Annals of the Institute of Statistical Mathematics (2021) 73:481–517
https://doi.org/10.1007/s10463-020-00755-2

Semiparametric methods for left-truncated and right-censored survival data with covariate measurement error

Li-Pang Chen[1] **· Grace Y. Yi**[1,2]

Received: 9 October 2019 / Revised: 25 February 2020 / Published online: 2 June 2020

Abstract

Many methods have been developed for analyzing survival data which are commonly right-censored. These methods, however, are challenged by complex features pertinent to the data collection as well as the nature of data themselves. Typically, biased samples caused by left-truncation (or length-biased sampling) and measurement error often accompany survival analysis. While such data frequently arise in practice, little work has been available to simultaneously address these features. In this paper, we explore valid inference methods for handling left-truncated and right-censored survival data with measurement error under the widely used Cox model. We first exploit a flexible estimator for the survival model parameters which does not require specification of the baseline hazard function. To improve the efficiency, we further develop an augmented nonparametric maximum likelihood estimator. We establish asymptotic results and examine the efficiency and robustness issues for the proposed estimators. The proposed methods enjoy appealing features that the distributions of the covariates and of the truncation times are left unspecified. Numerical studies are reported to assess the finite sample performance of the proposed methods.

Keywords Cox model · Efficiency · Left-truncation · Measurement error · Right censoring

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-020-00755-2) contains supplementary material, which is available to authorized users.

✉ Grace Y. Yi
gyi5@uwo.ca

Li-Pang Chen
lchen723@uwo.ca

1 Department of Statistical and Actuarial Sciences, University of Western Ontario, 1151 Richmond St, London, ON N6A 3K7, Canada

2 Department of Computer Science, University of Western Ontario, 1151 Richmond St, London, ON N6A 3K7, Canada

1 Introduction

Survival analysis has been proven useful in many areas including cancer research, clinical trials, epidemiological studies, actuarial science, and so on. A large body of methods have been developed under various survival models. Among them, methods on the Cox proportional hazards model have attracted the most research attention. Comprehensive discussion on those methods can be found in Kalbfleisch and Prentice (2002), Lawless (2003), and the references therein.

Those methods, however, break down in the presence of complex features pertinent to the data collection and the nature of variables. Left-truncation is a common characteristic arising from survival studies when subjects do not enter the study at the same time (e.g., Kalbfleisch and Prentice 2002, Sect. 1.3; Lawless 2003, Sect. 2.4). In the presence of left-truncation, individuals with shorter survival times are less likely to be recruited for the study, thus resulting in a biased sample. Many methods have been available for analyzing such data. For instance, Qin and Shen (2010) proposed the weighted estimating equation approach. Qin et al. (2011) described an EM algorithm for estimation involving infinite dimensional parameters. Huang et al. (2012) examined a profile likelihood method for parameter estimation for which the distribution of the truncation time was restricted as a uniform distribution. Wu et al. (2018) proposed a pairwise likelihood method for handling left-truncated data. With joint modeling of longitudinal covariates and survival outcomes, Su and Wang (2012) proposed a semiparametric method to handle the feature of left-truncation, where the linear mixed effects model is employed with the latent variable assumed to follow a normal distribution.

On the other hand, measurement error in covariates is ubiquitous, and a large number of research papers have emerged for handling error-contaminated survival data since the seminal work of Prentice (1982). To name a few, Nakamura (1992) developed an approximate corrected partial likelihood method which was extended by Kong and Gu (1999), Buzas (1998), and Hu and Lin (2002). Huang and Wang (2000) proposed a nonparametric approach for settings with repeated measurements for error-prone covariates. Xie et al. (2001) explored a least squares method to calibrate the induced hazard function. Song and Huang (2005) presented a conditional score approach for estimation of the model parameters. Other approaches include Augustin (2004), Greene and Cai (2004), Li and Ryan (2006), Küchenoff et al. (2007), and the references therein. A review on this topic was given by Yi (2017, Chapter 3).

While there have been methods for dealing with left-truncated survival data or error-contaminated survival data, methods of handling those features simultaneously are rather limited. To the best of our knowledge, Yi and Lawless (2007, Sect. 4.1) is the only work which directly touches on this topic, but their discussion is only a sketchy outline of ideas without rigorous or systematic development. Simultaneous presence of biased samples caused by left-truncation (or length-biased sampling) and measurement error in covariates poses considerable challenges in survival analysis. In this paper, we systematically investigate this important problem and develop valid inference methods for analysis of left-truncated and right-censored (LTRC) survival data with measurement error. To delineate the survival process, we employ the most widely used framework—the Cox proportional hazards model; to postulate the measurement error process, we extend the classical additive model, the most popularly considered in the literature of measurement error models, to facilitate measurement error that is induced from both a systematic way and a random manner. We exploit a flexible estimator for the survival model parameters which does not require the specification of the baseline hazard function. To improve the efficiency, we further develop an augmented nonparametric maximum likelihood estimator.

While the proposed methods generalize the scope of existing work on survival data, the extensions turn out neither trivial nor straightforward. The establishment of solid theoretical results is effort taking. In this paper, we provide asymptotic results for the proposed estimators including the consistency, asymptotic distributions, and the efficiency comparison. The proposed methods enjoy appealing features that the distributions of the true covariates and of the truncation times are left unspecified.

Our work is partially motivated by the Worcester Heart Attack Study (WHAS500) data (Hosmer et al. 2008) which involve both left-truncation and right-censoring. Three types of time are recorded: time of the hospital admission, time of the hospital discharge, and time of the last follow-up (which is either death or censoring time). The total follow-up length is defined as the time gap between the hospital admission and the last follow-up, and the hospital stay time is defined as the time length between the hospital admission and the hospital discharge. Data can only be collected for those individuals whose total follow-up length is larger than the hospital stay time. It is interesting to study how the risk factors are associated with the survival times after the patients are discharged from the hospital. To conduct sensible analyses, it is imperative to account for possible measurement error effects that are induced from error-prone covariates.

The remainder is organized as follows. In Sect. 2, we introduce the basic notation and the framework. In Sect. 3, we present the conditional likelihood method and provide the asymptotic properties. In Sect. 4, we develop an augmented estimation method to improve the efficiency of the estimator in Sect. 3. The estimators in Sects. 3 and 4 are developed under the assumption that the parameters associated with the measurement error model are known. When this assumption is untrue, in Sect. 5, we develop a two-stage procedure which includes estimation of the parameters for the measurement error model as well as estimation of the parameters for the survival model, and we establish the asymptotic results accordingly. Simulation results and a real data analysis are provided in Sect. 6. We conclude the paper with discussions in the last section.

2 Notation and model

For an individual in the target disease population, let ξ be the calendar time of the recruitment (e.g., the recruitment starts right at the hospital discharge) and let u and r denote the calendar time of the initiating event (e.g., hospital admission) and the failure event (e.g., death), respectively, where $u < r$, and $u < \xi < r$. Let $T^* = r - u$ be the lifetime (e.g., the time length between the hospital admission and the failure), and let $A^* = \xi - u$ be the truncation time (e.g., the time length between the hospital admission and the hospital discharge).

If $T^* < A^*$, this individual is not included in the study to contribute any information. We are only able to recruit an individual when $T^* \geq A^*$; in this case, we let $\{A, T\}$ replace $\{A^*, T^*\}$ to emphasize such an individual is eligible for the recruitment, consistent with the notation considered by Wu et al. (2018).

We define C as the censoring time for a recruited subject. Let $Y = \min\{T, A + C\}$ be the observed time and let $\Delta = I(T \leq A + C)$ be the indicator of a failure event. Figure 1 gives an illustration of the relationship among those variables. For an individual in the study, let X and Z be the associated covariates of dimensions $p \times 1$ and $q \times 1$, respectively, and write $V = (X^\top, Z^\top)^\top$. Let $h(a)$ be the probability density function of A^* which is unknown, and let $H(a) = \int_0^a h(u)\mathrm{d}u$ be the corresponding distribution function. Let $f(t/v)$ and $S(t/v)$ be the density function and the survivor function of the lifetime T^*, given $V = v$, respectively.

2.1 Cox model and inference

Suppose that we have a sample of n subjects where for $i = 1, \ldots, n$, $(Y_i, A_i, \Delta_i, V_i)$ has the same distribution as (Y, A, Δ, V), and $(y_i, a_i, \delta_i, v_i)$ represents realizations of $(Y_i, A_i, \Delta_i, V_i)$. Consider the Cox model for survival times T^* with the hazard function

$$\lambda(t|v_i) = \lambda_0(t) \exp(v_i^\top \beta), \tag{1}$$

where $\lambda_0(\cdot)$ is the unknown baseline hazards function, and β is the vector of parameters of primary interest.

Let

$$L_C = \prod_{i=1}^{n} \frac{f(y_i|v_i)^{\delta_i} S(y_i|v_i)^{1-\delta_i}}{S(a_i|v_i)} \tag{2}$$

be the conditional likelihood of Y_i, given $V_i = v_i$ and $A_i = a_i$, and let

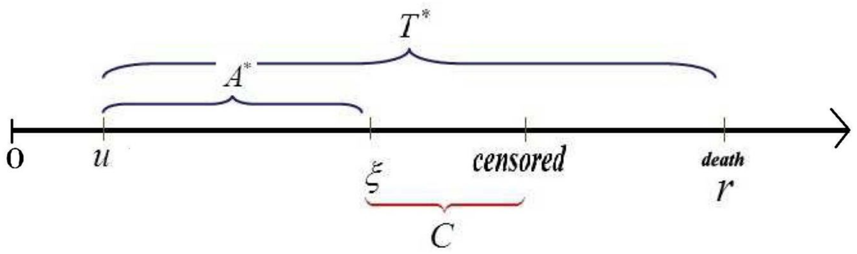

Fig. 1 Schematic depiction of LTRC data for $T^* \geq A^*$

$$L_M = \prod_{i=1}^{n} \frac{S(a_i|v_i)\mathrm{d}H(a_i)}{\int_0^\infty S(\alpha|v_i)\mathrm{d}H(\alpha)} \tag{3}$$

be the marginal likelihood of A_i, given $V_i = v_i$, where $S(t|v_i) = \exp\left\{-\Lambda_0(t) \times \exp\left(v_i^\top \beta\right)\right\}$, and $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ is the cumulative baseline hazards function.

Inference about β is then carried out by maximizing the likelihood function

$$L \propto L_C \times L_M = \prod_{i=1}^{n} \frac{f(y_i|v_i)^{\delta_i} S(y_i|v_i)^{1-\delta_i}\mathrm{d}H(a_i)}{\int_0^\infty S(u|v_i)\mathrm{d}H(u)} \tag{4}$$

with respect to the model parameters.

2.2 Measurement error model

In applications, covariates are often subject to measurement error. For $i = 1, \ldots, n$, suppose that Z_i is precisely observed and that X_i is subject to measurement error; let W_i denote an observed value or surrogate measurement of X_i.

We first consider the widely used classical additive measurement error model (e.g., Carroll et al. 2006; Yi 2017):

$$W_i = X_i + \epsilon_i, \tag{5}$$

where ϵ_i is independent of $\left\{X_i, Z_i, C_i, A_i, T_i\right\}$, and $\epsilon_i \sim N(0, \Sigma_\epsilon)$ with covariance matrix Σ_ϵ. Thus, the moment generation function of ϵ_i is given by $m(t) = \exp(\frac{1}{2}t^\top \Sigma_\epsilon t)$, and

$$E\left\{\exp\left(t^\top W_i\right)\right\} = m(t)\exp\left(t^\top X_i\right).$$

Starting with measurement error model (5), we next form a more flexible model using the idea of *exponential tilting* (e.g., Goutis and Casella 1999, Sect. 3.2.1). To be specific, let $f(w_i|x_i;\Sigma_\epsilon)$ denote the conditional density of W_i, given $X_i = x_i$, for model (5). Then we "tilt" model (5) by multiplying $f(w_i|x_i;\Sigma_\epsilon)$ by an exponential term $\exp\left(\alpha^\top w_i\right)$, where α is a vector of parameters. To make the resulting function $\exp\left(\alpha^\top w_i\right) \cdot f(w_i|x_i;\Sigma_\epsilon)$ be a legitimate density function, we need to attach a normalizing constant, which is given by $\left[E\left\{\exp\left(\alpha^\top W_i\right)\right\}\right]^{-1}$ with the expectation taken with respect to (5). Therefore, the "exponential tilting" model derived from (5) is given by

$$N(\Sigma_\epsilon \alpha + X_i, \Sigma_\epsilon)$$

with a p-dimensional vector α, suggesting that W_i and X_i are characterized by

$$W_i = \Sigma_\epsilon \alpha + X_i + \epsilon_i, \tag{6}$$

where ϵ_i is characterized as in (5).

Model (6) allows us to consider a broader class of settings than (5) does and it also embraces (5) as a special case by setting $\alpha = 0$. Model (6) describes a situation where W_i and X_i are different not only by a *random* amount ϵ_i but also systematically by a *fixed* amount indicated by $\Sigma_\epsilon \alpha$. While it is possible to express the amount $\Sigma_\epsilon \alpha$ of the systematic error by a single parameter vector, say γ, we retain the use of $\Sigma_\epsilon \alpha$ in model (6) because a value of α can directly reveal the degree of the systematic error relative to the degree of the random error (i.e., Σ_ϵ), but γ cannot.

In the following sections, we develop estimation methods with the measurement error model given by (6). Let $\Sigma_{\epsilon;0}$ and α_0 denote the true values of Σ_ϵ and α, respectively. To highlight the idea, in Sects. 3 and 4 we assume that the true values $\Sigma_{\epsilon;0}$ and α_0 for the parameters in (6) are known; discussion on handling unknown $\Sigma_{\epsilon;0}$ and α_0 is provided in Sect. 5. Let $W_i^* = W_i - \Sigma_\epsilon \alpha$. Then

$$E(W_i^* | X_i) = E(W_i - \Sigma_\epsilon \alpha | X_i) = X_i \tag{7}$$

and

$$E\left\{ \exp\left(t^\top W_i^* - \frac{1}{2} t^\top \Sigma_\epsilon t \right) \middle| X_i \right\} = \exp\left(t^\top X_i \right). \tag{8}$$

3 Conditional profile-likelihood method

3.1 Estimation method

We begin with a simple perspective by examining the conditional likelihood L_C, determined by (2), which allows us to ignore modeling of the truncation times. Let $\ell_C = \log L_C$. Since ℓ_C contains the X_i whose measurements are unavailable, we want to modify ℓ_C to be a new function, say ℓ_C^*, of the observed measurements and the model parameters so that its conditional expectation equals to ℓ_C:

$$E(\ell_C^* | \mathbb{X}, \mathbb{Z}, \mathbb{C}, \mathbb{A}, \mathbb{T}) = \ell_C, \tag{9}$$

where the expectation is taken with respective to the conditional distribution of $\mathbb{W}$ given $\{\mathbb{X}, \mathbb{Z}, \mathbb{C}, \mathbb{A}, \mathbb{T}\}$, with $\mathbb{W} = \{W_1, \ldots, W_n\}$, $\mathbb{X} = \{X_1, \ldots, X_n\}$, $\mathbb{Z} = \{Z_1, \ldots, Z_n\}$, $\mathbb{C} = \{C_1, \ldots, C_n\}$, $\mathbb{A} = \{A_1, \ldots, A_n\}$, and $\mathbb{T} = \{T_1, \ldots, T_n\}$. Such a strategy is useful in yielding an unbiased estimating function and is sometimes called the *"corrected" likelihood method* (e.g., Nakamura 1992; Yi and Lawless 2007) or the *insertion correction approach* (e.g., Yi 2017, Chapter 2).

Noticing that the X_i appear in ℓ_C in linear and exponential forms, we define

$$\begin{aligned} \ell_C^* = \sum_{i=1}^n \Big[& \delta_i \log \lambda_0(y_i) + \delta_i (w_i^{*\top} \beta_x + z_i^\top \beta_z) \\ & - \left\{ \Lambda_0(y_i) - \Lambda_0(a_i) \right\} \exp(w_i^{*\top} \beta_x + z_i^\top \beta_z) \left\{ m(\beta_x) \right\}^{-1} \Big], \end{aligned} \tag{10}$$

where w_i^* and z_i represent realizations of W_i^* and Z_i, respectively. It is easily seen that ℓ_C^* satisfies (9).

To use (10) to derive an estimator of (β_x, β_z), we need to deal with the baseline hazard function $\lambda_0(\cdot)$ and its cumulative function $\Lambda_0(\cdot)$. Different from the piecewise constant method considered by Yi and Lawless (2007), we discretize $\Lambda_0(\cdot)$ so that $\lambda_0(\cdot)$ has a nonzero value if $t = y_i$ for $i = 1, \ldots, n$; otherwise, $\lambda_0(t) = 0$. Let λ_i denote $\lambda_0(y_i)$ for $i = 1, \ldots, n$. Then, $\Lambda_0(t)$ is taken as $\sum_{i=1}^{n} I(y_i \leqslant t)\lambda_i$. Given β_x and β_z, we solve $\frac{\partial \ell_C^*}{\partial \lambda_i} = 0$ for $i = 1, \ldots, n$, which leads to an estimator of λ_i, given by

$$\widehat{\lambda}_i = \frac{\delta_i}{\sum_{k=1}^{n} I(a_k \le y_i \le y_k) \exp\left(w_k^{*\top}\beta_x + z_k^\top \beta_z\right)\left\{m(\beta_x)\right\}^{-1}} \quad \text{for } i = 1, \ldots, n; \tag{11}$$

and the corresponding estimate of the cumulative baseline hazard function:

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^{n} I(y_i \le t)\widehat{\lambda}_i \,. \tag{12}$$

Plugging (11) and (12) into (10) gives the function

$$\begin{aligned}\widehat{\ell}_C^* = \sum_{i=1}^{n} \Big[&\delta_i \log \widehat{\lambda}_i + \delta_i (w_i^{*\top}\beta_x + z_i^\top \beta_z) \\ &- \left\{ \widehat{\Lambda}_0(y_i) - \widehat{\Lambda}_0(a_i) \right\} \exp(w_i^{*\top}\beta_x + z_i^\top \beta_z)\left\{m(\beta_x)\right\}^{-1} \Big].\end{aligned} \tag{13}$$

An estimator of β, called the conditional estimator of β, is then obtained by maximizing $\widehat{\ell}_C^*$:

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmax}}\ \widehat{\ell}_C^*. \tag{14}$$

3.2 Asymptotic results

Let $\beta_0 = (\beta_{x0}^\top, \beta_{z0}^\top)^\top$ denote the true value of β and let Θ denote the parameter space of β. Consistent with others such as Huang et al. (2012), we assume that T_i^* has a finite maximal support τ, where $\tau = \sup\left\{t : P(T_i^* \le t) < 1\right\} < \infty$, implying that τ is also a maximal support of truncation time. Let $N_i(t) = \Delta_i I(Y_i \le t)$ be the counting process of the observed failure events for subject i. Let $\widetilde{V}_i = \left(W_i^{*\top}, Z_i^\top\right)^\top$. Define $S^{(k)}(u, \beta) = n^{-1} \sum_{i=1}^{n} \widetilde{v}_i^{\otimes k} \exp(\widetilde{v}_i^\top \beta) I(a_i \le u \le y_i)$ for $k = 0, 1, 2$, where $a^{\otimes 2}$ means $aa^\top$ for the column vector a . Let $\mathcal{S}^{(k)}(u, \beta) = E\left[\widetilde{V}_i^{\otimes k} \exp(\widetilde{V}_i^\top \beta) I(A_i \le u \le Y_i)\right]$ be the expectation of $S^{(k)}(u, \beta)$. Using these symbols, we express (12) as

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^{n} \mathrm{d}N_i(u)}{\sum_{i=1}^{n} \exp(\widetilde{v}_i^\top \beta) I(a_i \le u \le y_i)\left\{m(\beta_x)\right\}^{-1}}. \tag{15}$$

The following theorems, whose proofs are included in the Supplementary Material, establish the asymptotic properties of $\widehat{\Lambda}_0(t)$ and $\widehat{\beta}$.

Theorem 1 *Under regularity conditions in* "Appendix", *we have that as* $n \to \infty$,

$$\sup_{\beta\in\Theta, t\in[0,\tau]} |\widehat{\Lambda}_0(t) - \Lambda_0(t)| \xrightarrow{a.s.} 0,$$

where $\Lambda_0(t) = \int_0^t \left\{\mathcal{S}^{(0)}(u,\beta_0)\right\}^{-1} m(\beta_{x0}) dP(\Delta_i = 1, Y_i \le u)$.

Let

$$\begin{aligned}\Phi\left(w_i, z_i, y_i, a_i\right) = &\int_0^\tau \left\{\widetilde{v}_i - \frac{\mathcal{S}^{(1)}(u,\beta_0)}{\mathcal{S}^{(0)}(u,\beta_0)} + \begin{pmatrix} \Sigma_{\epsilon;0}\beta_{x0} \\ 0_q \end{pmatrix}\right\} \mathrm{d}N_i(u) \\ &- \int_0^\tau \frac{\exp\left(\widetilde{v}_i^\top \beta_0\right) I(a_i \le u \le y_i)}{\mathcal{S}^{(0)}(u,\beta_0)} \left(\widetilde{v}_i - \frac{\mathcal{S}^{(1)}(u,\beta_0)}{\mathcal{S}^{(0)}(u,\beta_0)}\right) \mathrm{d}E\left\{N_i(u)\right\}.\end{aligned}$$

Define

$$\mathcal{J} = E\left\{\Phi^{\otimes 2}\left(W_i, Z_i, Y_i, A_i\right)\right\}$$

and

$$\mathcal{I} = \int_0^\tau \left[\left\{\frac{\mathcal{S}^{(2)}(u,\beta_0)}{\mathcal{S}^{(0)}(u,\beta_0)} - \left(\frac{\mathcal{S}^{(1)}(u,\beta_0)}{\mathcal{S}^{(0)}(u,\beta_0)}\right)^{\otimes 2}\right\} - \begin{pmatrix} \Sigma_{\epsilon;0} & 0_{p\times q} \\ 0_{q\times p} & 0_{q\times q} \end{pmatrix}\right] \mathrm{d}E\left\{N_i(u)\right\}, \quad (16)$$

where $0_{p\times q}$ represents a $p \times q$ matrix with all entries 0, and 0_p stands for a $p \times 1$ vector with all entries 0.

Theorem 2 *Under regularity conditions given in* "Appendix", *the estimator* $\widehat{\beta}$ *obtained from* (14) *has the following asymptotic properties:*

(1) $\widehat{\beta} \xrightarrow{p} \beta_0$ as $n \to \infty$;
(2) $\sqrt{n}\left(\widehat{\beta} - \beta_0\right) \xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1})$ as $n \to \infty$.

4 Augmented pseudo-likelihood method

Estimator $\widehat{\beta}$ is easily formulated from (14), which basically hinges on the availability of ℓ_C^* satisfying (9). However, $\widehat{\beta}$ can be inefficient since it uses only the conditional likelihood L_C with the marginal likelihood L_M ignored, as shown by the likelihood (4) formulated in Sect. 2.1. To improve the efficiency of $\widehat{\beta}$, now we develop an augmented estimator. The basic idea, driven by the form of the likelihood (4), is to include the marginal likelihood L_M for the truncation times in the estimation procedure. In addition to containing the distribution function $H(\cdot)$ of A^*, the marginal likelihood L_M in (3) involves the unobserved covariate X_i. Due to the complexity of L_M, it is difficult to directly apply the insertion correction approach to construct an workable function like ℓ_C^*. To get around this difficulty, we employ the *regression calibration* strategy (Prentice 1982; Yi 2017, p.105) to handle the unobserved true covariate X_i in L_M, as elaborated in the following subsection.

4.1 Estimation method

Let μ_X and Σ_X be the mean vector and variance–covariance matrix of X_i, respectively. Let $W_i^* = W_i - \Sigma_\epsilon \alpha$ as in (7), then model (6) gives that $W_i^* = X_i + \epsilon_i$ with $\epsilon_i \sim N\left(0, \Sigma_\epsilon\right)$, yielding that

$$E(X_i|W_i^* = w_i^*) = \mu_X + \left(\Sigma_{W^*} - \Sigma_\epsilon\right)^\top \Sigma_{W^*}^{-1}(w_i^* - \mu_{W^*}), \tag{17}$$

where μ_{W^*} and Σ_{W^*} represent the mean and covariance matrix of W_i^*, respectively; we let $\widetilde{x}_i$ denote (17) for ease of notation. Using the method of moments, (17) is estimated by

$$\widehat{x}_i = \widehat{\mu}_{W^*} + \left(\widehat{\Sigma}_{W^*} - \Sigma_\epsilon\right)^\top \widehat{\Sigma}_{W^*}^{-1}(w_i^* - \widehat{\mu}_{W^*}) \tag{18}$$

with $\widehat{\mu}_{W^*} = \frac{1}{n}\sum_{i=1}^n w_i^*$ and $\widehat{\Sigma}_{W^*} = \frac{1}{n-1}\sum_{i=1}^n (w_i^* - \widehat{\mu}_{W^*})(w_i^* - \widehat{\mu}_{W^*})^\top$.

As a result, replacing $v_i = \left(x_i^\top, z_i^\top\right)^\top$ with $\left(\widetilde{x}_i^\top, z_i^\top\right)^\top$ in the likelihood function (3) gives

$$L_M^* = \prod_{i=1}^n \frac{S(a_i|\widetilde{x}_i, z_i)\mathrm{d}H(a_i)}{\int_0^\infty S(u|\widetilde{x}_i, z_i)\mathrm{d}H(u)}, \tag{19}$$

where $S(a_i|\widetilde{x}_i, z_i) = \exp\left\{-\Lambda_0(a_i)\exp\left(\widetilde{x}_i^\top \beta_x + z_i^\top \beta_z\right)\right\}$.

To use (19) for inference about β, we next estimate the distribution function $H(\cdot)$. Directly applying the kernel estimation (Silverman 1978) to the observed truncation times to estimate $\mathrm{d}H(\cdot)$ is not suitable since the observed truncation times form a biased sample. Instead, we use the nonparametric maximum likelihood estimator (NPMLE) (e.g., Wang 1991) to estimate the distribution function of A^*. For a fixed parameter β, the NPMLE of $H(a)$ in (19) is given by

$$\widehat{H}(a) = \left(\sum_{i=1}^n \frac{1}{\widehat{S}(a_i|\widehat{x}_i, z_i)}\right)^{-1} \sum_{i=1}^n \frac{I(a_i \le a)}{\widehat{S}(a_i|\widehat{x}_i, z_i)}, \tag{20}$$

where $\widehat{S}(a_i|\widehat{x}_i, z_i) = \exp\left\{-\widehat{\Lambda}_0(a_i)\exp\left(\widehat{x}_i^\top \widehat{\beta}_x + z_i^\top \widehat{\beta}_z\right)\right\}$, and $\widehat{\Lambda}_0(\cdot)$ and $\widehat{\beta}$ are consistent estimators of $\Lambda_0(\cdot)$ and β, respectively, proposed in Sect. 3.

Then replacing $H(a)$ by $\widehat{H}(a)$ in (19) gives $\widehat{L}_M^*$; let $\widehat{\ell}_M^* = \log \widehat{L}_M^*$, which is given by

$$\begin{aligned}\widehat{\ell}_M^* = &\sum_{i=1}^n \log\left\{\mathrm{d}\widehat{H}(a_i)\right\} - \sum_{i=1}^n \widehat{\Lambda}_0(a_i)\exp\left(\widehat{x}_i^\top \beta_x + z_i^\top \beta_z\right) \\ &- \sum_{i=1}^n \log\left[\int_0^\infty \exp\left\{-\widehat{\Lambda}_0(\alpha)\exp\left(\widehat{x}_i^\top \beta_x + z_i^\top \beta_z\right)\right\}\mathrm{d}\widehat{H}(\alpha)\right].\end{aligned} \tag{21}$$

Finally, we consider the pseudo-likelihood function

$$\widehat{\ell}^* = \widehat{\ell}_C^* + \widehat{\ell}_M^*; \tag{22}$$

maximizing $\widehat{\ell}^*$ with respect to β gives an estimator of β:

$$\widetilde{\beta} = \underset{\beta}{\operatorname{argmax}}\,(\widehat{\ell}_C^* + \widehat{\ell}_M^*), \tag{23}$$

which is called a pseudo-likelihood estimator of β.

4.2 Asymptotic results

For ease of notation, we let $\frac{\partial f(\beta_0)}{\partial \beta_0}$ and $\frac{\partial^2 f(\beta_0)}{\partial \beta_0 \partial \beta_0^\top}$, (or $\frac{\partial f}{\partial \beta_0}$ and $\frac{\partial^2 f}{\partial \beta_0 \partial \beta_0^\top}$), respectively, denote the first- and second-order partial derivatives of a function $f(\beta)$ with respect to β which are evaluated at β_0. Let

$$\mu\left(\widetilde{x}_i, z_i\right) = \int_0^\tau \exp\left\{-\Lambda_0(u)\exp(\widetilde{x}_i^\top \beta_{x0} + z_i^\top \beta_{z0})\right\}\mathrm{d}H(u). \tag{24}$$

Let $\mathcal{N}(t) = P(\Delta_i = 1, Y_i \leq t)$, $S(\xi|\widetilde{x}, z) = \exp\left\{-\Lambda_0(\xi)\exp(\widetilde{x}^\top \beta_{x0} + z^\top \beta_{z0})\right\}$, and

$$\begin{aligned}\psi_i(\beta_0|\widetilde{x}, z) = &\int_0^\tau \int_0^\tau S(\xi|\widetilde{x}, z)\left\{\frac{\mathrm{d}N_i(u)}{\mathcal{S}^{(0)}(u, \beta_0)}\right. \\ &\left. - \frac{\mathrm{d}\mathcal{N}(u)\exp\left(w_i^{*\top}\beta_{x0} + z_i^\top \beta_{z0}\right)I(a_i \leq u \leq y_i)}{\left\{\mathcal{S}^{(0)}(u, \beta_0)\right\}^2}\right\} m(\beta_{x0}) \\ &\times \exp\left(\widetilde{x}^\top \beta_{x0} + z^\top \beta_{z0}\right)\mathrm{d}H(\xi) + o_p(1).\end{aligned} \tag{25}$$

Let $G(a, \widehat{v})$ denote the joint distribution of A_i and $\widehat{V}_i$ where $\widehat{V}_i = \left(\widetilde{X}_i^\top, Z_i^\top\right)^\top$ and define

$$\begin{aligned}\Psi_{M1}\left(w_i, \widetilde{x}_i, z_i, y_i, a_i\right) = \frac{\partial}{\partial \beta_0}\Bigg[&\int_{-\infty}^\infty \int_0^\tau \left\{\frac{dN_i(u)}{\mathcal{S}^{(0)}(u, \beta_0)} - \frac{\mathrm{d}\mathcal{N}(u)\exp\left(\widetilde{v}_i^\top \beta_0\right)I(a_i \leq u \leq y_i)}{\left\{\mathcal{S}^{(0)}(u, \beta_0)\right\}^2}\right\} m(\beta_{x0}) \\ &\times \exp\left(\widehat{v}^\top \beta_0\right)I(u \leq a \leq \tau)\Bigg]\mathrm{d}G(a, \widehat{v}).\end{aligned}$$

Let

$$
\begin{aligned}
&\boldsymbol{\Psi}\left(w_i, \widetilde{x}_i, z_i, a_i, y_i\right) \\
&= \boldsymbol{\Phi}\left(w_i, z_i, y_i, a_i\right) - \boldsymbol{\Psi}_{M1}\left(w_i, \widetilde{x}_i, z_i, y_i, a_i\right) + \left[\int_{-\infty}^{\infty}\int_0^{\tau}\left\{\frac{1}{\mu\left(\widetilde{x}, z\right)}\frac{\partial}{\partial\beta_0}\psi_i(\beta_0|\widetilde{x}, z)\right.\right. \\
&\left.\left.-\frac{\partial\mu\left(\widetilde{x}, z\right)}{\partial\beta_0}\frac{1}{\mu^2\left(\widetilde{x}, z\right)}\psi_i(\beta_0|\widetilde{x}, z)\right\}\mathrm{d}G(a, \widehat{v})\right] - \frac{\partial}{\partial\beta_0}\Lambda_0(a_i)\exp\left(\widehat{v}_i^{\top}\beta_0\right) \\
&- \frac{1}{\mu\left(\widetilde{x}_i, z_i\right)}\frac{\partial}{\partial\beta_0}\mu\left(\widetilde{x}_i, z_i\right),
\end{aligned}
\tag{26}
$$

and

$$
\begin{aligned}
\mathcal{A}_M = E\Bigg[&\frac{\partial^2}{\partial\beta_0\partial\beta_0^{\top}}\Lambda_0(A_i)\exp\left(\widehat{V}_i^{\top}\beta_0\right) \\
&+\left\{\mu\left(\widetilde{X}_i, Z_i\right)\right\}^{-2}\left\{\mu\left(\widetilde{X}_i, Z_i\right)\frac{\partial\left\{\mu\left(\widetilde{X}_i, Z_i\right)\right\}^2}{\partial\beta_0\partial\beta_0^{\top}} - \left(\frac{\partial\mu\left(\widetilde{X}_i, Z_i\right)}{\partial\beta_0}\right)^{\otimes 2}\right\}\Bigg].
\end{aligned}
\tag{27}
$$

The following theorem shows the asymptotic results of $\widetilde{\beta}$; the proof is placed in the Supplementary Material.

Theorem 3 *Under regularity conditions given in "Appendix", estimator $\widetilde{\beta}$ obtained from (23) has the following properties:*

(1) $\widetilde{\beta}\xrightarrow{p}\beta_0$ *as* $n\to\infty$;
(2) $\sqrt{n}\left(\widetilde{\beta}-\beta_0\right)\xrightarrow{d}N(0, \mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1})$ *as* $n\to\infty$;

where $\mathcal{B}=E(\boldsymbol{\Psi}_i^{\otimes 2})$ *with* $\boldsymbol{\Psi}_i=\boldsymbol{\Psi}\left(W_i, \widetilde{X}_i, Z_i, A_i, Y_i\right)$, *and* $\mathcal{A}=\mathcal{I}+\mathcal{A}_M$ with $\mathcal{I}$ *and* $\mathcal{A}_M$ *determined by (16) and (27), respectively.*

The following theorem compares the efficiency between the estimators $\widehat{\beta}$ and $\widetilde{\beta}$ whose proof is given in the Supplementary Material.

Theorem 4 *Under regularity conditions given in "Appendix", the estimator $\widetilde{\beta}$ obtained from (23) is more efficient than the estimator $\widehat{\beta}$ determined by (14). That is,* $\operatorname{var}\left(\widehat{\beta}\right)-\operatorname{var}\left(\widetilde{\beta}\right)$ *is a positive definite matrix.*

We conclude this section with comments. The consistency of $\widehat{\beta}$ (determined by (14)) and its asymptotic distribution (i.e., Theorem 2) basically come as a result of unbiased estimating functions (e.g., Yi 2017, pp.12–13), where the unbiasedness of $\frac{\partial\ell_C}{\partial\beta}$ together with (9) is typically used. On the other hand, when establishing the consistency (as well as the asymptotic result) of $\widetilde{\beta}$ (determined by (23)), we are unable to directly employ the theory for unbiased estimating functions because the derivative $\frac{\partial \log L_M^*}{\partial \beta}$ does not necessarily have zero mean.

In the augmented pseudo-likelihood method, the first term $\widehat{\ell}_C^*$ in (22) links to the log-likelihood function ℓ_C via (9), and the second term $\widehat{\ell}_M^*$ in (22) is obtained by applying the regression calibration idea. While purely using the regression calibration method does not always ensure a consistent estimator (Wang 1999; Yi 2017, p.106; Zhao and Prentice 2014), our construction of $\log L_M^*$ allows its derivative with respect to β to approximately have a zero expectation at β_0, where the relevant conditions and detailed descriptions are summarized in Appendix A, including Lemmas 1 and 2. As a result, the expectation of the log-likelihood function (22) has the "approximate" maximizer at β_0. To show the consistency of $\widetilde{\beta}$, we utilize that the true value β_0 is the maximizer of the expectation of the likelihood function, so that the maximizer $\widetilde{\beta}$ of $\widehat{\ell}_C^* + \widehat{\ell}_M^*$ is expected to converge to the maximizer β_0 of $E\left\{ \frac{1}{n} \log(L_C^*) + \frac{1}{n} \log(L_M^*) \right\}$ (van der Vaart 1998, Sect. 5.2).

5 Inference with main/validation data

5.1 Estimation of parameters for measurement error model

In practice, the covariance matrix Σ_ϵ and parameter α for the measurement error model (6) are often unknown, and they need to be estimated from additional data sources.

To estimate the parameters in model (6), we assume the availability of an external validation sample. Let $\mathcal{M}$ and $\mathcal{V}$ denote the subject sets for the main study and the external validation study containing n and m subjects, respectively, where $\mathcal{M}$ and $\mathcal{V}$ do not overlap. That is, the available data contain measurements $\left\{ \left(y_i, a_i, \delta_i, w_i, z_i \right) : i \in \mathcal{M} \right\}$ from the main study and $\left\{ \left(w_i, z_i, x_i \right) : i \in \mathcal{V} \right\}$ from the validation sample. Hence, for the measurement error model, we have

$$W_i = \Sigma_\epsilon \alpha + X_i + \epsilon_i$$

for $i \in \mathcal{M} \cup \mathcal{V}$, where the ϵ_i are independent and identically distributed with mean zero and covariance matrix Σ_ϵ, and are independent of $\left\{ X_i, Z_i, C_i, A_i, T_i \right\}$. The distribution of (W_i, X_i, Z_i) for $i \in \mathcal{M}$ is assumed to be identical to that for $i \in \mathcal{V}$, essentially saying that the transportability of the distributions of the true covariates and their surrogate measurements is needed for the validation study and the main study. This assumption is typically imposed in applications (e.g., Carroll et al. 2006, p. 29; Yi et al. 2015) so that the information carried by the validation study can be used to delineate the measurement error degree involved in the main study data. We assume that $\lim\limits_{n\to\infty} \frac{m}{n}$ exists and is greater than 0, and let ρ denote this limit.

Estimation of α and Σ_ϵ can be carried out using the least squares regression method. Write $\gamma = \Sigma_\epsilon \alpha$ and define

$$Q(\gamma) = \sum_{i \in \mathcal{V}} \|W_i - X_i - \gamma\|_2^2 \tag{28}$$

where $\|v\|_2^2 = v^\top v$ for a column vector v. Then solving

$$\frac{\partial Q(\gamma)}{\partial \gamma} = 0$$

for γ yields

$$\hat{\gamma} = \frac{1}{m}\left(\sum_{i \in \mathcal{V}} W_i - \sum_{i \in \mathcal{V}} X_i\right). \tag{29}$$

For $i \in \mathcal{V}$, let $e_i = W_i - X_i - \hat{\gamma}$ be the residual. Since $E\left(e_i e_i^\top\right) = \frac{m-1}{m}\Sigma_\epsilon$ for $i \in \mathcal{V}$, we obtain that $E\left(\sum_{i \in \mathcal{V}} e_i e_i^\top\right) = (m-1)\Sigma_\epsilon$, which yields the unbiased estimator of Σ_ϵ:

$$\hat{\Sigma}_\epsilon = \frac{1}{m-1}\sum_{i \in \mathcal{V}} e_i e_i^\top. \tag{30}$$

Finally, since $\alpha = \Sigma_\epsilon^{-1}\gamma$, we obtain an estimator of α:

$$\hat{\alpha} = \hat{\Sigma}_\epsilon^{-1}\hat{\gamma}.$$

5.2 Two-stage estimation of parameter for survival model

To estimate the parameter β, we carry out a two-stage estimation procedure. At the first stage, we use (29) and (30) to, respectively, estimate γ and Σ_ϵ for the measurement error model, as described in Sect. 5.1. At the second stage, we estimate β using a modified version of (13) or (22), given by

$$\hat{\ell}^*_{val} = \hat{\ell}^*_{val,C} + \hat{\ell}^*_{val,M}, \tag{31}$$

where $\hat{\ell}^*_{val,C}$ and $\hat{\ell}^*_{val,M}$ are, respectively, $\hat{\ell}^*_C$ and $\hat{\ell}^*_M$ with the parameters of the measurement error model (6) replaced by their estimates obtained in the first stage. That is,

$$\begin{aligned}\hat{\ell}^*_{val,C} = \sum_{i \in \mathcal{M}} \Big[&\delta_i \log \hat{\lambda}_i + \delta_i\left\{(w_i - \hat{\gamma})^\top \beta_x + z_i^\top \beta_z\right\} \\ &- \left\{\hat{\Lambda}_0(y_i) - \hat{\Lambda}_0(a_i)\right\} \exp\left\{(w_i - \hat{\gamma})^\top \beta_x + z_i^\top \beta_z\right\}\left\{\hat{m}(\beta_x)\right\}^{-1}\Big]\end{aligned} \tag{32}$$

and

$$\begin{aligned}\widehat{\ell}^*_{val,M} =& \sum_{i\in\mathcal{M}} \log\left\{ d\widehat{H}_{val}(a_i)\right\} - \sum_{i=1}^{n} \widehat{\Lambda}_0(a_i)\exp\left(\widehat{x}_{val,i}^\top \beta_x + z_i^\top \beta_z\right) \\ &- \sum_{i\in\mathcal{M}} \log\left[\int_0^\infty \exp\left\{-\widehat{\Lambda}_0(u)\exp\left(\widehat{x}_{val,i}^\top \beta_x + z_i^\top \beta_z\right)\right\} d\widehat{H}_{val}(u)\right],\end{aligned} \tag{33}$$

where $\widehat{m}(\beta_x) = \exp\left(\frac{1}{2}\beta_x^\top \widehat{\Sigma}_\epsilon \beta_x\right)$, $\widehat{x}_{val,i} = \widehat{\mu}_{W^*} + \left(\widehat{\Sigma}_{W^*} - \widehat{\Sigma}_\epsilon\right)^\top \widehat{\Sigma}_{W^*}^{-1}(w_i^* - \widehat{\mu}_{W^*})$, and

$$\widehat{H}_{val}(a) = \left(\sum_{i=1}^{n} \frac{1}{\widehat{S}(a_i|\widehat{x}_{val,i}, z_i)}\right)^{-1} \sum_{i=1}^{n} \frac{I(a_i \le a)}{\widehat{S}(a_i|\widehat{x}_{val,i}, z_i)}.$$

By analogy to (14) and (23), two estimators of β can then be obtained by maximizing (32) and the pseudo-likelihood (31), respectively, given by

$$\widehat{\beta}_{val} = \underset{\beta}{\text{argmax}}\ \widehat{\ell}^*_{val,C}, \tag{34}$$

and

$$\widetilde{\beta}_{val} = \underset{\beta}{\text{argmax}}\,(\widehat{\ell}^*_{val,C} + \widehat{\ell}^*_{val,M}). \tag{35}$$

5.3 Asymptotic properties

We now explore the asymptotic results for the two estimators of β described in Sect. 5.2; the proofs are placed in the Supplementary Material. Different from the setting in Sects. 3 and 4 where the true values of the parameters in the measurement error model (6) are assumed to be known, here the true values of the parameters Σ_ϵ and α for model (6) are unknown and must be estimated from the validation sample $\mathcal{V}$. Let ζ_i be the indicator whether or not subject i belongs to the validation sample $\mathcal{V}$, i.e., $\zeta_i = 1$ if $i \in \mathcal{M}$ and $\zeta_i = 0$ if $i \in \mathcal{V}$.

Let $\gamma_0 = \Sigma_{\epsilon;0}\alpha_0$ denote the true value of γ which is defined before (28). Define

$$\begin{aligned}\mathbb{V} =& E\left\{N_i(\tau)\right\}(1-\zeta_i) \times \left[\left\{\begin{pmatrix}-1\\0\end{pmatrix} + \frac{1}{\mathcal{S}^{(0)}(u;\beta_0)}\frac{\partial \mathcal{S}^{(1)}(u;\beta_0)}{\partial \gamma_0}\right\}\right. \\ &\left.\times\left(W_i - X_i\right) + \frac{m}{m-1}\left\{\epsilon_i\epsilon_i^\top - (m-1)\Sigma_{\epsilon;0}\right\}\beta_{x0}\right]\end{aligned}$$

and

$$\mathcal{B}_{val1,i} = \sqrt{1+\rho}\,\zeta_i \boldsymbol{\Phi}\left(W_i, Z_i, Y_i, A_i\right) + \frac{\sqrt{1+\rho}}{\rho}\mathbb{V}. \tag{36}$$

Theorem 5 *Under regularity conditions in "Appendix", we have that as $n \to \infty$,*

(1) $\widehat{\beta}_{val} \xrightarrow{p} \beta_0$;

(2) $\sqrt{n}\left(\widehat{\beta}_{val} - \beta_0\right) \xrightarrow{d} N\left(0, \mathcal{I}^{-1}\mathcal{J}_{val}\mathcal{I}^{-1}\right)$,

where $\mathcal{J}_{val} = E\left\{\left(\mathcal{B}_{val1,i}\right)^{\otimes 2}\right\}$ *and* $\mathcal{I}$ *is given by* (16).

Theorem 5 resembles Theorem 2 in that both theorems describe the asymptotic properties for the estimators, $\widehat{\beta}$ and $\widehat{\beta}_{val}$, of β, respectively, based on using $\widehat{\ell}^*_C$ in (13) and $\widehat{\ell}^*_{val,C}$ in (32). The asymptotic covariance matrices for $\widehat{\beta}$ and $\widehat{\beta}_{val}$ share the same sandwich form with the same side matrix $\mathcal{I}$ but different inner matrices $\mathcal{J}$ and $\mathcal{J}_{val}$. Matrix $\mathcal{J}$ is formulated for the setting where the true value of the parameters α and Σ_ϵ associated with the measurement error model are known, while $\mathcal{J}_{val}$ is constructed for the scenario where α and Σ_ϵ are estimated from a validation sample. The induced variability of estimating α and Σ_ϵ is reflected by the inclusion of $\mathbb{V}$ as well as ρ in $\mathcal{J}_{val}$. A general discussion on these aspects can be found in Yi (2017, pp.25–27).

Let

$$\mathcal{E}_{val,1} = E\left[\frac{\partial}{\partial\beta_0}\int_0^\tau \mathrm{d}\mathcal{N}(u)\frac{1}{\left\{\mathcal{S}^{(0)}(u;\beta_0)\right\}^2}\frac{\partial\mathcal{S}^{(0)}(u;\beta_0)}{\partial\gamma_0}m(\beta_{x0})\times\exp\left\{\widetilde{X}_i^\top\beta_{x0} + Z_i^\top\beta_{z0}\right\}I(u \le A_i \le \tau)\right]$$

and

$$\varphi_{val,i} = \left[\frac{\sqrt{1+\rho}}{\rho}\left(W_i - X_i\right)\int_0^\tau\int_0^\tau -\left\{S(v|\widetilde{x},z)\frac{\mathrm{d}\mathcal{N}(t)}{\left\{\mathcal{S}^{(0)}(t;\beta_0)\right\}^2}\frac{\partial\mathcal{S}^{(0)}(t;\beta_0)}{\partial\gamma_0}\right.\right.$$
$$\left.\left.\times m(\beta_{x0})\exp\left(\widetilde{x}^\top\beta_{x0} + z^\top\beta_{z0}\right)\mathrm{d}H(v)\right\} + \sqrt{1+\rho}\psi_i(\beta_0|\widetilde{x},z)\right].$$

Define

$$\mathcal{B}_{val2,i} = -\sqrt{1+\rho}\zeta_i\Psi_{M1}\left(W_i,\widetilde{X}_i,Z_i,Y_i,A_i\right) + \frac{\sqrt{1+\rho}}{\rho}\mathcal{E}_{val,1}(1-\zeta_i)(W_i - X_i), \tag{37}$$

$$\mathcal{B}_{val3,i} = \left[\int_{-\infty}^{\infty}\int_0^\tau\left\{\frac{1}{\mu\left(\widetilde{x},z\right)}\frac{\partial}{\partial\beta_0}\varphi_{val,i} - \frac{\partial\mu\left(\widetilde{x},z\right)}{\partial\beta_0}\frac{1}{\mu^2\left(\widetilde{x},z\right)}\varphi_{val,i}\right\}\right]\mathrm{d}G(a,\widehat{v}), \tag{38}$$

and

$$\widetilde{U}_{M,val,i} = -\frac{\partial}{\partial\beta_0}\Lambda_0(A_i)\exp\left(\widehat{V}_i^\top\beta_{x0}\right) - \frac{1}{\mu\left(\widetilde{X}_i,Z_i\right)}\frac{\partial\mu\left(\widetilde{X}_i,Z_i\right)}{\partial\beta_0}.$$

Theorem 6 *Under regularity conditions in* "Appendix", *we have that as* $n \to \infty$,

(1) $\widetilde{\beta}_{val} \xrightarrow{p} \beta_0$;

(2) $\sqrt{n}\left(\widetilde{\beta}_{val} - \beta_0\right) \xrightarrow{d} N\left(0, \mathcal{A}^{-1}\mathcal{B}_{val}\mathcal{A}^{-1}\right)$,

where

$$\mathcal{B}_{val} = E\left\{ \left(\mathcal{B}_{val1,i} + \mathcal{B}_{val2,i} + \mathcal{B}_{val3,i} + \sqrt{1+\rho}\zeta_i \widetilde{U}_{M,val,i}\right)^{\otimes 2} \right\} \tag{39}$$

and $\mathcal{A}$ is defined in Theorem 3.

Similar to the comments after Theorem 5, Theorem 6 shares some similarity to Theorem 3, and their difference is reflected by the different expressions of $\mathcal{B}$ and $\mathcal{B}_{val}$. Theorems 5 and 6 establish the asymptotic results for the two estimators $\widehat{\beta}_{val}$ and $\widetilde{\beta}_{val}$. These results offer the basis of conducting inference about β such as calculating confidence intervals or performing hypothesis testing. While both $\widehat{\beta}_{val}$ and $\widetilde{\beta}_{val}$ are consistent estimators of β, their efficiencies are different, as shown in the following theorem.

Theorem 7 *Under regularity conditions given in* "Appendix", *the estimator $\widetilde{\beta}_{val}$ obtained from* (35) *is more efficient than the estimator $\widehat{\beta}_{val}$ determined by* (34). *That is*, $\mathrm{var}\left(\widehat{\beta}_{val}\right) - \mathrm{var}\left(\widetilde{\beta}_{val}\right)$ *is a positive definite matrix.*

6 Numerical studies

We conduct simulation studies to assess the finite sample performance of the proposed estimators $\widehat{\beta}$ and $\widetilde{\beta}$ under a variety of settings. In contrast, we also report the performance of the naive estimator which is obtained by disregarding the feature of measurement error.

6.1 Design setup

We consider the setting where the baseline hazard function is set as $\lambda_0(t) = 2t$ and the truncation time A^* is generated from the exponential distribution with mean 10. We consider two scenarios of generating the true covariates $V = (X, Z)^\top$. In Scenario 1, X and Z are generated from a bivariate normal distribution with mean zero and covariance matrix Σ; in Scenario 2, X is generated from an exponential distribution with mean one and Z is independently generated from the standard normal distribution.

Given $\lambda_0(t)$ and $(X, Z)^\top$, the failure time T^* is generated from the model:

$$\lambda(T^*|X, Z) = 2T^* \exp\left(X\beta_{x0} + Z\beta_{z0}\right),$$

where $\beta_0 = (\beta_{x0}, \beta_{z0})^\top$ is the vector parameters. That is, T^* is set as

$$\left\{-\exp\left(X\beta_{x0}+Z\beta_{z0}\right)\log(1-U)\right\}^{1/2},$$

where U is simulated from the uniform distribution $U(0, 1)$. The untruncated data $\{A, T\}$ is collected from $\{A^*, T^*\}$ by conditioning on that $T^* \geq A^*$.

We consider three censoring rates, 0%, 25%, and 50%, and let the censoring time C be generated from the uniform distribution $U(0, c)$, where c is determined by a given censoring rate. Consequently, Y and Δ are determined by $Y = \min\{T, A + C\}$ and $\Delta = I(T \leq A + C)$. One thousand five hundred simulations are run for each of the following settings:

Setting 1: $n = 200$, Scenario 1 with $\Sigma = \begin{pmatrix} 4 & 0.5 \\ 0.5 & 36 \end{pmatrix}$, and $(\beta_{x0}, \beta_{z0})^\top = (0.3, 1)^\top$.

Setting 2: This is the same as Setting 1 except that n is increased to $n = 1000$.

Setting 3: This is the same as Setting 1 except that β_{x0} is increased to be $\beta_{x0} = \log 3$.

Setting 4: This is the same as Setting 1 except that Σ is changed to be $\Sigma = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}$.

Setting 5: Simultaneously increase the values in Setting 1 to be

$$n = 1000, \Sigma = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}, \text{ and } \beta_{x0} = \log 3.$$

Setting 6: This is the same as Setting 1 except that the covariates are generated as in Scenario 2.

For the measurement error process, we consider model (6) with error $\epsilon \sim N\left(0, \Sigma_\epsilon\right)$, where α is set as 0, 10, 50, or 100; and variance Σ_ϵ is taken as 0.01, 0.5, or 0.75, describing different degrees of measurement error. These values of Σ_ϵ, respectively, yield the reliability ratio to be 0.997, 0.888, and 0.842 for Settings 1–3, and 0.990, 0.667, and 0.571 for Settings 4–6, where the reliability ratio is defined as Σ_X/Σ_W with Σ_X and Σ_W representing the variances of X and W, respectively.

6.2 Performance of proposed estimators: α and Σ_ϵ are known

We analyze the simulated data using the estimation methods described in Sects. 3 and 4 with the parameters of the measurement error model (6) assumed known. As a comparison, we conduct naive analysis with measurement error ignored.

We report finite sample biases (Bias) of the estimates, empirical standard errors (SEE), model-based standard errors (SEM), and the mean squared errors (MSE) under the measurement error model (6) with $\alpha = 0$. The results for Settings 1, 5, and 6 are reported in Tables 1, 2 and 3, respectively. Other results are reported in the Supplementary Material to save space.

For each setting with a given measurement error degree Σ_ϵ and a given censoring percentage, the three methods perform differently. The naive method performs the worst with tremendous biases, but the proposed conditional profile likelihood and

Table 1 Simulation results under Setting 1 and measurement error model (6) with $\alpha = 0$

Σ_ϵ	cr	Method	Estimator of β_x					Estimator of β_z				
			Bias	SEE	SEM	MSE	CP (%)	Bias	SEE	SEM	MSE	CP (%)
0.01	0%	Naive	0.113	0.031	0.030	0.014	19.8	0.106	0.042	0.040	0.013	56.7
		Conditional $(\widehat{\beta})$	0.008	0.048	0.043	0.002	95.4	0.017	0.066	0.065	0.004	94.9
		Full $(\widetilde{\beta})$	0.006	0.043	0.042	0.002	95.3	0.019	0.048	0.047	0.003	95.1
	25%	Naive	0.117	0.036	0.034	0.015	15.0	0.113	0.058	0.056	0.016	42.6
		Conditional $(\widehat{\beta})$	0.010	0.051	0.049	0.003	95.0	0.018	0.073	0.073	0.006	94.9
		Full $(\widetilde{\beta})$	0.007	0.046	0.045	0.002	95.0	0.018	0.065	0.064	0.004	95.1
	50%	Naive	0.128	0.051	0.050	0.019	8.5	0.125	0.064	0.063	0.020	32.3
		Conditional $(\widehat{\beta})$	0.010	0.061	0.058	0.004	95.0	0.017	0.098	0.095	0.010	95.0
		Full $(\widetilde{\beta})$	0.007	0.057	0.055	0.003	94.8	0.017	0.067	0.066	0.005	94.8
0.5	0%	Naive	0.124	0.039	0.039	0.018	6.2	0.112	0.051	0.050	0.015	45.1
		Conditional $(\widehat{\beta})$	0.024	0.052	0.048	0.003	94.6	0.026	0.076	0.071	0.006	94.6
		Full $(\widetilde{\beta})$	0.011	0.045	0.043	0.002	94.9	0.017	0.067	0.057	0.005	95.4
	25%	Naive	0.126	0.040	0.039	0.018	5.9	0.126	0.066	0.056	0.020	35.2
		Conditional $(\widehat{\beta})$	0.026	0.058	0.055	0.004	94.9	0.025	0.083	0.080	0.008	94.7
		Full $(\widetilde{\beta})$	0.013	0.054	0.052	0.002	95.1	0.015	0.082	0.079	0.007	95.0
	50%	Naive	0.132	0.053	0.032	0.023	5.1	0.135	0.073	0.073	0.024	25.1
		Conditional $(\widehat{\beta})$	0.031	0.072	0.068	0.006	94.4	0.033	0.100	0.099	0.011	94.6
		Full $(\widetilde{\beta})$	0.017	0.057	0.054	0.003	95.4	0.026	0.087	0.085	0.009	95.0

Table 1 (continued)

Σ_ϵ	cr	Method	Estimator of β_x					Estimator of β_z				
			Bias	SEE	SEM	MSE	CP (%)	Bias	SEE	SEM	MSE	CP (%)
0.75	0%	Naive	0.126	0.040	0.039	0.016	5.6	0.133	0.053	0.052	0.020	37.5
		Conditional $\left(\widehat{\beta}\right)$	0.040	0.057	0.056	0.005	93.8	0.036	0.078	0.077	0.007	94.0
		Full $\left(\widetilde{\beta}\right)$	− 0.013	0.053	0.051	0.002	95.0	− 0.021	0.068	0.066	0.005	94.8
	25%	Naive	0.130	0.042	0.041	0.020	4.2	0.134	0.071	0.070	0.022	33.0
		Conditional $\left(\widehat{\beta}\right)$	0.042	0.067	0.064	0.006	93.9	0.044	0.090	0.088	0.010	93.7
		Full $\left(\widetilde{\beta}\right)$	− 0.015	0.055	0.054	0.003	94.8	0.023	0.087	0.083	0.008	94.6
	50%	Naive	0.135	0.055	0.054	0.020	3.8	0.142	0.076	0.076	0.026	22.4
		Conditional $\left(\widehat{\beta}\right)$	0.040	0.094	0.092	0.008	94.6	0.045	0.106	0.104	0.013	94.0
		Full $\left(\widetilde{\beta}\right)$	− 0.018	0.058	0.056	0.003	95.4	0.029	0.088	0.086	0.008	94.6

Table 2 Simulation results under Setting 5 and measurement error model (6) with $\alpha = 0$

Σ_ϵ	cr	Method	Estimator of β_x					Estimator of β_z				
			Bias	SEE	SEM	MSE	CP (%)	Bias	SEE	SEM	MSE	CP (%)
0.01	0%	Naive	0.095	0.021	0.020	0.010	9.9	0.114	0.023	0.022	0.014	10.3
		Conditional $\left(\widehat{\beta}\right)$	0.024	0.036	0.035	0.002	94.8	0.020	0.037	0.035	0.002	94.8
		Full $\left(\widetilde{\beta}\right)$	0.022	0.025	0.024	0.001	94.9	0.016	0.029	0.027	0.001	95.1
	25%	Naive	0.109	0.025	0.025	0.013	8.5	0.125	0.032	0.030	0.017	9.8
		Conditional $\left(\widehat{\beta}\right)$	0.023	0.040	0.040	0.002	94.8	0.018	0.038	0.036	0.002	94.7
		Full $\left(\widetilde{\beta}\right)$	0.022	0.026	0.026	0.001	94.7	0.015	0.035	0.035	0.001	94.8
	50%	Naive	0.115	0.032	0.032	0.014	7.6	0.133	0.036	0.036	0.019	8.8
		Conditional $\left(\widehat{\beta}\right)$	0.024	0.041	0.041	0.002	94.7	0.023	0.044	0.044	0.002	94.6
		Full $\left(\widetilde{\beta}\right)$	0.023	0.036	0.036	0.003	94.8	0.017	0.040	0.039	0.002	94.7
0.5	0%	Naive	0.117	0.026	0.025	0.014	9.7	0.124	0.033	0.032	0.016	9.7
		Conditional $\left(\widehat{\beta}\right)$	0.056	0.037	0.036	0.005	93.5	0.048	0.048	0.047	0.005	93.6
		Full $\left(\widetilde{\beta}\right)$	− 0.028	0.031	0.032	0.002	95.3	0.019	0.038	0.036	0.002	95.1
	25%	Naive	0.128	0.028	0.025	0.017	8.3	0.129	0.033	0.033	0.018	8.9
		Conditional $\left(\widehat{\beta}\right)$	0.058	0.041	0.041	0.005	93.7	0.051	0.050	0.049	0.006	93.5
		Full $\left(\widetilde{\beta}\right)$	0.029	0.031	0.029	0.002	95.3	0.019	0.040	0.040	0.002	94.5
	50%	Naive	0.125	0.033	0.032	0.017	7.3	0.136	0.036	0.035	0.020	8.5
		Conditional $\left(\widehat{\beta}\right)$	0.058	0.045	0.045	0.005	93.3	0.059	0.055	0.054	0.005	93.0
		Full $\left(\widetilde{\beta}\right)$	0.028	0.037	0.035	0.002	94.8	0.020	0.043	0.042	0.002	94.7

Table 2 (continued)

Σ_ϵ	cr	Method	Estimator of β_x					Estimator of β_z				
			Bias	SEE	SEM	MSE	CP (%)	Bias	SEE	SEM	MSE	CP (%)
0.75	0%	Naive	0.163	0.029	0.028	0.027	8.3	0.175	0.034	0.034	0.032	8.9
		Conditional $\left(\widehat{\beta}\right)$	0.068	0.040	0.040	0.006	93.7	0.066	0.050	0.050	0.008	93.7
		Full $\left(\widetilde{\beta}\right)$	− 0.039	0.034	0.033	0.003	94.7	− 0.020	0.040	0.040	0.002	95.4
	25%	Naive	0.174	0.033	0.032	0.031	6.3	0.176	0.035	0.034	0.032	5.5
		Conditional $\left(\widehat{\beta}\right)$	0.071	0.042	0.040	0.007	93.3	0.072	0.063	0.061	0.011	93.4
		Full $\left(\widetilde{\beta}\right)$	− 0.047	0.036	0.036	0.003	94.7	− 0.023	0.043	0.042	0.002	95.2
	50%	Naive	0.186	0.035	0.033	0.036	4.3	0.182	0.039	0.038	0.035	3.4
		Conditional $\left(\widehat{\beta}\right)$	0.076	0.049	0.046	0.008	93.0	0.074	0.064	0.063	0.007	93.4
		Full $\left(\widetilde{\beta}\right)$	− 0.046	0.041	0.040	0.004	94.5	− 0.022	0.047	0.046	0.003	94.4

Table 3 Simulation results under Setting 6 and measurement error model (6) with $\alpha = 0$

Σ_ϵ	cr	Method	Estimator of β_x					Estimator of β_z				
			Bias	SEE	SEM	MSE	CP (%)	Bias	SEE	SEM	MSE	CP (%)
0.01	0%	Naive	0.144	0.029	0.028	0.023	10.8	0.141	0.036	0.035	0.021	11.3
		Conditional $\left(\widehat{\beta}\right)$	0.007	0.054	0.054	0.003	95.2	− 0.004	0.060	0.060	0.004	95.3
		Full $\left(\widetilde{\beta}\right)$	0.004	0.040	0.039	0.002	95.3	0.007	0.047	0.046	0.002	95.1
	25%	Naive	0.146	0.029	0.029	0.022	10.7	0.146	0.038	0.038	0.023	10.5
		Conditional $\left(\widehat{\beta}\right)$	0.011	0.056	0.056	0.003	95.0	0.010	0.067	0.067	0.005	94.8
		Full $\left(\widetilde{\beta}\right)$	0.008	0.045	0.042	0.002	95.0	0.007	0.053	0.051	0.003	94.7
	50%	Naive	0.149	0.029	0.029	0.023	10.1	0.168	0.041	0.040	0.030	8.4
		Conditional $\left(\widehat{\beta}\right)$	0.012	0.058	0.057	0.003	95.0	0.014	0.069	0.069	0.005	94.8
		Full $\left(\widetilde{\beta}\right)$	0.010	0.053	0.052	0.003	94.8	0.009	0.054	0.054	0.003	94.7
0.5	0%	Naive	0.162	0.033	0.032	0.027	9.9	0.145	0.041	0.041	0.023	9.8
		Conditional $\left(\widehat{\beta}\right)$	0.041	0.057	0.057	0.005	94.8	0.027	0.083	0.082	0.007	94.9
		Full $\left(\widetilde{\beta}\right)$	0.013	0.048	0.047	0.002	95.2	0.012	0.052	0.050	0.003	95.0
	25%	Naive	0.168	0.037	0.036	0.030	9.4	0.154	0.048	0.047	0.026	8.5
		Conditional $\left(\widehat{\beta}\right)$	0.047	0.067	0.066	0.007	94.6	0.029	0.084	0.084	0.008	94.6
		Full $\left(\widetilde{\beta}\right)$	0.015	0.050	0.050	0.003	95.1	0.015	0.054	0.054	0.003	95.1
	50%	Naive	0.186	0.046	0.045	0.037	8.2	0.170	0.050	0.050	0.031	5.6
		Conditional $\left(\widehat{\beta}\right)$	0.050	0.068	0.068	0.006	94.5	0.033	0.089	0.088	0.009	94.7
		Full $\left(\widetilde{\beta}\right)$	0.019	0.057	0.054	0.004	94.8	0.015	0.060	0.058	0.004	94.9

Table 3 (continued)

Σ_ϵ	cr	Method	Estimator of β_x					Estimator of β_z				
			Bias	SEE	SEM	MSE	CP (%)	Bias	SEE	SEM	MSE	CP (%)
0.75	0%	Naive	0.181	0.042	0.041	0.034	5.6	0.149	0.047	0.046	0.024	6.9
		Conditional $\left(\hat{\beta}\right)$	0.057	0.070	0.070	0.008	94.8	0.033	0.098	0.098	0.010	95.1
		Full $\left(\tilde{\beta}\right)$	0.019	0.055	0.055	0.002	95.2	0.016	0.061	0.058	0.004	95.1
	25%	Naive	0.232	0.044	0.044	0.056	4.1	0.168	0.052	0.050	0.031	6.8
		Conditional $\left(\hat{\beta}\right)$	0.063	0.073	0.072	0.009	94.3	0.036	0.106	0.105	0.012	94.8
		Full $\left(\tilde{\beta}\right)$	0.021	0.056	0.055	0.002	95.0	0.022	0.063	0.062	0.004	95.0
	50%	Naive	0.266	0.047	0.046	0.073	2.5	0.187	0.053	0.052	0.038	5.6
		Conditional $\left(\hat{\beta}\right)$	0.066	0.077	0.077	0.010	94.0	0.046	0.107	0.107	0.013	94.5
		Full $\left(\tilde{\beta}\right)$	0.025	0.061	0.060	0.003	94.8	0.023	0.065	0.065	0.005	94.7

the augmented pseudo-likelihood methods give satisfactory performance. Regarding SEE and SEM, they are fairly comparable for each method, though the discrepancies seem to be more noticeable for the naive method in some settings. As expected, the methods of correcting for measurement error effects yield larger SEE and SEM than the naive approach does. This is the price paid to remove biases in point estimators; this phenomenon is typical in the literature of measurement error models. Moreover, mean squared errors produced by the two proposed methods are a lot smaller than those obtained from the naive method, suggesting that the efforts of adjusting for measurement error effects are worthwhile. In terms of the performance of the two correction methods for measurement error effects, simulation results confirm that the augmented pseudo-likelihood method is more efficient than the conditional profile likelihood approach.

For a setting with a given censoring percentage, the degree of measurement error affects the performance of each method. As the measurement error degree increases, biases of the point estimate produced by the naive and proposed methods also tend to increase while the associated standard errors do not seem to considerably change. Within a setting with a given measurement error degree, the performance of each method is influenced by the censoring percentage. The performance tends to deteriorate as the censoring proportion increases. Furthermore, comparing the results in different settings, we see the impact of the sample size, the magnitude of covariate effects and the correlation among the covariates on the performance of the three methods. The results for Setting 6 show that when the error-prone covariate X follows a skewed distribution, the two proposed methods still have satisfactory performance with reasonably small finite sample biases and SEEs.

In summary, simulation results demonstrate that the naive method produces considerable finite sample biases with coverage rates of 95% confidence intervals significantly departing from the nominal level. Both the conditional profile likelihood and the augmented pseudo-likelihood methods output satisfactory estimates with small finite sample biases and reasonable coverage rates of 95% confidence intervals. The augmented pseudo-likelihood method is more efficient than the conditional profile likelihood method.

6.3 Performance with validation data

In this subsection, we evaluate the performance of the proposed method in Sect. 5 for situations where the main study and the validation study are available; the data from the main study are generated as Setting 1 in Sect. 6.1, and the external validation data with size $|\mathcal{V}| = 100$ are generated independently following the procedure in Sect. 6.1, where the true parameter values of the measurement error model (6) are set as $\alpha = 0, 10, 50$ or 100 and $\Sigma_\epsilon = 0.010$, 0.500, or 0.750, corresponding to increasing degrees of measurement error. In Table 4, we report the results for the settings with $\alpha = 100$ and defer the results for $\alpha = 0, 10$, and 50 to the Supplementary Material.

We first apply the estimation procedure described in Sect. 5.1 to estimate α and Σ_ϵ. Corresponding to $\Sigma_\epsilon = 0.010$, 0.500, and 0.750, we obtain estimates of Σ_ϵ:

Table 4 Simulation results under Setting 1 and measurement error model (6) with $\alpha = 100$ in the presence of validation data

Σ_ϵ	cr	Method	Estimator of β_x					Estimator of β_z				
			Bias	SEE	SEM	MSE	CP (%)	Bias	SEE	SEM	MSE	CP (%)
0.01	0%	Naive	0.079	0.033	0.030	0.007	12.5	0.107	0.045	0.045	0.013	15.4
		Conditional $\left(\widehat{\beta}_{val}\right)$	0.003	0.043	0.041	0.002	95.5	0.017	0.070	0.066	0.005	95.4
		Full $\left(\widetilde{\beta}_{val}\right)$	0.004	0.042	0.040	0.002	95.4	0.010	0.068	0.065	0.005	95.2
	25%	Naive	0.091	0.039	0.037	0.010	11.4	0.108	0.048	0.045	0.014	14.9
		Conditional $\left(\widehat{\beta}_{val}\right)$	− 0.003	0.050	0.047	0.003	95.3	0.018	0.075	0.074	0.006	95.2
		Full $\left(\widetilde{\beta}_{val}\right)$	0.002	0.045	0.043	0.002	95.3	0.015	0.074	0.072	0.006	95.0
	50%	Naive	0.117	0.040	0.040	0.015	10.0	0.126	0.050	0.048	0.018	11.2
		Conditional $\left(\widehat{\beta}_{val}\right)$	0.009	0.064	0.057	0.004	95.2	0.020	0.097	0.089	0.010	95.2
		Full $\left(\widetilde{\beta}_{val}\right)$	0.005	0.057	0.057	0.003	95.0	0.019	0.080	0.080	0.007	95.0
0.5	0%	Naive	0.145	0.040	0.040	0.023	9.4	0.117	0.050	0.050	0.016	13.8
		Conditional $\left(\widehat{\beta}_{val}\right)$	0.028	0.048	0.047	0.003	94.6	0.027	0.073	0.070	0.006	94.7
		Full $\left(\widetilde{\beta}_{val}\right)$	0.012	0.044	0.043	0.002	95.0	0.012	0.071	0.069	0.005	95.4
	25%	Naive	0.147	0.042	0.042	0.023	8.8	0.143	0.053	0.053	0.023	8.4
		Conditional $\left(\widehat{\beta}_{val}\right)$	0.029	0.057	0.054	0.004	94.6	0.027	0.086	0.080	0.008	94.6
		Full $\left(\widetilde{\beta}_{val}\right)$	0.015	0.047	0.045	0.002	95.0	0.017	0.078	0.078	0.007	95.4
	50%	Naive	0.148	0.044	0.044	0.022	8.4	0.191	0.056	0.055	0.040	7.6
		Conditional $\left(\widehat{\beta}_{val}\right)$	0.033	0.066	0.065	0.006	93.4	0.035	0.111	0.095	0.014	94.0
		Full $\left(\widetilde{\beta}_{val}\right)$	0.020	0.057	0.053	0.004	95.3	0.021	0.084	0.084	0.008	94.8

Table 4 (continued)

Σ_ϵ	cr	Method	Estimator of β_x					Estimator of β_z				
			Bias	SEE	SEM	MSE	CP (%)	Bias	SEE	SEM	MSE	CP (%)
0.75	0%	Naive	0.153	0.043	0.043	0.020	8.2	0.124	0.053	0.053	0.016	10.0
		Conditional$\left(\widehat{\beta}_{val}\right)$	0.042	0.061	0.060	0.006	94.2	0.041	0.083	0.076	0.009	94.8
		Full$\left(\widetilde{\beta}_{val}\right)$	0.017	0.051	0.051	0.002	95.0	0.016	0.074	0.074	0.005	95.0
	25%	Naive	0.157	0.045	0.044	0.019	7.6	0.147	0.055	0.054	0.025	7.3
		Conditional$\left(\widehat{\beta}_{val}\right)$	0.043	0.065	0.062	0.006	94.0	0.043	0.089	0.085	0.009	94.8
		Full$\left(\widetilde{\beta}_{val}\right)$	0.021	0.057	0.057	0.002	95.1	0.019	0.078	0.078	0.006	95.3
	50%	Naive	0.167	0.048	0.047	0.030	7.6	0.239	0.059	0.058	0.060	5.9
		Conditional$\left(\widehat{\beta}_{val}\right)$	0.048	0.080	0.076	0.009	93.6	0.050	0.113	0.113	0.015	94.2
		Full$\left(\widetilde{\beta}_{val}\right)$	0.023	0.058	0.056	0.004	95.0	0.024	0.085	0.082	0.007	94.8

$\widehat{\Sigma}_\epsilon = 0.010$, 0.497, and 0.743, respectively, with the corresponding standard errors 0.001, 0.035, and 0.051; and the corresponding estimates of α are 100.746, 101.154, and 101.492, with the associated standard errors 7.062, 7.029, 7.030, respectively. Then, we analyze the data from the main study using the estimators $\widehat{\beta}_{val}$ and $\widetilde{\beta}_{val}$ derived from (34) and (35), respectively.

The results uncover similar findings to those revealed in Sect. 6.2 and demonstrate satisfactory finite sample performance of the proposed estimators $\widehat{\beta}_{val}$ and $\widetilde{\beta}_{val}$. The results also confirm that $\widetilde{\beta}_{val}$ is more efficient than $\widehat{\beta}_{val}$. Finally, we note that the estimator $\widehat{\beta}$ (or $\widetilde{\beta}$), derived under the measurement error model with its parameters *given*, is not more efficient than $\widehat{\beta}_{val}$ (or $\widetilde{\beta}_{val}$), derived from that the parameters of the measurement error model are *estimated*. This counterintuitive phenomenon has been observed for various settings by many authors, including Robins et al. (1994), Yan and Yi (2015), and Ning et al. (2018). This paradox does not appear when estimation is based on a likelihood method but may occur when using estimation equation methods (Henmi and Eguchi 2004; Yi 2017, Sect. 1.3.4).

6.4 Analysis of Worcester Heart Attack Study

In this section, we apply the proposed methods to analyze the data arising from the Worcester Heart Attack Study (WHAS500), which are described in Sect. 1. As in Hosmer et al. (2008), a survival time was defined as the time since a subject was admitted to the hospital. We are interested in studying survival times of patients who were discharged alive from the hospital. Hence, a selection criterion was imposed that only those subjects who were discharged alive were eligible to be included in the analysis. That is, individuals were not enrolled in the analysis if they died before discharging from the hospital, hence left-truncation occurs. With such a criterion, a sample of size 461 was available. In this data set, the censoring rate is 61.8%. To be more specific, the total length of follow-up (lenfol) is the last event time (i.e., $Y_i = \min\left(T_i, C_i\right)$), the length of hospital stay (los) is the truncation time (i.e., A_i), and the vital status at last follow-up (fstat) is Δ_i.

In our analysis, the covariates include the body mass index (BMI) and the initial heart rate (HR) of a patient, where BMI is regarded to be error-contaminated. This consideration is driven by the fact that measurement error in BMI is commonly considered in the literature. For example, in analyzing the Korean Longitudinal Study of Aging (KLoSA) data, Xu et al. (2017, p.196) discussed error-prone BMI. Wang (2000, Sect. 5) considered a study of childhood growth where BMI is taken as measurement error involved. Carroll and Li (1992, Sect. 8) studied the breast cancer data which include body mass subject to mismeasurement. Furthermore, as commented by Rothman (2008), BMI may be subject to mismeasurement due to errors in self-reported data. In our analysis here, we let X denote the true value of BMI and let W denote its observed value, and they are assumed to follow the measurement error model (6). Let Z denote HR.

In this data set, while we are able to calculate the sample variance $\widehat{\Sigma}_W$ for the observed value W (which is 0.041), there is no additional data source such as a validation subsample to allow us to quantify the degree of measurement error. To get around this, we carry out sensitivity analyses, a common and viable strategy to understand the impact of measurement error on estimation. That is, given a range of representative values for Σ_ϵ and α, we estimate β using $\widehat{\beta}$ and $\widetilde{\beta}$ via (14) and (23), respectively; and we want to assess how sensitive the results are to different degrees of measurement error. Here, we particularly consider the case where α is set as 0, 10, 50, or 100, and the specification of Σ_ϵ is guided by using the reliability ratio $R = \frac{\Sigma_X}{\Sigma_W}$ where Σ_X and Σ_W are the variances of X and W, respectively (Yi 2017, p.46). Noting that $\Sigma_W = \Sigma_X + \Sigma_\epsilon$, i.e., Σ_X is no bigger than Σ_W, we take Σ_X to be a value about 10% smaller than Σ_W, bearing in mind that other percentages can be considered as well. As Σ_W is unknown, we use the sample variance $\widehat{\Sigma}_W$ to specify the value of Σ_X, $\Sigma_X = 90\% \cdot \widehat{\Sigma}_W$. To feature different degrees of measurement error in a possibly plausible range, we let R take a value in [0.6, 0.9]. That is, Σ_ϵ is specified as $\Sigma_X\left(R^{-1} - 1\right)$, taking a value in the range $\left[0.9 \times 0.041 \times \left(\frac{1}{0.9} - 1\right), 0.9 \times 0.041 \times \left(\frac{1}{0.6} - 1\right)\right] = [0.004, 0.025]$. The analysis results for $\alpha = 0, 10, 50$ and 100 are shown in Figs. 2 and 3, respectively. The performance of $\widetilde{\beta} = \left(\widetilde{\beta}_x, \widetilde{\beta}_z\right)^\top$ and $\widehat{\beta} = \left(\widehat{\beta}_x, \widehat{\beta}_z\right)^\top$ shows insensitivity to different values of α. The estimates obtained from the estimator $\widetilde{\beta}$ (defined in (23)) seem to remain stable regardless of changes of Σ_ϵ, while the estimates derived from $\widehat{\beta}$ (defined in (14)) tend to decrease as Σ_ϵ increases.

To see the analysis results more closely, in Table 5, we further report the point estimates (EST), model-based standard errors (SEM) determined by the asymptotic variances in Theorems 2 and 3, and p-values produced by $\widehat{\beta}$ and $\widetilde{\beta}$ for the cases with $\Sigma_\epsilon = 0.004$, 0.010 and 0.018 as well as the case without error (i.e., $\Sigma_\epsilon = 0$). While the point estimates produced by the two approaches are different, the associated standard errors for $\widetilde{\beta}$ are smaller than those of $\widehat{\beta}$, confirming the theoretical result established in Theorem 4. It is seen that when $\Sigma_\epsilon = 0$, the results for both methods remain unchanged to the four values of α. The estimator $\widehat{\beta}_x$ appears to change more noticeably than other estimators (i.e., $\widehat{\beta}_z$, $\widetilde{\beta}_z$ and $\widetilde{\beta}_z$) when measurement degree changes. Overall, at the significance level 0.05, both methods suggest evidence of supporting significant effects of BMI and HR regardless of different degrees of measurement error we consider, and the augmented pseudo-likelihood method identifies a lot stronger evidence than the conditional profile-likelihood method does.

As opposed to taking $\Sigma_X = 90\%\widehat{\Sigma}_W$ earlier, we also conduct sensitivity studies for a different value of Σ_X by setting $\Sigma_X = 10\%\widehat{\Sigma}_W$, following the suggestion of a referee. We take $\alpha = 0, 10, 50$, or 100, and let R take a value in [0.6, 0.9]. The results are report in Table S.7 and Appendix F.2 of the Supplementary Material, which reveal patterns similar to those displayed in Table 1.

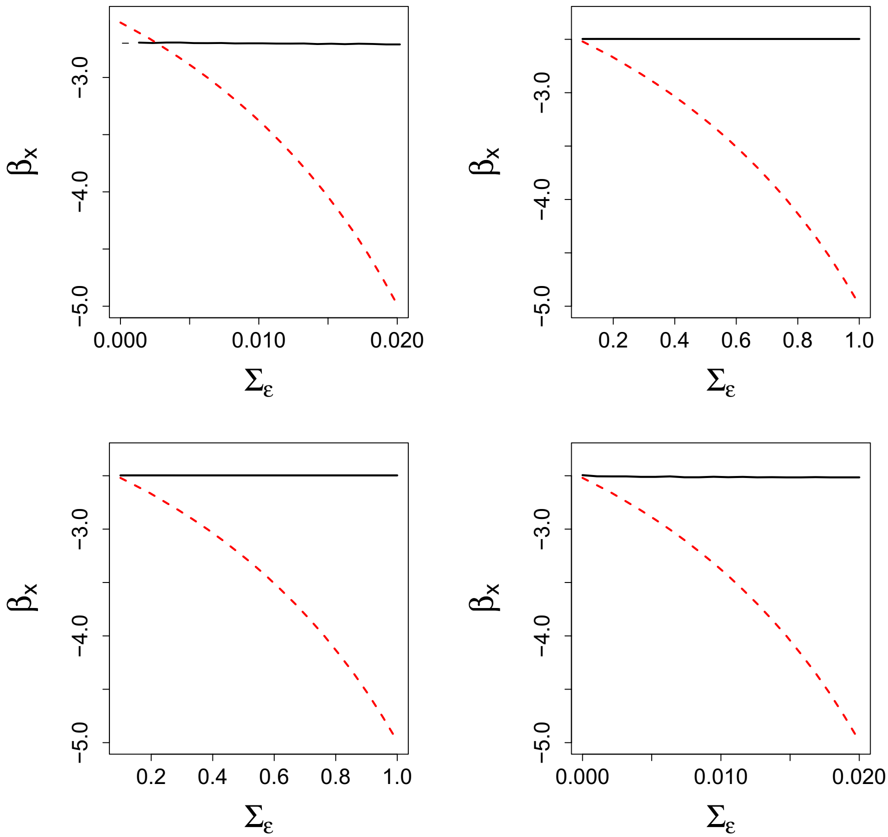

Fig. 2 Sensitivity analysis for the estimators of β_x, $\widehat{\beta}_x$ and $\widetilde{\beta}_x$, against different degrees of measurement error: solid curves are for $\widetilde{\beta}_x$ (derived from (23)) and dashed curves are for $\widehat{\beta}_x$ (obtained from (14)); in the first row, the left panel is for $\alpha = 0$ and the right panel is for $\alpha = 10$; in the second row, the left panel is for $\alpha = 50$ and the right panel is for $\alpha = 100$

7 Discussion

Although survival analysis has proven useful and many methods have been developed for analyzing survival data with individual features, there has been little work of addressing these features simultaneously in inferential procedures, as noted by Yi and Lawless (2007). In this article, we develop two estimation methods to handle left-truncated and right-censored survival data with measurement error in covariates. We establish asymptotic results for the proposed methods rigorously and explore the issues of efficiency of the proposed methods. We demonstrate satisfactory finite sample performance of our methods using simulation studies.

The proposed methods can also accommodate length-biased survival data with covariate measurement error. Length-biased data arise commonly from many fields including epidemiological studies, cancer research, and etiology studies, and many

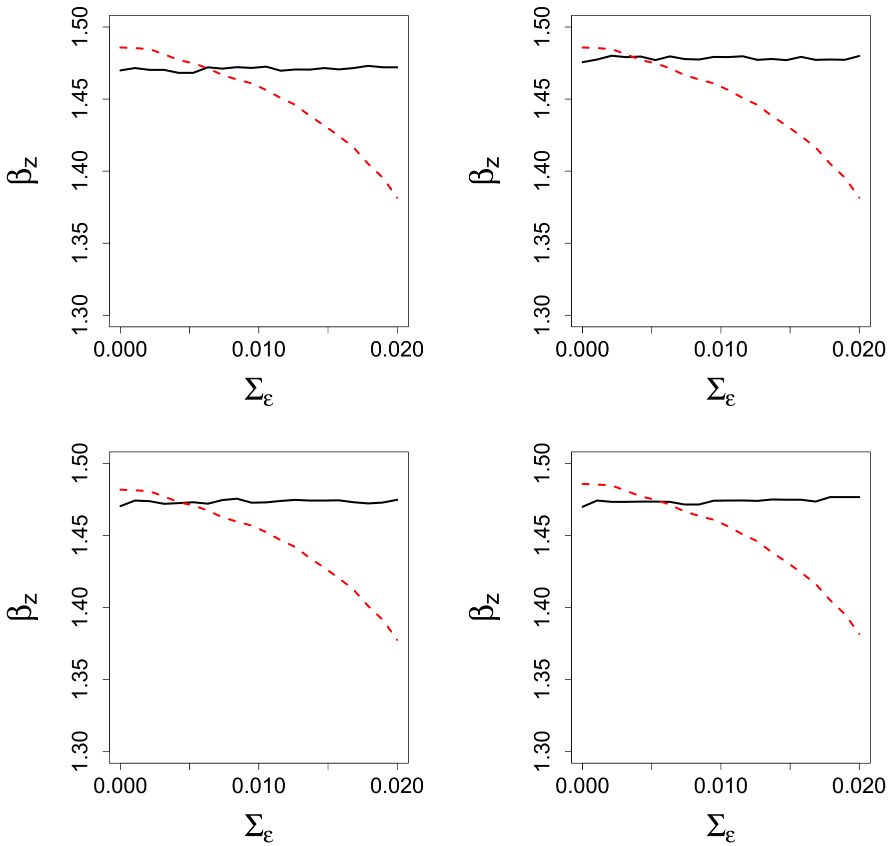

Fig. 3 Sensitivity analysis for the estimators of β_z, $\widehat{\beta}_z$ and $\widetilde{\beta}_z$, against different degrees of measurement error: solid curves are for $\widetilde{\beta}_z$ [derived from (23)] and dashed curves are for $\widehat{\beta}_z$ [obtained from (14)]; in the first row, the left panel is for $\alpha = 0$ and the right panel is for $\alpha = 10$; in the second row, the left panel is for $\alpha = 50$ and the right panel is for $\alpha = 100$

methods have been developed for analysis of such data. However, the validity of these methods is limited due to the key assumption that data must be accurately collected. In application, measurements of the variables usually are error-contaminated. Accommodating the feature of measurement error, our proposed methods generalize the scope of usual methods of handling length-biased survival data. Detailed discussions are given in Appendix A of the Supplementary Material.

We comment that in the development here, we consider only the measurement error processes which can be feasibly characterized by model (6), bearing in mind more general models, such as regression models with various forms of covariates, may be used to feature different types of measurement error processes. In applications, if there is an internal validation sample with measurements of W and X together with those of Z, Y and Δ, we may develop a model diagnostic procedure to assess the feasibility of model (6). Without such data, it is generally impossible to

Table 5 Sensitivity analyses of the Worcester Heart Attack Study Data

α	Σ_ϵ	parameter	$\widehat{\beta}=\left(\widehat{\beta}_x,\widehat{\beta}_z\right)^\top$			$\widetilde{\beta}=\left(\widetilde{\beta}_x,\widetilde{\beta}_z\right)^\top$		
			EST	SEM	p-value	EST	SEM	p-value
0	0.000	β_x	− 2.520	0.867	0.004	− 2.494	0.473	1.344e−07
		β_z	1.486	0.659	0.024	1.470	0.311	2.282e−06
	0.004	β_x	− 2.807	1.153	0.022	− 2.499	0.561	8.763e−06
		β_z	1.478	0.671	0.028	1.470	0.471	0.002
	0.010	β_x	− 3.319	1.542	0.031	− 2.501	0.543	4.106e−06
		β_z	1.461	0.699	0.036	1.472	0.456	0.001
	0.018	β_x	− 4.544	1.859	0.015	-2.507	0.575	1.301e−05
		β_z	1.405	0.716	0.049	1.473	0.529	0.005
10	0.000	β_x	− 2.520	0.867	0.004	− 2.494	0.473	1.344e−07
		β_z	1.486	0.659	0.024	1.470	0.311	2.282e−06
	0.004	β_x	− 2.823	1.152	0.014	− 2.507	0.565	9.115e−06
		β_z	1.487	0.677	0.028	1.497	0.478	0.002
	0.010	β_x	− 3.319	1.543	0.031	− 2.515	0.551	5.009e−06
		β_z	1.471	0.694	0.034	1.476	0.493	0.003
	0.018	β_x	− 4.544	1.866	0.014	− 2.516	0.578	1.343e−05
		β_z	1.415	0.713	0.047	1.478	0.521	0.005
50	0.000	β_x	− 2.520	0.867	0.004	− 2.494	0.473	1.344e−07
		β_z	1.486	0.659	0.024	1.470	0.311	2.282e−06
	0.004	β_x	− 2.823	1.155	0.014	− 2.503	0.566	9.767e−06
		β_z	1.497	0.673	0.026	1.476	0.483	0.002
	0.010	β_x	− 3.319	1.533	0.030	− 2.508	0.569	1.044e−05
		β_z	1.481	0.693	0.032	1.476	0.514	0.004
	0.018	β_x	− 4.544	1.833	0.013	− 2.513	0.577	1.329e−05
		β_z	1.425	0.711	0.045	1.477	0.523	0.005
100	0.000	β_x	− 2.520	0.867	0.004	− 2.494	0.473	1.344e−07
		β_z	1.486	0.659	0.024	1.470	0.311	2.282e−06
	0.004	β_x	− 2.807	1.150	0.022	− 2.506	0.569	1.061e−05
		β_z	1.478	0.674	0.027	1.473	0.498	0.003
	0.010	β_x	− 3.319	1.441	0.021	− 2.510	0.558	6.852e−06
		β_z	1.461	0.684	0.033	1.474	0.502	0.003
	0.018	β_x	− 4.544	1.776	0.011	− 2.515	0.572	1.098e−05
		β_z	1.405	0.709	0.047	1.477	0.504	0.003

assess the validity of model (6), which is just like the situation we commonly face with statistical modeling when handling noisy data with measurement error.

Finally, we comment on the differences between our "measurement error models" and "linear mixed effects models" considered by Su and Wang (2012), which were questioned by a referee. While it might be an angle to view the additive measurement error model to be a special case of the linear mixed effects model used by Su and Wang (2012), this perspective is often not being taken in the literature due to several reasons. First, in joint modeling of longitudinal covariates and survival outcomes, latent variables are commonly used to combine the model for the longitudinal process and the model for the survival process, and accordingly, inferences are derived from taking integration with respect to the latent variables, which is essentially a likelihood-based method. Such a joint modeling scheme essentially puts longitudinal covariates and survival outcomes on equal footing.

On the contrary, in handling measurement error problems, our emphasis is to use covariates to explain the outcome variable by using a regression model in terms of the true covariates X. Since X may not be precisely measured, but its observed version or the so-called "surrogate covariate" W can be collected, our inferences would be based on using measurement W with suitable adjustments to facilitate the possible differences between X and W. The key difficulties here are to develop a proper adjustment to fit each specific model for the response process as well as the measurement error process, and the likelihood-based methods are not the only approach.

A second noticeable difference lies in the interpretation and nature of the variables. Latent variables are random variables which *can never* be observed; their behavior is mainly featured by an assumed distribution which cannot be testified. On the other hand, for the problems with measurement error, although the true covariate X may not be observed for *every* subject in the study, it is possible to obtain the true value of X in situations where validation data are available. In addition, X does not have to be always taken as a random variable and its distribution does not have to be specified when conducting inferences (Carroll et al. 2006; Yi 2017).

Acknowledgements The authors thank the review team for the comments on the initial submission. This research was supported by the Natural Sciences and Engineering Research Council of Canada and partially supported by a Collaborative Research Team Project of the Canadian Statistical Sciences Institute. Yi is Canada Research Chair in Data Science (Tier 1). Her research was undertaken, in part, thanks to funding from the Canada Research Chairs program.

Appendix: Regularity conditions

Like any other asymptotic results, the validity of our results requires regularity conditions imposed on the processes of survival, censoring, measurement error and covariates as well as the sampling scheme. Basically, our regularity conditions pertain to those in Andersen and Gill (1982), Huang et al. (2012), and Yan and Yi (2016), including the following assumptions:

(C1) Θ is a compact set, and the true parameter value β_0 is an interior point of Θ.
(C2) $\int_0^\tau \lambda_0(t)\mathrm{d}t < \infty$, where τ is the finite maximum support of the failure time.
(C3) The $\left\{N_i(t), Y_i(t), Z_i, X_i\right\}$ are independent and identically distributed for $i = 1, \ldots, n$.
(C4) The covariates Z_i and X_i are bounded.
(C5) Conditional on V_i^*, $\left(T_i^*, V_i^*\right)$ are independent of A_i^*.
(C6) Censoring time C_i is non-informative. That is, the failure time T_i and the censoring time C_i are independent, given the covariates $\{Z_i, X_i\}$.

(C7) Matrices $E\left(-\frac{1}{n}\frac{\partial^2 \ell_C^*}{\partial\beta\partial\beta^\top}\right)$ and $E\left(-\frac{1}{n}\frac{\partial^2 \ell_M^*}{\partial\beta\partial\beta^\top}\right)$ are positive definite, where ℓ_C^* is defined in (10) and ℓ_M^* is the logarithm of the likelihood function (19).

(C8) The operations of differentiation and integration are exchangeable.

Condition (C1) is a basic condition that is used to derive the maximizer of the target function (e.g., Huang et al. 2012, p.203). (C2) to (C6) are standard conditions for survival analysis, which allow us to obtain the sum of independent and identically distributed random variables and hence to derive the asymptotic properties of the estimators (e.g., Andersen and Gill 1982). The requirement of positive definite matrices in Condition (C7) is standard which ensures asymptotic covariance matrices of ℓ_C^* and ℓ_M^* meaningful. Condition (C8) is a routine requirement for deriving asymptotic results.

Lemma 1 *Let*

$$\widehat{\ell}_P^* = \sum_{i=1}^n \int_0^\tau \left[\widetilde{v}_i^\top \beta + \frac{1}{2}\beta_x^\top \Sigma_\epsilon \beta_x - \log\left\{\sum_{j=1}^n \exp(\widetilde{v}_j^\top \beta) I(a_j \le u \le y_j)\right\}\right] \mathrm{d}N_i(u). \tag{40}$$

Then (10) *and* (40) *yield the same maximum likelihood estimator of* β.

The proof is given in Appendix B of the Supplementary Material. The following lemma is used to establish the consistency of the estimators $\widehat{\beta}$ and $\widetilde{\beta}$, respectively, given in Theorems 2 and 3.

Lemma 2 *Define*

$$\kappa_P = E\left(\frac{1}{n}\widehat{\ell}_P^*\right)$$

and let

$$\kappa = \kappa_P + E\left\{\frac{1}{n}\log\left(L_M^*\right)\right\},$$

where $\widehat{\ell}_P^*$ *and* L_M^* *are determined by* (40) *and* (19), *respectively, with the data* $\{\widetilde{v}_i, a_i, y_i, z_i\}$ *replaced by the corresponding random variables* $\{\widetilde{V}_i, A_i, Y_i, Z_i\}$. *Then* β_0 *is the unique maximizer of* κ_P *and* κ.

Proof

Part 1: We show that β_0 is the unique maximizer of κ_P.

Recall that ℓ_C is the logarithm of the likelihood function (2) based on the true covariates X. In the absence of measurement error, i.e., based on the true covariates X, Huang et al. (2012, p.208) showed that the true value β_0 is the unique maximizer of $E(\ell_C)$. Noting that by (9), ℓ_C and ℓ_C^*, defined in (10), have the relationship

$$E(\ell_C^*) = E(\ell_C). \tag{41}$$

We conclude that β_0 is also the unique maximizer of $E(\ell_C^*)$. By Lemma 1, we conclude that β_0 is the unique maximizer of κ_P. With regularity conditions including (C8),

$$\beta_0 \text{ is the unique solution of } E\left(\frac{1}{n}\frac{\partial \widehat{\ell}_P^*}{\partial \beta}\right) = 0, \tag{42}$$

and

$$\left.\frac{\partial^2 \kappa_P}{\partial \beta \partial \beta^\top}\right|_{\beta=\beta_0} \text{ is negative definite.} \tag{43}$$

<u>Part 2:</u> We show that β_0 is the unique maximizer of κ.

Let $\ell_M(\beta;X,Z)$ denote the logarithm of the likelihood function (3) based on the true covariates X, and let $\ell_M(\beta;\widetilde{X},Z)$ be $\ell_M(\beta;X,Z)$ with X replaced by $\widetilde{X} = E(X|W^*)$, where $W^* = W - \Sigma_\epsilon \alpha$ as defined before (7). Define $U_M(\beta;X,Z) = \frac{\partial \ell_M(\beta;X,Z)}{\partial \beta}$ and let $\mathcal{U}_M(\beta;X,Z) = E\left\{\frac{1}{n}U_M(\beta;X,Z)\right\}$.

Recall that $\mu_X = E(X)$ defined before (17), then by (7), we have that $E(\widetilde{X}) = \mu_X$. Let $\mu_Z = E(Z)$. Then by the linear approximation around μ_X and μ_Z, we express $U_M(\beta;\widetilde{X},Z)$ and $U_M(\beta;X,Z)$, respectively, as

$$\begin{aligned} U_M(\beta;\widetilde{X},Z) \approx U_M(\beta;\mu_X,\mu_Z) + \frac{\partial U_M(\beta;\mu_X,\mu_Z)}{\partial \mu_X}(\widetilde{X} - \mu_X) \\ + \frac{\partial U_M(\beta;\mu_X,\mu_Z)}{\partial \mu_Z}(Z - \mu_Z) \end{aligned} \tag{44}$$

and

$$\begin{aligned} U_M(\beta;X,Z) \approx U_M(\beta;\mu_X,\mu_Z) + \frac{\partial U_M(\beta;\mu_X,\mu_Z)}{\partial \mu_X}(X - \mu_X) \\ + \frac{\partial U_M(\beta;\mu_X,\mu_Z)}{\partial \mu_Z}(Z - \mu_Z), \end{aligned} \tag{45}$$

where $\frac{\partial U_M(\beta;\mu_X,\mu_Z)}{\partial \mu_X}$ represents the partial derivative $\frac{\partial U_M(\beta;a,b)}{\partial a}$ evaluated at $(a,b) = (\mu_X,\mu_Z)$, and $\frac{\partial U_M(\beta;\mu_X,\mu_Z)}{\partial \mu_Z}$ represents the partial derivative $\frac{\partial U_M(\beta;a,b)}{\partial b}$ evaluated at $(a,b) = (\mu_X,\mu_Z)$. Here $U_M(\beta;a,b)$ has the same functional form as $U_M(\beta;X,Z)$ except that the former is a real-valued function with arguments β, a and b, while the latter case is a function of random variables X and Z together with β.

Combining (44) and (45) gives that

$$U_M(\beta;\widetilde{X},Z) \approx U_M(\beta;X,Z) + \frac{\partial U_M(\beta;\mu_X,\mu_Z)}{\partial \mu_X}(\widetilde{X} - X). \tag{46}$$

Therefore, taking expectation on both sides of (46) and replacing β by β_0 give

$$\mathcal{U}_M(\beta_0;\widetilde{X},Z) \approx 0 \tag{47}$$

because that $E(\widetilde{X}) - E(X) = \mu_X - \mu_X = 0$ and $\mathcal{U}_M(\beta_0;X,Z) = 0$ (e.g., Huang et al. 2012, p.208).

By (42) and (47),

$$E\left(\frac{1}{n}\frac{\partial \widehat{\ell}_P^*}{\partial \beta}\bigg|_{\beta=\beta_0}\right) + \mathcal{U}_M(\beta_0;\widetilde{X},Z) \approx 0. \tag{48}$$

By definition of $U_M(\beta;X,Z)$ and (19) together with (17), we have $U_M(\beta;\widetilde{X},Z) = \frac{\partial \log(L_M^*)}{\partial \beta}$, and thus, $\mathcal{U}_M(\beta;\widetilde{X},Z) = E\left\{\frac{1}{n}\frac{\partial \log(L_M^*)}{\partial \beta}\right\}$ and $\frac{\partial \mathcal{U}_M(\beta;\widetilde{X},Z)}{\partial \beta} = E\left\{\frac{1}{n}\frac{\partial^2 \log(L_M^*)}{\partial \beta \partial \beta^\top}\right\}$. Then applying (48) gives

$$\frac{\partial \kappa}{\partial \beta}\bigg|_{\beta=\beta_0} \approx 0. \tag{49}$$

Next, by taking expectation on (46) and then taking the partial derivative with respect to β give that

$$\frac{\partial \mathcal{U}_M(\beta;\widetilde{X},Z)}{\partial \beta} \approx \frac{\partial \mathcal{U}_M(\beta;X,Z)}{\partial \beta}. \tag{50}$$

By the derivations similar to Huang et al. (2012, p.208), $\frac{\partial \mathcal{U}_M(\beta;X,Z)}{\partial \beta}$ is negative definite at $\beta = \beta_0$, and thus, by (50), $\frac{\partial \mathcal{U}_M(\beta;\widetilde{X},Z)}{\partial \beta}$ is also negative definite at $\beta = \beta_0$. Then combining with (43) gives that $\frac{\partial^2 \kappa}{\partial \beta \partial \beta^\top} = \frac{\partial^2 \kappa_P}{\partial \beta \partial \beta^\top} + E\left\{\frac{1}{n}\frac{\partial^2 \log(L_M^*)}{\partial \beta \partial \beta^\top}\right\}$ is negative definite at $\beta = \beta_0$. Therefore, combining with (49), we conclude that β_0 is approximately the maximizer of κ. $\square$

References

Andersen, P. K., Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, *10*, 1100–1120.

Augustin, T. (2004). An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions. *Scandinavian Journal of Statistics*, *31*, 43–50.

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

Carroll, R. J., Li, K.-C. (1992). Measurement error regression with unknown link: Dimension reduction and data visualization. *Journal of the American Statistical Association*, *87*, 1040–1050.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). *Measurement error in nonlinear model*. New York: Chapman & Hall/CRC.

Goutis, C., Casella, G. (1999). Explaining the saddlepoint approximation. *The American Statistician*, *53*, 216–224.

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

Henmi, M., Eguchi, S. (2004). A paradox concerning nuisance parameters and projected estimating functions. *Biometrika*, *91*, 929–941.
Hosmer, D. W., Lemeshow, S., May, S. (2008). *Applied survival analysis*: *Regression modeling of time to event data*. New York: Wiley.
Hu, C., Lin, D. Y. (2002). Cox regression with covariate measurement error. *Scandnavian Journal of Statistics*, *29*, 637–655.
Huang, C. Y., Qin, J., Follmann, D. A. (2012). A maximum pseudo-profile likelihood estimator for the Cox model under length-biased sampling. *Biometrika*, *99*, 199–210.
Huang, Y., Wang, C. Y. (2000). Cox regression with accurate covariates unascertainable: A nonparametric correction approach. *Journal of the American Statistical Association*, *95*, 1209–1219.
Kalbfleisch, J. D., Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
Kong, F. H., Gu, M. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statistica Sinica*, *9*, 953–969.
Küchenoff, H., Bender, R., Langner, I. (2007). Effect of Berkson measurement error on parameter estimates in Cox regression models. *Lifetime Data Analysis*, *13*, 261–272.
Lawless, J. F. (2003). *Statistical models and methods for lifetime data*. New York: Wiley.
Li, Y., Ryan, L. (2006). Inference on survival data with covariate measurement error: An imputation-based approach. *Scandinavian Journal of Statistics*, *33*, 169–190.
Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics*, *48*, 829–838.
Ning, Y., Yi, G. Y., Reid, N. (2018). A class of weighted estimating equations for semiparametric transformation models with missing covariates. *Scandinavian Journal of Statistics*, *45*, 87–109.
Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, *69*, 331–342.
Qin, J., Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics*, *66*, 382–392.
Qin, J., Ning, J., Liu, H., Shen, Y. (2011). Maximum likelihood estimations and EM algorithms with length-biased data. *Journal of the American Statistical Association*, *106*, 1434–1449.
Robins, J. M., Rotnitzky, A., Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, *89*, 846–866.
Rothman, K. J. (2008). BMI-related errors in the measurement of obesity. *International Journal of Obesity*, *32*, 56–59.
Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivative. *The Annals of Statistics*, *6*, 177–184.
Song, X., Huang, Y. (2005). On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics*, *61*, 702–714.
Su, Y., Wang, J. (2012). Modeling left-truncated and right-censored survival data with longitudinal covariates. *The Annals of Statistics*, *40*, 1465–1488.
van der Vaart, A. W. (1998). *Asymptotic statistics*. New York: Cambridge University Press.
Wang, C. Y. (1999). Robust sandwich covariance estimation for regression calibration estimator in Cox regression with measurement error. *Statistics & Probability Letters*, *45*, 371–378.
Wang, C. Y. (2000). Flexible regression calibration for covariate measurement error with longitudinal surrogate variables. *Statistica Sinica*, *10*, 905–921.
Wang, M. C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*, *86*, 130–143.
Wu, F., Kim, S., Qin, J., Saran, R., Li, Y. (2018). A pairwise likelihood augmented Cox estimator for left-truncated data. *Biometrics*, *74*, 100–108.
Xie, S. H., Wang, C. Y., Prentice, R. L. (2001). A risk set calibration method for failure time regression by using a covariate reliability sample. *Journal of the Royal Statistical Society, Series B*, *63*, 855–870.
Xu, Y., Kim, J. K., Li, Y. (2017). Semiparametric estimation for measurement error models with validation data. *The Canadian Journal of Statistics*, *45*, 185–201.
Yan, Y., Yi, G. Y. (2015). A corrected profile likelihood method for survival data with covariate measurement error under the Cox model. *The Canadian Journal of Statistics*, *43*, 454–480.
Yan, Y., Yi, G. Y. (2016). A class of functional methods for error-contaminated survival data under additive hazards models with replicate measurements. *Journal of the American Statistical Association*, *111*, 684–695.

Yi, G. Y. (2017). *Statistical analysis with measurement error and misclassication*: *Strategy, method and application*. New York: Springer.
Yi, G. Y., Lawless, J. F. (2007). A corrected likelihood method for the proportional hazards model with covariates subject to measurement error. *Journal of Statistical Planning and Inference*, *137*, 1816–1828.
Yi, G. Y., Ma, Y., Spiegelman, D., Carroll, R. J. (2015). Functional and structural methods with mixed measurement error and misclassification in covariates. *Journal of the American Statistical Association*, *110*, 681–696.
Zhao, S., Prentice, R. L. (2014). Covariate measurement error correction methods in mediation analysis with failure time data. *Biometrics*, *70*, 835–844.