \text{IQR}(X)/1.34)\, n_v^{-1/5}$, where $\hat{\sigma}_X$ is the sample standard deviation of X and IQR is the interquartile range (Silverman, 1986, Chap 3.4), satisfying the bandwidth conditions in Theorem 3. The nonparametric bootstrap was used to estimate the variance of $\hat{\beta}_{FCG}$ and $\hat{\beta}_{FCE}$ with 100 replacement resamples.

Table 1 summarizes the main results of estimating the regression coefficient $\beta_0 = 1$, based on 1000 replicates. The naive Cox estimator, $\hat{\beta}_{NC}$, was always biased toward the null. The ordinary regression calibration estimator, $\hat{\beta}_{ORC}$, yielded smaller bias, but had larger variance compared to $\hat{\beta}_{NC}$. The bias in $\hat{\beta}_{ORC}$ became larger when the truncated surrogate model (Model B) was considered (see Table 1). In addition, in the common disease case, $\hat{\beta}_{ORC}$ did not perform as well as in the rare disease case. On the other hand, the proposed frailty-copula estimators, $\hat{\beta}_{FCG}$ and $\hat{\beta}_{FCE}$, performed consistently well in all scenarios studied. With both linear and non-linear measurement error, the proposed estimators were virtually unbiased. Although the variances of $\hat{\beta}_{FCG}$ and $\hat{\beta}_{FCE}$ were larger than those for $\hat{\beta}_{ORC}$, by the mean-squared error (MSE) criterion, the frailty-copula estimators were among the best in all the scenarios considered. Finally, we note that the coverage probabilities for the frailty-copula

estimators under both the common and rare diseases scenarios lay in a reasonable range around the nominal 0.95.

4.2 Motivating Example

We applied our method to a study of the effect of physical activity on survival after breast cancer diagnosis in the Nurses' Health Study (NHS) (Holmes et al., 2005). The key features of this study are very similar to the simulation study described in Section 4.1, especially for Model B and the rare disease case shown in Table 1. The NHS is a prospective observational study, following 121,700 female registered nurses since 1976 who were 30-55 years of age at the start of follow-up. Our analysis focused on the 2987 women who were diagnosed with breast cancer in stages I, II or III between 1984 and 1998. These women were followed until death or June 2002, with a median follow-up time of 8 years, and 280 women (9.4%) died from breast cancer during the follow-up period. Physical activity, the primary exposure, was assessed as metabolic equivalent task (MET) hours per week at least 2 years after cancer diagnosis (a median of 3.2 years) to avoid bias due to declining physical activity immediately prior to and after cancer diagnosis.

An external validation study was conducted in the Nurses' Health Study II cohort, where the validity of the self-administered physical activity measure based on questionnaires (i.e., surrogate) was assessed using a detailed activity diary (i.e., the gold standard). The validation data from these 149 women were available to build the measurement error model (Wolf et al., 1994). Because preliminary analyses revealed that there was larger variation in the surrogate than the true measure, we used a functional measurement error model as in Section 4.1. The challenge is two-fold: the assessment of physical activity based on self-reported questionnaires was subject to measurement error, and based on the validation study, the relationship between the true physical activity and its surrogate measure could not be described by a simple linear function; hence, the classical error model would not be appropriate to correct the measurement error.

Table 1: Simulation results ($\beta = 1$)

		Model A		**Model B**	
		Common[1]	Rare[2]	Common[1]	Rare[2]
$\hat{\beta}_{NC}$	Bias	-0.412	-0.357	-0.351	-0.318
	SSD	0.054	0.048	0.059	0.048
	SEE	0.054	0.047	0.059	0.047
	CP (%)	0.0	0.0	0.1	0.0
	MSE$\times 10^2$	17.3	13.0	12.7	10.4
$\hat{\beta}_{ORC}$	Bias	-0.114	-0.032	-0.158	-0.115
	SSD	0.104	0.085	0.105	0.080
	SEE	0.105	0.086	0.105	0.081
	CP (%)	75.7	91.5	60.9	67.0
	MSE$\times 10^2$	2.4	0.8	3.6	2.0
$\hat{\beta}_{FCG}$	Bias	-0.026	-0.024	-0.002	-0.019
	SSD	0.153	0.102	0.169	0.106
	SEE	0.166	0.109	0.160	0.101
	CP (%)	93.6	93.0	91.9	92.0
	MSE$\times 10^2$	2.4	1.1	2.8	1.2
$\hat{\beta}_{FCE}$	Bias	-0.031	-0.016	0.027	-0.003
	SSD	0.164	0.112	0.172	0.116
	SEE	0.204	0.131	0.169	0.104
	CP (%)	95.8	95.2	94.7	92.1
	MSE$\times 10^2$	2.8	1.3	3.0	1.3

[1] Common disease setting: $n = 500$, $n_v = 100$, 60% event rate
[2] Rare disease setting: $n = 3000$, $n_v = 200$, 10% event rate
NOTE: $\hat{\beta}_{NC}$ is the naive Cox estimator; $\hat{\beta}_{ORC}$ is the ordinary regression calibration estimator; $\hat{\beta}_{FCG}$ and $\hat{\beta}_{FCE}$ are the proposed frailty-copula estimators, using the Gaussian and Epanechnikov kernel smoothings, respectively; SSD is the sample standard deviation; SEE is the standard error estimate; CP is the coverage probability of the 95% confidence interval; MSE is the mean squared error.

Table 2 shows the estimated effect of physical activity on breast cancer survival using the new frailty-copula estimators, compared to the naive Cox approach. The univariate analysis results were from a model with exposure only, and the multivariate analysis results were adjusted for possible other confounders, including age at diagnosis, body mass index (BMI) and cancer stage. In both analyses, increasing average daily physical activity had a significant protective effect on breast cancer survival, and the magnitude of the effect was attenuated (i.e., hazard ratio closer to 1) when the surrogate measure was used without adjusting for the measurement error (see results for $\hat{\beta}_{NC}$). In contrast, the proposed frailty-copula approach substantially corrected for the attenuation effect (see results for $\hat{\beta}_{FCG}$ and $\hat{\beta}_{FCE}$). Measurement error in physical activity had minimal impact on the estimated effect of age and BMI because they were not confounders. In some cases where the variables measured with error are more highly correlated with other covariates, however, measurement error will induce bias on the other estimated model coefficients too.

5 Conclusion

This paper reveals a previously undisclosed inherent relationship between the survival model with covariate measurement error and the frailty model. We have exploited this relationship to propose a frailty-copula approach for consistent estimation of the effect of an error-prone covariate in the Cox model through the derivation of simple unbiased estimating equations. The proposed approach is applicable both when the distribution of the frailty term given the surrogate is known up to a parametric form and when this distribution is unknown. When a parametric distribution for the frailty can be assumed, measurement error model parameters can be estimated by maximum likelihood. When the parametric form for the frailty distribution is unknown, a semi-parametric density estimator arising from the copula is used to estimate it. Our proposed framework is general – it accommodates flexible general measurement error models, including the commonly used classical measurement error model

Table 2: Analysis results for the study of physical activity in relation to breast cancer mortality in the Nurses' Health Study

	Univariate Analysis		**Multivariate Analysis**	
Effect	HR (95% CI)	P	HR (95% CI)	P
	Naive Cox Estimator ($\hat{\beta}_{NC}$)			
Physical activity [1]	0.788 (0.645, 0.963)	.019	0.794 (0.651, 0.970)	.024
Age at diagnosis (10 year) [2]			1.107 (0.946 1.295)	.210
Overweight (BMI $\geq$ 25) [2]			1.043 (0.819, 1.327)	.740
Cancer stage II or III [2]			3.815 (2.934, 4.961)	<.001
	Frailty-copula estimator using the Gaussian kernel ($\hat{\beta}_{FCG}$)			
Physical activity [1]	0.445 (0.201, 0.986)	.023	0.453 (0.201, 1.020)	.028
Age at diagnosis (10 year) [2]			1.099 (0.933, 1.295)	.129
Overweight (BMI $\geq$ 25) [2]			1.023 (0.799, 1.310)	.429
Cancer stage II or III [2]			3.846 (2.949, 5.015)	<.001
	Frailty-copula estimator using the Epanechnikov kernel ($\hat{\beta}_{FCE}$)			
Physical activity [1]	0.433 (0.229, 0.818)	.005	0.448 (0.236, 0.849)	.007
Age at diagnosis (10 year) [2]			1.102 (0.934, 1.301)	.124
Overweight (BMI $\geq$ 25) [2]			1.023 (0.799, 1.312)	.427
Cancer stage II or III [2]			3.844 (2.953. 5.003)	<.001

[1] error-prone covariate; per 20 MET-hrs/wk

[2] error-free covariate

[3] HR = Hazard Ratio; P = p-value

as a special case.

Other attractive features of the method include its applicability to external validation studies, for which very few existing methods are available. Moreover, compared to the alternative methods such as regression calibration, this new frailty approach shows good performance.

While the method has been restricted to a univariate error-prone covariate, extending the methodology to allow multiple mis-measured exposures is straightforward, if the multi-dimensional distribution of exposures is known or can be estimated from validation or reliability data. Future efforts will be devoted to estimating the multi-dimensional conditional distribution of the exposures given the surrogates using approaches for conditional copulas, for example. To save computation time, we have studied several types of copulas without considering model selection. However, it is worth considering methods for optimal copula selection, such as goodness-of-fit tests based on the empirical copula (Durrleman et al., 2000), the Kendall process (Genest and Rivest, 1993; Genest et al., 2009), and kernel density estimation (Fermanian, 2005). Based on our simulation investigation, results were robust to the choice of kernel function and its bandwidth, in terms of the mean squared errors of $\hat{\beta}$. For example, regardless of whether Gaussian, Epanechnikov or biweight kernel functions were applied with different bandwidths $b = c \times \min(\hat{\sigma}_X, \text{IQR}(X)/1.34)\, n_v^{-1/5}$, where $c = 0.9$, 2.34, or 2.78, the MSEs of $\hat{\beta}$ changed by no more than 0.004. However, it will be worthwhile to investigate whether efforts to reduce the asymptotic mean integrated squared error of the frailty distribution function itself via bandwidth selection tools such as cross-validation would appreciably improve the overall performance of β estimation.

This novel frailty-copula approach for solving the covariate measurement error problem in Cox regression models establishes a previously unnoticed linkage between the frailty concept and the measurement error framework, facilitating cross-fertilization between these two fields. With this formulation, we have simultaneously advanced standard frailty models by eliminating the restriction of the frailty term to a parametric distribution or to an assumed

independence from model covariates. Future research will investigate the application of these new developments to frailty models in survival data analysis.

Finally, although we mainly work on $f_X(x|z)$ in the paper, with the identity $f_Z(z|x) = f_X(x|z)f_Z(z)/f_X(x)$, our Copula-based approach can also handle $f_Z(z|x)$ when both X and Z are continuous variables. This way, our formulation would encompass almost all the major measurement error models.

6 Supplementary Materials

Web Appendices referenced in Section 3 are available from the corresponding author.

Acknowledgements

The authors gratefully acknowledge the support of NIH/NCI grant R01 CA050597 and NIH/NIEHS grant R01 ES009411.

References

Chen, Y.-H. (2002). Cox regression in cohort studies with validation sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64,** 51–62.

Durrleman, V., Nikeghbali, A., and Roncalli, T. (2000). Which copula is the right one? *Techinical report, Groupe de Recherche Operationnelle, Credit Lyonnais* .

Eubank, R. L. (1988). *Spline smoothing and nonparametric regression*. Dekker, New York.

Fermanian, J.-D. (2005). Goodness-of-fit tests for copulas. *Journal of multivariate analysis* **95,** 119–152.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics* **44,** 199–213.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association* **88,** 1034–1043.

Holmes, M. D., Chen, W. Y., Feskanich, D., Kroenke, C. H., and Colditz, G. A. (2005). Physical activity and survival after breast cancer diagnosis. *Journal of the American Medical Association* **293,** 2479–2486.

Hutchinson, T. and Lai, C. D. (1990). *Continuous bivariate distributions, emphasising applications*. Rumsby Scientific Publishing Adelaide, Sydney.

Nelsen, R. (2006). *An introduction to copulas*. Springer Verlag, New York.

Pipper, C. B. and Martinussen, T. (2004). An estimating equation for parametric shared frailty models with marginal additive hazards. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66,** 207–220.

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

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall, London.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8,** 229–231.

Spiegelman, D., McDermott, A., and Rosner, B. (1997). The regression calibration method for correcting measurement error bias in nutritional epidemiology. *American Journal of Clinical Nutrition* **65,** 1179S–1186S.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*. Chapman & Hall, London.

Wang, C., Hsu, L., Feng, Z., and Prentice, R. L. (1997). Regression calibration in failure time regression. *Biometrics* **53,** 131–145.

Wolf, A. M., Hunter, D. J., Colditz, G. A., Manson, J. E., Stampfer, M. J., Corsano, K. A., Rosner, B., Kriska, A., and Willett, W. C. (1994). Reproducibility and validity of a self-administered physical activity questionnaire. *International Journal of Epidemiology* **23,** 991–999.

Xie, S. X., Wang, C., and Prentice, R. L. (2001). A risk set calibration method for failure time regression by using a covariate reliability sample. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63,** 855–870.

Zhou, H. and Pepe, M. S. (1995). Auxiliary covariate data in failure time regression. *Biometrika* **82,** 139–149.

Zhou, H. and Wang, C. Y. (2000). Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62,** 657–665.

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

Web-based Supplementary Materials for "A frailty approach for survival analysis with error-prone covariate" by

Sehee Kim*, Yi Li, and Donna Spiegelman

*email: seheek@umich.edu

Proofs of Asymptotic Properties

This section proves the asymptotic properties of the inference procedures proposed in Section 3. Let θ be the vector of (β, γ^T) and let $(\theta_0, \Lambda_0(\cdot))$ be the true parameter values of $(\theta, \Lambda(\cdot))$. We impose the following regularity conditions:

(C1) The parameter value of θ_0 belongs to the interior of a compact set Θ in the domain of θ.

(C2) The function $\Lambda_0(t)$ is strictly increasing and continuously differentiable with derivative $\lambda_0(t) > 0$ for every $t \in [0, \tau]$, where τ is the duration of the study. The baseline hazard function $\lambda_0(t)$ is bounded above by some constant λ_{max} for all $t \in [0, \tau]$.

(C3) X, Z, and W are bounded.

(C4) With probability 1, there exists a positive constant δ_0 such that $P(C \geq \tau) > \delta_0$.

(C5) The limiting value $D(\theta_0)$ of $-n^{-1}\, \partial U(\theta, \Lambda_0)/\partial\theta|_{\theta=\theta_0}$ is positive definite with probability 1.

(C6) The function $\partial f_X(x|z; \xi)/\partial\xi$ is absolutely integrable.

(C7) The density function $f_X(x|z;\,\xi) = f_X(x|z;\,\xi_0)$ almost surely if and only if $\xi = \xi_0$. In addition, if $\nu^T \dot{f}_{Xz}(x|z;\,\xi_0) = 0$ holds for any vector ν almost surely, then $\nu = 0$, where $\dot{f}_{Xz}$ denotes the derivative of $f_X(x|z)$ with respect to ξ.

(C8)

$$\lim_{n_v\to\infty} n_v b = \infty, \quad \lim_{n_v\to\infty} n_v b^5 = 0, \quad \text{and} \quad \lim_{n_v\to\infty} b\log n_v = 0.$$

(C9) $\hat{\xi}$ is an estimator for ξ_0 satisfying

$$\|\hat{\xi} - \xi_0\| = O(\sqrt{\log n_v/n_v}) \quad \text{a.s.} \quad \text{and} \quad E(\hat{\xi} - \xi_0)^2 = O(n_v^{-1}\log n_v).$$

(C10) C'_ξ is bounded on $[0,\,1]^2$ and

$$|C'_{\xi_1}(u_1,\,v_1) - C'_{\xi_2}(u_2,\,v_2)| \le C_1(|u_1 - u_2| + |v_1 - v_2| + |\xi_1 - \xi_2|),$$

where $u_j = F_X(x_j)$, $v_j = F_Z(z_j)$, $(u_j,\,v_j) \in J \subset [0,\,1]^2$ and ξ_j belongs to a compact set $\tilde{D} \subset \mathbb{R}$ for $j = 1, 2$. Here $C_1 > 0$ is a constant and J is the intersection of an open set and $[0,\,1]^2$.

(C11) $(n_v b)/n \to C_3$ for a constant $C_3 > 0$.

In Condition (C8), the assumption that the $\lim_{n_v\to\infty} n_v b^5 = 0$ is used to make the bias in $\hat{f}_X(x)$ asymptotically negligible. Conditions (C9) - (C10) guarantee the consistency of the copula parameter estimator $\hat{\xi}$, and Condition (C11) establishes the weak convergence of the semi-parametric estimator $\sqrt{n}(\hat{\theta} - \theta_0)$ in Theorem 3.

For simplicity, we rewrite the proposed estimating equation (7) as

$$U(\theta,\,\Lambda_0) = \sum_{i=1}^{n} \int_0^\tau \left[Q_i - \frac{\sum_{i=1}^n Y_i(t)\,\eta_i(\theta, \Lambda_0(t))\,Q_i}{\sum_{i=1}^n Y_i(t)\,\eta_i(\theta, \Lambda_0(t))} \right] dN_i(t), \tag{A.1}$$

and simple algebraic manipulation yields

$$U(\theta_0,\, \Lambda_0) = \sum_{i=1}^{n} \int_0^{\tau} \left[Q_i - \frac{\sum_{i=1}^{n} Y_i(t)\, \eta_i(\theta_0, \Lambda_0(t))\, Q_i}{\sum_{i=1}^{n} Y_i(t)\, \eta_i(\theta_0, \Lambda_0(t))} \right] dM_i(t), \tag{A.2}$$

where $Q_i = (\mu(Z_i),\, W_i^T)^T$.

Web Appendix A. Consistency of $\hat{\theta}$

We introduce the notation

$$\begin{aligned}
S^{(k)}(t;\, \theta, \Lambda) &= n^{-1} \sum_{i=1}^{n} Y_i(t)\, Q_i^{\otimes k} \eta_i(\theta, \Lambda(t)), \\
S_\theta^{(k)}(t;\, \theta, \Lambda) &= n^{-1} \sum_{i=1}^{n} Y_i(t)\, Q_i^{\otimes k} \dot{\eta}_{\theta i}(\theta, \Lambda(t)), \\
S_\Lambda^{(k)}(t;\, \theta, \Lambda) &= n^{-1} \sum_{i=1}^{n} Y_i(t)\, Q_i^{\otimes k} \dot{\eta}_{\Lambda i}(\theta, \Lambda(t)),
\end{aligned}$$

for $k = 0,\, 1,\, 2$, where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, $a^{\otimes 2} = aa^T$, $\dot{\eta}_{\theta i} = \partial \eta_i(\theta, \Lambda(t))/\partial \theta$ and $\dot{\eta}_{\Lambda i} = \partial \eta_i(\theta, \Lambda(t))/\partial \Lambda(t)$ at a fixed time t. In addition, we define $s^{(0)}(t)$, $s^{(1)}(t)$, $s_\theta^{(0)}(t)$, $s_\theta^{(1)}(t)$, $s_\Lambda^{(0)}(t)$, and $s_\Lambda^{(1)}(t)$ as the corresponding expected values at $(\theta,\, \Lambda) = (\theta_0,\, \Lambda_0)$. It then follows that $n^{-1}\, U(\theta,\, \Lambda_0)$ converges a.s. uniformly in $\theta \in \Theta$ to its limit $u(\theta,\, \Lambda_0)$. We can show that $\sup_{\theta \in \Theta} || n^{-1}\, U(\theta,\, \hat{\Lambda}) - n^{-1}\, U(\theta,\, \Lambda_0) || \to 0$ as $n \to \infty$ from the uniform convergence of $\hat{\Lambda}(t)$ to $\Lambda_0(t)$ in t, as proven by Zucker (2005, A.3), together with the uniform Lipschitz continuity of $\eta(\theta, \Lambda)$, $\dot{\eta}_\theta(\theta, \Lambda)$, and $\dot{\eta}_\Lambda(\theta, \Lambda)$ with respect to any fixed continuous Λ, which consists of monotone and bounded functions. Hence, we have that $\sup_{\theta \in \Theta} || n^{-1}\, U(\theta,\, \hat{\Lambda}) - u(\theta,\, \Lambda_0) || \to 0$ almost surely. Moreover, we can show that $-n^{-1}\, \partial U(\theta,\, \hat{\Lambda})/\partial \theta$ converges in probability to $-\partial u(\theta,\, \Lambda_0)/\partial \theta = D(\theta)$ uniformly in θ, which is positive definite at $\theta = \theta_0$ by Condition (C5). Finally, since

$$u(\theta_0,\, \Lambda_0) = E\left[\int_0^{\tau} \left\{ Q_1 - \frac{S^{(1)}(t;\, \theta_0, \Lambda_0(t))}{S^{(0)}(t;\, \theta_0, \Lambda_0(t))} \right\} Y_1(t)\, \eta_1(\theta_0, \Lambda_0(t)) d\Lambda_0(t) \right] = 0,$$

and $n^{-1}\,U(\theta_0,\,\hat{\Lambda}) \to 0$ as $n \to \infty$, by applying arguments in Foutz (1977), there exists a unique consistent solution $\hat{\theta}$ such that $U(\hat{\theta},\,\hat{\Lambda}) = 0$ with probability one.

Web Appendix B. Weak convergence of $\hat{\theta}$ when $f_X(x|z)$ is completely known

Here, we derive the asymptotic normality of $\hat{\theta}$ when $f_x(x|z)$ is completely known.

By a Taylor series expansion of $\mathrm{U}(\hat{\theta},\,\hat{\Lambda})$ at $\mathrm{U}(\theta,\,\hat{\Lambda})$ and the consistency of $\hat{\theta}$, we obtain

$$D(\theta_0)\,\sqrt{n}\,(\hat{\theta}-\theta_0) = n^{-1/2}\,U(\theta_0,\,\Lambda_0) + n^{-1/2}\,[U(\theta_0,\,\hat{\Lambda}) - U(\theta_0,\,\Lambda_0)] + o_p(1), \tag{A.3}$$

where $D(\theta_0) \equiv E\,[\int_0^\tau \{s_\theta^{(1)}/s^{(0)}(t) - s^{(1)}s_\theta^{(0)}/(s^{(0)})^2(t)\}\,dN_1(t)]$, which is consistently estimated by

$$\begin{aligned}\hat{D} &= -n^{-1}\,\partial U(\theta,\,\hat{\Lambda})/\partial\theta\,|_{\theta=\hat{\theta}} \\ &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \{S_\theta^{(1)}/S^{(0)}(t;\,\hat{\theta},\hat{\Lambda}) - S^{(1)}S_\theta^{(0)}/(S^{(0)})^2(t;\,\hat{\theta},\hat{\Lambda})\}\,dN_i(t).\end{aligned}$$

It follows from the martingale representation of $\mathrm{U}(\theta_0,\,\Lambda_0)$ in (A.2) and from the multivariate central limit theorem that $n^{-1/2}\,U(\theta_0,\,\Lambda_0)$ is asymptotically zero-mean normal with covariance matrix $D(\theta_0)$.

We derive the limiting distribution of $n^{-1/2}\,[U(\theta_0,\,\hat{\Lambda}) - U(\theta_0,\,\Lambda_0)]$ by following arguments similar to those given by Zucker (2005, A.4-A.5). Applying the mean value theorem, we obtain

$$\begin{aligned}U(\theta_0,\,\hat{\Lambda}) - U(\theta_0,\,\Lambda_0) &= \sum_{i=1}^{n}\int_0^\tau \phi(t;\,\theta_0,\Lambda_0)\,\{\hat{\Lambda}(t) - \Lambda_0(t)\}\,dN_i(t) \\ &\quad + O(\sup_{t\in[0,\tau]}||\hat{\Lambda}(t) - \Lambda_0(t)||),\end{aligned} \tag{A.4}$$

where $\phi(t;\,\theta,\Lambda) = S^{(1)}S_\Lambda^{(0)}/(S^{(0)})^2(t;\,\theta,\Lambda) - S_\Lambda^{(1)}/S^{(0)}(t;\,\theta,\Lambda)$. Now, from the Taylor series

expansion and the fact that $dM_i(t) = dN_i(t) - Y_i(t)\,\eta_i(\theta_0, \Lambda_0(t))\,d\Lambda_0(t)$, $\{\hat{\Lambda}(t) - \Lambda_0(t)\}$ can be approximated by

$$n^{-1}\sum_{i=1}^{n}\int_0^t \frac{dM_i(u)}{S^{(0)}(u;\,\theta_0,\Lambda_0)} - \frac{S_\Lambda^{(0)}}{(S^{(0)})^2}(u;\,\theta,\Lambda)\{\hat{\Lambda}(u) - \Lambda_0(u)\}\,dN_i(u),$$

which has the solution (Yang & Prentice, 1999)

$$\hat{\Lambda}(t) - \Lambda_0(t) \approx \frac{1}{R_0(t)}\sum_{i=1}^{n}\int_0^t \frac{R_0(u-)}{n\,S^{(0)}(u;\,\theta_0,\Lambda_0)}\,dM_i(u), \tag{A.5}$$

where

$$R_0(t) = \prod_{u\le t}\left\{1 + \sum_{i=1}^{n} nS_\Lambda^{(0)}/(S^{(0)})^2(u;\,\theta_0,\Lambda_0)\,dN_i(u)\right\}.$$

Since $\sup_{t\in[0,\tau]}||\hat{\Lambda}(t) - \Lambda_0(t)|| = o_p(1)$, replacing $\{\hat{\Lambda}(t) - \Lambda_0(t)\}$ in (A.4) with (A.5) yields

$$n^{-1/2}\,[U(\theta_0,\,\hat{\Lambda}) - U(\theta_0,\,\Lambda_0)] = n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \frac{G(u;\,\theta_0,\Lambda_0)\,R_0(u-)}{n\,S^{(0)}(u;\,\theta_0,\Lambda_0)}\,dM_i(u) + o_p(1),$$

where $G(u;\,\theta_0,\Lambda_0) = \sum_{i=1}^{n}\int_u^\tau \phi(t;\,\theta_0,\Lambda_0)/R_0(t)\,dN_i(t)$. This is a sum of n independent mean-zero random vectors plus an asymptotically negligible term. Therefore, by the central limit theorem, we can show that $n^{-1/2}\,[U(\theta_0,\,\hat{\Lambda}) - U(\theta_0,\,\Lambda_0)]$ converges in distribution to a mean-zero normal random vector with covariance matrix H, where

$$H = E\left[\int_0^\tau \frac{G(t;\,\theta_0,\Lambda_0)\,R_0(t-)}{\{\sum_{i=1}^{n} Y_i(t)\,\eta_i(\theta_0,\Lambda_0(t))\}^2}\,dM_1(t)\right]^{\otimes 2},$$

which can be consistently estimated by

$$\hat{H} = n^{-1}\sum_{i=1}^{n}\int_0^\tau \frac{G(t;\,\hat{\theta},\hat{\Lambda})^{\otimes 2}\,\hat{R}(t-)^2}{\{\sum_{i=1}^{n} Y_i(t)\,\eta_i(\hat{\theta},\hat{\Lambda}(t))\}^2}\,dN_i(t).$$

Finally, since the first and second terms in (A.3) are asymptotically independent by an argument similar to that given by Zucker (2005, A.5), the desired asymptotic distribution

of $\sqrt{n}\,(\hat{\theta}-\theta_0)$ can be established as a zero-mean normal distribution with covariance matrix $D^{-1}+D^{-1}HD^{-1}$.

Web Appendix C. Weak convergence of $\hat{\theta}$ when $f_X(x|z)$ is known up to a parametric form

We now study the case in which $f_X(x|z)$ belongs to a parametric family indexed by a vector parameter ξ in $\mathbb{R}^p$. That is, the conditional density $f_X(x|z)$ may be written as

$$f_X(x|z;\,\xi) = C'_{\xi_3}(F_X(x;\,\xi_1),\,F_Z(z;\,\xi_2))\,f_X(x;\,\xi_1),$$

where $\xi=(\xi_1^T,\,\xi_2^T,\,\xi_3^T)^T$, the margins F_X and F_Z and their corresponding univariate densities f_X and f_Z are indexed by parameter vectors ξ_1 and ξ_2, respectively, and C'_{ξ_3} denotes the copula density function with an unknown parameter ξ_3.

Suppose we observe $\{Z_i;\ i=1,\ldots,n\}$ in the main study and $\{X_j,Z_j;\ j=1,\ldots,n_v\}$ in the external validation study. Under our assumptions, $\{Z_i\}$ and $\{X_j,Z_j\}$ are i.i.d. random vectors, and since the log-likelihood of the measurement error model is

$$\ell(\xi)=\sum_{i=1}^{n}\log f_Z(z_i;\,\xi_2)+\sum_{j=1}^{n_v}\{\log C'_{\xi_3}(F_X(x_j;\,\xi_1),\,F_Z(z_j;\,\xi_2))+\log f_X(x_j;\,\xi_1)+\log f_Z(z_j;\,\xi_2)\},$$

$\hat{\xi}$ is the maximizer of $\ell(\xi)$. Then, following standard maximum likelihood theory, the consistency of $\hat{\xi}$ to the true value ξ_0 as well as the asymptotic normality of $\sqrt{n}\,(\hat{\xi}-\xi_0)$ with covariance matrix Ω follows.

The estimating equation for θ with unknown parameter ξ is denoted by $U(\theta,\Lambda_0,\xi)$, which can be obtained by replacing $\mu(Z_i)$, Q_i and $\eta_i(\theta,\,\Lambda_0(t))$ with $\mu(Z_i;\,\xi)$, $Q_i(\xi)$ and $\eta_i(\theta,\,\Lambda_0(t),\,\xi)$, respectively, in (A.1).

By the functional delta method and the fact that $\hat{\theta}\to\theta_0$ in probability, the estimating

equation $n^{-1/2}\,U(\hat{\theta},\,\hat{\Lambda},\,\hat{\xi})$ can be expressed as

$$\begin{aligned} 0 &= n^{-1/2}\,U(\hat{\theta},\,\hat{\Lambda},\,\hat{\xi}) \\ &= n^{-1/2}\,U(\theta_0,\,\hat{\Lambda},\,\hat{\xi}) + \frac{1}{n}\frac{\partial U(\theta_0,\,\hat{\Lambda},\,\hat{\xi})}{\partial\theta}\sqrt{n}\,(\hat{\theta}-\theta_0) + o_p(1), \end{aligned}$$

and by $\hat{\xi} \to \xi_0$,

$$n^{-1/2}\,U(\theta_0,\,\hat{\Lambda},\,\hat{\xi}) = n^{-1/2}\,U(\theta_0,\,\hat{\Lambda},\,\xi_0) + \frac{1}{n}\frac{\partial U(\theta_0,\,\hat{\Lambda},\,\xi_0)}{\partial\xi}\sqrt{n}\,(\hat{\xi}-\xi_0) + o_p(1). \tag{A.6}$$

We can demonstrate the consistency of $n^{-1}\partial U(\theta_0,\,\hat{\Lambda},\,\hat{\xi})/\partial\theta$ and $n^{-1}\partial U(\theta_0,\,\hat{\Lambda},\,\xi_0)/\partial\xi$ in the same way as shown in Web Appendix B, and hence we can show that $-n^{-1}\partial U(\theta_0,\,\hat{\Lambda},\,\hat{\xi})/\partial\theta$ converges in probability to $D(\theta_0)$, and that $n^{-1}\partial U(\theta_0,\,\hat{\Lambda},\,\xi_0)/\partial\xi$ converges in probability to $V(\xi_0)$, which can be consistently estimated by

$$\begin{aligned} \hat{V} &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \left[\dot{Q}_{\xi i}(\hat{\xi}) + \frac{\sum_i Y_i(t)\,\eta_i(\hat{\theta},\hat{\Lambda},\hat{\xi})\,Q_i(\hat{\xi})\,\sum_i Y_i(t)\,\dot{\eta}_{\xi i}(\hat{\theta},\hat{\Lambda},\hat{\xi})^T}{\{\sum_i Y_i(t)\,\eta_i(\hat{\theta},\hat{\Lambda},\hat{\xi})\}^2}\right] dN_i(t) \\ &\quad -n^{-1}\sum_{i=1}^{n}\int_0^\tau \left[\frac{\sum_i Y_i(t)\,Q_i(\hat{\xi})\,\dot{\eta}_{\xi i}(\hat{\theta},\hat{\Lambda},\hat{\xi})^T\,\sum_i Y_i(t)\,\eta_i(\hat{\theta},\hat{\Lambda},\hat{\xi})\,\dot{Q}_{\xi i}(\hat{\xi})}{\sum_i Y_i(t)\,\eta_i(\hat{\theta},\hat{\Lambda},\hat{\xi})}\right] dN_i(t), \end{aligned}$$

where $\dot{Q}_{\xi i}$ and $\dot{\eta}_{\xi i}$ are partial derivatives of $Q_i(\xi)$ and $\eta_i(\theta,\Lambda,\xi)$ with respect to ξ. Finally, since the first term and the second term in (A.6) are asymptotically independent, we have just proven that the limiting distribution of $\sqrt{n}\,(\hat{\theta}-\theta_0)$ is a mean-zero normal distribution with covariance matrix $D^{-1}(\theta_0) + D^{-1}(\theta_0)[H + V(\xi_0)\,\Omega\,V(\xi_0)^T]D^{-1}(\theta_0)$.

Web Appendix D. Weak convergence of $\hat{\theta}$ when a parametric form of $f_X(x|z)$ is unknown

When the parametric form of $f_X(x|z)$ is unknown, we propose to use a semi-parametric estimator $\hat{f}_X(x|z) = C'_{\hat{\xi}}(\hat{F}_X(x),\,\hat{F}_Z(z))\,\hat{f}_X(x)$ as described in Section 3.2. We rewrite the

estimating equation for θ to emphasize that it is a function of $\hat{f}_X(x|z)$ as follows:

$$\begin{aligned} 0 &= U(\hat{\theta},\, \hat{\Lambda},\, \hat{f}_{X|Z}) \\ &= \{U(\hat{\theta},\, \hat{\Lambda},\, \hat{f}_{X|Z}) - U(\theta,\, \hat{\Lambda},\, \hat{f}_{X|Z})\} + \{U(\theta,\, \hat{\Lambda},\, \hat{f}_{X|Z}) - U(\theta,\, \Lambda,\, \hat{f}_{X|Z})\} \\ &\quad + \{U(\theta,\, \Lambda,\, \hat{f}_{X|Z}) - U(\theta,\, \Lambda,\, f_{X|Z})\} + U(\theta,\, \Lambda,\, f_{X|Z}). \end{aligned} \tag{A.7}$$

The asymptotic properties of $\hat{\theta}$ can be established by analyzing the four terms of (A.7). If we can establish the asymptotic properties of the third term of (A.7), the asymptotic properties of the remaining terms can be found as in Web Appendices A and B, and using Lemma 1 which we will prove later. First, we focus on deriving the asymptotic properties of $\sqrt{n}\,[U(\theta,\, \Lambda,\, \hat{f}_{X|Z}) - U(\theta,\, \Lambda,\, f_{X|Z})]$.

We start by proving the consistency of $\hat{f}_X(x|z)$.

Lemma 1 *Let $f_X(x|z)$ be a continuous and bounded probability density function. Under Conditions (C8) - (C10), for any compact set $D \subset \{(x,z) \in \mathbb{R}^2 : (F_X(x), F_Z(z)) \in J\}$,*

$$\sup_{(x,z)\in D} |\hat{f}_X(x|z) - f_X(x|z)| \to 0 \quad a.s. \qquad as \quad n_v \to \infty.$$

For ease of notation, we let $F_X(x) = F_1$, $F_Z(z) = F_2$, and $f_X(x) = f_1$. Then,

$$\begin{aligned} \hat{f}_X(x|z) - f_X(x|z) &= C'_{\hat{\xi}}(\hat{F}_1,\, \hat{F}_2)\hat{f}_1 - C'_{\xi_0}(F_1,\, F_2)f_1 \\ &= \{C'_{\hat{\xi}}(\hat{F}_1,\, \hat{F}_2) - C'_{\xi_0}(F_1,\, F_2)\}\hat{f}_1 + C'_{\xi_0}(F_1,\, F_2)(\hat{f}_1 - f_1). \end{aligned} \tag{A.8}$$

Under the continuity of the distribution functions F_1 and F_2, we have $\sup_{x,z\in\mathbb{R}} |\hat{F}_j - F_j| = O(\sqrt{\log(\log n_v)/n_v})$ $(j = 1,\, 2)$ almost surely (see Shorack & Wellner, 2009, Chap 13). Then, by Conditions (C9) - (C10),

$$\sup_{(x,z)\in D} |C'_{\hat{\xi}}(\hat{F}_1,\, \hat{F}_2) - C'_{\xi_0}(F_1,\, F_2)| = O(n_v^{-1/2}\sqrt{\log n_v}) \quad \text{a.s.}$$

For x in a compact set $\tilde{D}$,

$$\sup_{x\in\tilde{D}} |\hat{f}_1 - f_1| = O((n_v b)^{-1/2}\sqrt{\log n_v} + b^p) \quad \text{a.s.}$$

for p times continuously differentiable f_1 on $\mathbb{R}$ for some $p \geq 2$ (Newey, 1994, see). Then, from (A.8),

$$\sup_{(x,z)\in D} |\hat{f}_X(x|z) - f_X(x|z)| = O((n_v b)^{-1/2}\sqrt{\log n_v} + b^p).$$

Therefore, by Condition (C8), Lemma 1 holds.

We next establish the asymptotic normality of $\hat{f}_X(x|z)$.

Lemma 2 *Suppose that f_X is twice continuously differentiable at $x \in \mathbb{R}$. Under Conditions (C8) - (C10), for any compact set $D \subset \{(x,z) \in \mathbb{R}^2 : (F_X(x), F_Z(z)) \in J\}$, we have*

$$\sqrt{n_v b}\,(\hat{f}_X(x|z) - f_X(x|z)) \xrightarrow{\mathcal{D}} N(0,\, \Sigma),$$

where $\Sigma = C'^2_{\xi_0}(F_X,\, F_Z)\, f_X \int K^2(u)\, du$.

From Conditions (C9) - (C10) and using the consistency of the density estimator $\hat{f}_1$, we have $|\{C'_{\hat{\xi}}(\hat{F}_1,\, \hat{F}_2) - C'_{\xi_0}(F_1,\, F_2)\}\hat{f}_1| = O_p(n_v^{-1/2}\sqrt{\log n_v})$. Thus, from (A.8),

$$\hat{f}_X(x|z) - f_X(x|z) = C'_{\xi_0}(F_1,\, F_2)(\hat{f}_1 - f_1) + O_p(n_v^{-1/2}\sqrt{\log n_v}). \tag{A.9}$$

The next step is to show the asymptotic normality of $\sqrt{n_v b}\,(\hat{f}_1 - f_1)$. Let $w_j = b^{-1/2}K((x - X_j)/b) - b^{-1/2}EK((x - X_j)/b)$. Under the assumptions that $n_v b \to \infty$ and $\sqrt{n_v b^5} \to 0$ as

$n_v \to \infty$, we can show

$$\begin{aligned}
\sqrt{n_v b}\,(\hat{f}_1 - f_1) &= n_v^{-1/2} \sum_{j=1}^{n_v} \{b^{-1/2} K((x - X_j)/b) - b^{-1/2} EK((x - X_j)/b)\} \\
&\quad + \sqrt{n_v b}\,\{b^{-1} EK((x - X_1)/b) - f_X(x)\} \\
&= n_v^{-1/2} \sum_{j=1}^{n_v} w_j + O_p((n_v b)^{1/2}\, b^2) \\
&= n_v^{-1/2} \sum_{j=1}^{n_v} w_j + o_p(1),
\end{aligned}$$

since $b^{-1} EK((x - X_j)/b) - f_X(x) = O(b^2)$. Now, we can show that as $n_v \to \infty$,

$$\begin{aligned}
Ew_j^2 &= b^{-1} EK^2((x - X_j)/b) - b^{-1}\{EK(x - X_j)/b)\}^2 \\
&= b^{-1} \int K^2((x - u)/b)\, f_X(u)\, du - b^{-1}\{b \int K(u)\, f_X(x - ub)\, du\}^2 \\
&= \int K^2(u)\, f_X(x - ub)\, du + O(b) \\
&\longrightarrow f_X(x) \int K^2(u)\, du
\end{aligned}$$

by the dominated convergence theorem. We can further show that for independent $w_j's\,(j = 1, \ldots, n_v)$, $Ew_j = 0$ and $n_v^{-\tilde{\delta}/2} E|w_j|^{2+\tilde{\delta}} \to 0$ for some $\tilde{\delta} > 0$. Therefore, by the Lyapunov central limit theorem, we obtain

$$\sqrt{n_v b}\,(\hat{f}_1 - f_1) \xrightarrow{\mathcal{D}} N(0,\, f_1 \int K^2(u)\, du).$$

Hence, from (A.9), we have proved that $\sqrt{n_v b}\,(\hat{f}_X(x|z) - f_X(x|z))$ follows a mean-zero normal distribution with covariance matrix $C'^2_{\xi_0}(F_1,\, F_2)\, f_1 \int K^2(u)\, du$.

Finally, using the functional delta method, we can show $\sqrt{n}\,[U(\theta,\, \Lambda,\, \hat{f}_{X|Z}) - U(\theta,\, \Lambda,\, f_{X|Z})]$ weakly converges to $U'_f(M)$, where U'_f is the Hadamard derivative of $U(\theta,\, \Lambda,\, f_{X|Z})$ at $f_{X|Z}$, and M is a random variable following the same distribution as the limiting distribution of $\sqrt{n_v b}\,\{\hat{f}_X(x|z) - f_X(x|z)\}$, i.e., a mean-zero normal with the covariance matrix Σ, assuming

$(n_v b)/n \to C_3$ for a constant $C_3 > 0$. Thus, the proof of the Theorem 3 is completed by combining this result with the asymptotic normality that have been proven for the other terms in (A.7).

References

Foutz, R. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association* **72**, 147–148.

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory* **10**, 233–253.

Shorack, G. R. & Wellner, J. A. (2009). *Empirical processes with applications to statistics*. J. Wiley & Sons, New York.

Yang, S. & Prentice, R. (1999). Semiparametric inference in the proportional odds regression model. *Journal of the American Statistical Association* **94**, 125–136.

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