Analysis of Proportional Odds Models with Censoring and Errors-in-Covariates

Samiran SINHA and Yanyuan MA*

Abstract

We propose a consistent method for estimating both the finite and infinite dimensional parameters of the proportional odds model when a covariate is subject to measurement error and time-to-events are subject to right censoring. The proposed method does not rely on the distributional assumption of the true covariate which is not observed in the data. In addition, the proposed estimator does not require the measurement error to be normally distributed or to have any other specific distribution, and we do not attempt to assess the error distribution. Instead, we construct martingale based estimators through inversion, using only the moment properties of the error distribution, estimable from multiple erroneous measurements of the true covariate. The theoretical properties of the estimators are established and the finite sample performance is demonstrated via simulations. We illustrate the usefulness of the method by analyzing a dataset from a clinical study on AIDS.

Key Words: Estimating equations; Functional approach; Martingale; Measurement error; Proportional odds model; U-statistics.

Short title: Proportional Odds Model with Error

*Samiran Sinha (sinha@stat.tamu.edu) is Associate Professor and Yanyuan Ma (ma@stat.tamu.edu) is Professor, Department of Statistics, Texas A&M University, 3143 TAMU, College Station TX 77843-3143.
The authors are grateful to the review team for insightful comments that led to a significant improvement of the manuscript. The work was partially supported by the National Science Foundation (DMS-1206693), the National Institute of Neurological Disorders and Stroke (R01-NS073671), and the National Cancer Institute (R03-CA176760).

1 Introduction

We consider the proportional odds model when the time to event is subject to right censoring and a covariate is measured with errors. Proportional odds model is a widely used model in survival analysis as an alternative to the popular Cox proportional hazard model. In comparison with the Cox model which assumes that the ratio of the hazards corresponding to different covariate values does not change with time, the proportional odds model allows the hazard ratio to vary over time. Time varying hazards ratios can arise frequently in practice. For example, the relative effect of the stages of a cancer at the time of diagnosis on survival may change with time. In studying the proportional odds model with right censored data, Murphy, Rossini and van der Vaart (1997) proposed a nonparametric maximum likelihood estimator. Huang (1995) and Rossini and Tsiatis (1996) constructed consistent estimators for current status data. Cheng, Wei and Ying (1995) used an estimating equation based approach in the linear transformation model, which includes the proportional odds model as a special case, for right censored data.

Despite of the large literature in proportional odds model when covariates are measured precisely, relatively few works are available in this model when covariates are measured with errors. Cheng and Wang (2001) considered the measurement error issue in the linear transformation model, but their method requires a parametric model for the pairwise difference between the true covariate values of any two subjects, a parametric model for the pairwise difference between the measurement errors of any two subjects, and similar supports of the censoring distribution and the time-to-event distribution. Thus, the method would fail to produce consistent estimators if any of three model assumptions is violated. For current status data, Wen and Chen (2012) proposed a conditional score method for handling errors in covariate in proportional odds models under the assumption that the measurement errors follow a normal distribution. This is in stark contrast with the situation in the Cox model, where extensive studies of errors in covariate have been conducted, see for example Prentice (1982), Nakamura (1992), Huang and Wang (2000), Zhou and Wang (2000), Hu and Lin (2002) and Zucker (2005).

In this article, we propose a semiparametric method to treat errors in covariates in the proportional odds model when the events are subject to right censoring. We first construct a class of estimating equations through designing special martingale integrals under the error free case. The design of the estimating equation class further allows us to invert these estimating equations when covariates are measured with errors. This type of treatment to measurement error models is commonly known as the "corrected score" approach. Despite of the name, the technique is applicable to general estimating functions that are not necessarily score functions in the error-free cases (Nakamura 1990). For example, Huang and Wang (2001) applied this approach to the logistic regression model. Buzas (1998) used this approach to correct the partial likelihood score to estimate regression parameters in the Cox proportional hazard model while assuming the measurement errors follow a normal distribution. Huang and Wang (2000) further relaxed the normality assumption on the measurement errors. Using an empirical process approach, they obtained a consistent and asymptotically normal estimator while the measurement errors are assumed to satisfy some minor regularity conditions. Song and Huang (2005) further refined the parametric and nonparametric corrected score method of Huang and Wang (2000) to achieve better finite sample properties. Although all these methods are based on the general idea of "corrected scores", the implementation of the idea in different models requires very different model specific treatment and techniques that can be difficult and by no means straightforward. In addition, the theoretical properties in different models can also be quite different and need to be studied individually and can be challenging depending on the specificity of the models. This also applies to the new method proposed in this article. In other words, the distinction between our work and the work of Huang and Wang (2000, 2001) is rooted in the different models that are considered in these works. Their subsequent estimation procedure, methodological development, theoretical properties and numerical implementation are in turn all different from ours.

One advantage of the proposed method is that we do not make any assumption on the distribution of the errors other than symmetry, and we do not make any distributional assumptions on the true covariate prone to errors, hence we work in the functional measurement error framework (Carroll et al., 2006). This is in stark contrast with Zucker (2005), which requires a correct model for the unobserved covariate given the observed covariates, hence is essentially a structural model (Carroll et al., 2006).

In summary, the proposed estimator is applicable in relatively weak assumptions, requiring only symmetric error distribution, making no distributional assumption on the covariate measured with error, allowing censoring dependent of the covariates and very large censoring proportion. In such generality, this is the only existing consistent estimator for proportional odds models. The critical idea of the estimator relies on constructing a martingale based estimating equation that is not naturally derived from the standard score function consideration, but has the key advantage of being invertable when measurement error presents. The asymptotic analysis of the estimating equation requires techniques involving martingale, nonparametric and semiparametric analysis.

The rest of the paper is organized as follows. We describe the details of the methodology in Section 2 and study the asymptotic properties of the estimator in Section 3. In Section 3, we also provide a method of estimating the asymptotic variance of the proposed estimator. We evaluate the finite sample performance of our estimator via simulation studies in Section 4. To illustrate the usefulness of the method, in Section 5, we analyze a data set from an AIDs clinical trial. Concluding remarks are given in Section 6 while all technical details are relegated to the Supplementary materials.

2 Methodology

2.1 Model

Suppose that the observed data are independent and identically distributed (iid) copies of $(V, \Delta, W_1^*, \cdots, W_m^*, \mathbf{Z})$, where $V = \min(T, C)$ is the minimum of the time-to-event T and the censoring time C, and $\Delta = I(T \leq C)$. Here $\mathbf{Z}$ is a $p \times 1$ vector of covariates measured precisely, while X is not observed. Instead of X, m repeated measurements of an unbiased surrogate W^* of X are available. We assume that T and C are independent conditional on

$(\mathbf{Z}, X)$. Let T be related to the covariates via the proportional odds model

$$\mathrm{pr}(T \leq t | \mathbf{Z}, X) = \frac{\Lambda(t) \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{Z} + \beta_2 X)}{1 + \Lambda(t) \exp(\boldsymbol{\beta}_1 \mathbf{Z} + \beta_2 X)}, \tag{1}$$

where $\Lambda(t)$ is a non-decreasing right-continuous function with $\Lambda(0) = 0$. Let $\Lambda(t-)$ be the left-hand limit of Λ at t. Define $\lambda(t) = \partial \Lambda(t)/\partial t$ if Λ is differentiable, otherwise $\lambda(t) \equiv \Lambda(t) - \Lambda(t-)$. Our interest is in consistent estimation of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\mathrm{T}}, \beta_2)^{\mathrm{T}}$ and Λ. To this end, we first propose a novel estimating equation when there is no measurement errors. We then modify this estimating equations when X is measured with errors.

2.2 Error free estimator

Define $\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{Z}_i + \beta_2 X_i)$, $N_i(u) = I(V_i \leq u, \Delta_i = 1)$ and $Y_i(u) = I(V_i \geq u)$. Without loss of generality, we assume $0 < V_1 \leq V_2 \leq \cdots \leq V_n < \tau < \infty$, where $\tau = \inf\{t : \mathrm{pr}(V > t) = 0\}$. Then,

$$M(t) = N(t) - \int_0^t Y(u) \frac{\lambda(u) \eta(X, \mathbf{Z}, \boldsymbol{\beta})}{1 + \Lambda(u) \eta(X, \mathbf{Z}, \boldsymbol{\beta})} du$$

is a martingale with respect to filtration $\{\mathcal{F}_t : t \geq 0\}$, where $\mathcal{F}_t = \sigma\{Y(u), N(u), X, \mathbf{Z}, u \leq t\}$. Consider the situation that X is observed in the data. Then one may consistently estimate $\boldsymbol{\beta}$ and Λ by solving $\mathbf{S}_{\beta_1} = \mathbf{0}$, $S_{\beta_2} = 0$, and $S_\Lambda(u) = 0$ for all $u \geq 0$, where for any function $f(\Lambda, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ predictable with respect to $\{\mathcal{F}_t : t \geq 0\}$ with $\boldsymbol{\alpha}$ being possible additional parameters, we define

$$\begin{aligned}
\mathbf{S}_{\beta_1} &= \sum_{i=1}^n \int_0^\tau \mathbf{Z}_i \{1 + \Lambda(u) \eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(u), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \left\{ dN_i(u) - \frac{Y_i(u) \lambda(u) \eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta}) du}{1 + \Lambda(u) \eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})} \right\} \\
&= \sum_{i=1}^n (\mathbf{Z}_i \Delta_i \{1 + \Lambda(V_i) \eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(V_i), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \\
&\quad - \mathbf{Z}_i \eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta}) \left[F\{\Lambda(V_i), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} - F(0, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}) \right]), \qquad (2) \\
S_{\beta_2} &= \sum_{i=1}^n (X_i \Delta_i \{1 + \Lambda(V_i) \eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(V_i), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \\
&\quad - X_i \eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta}) \left[F\{\Lambda(V_i), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} - F(0, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}) \right]), \qquad (3) \\
S_\Lambda(u) &= \sum_{i=1}^n \{1 + \Lambda(u) \eta(X_i, Z_i, \boldsymbol{\beta})\} \left\{ dN_i(u) - Y_i(u) \frac{\lambda(u) \eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta}) du}{1 + \Lambda(u) \eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})} \right\}
\end{aligned}$$

$$= \sum_{i=1}^{n} \left[\{1 + \Lambda(u)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\} dN_i(u) - Y_i(u)\lambda(u)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta}) du \right], \text{ for all } u > 0. \qquad (4)$$

Here $F(\Lambda, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ satisfies $\partial F(\Lambda, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})/\partial \Lambda = f(\Lambda, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})$. Assuming that the observed failure times are $0 < t_{n_1} < \cdots < t_{n_k}$, then from (4) we obtain

$$S_\Lambda(t_{n_1}) = \sum_{i=1}^{n} \{1 + \Lambda(t_{n_1})\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\} dN_i(t_{n_1}) - \sum_{i=1}^{n} Y_i(t_{n_1})\{\Lambda(t_{n_1}) - \Lambda(t_{n_1}-)\}\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta}),$$
$$\vdots$$
$$S_\Lambda(t_{n_k}) = \sum_{i=1}^{n} \{1 + \Lambda(t_{n_k})\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\} dN_i(t_{n_k}) - \sum_{i=1}^{n} Y_i(t_{n_k})\{\Lambda(t_{n_k}) - \Lambda(t_{n_k}-)\}\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta}).$$

Using $\Lambda(t_{n_1}-) = 0$ in $S_\Lambda(t_{n_1}) = 0$, we obtain

$$\widehat{\Lambda}(t_{n_1}) = \frac{\sum_{i=1}^{n} dN_i(t_{n_1})}{\sum_{i=1}^{n} \eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\{Y_i(t_{n_1}) - dN_i(t_{n_1})\}},$$

and $\Lambda(t_{n_j})$'s can be estimated recursively as

$$\widehat{\Lambda}(t_{n_j}) = \frac{\sum_{i=1}^{n} dN_i(t_{n_j}) + \widehat{\Lambda}(t_{n_{(j-1)}}) \sum_{i=1}^{n} Y_i(t_{n_j})\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})}{\sum_{i=1}^{n} \{Y_i(t_{n_j}) - dN_i(t_{n_j})\}\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})}, \text{ for } j = 1, \cdots, k.$$

We did not include $f\{\Lambda(u), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}$ in $S_\Lambda(u)$, which simplifies the computation in obtaining $\widehat{\Lambda}(t, \boldsymbol{\beta})$. When the last observation happens to be an event, we replace $\widehat{\Lambda}(t_{n_k})$ with a large value, larger than $\widehat{\Lambda}(t_{n_{k-1}})$, to facilitate further analysis.

We point out that these estimating equations are the building blocks of our method, and form one of the important contributions of our work. In addition, the estimating equations involve both finite and infinite dimensional parameters, hence the derivation of the subsequent asymptotic theory is much more challenging. This is in contrast with Huang and Wang (2000), who benefits from the existing partial likelihood score functions, which do not involve infinite dimensional parameters, and hence are relatively easy to analyze.

Remark 1. We have left $f\{\Lambda(u), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}$ to be an arbitrary function in the above description. The flexibility in choosing $f\{\Lambda(u), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}$ leads to a broad class of consistent estimators. In the regularity condition C1, we specify the requirement on f so that the estimating equations will lead to a unique estimator in large samples. Note that when there is no measurement error, the score functions for the maximum likelihood estimator (Murphy et al., 1997) are obtained if we replace $f\{\Lambda(u), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}$ by $1/\{1+\Lambda(u)\eta(X, \mathbf{Z}, \boldsymbol{\beta})\}^2$ and multiply each summand of S_Λ by $1/\{1+\Lambda(u)\eta(X, \mathbf{Z}, \boldsymbol{\beta})\}$. However, the presence of X in the expression $1/\{1+\Lambda(u)\eta(X, \mathbf{Z}, \boldsymbol{\beta})\}^2$ will cause difficulties as soon as X becomes unobservable. To circumvent this issue we shall take f free-from X, so that the resulting estimating equations are invertible and we can construct "corrected" estimating equations in the presence of measurement errors. In the next subsection we discuss the choices of f when X is unobserved, and discusss the concept of "corrected" estimating equations.

2.3 Estimator under measurement error

Now we consider the case when X is not observed in the data, and instead, we observe a surrogate variable W^* multiple times, such that

$$W^*_{ij} = X_i + U^*_{ij},\ j = 1, \ldots, m,\ i = 1, \ldots, n.$$

Here the U^*_{ij}'s are iid copies of the random variable U^* that is symmetrically distributed. Furthermore, U^* is assumed to be independent of $V, \Delta, X, \mathbf{Z}$, a commonly used assumption (Huang and Wang, 2000). Define $W_i = m^{-1}\sum_{j=1}^m W^*_{ij}$. Following Li and Vuong (1998), the pdf of U^*_{ij} and X_i are both identifiable. Thus, the likelihood of a single observation $(W_{i1}, \mathbf{Z}_i, Y_i, \Delta_i)$ has the form

$$\left\{\int f_{Y|\mathbf{Z},X}(y_i, \mathbf{z}_i, x_i) f_U(w_{i1} - x_i) f_X(x_i) dx_i\right\}^{\Delta_i} \left\{\int S_{Y|\mathbf{Z},X}(y_i, \mathbf{z}_i, x_i) f_U(w_{i1} - x_i) f_X(x_i) dx_i\right\}^{1-\Delta_i}.$$

This can be viewed as a convolution of $f_{Y|\mathbf{Z},X}(y_i, \mathbf{z}_i, \cdot) f_X(\cdot)$ with $f_U(\cdot)$ when $\Delta_i = 1$, or a convolution of $S_{Y|\mathbf{Z},X}(y_i, \mathbf{z}_i, \cdot) f_X(\cdot)$ with $f_U(\cdot)$ when $\Delta_i = 0$. Thus, via deconvolution we can show that the Fourier transform of $f_{Y|\mathbf{Z},X}(y_i, \mathbf{z}_i, \cdot) f_X(\cdot)$ or $S_{Y|\mathbf{Z},X}(y_i, \mathbf{z}_i, \cdot) f_X(\cdot)$ is unique, hence $f_{Y|\mathbf{Z},X}(y_i, \mathbf{z}_i, x_i)$ is unique if $\Delta_i = 1$, and $S_{Y|\mathbf{Z},X}(y_i, \mathbf{z}_i, x_i)$ is unique if $\Delta_i = 0$. Thus, we obtain the identifiability of $\boldsymbol{\beta}$ and Λ. Now we propose to estimate $\boldsymbol{\beta}$ and Λ by solving

$$\begin{aligned}\mathbf{S}^{\mathrm{me}}_{\beta_1} = \sum_{i=1}^n &(\Delta_i \mathbf{Z}_i \{1 + \Lambda(V_i) g_1(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(V_i), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \\ &- \mathbf{Z}_i g_1(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) \left[F\{\Lambda(V_i), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} - F(0, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha})\right]) = \mathbf{0},\end{aligned}$$

$$
\begin{aligned}
S^{\text{me}}_{\beta_2} &= \sum_{i=1}^{n} (\Delta_i\{W_i + \Lambda(V_i) g_2(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(V_i), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \\
&\quad - g_2(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) \left[F\{\Lambda(V_i), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} - F(0, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha})\right]) = 0, \qquad (5) \\
S^{\text{me}}_{\Lambda} &= \sum_{i=1}^{n} [\{1 + \Lambda(u) g_1(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} dN_i(u) - Y_i(u)\lambda(u) g_1(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) du] = 0,
\end{aligned}
$$

where

$$
g_1(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) = \frac{\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})}{\gamma_1}, \;\; g_2(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) = \frac{\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})}{\gamma_1^2}(\gamma_1 W - \gamma_2),
$$

$\gamma_1 = E\{\exp(\beta_2 U_i)\}$, $\gamma_2 = E\{U_i \exp(\beta_2 U_i)\}$, and $U_i = \sum_{j=1}^{m} U^*_{ij}/m$. It is easy to verify that $E(g_1 \mid X, \mathbf{Z}) = \eta(X, \mathbf{Z}, \boldsymbol{\beta})$ and $E(g_2 \mid X, \mathbf{Z}) = X\eta(X, \mathbf{Z}, \boldsymbol{\beta})$. Consequently $E(\mathbf{S}^{\text{me}}_{\beta_1} | V, \Delta, X, \mathbf{Z}) = \mathbf{S}_{\beta_1}$, $E(S^{\text{me}}_{\beta_2} | V, \Delta, X, \mathbf{Z}) = S_{\beta_2}$, and $E(S^{\text{me}}_{\Lambda} | V, \Delta, X, \mathbf{Z}) = S_{\Lambda}$. The last three equalities lead to the notion of "corrected score", in the sense that the effect of the measurement error is corrected because the original "scores" are recovered via the intermediate conditional expectation step. As a result, as long as the original "scores" have mean zero, the "corrected" ones will also yield a consistent estimator.

Here we take $f\{\Lambda(u), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\} = 1/\{1 + \Lambda(u)\eta(X^*, \mathbf{Z}, \boldsymbol{\beta})\}^2$, where $E^*(X \mid \mathbf{Z})$ indicates the expectation of X conditional on $\mathbf{Z}$ calculated using a proposed model for X given $\mathbf{Z}$. This is a logical choice for $f\{\Lambda(u), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}$ and it bears similar spirit as the regression calibration idea (Carroll et al., 2006, Chapter 4). If we knew the distribution of X given $\mathbf{Z}$, a natural replacement of X would be $E(X \mid \mathbf{Z})$. Since we do not make any distributional assumption regarding X, we adopt a proposed model, which may be mis-specified, and replace the unobservable X with the corresponding conditional mean of X under the proposed model. However, unlike in the classical regression calibration treatment, our estimator will remain consistent whether the proposed model is correct or incorrect. This choice of $f\{\Lambda(u), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}$ is our recommended choice in practice. It is important to note that the method is consistent for any X^* that is a function of $\mathbf{Z}$.

To obtain X^*, one can further bypass the specification of a model for the distribution of X given $\mathbf{Z}$, and directly assume a model $E^*(X \mid \mathbf{Z}) = \mu(\mathbf{Z}, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is the additional parameter of the model if necessary. In this case, a natural estimator of $\boldsymbol{\alpha}$ can be obtained through solving

$$\sum_{i=1}^{n} \frac{\partial \mu(\mathbf{Z}_i, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \{W_i - \mu(\mathbf{Z}_i, \boldsymbol{\alpha})\} = \mathbf{0}. \tag{6}$$

Again, we point out that in fact, any arbitrary choice of $\boldsymbol{\alpha}$ will lead to a consistent estimator for $\boldsymbol{\beta}$ and Λ, hence the procedure is very robust.

2.4 Estimation of γ_1 and γ_2

To make use of the estimating equations in (5), we need to estimate γ_1 and γ_2. Observe that $\gamma_1 = E\{\exp(\beta_2 U_i)\} = \{\mathcal{M}(\beta_2/m)\}^m$, where $\mathcal{M}(\cdot)$ denotes the moment generating function of U_{ij}^*. Due to the symmetry assumption of the distribution of U_{ij}^*, $\mathcal{M}(\beta_2/m) = (2\sum_{j,k=1,j<k}^m E\left[\exp\{(W_{ij}^* - W_{ik}^*)\beta_2/m\}\right]/m(m-1))^{1/2}$. Therefore, we estimate γ_1 by

$$\widehat{\gamma}_1 = \left[\frac{2}{nm(m-1)} \sum_{j,k=1,j<k}^{m} \sum_{i=1}^{n} \exp\{(W_{ij}^* - W_{ik}^*)\beta_2/m\}\right]^{m/2}. \tag{7}$$

Further, since $\gamma_2 = E\{U_i \exp(\beta_2 U_i)\} = \partial E\{\exp(\beta_2 U_i)\}/\partial \beta_2$, we can write γ_2 as

$$\frac{1}{2} E^{(m/2-1)}\left[\frac{\sum_{j,k=1,j<k}^{m} 2\exp\{(W_{ij}^* - W_{ik}^*)\beta_2/m\}}{m(m-1)}\right] E\left[\frac{\sum_{j,k=1,j<k}^{m} 2(W_{ij}^* - W_{ik}^*)\exp\{(W_{ij}^* - W_{ik}^*)\beta_2/m\}}{m(m-1)}\right],$$

and we estimate γ_2 by

$$\widehat{\gamma}_2 = \left(\widehat{\gamma}_1\right)^{(m-2)/m} \times \frac{1}{nm(m-1)} \sum_{j,k=1,j<k}^{m} \sum_{i=1}^{n} (W_{ij}^* - W_{ik}^*)\exp\{(W_{ij}^* - W_{ik}^*)\beta_2/m\}. \tag{8}$$

A detailed derivation of $\widehat{\gamma}_2$ is given in the Supplementary materials S1. Now we are in the position to describe the steps of estimating the model parameters $\boldsymbol{\beta}$ and Λ in detail.

2.5 The complete estimation procedure

Taking into account the above derivations, we propose the complete estimation procedure as the following:

Step 0. Form $W_i = m^{-1}\sum_{j=1}^m W_{ij}^*$ for $i = 1, \ldots, n$. Obtain $\widehat{\boldsymbol{\alpha}}$ through solving (6).

Step 1. Form $\widehat{\gamma}_1(\boldsymbol{\beta})$ and $\widehat{\gamma}_2(\boldsymbol{\beta})$, both are functions of $\boldsymbol{\beta}$, following (7) and (8).

Step 2. For fixed $\boldsymbol{\beta}$ and $\widehat{\gamma}_1(\boldsymbol{\beta})$, form

$$\widehat{\Lambda}\{t_{n_1}; \boldsymbol{\beta}, \widehat{\gamma}_1(\boldsymbol{\beta})\} = \frac{\sum_{i=1}^n \widehat{\gamma}_1(\boldsymbol{\beta}) dN_i(t_{n_1})}{\sum_{i=1}^n \eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\{Y_i(t_{n_1}) - dN_i(t_{n_1})\}}$$

and

$$\widehat{\Lambda}\{t_{n_j},\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\} = \frac{\sum_{i=1}^n \{\widehat{\gamma}_1(\boldsymbol{\beta})dN_i(t_{n_j}) + Y_i(t_{n_j})\widehat{\Lambda}\{t_{n_{j-1}},\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\}\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\}}{\sum_{i=1}^n \eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\{Y_i(t_{n_j}) - dN_i(t_{n_j})\}}$$

as functions of $\boldsymbol{\beta}$ for $j = 2,\ldots,k$. These are the results from solving $S_\Lambda^{\mathrm{me}}\{u;\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\} = 0$ sequentially at $u = t_{n_1},\ldots,t_{n_k}$.

Step 3. We obtain $\widehat{\boldsymbol{\beta}}$ through solving

$$\sum_{i=1}^n \boldsymbol{\phi}[\mathbf{O}_i;\boldsymbol{\beta},\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\widehat{\boldsymbol{\alpha}}] = \mathbf{0},$$

where $\mathbf{O}_i = (W_i,\mathbf{Z}_i,V_i,\Delta_i)$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^{\mathrm{T}},\phi_2)^{\mathrm{T}}$, and

$$\begin{aligned}
\boldsymbol{\phi}_1\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma},\boldsymbol{\alpha}\} &= \mathbf{Z}_i\Delta_i\{\gamma_1 + \Lambda(V_i)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\}f\{\Lambda(V_i),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\} \\
&\quad -\mathbf{Z}_i\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})[F\{\Lambda(V_i),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\} - F(0,\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha})], \\
\phi_2\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma},\boldsymbol{\alpha}\} &= \Delta_i\{W_i\gamma_1^2 + \Lambda(V_i)(\gamma_1 W_i - \gamma_2)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\}f\{\Lambda(V_i),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\} \\
&\quad -(\gamma_1 W_i - \gamma_2)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\left[F\{\Lambda(V_i),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\} - F(0,\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha})\right].
\end{aligned}$$

Step 4. Go to Steps 1 and 2 to obtain $\gamma_1(\widehat{\boldsymbol{\beta}})$ and $\widehat{\Lambda}\{u,\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\}$ respectively.

In Step 3, $\widehat{\gamma}_1(\boldsymbol{\beta})$, $\widehat{\gamma}_2(\boldsymbol{\beta})$ and $\widehat{\Lambda}\{t_{n_j},\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\}$ are functions of $\boldsymbol{\beta}$, hence the resulting estimating equations $\sum_{i=1}^n \boldsymbol{\phi}[\mathbf{O}_i;\boldsymbol{\beta},\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\widehat{\boldsymbol{\alpha}}] = \mathbf{0}$ contain $\boldsymbol{\beta}$ as the only unknown quantity and are solved to obtain $\widehat{\boldsymbol{\beta}}$, and the estimator is referred to as error corrected estimator. This estimation procedure is a typical profiling procedure, hence we do not need to iterate the above steps. One can of course choose to use a backfitting procedure instead of profiling, where iteratively solving for $\boldsymbol{\beta}$ at fixed $\widehat{\Lambda}$, $\widehat{\gamma}_1$, $\widehat{\gamma}_2$, and solving for γ_1,γ_2 and Λ at fixed $\widehat{\boldsymbol{\beta}}$ will be required.

To solve the estimating equations in Step 3, we used a standard Newton-Raphson procedure which requires an initial value for $\boldsymbol{\beta}$. In both the simulation and the data example, we used the classical regression calibration estimates as the initial value. We also experimented with using the naive estimator as the initial value and the results are identical.

3 Asymptotic properties

3.1 Asymptotic properties

To present the asymptotic properties of the proposed error corrected method, we first need to introduce some necessary notations. For any vector or matrix a, we denote aa^{T} by $a^{\otimes 2}$, and we use $f_{\boldsymbol{\beta}}(\Lambda, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ and $F_{\boldsymbol{\beta}}(\Lambda, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ to represent the partial derivative of f and F with respect to $\boldsymbol{\beta}$. Define $\boldsymbol{\gamma} = [\gamma_1(\beta_2), \gamma_2(\beta_2)]^{\mathrm{T}} = (\gamma_1, \gamma_2)^{\mathrm{T}}$, $\kappa_1 = E\{\exp(2\beta_2 U)\}$, $\kappa_2 = E\{U \exp(2\beta_2 U)\}$ and let $\mathbf{f}_\gamma = (f_{\gamma,1}, f_{\gamma,2})^{\mathrm{T}}$ with

$$
\begin{aligned}
f_{\gamma,1}(\mathbf{W}_i^*, \boldsymbol{\beta}) =& \frac{\mathcal{M}^{(m-2)}(\beta_2/m)}{m-1} \sum_{j,k=1,j<k}^{m} \left[\exp\left\{\frac{(W_{ij}^* - W_{ik}^*)\beta_2}{m}\right\} - \mathcal{M}^2\left(\frac{\beta_2}{m}\right)\right], \\
f_{\gamma,2}(\mathbf{W}_i^*, \boldsymbol{\beta}) =& \left\{\frac{m(m-2)}{4}\right\} \mathcal{M}^{(m-4)}\left(\frac{\beta_2}{m}\right) \left[\frac{2}{m(m-1)} \sum_{j,k=1,j<k}^{m} \exp\left\{\frac{(W_{ij}^* - W_{ik}^*)\beta_2}{m}\right\}\right. \\
& \left. - \mathcal{M}^2\left(\frac{\beta_2}{m}\right)\right] \frac{\partial \mathcal{M}^2(\beta_2/m)}{\partial \beta_2} + \mathcal{M}^{(m-2)}\left(\frac{\beta_2}{m}\right) \left[\frac{1}{m(m-1)}\right. \\
& \times \left. \sum_{j,k=1,j<k}^{m} (W_{ij}^* - W_{ik}^*) \exp\left\{\frac{(W_{ij}^* - W_{ik}^*)\beta_2}{m}\right\} - \frac{m}{2} \frac{\partial \mathcal{M}^2(\beta_2/m)}{\partial \beta_2}\right].
\end{aligned}
$$

Further define $C_1(s) = E\{Y(s)\eta(W, \mathbf{Z}, \boldsymbol{\beta})\}$, $C_2(s) = E[Y(s)\eta(W, \mathbf{Z}, \boldsymbol{\beta})\lambda(s)\eta(X, \mathbf{Z}, \boldsymbol{\beta})/\{1 + \Lambda(s)\eta(X, \mathbf{Z}, \boldsymbol{\beta})\}] = E\{\eta(W, \mathbf{Z}, \boldsymbol{\beta}) dN(s)/ds\}$, $C_3(s) = E\{dN(s)/ds\}$, $\mathbf{C}_4(s) = E\{(\mathbf{Z}^{\mathrm{T}}, W)^{\mathrm{T}} Y(s)\eta(W, \mathbf{Z}, \boldsymbol{\beta})\}$, $\mathbf{C}_5(s) = E[(\mathbf{Z}^{\mathrm{T}}, W)^{\mathrm{T}} Y(s)\eta(W, \mathbf{Z}, \boldsymbol{\beta})\lambda(s)\eta(X, \mathbf{Z}, \boldsymbol{\beta})/\{1+\Lambda(s)\eta(X, \mathbf{Z}, \boldsymbol{\beta})\}] = E\{(\mathbf{Z}^{\mathrm{T}}, W)^{\mathrm{T}} Y(s)\eta(W, \mathbf{Z}, \boldsymbol{\beta})\, dN(s)/ds\}$, $D_1(s) = \exp[-\int_0^s \{C_2(u)/C_1(u)\} du]$, $D_2(s) = \int_0^s [D_1(u) C_3(u)/\{D_1(s) C_1(u)\}] du$, $\mathbf{D}_3(s) = \int_0^s \{D_1(u)[\Lambda(u, \boldsymbol{\beta}, \gamma_1)\{C_1(u)\mathbf{C}_5(u) - C_2(u)\mathbf{C}_4(u)\} - \gamma_1 C_3(u) \mathbf{C}_4(u)]\}\{D_1(s) C_1^2(u)\}^{-1} du$, $\boldsymbol{\phi}_\gamma = E[\partial \boldsymbol{\phi}\{\mathbf{O}; \boldsymbol{\beta}, \Lambda(V), \boldsymbol{\gamma}, \boldsymbol{\alpha}\}/\partial \boldsymbol{\gamma}^{\mathrm{T}}]$, and the elements of $\boldsymbol{\phi}_\gamma$ are $\boldsymbol{\phi}_{\gamma,11} = E[\mathbf{Z} \Delta f\{\Lambda(V), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}]$, $\boldsymbol{\phi}_{\gamma,12} = \mathbf{0}$, $\boldsymbol{\phi}_{\gamma,21} = E(\Delta W f\{\Lambda(V), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}\{2\gamma_1 + \Lambda(V)\eta(W, \mathbf{Z}, \boldsymbol{\beta})\} - W\eta(W, \mathbf{Z}, \boldsymbol{\beta}) \times [F\{\Lambda(V), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\} - F(0, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})])$, $\boldsymbol{\phi}_{\gamma,22} = E(-\eta(W, \mathbf{Z}, \boldsymbol{\beta}) [\Delta f\{\Lambda(V), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}\Lambda(V) - F\{\Lambda(V), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\} + F(0, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})])$. Also,

$$
\begin{aligned}
\boldsymbol{\gamma}_\beta =& \frac{\partial \boldsymbol{\gamma}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{\mathrm{T}}} = \begin{bmatrix} \mathbf{0} & E\{\exp(\beta_2 U) U\} \\ \mathbf{0} & E\{\exp(\beta_2 U) U^2\} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \gamma_2 \\ \mathbf{0} & E\{\exp(\beta_2 U) U^2\} \end{bmatrix}, \\
\boldsymbol{\phi}_\beta =& E\left[\frac{\partial \boldsymbol{\phi}\{\mathbf{O}; \boldsymbol{\beta}, \Lambda(V), \boldsymbol{\gamma}, \boldsymbol{\alpha}\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\right] \\
=& E\Big([\Delta \Lambda(V) f\{\Lambda(V), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\} - F\{\Lambda(V), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\} + F(0, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})] \eta(W, \mathbf{Z}, \boldsymbol{\beta})
\end{aligned}
$$

$$
\begin{aligned}
&\times \begin{pmatrix} \mathbf{Z} \\ \gamma_1 W - \gamma_2 \end{pmatrix} \begin{pmatrix} \mathbf{Z} \\ W \end{pmatrix}^{\mathrm{T}} \Bigg) \\
&+ E \left[\begin{array}{c} \mathbf{Z}\Delta\{\gamma_1 + \Lambda(V)\eta(W,\mathbf{Z},\boldsymbol{\beta})\} f_{\boldsymbol{\beta}}^{\mathrm{T}}\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} \\ \Delta\{W\gamma_1^2 + \Lambda(V)(\gamma_1 W - \gamma_2)\eta(W,\mathbf{Z},\boldsymbol{\beta})\} f_{\boldsymbol{\beta}}^{\mathrm{T}}\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} \end{array} \right] \\
&- E \left[\begin{array}{c} \mathbf{Z}\eta(W,\mathbf{Z},\boldsymbol{\beta})[F_{\boldsymbol{\beta}}^{\mathrm{T}}\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} - F_{\boldsymbol{\beta}}^{\mathrm{T}}(0,\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha})] \\ (\gamma_1 W - \gamma_2)\eta(W,\mathbf{Z},\boldsymbol{\beta}) \left[F_{\boldsymbol{\beta}}^{\mathrm{T}}\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} - F_{\boldsymbol{\beta}}^{\mathrm{T}}(0,\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha})\right] \end{array} \right],
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\phi}_\Lambda(\mathbf{O}) =& \frac{\partial \boldsymbol{\phi}\{\mathbf{O};\boldsymbol{\beta},\Lambda(V),\boldsymbol{\gamma},\boldsymbol{\alpha}\}}{\partial \Lambda(V)} \\
=& \begin{pmatrix} \mathbf{Z}\eta(W,\mathbf{Z},\boldsymbol{\beta})[(\Delta-1)f\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} + f'\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\}\Delta\Lambda(V)] \\ (\gamma_1 W - \gamma_2)\eta(W,\mathbf{Z},\boldsymbol{\beta})[(\Delta-1)f\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} + f'\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\}\Delta\Lambda(V)] \end{pmatrix} \\
&+ \left[\begin{array}{c} \mathbf{Z}\Delta\gamma_1 f'\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} \\ f'\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\}\Delta W \gamma_1^2 \end{array} \right],
\end{aligned}
$$

$$
\mathbf{D}_4(s) = E\left\{ \frac{Y(s)\boldsymbol{\phi}_\Lambda(\mathbf{O})D_1(s)}{D_1(V)C_1(s)} \right\}, \ \mathbf{g}(s,W,\mathbf{Z}) = \mathbf{D}_4(s)\left\{\gamma_1(\boldsymbol{\beta}) + \Lambda(s)\eta(W,\mathbf{Z},\boldsymbol{\beta})\right\},
$$

$\psi_1(s,t,W_i,\mathbf{Z}_i) = I(0 \le s \le t)\{D_1(s)/C_1(s)D_1(t)\}\{\gamma_1(\boldsymbol{\beta}) + \Lambda(s)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\} - \{D_2(t)(\mathbf{0}^{\mathrm{T}},\gamma_2) + \mathbf{D}_3^{\mathrm{T}}(t)\}\Sigma_H^{-1}\mathbf{g}(s,W_i,\mathbf{Z}_i)$, $\psi_2(t,X_i,U_i,W_i^*,\mathbf{Z}_i,Y_i) = D_2(t)f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta}) - \left\{D_2(t)(\mathbf{0},\gamma_2) + \mathbf{D}_3^{\mathrm{T}}(t)\right\}\Sigma_H^{-1}$ $[\boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma},\boldsymbol{\alpha}\} + \boldsymbol{\phi}_\gamma \mathbf{f}_\gamma(\mathbf{W}_i^*,\boldsymbol{\beta}) + E\{\boldsymbol{\phi}_\Lambda(\mathbf{O})D_2(V)\}f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta})]$, where

$$
\Sigma_H = \boldsymbol{\phi}_\beta + \boldsymbol{\phi}_\gamma \boldsymbol{\gamma}_\beta + E\left[\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)\mathbf{D}_3(V_i) + \left\{\mathbf{0}_{(p+1)\times p}, \gamma_2\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)D_2(V_i)\right\}\right].
$$

Finally, define

$$
\begin{aligned}
\Sigma_M =& E\Bigg[\boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma},\boldsymbol{\alpha}\} + \int_0^\tau \mathbf{g}(s,W,\mathbf{Z})dM(s) \\
&+ E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O})D_2(V)\right\} f_{\gamma,1}(\mathbf{W}^*,\boldsymbol{\beta}) + \boldsymbol{\phi}_\gamma \mathbf{f}_\gamma(\mathbf{W}^*,\boldsymbol{\beta})\Bigg]^{\otimes 2}.
\end{aligned}
$$

The following theorems establish the consistency and the asymptotic normality of the estimator in terms of its first order asymptotic properties. The regularity conditions and the proofs are given in the Supplementary materials S2, S3, S4, and S5.

Theorem 1. *Assume the regularity conditions hold. When $n \to \infty$, $|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}| \to \mathbf{0}$ in probability and $\sup_{u\in[0,\tau]} |\widehat{\Lambda}\{u,\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\} - \Lambda(u)| \to 0$ in probability.*

Theorem 2. *Assume the regularity conditions hold. When $n \to \infty$,*

i) $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to \mathrm{Normal}(\mathbf{0}, \Sigma_H^{-1}\Sigma_M\Sigma_H^{-T})$ in distribution;

ii) $\sqrt{n}[\widehat{\Lambda}\{t,\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\}-\Lambda(t)]$ *follows a zero-mean Gaussian process with the covariance kernel*

$$\Omega(t,t') = E\left\{\int_0^\tau \psi_1(s,t,W,\mathbf{Z})dM(s) + \psi_2(t,X,U,W^*,\mathbf{Z},Y)\right\}^{\otimes 2}.$$

Since the estimating equations in Section 2.3 reduce to (2), (3) and (4) when $W_i = X_i$, $\gamma_1 = 1$ and $\gamma_2 = 0$, the corresponding variance formula for the error-free case can be directly derived from the results in these theorems.

3.2 Estimation of the asymptotic variance

We now further study how to estimate the asymptotic variance of $\widehat{\boldsymbol{\beta}}$. We first write out the relation between X and $\mathbf{Z}$ as $X = \vartheta_1(\mathbf{Z},\boldsymbol{\zeta}_1) + \vartheta_2^{1/2}(\mathbf{Z},\boldsymbol{\zeta}_2)e_x$, where $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ are unknown parameters, and $E(e_x) = 0$. Here, for simplicity, we used parametric forms for the mean and variance function, while nonparametric model can also be used for increased flexibility. Define two weighted averages (Hall and Ma, 2007) of $(W_{i1}^*,\ldots,W_{im}^*)$, $W_{ia} = \sum_{j=1}^m a_j W_{ij}^*$ and $W_{ib} = \sum_{j=1}^m b_j W_{ij}^*$, where $a_j = b_j = 1/(2[m/2])$ for $j = 1,\ldots,[m/2]$, and $a_j = -b_j = 1/(2m-2[m/2])$ for $j = [m/2]+1,\ldots,m$. Here $[m/2]$ denotes the largest integer $\leq m/2$. Note that $\sum_{j=1}^m a_j = 1$ and $\sum_{j=1}^m b_j = 0$ and $\sum_{j=1}^m a_j^2 = \sum_{j=1}^m b_j^2$. To estimate $\boldsymbol{\zeta}_1$ we shall solve $\sum_{i=1}^n \partial\{\vartheta_1(\mathbf{Z}_i,\boldsymbol{\zeta}_1)/\partial\boldsymbol{\zeta}_1\}\{W_{ia} - \vartheta_1(\mathbf{Z}_i,\boldsymbol{\zeta}_1)\} = 0$. To estimate $\boldsymbol{\zeta}_2$ we shall solve $\sum_{i=1}^n \partial\{\vartheta_2(\mathbf{Z}_i,\boldsymbol{\zeta}_2)/\partial\boldsymbol{\zeta}_2\}[\{W_{ia} - \vartheta_1(\mathbf{Z}_i,\widehat{\boldsymbol{\zeta}}_1)\}^2 - \widehat{\sigma}_a^2 - \vartheta_2(\mathbf{Z},\boldsymbol{\zeta}_2)] = 0$, where $\widehat{\sigma}_a^2 = \sum_{i=1}^n W_{ib}^2/n$. Observe that due to symmetry, the distribution of $U_{ia} = \sum_{j=1}^m a_j U_{ij}^*$ is the same as that of $W_{ib} = U_{ib} = \sum_{j=1}^m b_j U_{ij}^*$, and it is a symmetric distribution. Thus, the density of U_{ia} can be estimated via $\widehat{f}_{U_a}(u) = (1/nh)\sum_{i=1}^n K\{(u-U_{ib})/h\}$, where we let $K(\cdot)$ be a symmetric kernel function and $h > 0$ be a bandwidth, and we select the optimal bandwidth via the plug-in bandwidth selection method given in Sheather and Jones (1991). Next we estimate $\boldsymbol{\omega}$ by maximizing the estimated likelihood of W_{ia} given $\mathbf{Z}_i$, i.e.,

$$\prod_{i=1}^n \sum_{l=1}^{L(n)} \omega_l \frac{1}{nh}\sum_{i_1=1}^n K[h^{-1}\{(W_{ia} - \vartheta_1(\mathbf{Z}_i,\widehat{\boldsymbol{\zeta}}_1) - \vartheta_2^{1/2}(\mathbf{Z}_i,\widehat{\boldsymbol{\zeta}}_2)e_l - U_{i_1 b}\}],$$

where we approximate the expectation with respect to e_x through adding the probability masses $\boldsymbol{\omega} = (\omega_1,\cdots,\omega_{L(n)})^T$ at points $e_1 < \cdots < e_{L(n)}$. Define $\mathcal{X}_{il} = \vartheta_1(\mathbf{Z}_i,\widehat{\boldsymbol{\zeta}}_1) +$

$\vartheta_2^{1/2}(\mathbf{Z}_i, \widehat{\boldsymbol{\zeta}}_2)e_l$, and let $\widehat{\boldsymbol{\omega}} = (\widehat{\omega}_1, \cdots, \widehat{\omega}_{L(n)})^T$,

$$\widehat{\kappa}_1 = \left[\frac{2}{nm(m-1)} \sum_{j,k=1,j<k}^{m} \sum_{i=1}^{n} \exp\{(W_{ij}^* - W_{ik}^*)2\widehat{\beta}_2/m\}\right]^{m/2},$$
$$\widehat{\kappa}_2 = \left(\frac{1}{2}\right)\left(\frac{m}{2}\right)(\widehat{\kappa}_1)^{(m-2)/m}\left[\frac{4}{nm^2(m-1)} \sum_{i=1}^{n} \sum_{j<k} (W_{ij}^* - W_{ik}^*) \exp\{(W_{ij}^* - W_{ik}^*)2\widehat{\beta}_2/m\}\right]$$

be the estimators of $\boldsymbol{\omega}$, κ_1 and κ_2, respectively. In Section S5 of the Supplementary materials, we express Σ_M as $\Sigma_M = \mathbf{G}^{(1)} + \mathbf{G}^{(2)} + \mathbf{G}^{(3)} + \mathbf{G}^{(4)} + \mathbf{G}^{(5)} + (\mathbf{G}^{(4)} + \mathbf{G}^{(5)})^T$. This expression allows us to construct a consistent estimator of the asymptotic variance of $\widehat{\boldsymbol{\beta}}$, which we provide in Corollary 1.

Corollary 1. *A consistent estimator of the asymptotic variance of* $\widehat{\boldsymbol{\beta}}$ *is* $n^{-1}\widehat{\Sigma}_H^{-1}\widehat{\Sigma}_M\widehat{\Sigma}_H^{-T}$, *where* $\widehat{\Sigma}_H \equiv \widehat{\boldsymbol{\phi}}_\beta + \widehat{\boldsymbol{\phi}}_\gamma\widehat{\boldsymbol{\gamma}}_\beta + n^{-1}\sum_{i=1}^n[\widehat{\boldsymbol{\phi}}_\Lambda(\mathbf{O}_i)\widehat{\mathbf{D}}_3(V_i) + \{\mathbf{0}_{(p+1)\times p}, \widehat{\gamma}_2\widehat{\boldsymbol{\phi}}_\Lambda(\mathbf{O}_i)\widehat{D}_2(V_i)\}]$, *and*

$$\widehat{\Sigma}_M = \widehat{\mathbf{G}}^{(1)} + \widehat{\mathbf{G}}^{(2)} + \widehat{\mathbf{G}}^{(3)} + \widehat{\mathbf{G}}^{(4)} + \widehat{\mathbf{G}}^{(5)} + (\widehat{\mathbf{G}}^{(4)} + \widehat{\mathbf{G}}^{(5)})^T$$

with $\widehat{\mathbf{G}}^{(1)} = n^{-1}\sum_{i=1}^n \boldsymbol{\phi}^{\otimes 2}\{\mathbf{O}_i; \widehat{\boldsymbol{\beta}}, \widehat{\Lambda}(V_i), \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\alpha}}\}$, $\widehat{\mathbf{G}}^{(2)} = n^{-1}\sum_{i=1}^n \Delta_i\mathbf{g}^{\otimes 2}(V_i, W_i, \mathbf{Z}_i)$, $\widehat{\mathbf{G}}^{(3)} = n^{-1}\sum_{i=1}^n \left[\boldsymbol{\phi}_\gamma\mathbf{f}_\gamma(\mathbf{W}_i^*, \widehat{\boldsymbol{\beta}}) + E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O})\widehat{D}_2(V)\right\} f_{\gamma,1}(\mathbf{W}_i^*, \widehat{\boldsymbol{\beta}})\right]^{\otimes 2}$, $\widehat{\mathbf{G}}^{(4)} = n^{-1}\sum_{i=1}^n \boldsymbol{\phi}\{\mathbf{O}_i; \widehat{\boldsymbol{\beta}}, \widehat{\Lambda}(V_i), \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\alpha}}\}[\boldsymbol{\phi}_\gamma\mathbf{f}_\gamma(\mathbf{W}_i^*, \widehat{\boldsymbol{\beta}}) + E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O})\widehat{D}_2(V)\right\} f_{\gamma,1}(\mathbf{W}_i^*, \widehat{\boldsymbol{\beta}})]^{\mathrm{T}}$, *and*

$$\begin{aligned}
\widehat{\mathbf{G}}^{(5)} =& \frac{1}{n}\sum_{i=1}^n \Delta_i\boldsymbol{\phi}\{\mathbf{O}_i; \widehat{\boldsymbol{\beta}}, \widehat{\Lambda}(V_i), \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\alpha}}\}\mathbf{g}^{\mathrm{T}}(V_i, W_i, \mathbf{Z}_i) \\
&+ \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} \mathbf{Z}_i\eta(W_i, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}) \\ \\ (\widehat{\gamma}_1 W_i - \widehat{\gamma}_2)\eta(W_i, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}) \end{bmatrix} \sum_{V_k:\Delta_k=1} f\{\widehat{\Lambda}(V_k), \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}\}\widehat{\lambda}(V_k)\Delta_i\mathbf{g}^T(V_i, W_i, \mathbf{Z}_i)Y_i(V_k) \\
&+ \frac{1}{n}\sum_{i=1}^n\sum_{l=1}^{L(n)}\sum_{V_k:\Delta_k=1}\left(\begin{bmatrix} \mathbf{Z}_i\eta(\mathcal{X}_{il}, \mathbf{Z}_i, \boldsymbol{\beta})\{\widehat{\gamma}_1^2 + \widehat{\Lambda}(V_k)\eta(\mathcal{X}_{il}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}})\widehat{\kappa}_1\} \\ \\ \widehat{\gamma}_1^3\eta(\mathcal{X}_{il}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}})X_i + \widehat{\Lambda}(V_k)\eta^2(\mathcal{X}_{il}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}})(\widehat{\gamma}_1\mathcal{X}_{il}\widehat{\kappa}_1 + \widehat{\gamma}_1\widehat{\kappa}_2 - \widehat{\gamma}_2\widehat{\kappa}_1) \end{bmatrix}\right. \\
&\times \left.\frac{\widehat{\mathbf{D}}_4^T(V_k)\widehat{\lambda}(V_k)\eta(\mathcal{X}_{il}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}})}{1 + \widehat{\Lambda}(V_k)\eta(\mathcal{X}_{il}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}})} \sum_{V_j \geq V_k} f\{\widehat{\Lambda}(V_j), \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}\}\widehat{\lambda}(V_j)Y_i(V_j)\right)\widehat{\omega}_l \\
&- \frac{1}{n}\sum_{i=1}^n\sum_{l=1}^{L(n)}\sum_{V_k:\Delta_k=1}\left(\begin{bmatrix} \mathbf{Z}_i[\widehat{\gamma}_1^2 + \widehat{\gamma}_1^2\{\widehat{\Lambda}(V_k) + \widehat{\Lambda}(V_i)\}\eta(\mathcal{X}_{il}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}) \\ +\widehat{\Lambda}(V_k)\widehat{\Lambda}(V_i)\eta^2(\mathcal{X}_{il}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}})\kappa_1] \\ \\ \widehat{\gamma}_1^3\mathcal{X}_{il} + \widehat{\gamma}_1^3\mathcal{X}_{il}\eta(\mathcal{X}_{il}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}})\Lambda(V_i) + \widehat{\gamma}_1^2\eta(\mathcal{X}_{il}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}})\widehat{\Lambda}(V_k)(\widehat{\gamma}_1\mathcal{X}_{il} + \widehat{\gamma}_2) \\ +\eta^2(\mathcal{X}_{il}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}})\widehat{\Lambda}(V_k)\widehat{\Lambda}(V_i)(\widehat{\gamma}_1\widehat{\kappa}_1\mathcal{X}_{il} + \widehat{\gamma}_1\widehat{\kappa}_2 - \widehat{\gamma}_2\widehat{\kappa}_1) \end{bmatrix}\right.
\end{aligned}$$

$$\times \frac{\widehat{\mathbf{D}}_4^T(V_k)\widehat{\lambda}(V_k)\eta(\mathcal{X}_{il}, \mathbf{Z}_i, \boldsymbol{\beta})}{1 + \Lambda(V_k)\eta(\mathcal{X}_{il}, \mathbf{Z}_i, \boldsymbol{\beta})} f\{\widehat{\Lambda}(V_i), \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}\}\Delta_i Y_i(V_k)\Big)\widehat{\omega}_l.$$

Note that other than the last two terms of $\mathbf{G}^{(5)}$, all other terms of Σ_M are estimated via empirical averages of obvervable random variables. While the root-n consistency and asymptotic normality are established in Theorems 1 and 2, the results in Corollary 1 further allow us to utilize these results to perform inference. All these results are established in the context of the proportional odds model subject to a symmetric, but otherwise unspecified covariate measurement errors, and without making any parametric assumption on the distribution of the unobserved covariate either. Thus, the estimation and inference are conducted in the functional measurement error framework (Carroll et al., 2006).

4 Simulation studies

We now investigate the finite sample performance of the proposed error corrected method through simulation studies. We simulated 1,000 data sets through generating Z from Normal$(0, 1)$, and generating X from a two-component mixture of normal distributions, $(1/3)$Normal$(-0.6, 0.5^2)$ + $(2/3)$Normal$(1.25, 0.5^2)$. The purpose of taking such a non-standard distribution for X is to show that the method can handle any distribution for X. The time-to-event T was generated from the proportional odds model (1) with $\Lambda(t) = t^2$, and $\beta_1 = \beta_2 = 1$. To generate censoring time C independent of X and Z, we used Exp$(e^{1.7})$ and Exp$(e^{0.2})$, the exponential distributions with mean $e^{1.7}$ and $e^{0.2}$, respectively. This results in an average of 20% and 50% censoring respectively. For censoring process dependent on X and Z, we generated the censoring time C from Exp$(e^{2.25-X-Z})$ and Exp$(e^{0.75-X-Z})$, resulting in 20% and 50% censoring, respectively. The two unbiased surrogate variables W_1^* and W_2^* were simulated by adding random noise U^* to X. In order to show that the proposed error corrected approach can handle any symmetric error distribution, we considered two different distributions for U^*, Normal$(0, 1)$ and Uniform$(-1.75, 1.75)$, and in both cases the error variances were equal to the variance of X.

We analyzed the simulated data sets using four methods, the naive method (NV), the regression calibration (RC), the method by Cheng and Wang (2001) (referred to as CW), and the error corrected (COR) method proposed in Section 2.3. For the naive method, we used the maximum likelihood method of Murphy et al. (1997) to estimate $\boldsymbol{\beta}$ and Λ, with X_i replaced by $W_i = (W_{i1}^* + W_{i2}^*)/2$. For the regression calibration method, we implemented the same maximum likelihood method, but with X_i replaced by $\widehat{X}_i$, where $\widehat{X}_i = (1/\widehat{\sigma}^2 + 1/\widehat{\sigma}_U^2)\{\overline{W}_i/\widehat{\sigma}_U^2 + (\widehat{\zeta}_0 + \widehat{\zeta}_1^{\mathrm{T}} Z_i)/\widehat{\sigma}^2\}$ with $\widehat{\sigma}^2$, $\widehat{\sigma}_U^2$, $\widehat{\zeta}_0$ and $\widehat{\zeta}_1$ being the estimators of σ^2, $\sigma_U^2 = \mathrm{var}(U)$, ζ_0 and ζ_1, respectively. Furthermore, ζ_0 and ζ_1 are the coefficients of the linear regression of X on Z, whereas σ^2 represents the conditional variance of X given Z. To implement the method by Cheng and Wang (2001), we estimated the parameters under the assumption that C is independent of any of T, X, Z or U^*, and we used normal models for both $X_i - X_{i'}$ and $U_{ij}^* - U_{i'j}^*$. Lastly, in our error corrected method, we used $f\{\Lambda(t), Z, \boldsymbol{\beta}, \boldsymbol{\alpha}\} = \{1 + \Lambda(t)\exp(Z\beta_1 + X^*\beta_2)\}^{-2}$, where $X^* = E(X \mid Z)$ was obtained from the linear model $\overline{W} = X + U = \alpha_0 + Z\alpha_1 + \epsilon + U$, and the standard errors of the estimator were estimated using the analytical formula given in Corollary 1 in Section 3. We used the Newton-Raphson procedure to solve the estimating equations, with the convergence criterion set to be either the absolute value of the estimating equations are smaller than 10^{-8} for each component, or the relative difference of the two latest iterations is smaller than 10^{-8} for each component in the $\boldsymbol{\beta}$. Both convergence criteria are standard in the usual statistical softwares. In estimating standard errors, integrals such as $\int_0^t f^* d\widehat{\Lambda}$ are replaced by $\sum_{k: t_k \le t, \Delta_k = 1} f^*(t_k)\widehat{\lambda}(t_k)$, for any generic function f^*.

Tables 1 and 2 contain the simulation results for the normal and uniform errors respectively. For both tables we took two different sample sizes $n = 500$ and $1,000$. We presented the bias, empirical standard error, median absolute deviation. In addition, for our error corrected method, we also provided the estimated standard error and the Wald type 95% coverage probability.

The general trend is the same in both tables. Overall the naive estimator is very biased. The regression calibration estimator has smaller bias, but its bias is still substantial compared with our error corrected estimator. In fact, the finite sample bias, especially in estimating β_2, is greatly reduced in the proposed error corrected method. The variance of the estimators decreases with the sample size n. Importantly, the estimated standard error based on our asymptotic results and the empirical standard error are quite close, and the coverage probabilities are reasonably close to the nominal level.

When the censoring time C is independent of both covariates and is generated from an Exponential distribution, the method of Cheng and Wang (2001) works surprisingly well (Tables 1 and 2) despite the fact that several model assumptions are violated in these simulations. However, as soon as the censoring mechanism depends on X and Z or the censoring rate is high, their method shows large bias. To further investigate this matter, in the uniform measurement error scenario with $n = 1,000$, we generated C in three different cases. In case 1, C followed Exp(0.22). In case 2, C followed Uniform$(0, 0.5)$. In case 3, C followed Exp$(e^{-0.8-X-Z})$. All three cases have about 85% censoring, roughly the same as in the data example. In case 1, C is independent of the covariates and the supports of the time-to-event T and censoring time C are similar. In case 2, C is also independent of the covariates, but the support of C is shorter than that of T. This is a common scenario in many clinical studies and is also the case for our real data example. In case 3, the censoring mechanism depends on the covariates and the supports of T and C are similar. This simulation results in Table 3 indicate dramatically large estimation bias and MSE of the Cheng and Wang (2001) method in comparison with our method for cases 2 and 3. In case 1, although the bias of CW is comparable with ours, their MSE is larger than ours. In conclusion, for heavy censoring, regardless of whether censoring is dependent on the covariates or not, the bias of the CW method is substantial. This study verifies the inconsistency of the estimating equations of Cheng and Wang (2001), similar to the inconsistency of Cheng et al. (1995) pointed out by Fine et al. (1998).

Finally, the computation of the proposed error corrected estimator is also much simpler and faster than that of the Cheng and Wang (2001) method. This is mainly because their method requires numerical integration and is hence very time consuming.

5 Real data analysis

For the purpose of illustration we now apply the proposed method to analyze a dataset from the ACTG 175 study, a clinical trial of HIV therapy (Hammer et al., 1996). This was a randomized double-blinded study to investigate the effect of a single nucleoside or two nucleosides among HIV-1 infected adults. We considered only $n = 1,036$ subjects who did not have antiretroviral treatment before this trial, and among them 262 received 600 mg of zidovudine (treatment 1), 257 received 600 mg of zidovudine plus 400 mg of didanosine (treatment 2), 260 received 600 mg of zidovudine plus 2.25 mg of zalcitabine (treatment 3), and 257 received 400 mg of didanosine (treatment 4). The primary clinical endpoints were progression to AIDS and/or death, thus we consider T as the time to AIDs or death from the date the treatment started. In our data, only 85 subjects experienced the events during an average follow-up time of 32 months. For all subjects, two ($m = 2$) baseline CD4 measurements that were taken prior to the treatment started, were available. CD4 cells help to fight infection. Therefore, low CD4 counts indicates weak immune system and it is used as a marker of the stage of HIV disease.

We fit model (1) to this data set, where the logarithm of the actual CD4 count at the baseline minus 5.89 is considered as X. The two baseline measurements are considered as two erroneous measurements for X. The three dummy variables corresponding to the four treatments are considered to be error free covariates $\mathbf{Z}$ where treatment 1 is considered as the reference category. We analyze the data set using four methods, NV, RC, CW, and COR described in the simulation section. For the CW method, T and C are assumed to be independent.

Table 4 contains the estimates and their corresponding standard errors. All methods indicate a statistically significant (at the 5% level) association between X and T. We also find that compared to the monotherapy with zidovudine, other three therapies have statistically significant association (at the 5% level) with T, in particular, the results indicate that the therapies tend to delay the time-to-event. Interestingly, after adjusting for the measurement errors, the CW estimate of the coefficient for CD4 counts, β_2, is substantially different from that of NV, RC, and COR methods, although the effect of the log(CD4) still turned out to be statistically significant. Our experience with the simulation studies indicates that the distinct result of the CW estimator is likely due to the high censoring percentage in the data (around 90%), shorter support of C compared to that of T as the subjects were followed for a maximum of three and half years, and the possible dependence between the covariates and the censoring mechanism, which violate the model assumption required by the CW estimator. The dependence between the covariates and the censoring mechanism is indicated when we fit the Cox model to the censoring distribution using $(V_i, (1-\Delta_i), \mathbf{Z}_i, W_i^\dagger), i = 1, \cdots, n$, where $W_i^\dagger = I(W_i < -0.4)$, with -0.4 being the 15th quantile of W_i. The results show statistically significant association (at the 1% level) between C and covariates ($\mathbf{Z}$ and $W^\dagger$). So, we also suspect that C and X are dependent as well.

Inspired by a referee's suggestion, we further estimated the parameters using the proposed method with $f = 1/\{1+\Lambda(u)\eta(X^*, \mathbf{Z}, \boldsymbol{\beta})\}^r$ for $r = 0, 1, 2, 3, 4, 5, 10, 15$. Based on the results in Table 5, although the estimates differ with r, the magnitude of the change is quite small. Our experience in more extensive numerical experiments not reported here also indicates that the variability in estimating $\boldsymbol{\beta}$ is somewhat insensitive to the choice of f.

Following a referee's request, we also conducted a model checking for this data example. Because there is no existing method to check proportional odds assumption when a covariate is measured with errors, we developed the following graphical tools, inspired by the graphical tools developed for the Cox proportional hazard model without measurement errors (Klein and Moeschberger 2003, Chapter 11.4, page 363). Note that the proportional odds model has the property $\mathrm{pr}(T \le t|X, \mathbf{Z})/\mathrm{pr}(T > t|X, \mathbf{Z}) = \log\{\Lambda(t)\} + \boldsymbol{\beta}_1^T\mathbf{Z} + \beta_2 X$. In the data example, $\mathbf{Z}$ is a nominal categorical variable. Define $X^\dagger$ to be zero when X is less than or equal to $r = -0.1$ and one otherwise, where $r = -0.1$ is the median of W_i. Define $\mathrm{pr}_{\mathbf{z}}(T \le t|X^\dagger) = \mathrm{pr}(T \le t|X^\dagger, \mathbf{Z} = \mathbf{z})$. Then in each category of $\mathbf{Z}$, we plot $\log\{\widehat{\mathrm{pr}}_{\mathbf{z}}(T \le t|X^\dagger)/\widehat{\mathrm{pr}}_{\mathbf{z}}(T > t|X^\dagger)\}$ as a function of time t. If the proportional odds assumption holds, then the two curves corresponding to $X^\dagger = 0$ and $X^\dagger = 1$ will have the same shape and they will differ only by a constant shift. Here $\log\{\widehat{\mathrm{pr}}_{\mathbf{z}}(T \le t|X^\dagger)\}$ is an estimator of $\log\{\mathrm{pr}_{\mathbf{z}}(T \le t|X^\dagger)\}$. When X

is measured with error, deterministic classification of the subjects into two groups $X^\dagger = 0$ and $X^\dagger = 1$ is not possible. Therefore, first we estimate $\mathrm{pr}(X_i^\dagger = 0|W_{ia}) = \mathrm{pr}(X_i \leq r, W_{ia})/f(W_{ia})$ through

$$q_i = \frac{\widehat{\mathrm{pr}}(X_i \leq r, W_{ia})}{\widehat{f}(W_{ia})} = \frac{\sum_{l=1}^{L(n)} \omega_l \sum_{k=1}^n K[h^{-1}\{(W_{ia} - \mathcal{X}_{il} - U_{kb}\}]I(\mathcal{X}_{il} \leq r)}{\sum_{l=1}^{L(n)} \omega_l \sum_{k=1}^n K[h^{-1}\{(W_{ia} - \mathcal{X}_{il} - U_{kb}\}]},$$

where $\mathcal{X}_{il} = \vartheta_1(\mathbf{Z}_i, \widehat{\boldsymbol{\zeta}}_1) + \vartheta_2^{1/2}(\mathbf{Z}_i, \widehat{\boldsymbol{\zeta}}_2)e_l$. To estimate the survival function $\mathrm{pr}_{\mathbf{z}}(T > v|X^\dagger = k)$ with uncertain membership, we use the procedure developed in Ma et al. (2011) for estimating a distribution function. Note that for any v, $E\{I(V_i > v)\} = q_i \mathrm{pr}_{\mathbf{z}}(T_i > v|X_i^\dagger = 0)\mathrm{pr}_{\mathbf{z}}(C_i > v|X_i^\dagger = 0) + (1-q_i)\mathrm{pr}_{\mathbf{z}}(T_i > v|X_i^\dagger = 1)\mathrm{pr}_{\mathbf{z}}(C_i > v|X_i^\dagger = 1)$ for $i = 1, \cdots, n$. Thus, using $I(V_i > v)$ as the observed response and $\boldsymbol{q}_i = (q_i, 1-q_i)^T$ as the observed predictor for the ith subject, the least square solutions of the unknowns are

$$\left\{ \begin{array}{l} \mathrm{pr}_{\mathbf{z}}(T > v|X^\dagger = 0)\mathrm{pr}_{\mathbf{z}}(C > v|X^\dagger = 0) \\ \mathrm{pr}_{\mathbf{z}}(T > v|X^\dagger = 1)\mathrm{pr}_{\mathbf{z}}(C > v|X^\dagger = 1) \end{array} \right\} = (\textstyle\sum_{i:\mathbf{Z}_i=\mathbf{z}} \boldsymbol{q}_i \boldsymbol{q}_i^T)^{-1} \sum_{i:\mathbf{Z}_i=\mathbf{z}} \boldsymbol{q}_i^T I(V_i > v). \quad (9)$$

To further handle the censoring issue and to extract the survival function $\mathrm{pr}_{\mathbf{z}}(T > v|X^\dagger = k)$ alone, we consider the following. Let $\lambda_{T0}(t)$ and $\lambda_{T1}(t)$ be the hazard of T when $\mathbf{Z} = \mathbf{z}$ and $X^\dagger = 0$ and $X^\dagger = 1$, respectively, and the corresponding hazard of the censoring variable are $\lambda_{C0}(t)$ and $\lambda_{C1}(t)$, respectively. Then $\mathrm{pr}_{\mathbf{z}}(T > v|X^\dagger = k) = \exp\{-\int_0^v \lambda_{Tk}(u)du\}$ and $\mathrm{pr}_{\mathbf{z}}(C > v|X^\dagger = k) = \exp\{-\int_0^v \lambda_{Ck}(u)du\}$ for $k = 0$ and 1. Since $N(t) - \int_0^t Y(s)\{q\lambda_{T0}(s) + (1-q)\lambda_{T1}(s)\}ds$ and $I(V \leq t, \Delta = 0) - \int_0^t Y(s)\{q\lambda_{C0}(s) + (1-q)\lambda_{C1}(s)\}ds$ are two martingale processes, for any v, we consider two sets of estimating equations

$$\sum_{i=1}^n dN_i(v) = \lambda_{T0}(v)\sum_{i=1}^n q_i Y_i(v) + \lambda_{T1}(v)\sum_{i=1}^n (1-q_i)Y_i(v), \quad (10)$$

$$\sum_{i=1}^n I(V_i = v, \Delta_i = 0) = \lambda_{C0}(v)\sum_{i=1}^n q_i Y_i(v) + \lambda_{C1}(v)\sum_{i=1}^n (1-q_i)Y_i(v). \quad (11)$$

Therefore, for each v we estimate the hazards from equations (9), (10), and (11). Once the hazards are estimated, we obtain $\widehat{\mathrm{pr}}_{\mathbf{z}}(T > v|X^\dagger = k) = \exp\{-\sum_{i:u_i \leq v, \Delta_i=1} \widehat{\lambda}_{Tk}(u_i)\}$ and produce the plots in Figure 1. None of the four plots indicates any striking violation of the proportional odds assumption, such as crossing of the curves. Thus, proportional odds model is a suitable model for this data set. Although the method was developed for a discretized X, the method is useful for detecting any major model violation.

6 Discussion

We have proposed an error corrected martingale based estimating equation to analyze the time-to-event data in the proportional odds model when both covariate measurement error and right censoring to event time occur. In contrast to the existing literature, we do not assume or estimate the distribution for the measurement errors or for the true unobservable covariates. We have merely required multiple measurements, which is needed even for identifiability in such models. Our results on the theoretical properties of the estimators show that the estimators have the desired asymptotic properties which facilitate further inference. Finally, although the estimator is designed for errors in covariates, it captures the usual error-free covariates case as well by simply allowing the error distribution to be a point mass at zero. This provides a new estimator in its own for the usual proportional odds models without measurement errors. As pointed out by a referee, small sample and large error variance can break any consistent estimator designed for measurement error problems. Thus, improving the finite sample performance under small sample size and large error, such as the one investigated in Song and Huang (2005), is definitely an important research question worth investigating.

Compared to the Cox model, other time-to-event models have received relatively less attention when an important covariate is measured with errors. The present article with nonparametric correction is the first attempt to break such barrier. However, the problem is far from being completely resolved. For example, estimation efficiency is not achieved in the estimator. In fact, our preliminary analysis indicates that even in the Cox model, efficient estimator can be hard to achieve. The main difficulties in achieving efficiency include the need to estimate the measurement error distribution, the need to estimate the distribution of the covariate subject to error, and the need to estimate the censoring process when covariates are not all observable. We envision that the proposed error corrected method and some of its apparent limitations will help generate new ideas. In particular, the proposed error corrected method will be useful for developing methods for handling measurement errors in multiple time-dependent covariates (Song et al., 2002) in the proportional odds model. Also, we believe that existing variable selection technique in the presence of measurement error (Ma and Li, 2010) can be integrated with our proposed error corrected method in the time-to-event model, in particular in the proportional odds model. Our method will also help to develop methodology for handling covariate measurement errors in multivariate failure time model (Greene and Cai, 2004).

It is well known that estimating equation based methods face the potential difficulty of having multiple roots in finite samples. Although there are some available methods for multiple roots of estimating equations in parametric models (Small and Wang, 2003; p. 163), as far as we are aware, in the current literature of semiparametric models like ours, multiple roots issue is handled through empirical analysis. For example, in the measurement error problems, one could compare the naive estimator and the regression calibration estimator with the multiple roots obtained from the estimating equations, and choose the root that is most sensible based on the knowledge that regression calibration estimator is an approximately consistent estimator that corrects the bias inherent in the naive estimator. Alternatively, in the presence of multiple roots, one may compute an estimated version of the likelihood

$$\begin{aligned} L &= \prod_{i=1}^{n} \left\{ \int f_{Y|\mathbf{Z},X}(V_i|X,\mathbf{Z}_i) f_{W,X|\mathbf{Z}}(W_{ia},X|\mathbf{Z}_i)dX \right\}^{\Delta_i} \\ &\quad \times \left\{ \int \mathrm{pr}(T_i \geq V_i|X,\mathbf{Z}_i) f_{W,X|\mathbf{Z}}(W_{ia},X|\mathbf{Z}_i)dX \right\}^{1-\Delta_i}, \end{aligned}$$

via

$$\begin{aligned} \widehat{L} &= \prod_{i=1}^{n} \sum_{l=1}^{L(n)} \sum_{k=1}^{n} \Bigg(\left[\frac{\sum_{j=1}^{k} \{\widehat{\Lambda}(t_{n_j}) - \widehat{\Lambda}(t_{n_{j-1}})\} I(t_{n_j} \leq V_i < t_{n_{j+1}}) \eta(\mathcal{X}_{il},\mathbf{Z}_i,\widehat{\boldsymbol{\beta}})}{\{1+\sum_{j=1}^{k} \widehat{\Lambda}(t_{n_j}) I(t_{n_j} \leq V_i < t_{n_{j+1}}) \eta(\mathcal{X}_{il},\mathbf{Z}_i,\widehat{\boldsymbol{\beta}})\}^2} \right]^{\Delta_i} \\ &\quad \times \left\{ \frac{1}{1+\sum_{j=1}^{k} \widehat{\Lambda}(t_{n_j}) I(t_{n_j} \leq V_i < t_{n_{j+1}}) \eta(\mathcal{X}_{il},\mathbf{Z}_i,\widehat{\boldsymbol{\beta}})} \right\}^{1-\Delta_i} \Bigg) \\ &\quad \times \left(\frac{\widehat{\omega}_l}{nh} \right) K\{h^{-1}(W_{ia} - \mathcal{X}_{il} - W_{kb})\}, \end{aligned}$$

and choose the root that maximizes $\widehat{L}$. The notations $\widehat{\omega}$, $\mathcal{X}_{il}$, W_{ia}, W_{ib}, K, and h are defined in Section 3.2.

References

Buzas, J. S. (1998). Unbiased scores in proportional hazards regression with covariate measurement error. *Journal of Statistical Planning and Inference*, **67**, 247–257.

Carroll, R. J., Ruppert, A., Stefanski, L. A. and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, Second Edition. London: CRC Press.

Cheng, S. C., Wei, L. J., Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, **82**, 835–845.

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

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

Greene, W. F. and Cai, J. (2004). Measurement error in covariates in the marginal hazards model for multivariate failure time data. *Biometrics*, **60**, 987–996.

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

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S. & Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *The new England Journal of Medicine* **335**, 1081-1090.

Hu, C. and Lin, D. Y. (2002). Cox regression with covariate measurement error. *The Scandinavian Journal of Statistics*, **29**, 637–655.

Huang, J. (1995). Maximum likelihood estimation for proportional odds regression model with current status data. *Analysis of Censored Data (IMS Lecture Notes Monograph*

Series, Volume 27), 129–145.

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

Huang, Y. and Wang, C. Y. (2001). Consistent functional methods for logistic regression with errors in covariates. *Journal of the American Statistical Association*, **96**, 1469–1482.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis, Techniques for Censored and Truncated Data*. Springer: New York.

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

Ma, Y., Hart, J. D., and Carroll, R. J. (2011). Density estimation in several populations with uncertain population membership. *Journal of the American Statistical Association*, **106**, 1180–1192.

Ma, Y. and Li, R. (2010). Variable selection in measurement error models. *Bernoulli*, **16**, 274–300.

Murphy, S. A., Rossini, A. J., and van der Vaart, A. W. (1997). MLE in the proportional odds model. *Journal of the American Statistical Association*, **92**, 968–976.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: methodology and application to generalized linear models. *Biometrika*, **77**, 127–13.

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

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

Rossini, A. J. and Tsiatis, A. A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *Journal of the American Statistical Association*, **91**, 713–721.

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

Small, C. G. and Wang, J. (2003). *Numerical Methods for Nonlinear Estimating Equations.* Oxford University Press: Oxford.

Song, X. and Huang, Y. (2005). On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics*, **61**, 702–714.

Song, X., Davidian, M., and Tsiatis, A. A. (2002). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics*, **3**, 511–528.

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

Zhou, H., and Wang, C. Y. (2001). Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society, Series B*, **62**, 657–665.

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

Table 1: Results of the simulation study with 1,000 replications. NV, RC, CW, and COR stand for the naive, regression calibration, Cheng and Wang (2001) and the error corrected estimators. SD, ESE and CP denote the standard deviation of the estimates, estimated standard error based on the formula, and the 95% coverage probabilities. The errors $U^* \sim \text{Normal}(0,1)$. All entries are multiplied by 10.

	$n=500$								$n=1,000$							
	NV		RC		CW		COR		NV		RC		CW		COR	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
	Censoring depends on X and Z								Censoring depends on X and Z							
	20% Censoring								20% Censoring							
Bias	−0.78	−3.75	−0.78	−1.67	−1.47	−0.76	0.21	0.31	−0.79	−3.76	−0.79	−1.68	−1.54	−0.83	0.11	0.35
SD	0.89	0.71	0.90	0.98	0.98	1.40	1.47	2.21	0.64	0.48	0.64	0.65	0.70	0.93	1.02	1.76
MAD	0.92	0.74	0.92	1.01	1.02	1.40	1.40	1.99	0.61	0.48	0.60	0.66	0.67	0.89	1.02	1.65
ESE							1.43	2.06							1.00	1.77
CP							9.24	9.46							9.42	9.49
	50% Censoring								50% Censoring							
Bias	−1.25	−3.90	−1.25	−1.87	−3.61	−2.34	0.33	0.54	−1.26	−3.91	−1.27	−1.89	−3.65	−2.38	0.25	0.50
SD	1.09	0.85	1.09	1.15	1.05	1.54	1.75	2.71	0.75	0.55	0.76	0.74	0.73	1.04	1.24	1.92
MAD	1.07	0.83	1.08	1.16	1.01	1.56	1.60	2.29	0.69	0.55	0.71	0.77	0.72	1.07	1.16	1.79
ESE							1.66	2.60							1.11	1.80
CP							9.61	9.23							9.51	9.58
	Censoring is independent of X and Z								Censoring is independent of X and Z							
	20% Censoring								20% Censoring							
Bias	−0.59	−3.72	−0.58	−1.62	0.07	0.23	0.19	0.30	−0.59	−3.73	−0.59	−1.63	0.03	0.17	0.09	0.32
SD	0.88	0.71	0.89	0.97	1.04	1.46	1.51	2.36	0.60	0.48	0.60	0.65	0.71	0.99	0.99	1.80
MAD	0.95	0.71	0.91	0.97	1.06	1.45	1.41	2.13	0.58	0.47	0.58	0.65	0.69	0.96	0.94	1.79
ESE							1.32	2.27							0.91	1.70
CP							9.53	9.30							9.43	9.44
	50% Censoring								50% Censoring							
Bias	−0.54	−3.78	−0.54	−1.69	0.08	0.27	0.18	0.31	−0.57	−3.80	−0.58	−1.73	0.04	0.17	0.11	0.37
SD	1.01	0.83	1.02	1.13	1.31	1.83	1.56	2.52	0.69	0.56	0.70	0.76	0.94	1.22	1.08	1.95
MAD	1.04	0.81	1.04	1.14	1.33	1.72	1.48	2.42	0.65	0.54	0.65	0.74	0.90	1.19	1.08	1.88
ESE							1.33	2.35							1.05	1.98
CP							9.42	9.32							9.51	9.59

Table 2: Results of the simulation study with 1,000 replications. NV, RC, CW, and COR stand for the naive, regression calibration, Cheng and Wang (2001) and the error corrected estimators. SD, ESE and CP denote the standard deviation of the estimates, estimated standard error based on the formula, and the 95% coverage probabilities. 95% coverage probabilities are denoted by CP. The errors $U^* \sim$ Uniform$(-1.75, 1.75)$. All entries are multiplied by 10.

	$n=500$								$n=1,000$							
	NV		RC		CW		COR		NV		RC		CW		COR	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
	Censoring depends on X and Z								Censoring depends on X and Z							
	20% Censoring								20% Censoring							
Bias	−0.78	−3.79	−0.79	−1.69	−1.46	−0.72	0.22	0.37	−0.81	−3.79	−0.82	−1.70	−1.53	−0.77	0.08	0.27
SD	0.90	0.73	0.90	0.98	0.99	1.40	1.33	2.34	0.63	0.49	0.63	0.69	0.69	0.99	0.97	1.60
MAD	0.89	0.72	0.91	1.00	1.01	1.38	1.30	2.26	0.62	0.51	0.62	0.71	0.67	1.04	0.95	1.59
ESE							1.21	2.40							0.84	1.72
CP							9.42	9.59							9.43	9.69
	50% Censoring								50% Censoring							
Bias	−1.26	−3.94	−1.26	−1.89	−3.60	−2.32	0.35	0.54	−1.29	−3.95	−1.29	−1.90	−3.63	−2.33	0.19	0.40
SD	1.10	0.89	1.11	1.14	1.05	1.51	1.68	2.62	0.75	0.57	0.76	0.78	0.72	1.08	1.18	1.76
MAD	1.09	0.89	1.07	1.21	1.05	1.56	1.56	2.43	0.72	0.58	0.71	0.80	0.73	1.12	1.12	1.67
ESE							1.65	2.54							1.11	1.76
CP							9.68	9.43							9.62	9.65
	Censoring is independent of X and Z								Censoring is independent of X and Z							
	20% Censoring								20% Censoring							
Bias	−0.59	−3.76	−0.58	−1.65	0.10	0.25	0.16	0.31	−0.60	−3.76	−0.68	−1.65	0.03	0.20	0.06	0.24
SD	0.89	0.71	0.90	0.97	1.10	1.53	1.40	2.04	0.60	0.49	0.60	0.68	0.71	1.04	0.94	1.61
MAD	0.95	0.72	0.93	0.98	1.11	1.53	1.35	1.96	0.55	0.52	0.57	0.73	0.69	1.08	0.93	1.57
ESE							1.30	2.00							0.92	1.46
CP							9.59	9.56							9.58	9.40
	50% Censoring								50% Censoring							
Bias	−0.55	−3.82	−0.55	−1.73	0.09	0.30	0.20	0.37	−0.58	−3.83	−0.59	−1.74	0.04	0.21	0.08	0.30
SD	1.02	0.81	1.03	1.10	1.36	1.84	1.55	2.40	0.69	0.56	0.70	0.78	0.93	1.24	1.03	1.87
MAD	1.03	0.82	1.04	1.09	1.36	1.79	1.54	2.22	0.65	0.56	0.64	0.81	0.88	1.25	1.00	1.83
ESE							1.42	2.30							0.90	1.61
CP							9.55	9.57							9.32	9.56

Table 3: Results of the simulation study with $2,000$ replications. NV, RC, CW, and COR stand for the naive, regression calibration, Cheng and Wang (2001) and the error corrected estimators. SD and MSE denote the standard deviation and the mean squared error, respectively. The errors $U^* \sim$ Uniform$(-1.75, 1.75)$. All entries are multiplied by 10.

Method		Case 1			Case 2			Case 3		
		Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
NV	β_1	−0.44	1.18	0.15	−0.41	1.04	0.13	−2.08	1.30	0.60
	β_2	−4.09	0.91	1.76	−4.12	0.79	1.76	−4.44	0.93	2.06
RC	β_1	−0.44	1.18	0.16	−0.42	1.05	0.13	−2.08	1.30	1.26
	β_2	−2.02	1.23	0.59	−2.13	1.08	0.58	−2.56	1.25	0.82
CW	β_1	−0.56	3.32	1.13	−3.51	1.22	1.38	−5.54	1.17	3.22
	β_2	−0.33	4.89	2.39	−3.82	1.45	1.67	−5.28	1.57	3.04
COR	β_1	0.10	1.49	0.23	0.09	1.29	0.17	0.44	1.96	0.40
	β_2	0.42	2.80	0.80	0.45	2.40	0.59	0.50	2.37	0.59

Table 4: Analysis of the ACTG 175 aids clinical trial data. Est. and SE stand for estimate and standard error, respectively. Z, Z+D, Z+Z, and D stand for zidovudine, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine, respectively. Here Est. and SE stand for the estimates and standard errors, respectively. For the NV, RC, and CW methods, the standard errors were calculated based on 5,000 bootstrap samples. For the COR method, the standard errors are based on asymptotic results.

Covariates	NV		RC		CW		COR	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Z+D (Ref: Z)	−0.780	0.325	−0.759	0.325	−0.164	0.125	−0.801	0.344
Z+Z (Ref: Z)	−1.004	0.340	−0.999	0.344	−0.267	0.103	−0.999	0.361
D (Ref: Z)	−0.748	0.313	−0.753	0.314	−0.217	0.112	−0.814	0.338
log(CD4)	−2.186	0.404	−2.576	0.477	−0.854	0.194	−2.700	0.570

Table 5: Analysis of the ACTG 175 aids clinical trial data using the error corrected method with $f\{\Lambda(u), \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\} = 1/\{1 + \Lambda(u)\eta(X^*, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})\}^r$. Est. and SE stand for estimate and standard error, respectively. Z, Z+D, Z+Z, and D stand for zidovudine, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine, respectively.

Covariates		$r=0$	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	$r=10$	$r=15$
Z+D (Ref: Z)	Est.	−0.783	−0.793	−0.803	−0.813	−0.823	−0.830	−0.869	−0.904
	SE	0.361	0.358	0.355	0.352	0.350	0.348	0.340	0.335
Z+Z (Ref: Z)	Est.	−0.979	−0.990	−1.002	−1.012	−1.023	−1.033	−1.080	−1.120
	SE	0.366	0.364	0.361	0.358	0.356	0.354	0.347	0.346
D (Ref: Z)	Est.	−0.794	−0.805	−0.815	−0.825	−0.834	−0.843	−0.879	−0.906
	SE	0.339	0.337	0.334	0.332	0.329	0.328	0.325	0.329
log(CD4)	Est.	−2.685	−2.691	−2.697	−2.702	−2.705	−2.708	−2.703	−2.679
	SE	0.567	0.561	0.557	0.554	0.551	0.550	0.557	0.584

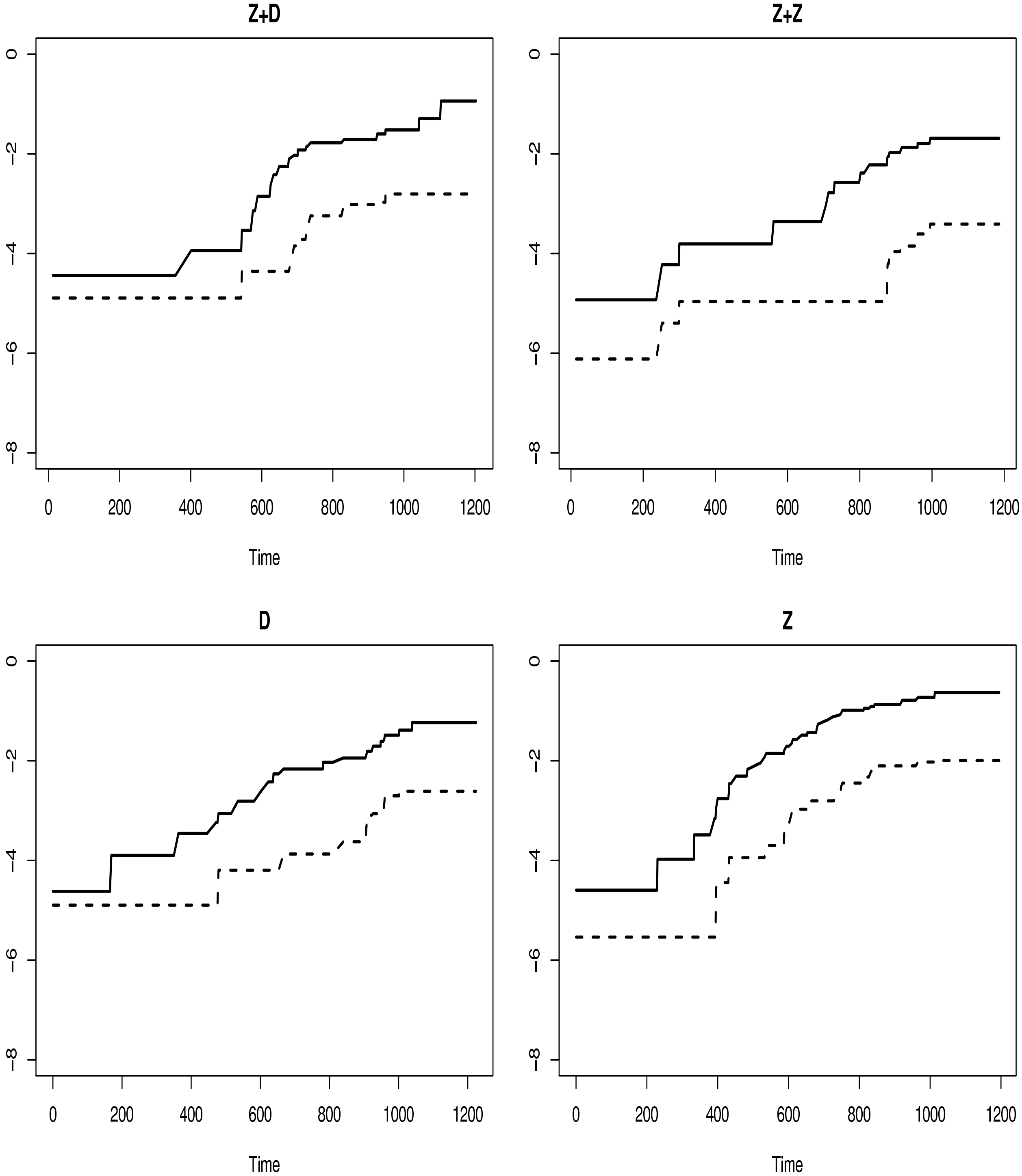

Figure 1: Plot of $\log\{\mathrm{pr}(T \leq t)/\mathrm{pr}(T > t)\}$ versus t for each treatment group. The solid and dotted curves correspond to the cases of $X \leq -0.1$ and $X > -0.1$ respectively, where $X = \log(\text{True CD4 count}) - 5.89$. Z, Z+D, Z+Z, and D represent zidovudine, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine, respectively.