

# Sufficient dimension reduction for survival data analysis with error-prone variables

Li-Pang Chen<sup>1</sup> and Grace Y. Yi<sup>\*2</sup>

<sup>1</sup>*Department of Statistics, National Chengchi University*  
e-mail: [lchen723@nccu.edu.tw](mailto:lchen723@nccu.edu.tw)

<sup>2</sup>*Department of Statistical and Actuarial Sciences, Department of Computer Science, University of Western Ontario*  
e-mail: [gyi5@uwo.ca](mailto:gyi5@uwo.ca)

**Abstract:** Sufficient dimension reduction (SDR) is an important tool in regression analysis which reduces the dimension of covariates without losing predictive information. Several methods have been proposed to handle data with either censoring in the response or measurement error in covariates. However, little research is available to deal with data having these two features simultaneously. In this paper, we examine this problem. We start with considering the cumulative distribution function in regular settings and propose a valid SDR method to incorporate the effects of censored data and covariates measurement error. Theoretical results are established, and numerical studies are reported to assess the performance of the proposed methods.

**Keywords and phrases:** Cross-validation, dimension reduction, error-prone variable, right-censoring, semiparametric estimation.

Received February 2021.

## 1. Introduction

Survival analysis has been proven useful in many areas including cancer research, clinical trials, epidemiological studies, actuarial science, and so on. A primary interest in survival analysis is to study the association between survival times and covariates of interests. In addition to traditionally used survival models such as the Cox proportional hazards model, the accelerated failure time model, the proportional odds model, and the additive hazards model, many parametric or semiparametric models have been proposed in recent years to handle more complex features in survival analysis. Those models include the partially linear hazard regression [3], the partially linear single-index survival model [11], and the partially linear transformation model. However, they may still be inadequate to handle real problems due to the lack of the knowledge of the suitability of a particular model.

---

\*Corresponding author.

Motivated by this, non-parametric regression models are employed in applications to offer the flexibility of modeling and protect against the risk of model misspecification. However, such models are hampered by the high dimension of covariates. To offer a flexible yet parsimonious model formulation, *sufficient dimension reduction* (SDR) becomes useful in reducing the dimension of covariates as well as preserving their predictive information (e.g., [15]).

For uncensored data, various methods have been proposed to reduce the dimension of covariates, including the inverse regression ([16]; [28]), the minimum average variance estimation [24], and the semiparametric framework [19]; some details can be found in [7] and [15]. For right-censored survival data, a number of methods have also been developed for dimension reduction. To name a few, [17] examined the sliced inverse regression method to estimate the central subspace (CS) of dimension reduction directions. [25] considered semiparametric models and proposed the minimum average variance estimation using the inverse censoring weighting scheme. [18] discussed the sliced inverse regression with inverse probability weights and implemented the variable selection approach for sparse data with a large dimension.

While those methods are useful for different settings, they are inapplicable for error-contaminated data, an ubiquitous feature in applications. As noted by [4], when covariates are subject to measurement error, misleading results are often yielded if measurement error effects are ignored when performing sufficient dimension reduction. To address measurement error effects in the SDR framework, [4] proposed the “corrected” covariates for the implementation of sliced inverse regression. [14] established the invariance law for correcting measurement error effects. [27] developed the cumulative slicing estimation method using the “corrected” covariates, which extended the development of [28]. In the presence of both censored data and measurement error in covariates, however, there has been no available work, to the best of our knowledge.

Driven by the lack of methods for handling such data, in this paper we develop estimation methods for handling dimension reduction for censored data with covariate measurement error. We consider the single-index conditional distribution model which covers many useful survival models, and based on them, we develop a semiparametric estimation procedure. Our method does not require the usual assumptions such as linearity and constant variance conditions that are imposed by other authors (e.g., [16]). Our method employs the “corrected” covariates to correct for measurement error effects and applies the conditional expectation scheme to remove the bias caused by censoring.

The remainder is organized as follows. In Section 2, we introduce the framework of right-censored survival data, the single-index conditional distribution model, and the measurement error model. In Section 3, we propose a valid estimation procedure to correct for the measurement error effects, estimate the parameters of interest, and reduce the dimension of covariates. Theoretical results are established in Section 4. Empirical studies, including simulation results and real data analysis, are provided in Section 5. We conclude the article with discussions in Section 6. Technical justifications are included in the appendices.

## 2. Preliminaries and framework

In this section, we introduce the preliminaries of SDR, survival analysis, and measurement error models.

### 2.1. SDR and conditional distribution

Let  $T > 0$  be the univariate response, and let  $X$  be the  $p$ -dimensional vector of covariates having a multivariate normal distribution, where  $p$  is often a large positive integer. The spirit of SDR is to find a  $p \times d$  matrix, say  $B = [\beta_1 \cdots \beta_d]$ , such that

$$T \perp\!\!\!\perp X | B^\top X, \quad (2.1)$$

where “ $\perp\!\!\!\perp$ ” stands for the statistical independence, and  $\beta_j$  is a  $p$ -dimensional column vector for  $j = 1, \dots, d$ . Here  $d$  can be viewed as the dimension of the reduced covariates and is smaller than  $p$ , and  $B$  is often called a basis.

Let  $\mathcal{S}(B)$  represent the SDR subspace which is spanned by the column vectors of  $B$ . [6] showed that the intersection of all such  $\mathcal{S}(B)$  exists. Consequently, such an intersection is called the *central subspace* (CS) for the regression of  $T$  on  $X$ . Let  $\mathcal{S}_{T|X}$  denote the CS with the structural dimension  $d \triangleq \dim(\mathcal{S}_{T|X})$  which is basically unknown.

We now consider the cumulative distribution function of  $T$  given  $X = x$ ,  $F_{T|X}(t|x) \triangleq P(T \leq t | X = x)$ , which is written as

$$F_{T|X}(t|x) = F_0(t, x) \quad (2.2)$$

to show its dependence on both values of  $T$  and  $X$ , where  $F_0(\cdot, \cdot)$  is an unknown nonnegative function.

As discussed in [19] and [24], in the absence of censoring, (2.1) yields that

$$P(T \leq t | X = x) = P(T \leq t | B^\top X = B^\top x)$$

for any  $t \in \mathbb{R}^1$  and  $X \in \mathbb{R}^p$ , suggesting that

$$F_{T|X}(t|x) = F_0(t, B^\top x). \quad (2.3)$$

We consider the setting where the response  $T$  represents the survival time for a subject and is associated with the covariates  $X$ . We are interested in finding the CS,  $\mathcal{S}_{T|X}$ , to study the relationship between the survival time  $T$  and covariates  $X$ . As noted by [19], the basis matrix  $B$  is not unique. To uniquely map CS to a basis matrix, we follow [19] and redefine  $B$  as

$$B = [I_{d \times d} \ D^\top]^\top \quad (2.4)$$

with  $D$  being a  $(p-d) \times d$  matrix having  $(p-d)d$  unknown parameters and  $I_{d \times d}$  being the  $d \times d$  identity matrix. Our objective is to estimate the basis matrix  $D$ . But for ease of exposition, we still refer to  $B$  in the following development.

## 2.2. Survival data with measurement error

In survival analysis,  $T$  is usually incomplete due to the presence of the censoring time for a subject, denoted  $C$ . We assume that given  $X$ ,  $C$  and  $T$  are independent and that  $C$  and  $X$  are independent (e.g., [22]); the assumption are listed as Condition (C2) in Appendix A.

Let  $Y = \min(T, C)$  and  $\Delta = \mathbb{I}(T \leq C)$ , where  $\mathbb{I}(\cdot)$  is the indicator function. Directly implementing the SDR methods to the observed variable  $Y$  and covariates  $X$  is equivalent to studying  $\mathcal{S}_{Y|X}$ , which is generally not equal to  $\mathcal{S}_{T|X}$ , as shown by [18] and [25] that

$$\mathcal{S}_{Y|X} \subseteq \mathcal{S}_{T|X} + \mathcal{S}_{C|X} = \{\mathbf{v}_1 + \mathbf{v}_2 : \mathbf{v}_1 \in \mathcal{S}_{T|X}; \mathbf{v}_2 \in \mathcal{S}_{C|X}\}.$$

On top of the issue of censoring, another challenge is posed by that covariates  $X$  are commonly error-contaminated. To facilitate this feature, let  $X^*$  denote the observed surrogate version of  $X$ . The dimension of  $X^*$ , denoted  $s$ , can be identical to or differ from  $p$ , though in applications,  $s$  is often equal to  $p$ . As in [4], we consider the measurement error model

$$X^* = \gamma + \Gamma X + \epsilon, \quad (2.5)$$

where  $\epsilon$  is independent of  $\{X, T, C\}$ ,  $\gamma$  is an  $s$ -dimensional vector of parameters, and  $\Gamma$  is an  $s \times p$  matrix of parameters which may be known, partially known, or unknown.

As in [14], we assume that  $\epsilon$  follows a normal distribution. Let  $\Sigma_{X^*}$ ,  $\Sigma_X$ , and  $\Sigma_\epsilon$  denote the covariance matrices of  $X^*$ ,  $X$ , and  $\epsilon$ , respectively. As discussed in (2.5) of [4], we consider

$$U = LX^* \quad (2.6)$$

as the “corrected” covariates expressed in terms of  $X^*$ , where

$$L \triangleq \text{cov}(X, X^*)\Sigma_{X^*}^{-1} = \Sigma_X\Gamma^\top\Sigma_{X^*}^{-1}. \quad (2.7)$$

With  $X$  and  $\epsilon$  following a multivariate normal distribution, [14] showed that  $\mathcal{S}_{T|X} = \mathcal{S}_{T|U}$ , suggesting that replacing  $X$  by  $U$  preserves the CS  $\mathcal{S}_{T|X}$ , and hence,

$$T \perp\!\!\!\perp X|B^\top X \text{ is equivalent to } T \perp\!\!\!\perp U|B^\top U. \quad (2.8)$$

## 2.3. Determination of “corrected” covariates

Suppose that we want to collect measurements  $\{\{Y_i, \Delta_i, X_i\} : i = 1, \dots, n\}$  of a sample of  $n$  subjects, where for  $i = 1, \dots, n$ ,  $\{Y_i, \Delta_i, X_i\}$  has the same distribution as  $\{Y, \Delta, X\}$ . Let  $\tau$  denote a finite value typically larger than the maximum observed times in the sample. As  $X_i$  is subject to measurement error, rather than having precise measurement of  $X_i$ , we often observe surrogate measurement  $X_i^*$  (or repeated measurements) for  $i = 1, \dots, n$ , where  $\{X_i^*, X_i\}$  has the same distribution as  $\{X^*, X\}$ .

Note that model (2.5) yields that  $\Sigma_{X^*} = \Gamma \Sigma_X \Gamma^\top + \Sigma_\epsilon$ , and that  $\Sigma_{X^*}$  can be estimated by its empirical estimator based on the available measurements of  $X^*$ . To estimate  $L$ , we need only to handle  $\Sigma_\epsilon$  and  $\Gamma$ . Consequently, we consider the following three scenarios.

**Scenario I** : Both  $\Sigma_\epsilon$  and  $\Gamma$  are known.

In this scenario,  $L$  is determined by (2.7), which allows us to directly calculate the “corrected” covariates  $U$  using (2.6). Such a scenario is useful for conducting sensitivity analyses to understand the impacts of measurement error in  $X$  on SDR.

**Scenario II** :  $\Gamma$  is known,  $\Sigma_\epsilon$  is unknown, and repeated measurements of  $X$  are available.

Suppose that two repeated measurements of  $X$ ,  $\{X_{ir}^* : r = 1, 2; i \in \mathcal{R}\}$ , are collected for additional  $m \triangleq |\mathcal{R}|$  subjects, where  $\mathcal{R}$  denotes the index set for those subjects. Consistent with [4], we take  $\Gamma$  to be  $I_{p \times p}$  for ease of discussions. Then the measurement error model based on repeated measurements is given by

$$X_{ir}^* = \gamma + X_i + \epsilon_{ir} \quad (2.9)$$

for  $i \in \mathcal{R}$  and  $r = 1, 2$ , where  $\epsilon_{ir} \sim N(0, \Sigma_\epsilon)$  and  $\epsilon_{ir}$  is independent of  $\{X_i, T_i, C_i\}$ .

By direct derivations, we obtain that

$$\Sigma_\epsilon = \frac{1}{2} \text{var}(X_{i1}^* - X_{i2}^*) \quad (2.10)$$

and

$$\Sigma_{X^*} = \frac{1}{4} \{\text{var}(X_{i1}^* + X_{i2}^*) - \text{var}(X_{i1}^* - X_{i2}^*)\}, \quad (2.11)$$

so that  $L$  can be expressed as

$$L = (\Sigma_{X^*} - \Sigma_\epsilon) \Sigma_{X^*}^{-1},$$

which is estimated by

$$\widehat{L} = \left( \widehat{\Sigma}_{X^*} - \widehat{\Sigma}_\epsilon \right) \widehat{\Sigma}_{X^*}^{-1} \quad (2.12)$$

with  $\widehat{\Sigma}_\epsilon$  and  $\widehat{\Sigma}_{X^*}$  being empirical estimators of (2.10) and (2.11), respectively.

**Scenario III** : Both  $\Sigma_\epsilon$  and  $\Gamma$  are unknown and validation data are available.

Suppose that  $\mathcal{M}$  is the set of  $n$  subjects for the main study and  $\mathcal{V}$  is the set of  $m$  subjects for the external validation study. That is,  $\mathcal{M}$  and  $\mathcal{V}$  do not overlap, and the available data for the main study and the validation sample are  $\{\{T_i, C_i, \Delta_i, X_i^*\} : i \in \mathcal{M}\}$  and  $\{\{X_i, X_i^*\} : i \in \mathcal{V}\}$ , respectively. Hence, the measurement error model (2.5) gives that

$$X_i^* = \gamma + \Gamma X_i + \epsilon_i$$

for  $i \in \mathcal{M} \cup \mathcal{V}$ , where  $\epsilon_i \sim N(0, \Sigma_\epsilon)$  and  $\epsilon_i$  is independent of  $\{X_i, T_i, C_i\}$  for  $i \in \mathcal{M} \cup \mathcal{V}$ . Here the transportability assumption [5, p. 30] is made for the validation and main study data.

Let  $\mu_X = E(X_i)$  and  $\mu_{X^*} = E(X_i^*)$ . Then using the validation data  $\{\{X_i, X_i^*\} : i \in \mathcal{V}\}$ , we estimate  $\mu_X$  and  $\mu_{X^*}$  by  $\hat{\mu}_X = \frac{1}{m} \sum_{i \in \mathcal{V}} X_i$  and  $\hat{\mu}_{X^*} = \frac{1}{m} \sum_{i \in \mathcal{V}} X_i^*$ , respectively. Then a consistent estimate of  $\text{cov}(X_i, X_i^*)$  is given by

$$\widehat{\text{cov}}(X_i, X_i^*) = \frac{1}{m} \sum_{i \in \mathcal{V}} (X_i - \hat{\mu}_X)(X_i^* - \hat{\mu}_{X^*})^\top,$$

and hence, by (2.7),  $L$  can be estimated by

$$\hat{L} = \widehat{\text{cov}}(X_i, X_i^*) \hat{\Sigma}_{X^*}^{-1}. \quad (2.13)$$

Once the estimator  $\hat{L}$  is obtained by either repeated measurements or validation data in Scenarios II or III, we adjust the surrogate covariate  $X_i^*$  by  $\hat{U}_i = \hat{L}X_i^*$  to accommodate measurement error effects in the following development.

### 3. Methodology

In this section, we consider model (2.3) and propose a method to estimate the basis matrix  $B$  for survival data subject to measurement error in covariates. To be more specific, we first apply (2.6) to correct for the measurement error effects and also use the inverse probability weighting scheme to adjust for the censoring effects due to censored responses; next, we propose valid inferential procedures to estimate  $B$  and  $d$  without imposing additional conditions, such as the linearity condition (e.g., [16]) which is commonly used in the conventional SDR methods.

#### 3.1. Adjustment for measurement error and censoring effects

In the presence of measurement error,  $X$  in (2.3) is often unavailable for estimation of  $B$ . We want to use the observed surrogate measurement  $X^*$  to estimate  $B$  with measurement error effects accounted for. The equivalence (2.8) suggests us to consider SDR based on  $U$ . Analogous to (2.2), we write  $P(T \leq t|U)$ , or  $E\{\mathbb{I}(T \leq t)|U\}$  as

$$E\{\mathbb{I}(T \leq t)|U\} = F(t, u)$$

for a nonnegative unknown function  $F(\cdot, \cdot)$ . By (2.8), we have that

$$E\{\mathbb{I}(T \leq t)|U\} = E\{\mathbb{I}(T \leq t)|B^\top U\} \quad (3.1)$$

for any  $t > 0$ , and hence, yielding that

$$E\{\mathbb{I}(T \leq t)|U\} = F(t, B^\top u). \quad (3.2)$$

The equality (3.2) offers us the basis of estimating  $B$  using the surrogate measurements of  $X_i$  together with the survival times  $T_i$ . In survival analysis, however,  $T_i$  is usually incomplete due to censoring; we have only  $(Y_i, \Delta_i)$  as described in Section 2.1. To accommodate censored responses, a useful strategy is to proceed with the inverse weighted scheme.

For any given  $c > 0$ , let  $G(c) \triangleq P(C \geq c)$ . Then for given  $y > 0$  and  $U = u$ ,

$$\begin{aligned} & E \left\{ \frac{\Delta \mathbb{I}(Y \leq y)}{G(Y)} \middle| B^\top U = B^\top u \right\} \\ = & E \left\{ \frac{\Delta \mathbb{I}(Y \leq y)}{G(Y)} \middle| U = u \right\} \\ = & E \left\{ \frac{\mathbb{I}(T \leq C) \mathbb{I}(Y \leq y)}{G(Y)} \middle| U = u \right\} \\ = & E \left[ E \left\{ \frac{\mathbb{I}(T \leq C) \mathbb{I}(T \leq y)}{G(T)} \middle| T, U = u \right\} \middle| U = u \right] \\ = & E \{ \mathbb{I}(T \leq y) | U = u \} \\ = & F(y, B^\top u), \end{aligned}$$

where the first step is due to (2.8), and the last step comes from (3.2). That is,

$$F(y, B^\top u) = E \left\{ \frac{\Delta \mathbb{I}(Y \leq y)}{G(Y)} \middle| B^\top U \right\} \quad (3.3)$$

for given  $y > 0$  and  $u$ .

The identity (3.3) allows us to estimate  $B$  using the observed time  $Y$  by examining an expectation conditional on the observed surrogate covariate  $X^*$ , where the weight  $\frac{\Delta}{G(Y)}$  is imposed to correct for the censoring effect, and  $U$  adjusts for the measurement error effects.

### 3.2. Estimation procedures

With (3.3), we estimate the function  $F(\cdot, \cdot)$  using the kernel estimation method:

$$\widehat{F}(y, B^\top u) = \frac{\sum_{i=1}^n \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) \mathcal{K}_h(B^\top U_i - B^\top u)}{\sum_{i=1}^n \mathcal{K}_h(B^\top U_i - B^\top u)} \quad (3.4)$$

with  $B$  given by (2.4), where  $\widehat{G}(\cdot)$  is a consistent estimator of  $G(\cdot)$  which can be obtained from the nonparametric Kaplan-Meier estimator (e.g., [22]);  $\mathcal{K}_h(u) = \prod_{j=1}^d \frac{1}{h} K\left(\frac{u_j}{h}\right)$ ;  $h$  is a positive bandwidth to be determined;  $K(v)$  is a user-specified  $q$ th-order kernel function with  $\int K(v)dv = 1$ ,  $\int v^k K(v)dv = 0$  for  $k = 1, \dots, q-1$ ,  $\int v^q K(v)dv < \infty$ ; and  $q$  is a positive integer.

With the estimator of  $F(\cdot, \cdot)$  by (3.4), we use the leave-one-out cross-validation (CV) criterion to calculate the CV value

$$CV(B, d, h) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) - \widehat{F}^{(-i)}(y, B^\top U_i) \right\}^2 d\widehat{F}_Y(y), \quad (3.5)$$

where  $\widehat{F}_Y(\cdot)$  is the empirical distribution function of  $Y_i$

$$\widehat{F}_Y(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(Y_i \leq y)$$

and  $\widehat{F}^{(-i)}(y, B^\top u)$  is the estimator of (3.4) with the  $i$ th subject being deleted. In the degenerate case with  $d = 0$ ,  $B$  is null, and the CV value in (3.5) can be calculated as

$$\frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) - \widehat{F}_0^{(-i)}(y) \right\}^2 d\widehat{F}_Y(y),$$

where  $\widehat{F}_0^{(-i)}(y) = \frac{1}{n-1} \sum_{j \neq i} \frac{\Delta_j}{\widehat{G}(Y_j)} \mathbb{I}(Y_j \leq y)$ .

The estimator of  $(B, d, h)$  can be derived by minimizing (3.5):

$$\left( \widehat{B}, \widehat{d}, \widehat{h} \right) = \underset{B, d, h}{\operatorname{argmin}} CV(B, d, h). \quad (3.6)$$

The implementation of the minimization problem (3.6) may proceed as follows. First, for given  $d \geq 1$ , let  $(\widehat{B}_d, \widehat{h}_d)$  denote the estimators which are obtained by minimizing (3.5):

$$\left( \widehat{B}_d, \widehat{h}_d \right) = \underset{B, h}{\operatorname{argmin}} CV(B, d, h). \quad (3.7)$$

We start with  $d = 1$  and implement (3.7) until  $d = \widehat{d}$  with

$$CV(\widehat{B}_{\widehat{d}+1}, \widehat{d} + 1, \widehat{h}_{\widehat{d}+1}) > CV(\widehat{B}_{\widehat{d}}, \widehat{d}, \widehat{h}_{\widehat{d}}).$$

Then the final estimator is given by  $(\widehat{B}, \widehat{d}, \widehat{h}) = (\widehat{B}_{\widehat{d}}, \widehat{d}, \widehat{h}_{\widehat{d}})$ .

#### 4. Theoretical results

In this section, we present theoretical results of the proposed method, and the proofs of the theoretical results are deferred to Appendix C.

Let  $\operatorname{vec}(\cdot)$  denote the vectorization operation that stacks the columns of a matrix from left to right, and let  $\|\cdot\|$  represent the Frobenius norm of a matrix. Define  $a^{\otimes 2} = aa^\top$  for any vector  $a$ . To emphasize the involvement of the estimator  $\widehat{L}$ , we write  $\widehat{U}_i = \widehat{L}X_i^*$  as defined in Section 2.3. For a differentiable function  $f(\alpha)$ , let  $\nabla_\alpha^j f(\alpha)$  denote the  $j$ th order derivative of the function  $f(\alpha)$  with respect to  $\alpha$ . Let  $B_0$  and  $d_0$  denote the true values of the parameter and



its structural dimension, respectively. Let  $h_0$  be the optimal bandwidth. We first present the consistency of the estimators  $(\widehat{B}, \widehat{d}, \widehat{h})$  whose proof is placed in Appendix C.1.

**Theorem 4.1.** *Under regularity conditions in Appendix A, for any  $\eta > 0$ , as  $n \rightarrow \infty$ ,*

$$\widehat{B} \xrightarrow{p} B_0 \text{ and } P \left( \widehat{d} = d_0, \left| \frac{\widehat{h}}{h_0} - 1 \right| < \eta \right) \rightarrow 1.$$

Theorem 4.1 says that the estimator  $\widehat{B}$  converges in probability to its true value as  $n$  goes to infinity, and both estimated structural dimension and bandwidth are close to the true structural dimension and the optimal bandwidth with probability approaching one, respectively.

For  $l = 0, 1$  and  $j = 0, 1, 2$ , let

$$\widehat{\mathbb{F}}_{l,B,L}^{(j)}(y, u) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) \right\}^l \nabla_{\text{vec}(B)}^j \mathcal{K}(B^\top U_i - u) \quad (4.1)$$

and

$$\widehat{\mathbb{F}}_{l,B,\widehat{L}}^{(j)}(y, u) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) \right\}^l \nabla_{\text{vec}(B)}^j \mathcal{K}(B^\top \widehat{U}_i - u). \quad (4.2)$$

Furthermore, let  $f_{B^\top U}(B^\top u)$  denote the density function of  $B^\top U$ , and define

$$\begin{aligned} & \mathbb{F}_{l,B,L}^{(j)}(y, u) \\ &= \nabla_{\text{vec}(B)}^j \left[ \{F(y, B^\top u)\}^l \right] E \left\{ (U_i - u)^{\otimes j} \mid B^\top U_i = B^\top u \right\} f_{B^\top U}(B^\top u), \end{aligned} \quad (4.3)$$

and

$$\widetilde{\mathbb{F}}_{i,l,B,L}^{(j)}(y, u) = \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) \right\}^l \nabla_{\text{vec}(B)}^j \mathcal{K}(B^\top U_i - u) - \mathbb{F}_{l,B,L}^{(j)}(y, u).$$

Define

$$\zeta_{i,B_0}^{(0)}(y, u) = \sum_{l=0}^1 \widetilde{\mathbb{F}}_{i,l,B,L}^{(0)}(y, u) \frac{\{-F(y, B_0^\top u)\}^{1-l}}{\mathbb{F}_{0,B,L}^{(0)}(y, B_0^\top u)}. \quad (4.4)$$

We now present the theoretical result for (3.4) with  $B$  replaced by its estimator, and the proof is deferred to Appendix C.2.

**Theorem 4.2.** *Under regularity conditions in Appendix A, then*

$$\sup_{y,u} \left| \widehat{F} \left( y, \widehat{B}^\top u \right) - F(y, B_0^\top u) - \frac{1}{n} \sum_{i=1}^n \zeta_{i,B_0}^{(0)}(y, u) \right| = O_p \left( \frac{1}{\sqrt{n}} \right).$$

Theorem 4.2 indicates that the difference between the estimated function and the true function is of order  $O_p(n^{-1/2})$  if the biased term (i.e., the third term in the left-hand-side of the identity) is removed.

Finally, we present the asymptotic distribution of the estimator  $\widehat{B}$ . Define

$$F^{(0)}(y, u) = \frac{\mathbb{F}_{1,B,L}^{(1)}(y, u)}{\mathbb{F}_{0,B,L}^{(0)}(y, u)}, \tag{4.5}$$

$$F^{(1)}(y, u) = \sum_{l=0}^1 \frac{\{-F(y, u)\}^l \mathbb{F}_{1-l,B,L}^{(1)}(y, u)}{\mathbb{F}_{0,B,L}^{(0)}(y, u)}, \tag{4.6}$$

and

$$F^{(2)}(y, u) = \sum_{l_1=0}^1 \sum_{l_2=0}^1 2^{l_1} \left\{ \frac{-\mathbb{F}_{0,B,L}^{(2-l_1)}(y, u)}{\mathbb{F}_{0,B,L}^{(0)}(y, u)} \right\}^{l_1+l_2} \left( \frac{\mathbb{F}_{1,B,L}^{(\{2-l_1\}\{1-l_2\})}(y, u)}{\mathbb{F}_{0,B,L}^{(0)}(y, u)} \right).$$

Let

$$U(B_0) = \int_0^\tau \left\{ \frac{\Delta_i}{G(Y_i)} \mathbb{I}(Y_i \leq y) - F^{(0)}(y, B_0^\top U_i) \right\} F^{(1)}(y, B_0^\top U_i) dF_Y(y)$$

and

$$\begin{aligned} \mathcal{A} = & 2E \left( \int_0^\tau \left[ \left\{ F^{(1)}(y, B_0^\top U_i) \right\}^{\otimes 2} \right. \right. \\ & \left. \left. - F^{(2)}(y, B_0^\top U_i) \left\{ \frac{\Delta_i}{G(Y_i)} \mathbb{I}(Y_i \leq y) - F^{(0)}(y, B_0^\top U_i) \right\} \right] dF_Y(y) \right). \end{aligned}$$

We define

$$\mathcal{T}(B_0) = E \left[ \int_0^\tau \left\{ F^{(1)}(y, B_0^\top U_i) \right\}^{\otimes 2} dF_Y(y) \right] \Sigma_{X^*}.$$

**Theorem 4.3.** *Suppose that regularity conditions in Appendix A hold.*

(a) *Assume that  $L$  is known, then as  $n \rightarrow \infty$ ,*

$$\sqrt{n} \left\{ \text{vec}(\widehat{B}) - \text{vec}(B_0) \right\} \xrightarrow{d} N(0, \mathcal{A}^{-1} \mathcal{B} \mathcal{A}^{-1}),$$

where  $\mathcal{B} = E \{ U^{\otimes 2}(B_0) \}$ .

(b) *Assume that  $L$  is unknown and estimated based on either repeated measurements or validation data. Let  $\Phi_i$  be*

$$\left\{ (X_{i1}^* - X_{i2}^*) (X_{i1}^* - X_{i2}^*)^\top - 2\Sigma_\epsilon \right\}$$

if  $L$  is estimated based on repeated measurements, and let  $\Phi_i$  be

$$\left\{ (X_i - \mu_X) (X_i^* - \mu_{X^*})^\top - \Sigma_{XX^*} \right\}$$

if  $L$  is estimated from validation data. Then as  $n \rightarrow \infty$ ,

$$\sqrt{n} \left\{ \text{vec}(\widehat{B}) - \text{vec}(B_0) \right\} \xrightarrow{d} N(0, \mathcal{A}^{-1} \mathcal{B}_L \mathcal{A}^{-1}),$$

where  $\mathcal{B}_L = E \left[ \{ U(B_0) + \mathcal{T}(B_0) \Phi_i \}^{\otimes 2} \right]$ .

Theorem 4.3 shows the asymptotic normality of the proposed estimator  $\widehat{B}$  under three different scenarios in Section 2.3. In particular, when  $\Sigma_\epsilon$  is unknown and additional information is available, there exist an additional term  $\mathcal{T}(B_0)\Phi_i$  in the asymptotic variance.

## 5. Numerical studies

In this section, we conduct simulation studies to assess the performance of the proposed estimators for a variety of settings. We first design the simulation settings and then present the simulation results. Finally, the methods are implemented to analyze a real dataset.

### 5.1. Simulation studies

#### 5.1.1. Simulation design

Let  $B_0$  be the true  $p \times d_0$  matrix. The  $p$ -dimensional covariates  $X$  is generated from the multivariate normal distribution  $N(0, \Sigma_X)$ , where  $\Sigma_X$  is the covariance matrix with diagonal entries being one and off-diagonal entries being 0.4. Given the covariates  $X$  and  $B_0$ , we use three models, the proportional hazards (PH), proportional odds (PO), and additive hazards (AH) models, to generate survival times. Specifically, the corresponding cumulative distribution functions  $F(\cdot, \cdot)$  are formulated, respectively, as

$$\begin{aligned} F_{PH}(t, B_0^\top X) &= 1 - \exp[-t^2 \exp\{\varphi(X, B_0)\}], \\ F_{PO}(t, B_0^\top X) &= \frac{\exp[t - \{\varphi(X, B_0)\}]}{1 + \exp[t - \{\varphi(X, B_0)\}]}, \end{aligned}$$

and

$$F_{AH}(t, B_0^\top X) = 1 - \exp[-\{t^2 + t\{\varphi(X, B_0)\}\}],$$

where  $\varphi(\cdot, \cdot)$  is a given function.

Let  $\mathcal{U}$  be generated from the uniform distribution UNIF(0, 1). Then survival times  $T$  based on the PH and PO models can be generated from

$$T = \sqrt{\exp\{-\varphi(X, B_0)\} \log(1 - \mathcal{U})}$$

and

$$T = \log\left\{(1 + \mathcal{U})^{-1} - 1\right\} + \{\varphi(X, B_0)\},$$

respectively, and survival times  $T$  based on the AH model can be obtained by solving the following equation

$$T^2 + T\{\varphi(X, B_0)\} + \log(1 - \mathcal{U}) = 0.$$

We consider the following two settings:

**Setting 1:**  $p = 10$ ;  $d_0 = 2$  with  $B_0 = [\beta_{10} \ \beta_{20}]$  being a  $p \times d_0$  matrix,  $\beta_{10} = (1, 0, 1, 0, 0_{p-4})^\top$  and  $\beta_{20} = (0, 1, 0, 1, 0_{p-4})^\top$ ; and  $\varphi(X, B_0) = (X^\top \beta_{10})^2 + 2(X^\top \beta_{20})$ , where  $0_r$  represents a zero vector of dimension  $r$ .

**Setting 2:**  $p = 30$ ;  $d_0 = 3$  with  $B_0 = [\beta_{10} \ \beta_{20} \ \beta_{30}]$  being a  $p \times d_0$  matrix,  $\beta_{10} = (1, 0, 0, 1, 0_{p-4})^\top$ ,  $\beta_{20} = (0, 1, 0, 0, 1, 0_{p-5})^\top$ , and  $\beta_{30} = (0, 0, 1, 0, 0, 1, 0_{p-6})^\top$ ; and  $\varphi(X, B_0) = (X^\top \beta_{10})^2 + 2(X^\top \beta_{20}) + (X^\top \beta_{30})$ .

Let  $C$  be the censoring time generated from the uniform distribution  $\text{UNIF}(0, c)$ , where  $c$  is a constant that is chosen to yield about 50% censoring. Consequently, we calculate  $Y = \min\{T, C\}$  and  $\Delta = \mathbb{I}(T \leq C)$ .

We consider three scenarios to generate surrogate measurements of  $X$  and estimate  $B$  using the methods described in Section 2.3 accordingly.

**Scenario 1:** Both  $\Sigma_\epsilon$  and  $\Gamma$  are assumed known.

For  $i = 1, \dots, n$ ,  $X_i^*$  is generated from (2.5) where  $\gamma = 0$ ,  $\Gamma$  is the identity matrix, and  $\epsilon_i$  assumes a distribution, denoted  $f(\epsilon_i)$ . To obtain the estimate  $\hat{B}$  of  $B$ , we employ (3.6) in combination with Scenario I in Section 2.3.

**Scenario 2:**  $\Gamma$  is known,  $\Sigma_\epsilon$  is unknown, and repeated measurements of  $X_i$  are available.

Suppose that two repeated measurements  $\{X_{ir}^* : r = 1, 2; i \in \mathcal{R}\}$  are generated for additional  $m$  subjects with  $m = 100$ , where  $X_i$  and  $\epsilon_{ir}$  are independently generated from  $N(0, \Sigma_X)$  and  $f(\epsilon_i)$ , respectively, and  $X_{ir}^*$  is generated from

$$X_{ir}^* = X_i + \epsilon_{ir}$$

for  $i \in \mathcal{R}$  and  $r = 1, 2$ . To obtain the estimate  $\hat{B}$  of  $B$ , we employ (3.6) in combination with Scenario II in Section 2.3.

**Scenario 3:** Both  $\Sigma_\epsilon$  and  $\Gamma$  are unknown and validation data are available.

For  $i = 1, \dots, n$ , we generate  $\epsilon_i$  from  $f(\epsilon_i)$  independently and set  $X_i^* = X_i + \epsilon_i$ . The main study data then consist of  $\{T_i, C_i, \Delta_i, X_i^*\} : i \in \mathcal{M}$  with  $\mathcal{M} = \{1, \dots, n\}$ . To generate an external validation sample  $\{X_i, X_i^*\} : i \in \mathcal{V}$ , we set  $|\mathcal{V}|$  to be 100, and for  $i \in \mathcal{V}$ , independently generate  $X_i$  from  $N(0, \Sigma_X)$  and  $\epsilon_i$  from  $f(\epsilon_i)$ . Setting  $X_i^* = X_i + \epsilon_i$  for  $i \in \mathcal{V}$  gives the validation data  $\{X_i, X_i^*\} : i \in \mathcal{V}$ . Then we obtain the estimator  $\hat{B}$  of  $B$  using (3.6) in combination with Scenario III in Section 2.3.

We compare the performance of the proposed methods to the *naive estimators* of  $F(\cdot)$  and  $B$ , which are derived by directly implementing the observed covariates  $X_i^*$  in (3.4) and (3.5). As a reference for comparisons, we also use the true values of  $X$  for the estimation, and denote this method as “true”. In implementing our methods, we take the kernel function to be the Epanechnikov function, given by

$$K(v) = \frac{3}{4}(1 - v^2) \text{ for } |v| \leq 1.$$

Consider the sample size  $n = 200, 300$ , or  $400$ . We repeat simulations 500 times for each setting. To assess the accuracy of the estimator  $\hat{B}$  of  $B$ , we

consider the Frobenius norm

$$\|\Delta_B\| = \sqrt{\sum_{j=1}^p \sum_{k=1}^d |\hat{B}_{ij} - B_{0,ij}|^2}$$

for  $\Delta_B = \hat{B} - B_0$ , where  $\hat{B}_{ij}$  and  $B_{0,ij}$  represent the  $(i,j)$  element of  $\hat{B}$  and  $B_0$ , respectively. To examine the performance of the estimator of  $F(\cdot)$ , we consider the mean integrated squared error (MISE)

$$\text{MISE}(\hat{F}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \hat{F}(y, \hat{B}^\top u_i) - F(y, B_0^\top u_i) \right\}^2 d\hat{F}_Y(y).$$

Further, we evaluate the standard error (S.E.) for the estimators of  $B$  and  $F(\cdot)$  which are calculated as the sample standard deviation for those estimates obtained from the 500 simulations.

For each setting, we report four proportions of the estimator  $\hat{d}$  for 500 simulations, given by

$$\frac{1}{500} \sum_{k=1}^{500} \mathbb{I}(A_{kj}) \text{ for } j = 1, 2, 3, 4,$$

where  $A_{k1}$ ,  $A_{k2}$ ,  $A_{k3}$  and  $A_{k4}$  represent  $\hat{d}_k \leq 1$ ,  $\hat{d}_k = 2$ ,  $\hat{d}_k = 3$ , and  $\hat{d}_k \geq 4$ , respectively, with  $\hat{d}_k$  standing for the estimate of  $d$  at the  $k$ th simulation. Those proportions are displayed under the headings  $\hat{d} \leq 1$ ,  $\hat{d} = 2$ ,  $\hat{d} = 3$ , and  $\hat{d} \geq 4$ , respectively.

### 5.1.2. Performance under normally distributed measurement errors

In this subsection we evaluate the performance of the proposed method under the circumstance where the distribution of measurement error term is correctly specified. That is, we set  $f(\epsilon_i)$  to be a normal distribution in the three scenarios of generating surrogate measurements. Specifically, we set  $f(\epsilon_i)$  to be  $N(0, \Sigma_\epsilon)$  with  $\Sigma_\epsilon$  being a diagonal matrix having diagonal entry  $\sigma_\epsilon^2 = 0.15, 0.5$ , or  $0.75$ .

We present the simulation results in the top panels of Tables 1-9. In terms of estimation of  $B$  and  $F(\cdot)$ , the naive method produces biased results, and the finite sample bias increases as the degree of measurement error increases; the proposed methods greatly outperform the naive approach, yielding results that are fairly close to those produced from the reference method by using the true measurements of the covariates. Agreeing with the phenomenon we observed in the literature of measurement error models, the standard errors associated with the proposed methods are larger than those obtained from the naive method, which is the price paid to correct for biases induced from measurement error in covariates. While the differences for estimation of  $h$  and  $d$  are not very striking between the naive method and our proposed methods, our approaches perform better than the naive approach.

Table 1: Simulation results under the PH model: Scenario 1 where  $\Sigma_\epsilon$  and  $\Gamma$  are assumed known.

Distribution <sup>a</sup>	Setting <sup>b</sup>	$\sigma_\epsilon^2$	Method	Estimator of $B$		Estimator of $F(\cdot)$		Estimator of $h$		Estimator of $d$			
				$\ \Delta_B\ $	S.E.	MISE( $F$ )	S.E.	$h$	S.E.	$d \leq 1$	$d = 2$	$d = 3$	$d \geq 4$
Normal	1	0.15	Naive	0.300	0.014	0.188	0.018	0.806	0.015	0.090	0.910	0.000	0.000
			Corrected	0.021	0.023	0.072	0.025	0.802	0.015	0.000	1.000	0.000	0.000
		0.50	Naive	0.315	0.016	0.191	0.021	0.839	0.014	0.096	0.904	0.000	0.000
			Corrected	0.023	0.027	0.073	0.026	0.798	0.015	0.020	0.980	0.000	0.000
		0.75	Naive	0.319	0.019	0.206	0.022	0.824	0.016	0.103	0.897	0.000	0.000
			Corrected	0.033	0.030	0.076	0.034	0.744	0.017	0.019	0.981	0.000	0.000
		True	0.019	0.019	0.062	0.021	0.808	0.016	0.000	1.000	0.000	0.000	
	2	0.15	Naive	0.345	0.017	0.193	0.018	0.987	0.016	0.005	0.085	0.910	0.000
			Corrected	0.021	0.028	0.075	0.025	0.994	0.017	0.000	0.000	1.000	0.000
		0.50	Naive	0.351	0.020	0.198	0.020	0.989	0.016	0.005	0.090	0.905	0.000
			Corrected	0.024	0.029	0.078	0.027	1.002	0.017	0.002	0.012	0.986	0.000
		0.75	Naive	0.356	0.024	0.207	0.026	0.993	0.018	0.008	0.093	0.899	0.000
Corrected			0.035	0.030	0.081	0.038	0.987	0.020	0.003	0.013	0.984	0.000	
	True	0.018	0.019	0.068	0.020	0.978	0.015	0.000	0.000	1.000	0.000		
Uniform	1	0.15	Naive	0.316	0.017	0.189	0.017	0.913	0.018	0.088	0.912	0.000	0.000
			Corrected	0.020	0.025	0.075	0.019	0.946	0.019	0.001	0.999	0.000	0.000
		0.50	Naive	0.322	0.020	0.195	0.023	0.926	0.018	0.090	0.910	0.000	0.000
			Corrected	0.023	0.027	0.079	0.028	0.952	0.019	0.009	0.991	0.000	0.000
		0.75	Naive	0.336	0.024	0.212	0.026	0.934	0.020	0.093	0.907	0.000	0.000
			Corrected	0.035	0.037	0.083	0.037	0.958	0.022	0.014	0.986	0.000	0.000
	2	0.15	Naive	0.351	0.019	0.196	0.019	0.894	0.018	0.000	0.091	0.909	0.000
			Corrected	0.022	0.025	0.077	0.022	0.965	0.020	0.000	0.007	0.993	0.000
		0.50	Naive	0.360	0.023	0.207	0.024	0.899	0.022	0.000	0.095	0.905	0.000
			Corrected	0.028	0.026	0.081	0.027	1.001	0.028	0.000	0.010	0.990	0.000
		0.75	Naive	0.373	0.029	0.224	0.028	0.933	0.025	0.000	0.098	0.902	0.000
			Corrected	0.037	0.038	0.087	0.031	0.978	0.030	0.000	0.015	0.985	0.000

a: "Normal" and "Uniform" correspond to distributions of  $\epsilon_i$  in Sections 5.1.1 and 5.1.2, respectively.b: Setting 1 has  $p = 10$  and  $d_0 = 2$ ; Setting 2 has  $p = 30$  and  $d_0 = 3$ .

Table 2: Simulation results under the PO model: Scenario 1 where  $\Sigma_\epsilon$  and  $\Gamma$  are assumed known.

Distribution <sup>a</sup>	Setting <sup>b</sup>	$\sigma_\epsilon^2$	Method	Estimator of $B$		Estimator of $F(\cdot)$		Estimator of $h$		Estimator of $d$				
				$\ \Delta_B\ $	S.E.	MISE( $F$ )	S.E.	$\hat{h}$	S.E.	$d \leq 1$	$d = 2$	$d = 3$	$d \geq 4$	
Normal	1	0.15	Naive	0.279	0.016	0.197	0.014	0.387	0.017	0.091	0.909	0.000	0.000	
			Corrected	0.016	0.032	0.079	0.016	0.544	0.018	0.000	1.000	0.000	0.000	
		0.50	Naive	0.297	0.017	0.198	0.017	0.409	0.018	0.096	0.904	0.000	0.000	
			Corrected	0.020	0.034	0.086	0.019	0.580	0.017	0.005	0.995	0.000	0.000	
		0.75	Naive	0.305	0.018	0.199	0.012	0.377	0.019	0.096	0.904	0.000	0.000	
			Corrected	0.028	0.036	0.094	0.022	0.576	0.021	0.010	0.990	0.000	0.000	
	2	0.15	True	0.010	0.019	0.068	0.010	0.558	0.018	0.000	1.000	0.000	0.000	
			Naive	0.293	0.020	0.203	0.014	0.974	0.019	0.003	0.090	0.907	0.000	
			Corrected	0.019	0.033	0.081	0.016	0.991	0.025	0.000	0.000	1.000	0.000	
		0.50	Naive	0.304	0.024	0.211	0.017	0.978	0.023	0.006	0.094	0.900	0.000	
			Corrected	0.024	0.036	0.086	0.020	0.880	0.027	0.000	0.003	0.997	0.000	
		0.75	Naive	0.315	0.027	0.223	0.020	0.980	0.025	0.006	0.096	0.894	0.000	
	Corrected		0.030	0.037	0.095	0.024	0.933	0.027	0.000	0.005	0.995	0.000		
	Uniform	1	0.15	Naive	0.282	0.020	0.205	0.015	0.806	0.019	0.089	0.911	0.000	0.000
				Corrected	0.022	0.033	0.086	0.016	0.967	0.022	0.003	0.997	0.000	0.000
			0.50	Naive	0.290	0.024	0.211	0.020	0.814	0.021	0.093	0.907	0.000	0.000
				Corrected	0.024	0.036	0.093	0.025	0.970	0.024	0.010	0.990	0.000	0.000
			0.75	Naive	0.297	0.029	0.223	0.023	0.829	0.023	0.097	0.903	0.000	0.000
Corrected				0.030	0.038	0.096	0.026	0.975	0.026	0.016	0.984	0.000	0.000	
2		0.15	Naive	0.289	0.025	0.212	0.015	0.827	0.022	0.000	0.091	0.909	0.000	
			Corrected	0.025	0.031	0.088	0.017	0.922	0.027	0.000	0.004	0.996	0.000	
		0.50	Naive	0.297	0.029	0.226	0.021	0.827	0.022	0.000	0.096	0.904	0.000	
			Corrected	0.031	0.034	0.095	0.027	0.928	0.029	0.000	0.013	0.987	0.000	
		0.75	Naive	0.304	0.033	0.234	0.025	0.850	0.026	0.000	0.102	0.898	0.000	
			Corrected	0.037	0.039	0.099	0.029	0.916	0.029	0.000	0.018	0.982	0.000	

a: "Normal" and "Uniform" correspond to distributions of  $\epsilon_i$  in Sections 5.1.1 and 5.1.2, respectively.  
b: Setting 1 has  $p = 10$  and  $d_0 = 2$ ; Setting 2 has  $p = 30$  and  $d_0 = 3$ .

Table 3: Simulation results under the AH model: Scenario 1 where  $\Sigma_\epsilon$  and  $\Gamma$  are assumed known.

Distribution <sup>a</sup>	Setting <sup>b</sup>	$\sigma_\epsilon^2$	Method	Estimator of $B$		Estimator of $F(\cdot)$		Estimator of $h$		Estimator of $d$			
				$\ \Delta_B\ $	S.E.	MISE( $F$ )	S.E.	$h$	S.E.	$d \leq 1$	$d = 2$	$d = 3$	$d \geq 4$
Normal	1	0.15	Naive	0.286	0.017	0.184	0.013	0.605	0.020	0.088	0.912	0.000	0.000
			Corrected	0.017	0.028	0.064	0.018	0.743	0.022	0.000	1.000	0.000	0.000
		0.50	Naive	0.295	0.018	0.191	0.016	0.590	0.019	0.092	0.908	0.000	0.000
			Corrected	0.023	0.028	0.061	0.020	0.733	0.027	0.000	1.000	0.000	0.000
		0.75	Naive	0.305	0.018	0.196	0.018	0.568	0.021	0.092	0.908	0.000	0.000
			Corrected	0.028	0.030	0.060	0.023	0.710	0.031	0.009	0.991	0.000	0.000
		True	0.012	0.023	0.037	0.015	0.745	0.017	0.000	1.000	0.000	0.000	
	2	0.15	Naive	0.295	0.023	0.195	0.016	0.855	0.021	0.003	0.085	0.912	0.000
			Corrected	0.017	0.029	0.067	0.020	0.985	0.025	0.000	0.001	0.999	0.000
		0.50	Naive	0.304	0.027	0.203	0.019	0.884	0.026	0.006	0.088	0.906	0.000
			Corrected	0.021	0.032	0.070	0.025	0.987	0.030	0.000	0.004	0.996	0.000
		0.75	Naive	0.312	0.029	0.211	0.023	0.903	0.028	0.006	0.090	0.904	0.000
Corrected			0.029	0.034	0.072	0.028	0.993	0.032	0.000	0.006	0.994	0.000	
	True	0.015	0.022	0.060	0.017	0.749	0.020	0.000	0.002	0.998	0.000		
Uniform	1	0.15	Naive	0.293	0.023	0.190	0.017	0.855	0.019	0.089	0.911	0.000	0.000
			Corrected	0.023	0.026	0.068	0.021	0.989	0.021	0.002	0.998	0.000	0.000
		0.50	Naive	0.298	0.024	0.195	0.019	0.879	0.021	0.094	0.906	0.000	0.000
			Corrected	0.025	0.026	0.069	0.024	0.987	0.025	0.007	0.993	0.000	0.000
		0.75	Naive	0.307	0.027	0.203	0.021	0.894	0.022	0.095	0.905	0.000	0.000
			Corrected	0.030	0.031	0.074	0.025	0.989	0.025	0.010	0.990	0.000	0.000
	2	0.15	Naive	0.310	0.023	0.199	0.019	0.893	0.020	0.000	0.086	0.914	0.000
			Corrected	0.026	0.027	0.074	0.023	0.974	0.024	0.000	0.003	0.997	0.000
		0.50	Naive	0.318	0.025	0.207	0.023	0.916	0.022	0.000	0.089	0.911	0.000
			Corrected	0.028	0.028	0.076	0.028	0.980	0.025	0.000	0.006	0.994	0.000
		0.75	Naive	0.325	0.028	0.214	0.025	0.934	0.023	0.000	0.093	0.907	0.000
			Corrected	0.033	0.031	0.082	0.029	0.972	0.026	0.000	0.009	0.991	0.000

a: "Normal" and "Uniform" correspond to distributions of  $\epsilon_i$  in Sections 5.1.1 and 5.1.2, respectively.b: Setting 1 has  $p = 10$  and  $d_0 = 2$ ; Setting 2 has  $p = 30$  and  $d_0 = 3$ .



Table 4: Simulation results under the PH model: Scenario 2 where  $\Gamma$  is assumed known and  $\Sigma_\epsilon$  is unknown.

Distribution <sup>a</sup>	Setting <sup>b</sup>	$\sigma_\epsilon^2$	Method	Estimator of $B$		Estimator of $F(\cdot)$		Estimator of $h$		Estimator of $d$				
				$\ \Delta_B\ $	S.E.	MISE( $\hat{F}$ )	S.E.	$\hat{h}$	S.E.	$d \leq 1$	$d = 2$	$d = 3$	$d \geq 4$	
Normal	1	0.15	Naive	0.292	0.013	0.193	0.017	0.712	0.015	0.010	0.990	0.000	0.000	
			Corrected	0.017	0.021	0.064	0.022	0.759	0.016	0.000	1.000	0.000	0.000	
		0.50	Naive	0.295	0.015	0.196	0.020	0.777	0.015	0.015	0.985	0.000	0.000	
			Corrected	0.025	0.024	0.070	0.024	0.610	0.018	0.000	1.000	0.000	0.000	
		0.75	Naive	0.298	0.016	0.207	0.021	0.765	0.016	0.016	0.984	0.000	0.000	
			Corrected	0.029	0.026	0.077	0.028	0.386	0.025	0.000	1.000	0.000	0.000	
	2	0.15	Naive	0.301	0.019	0.200	0.017	0.886	0.016	0.002	0.018	0.980	0.000	
			Corrected	0.022	0.023	0.077	0.021	0.990	0.019	0.000	0.000	1.000	0.000	
		0.50	Naive	0.313	0.021	0.211	0.020	0.893	0.020	0.003	0.020	0.977	0.000	
			Corrected	0.030	0.025	0.080	0.024	0.985	0.023	0.000	0.000	1.000	0.000	
		0.75	Naive	0.321	0.023	0.224	0.023	0.898	0.020	0.003	0.025	0.972	0.000	
			Corrected	0.039	0.028	0.086	0.026	0.982	0.024	0.000	0.002	0.998	0.000	
	Uniform	1	0.15	Naive	0.303	0.017	0.200	0.018	0.820	0.016	0.009	0.991	0.000	0.000
				Corrected	0.021	0.026	0.092	0.022	0.945	0.019	0.000	1.000	0.000	0.000
			0.50	Naive	0.310	0.020	0.211	0.020	0.846	0.018	0.014	0.986	0.000	0.000
				Corrected	0.028	0.029	0.100	0.025	0.921	0.021	0.002	0.998	0.000	0.000
			0.75	Naive	0.318	0.026	0.223	0.023	0.851	0.020	0.017	0.983	0.000	0.000
				Corrected	0.031	0.033	0.104	0.027	0.858	0.025	0.008	0.998	0.000	0.000
2		0.15	Naive	0.309	0.020	0.206	0.019	0.886	0.018	0.003	0.019	0.978	0.000	
			Corrected	0.027	0.026	0.090	0.022	0.961	0.020	0.000	0.002	0.998	0.000	
		0.50	Naive	0.314	0.024	0.215	0.022	0.891	0.021	0.003	0.023	0.974	0.000	
			Corrected	0.029	0.029	0.096	0.027	0.916	0.025	0.000	0.003	0.997	0.000	
		0.75	Naive	0.322	0.028	0.229	0.025	0.905	0.024	0.004	0.026	0.970	0.000	
			Corrected	0.030	0.031	0.105	0.027	0.819	0.027	0.000	0.005	0.995	0.000	

a: "Normal" and "Uniform" correspond to distributions of  $\epsilon_i$  in Sections 5.1.1 and 5.1.2, respectively.  
b: Setting 1 has  $p = 10$  and  $d_0 = 2$ ; Setting 2 has  $p = 30$  and  $d_0 = 3$ .

Table 5: Simulation results under the PO model: Scenario 2 where  $\Gamma$  is assumed known and  $\Sigma_\epsilon$  is unknown.

Distribution <sup>a</sup>	Setting <sup>b</sup>	$\sigma_\epsilon^2$	Method	Estimator of $B$		Estimator of $F(\cdot)$		Estimator of $h$		Estimator of $d$			
				$\ \Delta_B\ $	S.E.	MISE( $F$ )	S.E.	$h$	S.E.	$d \leq 1$	$d = 2$	$d = 3$	$d \geq 4$
Normal	1	0.15	Naive	0.263	0.013	0.189	0.010	0.420	0.018	0.009	0.991	0.000	0.000
			Corrected	0.012	0.026	0.044	0.013	0.514	0.020	0.000	1.000	0.000	0.000
		0.50	Naive	0.305	0.016	0.194	0.011	0.429	0.020	0.011	0.989	0.000	0.000
			Corrected	0.016	0.028	0.053	0.014	0.017	0.021	0.002	0.997	0.001	0.000
		0.75	Naive	0.313	0.016	0.194	0.011	0.492	0.020	0.013	0.987	0.000	0.000
			Corrected	0.024	0.028	0.061	0.019	0.142	0.024	0.003	0.996	0.001	0.000
	2	0.15	Naive	0.288	0.017	0.195	0.013	0.804	0.014	0.001	0.014	0.985	0.000
			Corrected	0.026	0.024	0.050	0.016	0.913	0.017	0.000	0.000	1.000	0.000
		0.50	Naive	0.310	0.022	0.206	0.014	0.822	0.021	0.005	0.020	0.975	0.000
			Corrected	0.031	0.030	0.054	0.016	0.981	0.026	0.000	0.002	0.998	0.000
		0.75	Naive	0.319	0.025	0.215	0.019	0.835	0.022	0.005	0.024	0.971	0.000
			Corrected	0.032	0.034	0.063	0.024	0.972	0.026	0.000	0.003	0.997	0.000
Uniform	1	0.15	Naive	0.277	0.018	0.195	0.011	0.902	0.015	0.012	0.988	0.000	0.000
			Corrected	0.018	0.029	0.053	0.015	0.947	0.017	0.000	1.000	0.000	0.000
0.50		Naive	0.285	0.023	0.211	0.013	0.911	0.016	0.014	0.986	0.000	0.000	
		Corrected	0.019	0.032	0.059	0.015	0.925	0.019	0.003	0.997	0.000	0.000	
0.75		Naive	0.285	0.023	0.211	0.013	0.911	0.016	0.017	0.983	0.000	0.000	
		Corrected	0.023	0.034	0.065	0.024	0.817	0.025	0.004	0.996	0.000	0.000	
2	0.15	Naive	0.291	0.024	0.212	0.014	0.945	0.019	0.002	0.018	0.980	0.000	
		Corrected	0.018	0.031	0.056	0.016	0.833	0.020	0.000	0.000	1.000	0.000	
	0.50	Naive	0.307	0.028	0.228	0.015	0.959	0.020	0.002	0.023	0.975	0.000	
		Corrected	0.021	0.033	0.058	0.016	0.885	0.022	0.000	0.002	0.998	0.000	
	0.75	Naive	0.326	0.031	0.240	0.021	0.974	0.023	0.003	0.026	0.971	0.000	
		Corrected	0.026	0.037	0.064	0.027	0.933	0.026	0.000	0.004	0.996	0.000	

a: "Normal" and "Uniform" correspond to distributions of  $\epsilon_i$  in Sections 5.1.1 and 5.1.2, respectively.b: Setting 1 has  $p = 10$  and  $d_0 = 2$ ; Setting 2 has  $p = 30$  and  $d_0 = 3$ .

Table 6: Simulation results under the AH model: Scenario 2 where  $\Gamma$  is assumed known and  $\Sigma_\epsilon$  is unknown.

Distribution <sup>a</sup>	Setting <sup>b</sup>	$\sigma_\epsilon^2$	Method	Estimator of $B$		Estimator of $F(\cdot)$		Estimator of $h$		Estimator of $d$				
				$\ \Delta_B\ $	S.E.	MISE( $\hat{F}$ )	S.E.	$\hat{h}$	S.E.	$d \leq 1$	$d = 2$	$d = 3$	$d \geq 4$	
Normal	1	0.15	Naive	0.270	0.016	0.176	0.013	0.620	0.020	0.005	0.994	0.001	0.000	
			Corrected	0.013	0.023	0.052	0.017	0.652	0.022	0.000	1.000	0.000	0.000	
		0.50	Naive	0.275	0.017	0.184	0.015	0.618	0.020	0.010	0.989	0.001	0.000	
			Corrected	0.015	0.025	0.055	0.017	0.460	0.022	0.006	0.994	0.000	0.000	
		0.75	Naive	0.295	0.018	0.186	0.016	0.638	0.019	0.011	0.987	0.002	0.000	
			Corrected	0.016	0.026	0.059	0.018	0.313	0.029	0.007	0.992	0.001	0.000	
	2	0.15	Naive	0.283	0.019	0.185	0.018	0.943	0.018	0.002	0.010	0.988	0.000	
			Corrected	0.017	0.024	0.055	0.023	0.996	0.021	0.000	0.000	1.000	0.000	
		0.50	Naive	0.290	0.022	0.192	0.020	0.956	0.021	0.004	0.016	0.980	0.000	
			Corrected	0.024	0.028	0.058	0.025	0.966	0.026	0.000	0.002	0.998	0.000	
		0.75	Naive	0.307	0.023	0.205	0.022	0.961	0.025	0.004	0.020	0.976	0.000	
			Corrected	0.028	0.028	0.063	0.026	0.998	0.032	0.000	0.004	0.996	0.000	
	Uniform	1	0.15	Naive	0.288	0.019	0.190	0.018	0.901	0.022	0.006	0.994	0.000	0.000
				Corrected	0.016	0.025	0.056	0.021	0.962	0.027	0.000	1.000	0.000	0.000
			0.50	Naive	0.295	0.024	0.205	0.020	0.919	0.024	0.008	0.992	0.000	0.000
				Corrected	0.017	0.029	0.063	0.023	0.937	0.029	0.002	0.998	0.000	0.000
			0.75	Naive	0.310	0.027	0.221	0.021	0.935	0.027	0.011	0.989	0.000	0.000
				Corrected	0.019	0.031	0.065	0.024	0.854	0.030	0.005	0.995	0.000	0.000
2		0.15	Naive	0.297	0.021	0.208	0.019	0.908	0.023	0.003	0.012	0.985	0.000	
			Corrected	0.017	0.025	0.060	0.021	0.964	0.025	0.000	0.000	1.000	0.000	
		0.50	Naive	0.306	0.024	0.214	0.022	0.925	0.025	0.003	0.017	0.980	0.000	
			Corrected	0.018	0.026	0.064	0.027	0.930	0.027	0.000	0.002	0.998	0.000	
		0.75	Naive	0.321	0.025	0.227	0.026	0.940	0.026	0.004	0.022	0.974	0.000	
			Corrected	0.022	0.028	0.066	0.030	0.833	0.028	0.000	0.004	0.996	0.000	

a: "Normal" and "Uniform" correspond to distributions of  $\epsilon_i$  in Sections 5.1.1 and 5.1.2, respectively.  
b: Setting 1 has  $p = 10$  and  $d_0 = 2$ ; Setting 2 has  $p = 30$  and  $d_0 = 3$ .

Table 7: Simulation results under the PH model: Scenario 3 where  $\Gamma$  and  $\Sigma_\epsilon$  are unknown.

Distribution <sup>a</sup>	Setting <sup>b</sup>	$\sigma_\epsilon^2$	Method	Estimator of $B$		Estimator of $F(\cdot)$		Estimator of $h$		Estimator of $d$			
				$\ \Delta_B\ $	S.E.	MISE( $F$ )	S.E.	$h$	S.E.	$d \leq 1$	$d = 2$	$d = 3$	$d \geq 4$
Normal	1	0.15	Naive	0.249	0.012	0.156	0.016	0.715	0.012	0.010	0.985	0.005	0.000
			Corrected	0.014	0.018	0.058	0.019	0.820	0.013	0.000	1.000	0.000	0.000
		0.50	Naive	0.251	0.013	0.163	0.019	0.747	0.016	0.020	0.973	0.007	0.000
			Corrected	0.018	0.022	0.063	0.023	0.825	0.014	0.000	1.000	0.000	0.000
		0.75	Naive	0.272	0.016	0.176	0.020	0.778	0.016	0.017	0.978	0.005	0.000
			Corrected	0.017	0.024	0.066	0.025	0.863	0.013	0.000	1.000	0.000	0.000
	2	0.15	Naive	0.253	0.013	0.165	0.016	0.876	0.011	0.001	0.013	0.986	0.000
			Corrected	0.019	0.017	0.072	0.017	0.991	0.014	0.000	0.000	1.000	0.000
		0.50	Naive	0.261	0.016	0.171	0.018	0.883	0.014	0.002	0.016	0.982	0.000
			Corrected	0.023	0.021	0.083	0.020	0.989	0.017	0.000	0.000	1.000	0.000
		0.75	Naive	0.269	0.022	0.189	0.021	0.897	0.019	0.002	0.020	0.978	0.000
			Corrected	0.025	0.026	0.084	0.024	0.991	0.025	0.000	0.002	0.998	0.000
Uniform	1	0.15	Naive	0.263	0.014	0.170	0.018	0.833	0.014	0.014	0.986	0.000	0.000
			Corrected	0.018	0.018	0.089	0.020	0.977	0.016	0.000	1.000	0.000	0.000
		0.50	Naive	0.285	0.020	0.184	0.023	0.865	0.016	0.019	0.981	0.000	0.000
			Corrected	0.025	0.027	0.091	0.028	0.988	0.018	0.000	1.000	0.000	0.000
		0.75	Naive	0.294	0.023	0.190	0.026	0.890	0.016	0.023	0.977	0.000	0.000
			Corrected	0.028	0.027	0.094	0.030	0.964	0.018	0.002	0.998	0.000	0.000
	2	0.15	Naive	0.271	0.016	0.183	0.020	0.845	0.014	0.002	0.018	0.980	0.000
			Corrected	0.020	0.019	0.098	0.021	0.986	0.016	0.000	0.000	1.000	0.000
		0.50	Naive	0.280	0.023	0.191	0.022	0.866	0.017	0.002	0.023	0.975	0.000
			Corrected	0.028	0.029	0.107	0.025	0.983	0.022	0.000	0.002	0.998	0.000
		0.75	Naive	0.293	0.026	0.205	0.026	0.893	0.023	0.003	0.024	0.973	0.000
			Corrected	0.031	0.029	0.112	0.032	0.988	0.028	0.000	0.003	0.997	0.000

a: "Normal" and "Uniform" correspond to distributions of  $\epsilon_i$  in Sections 5.1.1 and 5.1.2, respectively.b: Setting 1 has  $p = 10$  and  $d_0 = 2$ ; Setting 2 has  $p = 30$  and  $d_0 = 3$ .

Table 8: Simulation results under the PO model: Scenario 3 where  $\Gamma$  and  $\Sigma_\epsilon$  are unknown.

Distribution <sup>a</sup>	Setting <sup>b</sup>	$\sigma_\epsilon^2$	Method	Estimator of $B$		Estimator of $F(\cdot)$		Estimator of $h$		Estimator of $d$				
				$\ \Delta_B\ $	S.E.	MISE( $F$ )	S.E.	$\hat{h}$	S.E.	$d \leq 1$	$d = 2$	$d = 3$	$d \geq 4$	
Normal	1	0.15	Naive	0.253	0.012	0.164	0.009	0.306	0.019	0.008	0.992	0.000	0.000	
			Corrected	0.011	0.023	0.040	0.012	0.506	0.021	0.002	0.998	0.000	0.000	
		0.50	Naive	0.259	0.016	0.172	0.011	0.374	0.021	0.011	0.989	0.000	0.000	
			Corrected	0.016	0.025	0.042	0.014	0.597	0.021	0.003	0.997	0.000	0.000	
		0.75	Naive	0.276	0.016	0.176	0.011	0.415	0.022	0.013	0.987	0.000	0.000	
			Corrected	0.021	0.027	0.047	0.017	0.618	0.022	0.003	0.997	0.000	0.000	
	2	0.15	True	0.009	0.015	0.030	0.010	0.467	0.019	0.003	0.997	0.000	0.000	
			Naive	0.266	0.017	0.176	0.011	0.901	0.019	0.002	0.018	0.980	0.000	
			Corrected	0.020	0.021	0.040	0.015	0.953	0.025	0.000	0.001	0.999	0.000	
		0.50	Naive	0.271	0.020	0.188	0.012	0.915	0.022	0.002	0.022	0.976	0.000	
			Corrected	0.025	0.025	0.044	0.015	0.942	0.026	0.000	0.000	1.000	0.000	
		0.75	Naive	0.284	0.023	0.195	0.014	0.920	0.024	0.003	0.025	0.972	0.000	
	Corrected		0.030	0.029	0.050	0.017	0.958	0.029	0.000	0.003	0.997	0.000		
	Uniform	1	0.15	Naive	0.260	0.018	0.175	0.012	0.856	0.019	0.010	0.990	0.000	0.000
				Corrected	0.016	0.024	0.045	0.015	0.956	0.023	0.002	0.998	0.000	0.000
			0.50	Naive	0.269	0.023	0.182	0.015	0.867	0.022	0.012	0.988	0.000	0.000
				Corrected	0.019	0.032	0.050	0.018	0.971	0.024	0.004	0.996	0.000	0.000
			0.75	Naive	0.280	0.029	0.195	0.020	0.893	0.024	0.015	0.985	0.000	0.000
Corrected				0.023	0.035	0.050	0.024	0.991	0.028	0.005	0.995	0.000	0.000	
2		0.15	Naive	0.278	0.022	0.189	0.013	0.898	0.021	0.001	0.019	0.980	0.000	
			Corrected	0.018	0.026	0.056	0.016	0.954	0.025	0.000	0.002	0.998	0.000	
		0.50	Naive	0.286	0.027	0.197	0.015	0.914	0.024	0.001	0.022	0.977	0.000	
			Corrected	0.020	0.035	0.057	0.019	0.958	0.026	0.000	0.005	0.995	0.000	
		0.75	Naive	0.297	0.033	0.213	0.020	0.935	0.026	0.003	0.026	0.971	0.000	
			Corrected	0.026	0.038	0.065	0.026	0.944	0.030	0.000	0.005	0.995	0.000	

a: "Normal" and "Uniform" correspond to distributions of  $\epsilon_i$  in Sections 5.1.1 and 5.1.2, respectively.  
b: Setting 1 has  $p = 10$  and  $d_0 = 2$ ; Setting 2 has  $p = 30$  and  $d_0 = 3$ .

Table 9: Simulation results under the AH model: Scenario 3 where  $\Gamma$  and  $\Sigma_\epsilon$  are unknown.

Distribution <sup>a</sup>	Setting <sup>b</sup>	$\sigma_\epsilon^2$	Method	Estimator of $B$		Estimator of $F(\cdot)$		Estimator of $h$		Estimator of $d$			
				$\ \Delta_B\ $	S.E.	MISE( $F$ )	S.E.	$h$	S.E.	$d \leq 1$	$d = 2$	$d = 3$	$d \geq 4$
Normal	1	0.15	Naive	0.263	0.012	0.159	0.013	0.631	0.018	0.010	0.990	0.000	0.000
			Corrected	0.010	0.022	0.044	0.015	0.742	0.018	0.003	0.997	0.000	0.000
		0.50	Naive	0.261	0.014	0.168	0.014	0.594	0.022	0.012	0.988	0.000	0.000
			Corrected	0.014	0.023	0.046	0.016	0.704	0.019	0.003	0.997	0.000	0.000
		0.75	Naive	0.281	0.016	0.172	0.014	0.652	0.019	0.011	0.989	0.000	0.000
			Corrected	0.015	0.025	0.053	0.017	0.779	0.015	0.003	0.997	0.000	0.000
	2	0.15	Naive	0.278	0.015	0.163	0.016	0.911	0.017	0.000	0.012	0.988	0.000
			Corrected	0.016	0.021	0.045	0.020	0.994	0.018	0.000	0.002	0.998	0.000
		0.50	Naive	0.285	0.018	0.176	0.019	0.923	0.020	0.001	0.015	0.984	0.000
			Corrected	0.020	0.022	0.047	0.022	0.995	0.024	0.000	0.002	0.998	0.000
		0.75	Naive	0.291	0.022	0.184	0.020	0.935	0.023	0.002	0.023	0.975	0.000
			Corrected	0.021	0.025	0.060	0.023	0.989	0.029	0.000	0.003	0.997	0.000
Uniform	1	0.15	Naive	0.281	0.022	0.170	0.016	0.911	0.018	0.011	0.989	0.000	0.000
			Corrected	0.016	0.031	0.045	0.020	0.995	0.019	0.002	0.998	0.000	0.000
		0.50	Naive	0.293	0.027	0.184	0.019	0.923	0.019	0.014	0.986	0.000	0.000
			Corrected	0.019	0.032	0.048	0.023	0.997	0.022	0.005	0.995	0.000	0.000
		0.75	Naive	0.313	0.029	0.196	0.022	0.948	0.023	0.019	0.981	0.000	0.000
			Corrected	0.021	0.034	0.055	0.025	0.982	0.026	0.006	0.994	0.000	0.000
	2	0.15	Naive	0.297	0.026	0.181	0.017	0.940	0.019	0.000	0.013	0.987	0.000
			Corrected	0.022	0.035	0.048	0.018	0.989	0.021	0.000	0.003	0.997	0.000
		0.50	Naive	0.302	0.031	0.196	0.017	0.963	0.021	0.000	0.018	0.982	0.000
			Corrected	0.023	0.037	0.053	0.019	0.985	0.024	0.000	0.004	0.996	0.000
		0.75	Naive	0.316	0.034	0.207	0.023	0.978	0.025	0.002	0.022	0.976	0.000
			Corrected	0.028	0.039	0.058	0.027	0.989	0.028	0.000	0.007	0.993	0.000

a: "Normal" and "Uniform" correspond to distributions of  $\epsilon_i$  in Sections 5.1.1 and 5.1.2, respectively.b: Setting 1 has  $p = 10$  and  $d_0 = 2$ ; Setting 2 has  $p = 30$  and  $d_0 = 3$ .

### 5.1.3. Misspecification of the measurement error distribution

To assess the sensitivity of the proposed methods to misspecification of the measurement error distribution, we repeat the simulation studies in Section 5.1.2, except that the normal distribution  $N(0, \Sigma_\epsilon)$  of  $\epsilon_i$  or  $\epsilon_{ir}$  is replaced by the uniform distribution  $\text{UNIF}(0, \sqrt{12}\sigma_\epsilon)$  when we generate data. Simulation results are summarized in the bottom panels of Tables 1-9.

The results demonstrate the same patterns observed in Section 5.1.2. The estimators of  $B$  and  $F(\cdot)$  produced from the proposed methods still outperform those without suitable correction for the measurement error effects. Standard errors for the associated methods exhibit the same patterns as those shown in Section 5.1.2. Compared to the results in Section 5.1.2, the performance of the proposed estimators appears to be fairly insensitive under the misspecification of the measurement error distribution we consider.

In summary, in the presence of measurement error, the naive method yields unsatisfactory results. The proposed methods appear to successfully correct for the measurement error effects for various settings.

## 5.2. Analysis of ACTG 175 dataset

We implement the proposed method to analyze the AIDS Clinical Trials Group (ACTG) 175 data which were discussed by [10]. The ACTG 175 study was a double-blind randomized clinical trial which evaluated the HIV treatment effects. The dataset is available in R package “speff2trial”. The dataset contains measurements on 26 variables for 2139 individuals; these variables are `age`, `wtkg`, `hemo`, `homo`, `drugs`, `karnof`, `oprior`, `z30`, `zprior`, `preanti`, `race`, `gender`, `str2`, `strat`, `symptom`, `treat`, `offtrt`, `cd40`, `cd420`, `cd496`, `r`, `cd80`, `cd820`, `cens`, `days`, and `arms`. Since the variable `cd496` contains missing values and `r` is its missing indicator, so we remove those two variables. In addition, we remove variables `zprior` and `treat` because `zprior` is the constant 1 for all subjects and `treat` indicates whether or not the subject received the zidovudine treatment, overlapping with `arms`. As a result, in addition to the survival time `days` and the censoring indicator `cens`, we have  $p = 20$  covariates in the dataset where `cd40` is error-prone. The censoring rate of this dataset is approximately 75.6%.

Forty-four subjects were measured once for `cd40` at the baseline, while 2095 subjects had two replicated baseline measurements of `cd40`. As discussed in [26, Section 3.6.4], let  $X$  denote  $\log(\text{cd40} + 1)$ . To implement the proposed method, we consider the measurement error model (2.9) to facilitate the availability of repeated measurements. Consequently, based on the discussion of Scenario II in Section 2.3, the estimates of  $\Sigma_\epsilon$  and  $\Sigma_{X^*}$  are given by  $\widehat{\Sigma}_\epsilon = 0.035$  and  $\widehat{\Sigma}_{X^*} = 0.114$ , respectively, yielding  $\widehat{L} = 0.693$  as indicated by (2.12). In this study, we consider the entire data with error correction by  $\widehat{L}$ . In addition to  $X$ , let  $Z$  denote the vector of the remaining 19 covariates that are assumed to

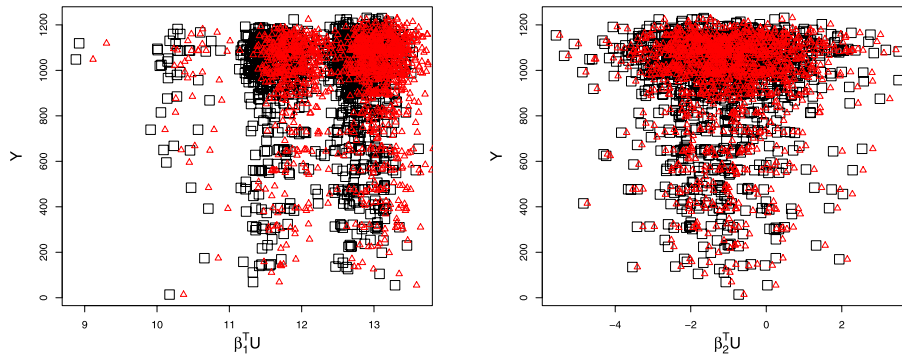


FIG 1. Scatter plots of the observed time  $Y$  against  $\beta_j^\top U$  with  $j = 1, 2$ . The left panel is  $\beta_1^\top U$  and the right panel is  $\beta_2^\top U$ . The first row with black boxes ( $\square$ ) is obtained from the naive approach, and the second row with red triangles ( $\triangle$ ) is obtained from the proposed method.

be precisely measured. Consequently, we let  $U = (\widehat{L}X^{*\top}, Z^\top)^\top$  denote a 20-dimensional vector of covariates to be implemented with the proposed method.

The naive method and the proposed method give  $\widehat{d} = 2$ , suggesting that there are two directions in the central subspace, say  $\beta_1$  and  $\beta_2$ . In Figure 1, we present the scatter plots of the observed time  $Y_i$  and  $\beta_j^\top U_i$  with  $j = 1, 2$ . While the naive method and the proposed method show fairly similar patterns, the scatter plot based on the naive method seems slightly more variable than that for the proposed method. The similarity may be pertinent to the small degree of measurement error, indicated by the estimate  $\widehat{\Sigma}_\epsilon = 0.035$  of  $\Sigma_\epsilon$  obtained from the replicate measurements. The more variability of the results obtained from the naive method agrees with the phenomenon that often exhibits in settings with measurement error (e.g., [4]; [5]; [26]). We also examine the estimated functions  $1 - \widehat{F}(\cdot)$  for subjects  $i = 1, 7$  and  $23$ , and display the curves in Figure 2. It is seen that the estimated curves based on the naive and the proposed methods have similar patterns.

## 6. Discussion

Sufficient dimension reduction is a useful tool in regression models, which mainly reduces the dimension of variables with the predictive information of covariates preserved. In this paper, we deal with SDR for censored data with measurement error using cumulative distribution models. We propose valid inferential procedures to correct for the measurement error effects and estimate the central subspace. Our methods are justified theoretically, and their finite sample performance is demonstrated to be satisfactory through simulation studies.

Some further extensions are worth exploring. In the development here the censoring time  $C_i$  is assumed to be independent of the covariates  $X_i$ . This consideration is basically driven by its common use in the literature, which enables



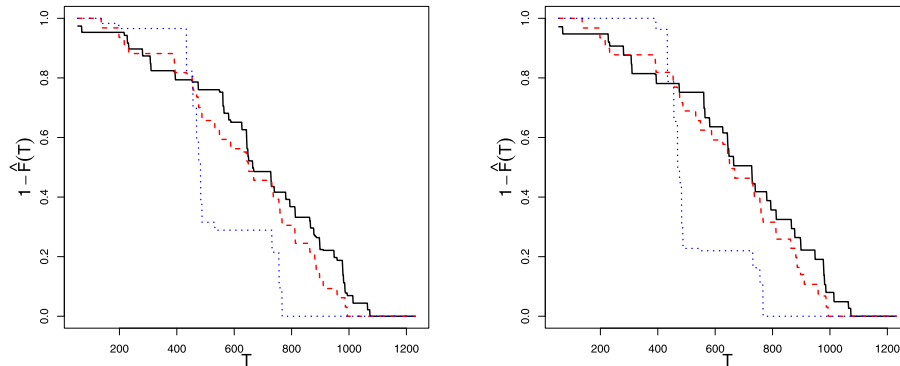


FIG 2. Estimated curves of  $1 - \hat{F}(y|U_i)$ . The solid curve is for subject  $i = 1$ , the dash curve is for subject  $i = 7$ , and the dot curve is for subject  $i = 23$ . The left panel is given by the proposed method, the right panel is given by the naive method.

us to use the Kaplan-Meier estimator to consistently estimate the survivor function of the censoring process. This assumption, however, is not essential. One may relax it by replacing the unconditional survivor function  $G(c)$  of  $C_i$  with the conditional survivor function,  $G(c|X_i) \triangleq P(C_i > c|X_i)$ , of  $C_i$  given  $X_i$ , and modify the development here accordingly. Estimation of  $G(c|X_i)$  may be performed with parametric or semiparametric regression models imposed (e.g., [18]).

As shown in Section 2.2, the normal distribution for  $X$  and  $\epsilon$  is required when replacing  $X^*$  by  $U$ , which ensures the invariance law to hold ([14]). When  $X$  does not follow a normal distribution but has a large dimension, Li and Yin [14, Section 6] discussed approximate invariance using the low dimensional projections of covariates. It would be interesting to examine whether this scheme can be combined with the estimation procedure in Section 3 to handle non-normally distributed and error-contaminated covariates. Another interesting topic is to relax the normality assumption of  $X$  by assuming  $X$  to be elliptical symmetric [8], as suggested by the Associate Editor.

The development here considers the case where the components of  $X$  are all continuous random variables. In applications, however, error-prone vector  $X$  may contain discrete components. It is interesting to generalize our methods to accommodating this circumstance, and such research warrants careful explorations.

## Appendix A: Regularity conditions

- (C1)  $P(Y_i < \tau) > 0$ , where  $\tau$  is an upper bound of survival times assumed to be finite.
- (C2) The survival time and the censoring time are conditionally independent, given the covariates. The censoring time  $C_i$  is independent of the covariates

$X_i$ .

- (C3) For any  $t > 0$ , the  $\{\mathbb{I}(Y_i \leq t), X_i^*\}$  are independent and identically distributed for  $i = 1, \dots, n$ .
- (C4) The bandwidth  $h$  is chosen from the interval  $H_{\kappa;n} = [h_l n^{-\kappa}, h_u n^{-\kappa}]$  for some constants  $h_l$  and  $h_u$  and  $\kappa \in (1/(4q), 1/\max\{2d+2, d+4\})$ .
- (C5)  $\mathcal{A}$  in Section 4 is a positive definite matrix.
- (C6)  $\widehat{G}(\cdot)$  is a consistent estimator of  $G(\cdot)$  which is bounded away from zero.

Conditions (C1) to (C3) are regular assumptions in survival analysis for the establishment of the asymptotic properties (e.g., [1]). The independence of  $C_i$  and  $X_i$  is commonly imposed in survival analysis (e.g., [22]). Condition (C4) guides the choice of a suitable bandwidth and is used to establish the  $\sqrt{n}$ -consistency of  $\widehat{B}$  [13]. Condition (C5) is a routine requirement that is used to establish the asymptotic distribution of  $\widehat{B}$ . Consistency of  $\widehat{G}(\cdot)$  in Condition (C6) is imposed to establish the asymptotic properties, which was also used by other authors such as [18]. Requiring it to be bounded away from zero avoids numerical instability in extreme situations, as noted by [29].

## Appendix B: Technical Lemmas

Taking the difference between (4.1) and (4.3) and between (4.2) and (4.3), we further define

$$\widetilde{\mathbb{F}}_{l,B,L}^{(j)}(y, u) = \widehat{\mathbb{F}}_{l,B,L}^{(j)}(y, B^\top u) - \mathbb{F}_{l,B,L}^{(j)}(y, u) \quad (\text{B.1})$$

and

$$\widetilde{\mathbb{F}}_{l,B,\widehat{L}}^{(j)}(y, u) = \widehat{\mathbb{F}}_{l,B,\widehat{L}}^{(j)}(y, B^\top u) - \mathbb{F}_{l,B,L}^{(j)}(y, u) \quad (\text{B.2})$$

for  $l = 0, 1$  and  $j = 0, 1, 2$ .

Define

$$\widetilde{F}^{(j)}(y, u) = \nabla_{\text{vec}(B)}^j \widehat{F}(y, B^\top u) - F^{(j)}(y, u) \quad (\text{B.3})$$

for  $j = 0, 1$ , where  $\widehat{F}(y, B^\top u)$  is given by (3.4), and  $F^{(j)}(y, u)$  with  $j = 0$  and 1 is defined in (4.5) and (4.6).

Noting that  $\nabla_{\text{vec}(B)}^j \widehat{F}(y, B^\top u)$  involves the measurement error correction term  $L$  which is assumed to be known or estimated, so here we add the subscript  $(B, L)$  or  $(B, \widehat{L})$  in (B.3) to show each case. That is, if  $L$  is known, then we re-write (B.3) as

$$\widetilde{F}_{B,L}^{(j)}(y, u) = \nabla_{\text{vec}(B),L}^j \widehat{F}(y, B^\top u) - F_{B,L}^{(j)}(y, u);$$

if  $L$  is unknown and is estimated by  $\widehat{L}$ , then we express (B.3) as

$$\widetilde{F}_{B,\widehat{L}}^{(j)}(y, u) = \nabla_{\text{vec}(B),\widehat{L}}^j \widehat{F}(y, B^\top u) - F_{B,L}^{(j)}(y, u).$$

In the following lemmas, we present the convergence rates of (B.1) and (B.3).

**Lemma B.1.** *Suppose that regularity conditions in Appendix A hold. For  $j = 0, 1, 2$  and  $l = 0, 1$ , if  $L$  is known, then*

$$\sup_{y,u,B} \left\| \tilde{\mathbb{F}}_{l,B,L}^{(j)}(y,u) \right\| = O_p(h^q) + o_p\left(\frac{\log n}{\sqrt{nh^{j+d}}}\right); \quad (\text{B.4})$$

if  $L$  is unknown and is estimated by an estimator  $\hat{L}$ , then

$$\sup_{y,u,B} \left\| \tilde{\mathbb{F}}_{l,B,\hat{L}}^{(j)}(y,u) \right\| = o_p\left(\frac{p \log m}{\sqrt{m}}\right) + O_p(h^q) + o_p\left(\frac{\log n}{\sqrt{nh^{j+d}}}\right). \quad (\text{B.5})$$

**Proof:**

We first show (B.4). Since  $\{\mathbb{I}(Y_i \leq y) : y \geq 0\}$ ,  $\{\mathcal{K}(B^\top U_i - u) : B \in \mathbb{R}^{p \times d}\}$  and  $\{(U_i - u)^{\otimes j} : j = 0, 1, 2\}$  are the Vapnik-Červonenkis (VC) classes by [9, p.911] and Lemma 2.4 in [20]. Further, Lemma 2.12 in [20] implies that those three classes are Euclidean, and thus, Lemma 2.14 in [20] indicates that  $\left\{ \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) \right\}^l \nabla_{\text{vec}(B)}^j \mathcal{K}(B^\top \widehat{U}_i - u) : y, u, B \right\}$  is also a Euclidean. As a result, by Theorem II.37 in [21] and derivations similar to [9], we have that

$$\widehat{\mathbb{F}}_{l,B,L}^{(j)}(y, B^\top u) - E \left\{ \widehat{\mathbb{F}}_{l,B,L}^{(j)}(y, B^\top u) \right\} = o_p\left(\frac{\log n}{\sqrt{nh^{j+d}}}\right) \quad (\text{B.6})$$

and

$$E \left\{ \widehat{\mathbb{F}}_{l,B,L}^{(j)}(y, B^\top u) \right\} - \mathbb{F}_{l,B,L}^{(j)}(y, B^\top u) = O_p(h^q) \quad (\text{B.7})$$

for  $l = 0, 1$  and  $j = 0, 1, 2$ . Therefore, combining (B.6) and (B.7) gives (B.4).

We next show (B.5). Consider

$$\begin{aligned} & \widehat{\mathbb{F}}_{l,B,\hat{L}}^{(j)}(y,u) - \widehat{\mathbb{F}}_{l,B,L}^{(j)}(y,u) \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) \right\}^l \nabla_{\text{vec}(B)}^j \left\{ \mathcal{K}(B^\top \widehat{U}_i - u) - \mathcal{K}(B^\top U_i - u) \right\} \right] \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) \right\}^l \nabla_{\text{vec}(B)}^j \left\{ \nabla_L^1 \mathcal{K}(B^\top U_i - u) \right\} \right] (\hat{L} - L). \end{aligned}$$

Since  $\hat{L}$  is estimated by either repeated measurements or a validation sample, by similar derivations of Theorem 1 in [27], we have that

$$\hat{L} - L = o_p\left(\frac{p \log m}{\sqrt{m}}\right).$$

As the result, we have

$$\sup_{y,u,B} \left\| \widehat{\mathbb{F}}_{l,B,\hat{L}}^{(j)}(y,u) - \widehat{\mathbb{F}}_{l,B,L}^{(j)}(y,u) \right\| = o_p\left(\frac{p \log m}{\sqrt{m}}\right),$$

and combining the result (B.4) gives the desired result of (B.5).  $\square$

Next, let  $\zeta_{i,B_0}^{(0)}(y, u)$  be as defined in (4.4), and define

$$\begin{aligned} & \zeta_{i,B_0}^{(1)}(y, u) \\ = & \sum_{j,l=0}^1 \left\{ \tilde{\mathbb{F}}_{i,l,B,L}^{(j)}(y, u) \frac{(-1)^{1+l} \mathbb{F}_{1-l,B,L}^{(0)}(y, B^\top u)}{\mathbb{F}_{0,B,L}^{(0)3-l}(y, B^\top u)} \right. \\ & \left. \times \left( \sum_{l'=0}^1 (l+l'-2) \mathbb{F}_{1-l',B,L}^{l'}(u, B^\top u) \mathbb{F}_{l',B,L}^{l'}(y, B^\top u) \mathbb{F}_{l',B,L}^1(y, B^\top u) \right)^{1-j} \right\}. \end{aligned} \quad (\text{B.8})$$

**Lemma B.2.** *Suppose that regularity conditions in Appendix A holds. For  $j = 0, 1$ , if  $L$  is known, then*

$$\sup_{y,u,B} \left\| \tilde{F}_{B,L}^{(j)}(y, u) - \frac{1}{n} \sum_{i=1}^n \zeta_{i,B}^{(j)}(y, u) \right\| = o_p \left( \frac{1}{\sqrt{n}} \right); \quad (\text{B.9})$$

if  $L$  is unknown and  $\hat{L}$  is the estimator, then

$$\sup_{y,u,B} \left\| \tilde{F}_{B,\hat{L}}^{(j)}(y, u) - \frac{1}{n} \sum_{i=1}^n \zeta_{i,B}^{(j)}(y, u) \right\| = o_p \left( \frac{1}{\sqrt{n}} \right) + o_p \left( \frac{1}{\sqrt{m}} \right). \quad (\text{B.10})$$

**Proof:**

For  $j = 0$ , by expressions (3.4) and (4.1), we observe that

$$\begin{aligned} \tilde{F}_{B,L}^{(0)}(y, u) &= \nabla_{\text{vec}(B),L}^0 \hat{F}(y, B^\top u) - F_{B,L}^{(0)}(y, u) \\ &= \frac{\hat{\mathbb{F}}_{1,B,L}^{(0)}(y, B^\top u)}{\hat{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u)} - \frac{\mathbb{F}_{1,B,L}^{(0)}(y, u)}{\mathbb{F}_{0,B,L}^{(0)}(y, u)} \\ &= \frac{\hat{\mathbb{F}}_{1,B,L}^{(0)}(y, B^\top u) \mathbb{F}_{0,B,L}^{(0)}(y, u) - \hat{\mathbb{F}}_{0,B,L}^{(0)}(y, u) \mathbb{F}_{1,B,L}^{(0)}(y, u)}{\hat{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u) \mathbb{F}_{0,B,L}^{(0)}(y, u)} \\ &= \frac{\mathbb{F}_{0,B,L}^{(0)}(y, u) \left\{ \hat{\mathbb{F}}_{1,B,L}^{(0)}(y, B^\top u) - \mathbb{F}_{1,B,L}^{(0)}(y, B^\top u) \right\}}{\hat{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u) \mathbb{F}_{0,B,L}^{(0)}(y, u)} \\ &\quad - \frac{\mathbb{F}_{1,B,L}^{(0)}(y, u) \left\{ \hat{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u) - \mathbb{F}_{0,B,L}^{(0)}(y, B^\top u) \right\}}{\hat{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u) \mathbb{F}_{0,B,L}^{(0)}(y, u)} \\ &= \frac{\tilde{\mathbb{F}}_{1,B,L}^{(0)}(y, B^\top u)}{\hat{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u)} - \frac{F(y, u) \tilde{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u)}{\hat{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u)}, \end{aligned} \quad (\text{B.11})$$

where the fourth equality is due to adding and subtracting the term  $\mathbb{F}_{0,B,L}^{(0)}(y, u) \mathbb{F}_{1,B,L}^{(0)}(y, u)$ , the fifth equality comes from (B.1) with  $j = 0$  and  $l = 0, 1$ .

Moreover, applying the Taylor series expansion to  $\left\{\widehat{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u)\right\}^{-1}$  gives

$$\begin{aligned} \frac{1}{\widehat{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u)} &= \frac{1}{\mathbb{F}_{0,B,L}^{(0)}(y, B^\top u)} - \frac{\widetilde{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u)}{\mathbb{F}_{0,B,L}^{(0)2}(y, B^\top u)} \\ &\quad + 2\frac{\left\{\widetilde{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u)\right\}^2}{\widehat{\mathbb{F}}_{0,B,L}^{(0)*3}(y, B^\top u)}, \end{aligned} \tag{B.12}$$

where  $\widehat{\mathbb{F}}_{0,B,L}^{(0)*}$  is “between”  $\widehat{\mathbb{F}}_{0,B,L}^{(0)}(y, B^\top u)$  and  $\mathbb{F}_{0,B,L}^{(0)}(y, B^\top u)$ . Combining (B.11) with (B.12) and applying Lemma B.1 give that

$$\begin{aligned} \widetilde{F}_{B,L}^{(0)}(y, u) &= \frac{1}{n} \sum_{i=1}^n \zeta_{i,B}^{(0)}(y, u) - \frac{\widetilde{\mathbb{F}}_{0,B,L}^{(0)}(y, u)\widetilde{\mathbb{F}}_{1,B,L}^{(0)}(y, u)}{\mathbb{F}_{0,B,L}^{(0)2}(y, B^\top u)} \\ &\quad + \frac{2\widetilde{\mathbb{F}}_{1,B,L}^{(0)}(y, B^\top u)\left\{\widetilde{\mathbb{F}}_{0,B,L}^{(0)}(y, u)\right\}^2}{\widehat{\mathbb{F}}_{0,B,L}^{(0)*3}(y, B^\top u)}, \end{aligned} \tag{B.13}$$

where  $\zeta_{i,B}^{(0)}(y, u)$  is given by (4.4).

By Lemma B.1, we have that

$$\begin{aligned} &\sup_{y,u,B} \left\| \frac{\widetilde{\mathbb{F}}_{0,B,L}^{(0)}(y, u)\widetilde{\mathbb{F}}_{1,B,L}^{(0)}(y, u)}{\mathbb{F}_{0,B,L}^{(0)2}(y, B^\top u)} - \frac{2\widetilde{\mathbb{F}}_{1,B,L}^{(0)}(y, B^\top u)\left\{\widetilde{\mathbb{F}}_{0,B,L}^{(0)}(y, u)\right\}^2}{\widehat{\mathbb{F}}_{0,B,L}^{(0)*3}(y, B^\top u)} \right\| \\ &= o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{B.14}$$

Therefore, combining with (B.13) and (B.14) gives (B.9) with  $j = 0$ . Similar procedure gives (B.9) with  $j = 1$ .

We next discuss the derivation of (B.10). Since

$$\begin{aligned} \widetilde{F}_{B,\widehat{L}}^{(j)}(y, u) &= \nabla_{\text{vec}(B),\widehat{L}}^j \widehat{F}(y, B^\top u) - F_{B,L}^{(j)}(y, u) \\ &= \left(\nabla_{\text{vec}(B),\widehat{L}}^j \widehat{F}(y, B^\top u) - \nabla_{\text{vec}(B),L}^j \widehat{F}(y, B^\top u)\right) \\ &\quad + \left(\nabla_{\text{vec}(B),L}^j \widehat{F}(y, B^\top u) - F_{B,L}^{(j)}(y, u)\right) \\ &= \left(\nabla_{\text{vec}(B),\widehat{L}}^j \widehat{F}(y, B^\top u) - \nabla_{\text{vec}(B),L}^j \widehat{F}(y, B^\top u)\right) \\ &\quad + \widetilde{F}_{B,L}^{(j)}(y, u) \end{aligned} \tag{B.15}$$

for  $j = 0, 1$ . The second term of (B.15) is derived, so the remaining target is the first term of (B.15). By the Taylor series expansion on  $\nabla_{\text{vec}(B),\widehat{L}}^j \widehat{F}(y, B^\top u)$  with respect to  $L$  gives

$$\nabla_{\text{vec}(B),\widehat{L}}^j \widehat{F}(y, B^\top u) - \nabla_{\text{vec}(B),L}^j \widehat{F}(y, B^\top u) = \left(\nabla_L^1 \nabla_{\text{vec}(B),L}^j \widehat{F}(y, B^\top u)\right) (\widehat{L} - L)$$

for  $j = 0, 1$ . By the result in Lemma B.1, we have

$$\sup_{y,u,B} \left\| \nabla_{\text{vec}(B), \widehat{L}}^j \widehat{F}(y, B^\top u) - \nabla_{\text{vec}(B), L}^j \widehat{F}(y, B^\top u) \right\| = o_p \left( \frac{1}{\sqrt{m}} \right). \tag{B.16}$$

Consequently, combining (B.16) and (B.9) with (B.15) gives (B.10).  $\square$

Let

$$\begin{aligned} \sigma_0^2 &= E \left\{ \left\| \mathbb{I}(Y_i \leq y) - F(y, B_0^\top U_i) \right\|^2 \right\} \\ b_0^2(B) &= E \left\{ \left\| F(y, B_0^\top U_i) - F(y, B^\top U_i) \right\|^2 \right\} \\ \mathcal{B}_B(y, u) &= \frac{\int v^q K(v) dv}{q!} (dh^q) \sum_{l=0}^1 \frac{\{-F(y, u)\}^{1-l} \nabla^q \mathbb{F}_{0,B,L}^{(0)}(y, u)}{\mathbb{F}_{0,B,L}^{(0)}(y, u)} \\ \mathcal{V}_B(y, u) &= \frac{(\int K(v) dv)^d F(y, u) \{1 - F(y, u)\}}{nh^d f_{B^\top U}(u)}. \end{aligned}$$

In addition, define

$$AMISE_B(h) = E \{ \mathcal{B}_B^2(y, B^\top u) + \mathcal{V}_B(y, B^\top u) \}$$

and

$$ECV(B, d, h) = \begin{cases} \sigma_0^2 + AMISE_B(h) & \text{if } \mathcal{S}_{T|U} \subseteq \mathcal{S}(B) \\ \sigma_0^2 + b_0^2(B) + AMISE_B(h) & \text{if } \mathcal{S}_{T|U} \not\subseteq \mathcal{S}(B). \end{cases}$$

In the next lemma, we examine the behavior of the CV value.

**Lemma B.3.** *Under regularity conditions in Appendix A, if  $\mathcal{S}_{T|U} \subseteq \mathcal{S}(B)$ , then*

$$\sup_{B,d,h} \frac{|CV(B, d, h) - ECV(B, d, h)|}{AMISE_B(h)} = o_p(1); \tag{B.17}$$

if  $\mathcal{S}_{T|U} \not\subseteq \mathcal{S}(B)$ , then

$$\sup_{B,d,h} \frac{|CV(B, d, h) - ECV(B, d, h)|}{b_0(B)AMISE_B(h)} = O_p(1). \tag{B.18}$$

**Proof:**

Let

$$\begin{aligned} \mathcal{E}_{1;i,Y_j} &= \mathbb{I}(Y_i \leq y) - F(Y_j, B_0^\top U_i), \\ \mathcal{E}_{2;i,Y_j} &= F(Y_j, B_0^\top U_i) - F(Y_j, B^\top U_i), \\ \mathcal{E}_{3;i,Y_j} &= F(Y_j, B^\top U_i) - \widehat{F}^{(-i)}(Y_j, B^\top U_i), \end{aligned}$$

and

$$\mathcal{E}_{4;i,Y_j} = \widehat{F}^{(-i)}(Y_j, B^\top U_i) - \widehat{F}^{(-i)}(Y_j, B^\top \widehat{U}_i).$$

Then the cross-validation criterion (3.5) can be decomposed by

$$\begin{aligned}
& CV(B, d, h) \\
&= \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{E}_{1;i,Y_j}^2 + \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{E}_{2;i,Y_j}^2 + \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{E}_{3;i,Y_j}^2 + \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{E}_{4;i,Y_j}^2 \\
&\quad + \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{E}_{1;i,Y_j} \mathcal{E}_{2;i,Y_j} + \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{E}_{1;i,Y_j} \mathcal{E}_{3;i,Y_j} + \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{E}_{1;i,Y_j} \mathcal{E}_{4;i,Y_j} \\
&\quad + \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{E}_{2;i,Y_j} \mathcal{E}_{3;i,Y_j} + \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{E}_{2;i,Y_j} \mathcal{E}_{4;i,Y_j} + \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{E}_{3;i,Y_j} \mathcal{E}_{4;i,Y_j} \\
&\triangleq S_1 + S_2 + S_3 + S_4 + R_1 + R_2 + R_3 + R_4 + R_5 + R_6. \tag{B.19}
\end{aligned}$$

To study the uniform consistency of  $CV(B, d, h)$  and  $ECV(B, d, h)$ , we consider the following two scenarios.

**Case 1:**  $\mathcal{S}_{T|U} \subseteq \mathcal{S}(B)$ .

In this case, we have that  $F(y, B_0^\top u) = F(y, B^\top u)$  (e.g., [13]). Thus, we immediately have that

$$S_2 = R_1 = R_4 = R_5 = 0.$$

Since  $S_1$  is the form of U-statistic, then applying the convergence property of U-statistic (e.g., [23, Ch 12]) gives that as  $n \rightarrow \infty$ ,

$$S_1 \xrightarrow{p} \sigma_0^2. \tag{B.20}$$

By Lemmas B.1 and B.2,  $S_3$  can be expressed as

$$\begin{aligned}
S_3 &= \left\{ \frac{1}{n^2(n-1)^2} \sum_{i \neq j_1 \neq j_2} \sum_{k=1}^n \zeta_{j_1, B}(Y_k, B^\top U_i) \zeta_{j_2, B}(Y_k, B^\top U_i) \right. \\
&\quad \left. + \frac{1}{n^2(n-1)^2} \sum_{i \neq j} \sum_{k=1}^n \zeta_{j, B}^2(Y_k, B^\top U_i) \right\} \{1 + o_p(1)\} \\
&\triangleq \{K_1(B) + K_2(B)\} \{1 + o_p(1)\}.
\end{aligned}$$

Since the class  $\{\zeta_{j_1, B}(y, u)\}$  is Euclidean ([20, Theorems 2.13 and 2.14]), then by the derivations similar to Theorem of [13], we can show that

$$\sup_B |K_1(B) - E\{\mathcal{B}_B^2(Y_k, B^\top U_i)\}| = o_p(1). \tag{B.21}$$

and

$$\sup_B |K_2(B) - E\{\mathcal{V}_B(Y_k, B^\top U_i)\}| = o_p(1). \tag{B.22}$$

As a result, by (B.21) and (B.22), we have

$$\sup_B |S_3 - AMISE_B(h)| = o_p\left(dh^{2q} + \frac{1}{nh^d}\right). \tag{B.23}$$

On the other hand, for the parameter corresponding to measurement error model, if  $L$  is known, then  $S_4 = R_3 = R_6 = 0$ . If  $L$  is unknown and  $\widehat{L}$  is the corresponding estimator, then

$$\begin{aligned} S_4 &= \frac{1}{n} \sum_{i=1}^n \int \left\{ \widehat{F}^{(-i)}(y, B^\top U_i) - \widehat{F}^{(-i)}(y, B^\top \widehat{U}_i) \right\}^2 dF_Y(y) \\ &= \frac{1}{n} \sum_{i=1}^n \int \left\{ \nabla_L^1 \widehat{F}^{(-i)}(y, B^\top U_i) \right\}^2 dF_Y(y) (\widehat{L} - L)^2. \end{aligned} \tag{B.24}$$

As shown in [14],  $\widehat{L}$  is the estimator of  $L$  with  $\sqrt{m}$ -rate. It means that  $\widehat{L} - L = O_p\left(\frac{1}{\sqrt{m}}\right)$ . As a result, applying Lemma B.1 gives  $S_4 = o_p(1)$ .

For  $R_3$  and  $R_6$ , similar to the derivations, we have that  $\sup_B |R_3| = \sup_B |R_6| = o_p(1)$ . Combining all results (B.20) – (B.24) with (B.19), we have that

$$\begin{aligned} CV(B, d, h) &= \sigma_0^2 + S_3 + o_p(1) \\ &= \sigma_0^2 + S_3 - AMISE(h) + AMISE_B(h) + o_p(1) \\ &= ECV(B, d, h) + \{S_3 - AMISE_B(h)\} + o_p(1), \end{aligned}$$

where  $ECV(B, d, h) = \sigma_0^2 + AMISE_B(h)$ . Consequently, by (B.23) and similar derivations of Proposition 2 in [13] that  $AMISE_B(h) = O_p\left(dh^{2q} + \frac{1}{nh^d}\right)$ , we have that

$$\sup_{d, B, h} \left| \frac{CV(B, d, h) - ECV(B, d, h)}{AMISE_B(h)} \right| = o_p(1).$$

**Case 2:**  $S_{T|U} \not\subseteq \mathcal{S}(B)$ .

In this case, the derivations of  $S_1, S_3, R_2, R_3$ , and  $R_6$  can be determined in Case 1, but  $S_2, R_1, R_4$ , and  $R_5$  do not equal zero. The main goal in this case is to discuss those remaining parts.

Noting that by Theorem 3.1 in [2], we have that

$$\sup_B |S_2 - b_0^2(B)| = o_p(1),$$

which is equivalent to

$$\sup_B \left| \frac{S_2}{b_0^2(B)} - 1 \right| = o_p(1). \tag{B.25}$$

By the Cauchy-Schwarz inequality, we have that  $R_1^2 \leq S_1 S_2$ . Then dividing  $b_0(B)$  on both sides and applying (B.20) and (B.25) yield

$$\sup_B \frac{|R_1|}{b_0(B)} = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Furthermore, by the Cauchy-Schwarz inequality again, Lemma B.1 and (B.25), we have that

$$\sup_B \frac{|R_4|}{b_0(B)} = O_p(h^q) + o_p\left(\frac{\log n}{\sqrt{nh^d}}\right)$$



and

$$\sup_B \frac{|R_5|}{b_0(B)} = o_p \left( \frac{p \log m}{\sqrt{m}} \right).$$

Consequently, combining those results with (B.19) yields

$$\sup_{B,d,h} \frac{|CV(B, d, h) - ECV(B, d, h)|}{b_0(B)AMISE(h)} = O_p(1),$$

where  $ECV(B, h) = \sigma_0^2 + b_0^2(B) + AMISE(h)$ . □

**Appendix C: Proofs of Theorems in Section 4**

*Appendix C.1. Proof of Theorem 4.1*

Note that  $\widehat{L}$  is the consistent estimator in the sense that  $\widehat{L} = L + o_p(1)$ , then we have that  $\widehat{U}_i = U_i + o_p(1)$ . Hence, in the remaining proof, we focus on  $U_i$ . In addition, we divide this proof into two steps. In step 1, we discuss the consistency of  $\widehat{B}$ , and then discuss the asymptotic performances of  $\widehat{h}$  and  $\widehat{d}$  in Step 2.

**Step 1:** The consistency of  $\widehat{B}$ .

Let  $\widehat{B}$  denote the minimizer of  $CV(B, d, h_0)$  with  $d \geq d_0$  and  $h = h_0$ . Then we have that

$$P \left\{ CV(\widehat{B}, d, h_0) < CV(B_0, d, h_0) \right\} = 1. \tag{C.1}$$

Let  $DCV(B, d, h_0) = |CV(B, d, h_0) - ECV(B, d, h_0)|$ . For any  $\xi > 0$ , we further have that

$$\begin{aligned} & \left\{ CV(\widehat{B}, d, h_0) < CV(B_0, d, h_0) \right\} \\ = & \left\{ b_0^2(\widehat{B}) < \xi, b_0^2(\widehat{B}) > \xi, CV(\widehat{B}, d, h_0) < CV(B_0, d, h_0) \right\} \\ \subseteq & \left\{ b_0^2(\widehat{B}) < \xi \right\} \\ & \cup \left\{ b_0^2(\widehat{B}) > \xi, DCV(\widehat{B}, d, h_0) + DCV(B_0, d, h_0) \right. \\ & \left. > ECV(\widehat{B}, d, h_0) - ECV(B_0, d, h_0) \right\}. \end{aligned} \tag{C.2}$$

By (C.1) and (C.2), we have that

$$\begin{aligned} 1 \leq & P \left\{ b_0^2(\widehat{B}) < \xi \right\} \\ & + P \left\{ b_0^2(\widehat{B}) > \xi, DCV(\widehat{B}, d, h_0) + DCV(B_0, d, h_0) \right. \\ & \left. > ECV(\widehat{B}, d, h_0) - ECV(B_0, d, h_0) \right\}. \end{aligned} \tag{C.3}$$

In particular, for the second term of (C.3), we further have that

$$P \left\{ b_0^2(\widehat{B}) > \xi, DCV(\widehat{B}, d, h_0) + DCV(B_0, d, h_0) \right.$$

$$\begin{aligned}
&> ECV(\widehat{B}, d, h_0) - ECV(B_0, d, h_0) \} \\
= & P \left\{ b_0^2(\widehat{B}) > \xi, \frac{DCV(\widehat{B}, d, h_0)}{b_0(\widehat{B})} + \frac{DCV(B_0, d, h_0)}{b_0(\widehat{B})} \right. \\
&> \left. \frac{ECV(\widehat{B}, d, h_0)}{b_0(\widehat{B})} - \frac{ECV(B_0, d, h_0)}{b_0(\widehat{B})} \right\} \\
\leq & P \left\{ b_0^2(\widehat{B}) > \xi, \frac{DCV(\widehat{B}, d, h_0)}{\sqrt{\xi}} + \frac{DCV(B_0, d, h_0)}{\sqrt{\xi}} \right. \\
&> \left. \frac{ECV(\widehat{B}, d, h_0)}{\sqrt{\xi}} - \frac{ECV(B_0, d, h_0)}{\sqrt{\xi}} \right\},
\end{aligned} \tag{C.4}$$

where the last step is due to the consideration of the event  $b_0^{-1}(\widehat{B}) < \xi^{-1/2}$ .

Moreover, since the event  $b_0^2(\widehat{B}) > \xi > 0$  in (C.4) implies  $\mathcal{S}_{T|U} \not\subseteq \mathcal{S}(B)$ , then  $ECV(B, d, h) = \sigma_0^2 + b_0^2(B) + AMISE(h)$ . Therefore, (C.4) becomes

$$\begin{aligned}
& P \left\{ b_0^2(\widehat{B}) > \xi, DCV(\widehat{B}, d, h_0) + DCV(B_0, d, h_0) \right. \\
&> \left. ECV(\widehat{B}, d, h_0) - ECV(B_0, d, h_0) \right\} \\
\leq & P \left\{ b_0^2(\widehat{B}) > \xi, O_p \{ AMISE_{\widehat{B}}(h) \} + o_p \{ AMISE_{B_0}(h) \} \right. \\
&> \left. \sqrt{\xi} + \frac{AMISE_{\widehat{B}}(h) - AMISE_{B_0}(h)}{\sqrt{\xi}} \right\} \\
\rightarrow & P \left\{ b_0^2(\widehat{B}) > \xi, 0 > \sqrt{\xi} \right\} \\
= & 0
\end{aligned} \tag{C.5}$$

as  $n \rightarrow \infty$ , where the first step is due to (B.17) and (B.18), and the second step is due to both  $O_p \{ AMISE_{B_0}(h) \} \rightarrow 0$  and  $o_p \{ AMISE_{B_0}(h) \} \rightarrow 0$  as  $n \rightarrow \infty$ . As a result, combining (C.3) and (C.5), we have that as  $n \rightarrow \infty$ ,

$$P \left\{ b_0^2(\widehat{B}) < \xi \right\} \rightarrow 1. \tag{C.6}$$

Therefore, by (C.6), we have that as  $n \rightarrow \infty$ ,

$$\widehat{B} \xrightarrow{p} B_0.$$

**Step 2:** The asymptotic performance of  $(\widehat{d}, \widehat{h})$ .

In (C.6), set  $\xi = \inf_{B:d < d_0} b_0^2(B)$ . Then we have that

$$P \left\{ b_0^2(\widehat{B}) < \inf_{B:d < d_0} b_0^2(B) \right\}$$

$$\begin{aligned}
&\leq P\left\{b_0^2(\widehat{B}) < \inf_{B:d < d_0} b_0^2(B), \widehat{d} < d_0\right\} + P\left\{b_0^2(\widehat{B}) < \inf_{B:d < d_0} b_0^2(B), \widehat{d} \geq d_0\right\} \\
&\leq P\left\{\widehat{d} \geq d_0\right\} \\
&\leq 1,
\end{aligned} \tag{C.7}$$

where the second inequality is due to the fact

$$P\left\{b_0^2(\widehat{B}) < \inf_{B:d < d_0} b_0^2(B), \widehat{d} \geq d_0\right\} \leq P\left\{\widehat{d} \geq d_0\right\}$$

as well as that  $P\left\{b_0^2(\widehat{B}) < \inf_{B:d < d_0} b_0^2(B), \widehat{d} < d_0\right\} = 0$  by the definition of  $\widehat{B}$ . Applying (C.6) to (C.7) gives that as  $n \rightarrow \infty$ ,

$$P\left\{\widehat{d} \geq d_0\right\} \rightarrow 1. \tag{C.8}$$

Furthermore, define

$$\begin{aligned}
W_1 &= \left\{b_0^2(\widehat{B}) < \frac{\log n}{n}, \widehat{d} = d_0, \left|\frac{\widehat{h}}{h_0} - 1\right| < \eta\right\}, \\
W_2 &= \left\{b_0^2(\widehat{B}) \geq \frac{\log n}{n}\right\}, \\
W_3 &= \left\{\widehat{d} < d_0\right\}, \\
W_4 &= \left\{b_0^2(\widehat{B}) < \frac{\log n}{n}, \widehat{d} \geq d_0, \left|\frac{\widehat{h}}{h_0} - 1\right| \geq \eta\right\}, \\
W_5 &= \left\{b_0^2(\widehat{B}) < \frac{\log n}{n}, \widehat{d} > d_0, \left|\frac{\widehat{h}}{h_0} - 1\right| < \eta\right\},
\end{aligned}$$

and

$$\begin{aligned}
W_{CV} &= \left\{DCV(\widehat{B}, d_0, h_0) + DCV(B_0, d_0, h_0) \right. \\
&\quad \left. > ECV(\widehat{B}, d_0, h_0) - ECV(B_0, d_0, h_0)\right\}.
\end{aligned}$$

By (C.8) and the partition of events  $W_{CV}$  and  $W_k$  for  $k = 1, \dots, 5$ , we have that

$$1 \leq P(W_1) + \sum_{k=2}^5 P(W_k \cap W_{CV}). \tag{C.9}$$

For  $k = 2$ , applying (C.6) with  $\xi = \frac{\log n}{n}$  gives that as  $n \rightarrow \infty$ ,

$$0 < P(W_2 \cap W_{CV}) \leq P(W_2) \rightarrow 0. \tag{C.10}$$

For  $k = 3$ , applying (C.8) gives that as  $n \rightarrow \infty$ ,

$$0 < P(W_3 \cap W_{CV}) \leq P(W_3) \rightarrow 0. \quad (\text{C.11})$$

For  $k = 4$ , we have that as  $n \rightarrow \infty$ ,

$$P(W_4 \cap W_{CV}) \rightarrow 0. \quad (\text{C.12})$$

For  $k = 5$ , we have that as  $n \rightarrow \infty$ ,

$$P(W_5 \cap W_{CV}) \rightarrow 0. \quad (\text{C.13})$$

Therefore, combining (C.10)-(C.13) with (C.9) yields that as  $n \rightarrow \infty$ ,

$$P \left\{ b_0^2(\hat{B}) < \frac{\log n}{n}, \hat{d} = d_0, \left| \frac{\hat{h}}{h_0} - 1 \right| < \eta \right\} \rightarrow 1,$$

and we conclude that  $(\hat{d}, \hat{h})$  are consistent estimators.  $\square$

### Appendix C.2. Proof of Theorem 4.2

Note that  $\hat{L}$  is a consistent estimator in the sense that  $\hat{L} = L + o_p(1)$ , then we have that  $\hat{U}_i = U_i + o_p(1)$ . Hence, in the remaining proof, we focus on  $U_i$ .

By adding and subtracting an additional term  $\hat{F}(y, B_0^\top u)$ , we have that

$$\begin{aligned} & \hat{F}(y, \hat{B}^\top u) - F(y, B_0^\top u) \\ &= \left\{ \hat{F}(y, \hat{B}^\top u) - \hat{F}(y, B_0^\top u) \right\} + \left\{ \hat{F}(y, B_0^\top u) - F(y, B_0^\top u) \right\} \\ &\triangleq A_1 + A_2. \end{aligned} \quad (\text{C.14})$$

For  $A_1$ , applying the first order Taylor series expansion at  $B_0$  gives

$$\begin{aligned} A_1 &= \hat{F}^{(1)}(y, B_0^\top u) (\hat{B} - B_0) \\ &= \left\{ \hat{F}^{(1)}(y, B_0^\top u) - F^{(1)}(y, B_0^\top u) \right\} (\hat{B} - B_0) + F^{(1)}(y, B_0^\top u) (\hat{B} - B_0) \\ &= o_p(1), \end{aligned} \quad (\text{C.15})$$

where the third equality is due to Lemma B.1 and Theorem 4.1. On the other hand, applying Lemma B.2 with  $j = 0$  for  $A_2$  gives that

$$\sup_{y, u} \left| \hat{F}(y, B_0^\top u) - F(y, B_0^\top u) - \frac{1}{n} \sum_{i=1}^n \zeta_{i, B_0}^{(0)}(y, u) \right| = O_p \left( \frac{1}{\sqrt{n}} \right). \quad (\text{C.16})$$

Finally, combining (C.15) and (C.16) with (C.14) gives the desired result in Theorem 4.2.  $\square$

### Appendix C.3. Proof of Theorem 4.3

Let  $\widehat{B}$  denote the minimizer of  $CV(B, d_0, h_0)$  which also satisfies  $\nabla_{\text{vec}(B)}^1 CV(\widehat{B}, d_0, h_0) = 0$ . Then by the first order Taylor series expansion, we have that

$$\begin{aligned} 0 &= \nabla_{\text{vec}(B)}^1 CV(\widehat{B}, d_0, h_0) \\ &= \nabla_{\text{vec}(B)}^1 CV(B_0, d_0, h_0) + \nabla_{\text{vec}(B)}^2 CV(B^*, d_0, h_0) \left\{ \text{vec}(\widehat{B}) - \text{vec}(B_0) \right\}, \end{aligned}$$

where  $B^*$  is “between”  $\widehat{B}$  and  $B_0$ . Equivalently, we have that

$$\begin{aligned} &\sqrt{n} \left\{ \text{vec}(\widehat{B}) - \text{vec}(B_0) \right\} \\ &= \left\{ \nabla_{\text{vec}(B)}^2 CV(B^*, d_0, h_0) \right\}^{-1} \sqrt{n} \nabla_{\text{vec}(B)}^1 CV(B_0, d_0, h_0). \quad (\text{C.17}) \end{aligned}$$

We first discuss the asymptotic result of  $\nabla_{\text{vec}(B)}^2 CV(B^*, d_0, h_0)$ . By Theorem 4.1, we have that  $B^* \xrightarrow{p} B_0$  as  $n \rightarrow \infty$ . By simple calculations,  $\nabla_{\text{vec}(B)}^2 CV(B_0, d_0, h_0)$  can be written as

$$\begin{aligned} &\nabla_{\text{vec}(B)}^2 CV(B_0, d_0, h_0) \\ &= \frac{2}{n} \sum_{i=1}^n \int_0^\tau \left[ \left\{ \nabla_{\text{vec}(B)}^1 \widehat{F}^{(-i)}(y, B_0^\top U_i) \right\}^{\otimes 2} \right. \\ &\quad \left. - \nabla_{\text{vec}(B)}^2 \widehat{F}^{(-i)}(y, B_0^\top U_i) \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) - \widehat{F}^{(-i)}(y, B_0^\top U_i) \right\} \right] d\widehat{F}_Y(y). \end{aligned}$$

By Lemma B.1 and the Law of Large Numbers, we have that as  $n \rightarrow \infty$ ,

$$\nabla_{\text{vec}(B)}^2 CV(B_0, d_0, h_0) \xrightarrow{p} \mathcal{A}, \quad (\text{C.18})$$

where

$$\begin{aligned} \mathcal{A} &= 2E \left( \int_0^\tau \left[ \left\{ F^{(1)}(y, B_0^\top U_i) \right\}^{\otimes 2} \right. \right. \\ &\quad \left. \left. - F^{(2)}(y, B_0^\top U_i) \left\{ \frac{\Delta_i}{G(Y_i)} \mathbb{I}(Y_i \leq y) - F(y, B_0^\top U_i) \right\} \right] dF_Y(y) \right). \end{aligned}$$

On the other hand, if  $L$  is known, then  $\nabla_{\text{vec}(B)}^1 CV(B_0, d_0, h_0)$  can be expressed as

$$\begin{aligned} &\nabla_{\text{vec}(B)}^1 CV(B_0, d_0, h_0) \\ &= \frac{2}{n} \sum_{i=1}^n \int_0^\tau \left[ \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) - \widehat{F}^{(-i)}(y, B_0^\top U_i) \right\} \right. \end{aligned}$$

$$\begin{aligned}
 & \times \left\{ -\nabla_{\text{vec}(B)}^1 \widehat{F}^{(-i)}(y, B_0^\top U_i) \right\} \Big] d\widehat{F}_Y(y) \\
 = & \frac{2}{n} \sum_{i=1}^n \int_0^\tau \left[ \left\{ \frac{\Delta_i}{\widehat{G}(Y_i)} \mathbb{I}(Y_i \leq y) - F(y, B_0^\top u) + F(y, B_0^\top u) \right. \right. \\
 & \left. \left. - \widehat{F}^{(-i)}(y, B_0^\top U_i) \right\} \right. \\
 & \left. \times \left\{ -\nabla_{\text{vec}(B)}^1 \widehat{F}^{(-i)}(y, B_0^\top U_i) + F^{(1)}(y, B_0^\top u) - F^{(1)}(y, B_0^\top u) \right\} \right] d\widehat{F}_Y(y) \\
 = & -\frac{2}{n} \sum_{i=1}^n \int_0^\tau \mathcal{E}_{1;iy} F^{(1)}(y, U_i) d\widehat{F}_Y(y) + \frac{2}{n} \sum_{i=1}^n \int_0^\tau \mathcal{E}_{3;iy} F^{(1)}(y, U_i) d\widehat{F}_Y(y) \\
 & - \frac{2}{n} \sum_{i=1}^n \int_0^\tau \mathcal{E}_{1;iy} \left\{ \nabla_{\text{vec}(B)}^1 \widehat{F}^{(-i)}(y, B_0^\top U_i) - F^{(1)}(y, U_i) \right\} d\widehat{F}_Y(y),
 \end{aligned}$$

where the second equality comes from adding and subtracting additional terms  $F(y, B_0^\top u)$  and  $F^{(1)}(y, B_0^\top u)$ , and the last step is due to

$$\begin{aligned}
 & \frac{2}{n} \sum_{i=1}^n \int_0^\tau \left[ \left\{ \widehat{F}^{(-i)}(y, B_0^\top U_i) - F(y, B_0^\top u) \right\} \right. \\
 & \left. \times \left\{ \nabla_{\text{vec}(B)}^1 \widehat{F}^{(-i)}(y, B_0^\top U_i) - F^{(1)}(y, U_i) \right\} \right] d\widehat{F}_Y(y) \\
 = & \frac{1}{n} \nabla_{\text{vec}(B)}^1 \sum_{i=1}^n \int_0^\tau \left\{ \widehat{F}^{(-i)}(y, B_0^\top U_i) - F(y, B_0^\top u) \right\}^2 d\widehat{F}_Y(y) \\
 = & 0.
 \end{aligned}$$

By the result in Lemma B.2,  $\nabla_{\text{vec}(B)}^1 CV(B_0, d_0, h_0)$  can be further written as

$$\begin{aligned}
 & \nabla_{\text{vec}(B)}^1 CV(B_0, d_0, h_0) \\
 = & -\frac{2}{n^2} \sum_{i \neq j} \mathcal{E}_{1;iy} F^{(1)}(Y_j, U_i) + \frac{2}{n^2(n-1)} \sum_{i \neq j \neq k} F^{(1)}(Y_j, U_i) \zeta_{j,B}^{(0)}(Y_k, U_i) \\
 & - \frac{2}{n^2(n-1)} \sum_{i \neq j \neq k} \mathcal{E}_{1;iy} \zeta_{j,B}^{(1)}(Y_k, U_i) + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 \triangleq & T_1 + T_2 + T_3 + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{C.19}
 \end{aligned}$$

For  $T_1$ , applying the convergence property of U-statistic in [12], we have that as  $n \rightarrow \infty$ ,

$$\sqrt{n}T_1 \xrightarrow{d} N(0, \mathcal{B}), \tag{C.20}$$

where  $\mathcal{B} = E \{U^{\otimes 2}(B_0)\}$  with

$$U(B_0) = \int_0^\tau \left\{ \frac{\Delta_i}{G(Y_i)} \mathbb{I}(Y_i \leq y) - F(y, B_0^\top U_i) \right\} F^{(1)}(y, U_i) dF_Y(y).$$

Since  $T_2$  and  $T_3$  contain  $\zeta_{i,B_0}^{(0)}(y, u)$  and  $\zeta_{i,B_0}^{(1)}(y, u)$  in (4.4) and (B.8), respectively, and involve  $\tilde{\mathbb{F}}_{i,l,B_0,L}^{(j)}(y, u)$  with  $j = 0, 1$ . Then by Lemma B.1, we can show that

$$\sqrt{n}T_k \xrightarrow{p} 0 \text{ for } k = 2, 3. \quad (\text{C.21})$$

Consequently, combining (C.20) and (C.21) with (C.19), we have that as  $n \rightarrow \infty$ ,

$$\nabla_{\text{vec}(B)}^1 CV(B_0, d_0, h_0) \xrightarrow{d} N(0, \mathcal{B}). \quad (\text{C.22})$$

Finally, combining (C.18) and (C.22) with (C.17) gives that as  $n \rightarrow \infty$ ,

$$\sqrt{n} \left\{ \text{vec}(\hat{B}) - \text{vec}(B_0) \right\} \xrightarrow{d} N(0, \mathcal{A}^{-1} \mathcal{B} \mathcal{A}^{-1}).$$

If  $L$  is unknown and  $\hat{L}$  is the estimator, then  $\nabla_{\text{vec}(B)}^1 CV(B_0, d_0, h_0)$  can be written as

$$\sqrt{n} \nabla_{\text{vec}(B)}^1 CV(B_0, d_0, h_0) = \sqrt{n}T_1 + \sqrt{n}T_2 + \sqrt{n}T_3 + \sqrt{n}T_4 + o_p(1),$$

where  $\sqrt{n}T_k$  with  $k = 1, 2, 3$  have been derived in (C.20) and (C.21), and  $T_4$  is

$$\begin{aligned} T_4 &= \frac{2}{n} \sum_{i=1}^n \int_0^\tau \mathcal{E}_{4;iy} F^{(1)}(y, B_0^\top U_i) d\hat{F}_Y(y) \\ &= \frac{2}{n} \sum_{i=1}^n \int_0^\tau \left\{ \hat{F}^{(-i)}(y, B_0^\top U_i) - \hat{F}^{(-i)}(y, B_0^\top \hat{U}_i) \right\} F^{(1)}(y, B_0^\top U_i) d\hat{F}_Y(y) \\ &= -\frac{2}{n} \sum_{i=1}^n \int_0^\tau \left\{ \nabla_{\text{vec}(B_0)}^1 \hat{F}^{(-i)}(y, B_0^\top U_i) \right\} F^{(1)}(y, B_0^\top U_i) d\hat{F}_Y(y) (\hat{L} - L). \end{aligned} \quad (\text{C.23})$$

If  $\hat{L}$  is determined by repeated measurements, then

$$\begin{aligned} \hat{L} - L &= \left( I_{s \times s} - \hat{\Sigma}_\epsilon \hat{\Sigma}_{X^*}^{-1} \right) - \left( I_{s \times s} - \Sigma_\epsilon \Sigma_{X^*}^{-1} \right) \\ &= - \left( \hat{\Sigma}_\epsilon \hat{\Sigma}_{X^*}^{-1} - \Sigma_\epsilon \Sigma_{X^*}^{-1} \right) \\ &= -\Sigma_{X^*}^{-1} \left( \hat{\Sigma}_\epsilon - \Sigma_\epsilon \right) \\ &= -\Sigma_{X^*}^{-1} \frac{1}{2m} \sum_{i=1}^m \left\{ (x_{i1}^* - x_{i2}^*) (x_{i1}^* - x_{i2}^*)^\top - 2\Sigma_\epsilon \right\} + o_p(1), \end{aligned}$$

and if  $\hat{L}$  is obtained by validation data, then

$$\hat{L} - L = \text{cov}(\widehat{X}_i, X_i^*) \hat{\Sigma}_{X^*}^{-1} - \text{cov}(X_i, X_i^*) \Sigma_{X^*}^{-1}$$

$$\begin{aligned}
&= -\Sigma_{X^*}^{-1} \left\{ \text{cov}(\widehat{X}_i, X_i^*) - \text{cov}(X_i, X_i^*) \right\} \\
&= -\Sigma_{X^*}^{-1} \frac{1}{m} \sum_{i=1}^m \left\{ (x_i - \mu_X) (x_i^* - \mu_{X^*})^\top - \Sigma_{XX^*} \right\} + o_p(1),
\end{aligned}$$

where  $\Sigma_{XX^*} = \text{cov}(X_i, X_i^*)$ .

Applying Lemma B.1 and combining (C.23) with the expression of  $\widehat{L} - L$  give that

$$\sqrt{n}T_4 = \frac{\sqrt{n}}{m} \sum_{i=1}^m \mathcal{T}(B_0)\Phi_i + o_p(1), \quad (\text{C.24})$$

where

$$\mathcal{T}(B_0) = E \left[ \int_0^\tau \left\{ F^{(1)}(y, B_0^\top U_i) \right\}^{\otimes 2} dF_Y(y) \right] \Sigma_{X^*},$$

$\Phi_i = (x_{i1}^* - x_{i2}^*)(x_{i1}^* - x_{i2}^*)^\top - 2\Sigma_\epsilon$  if  $L$  is estimated based on repeated measurements, and  $\Phi_i = (x_i - \mu_X)(x_i^* - \mu_{X^*})^\top - \Sigma_{XX^*}$  if  $L$  is estimated based on validation data.

Therefore, combining (C.20), (C.21), and (C.24) with (C.23) gives that as  $n \rightarrow \infty$ ,

$$\nabla_{\text{vec}(B)}^1 CV(B_0, d_0, h_0) \xrightarrow{d} N(0, \mathcal{B}_L), \quad (\text{C.25})$$

where  $\mathcal{B}_L = E \left[ \{U(B_0) + \mathcal{T}(B_0)\Phi_i\}^{\otimes 2} \right]$ . As a result, by the Slutsky Theorem on (C.18) and (C.25), we have that as  $n \rightarrow \infty$ ,

$$\sqrt{n} \left\{ \text{vec}(\widehat{B}) - \text{vec}(B_0) \right\} \xrightarrow{d} N(0, \mathcal{A}^{-1} \mathcal{B}_L \mathcal{A}^{-1}). \quad \square$$

## Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and partially supported by a Collaborative Research 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.

## References

- [1] Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10, 1100–1120. [MR0673646](#)
- [2] Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *The Annals of Probability*, 21, 1494–1542. [MR1235426](#)
- [3] Cai, J., Fan, J., Jiang, J., Zhou, H. (2006). Partially linear hazard regression for multivariate survival data. *Journal of the American Statistical Association*, 102, 538–551. [MR2370851](#)



- [4] Carroll, R. J. and Li, K.-C. (1992). Measurement error regression with unknown link: dimension reduction and data visualization. *Journal of the American Statistical Association*, 87, 1040–1050. [MR1209565](#)
- [5] Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Model*. Chapman & Hall/CRC [MR2243417](#)
- [6] Cook, R. D. (1994). On the interpretation of regression plots. *Journal of the American Statistical Association*, 89, 177–189. [MR1266295](#)
- [7] Cook, R. D. (1998). *Regression Graphics: Ideas for Studying Regressions through Graphics*. Wiley, New York. [MR1645673](#)
- [8] Eftekhari, H., Banerjee, M., and Ritov, Y. (2021). Inference in high-dimensional single-index models under symmetric designs. *Journal of Machine Learning Research*, 22, 1–63. [MR4253720](#)
- [9] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré*, 38, 907–921. [MR1955344](#)
- [10] Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundaker, H., Schooley, R. T., Haubrich, R. H. et al. (1996). 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*, 335, 1081–1090.
- [11] He, W. and Yi, G. Y. (2020). Parametric and semiparametric estimation methods for survival data under a flexible class of models. *Lifetime Data Analysis*, 26, 369–388. [MR4079671](#)
- [12] Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19, 293–325. [MR0026294](#)
- [13] Huang, M.-Y. and Chiang, C.-T. (2017). An effective semiparametric estimation approach for the sufficient dimension reduction model. *Journal of the American Statistical Association*, 112, 1296–1310. [MR3735378](#)
- [14] Li, B. and Yin, X. (2007). On surrogate dimension reduction for measurement error regression: an invariance law. *The Annals of Statistics*, 35, 2143–2172. [MR2363966](#)
- [15] Li, B. (2018). *Sufficient Dimension Reduction: Methods and Applications with R*. CRC Press. [MR3838449](#)
- [16] Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86, 316–327. [MR1137117](#)
- [17] Li, K.-C., Wang, J.-L., and Chen, C.-H. (1999). Dimension reduction for censored regression data. *The Annals of Statistics*, 27, 1–23. [MR1701098](#)
- [18] Lu, W. and Li, L. (2011). Sufficient dimension reduction for censored regressions. *Biometrics*, 67, 513–523. [MR2829020](#)
- [19] Ma, Y. and Zhu, L. (2013). Efficient estimation in sufficient dimension reduction. *The Annals of Statistics*, 41, 250–268. [MR3059417](#)
- [20] Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57, 1027–1057. [MR1014540](#)
- [21] Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New

- York. [MR0762984](#)
- [22] Rubin, D. and van der Laan, M. J. (2007). A doubly robust censoring unbiased transformation. *The International Journal of Biostatistics*, 3, 1–21 [MR2306842](#)
- [23] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York. [MR1652247](#)
- [24] Wang, H. and Xia, Y. (2008). Sliced regression for dimension reduction. *Journal of the American Statistical Association*, 103, 811–821. [MR2524332](#)
- [25] Xia, Y., Zhang, D., and Xu, J. (2010). Dimension reduction and semiparametric estimation of survival models. *Journal of the American Statistical Association*, 105, 278–290. [MR2656052](#)
- [26] Yi, G. Y. (2017). *Statistical Analysis with Measurement Error and Misclassification: Strategy, Method and Application*. Springer. [MR3676914](#)
- [27] Zhang, J., Zhu, L., and Zhu, L. (2014). Surrogate dimension reduction in measurement error regressions. *Statistica Sinica*, 24, 1341–1363. [MR3241291](#)
- [28] Zhu, L., Zhu, L. and Feng, Z. (2010). Dimension reduction in regression through cumulative slicing estimation. *Journal of the American Statistical Association*, 105, 1455–1466. [MR2796563](#)
- [29] Zhu, R. and Kosorok, M. R. (2012). Recursively imputed survival trees. *Journal of the American Statistical Association*, 107, 331–340. [MR2949363](#)