Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Bayesian Analysis for Partly Linear Cox Model with Measurement Error and Time-Varying Covariate Effect

Anqi Pan | Xiao Song* | Hanwen Huang*

[1]Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Georgia, United States

Correspondence
*Xiao Song, Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, 101 Buck Road, Athens, GA 30602. Email: xsong@uga.edu
*Hanwen Huang, Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, 101 Buck Road, Athens, GA 30602. Email: huanghw@uga.edu

Present Address
Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, 101 Buck Road, Athens, GA 30602.

Summary

The Cox proportional hazards model is commonly used to estimate the association between time-to-event and covariates. Under the proportional hazards assumption, covariate effects are assumed to be constant in the follow-up period of study. When measurement error presents, common estimation methods that adjust for an error-contaminated covariate in the Cox proportional hazards model assume that the true function on the covariate is parametric and specified. We consider a semi-parametric partly linear Cox model that allows the hazard to depend on an unspecified function of an error-contaminated covariate and an error-free covariate with time-varying effect, which simultaneously relaxes the assumption on the functional form of the error-contaminated covariate and allows for non-constant effect of the error-free covariate. We take a Bayesian approach and approximate the unspecified function by a B-spline. Simulation studies are conducted to assess the finite sample performance of the proposed approach. The results demonstrate that our proposed method has favorable statistical performance. The proposed method is also illustrated by an application to data from the AIDS Clinical Trials Group (ACTG) Protocol 175.

KEYWORDS:
Cox model, Measurement error model, time-to-event outcome, Bayesian methods, Semi-parametric regression, time-varying coefficient

1 | INTRODUCTION

In survival data analysis, the Cox proportional hazards model is widely used to investigate the relationship between covariates and censored survival time.[1] In many biomedical applications, however, covariates are subject to measurement error and not directly observable due to natural biological fulctuation and instrument error.[2,3,4,5] Measurement errors of biomarkers in clinical studies include both laboratory error and variations due to sampling, storage and within-subject variability.[6,7] When covariates are measured with error, characterizing the association between the survival time and the true underlying covariates is crucial for drawing inference correctly. In practice, for regression analysis, it is common to naively use the mismeasured covariates, which can lead to substantial bias to the estimates of the parameters of interest and yield misleading conclusions.[8]

For the recent decades, people have been exploring alternative strategies to account for the measurement error for the Cox proportional hazards model. Among those methods, approximation methods like regression calibration and SIMEX are commonly used.[9,10] Score methods are also popular, including parametric corrected score, nonparametric corrected score, conditional score

[0]**Abbreviations:** Add if there is any abbreviations.

and refined corrected score.[11,12,13,14] There are some other widely used methods, such as Bayesian methods and seminonparametric(SNP) likelihood.[3,15] But the true function on the error-contaminated covariate is usually specified in the model.[12,16,17,18] It is common to specify a linear term for the error-contaminated covariate.[12,13,14,19] Some allow including a specified nonlinear function of the error contaminated covariates.[20,21] However, using a specified function for the unobserved covariate can incur biased estimations when the function is misspecified. For example, when the true relationship between the error-contaminated covariate and the log hazard is nonlinear, assuming a simple linear relationship has an unclear effect on the model parameter estimates.

Under the condition of no measurement error presenting in the data, researchers have considered modeling the nonlinear covariate effect using the partly linear additive Cox model, which allows flexibility on modeling the covariate effects.[22,23,24,25] With measurement errors presenting, Bayesian approaches have been developed to deal with the measurement error under this model.[26,27] However, the linear covariate effect is assumed constant over time. But in practice, the covariate effect may change over time and the constant effect assumption may not hold.[28,29,30] A common example is that the treatment effect on survival may fade over time. Statisticians have proposed using time-varying coefficients for such cases.[29,31,32] Nevertheless, to the best of our knowledge, no existing methods consider nonparametric error-contaminated covariate and time-varying error free covariate effect simultaneously.

In this article, we develop a Bayesian approach for the partly linear Cox model that allows the hazard to depend on an unspecified function of an error-prone covariate and a linear error-free covariate with time-varying effect. The unknown function is approximated by a linear B-spline and the time-varying effect is approximated by a piecewise constant function. Our contribution lies in two aspects. First, we simultaneously consider both nonparametric error-contaminated covariate and time-varying error-free covariate effect, which is much more challenging than handling only one of these complexities. Second, we investigated knot selection using the deviance information criterion (DIC), while existing Bayesian spline models for measurement-error problems usually fix the number of knots. For example, Bhadra and Carroll[33] set the number of knots at 25, and Cheng and Crainiceanu[26] experimented with fixing it at 10 and 20.

The remainder of this article is organized as follows: in Section 2, we present the model. The likelihood function, Bayesian algorithm for parameter estimation and model selection criteria are described in Section 3. In Section 4, we present the simulation studies. In Section 5, we illustrate the proposed method on AIDS Clinical Trials Group (ACTG) Protocol 175 data. Finally, in Section 6, we present conclusions and discussions.

2 | JOINT MODELS

In this section, we introduce joint models for measurement error and time-to-event outcome.

2.1 | Measurement Error Model

We define n to be the sample size and m_i to be the number of repeated measures for subject i for $i = 1, \dots, n$. For subject i, let X_i be an error-prone covariate, Z_i be a vector of error-free covariates, and W_{il}, $l = 1, \dots, m_i$, be the replicated measurements of X_i. For simplicity of presentation, we consider the case where Z_i is one-dimentional. We consider a classical additive measurement error model,

$$W_{il} = X_i + \epsilon_{il}, \quad i = 1, \dots, n, l = 1, \dots, m_i, \tag{1}$$

and the errors ϵ_{il} are independently and identically distributed normal random variable with mean 0 and variance σ^2.

2.2 | The Survival Model

Let T_i^* and C_i denote the true event time and censoring time for subject i. The observed event time is defined as $T_i = min(T_i^*, C_i)$. The event indicator is defined as $\delta_i = I(T_i^* \leq C_i)$.

We consider a partly linear Cox model and assume the hazard function as follows:

$$h\left(t|X_i, Z_i\right) = h_0(t) \exp\left(g(X_i) + \beta(t) Z_i\right), \tag{2}$$

where $h_0(t)$ is an unspecified baseline hazard, $g(\cdot)$ denotes an unknown function of X_i, and $\beta(t)$ denotes the unspecified time varying coefficient of Z_i. For identifiability of the parameters, we set $g(x_0) = 0$ for a fixed x_0.

3 | METHODS

In this section, we present the Bayesian algorithm for parameter estimation and inference. Combining the measurement error and survival models, we derive the likelihood function for the joint models. We specify the prior distribution and derive the conditional posterior distribution for each parameter. We present the method for selecting number of knots in B-splines, baseline hazard, time-varying coefficient and how to place these knots accordingly.

3.1 | Approximation

The baseline hazard $h_0(t)$ and the time varying coefficient $\beta(t)$ are approximated by piecewise constant functions $h_0^*(t) = \sum_{j=1}^{J} I(t \in I_j)\lambda_j$ and $\beta^*(t) = \sum_{j=1}^{J} I(t \in I_j)\beta_j$, which are constants λ_j and β_j, respectively, over time intervals $I_j = (c_{j-1}, c_j]$, $j = 1, \ldots, J$, $c_0 = 0 < c_1 < \ldots < c_J < c_{J+1} = \infty$.

We use a degree one B-spline $g^*(\cdot)$ to approximate the unknown function $g(\cdot)$. The B-spline is defined on an interval $[\omega_0, \omega_K]$ with $\omega_0 = min_{il}(W_{il})$ and $\omega_K = max_{il}(W_{il})$ for $i = 1, \ldots, n$, and $l = 1, \ldots, m_i$. And $\omega_1, \ldots, \omega_{K-1}$ are $K - 1$ internal knots. Then we have $g^*(x) = \sum_{k=1}^{K+1} B_k(x)\theta_k$, where $\theta_1, \ldots, \theta_{K+1}$ are coefficients of B-splines, $B(x) = \{B_1(x), \ldots, B_{K+1}(x)\}$ denotes the basis-functions of degree one B-spline, and

$$B_1(x) = \frac{\omega_1 - x}{\omega_1 - \omega_0} I(\omega_0 \le x < \omega_1),$$

$$B_{K+1}(x) = \frac{x - \omega_{K-1}}{\omega_K - \omega_{K-1}} I(\omega_{K-1} \le x < \omega_K),$$

and

$$B_k(x) = \frac{x - \omega_{k-2}}{\omega_{k-1} - \omega_{k-2}} I(\omega_{k-2} \le x < \omega_{k-1}) + \frac{\omega_k - x}{\omega_k - \omega_{k-1}} I(\omega_{k-1} \le x < \omega_k),$$

for $k = 2, \ldots, K$.[33]

Then the hazard function is approximated as

$$\begin{aligned} h\left(t|X_i, Z_i\right) &= h_0(t) \exp\left(g(X_i) + \beta(t)Z_i\right) \\ &\cong h_0^*(t) \exp\left(g^*(X_i) + \beta^*(t)Z_i\right). \end{aligned}$$

3.2 | Joint Likelihood Function

The joint likelihood can be written as

$$L = L_1 \cdot L_2, \tag{3}$$

where the likelihood for the measurement error sub-model is written as

$$L_1 \propto \prod_{i=1}^{n} \prod_{l=1}^{m_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(W_{il} - X_i\right)^2\right],$$

and the likelihood for the survival sub-model is written as

$$L_2 \propto \prod_{\{i:\delta_i=1\}} h(t_i) \prod_{i=1}^{n} S(t_i),$$

where

$$\prod_{\{i:\delta_i=1\}} h(t_i) \cong \prod_{j=1}^{J} \lambda_j^{d_j} \prod_{i=1}^{n} \exp\left\{\delta_i \left[g^*(X_i) + \beta^*(t_i)Z_i\right]\right\}, \tag{4}$$

and

$$\begin{aligned}\prod_{i=1}^{n} S(t_i) &= \prod_{i=1}^{n} \exp\left[-\int_0^{t_i} h(t)dt\right] \\ &\cong \prod_{i=1}^{n} \exp\left[-\int_0^{t_i} h_0^*(t)\exp\left(g^*(X_i)+\beta^*(t)Z_i\right)dt\right] \\ &\cong \prod_{i=1}^{n} \exp\left[\exp\left(g^*(X_i)\right)\left(-\sum_{j=1}^{J}\lambda_j \exp(\beta_j Z_i)\Delta t_{ij}\right)\right].\end{aligned} \tag{5}$$

Combining (4) and (5), we have

$$L_2 \propto \prod_{j=1}^{J}\lambda_j^{d_j}\prod_{i=1}^{n}\exp\left\{\delta_i\left[g^*(X_i)+\beta^*(t_i)Z_i\right]\right\}\cdot\prod_{i=1}^{n}\exp\left[\exp\left(g^*(X_i)\right)\left(-\sum_{j=1}^{J}\lambda_j\exp(\beta_j Z_i)\Delta t_{ij}\right)\right],$$

where for $i = 1,\dots,n$ and $j = 1,\dots,J$, $\Delta t_{ij} = 0$ when $t_i \le c_{j-1}$, $\Delta t_{ij} = t_i - c_{j-1}$ when $t_i \in (c_{j-1}, c_j]$, $\Delta t_{ij} = c_j - c_{j-1}$ when $t_i > c_j$, and d_j is the number of events in the jth interval.

3.3 | Bayesian Approaches for Joint Model with B-spline

The complete joint likelihood under Bayesian framework can be described by (3). The observed data are $D = \{(T_i, \delta_i, W_{i1},\dots,W_{im_i}, Z_i): i = 1,\dots,n\}$, and the vector of parameters is $\Theta = \left(\sigma^2, \lambda_1,\dots,\lambda_J,\beta_1,\dots,\beta_J,\theta_1,\dots,\theta_{K+1},X_1,\dots,X_n\right)^T$.

In Bayesian analysis, random walk priors can be used with spline models as local smoothers to avoid overfitting and enforce smoothness across the coefficients.[26,34,35,36] In the model fitting, we used the first-order random walk priors (RW1) for $\beta^*(t)$ and $g^*(X_i)$. Specifically, for $\beta^*(t) = \sum_{j=1}^{J} I(t \in I_j)\beta_j$, the prior of $(\beta_1,\dots,\beta_J)$ is set as follows: $\beta_1 \sim N\left(0, v_1^2\right)$ and $\beta_j - \beta_{j-1} \sim N\left(0,\sigma_\beta^2\right)$ for $j = 2,\dots,J$. Similarly, for $g^*(x) = \sum_{k=1}^{K+1} B_k(x)\theta_k$, the prior of $(\theta_1,\dots,\theta_{K+1})$ satisfies $\theta_1 \sim N\left(0, v_2^2\right)$ and $\theta_k - \theta_{k-1} \sim N\left(0,\sigma_\theta^2\right)$ for $k = 2,\dots,K+1$. Consequently, the prior density of $(\beta_1,\dots,\beta_J)$ is

$$\frac{1}{\sqrt{2\pi v_1^2}}\exp\left[-\frac{\beta_1^2}{2v_1^2}\right]\cdot\prod_{j=2}^{J}\frac{1}{\sqrt{2\pi\sigma_\beta^2}}\exp\left[-\frac{1}{2\sigma_\beta^2}\left(\beta_j-\beta_{j-1}\right)^2\right],$$

and the prior density of $(\theta_1,\dots,\theta_{K+1})$ is

$$\frac{1}{\sqrt{2\pi v_2^2}}\exp\left[-\frac{\theta_1^2}{2v_2^2}\right]\cdot\prod_{k=2}^{K+1}\frac{1}{\sqrt{2\pi\sigma_\theta^2}}\exp\left[-\frac{1}{2\sigma_\theta^2}\left(\theta_k-\theta_{k-1}\right)^2\right].$$

To obtain conjugacy, we use the normal prior for the normal mean, the inverse-gamma priors for the normal variances and gamma priors for the baseline hazard parameters.[16,17,37] We chose the values of the hyperparameters following the common practice in Bayesian analysis for vague and minimal informative priors, e.g. a normal distribution with zero mean and very large variance or a gamma or an inverse-gamma distribution with very small shape and rate.[16,17,37,38,39,40] For the measurement error variance σ^2, we use an inverse gamma prior $IG(e_0, f_0)$. For baseline hazard parameters λ_j $(j = 1,\dots,J)$, we use a gamma prior $G(e_1, f_1)$. For the random walk prior variances σ_β^2 and σ_θ^2, we use inverse gamma priors $IG(e_2, f_2)$ and $IG(e_3, f_3)$, respectively. For unobserved X_i $(i = 1,\dots,n)$, we use a normal prior $N(\mu_0, \sigma_0^2)$. The hyperprior for μ_0 is chosen as a normal $N(0, v_0^2)$ and the hyperprior for σ_0^2 is chosen as an inverse gamma $IG(e_4, f_4)$. We set the values of the hyperparameters above as $e_0 = 0.001$, $f_0 = 0.001$, $e_1 = 0.001$, $f_1 = 0.001$, $e_2 = 0.001$, $f_2 = 0.001$, $e_3 = 0.001$, $f_3 = 0.001$, $e_4 = 0.001$, $f_4 = 0.001$, $v_0 = 1000$, $v_1 = 1000$, $v_2 = 1000$. We derive the conditional posterior distributions as follows, where *rest* denotes the other parameters in the model.

1.

$$\left(\sigma^2 \mid D, rest\right) \sim \text{Inv}-\text{Gamma}\left(e_0 + \frac{\sum_{i=1}^{n} m_i}{2}, f_0 + \frac{\sum_{i=1}^{n}\sum_{l=1}^{m_i}\left(W_{il}-X_i\right)^2}{2}\right) \tag{6}$$

2.
$$\left(\lambda_j \mid D, rest\right) \sim \text{Gamma}\left(d_j + e_1, \left(\sum_{i=1}^{n} \Delta t_{ij} \exp\left(g^*(X_i) + \beta_j Z_i\right)\right) + f_1\right) \tag{7}$$

3.
$$\left(\beta_j \mid D, rest\right) \propto N\left(\beta_j \mid \cdot\right) L(\beta_j \mid \cdot), \tag{8}$$
where $N(\beta_1 \mid \cdot) \propto N(\beta_1 \mid 0, v_1^2) N(\beta_2 - \beta_1 \mid 0, \sigma_\beta^2)$, $N(\beta_J \mid \cdot) \propto N(\beta_J - \beta_{J-1} \mid 0, \sigma_\beta^2)$, $N(\beta_j \mid \cdot) \propto N(\beta_{j+1} - \beta_j \mid 0, \sigma_\beta^2) N(\beta_j - \beta_{j-1} \mid 0, \sigma_\beta^2)$ for $j = 2, \ldots, J-1$, and
$$L(\beta_j \mid \cdot) \propto \prod_{\{i : t_i \in (c_{j-1}, c_j]\}} \exp\left\{\delta_i \left[g^*(X_i) + \beta_j Z_i\right]\right\} \cdot \prod_{i=1}^{n} \exp\left[\exp\left(g^*(X_i)\right)\left(-\sum_{j=1}^{J} \lambda_j \exp(\beta_j Z_i) \Delta t_{ij}\right)\right].$$
Since $N\left(\beta_j \mid \cdot\right)$ and $L(\beta_j \mid \cdot)$ are both log-concave, $(\beta_j \mid D, rest)$ is also log-concave.

4.
$$\left(\sigma_\beta^2 \mid D, rest\right) \sim \text{Inv} - \text{Gamma}\left(e_2 + \frac{J-1}{2}, f_2 + \frac{\sum_{j=2}^{J} \left(\beta_j - \beta_{j-1}\right)^2}{2}\right) \tag{9}$$

5.
$$\left(\theta_k \mid D, rest\right) \propto N\left(\theta_k \mid \cdot\right) L\left(\theta_k \mid \cdot\right), \tag{10}$$
where $N(\theta_1 \mid \cdot) \propto N(\theta_1 \mid 0, v_2^2) N(\theta_2 - \theta_1 \mid 0, \sigma_\theta^2)$, $N(\theta_{K+1} \mid \cdot) \propto N(\theta_{K+1} - \theta_K \mid 0, \sigma_\theta^2)$, $N(\theta_k \mid \cdot) \propto N(\theta_{k+1} - \theta_k \mid 0, \sigma_\theta^2) N(\theta_k - \theta_{k-1} \mid 0, \sigma_\theta^2)$ for $k = 2, \ldots, K$, and
$$\begin{aligned} L\left(\theta_k \mid \cdot\right) &\propto \prod_{i=1}^{n} \exp\left\{\delta_i \left[g^*(X_i) + \beta^*(t_i) Z_i\right]\right\} \cdot \prod_{i=1}^{n} \exp\left[\exp\left(g^*(X_i)\right)\left(-\sum_{j=1}^{J} \lambda_j \exp(\beta_j Z_i) \Delta t_{ij}\right)\right] \\ &\propto \prod_{i=1}^{n} \exp\left\{\delta_i \left[\sum_{k=1}^{K+1} \mathrm{B}_k\left(X_i\right) \theta_k + \beta^*(t_i) Z_i\right]\right\} \\ &\cdot \prod_{i=1}^{n} \exp\left[\exp\left(\sum_{k=1}^{K+1} \mathrm{B}_k\left(X_i\right) \theta_k\right)\left(-\sum_{j=1}^{J} \lambda_j \exp(\beta_j Z_i) \Delta t_{ij}\right)\right]. \end{aligned}$$
Since $N\left(\theta_k \mid \cdot\right)$ and $L\left(\theta_k \mid \cdot\right)$ are both log-concave, $\left(\theta_k \mid D,\ rest\right)$ is also log-concave.

6.
$$\left(\sigma_\theta^2 \mid D, rest\right) \sim \text{Inv} - \text{Gamma}\left(e_3 + \frac{K}{2}, f_3 + \frac{\sum_{k=2}^{K+1} \left(\theta_k - \theta_{k-1}\right)^2}{2}\right) \tag{11}$$

7.
$$\left(\mu_0 \mid \sigma_0^2, X_1, \ldots, X_n\right) \sim N\left(\left(n\sigma_0^{-2} + v_0^{-2}\right)^{-1} n\overline{X}\sigma_0^{-2}, \left(n\sigma_0^{-2} + v_0^{-2}\right)^{-1}\right) \tag{12}$$

8.
$$\left(\sigma_0^2 \mid \mu_0, X_1, \ldots, X_n\right) \sim \text{Inv} - \text{Gamma}\left(e_4 + \frac{n}{2}, f_4 + \frac{1}{2}\sum_{i=1}^{n} \left(X_i - \mu_0\right)^2\right) \tag{13}$$

9.
$$\left(X_i \mid D,\ \text{rest}\ \right) \propto N\left(\mu_0, \sigma_0^2\right) L\left(X_i \mid \cdot\right), \tag{14}$$
where

$$
\begin{aligned}
L\left(X_i \mid \cdot\right) &\propto \prod_{l=1}^{m_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(W_{il}-X_i\right)^2\right] \cdot \exp\left\{\delta_i\left[g^*(X_i)+\beta^*(t_i)Z_i\right]\right\} \\
&\quad \cdot \exp\left[\exp\left(g^*(X_i)\right)\left(-\sum_{j=1}^{J}\lambda_j \exp(\beta_j Z_i)\Delta t_{ij}\right)\right] \\
&\propto \prod_{l=1}^{m_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(W_{il}-X_i\right)^2\right] \cdot \exp\left\{\delta_i\left[\sum_{k=1}^{K+1} \mathrm{B}_k\left(X_i\right)\theta_k + \beta^*(t_i)Z_i\right]\right\} \\
&\quad \cdot \exp\left[\exp\left(\sum_{k=1}^{K+1} \mathrm{B}_k\left(X_i\right)\theta_k\right)\left(-\sum_{j=1}^{J}\lambda_j \exp(\beta_j Z_i)\Delta t_{ij}\right)\right].
\end{aligned}
$$

Since $N\left(\mu_0, \sigma_0^2\right)$ is log-concave and $L\left(X_i \mid \cdot\right)$ is piecewise log-concave, $\left(X_i \mid D,\ rest\right)$ is also piecewise log-concave.

3.4 | Bayesian Algorithm

Since $\sigma^2, \lambda_1, \ldots, \lambda_J, \sigma_\beta^2, \sigma_\theta^2, \mu_0$ and σ_0^2 have conjugate priors, they can be sampled directly from their posteriors (6), (7), (9), (11), (12), (13). For $\beta_1, \ldots, \beta_J, \theta_1, \ldots, \theta_{K+1}, X_1, \ldots, X_n$, since their corresponding conditional posterior densities are log-concave or piecewise log-concave, they can be sampled using an adaptive rejection algorithm.[41]

Algorithm:

Step 0: Initialize parameters. Sample from the joint posterior

$$\Theta = \left(\sigma^2, \lambda_j, \beta_j, \theta_k, X_i \mid T_i, \delta_i, W_{i1}, \ldots, W_{im_i}, Z_i\right), i = 1, \ldots, n, j = 1, \ldots, J, k = 1, \ldots, K+1.$$

Step 1: Simulate $\sigma^2 \mid rest$ from posterior (6).

Step 2: Simulate $\lambda_j \mid rest$ from posterior (7).

Step 3: Simulate $\beta_j \mid rest$ from log-concave posterior density (8).

Step 4: Simulate $\sigma_\beta^2 \mid rest$ from posterior (9).

Step 5: Simulate $\theta_k \mid rest$ from log-concave posterior density (10).

Step 6: Simulate $\sigma_\theta^2 \mid rest$ from posterior (11).

Step 7: Simulate $\mu_0 \mid rest$ from posterior (12).

Step 8: Simulate $\sigma_0^2 \mid rest$ from posterior (13).

Step 9: By numerical integration of (14), we first compute the posterior probability of X_i in each interval $\left[w_0, w_1\right], \ldots, \left[w_{K-1}, w_K\right]$, and simulate the index k of the interval that X_i belongs to from the multinomial distribution with the calculated probabilities. We then simulate $\left(X_i \mid rest\right)$ from the piecewise log-concave posterior density (14) truncated by the interval $\left[w_{k-1}, w_k\right]$.

Repeat steps 1-9 until convergence is achieved.

We conducted parameter estimation via the Markov chain Monte Carlo algorithm under a Bayesian framework in R, version 3.3.4. Adaptive rejection algorithm was implemented via R package 'Runuran'.[42] We discarded the first 1000 Markov chain Monte Carlo samples, and inference was conducted using a total of 4000 posterior samples after burn-in. Convergence was evaluated on the basis of deviance and trace plots of model parameters. Convergence was achieved within the burn-in iterations.

3.5 | Inference for Goodness-of-fit and Selection of Number of Knots

We used the DIC to assess the trade-off between model fit and complexity and to select the number of knots for B-splines.[43] The DIC is defined in R2WInBUGS[44] as

$$DIC = \overline{D(\theta^*)} + p_D,$$

where $\overline{D(\theta^*)}$ is the posterior mean deviance to summarize fit, p_D is the additional term to summarize complexity. The posterior mean deviance $\overline{D(\theta^*)}$ is a Bayesian measure of how well the model fits the data, the smaller it is, the better the model fits. The term p_D measures the effective number of parameters included in the model, the larger the effective number of parameters is, the easier it is for the model to fit the data, and the larger deviance should be penalized. Therefore, DIC is considered as a Bayesian measure of fit, penalized by an additional complexity term.[43]

Here, the deviance is defined as

$$D(\theta^*) = -2log\,L(Data|\theta^*),$$

the additional complexity term is defined as

$$p_D = p_v = \frac{\text{Var}\,(D\,(\theta^*))}{2},$$

which is an alternative measure for model complexity.[37,45,46] It is positive, invariant to parametrization and considered robust and accurate in estimating the effective number of model parameters.[37,45,46] For the proposed model, the likelihood used to compute the deviance is the complete likelihood as in the DIC_5 from Celeux et al. (2006)[47]:

$$\begin{aligned} L &= L\,(W, X, Z \mid \theta^*) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{1}{2\sigma_0^2}\left(X_i - \mu_0\right)^2\right] \cdot \prod_{i=1}^{n}\prod_{l=1}^{m_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(W_{il} - X_i\right)^2\right] \\ &\cdot \prod_{j=1}^{J} \lambda_j^{d_j} \prod_{i=1}^{n} \exp\left\{\delta_i\left[g^*(X_i) + \beta^*(t_i)Z_i\right]\right\} \cdot \prod_{i=1}^{n} \exp\left[\exp\left(g^*(X_i)\right)\left(-\sum_{j=1}^{J} \lambda_j \exp(\beta_j Z_i)\Delta t_{ij}\right)\right]. \end{aligned}$$

For the naive model, the likelihood used to compute the deviance is the same as the original definition in Spiegelhalter et al. (2002)[43]:

$$\begin{aligned} L = L\left(\overline{W}, Z \mid \theta^*\right) &= \prod_{j=1}^{J} \lambda_j^{d_j} \prod_{i=1}^{n} \exp\left\{\delta_i\left[g^*(\overline{W}_{i\cdot}) + \beta^*(t_i)Z_i\right]\right\} \\ &\cdot \prod_{i=1}^{n} \exp\left[\exp\left(g^*(\overline{W}_{i\cdot})\right)\left(-\sum_{j=1}^{J} \lambda_j \exp(\beta_j Z_i)\Delta t_{ij}\right)\right], \end{aligned}$$

where $\overline{W}_{i\cdot} = m_i^{-1}\sum_{l=1}^{m_i} W_{il}$ is the mean of the replicated measures for subject i. For the ideal model, substitute $\overline{W}_{i\cdot}$ with X_i.

One may choose knots that are equally-spaced or quantile-based. In this paper, we choose locations of knots on the quantiles of the observed measurements of X with approximately equal number of events. To be more specific, for $k = 1, \ldots, K$, the number of events in the interval $I_k = \left[w_{k-1}, w_k\right]$ is $d_k = \sum_{i=1}^{n}\sum_{l=1}^{m_i} \delta_i \cdot I(W_{il} \in I_k)$. Knots were placed accordingly such that $d_1 \approx \ldots \approx d_K$. We performed experiments with other knots placement strategies, such as equally-spaced or at the quantiles of observed measurements of X. The current strategy worked best in our study.

For the piecewise constant functions that approximate $h_0(t)$ and $\beta(t)$, we set the number of intervals J proportional to the square-root of the number of events. Denoting V as the number of events, we have $J = c\sqrt{V}$. By experiment, we found $c = 0.3$ works well in our study (see Section S.1 in the supplementary materials). The $J - 1$ internal knots were placed on the quantiles of time with approximately equal number of events.

4 | SIMULATION STUDY

We conducted Monte Carlo (MC) simulations to evaluate the performance of our proposed method. We compare our proposed method with the ideal method that uses the true value of the unobserved latent covariate and the naive method that replaces X_i by $\overline{W}_{i\cdot}$.

TABLE 1 Simulation Scenario (1): X is with normal error and independent of Z.

			n=1000					n=2000				
	σ	Method	Mean Bias	Mean SE	Emp. SE	95% CR	MSE	Mean Bias	Mean SE	Emp. SE	95% CR	MSE
		Ideal	0.064	0.157	0.129	96	0.022	0.039	0.124	0.101	97	0.013
	0.3	Proposed	0.073	0.166	0.140	94	0.027	0.046	0.130	0.107	97	0.015
$\beta(t)$		Naive	0.206	0.149	0.124	71	0.066	0.192	0.116	0.095	58	0.052
	0.5	Proposed	0.078	0.175	0.143	95	0.028	0.048	0.138	0.110	97	0.015
		Naive	0.338	0.147	0.123	35	0.142	0.330	0.114	0.094	22	0.128
		Ideal	0.042	0.161	0.186	88	0.039	0.027	0.121	0.140	88	0.022
	0.3	Proposed	0.058	0.243	0.231	93	0.061	0.055	0.179	0.179	89	0.038
$g(x)$		Naive	0.157	0.166	0.178	75	0.069	0.139	0.123	0.133	71	0.049
	0.5	Proposed	0.078	0.298	0.246	96	0.072	0.069	0.230	0.213	92	0.054
		Naive	0.286	0.162	0.162	58	0.158	0.272	0.121	0.120	50	0.136

TABLE 2 Simulation Scenario (2): X is with non-normal error and independent of Z.

			n=1000					n=2000				
	σ	Method	Mean Bias	Mean SE	Emp. SE	95% CR	MSE	Mean Bias	Mean SE	Emp. SE	95% CR	MSE
		Ideal	0.064	0.157	0.129	96	0.022	0.039	0.124	0.101	97	0.013
	0.3	Proposed	0.059	0.166	0.155	94	0.029	0.045	0.132	0.104	98	0.014
$\beta(t)$		Naive	0.157	0.150	0.139	79	0.049	0.172	0.118	0.094	69	0.044
	0.5	Proposed	0.062	0.176	0.163	93	0.032	0.050	0.140	0.112	98	0.016
		Naive	0.269	0.148	0.138	52	0.103	0.291	0.115	0.090	28	0.102
		Ideal	0.042	0.161	0.186	88	0.039	0.027	0.121	0.140	88	0.022
	0.3	Proposed	0.072	0.249	0.227	93	0.060	0.067	0.185	0.185	90	0.043
$g(x)$		Naive	0.126	0.170	0.200	77	0.063	0.114	0.122	0.133	74	0.039
	0.5	Proposed	0.098	0.289	0.271	93	0.090	0.110	0.219	0.200	91	0.060
		Naive	0.223	0.168	0.196	63	0.115	0.207	0.128	0.131	60	0.090

TABLE 3 Simulation Scenario (3): X is with normal error and correlated with Z.

			n=1000					n=2000				
	σ	Method	Mean Bias	Mean SE	Emp. SE	95% CR	MSE	Mean Bias	Mean SE	Emp. SE	95% CR	MSE
		Ideal	0.040	0.187	0.158	97	0.027	0.038	0.141	0.129	95	0.019
	0.3	Proposed	0.044	0.203	0.173	96	0.033	0.047	0.152	0.139	95	0.022
$\beta(t)$		Naive	0.231	0.181	0.160	76	0.082	0.246	0.136	0.129	54	0.080
	0.5	Proposed	0.047	0.222	0.195	96	0.041	0.051	0.163	0.152	95	0.027
		Naive	0.458	0.175	0.160	25	0.238	0.469	0.131	0.132	9	0.240
		Ideal	0.066	0.196	0.220	87	0.060	0.053	0.143	0.171	86	0.038
	0.3	Proposed	0.075	0.239	0.249	89	0.075	0.066	0.190	0.199	87	0.049
$g(x)$		Naive	0.128	0.190	0.206	82	0.069	0.114	0.138	0.159	76	0.050
	0.5	Proposed	0.100	0.289	0.296	89	0.107	0.085	0.216	0.223	88	0.063
		Naive	0.198	0.183	0.194	70	0.125	0.188	0.130	0.151	65	0.113

For each data set, we simulated the time-to-event endpoint from a Cox model with a Weibull baseline hazard function $h_0(t) = abt^{b-1}$, where $a = 0.1$ is the rate parameter and $b = 1.1$ is the shape parameter. We set $\beta(t) = c(t+2)^d$ for the time varying coefficient for the baseline covariate, where $c = -1.5$ and $d = 0.1$. We set $x_0 = 0$ and

$$g(x) = \frac{3\sin(\pi x/2)}{1 + 2x^2\{2 - 1.5\,\text{sign}(x)\}},$$

where sign(x) =1 if x >0, sign(x) =-1 if x <0, and sign(x) =0 if x =0.

We consider three simulation scenarios as follows:

(1) $X \perp Z$, $X_i \overset{i.i.d.}{\sim} N(0,1)$, $\epsilon_{il} \overset{i.i.d.}{\sim} N(0,\sigma^2)$.

(2) $X \perp Z$, $X_i \overset{i.i.d.}{\sim} N(0,1)$, $\epsilon_{il} \overset{i.i.d.}{\sim}$ scaled t_3 with variance σ^2 .

(3) If $Z_i = 0$, $X_i \overset{i.i.d.}{\sim} N(-1,1)$, else $X_i \overset{i.i.d.}{\sim} N(1,1)$, $\epsilon_{il} \overset{i.i.d.}{\sim} N(0,\sigma^2)$.

Here the covariate Z_i was simulated from a Bernoulli distribution with the probability equals to $p_z = 0.5$. These three scenarios represent three situations where X is subject to normal error and independent of Z, X is subject to non-normal error and independent of Z, and X is subject to normal error and correlated with Z. For the third scenario, to account for the correlation between X and Z in the model, we assume that X has two different normal priors respectively for $Z = 0$ and $Z = 1$ in the Bayesian estimation steps. We simulated the censoring times independently from a uniform distribution and truncated them by the study endpoint. The censoring rate is around 40%. For each of the three scenarios, we consider sample sizes n=1000 and 2000, measurement error variances $\sigma^2 = 0.3^2$ and 0.5^2, and two repeated measures of X_i were simulated for each subject based on (1). A total of 100 replicates were performed for each scenario. We used estimates from the naive method as the initial values in our proposed approach. On average, the proposed Bayesian method converges fast. The modeling performance is evaluated based on five metrics, including the mean bias, the mean standard error (SE), the empirical standard error (Emp. SE), the coverage probabilities of the 95% confidence intervals (95% CR), where the confidence intervals are Bayesian highest density interval (HDI), and the mean squared error (MSE). For each function of $\beta(t)$ and $g(x)$, we computed these metrics on an equally-spaced 100 grid points. Specifically, for the mean bias, we first computed the mean bias at each of the 100 grid points and then took the average of their absolute values; for the mean standard error, the empirical standard error, 95% CR, and the MSE, we first computed the metrics at each of the 100 grid points and then took the average.

Tables 1-3 and Figures 1-3 present the simulation results. The naive approach had a moderate performance when variance and sample size were small and its performance worsened as the error variance increased. On average, the naive estimates of $\beta(t)$ had large bias and the coverage probabilities of the 95% confidence intervals were far from the nominal levels as the error variance increasing. In contrast, the proposed approach generally performed well in all scenarios. The estimated parameters had small bias and the coverage probabilities were either better or comparable to the ideal method. Compared to the naive approach, the MSEs of the proposed approach were usually smaller, but could be slightly larger for estimating $g(x)$ when σ^2 is small since the naive estimates had smaller variances. Figures 1-3 show that the estimated $g(x)$ from the proposed approach are very close to the true curve. The coverage probabilities of $g(x)$ from the ideal approach were similar for $n = 1000$ and 2000 and ranged between 86% and 88%, which were lower than the nominal level. One possible reason is that the variations from the model selections were not taken into account in constructing the confidence intervals. In addition, the placement of the knots did not take into account of the shape of $g(x)$, which makes the coverage probabilities lower in intervals with high curvature in $g(x)$ and low event density. These aspects might also contribute to the decrease of the coverage probabilities of the proposed method when sample size increased from 1000 to 2000. To explore the influence of the function shape, we conducted simulations when $g(x)$ was a function with lower curvature and obtained the ideal estimates. The coverage probabilities were closer to the nominal level (see Section S.2 in the supplementary materials).

5 | REAL DATA APPLICATION ON ACTG

We applied our proposed approach to the AIDS Clinical Trials Group (ACTG) Protocol 175 data.[48] The ACTG 175 is a double-blind randomized controlled trial evaluating different treatment regimens effectiveness in HIV-infected patients. It compared four treatment regimens, zidovudine alone, zidovudine plus didanosine, zidovudine plus zalcitabine and didanosine alone, in HIV-infected participants based on the time to progression to AIDS-defining event or death.[48] A total of 2467 participants were recruited in the study between December 1991 and October 1992. By following the participants until November 1994, researchers recored a total of 308 events. Since the primary study suggested that zidovudine alone is less effective compared to the other three treatment regimens, the aim of further investigations is to compare two treatment groups, which are zidovudine alone and the combination of the other three.

We are interested in investigating the effect of treatment on survival time adjusted for baseline CD4 counts. CD4 counts are commonly used as biomarker in clinical trials to assess HIV-infected patients for treatment eligibility and monitor antiretroviral response to treatment.[49] CD4 measurements is known to be measured with error, mainly due to physiologic biologic variation

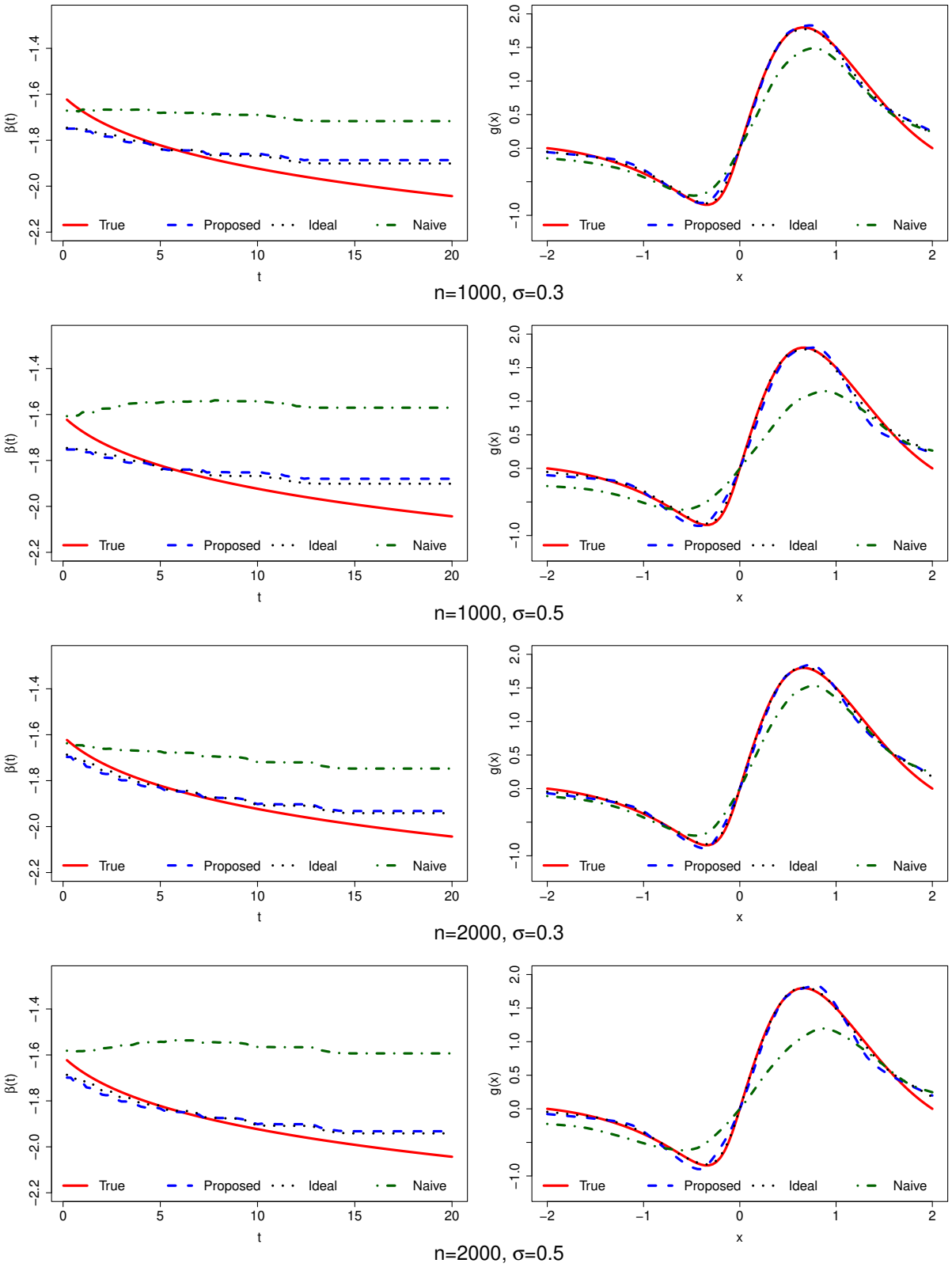

FIGURE 1 Simulation Scenario (1): Estimated $\beta(\cdot)$ left and $g(\cdot)$ right.

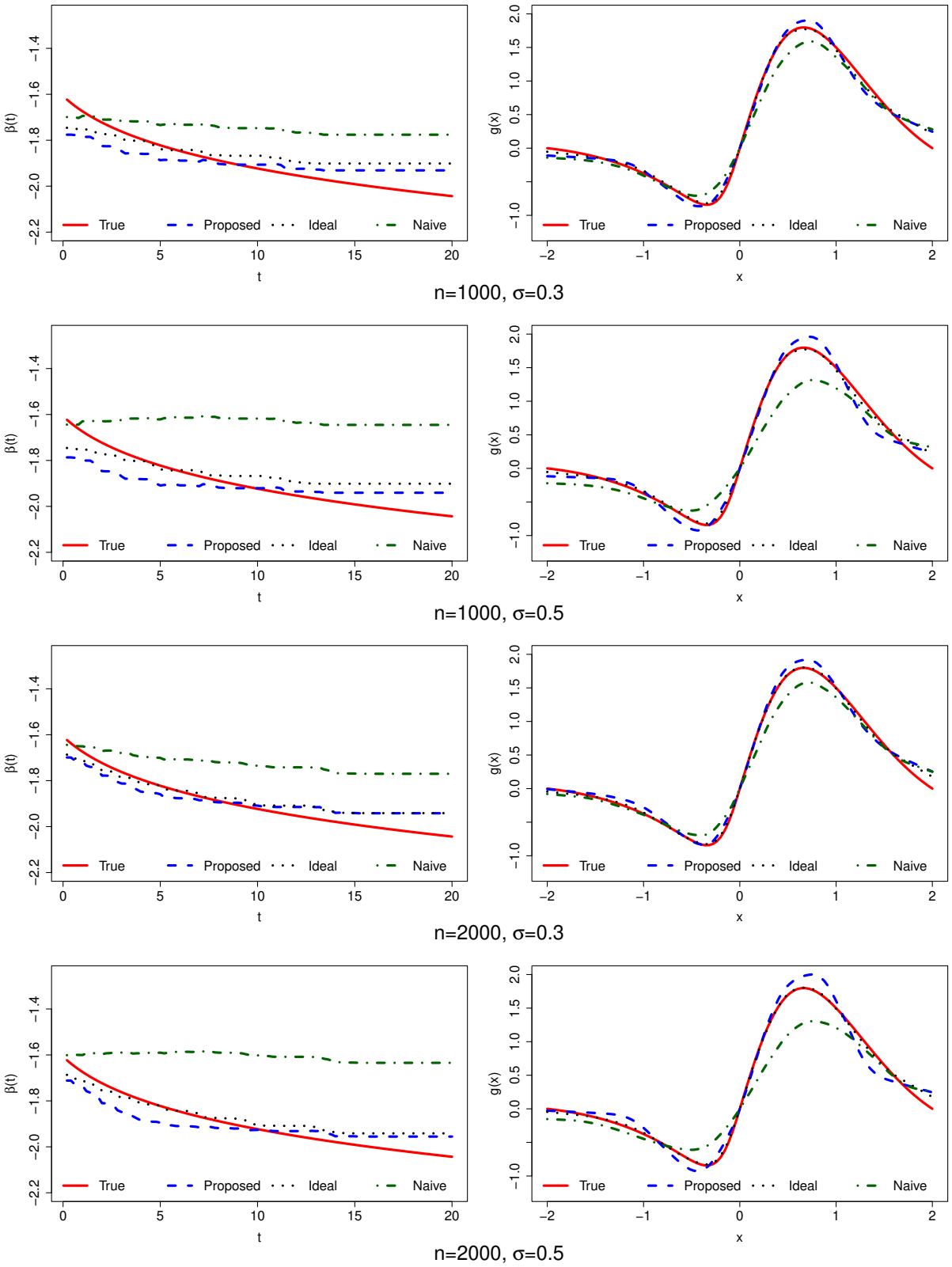

FIGURE 2 Simulation Scenario (2): Estimated $\beta(\cdot)$ left and $g(\cdot)$ right.

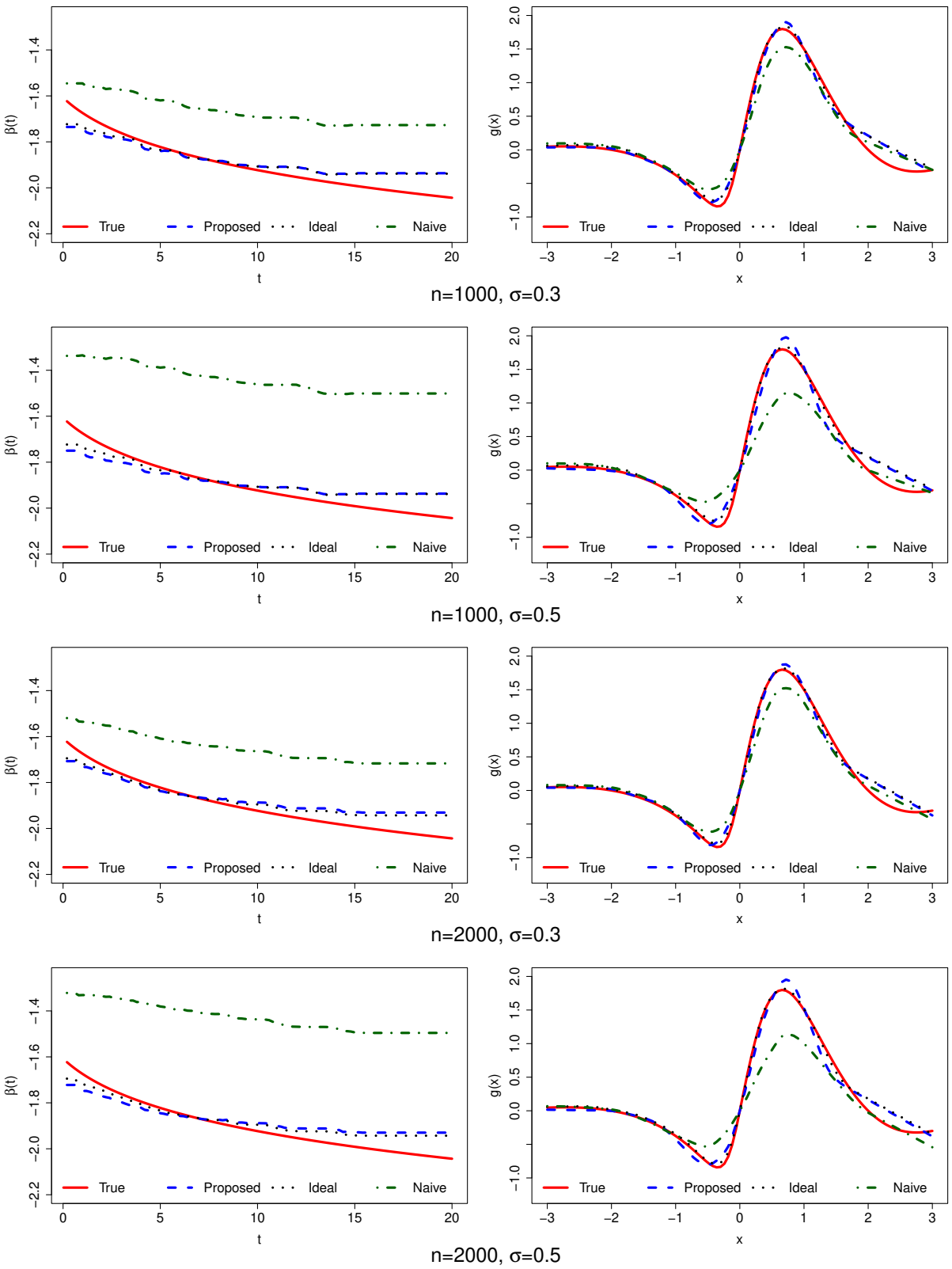

FIGURE 3 Simulation Scenario (3): Estimated $\beta(\cdot)$ left and $g(\cdot)$ right.

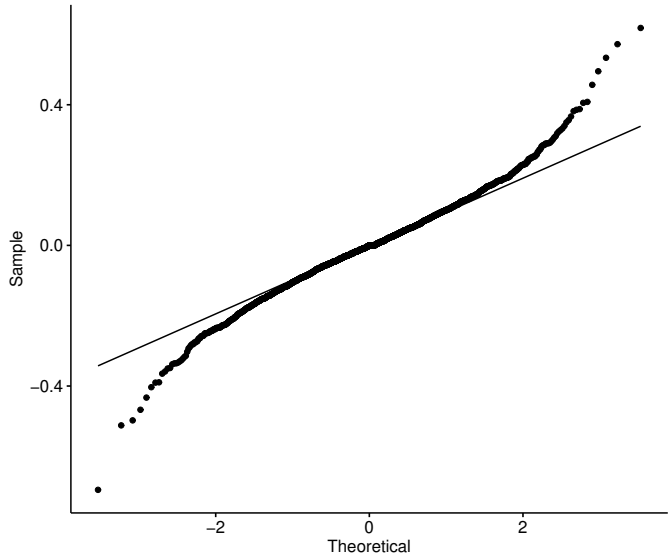

FIGURE 4 ACTG: Normal Q-Q plot for within-person measurement differences.

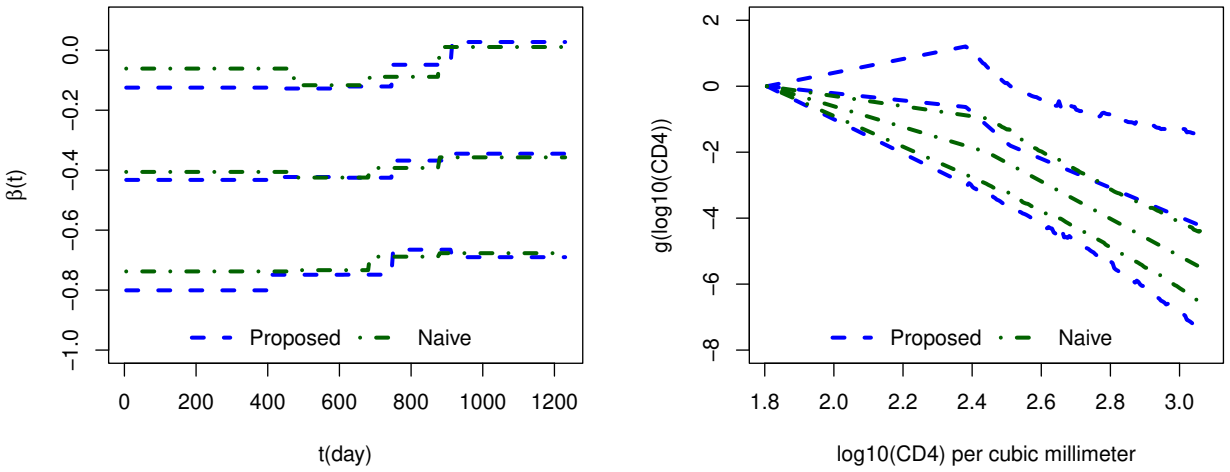

FIGURE 5 ACTG: Estimated $\beta(\cdot)$ and $g(\cdot)$ (center curves) and 95% confidence intervals (outer curves).

and assay performance.[50,51] Therefore, the measured CD4 counts may not reflect the true underlying CD4 counts. Most participants in ACTG 175 study had replicated CD4 measurements before receiving treatment. After excluding 4 CD4 observation outliners, for each participant, we used two repeated measurements between 3 weeks before randomization and 1 week after randomization as replicates for the baseline CD4 measurements, those with only 1 measurement were also included. To achieve approximate constant variance, we applied the logarithmic 10 transformation to the CD4 counts. We have a total of 2463 analytical participants, of which 307 events were observed and 2156 were censored. To investigate the normality assumption on the measurement error, we produced a normal Q-Q plot for within-person CD4 measurement differences (Figure 4). It shows a slight deviation from normality in the tails, which is similar to the t distribution.

We used DIC to select among models with $g^*(\cdot)$ being a B-spline of 1 to 9 internal knots or a linear function to determine whether the true unknown $g^*(\cdot)$ for CD4 is linear or nonlinear, and with both $h^*(\cdot)$ and $\beta^*(\cdot)$ being a piecewise constant function of 1 to 5 internal knots. And X_i is the logarithmic 10 transformed CD4, Z_i is the indicator variable $I(\text{treatment} \neq \text{zidovudine alone})$. We set $x_0 = 1.806$, which is the minimum observed log_{10}CD4.

We compared models according to DIC: a smaller DIC suggests the model has a better fit. The model with 2 internal knots for $g^*(\cdot)$ and 4 internal knots for $h^*(\cdot)$ and $\beta^*(\cdot)$ has the smallest DIC (-4353). Denote this model by $M_{\beta(t)}$. The corresponding estimated functions on the treatment effect and the unobserved true CD4 measures are presented in Figure 5. The estimated measurement error variance is 0.0066 (95% CI: [0.0062, 0.0069]). We also fixed $\beta^*(\cdot)$ as a non-time-varying constant β and used DIC to select among models with $g^*(\cdot)$ being a B-spline of 1 to 9 internal knots or a linear function and with $h^*(\cdot)$ a piecewise constant function of 1 to 5 internal knots. The model with 5 internal knot for $g^*(\cdot)$ and 4 internal knots for $h^*(\cdot)$, denoted by M_β, has the smallest DIC (-4273), which is larger than the DIC of $M_{\beta(t)}$. To further evaluate the evidence of time-varying covariate effect, we also compared the two models $M_{\beta(t)}$ and M_β using the Bayes factor.[52,53] The Bayes factor of $M_{\beta(t)}$ to M_β is 494. This result indicates decisive evidence for the model with time-varying $\beta(\cdot)$.[53]

We found that the use of different treatment regimens is significantly associated with the risk of progression to AIDS-defining events or death after adjusting for CD4 counts. The difference between the two treatment groups is stable at the early stage and then decrease over time. A potential reason can be drug effect eliminating across time. It can be seen from Figure 5 that log hazard decreases when log_{10}CD4 count increases. There are obvious changes of rates when the log_{10}CD4 count is between 2.4 to 2.5 or the CD4 count is between 250 to 315. We also repeat the above analyses using the naive method. The model with 1 internal knot for $g^*(\cdot)$ and 3 internal knots for $h^*(\cdot)$ and $\beta^*(\cdot)$ was selected according to DIC (5866). The difference between the two treatment groups seems to slightly increase at the early stage and then decrease over time. On average, we observe similar results on the estimate of treatment effect compared to the proposed method. A potential reason could be that the measurement error is relatively small in the real data. The naive estimate of $g(\cdot)$ is closer to a straight line. The estimated rate of log hazard decrease with log_{10}CD4 count is larger for low CD4 count (fewer than around 250) than the proposed estimate, and smaller for median range CD4 count (between around 250 to 315), and similar for higher CD4 count (more than around 315).

6 | CONCLUSION

We considered a Cox-type model that depends nonparametrically on an error-contaminated covariate and allows for time-varying covariate effect. The nonlinear effect and time-varying covariate effect are estimated via nonparametric Bayesian approach. The simulation study shows that our proposed method has significantly better modeling performance compared to the naive method under all simulation scenarios. As it is shown in the simulation studies, our proposed method also works well when the measurement error follows a t distribution with degree of freedom equals 3.

Our current work focused on survival models with normally distributed covariates contaminated with normal error. The proposed method can potentially be extended to survival models with non-normally distributed covariates contaminated with non-normal error. Other survival models such as frailty models and recurrent event models and other censoring mechanisms such as interval censoring may also be explored. Another potential future research topic is to extend the proposed method to allow for different measurement error models.

Our proposed method can be applied to a wide range of clinical trials and epidemiology studies where modeling the association between the unobserved covariate and the survival outcome plays an important role.

ACKNOWLEDGEMENTS

This research was partially supported by NSF grant DMS-1916411 (Song and Huang) and NIH grant CA239168 (Song).

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

The ACTG 175 data used in Section 5 are available from request to ACTG (https://actgnetwork.org/).

References

1. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 1972; 34(2): 187–202.

2. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *The annals of statistics* 1982: 1100–1120.

3. Hu P, Tsiatis AA, Davidian M. Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* 1998: 1407–1419.

4. Hu C, Lin D. Cox regression with covariate measurement error. *Scandinavian Journal of Statistics* 2002; 29(4): 637–655.

5. Song X, Wang CY. Proportional hazards model with covariate measurement error and instrumental variables. *Journal of the American Statistical Association* 2014; 109(508): 1636–1646.

6. Tworoger SS, Hankinson SE. Use of biomarkers in epidemiologic studies: minimizing the influence of measurement error in the study design and analysis. *Cancer Causes & Control* 2006; 17(7): 889–899.

7. White E. Measurement error in biomarkers: sources, assessment, and impact on studies.. *IARC scientific publications* 2011(163): 143–161.

8. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement error in nonlinear models: a modern perspective*. CRC press . 2006.

9. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 1982; 69(2): 331–342.

10. Li Y, Lin X. Functional inference in frailty measurement error models for clustered survival data using the SIMEX approach. *Journal of the American Statistical Association* 2003; 98(461): 191–203.

11. Nakamura T. Proportional hazards model with covariates subject to measurement error. *Biometrics* 1992: 829–838.

12. Huang Y, Wang C. Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *Journal of the American Statistical Association* 2000; 95(452): 1209–1219.

13. Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* 2001; 88(2): 447–458.

14. Song X, Huang Y. On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics* 2005; 61(3): 702–714.

15. Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in medicine* 1996; 15(15): 1663–1685.

16. R. Brown E, G. Ibrahim J. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* 2003; 59(2): 221–228.

17. Ibrahim JG, Chen MH, Sinha D. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica* 2004: 863–883.

18. Song X, Wang L, Ma S, Huang H. Variable selection for partially linear proportional hazards model with covariate measurement error. *Journal of nonparametric statistics* 2019; 31(1): 196–220.

19. Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* 2002; 58(4): 742–753.

20. Song X, Davidian M, Tsiatis AA. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics* 2002; 3(4): 511–528.

21. Xu Y, Li Y, Song X. Locally efficient semiparametric estimators for proportional hazards models with measurement error. *Scandinavian Journal of Statistics* 2016; 43(2): 558–572.

22. Sasieni P. Non-orthogonal projections and their application to calculating the information in a partly linear Cox model. *Scandinavian Journal of Statistics* 1992: 215–233.

23. Huang J, others . Efficient estimation of the partly linear additive Cox model. *Annals of Statistics* 1999; 27(5): 1536–1563.

24. Heller G. The Cox proportional hazards model with a partly linear relative risk function. *Lifetime data analysis* 2001; 7(3): 255–277.

25. Wu Q, Zhao H, Zhu L, Sun J. Variable selection for high-dimensional partly linear additive Cox model with application to Alzheimer's disease. *Statistics in Medicine* 2020; 39(23): 3120–3134.

26. Cheng YJ, Crainiceanu CM. Cox models with smooth functional effect of covariates measured with error. *Journal of the American Statistical Association* 2009; 104(487): 1144–1154.

27. Kneib T, Brezger A, Crainiceanu CM. Generalized semiparametric regression with covariates measured with error. In: Springer. 2010 (pp. 133–154).

28. Zucker DM, Karr AF. Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *The Annals of Statistics* 1990; 18(1): 329–353.

29. Murphy SA. Testing for a time dependent coefficient in Cox's regression model. *Scandinavian journal of Statistics* 1993: 35–50.

30. Lehr S, Schemper M. Parsimonious analysis of time-dependent effects in the Cox model. *Statistics in medicine* 2007; 26(13): 2686–2698.

31. Martinussen T, Scheike TH, Skovgaard IM. Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics* 2002; 29(1): 57–74.

32. Cai Z, Sun Y. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics* 2003; 30(1): 93–111.

33. Bhadra A, Carroll RJ. Exact sampling of the unobserved covariates in Bayesian spline models for measurement error problems. *Statistics and computing* 2016; 26(4): 827–840.

34. Lang S, Brezger A. Bayesian P-splines. *Journal of computational and graphical statistics* 2004; 13(1): 183–212.

35. Brezger A, Lang S. Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics & Data Analysis* 2006; 50(4): 967–991.

36. Wang X, Yue Y, Faraway JJ. *Bayesian regression modeling with INLA*. Chapman and Hall/CRC . 2018.

37. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian data analysis*. CRC press . 2013.

38. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS user manual. *Cambridge: MRC Biostatistics Unit* 2003.

39. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis* 2006; 1(3): 515–534.

40. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. The BUGS book. *A Practical Introduction to Bayesian Analysis, Chapman Hall, London* 2013.

41. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 1992; 41(2): 337–348.

42. Leydold J, Wolfgang H, Leydold MJ. Package 'Runuran'. 2020.

43. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)* 2002; 64(4): 583–639.

44. Sturtz S, Ligges U, Gelman A. R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software* 2005; 12(3): 1–16.

45. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis Chapman & Hall. *CRC Texts in Statistical Science* 2004.

46. Spiegelhalter DJ, Best NG, Carlin BP, Linde V. dA. The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B: Statistical Methodology* 2014: 485–493.

47. Celeux G, Forbes F, Robert CP, Titterington DM, others . Deviance information criteria for missing data models. *Bayesian analysis* 2006; 1(4): 651–673.

48. Hammer SM, Katzenstein DA, Hughes MD, et al. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine* 1996; 335(15): 1081–1090.

49. Daneau G, Buyze J, Wade D, et al. CD4 results with a bias larger than hundred cells per microliter can have a significant impact on the clinical decision during treatment initiation of HIV patients. *Cytometry Part B: Clinical Cytometry* 2017; 92(6): 476–484.

50. Raboud J, Haley L, Montaner J, Murphy C, Januszewska M, Schechter M. Quantification of the variation due to laboratory and physiologic sources in CD4 lymphocyte counts of clinically stable HIV-infected individuals.. *Journal of acquired immune deficiency syndromes and human retrovirology: official publication of the International Retrovirology Association* 1995; 10: S67–73.

51. Raboud J, Montaner J, Conway B, et al. Variation in plasma RNA levels, CD4 cell counts, and p24 antigen levels in clinically stable men with human immunodeficiency virus infection. *Journal of Infectious Diseases* 1996; 174(1): 191–194.

52. Kass RE, Raftery AE. Bayes factors. *Journal of the american statistical association* 1995; 90(430): 773–795.

53. Jeffreys H. *The theory of probability*. OUP Oxford . 1998.

How to cite this article: Pan A., X. Song, and H. Huang (202X), Bayesian Analysis For Partly Linear Cox Model with Measurement Error and Time Varying Covariate Effect, *Statistics in Medicine*, *202X; XX:X–X*.