Electronic Journal of Statistics
Vol. 14 (2020) 4054–4109
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1762

Model selection and model averaging for analysis of truncated and censored data with measurement error*

Li-Pang Chen† **and Grace Y. Yi**†,‡,§

Abstract: Model selection plays a critical role in statistical inference and a large literature has been devoted to this topic. Despite extensive research attention on model selection, research gaps still remain. An important but relatively unexplored problem concerns truncated and censored data with measurement error. Although analysis of left-truncated and right-censored (LTRC) data has received extensive interests in survival analysis, there has been no research on model selection for LTRC data with measurement error. In this paper, we take up this important problem and develop inferential procedures to handle model selection for LTRC data with measurement error in covariates. Our development employs the local model misspecification framework ([6]; [10]) and emphasizes the use of the focus information criterion (FIC). We develop valid estimators using the model averaging scheme and establish theoretical results to justify the validity of our methods. Numerical studies are conducted to assess the performance of the proposed methods.

Keywords and phrases: Focus information criterion, left-truncation, measurement error, model averaging, model selection, survival analysis.

Received August 2019.

1. Introduction

Model selection plays an important role in statistical inference, and various model selection criteria have been proposed, including the Akaike information criterion (AIC), Bayesian information criterion (BIC), and cross validation. To focus on the quantity of interest, [6] proposed the focus information criterion (FIC) for model selection by selecting the best candidate model using the smallest mean squared error for the corresponding estimator of the focus parameter under regression models. Several extensions of the FIC method have been developed for distinct settings. For example, [5] studied the FIC method for the

*This research was supported by the Natural Sciences and Engineering Research Council of Canada and partially supported by a Collaborative Research Team Project of the Canadian Statistical Sciences Institute. Yi is Canada Research Chair in Data Science (Tier 1). Her research was undertaken, in part, thanks to funding from the Canada Research Chairs Program.

†Department of Statistics and Actuarial Science, University of Waterloo

‡Department of Statistical and Actuarial Sciences, Department of Computer Science, University of Western Ontario

§Corresponding author. Email: gyi5@uwo.ca.

partially linear model. [30] implemented the FIC method to generalized additive partial linear models. [26] discussed the FIC method based on weighted composite quantile regression.

Traditional statistical analysis often first builds the model by selecting important variables and then, based on the model, carries out statistical inferences. This procedure, however, as pointed out by [7] and [8], among others, ignores the uncertainty induced from the variable selection process, thus producing estimators with invalid characterization of the associated variability. To mitigate this issue, researchers came up with the *model averaging* strategy based on averaging a set of candidate models with suitable weights attached and then producing a compromise estimator of the model parameter accordingly. The specification of the weights basically hinges on the choice of the model selection criterion, which induces uncertainty to inferential procedures. Detailed discussions can be found in [7].

With censored data, variable selection methods were developed for various settings by many authors including [4], [9], [11], and [16]. While those extensions branch out the scope of the FIC method, they are inadequate in handling data with concurrent features of truncation, censoring, and measurement error, which commonly arise from many areas including clinical trials, epidemiological studies, actuarial science, and so on. Little work has been available to address those features simultaneously as noted by [27] and [29].

In this paper, we investigate this problem and develop valid inference methods which simultaneously accommodate measurement error effects and sampling issues as well as model building for censored survival data. Our model selection development takes the local model misspecification framework [6, 10] for which we modify the FIC criterion of [6] and [11] by accommodating the effects due to the concurrent presence of measurement error, left-truncation and right-censoring in data. To estimate the focus parameters, we adopt the frequentist model average framework parallel to [10] and [30]. We establish asymptotic results for the proposed estimators.

The remainder is organized as follows. In Section 2, we introduce the notation and model setup for LTRC data. In Section 3, we propose a pseudo likelihood for each candidate model using an adjusted conditional log-likelihood together with regression calibration. The asymptotic properties of the estimators under candidate models and the model averaging estimators are presented in Sections 4 and 5. Empirical studies are provided in Section 6. We conclude the article with discussion in the last section. Technical justifications are outlined in the appendices.

2. Notation and model

When we are interested in studying the survival process for a population of individuals having a disease, we may consider a prevalent cohort which consists of individuals having the disease at enrollment to the study. Suppose the disease progression involves two chronologically ordered events, called the *initiating* *event* and the *failure event* (e.g., [24]). For an individual in the target population, let u and r denote the calendar times for the initiating event and the failure event, respectively, where $u < r$. We define $\widetilde{T} = r - u$ as the lifetime since the initiating event, which is of prime interest. In the process of collecting data, only those individuals who have experienced the initiating event can be recruited in the study, i.e., the study subjects can only include those with $u < \xi$, where ξ represents the calendar time of the recruitment. In this instance, $\widetilde{A} = \xi - u$ represents the truncation time, as shown in Figure 1. Let $h(a)$ be the probability density function of $\widetilde{A}$ which is typically unknown, and let $H(a) = \int_0^a h(u)du$ denote the corresponding distribution function.

For an individual with $\widetilde{T} \geq \widetilde{A}$, we let (A, T) denote $\left(\widetilde{A}, \widetilde{T}\right)$ to indicate that such an individual is eligible for the recruitment so that measuring (A, T) is possible. If $\widetilde{T} < \widetilde{A}$, then such an individual is not included in the study to contribute any information. In addition, we define C as the censoring time after the recruitment. That is, once individuals are enrolled in the study, one may collect either residual survival time $T - A$ or censoring time C. Let $Y = \min\{T, A+C\}$ be the observed time and let $\Delta = I(T \leq A + C)$ be the indicator of a failure event, where $I(\cdot)$ is the indicator function. For an individual in the study, let X and Z denote the associated covariates of dimension $p \times 1$ and $q \times 1$, respectively, and write $V = (X^\top, Z^\top)^\top$.

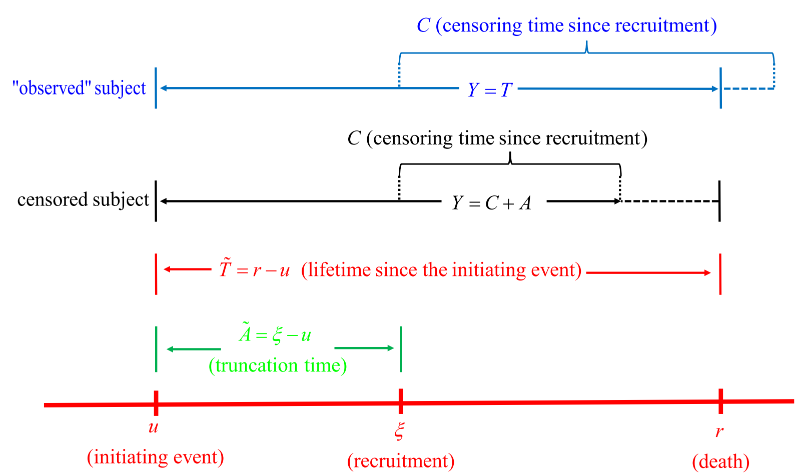

FIG 1. *Schematic depiction of LTRC data for* $\widetilde{T} \geq \widetilde{A}$

2.1. Cox model and inference

Suppose that we have a sample of n subjects and for $i = 1, \cdots, n$, $\{Y_i, A_i, \Delta_i, V_i\}$ has the same distribution as $\{Y, A, \Delta, V\}$. Let the lowercase letters y_i, a_i, δ_i, and $v_i = (x_i^\top, z_i^\top)^\top$ represent realizations of $Y_i, A_i, \Delta_i,$ and $V_i = (X_i^\top, Z_i^\top)^\top$, respectively. Consider the Cox model for survival times $\widetilde{T}$ whose hazard function is modeled as

$$\lambda(t|v_i) = \lambda_0(t) \exp(v_i^\top \beta), \tag{2.1}$$

where $\lambda_0(\cdot)$ is an unknown baseline hazard function, and β is the vector of the parameters that are of interest.

Let $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ be the cumulative baseline hazard function. Let $\mathcal{F}(t|v_i) = \exp\left\{-\Lambda_0(t)\exp\left(v_i^\top \beta\right)\right\}$ denote the survivor function of $\widetilde{T}$ given the covariates and let $f(t|v_i) = -\frac{d}{dt}\mathcal{F}(t|v_i)$. By Assumptions (C5) and (C6) in Appendix A, the likelihood function is given by

$$L = \prod_{i=1}^{n} \frac{f(y_i|v_i)^{\delta_i} \mathcal{F}(y_i|v_i)^{1-\delta_i} dH(a_i)}{\int_0^\infty \mathcal{F}(u|v_i)dH(u)}, \tag{2.2}$$

which can be equivalently re-written as the product of the conditional likelihood

$$L_{\mathrm{C}} = \prod_{i=1}^{n} \frac{f(y_i|v_i)^{\delta_i} \mathcal{F}(y_i|v_i)^{1-\delta_i}}{\mathcal{F}(a_i|v_i)} \tag{2.3}$$

and the marginal likelihood

$$L_{\mathrm{M}} = \prod_{i=1}^{n} \frac{\mathcal{F}(a_i|v_i)dH(a_i)}{\int_0^\infty \mathcal{F}(u|v_i)dH(u)}. \tag{2.4}$$

Discussion on this formulation can be found in [3] and [24]. In principle, estimation of the model parameters may proceed with maximizing L with respect to the model parameters.

2.2. Framework with submodels

In specifying the model (2.1) we include all the covariates in the model without discretion; irrelevant or unimportant covariates may be included in the model. To feature this, we consider the *local model misspecification* framework, initiated by [11]. Let Z_i represent the vector of important covariates that are always being included in the model, and let X_i represent the vector of covariates which may be subject to exclusion when building a model. Write $X_i = (X_{i1}, \ldots, X_{ip})^\top$ and $Z_i = (Z_{i1}, \ldots, Z_{iq})^\top$. Let $\beta = (\beta_x^\top, \beta_z^\top)^\top$ be a vector with dimension $d = p + q$, where β_x is the p-dimensional parameter vector for which we are unsure whether or not all of its components should be included in the model and β_z is the q-dimensional parameter vector which should be used in the model. Let $\beta_0 = \left(\frac{\eta^\top}{\sqrt{n}}, \beta_{z0}^\top\right)^\top$ represent the parameter value perturbed around the null model with the parameter value $(0_p^\top, \beta_{z0}^\top)^\top$, where η is the p-dimensional parameter, $\frac{\eta}{\sqrt{n}}$ represents the degree of the departure of the corresponding model from the null model, and 0_p is a p-dimensional zero vector.

Let $\mathcal{S}$ be the class of all subsets of $\{1, 2, \cdots, p\}$. For any $S \in \mathcal{S}$, let $|S|$ denote the number of the elements in S. Set S with $|S| = 0$ to be the empty set; if $|S| = p$, then such an S is called the *full* set. Let $\beta_S = \left(\beta_{x,S}^\top, \beta_{z,S}^\top\right)^\top$ denote the parameter vector for the candidate model S which corresponds to the covariates indexed by S, with $\beta_{x,S}$ being an $|S| \times 1$ subvector of β_x. Although covariate Z_i is always included in the model, the subscript S in $\beta_{z,S}$ is used to emphasize that this is the parameter corresponding to Z_i under the candidate model associated with S.

We now define a projection operator. For any S, let π_S be an $|S| \times p$ matrix with element 0 or 1; in each row there is one and only one element which takes value 1 and in each column there is at most one element taking value 1. More specifically, if $S = \left(j_1, j_2, \cdots, j_{|S|}\right)$ with $1 \leqslant j_1 < j_2 < \cdots < j_{|S|} \leqslant p$, then the (k, j_k) element of π_S takes value 1 for $k = 1, \cdots, |S|$; other elements of π_S take value 0. Let $\Pi_S = \begin{pmatrix} \pi_S & 0_{|S| \times q} \\ 0_{q \times p} & I_{q \times q} \end{pmatrix}$, where $0_{p \times q}$ is the $p \times q$ matrix with entries zero, and $I_{q \times q}$ is the $q \times q$ identity matrix. Then applying Π_S to $\left(X_i^\top, Z_i^\top\right)^\top$ gives the $(|S| + q) \times 1$ vector, $\Pi_S \left(X_i^\top, Z_i^\top\right)^\top$, which merely includes the covariates in the candidate model S.

2.3. Measurement error model

In applications, some covariates are subject to measurement error. Here we consider the case where X_i is error-contaminated and Z_i is precisely measured. For the case where Z_i is subject to measurement error and X_i is precisely measured, or both X_i and Z_i are subject to measurement error, the following development needs to be modified. If X_i is precisely measured but Z_i is error-prone, then the development here can be readily modified. But when both X_i and Z_i are error-contaminated, the technical details would be more notationally complex.

Let X_i^* denote the observed or surrogate measurement of X_i. Consistent with the most work in the literature (e.g., [2, 29]), we consider that the X_i are continuous and linked with the X_i^* by the additive measurement error model:

$$X_i^* = X_i + \epsilon_i, \tag{2.5}$$

where ϵ_i is independent of $\{Y_i, A_i, \Delta_i, V_i\}$, $\epsilon_i \sim N(0_p, \Sigma_\epsilon)$ with covariance matrix Σ_ϵ, and 0_p represents the $p \times 1$ zero vector. To highlight the key ideas, here we focus our attention on estimation of parameters associated with the survival model and assume Σ_ϵ for the measurement error model (2.5) to be known.

In situations where the parameters for the measurement error model (2.5) must be estimated, we may utilize the information carried with additional data sources such as repeated measurements or validation subsamples. For instance, with the availability of repeated measurements, let X_{ij}^* denote the jth repeated observation of X_i with $j = 1, \cdots, n_i$, where n_i is the number of the replicates for subject i. Define $\bar{X}_i^* = \frac{1}{n_i} \sum\limits_{j=1}^{n_i} X_{ij}^*$ for $i = 1, \cdots, n$. Then Σ_ϵ can be estimated as ([2], p.71)

$$\widehat{\Sigma}_\epsilon = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(X_{ij}^* - \bar{X}_i^*\right)\left(X_{ij}^* - \bar{X}_i^*\right)^\top}{\sum_{i=1}^{n}(n_i-1)}.$$

When validation data are available, one may adapt the discussion of [28] to incorporate estimation of the parameters for the measurement error model (2.5) into inferential procedures.

3. Methodology for the correction of measurement error effects

In this section we discuss estimation procedures which accommodate measurement error effects. We employ the framework considered by [3].

3.1. Correction for conditional log-likelihood function

For any candidate model S, let $L_{\mathrm{C},S}$ denote the derived conditional likelihood function, which, similar to the expression of L_{C} in (2.3), leads to its logarithm

$$\begin{aligned}\ell_{\mathrm{C},S} \quad = \quad & \sum_{i=1}^{n}\left[\delta_i \log\lambda_0(y_i) + \delta_i\left\{(\pi_S x_i)^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right\}\right. \\ & \left. - \left\{\Lambda_0(y_i) - \Lambda_0(a_i)\right\}\exp\left\{(\pi_S x_i)^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right\}\right], \qquad (3.1)\end{aligned}$$

showing that $(\pi_S X_i)^\top \beta_{x,S}$ and $\exp\left\{(\pi_S X_i)^\top \beta_{x,S}\right\}$ are the only terms involving error-prone covariates.

To correct for the measurement error effects, we first manipulate the measurement error model (2.5) as

$$\pi_S X_i^* = \pi_S X_i + \pi_S \epsilon_i, \qquad (3.2)$$

where $\pi_S\epsilon_i \sim N\left(0_{|S|}, \pi_S\Sigma_\epsilon\pi_S^\top\right)$, yielding the moment generating function $m_S(t) = E\left\{\exp\left(t^\top\pi_S\epsilon_i\right)\right\} = \exp\left(\frac{1}{2}t^\top\pi_S\Sigma_\epsilon\pi_S^\top t\right)$, where t is a $|S|\times 1$ vector of real numbers. Consequently,

$$E\left(\pi_S X_i^* | X_i, Z_i\right) = \pi_S X_i \qquad (3.3)$$

and

$$E\left\{\exp\left(\beta_{x,S}^\top\pi_S X_i^* - \frac{\beta_{x,S}^\top\pi_S\Sigma_\epsilon\pi_S^\top\beta_{x,S}}{2}\right)\middle| X_i, Z_i\right\} = \exp\left(\beta_{x,S}^\top\pi_S X_i\right). \qquad (3.4)$$

Define

$$\ell_{\mathrm{C},S}^* \quad = \quad \sum_{i=1}^{n}\left[\delta_i\log\lambda_0(y_i) + \delta_i\left(\left(\pi_S x_i^*\right)^\top\beta_{x,S} + z_i^\top\beta_{z,S}\right)\right. \qquad (3.5)$$

$$- \quad \{\Lambda_0(y_i) - \Lambda_0(a_i)\} \exp\left\{(\pi_S x_i^*)^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right\} \{m_S(\beta_{x,S})\}^{-1} \Bigg].$$

Then applying (3.3) and (3.4) to (3.5) yields that

$$E\left(\ell^*_{\mathrm{C},S} | X_i, Z_i\right) = \ell_{\mathrm{C},S};$$

this property ensures that working with the function $\ell^*_{\mathrm{C},S}$ allows us to recover the information carried by $\ell_{\mathrm{C},S}$. Such a strategy, called the "corrected" likelihood method or the insertion correction approach (e.g., [29], Chapter 2), is useful in yielding an unbiased estimating function, which captalizes on the identities (3.3) and (3.4) derived from the moment generating function for the error term in model (2.5).

$\ell^*_{\mathrm{C},S}$ differs from $\ell_{\mathrm{C},S}$ in the availability of measurements for the X_i^* and X_i. The inclusion of the term $\{m_S(\beta_{x,S})\}^{-1}$ in $\ell^*_{\mathrm{C},S}$ adjusts the effects of replacing X_i in $\ell_{\mathrm{C},S}$ with the surrogate version X_i^*. The function $\ell^*_{\mathrm{C},S}$ provides the basis for developing the following estimation procedure for β_S. With $(\beta_{x,S}^\top, \beta_{z,S}^\top)^\top$ fixed, maximizing (3.5) with respect to $\lambda_0(y_i)$, we derive the estimated cumulative baseline function as

$$\widehat{\Lambda}_{0,S}(t) = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n dN_i(u)}{\{m_S(\beta_{x,S})\}^{-1} G_S^{(0)}(u, \beta_{x,S}, \beta_{z,S})}, \tag{3.6}$$

where

$$G_S^{(0)}(u, \beta_{x,S}, \beta_{z,S}) = \frac{1}{n}\sum_{i=1}^n Y_i(u) \exp\left\{(\pi_S x_i^*)^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right\}, \tag{3.7}$$

$Y_i(u) = I(A_i \leqslant u \leqslant Y_i)$, and $N_i(u) = I(Y_i \leqslant u)$.

The estimator (3.6) is similar to the Breslow estimator (e.g., [15], p.385) in the formulation but involves additional terms $\{m_S(\beta_{x,S})\}^{-1}$ and $Y_i(u)$ to reflect the adjustments of the effects of measurement error and left-truncation. Plugging (3.6) into (3.5), we let $\widehat{\ell}^*_{\mathrm{C},S}$ denote the resulting function which is to be used for estimation of β_S.

3.2. Augmented pseudo-likelihood estimation

The formulation of the marginal likelihood (2.4) for the candidate model S is

$$L_{\mathrm{M},S} = \prod_{i=1}^n \frac{\mathcal{F}(a_i|\pi_S x_i, z_i) dH(a_i)}{\int_0^\infty \mathcal{F}(\alpha|\pi_S x_i, z_i) dH(\alpha)}, \tag{3.8}$$

where $\mathcal{F}(a_i|\pi_S x_i, z_i) = \exp\left[-\Lambda_0(a_i)\exp\left\{(\pi_S x_i)^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right\}\right]$. Noting that the marginal likelihood (3.8) involves the unobserved covariate X_i, we now construct a modified version of (3.8) to address the measurement error effects.

Let μ_X and Σ_X be the mean vector and covariance matrix of X_i, respectively. Let $X^*_{i,S} = \pi_S X^*_i$ as in (3.3), then model (3.2) gives that

$$E(\pi_S X_i | X^*_{i,S} = x^*_{i,S}) = \pi_S \mu_X + \left(\Sigma_{X^*_S} - \Sigma_{\epsilon,S}\right)^\top \Sigma^{-1}_{X^*_S} (x^*_{i,S} - \mu_{X^*_S}), \tag{3.9}$$

where $\Sigma_{\epsilon,S} = \pi_S \Sigma_\epsilon \pi_S^\top$, and $\mu_{X^*_S}$ and $\Sigma_{X^*_S}$ represent the mean and covariance matrix of $X^*_{i,S}$, respectively.

We let $\widetilde{x}_{i,S}$ denote (3.9) for ease of notation. Using the method of moments, (3.9) is estimated by

$$\widehat{x}_{i,S} = \widehat{\mu}_{X^*_S} + \left(\widehat{\Sigma}_{X^*_S} - \Sigma_{\epsilon,S}\right)^\top \widehat{\Sigma}^{-1}_{X^*_S} \left(x^*_{i,S} - \widehat{\mu}_{X^*_S}\right) \tag{3.10}$$

with $\widehat{\mu}_{X^*_S} = \frac{1}{n}\sum_{i=1}^n x^*_{i,S}$ and $\widehat{\Sigma}_{X^*_S} = \frac{1}{n-1}\sum_{i=1}^n (x^*_{i,S} - \widehat{\mu}_{X^*_S})(x^*_{i,S} - \widehat{\mu}_{X^*_S})^\top$. Then replacing $\pi_S x_i$ with $\widehat{x}_{i,S}$ in likelihood function (3.8) gives

$$L^*_{\mathrm{M},S} = \prod_{i=1}^n \frac{\mathcal{F}(a_i|\widehat{x}_{i,S}, z_i) dH(a_i)}{\int_0^\infty \mathcal{F}(\alpha|\widehat{x}_{i,S}, z_i) dH(\alpha)}, \tag{3.11}$$

where

$$\mathcal{F}(a_i|\widehat{x}_{i,S}, z_i) = \exp\left\{-\Lambda_0(a_i)\exp\left(\widehat{x}^\top_{i,S}\beta_{x,S} + z_i^\top \beta_{z,S}\right)\right\}. \tag{3.12}$$

To use (3.11) for inference about β_S, we first use the nonparametric maximum likelihood estimator (NPMLE) (e.g., [23]) to estimate the distribution function $H(\cdot)$ of $\widetilde{A}$. Specifically, the NPMLE of $H(a)$ under a candidate model S in (3.11) is given by

$$\widehat{H}_S(a) = \left(\sum_{i=1}^n \frac{1}{\widehat{\mathcal{F}}(a_i|\widehat{x}_{i,S}, z_i)}\right)^{-1} \sum_{i=1}^n \frac{I(a_i \le a)}{\widehat{\mathcal{F}}(a_i|\widehat{x}_{i,S}, z_i)}, \tag{3.13}$$

where $\widehat{\mathcal{F}}(a_i|\widehat{x}_{i,S}, z_i)$ is determined by (3.12) with $\Lambda_0(\cdot)$ and $\left(\beta^\top_{x,S}, \beta^\top_{z,S}\right)^\top$ replaced by $\widehat{\Lambda}_{0,S}(\cdot)$ and $\widehat{\beta}_{CS} \triangleq \left(\widehat{\beta}^\top_{x,CS}, \widehat{\beta}^\top_{z,CS}\right)^\top$, respectively, given by (3.6), and $\widehat{\beta}_{CS} = \underset{\beta_S}{\operatorname{argmax}}\, \widehat{\ell}^*_{\mathrm{C},S}$.

Then replacing $H(a)$ in (3.11) with $\widehat{H}_S(a)$ gives $\widehat{L}^*_{\mathrm{M},S}$, and letting $\widehat{\ell}^*_{\mathrm{M},S} = \log \widehat{L}^*_{\mathrm{M},S}$ gives

$$\begin{aligned} \widehat{\ell}^*_{\mathrm{M},S} &= \sum_{i=1}^n \log\left\{d\widehat{H}_S(a_i)\right\} - \sum_{i=1}^n \widehat{\Lambda}_{0,S}(a_i)\exp\left(\widehat{x}^\top_{i,S}\beta_{x,S} + z_i^\top\beta_{z,S}\right) \\ &- \sum_{i=1}^n \log\left[\int_0^\infty \exp\left\{-\widehat{\Lambda}_{0,S}(\alpha)\exp\left(\widehat{x}^\top_{i,S}\beta_{x,S} + z_i^\top\beta_{z,S}\right)\right\} d\widehat{H}_S(\alpha)\right]. \end{aligned} \tag{3.14}$$

Consequently, the model parameter β_S for the candidate model S can be estimated by

$$\widehat{\beta}_S \triangleq \left(\widehat{\beta}^\top_{x,S}, \widehat{\beta}^\top_{z,S}\right)^\top = \underset{\beta_S}{\operatorname{argmax}}(\widehat{\ell}^*_{\mathrm{C},S} + \widehat{\ell}^*_{\mathrm{M},S}). \tag{3.15}$$

When $|S| = p$, i.e., all the variables $\{X_i, Z_i\}$ are included in the model, we have that $\pi_S = I_{p\times p}$ and hence $\pi_S X_i^* = X_i^*$, leading to the terms for the full model for which the subscript S is removed from the notation. For instance, specifying S as the full model, we let $\widehat{\Lambda}_0(\cdot)$ and $\widehat{\ell}_{\mathrm{M}}^*$ represent (3.6) and (3.14), respectively, and we write $\widehat{\ell}_{\mathrm{C}}^*$ as (3.5) with $\Lambda_0(\cdot)$ replaced by $\widehat{\Lambda}_0(\cdot)$.

Consequently, the estimator of β based on the full dataset is

$$\widehat{\beta}_{\text{full}} \triangleq \left(\widehat{\beta}_{x,\text{full}}^\top, \widehat{\beta}_{z,\text{full}}^\top\right)^\top = \underset{\beta}{\operatorname{argmax}} \left(\widehat{\ell}_{\mathrm{C}}^* + \widehat{\ell}_{\mathrm{M}}^*\right). \tag{3.16}$$

4. Asymptotic results under different settings

This section presents asymptotic results for different estimators derived from different models or formed for different parameters. In the first subsection, the parameter form is given, and we describe asymptotic results by considering different candidate models. In contrast, in the second subsection, we focus on a given candidate model and discuss asymptotic results for estimators of differently formed focus parameters. This section provides the basis for the development in Section 5 with the effects of concurrent features of measurement error, left-truncation, and right-censoring taken into account.

4.1. Asymptotic results concerning candidate models

Given a candidate model S, we define

$$\begin{aligned} \ell_{\mathrm{P},S}^* \;=\; & \sum_{i=1}^n \Bigg[\delta_i \left\{\left(\pi_S x_i^*\right)^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right\} + \frac{1}{2}\delta_i \log\left\{m_S(\beta_{x,S})\right\} \\ & -\delta_i \log\left\{\sum_{j=1}^n \exp\left(\left(\pi_S x_i^*\right)^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right) I(a_j \le y_i \le y_j)\right\}\Bigg], \end{aligned} \tag{4.1}$$

which has the same maximum likelihood estimator as obtained from (3.5); the relevant detail can be found in [3] and [24]. Let ℓ_{P}^* denote (4.1) with S being the full model.

Let $U_{\mathrm{P},S}(\beta_{x,S}, \beta_{z,S}) = \frac{\partial \ell_{\mathrm{P},S}^*}{\partial \beta_S}$, $U_{\mathrm{M},S}(\beta_{x,S}, \beta_{z,S}) = \frac{\partial \widehat{\ell}_{\mathrm{M},S}^*}{\partial \beta_S}$, $U_{\mathrm{P}}(\beta_x, \beta_z) = \frac{\partial \ell_{\mathrm{P}}^*}{\partial \beta}$ and $U_{\mathrm{M}}(\beta_x, \beta_z) = \frac{\partial \widehat{\ell}_{\mathrm{M}}^*}{\partial \beta}$, where $\ell_{\mathrm{P},S}^*$ and ℓ_{P}^* are determined by (4.1) with S being a submodel and the full model, respectively, and $\widehat{\ell}_{\mathrm{M},S}^*$ and $\widehat{\ell}_{\mathrm{M}}^*$ are given by (3.14) with S specified as a submodel and the full model, respectively. The following lemmas present the relationship between the candidate model S and the full model, which are useful for deriving asymptotic results of (3.15). The proofs of Lemmas 4.1 and 4.2 are given in Appendices B.1 and B.2, respectively.

Lemma 4.1. *For any candidate model S, let $\Sigma_{X_S^*}$ be the covariance matrix of X_S^* and let Σ_{X^*} be the covariance matrix of X^*. Then*

$$\pi_S^\top \Sigma_{X_S^*}^{-1} \pi_S = \Sigma_{X^*}^{-1}.$$

Lemma 4.2. *Under regularity conditions in Appendix A, the following results hold for any candidate model S:*

(a) $U_{P,S}\left(0_{|S|}, \beta_{z0}\right) = \Pi_S U_P\left(0_p, \beta_{z0}\right)$*;*
(b) $U_{M,S}\left(0_{|S|}, \beta_{z0}\right) = \Pi_S U_M\left(0_p, \beta_{z0}\right)$*.*

With the candidate model S, let W_S denote a random vector having the normal distribution $N\left(\mathcal{A}_S^{-1}\Pi_S\mathcal{A}\begin{pmatrix}\eta\\ 0_q\end{pmatrix},\ \mathcal{A}_S^{-1}\mathcal{B}_S\mathcal{A}_S^{-1}\right)$, where $\mathcal{A}$, $\mathcal{A}_S$ and $\mathcal{B}_S$ are defined in Appendix B.3.1.

Theorem 4.1. *Assume regularity conditions in Appendix A and the candidate model S. As $n\to\infty$, we have that*

(a) $\sqrt{n}\begin{pmatrix}\widehat{\beta}_{x,S}\\ \widehat{\beta}_{z,S}-\beta_{z0}\end{pmatrix}\xrightarrow{d} W_S$*;*

(b) $\sqrt{n}\left\{\widehat{\Lambda}_{0,S}(t)-\Lambda_0(t)\right\}\xrightarrow{d}\mathcal{V}(t)-\begin{pmatrix}F_{x,S}(t)\\ F_z(t)\end{pmatrix}^{\top}W_S+F_x(t)^{\top}\eta$*, where $\mathcal{V}(t)$ is the Gaussian process with mean zero, $F_{x,S}(t)$, $F_x(t)$ and $F_z(t)$ are given in Appendix B.3.2, and η is the parameter defined in Section 2.2.*

4.2. Asymptotic results for estimators of focused parameters

Rather than examining the model parameters individually, in applications we are often interested in their combined forms or functions of those parameters. To facilitate such settings, we let $\mu = \mu\left(\beta_x, \beta_z, \Lambda_0(\cdot)\right)$ be a scalar function of parameter $\beta = \left(\beta_x^\top, \beta_z^\top\right)^\top$ and function $\Lambda_0(t)$. The new parameter μ plays the role of using a simple *scalar* measure to express certain combined information of the original multi-dimensional parameters; it is called the *focus* parameter [6, 7, 11]. The choice of the function $\mu(\cdot)$ is often driven by the nature of individual problems (to be discussed in Section 5.1). In contrast to the notation $\beta_0 = (\frac{\eta^\top}{\sqrt{n}}, \beta_{z0}^\top)$ defined in Section 2.2, we let $\mu_{\text{true}} = \mu\left(\frac{\eta}{\sqrt{n}}, \beta_{z0}, \Lambda_0(\cdot)\right)$ denote the corresponding value of the focus parameter μ. By the invariance property of the maximum likelihood estimator, $\widehat{\mu}_S = \mu\left(\widehat{\beta}_{x,S}, \widehat{\beta}_{z,S}, \widehat{\Lambda}_0(\cdot)\right)$ can be taken as the estimated focus parameter corresponding to the candidate model S. We comment that although the density function $h(\cdot)$ of the truncation time is unknown, we do not include it when defining the focus parameter.

When S is the full model, we let $\mathcal{B}$ denote $\mathcal{B}_S$. For $\mathcal{A}$ defined in Section 4.1 and $\mathcal{B}$, we express them as block matrices according to the dimension of the covariates X_i and Z_i: $\mathcal{A} = \begin{pmatrix}A_{xx} & A_{xz}\\ A_{zx} & A_{zz}\end{pmatrix}$ and $\mathcal{B} = \begin{pmatrix}B_{xx} & B_{xz}\\ B_{zx} & B_{zz}\end{pmatrix}$. Let $\mathcal{A}^{-1} = \begin{pmatrix}A^{xx} & A^{xz}\\ A^{zx} & A^{zz}\end{pmatrix}$ denote the inverse matrix of $\mathcal{A}$. We now present the asymptotic properties for the estimators of the focus parameters whose proof is placed in Appendix B.4.

Theorem 4.2. *Assume that the conditions in Theorem 4.1 hold and consider the candidate model S. Suppose that the focus parameter $\mu = \mu(\beta_x, \beta_z, \Lambda_0(t))$ is a continuously differentiable function of parameter β and the cumulative baseline hazard function at a given time point t. Then as $n \to \infty$,*

$$\begin{aligned}\sqrt{n}\,(\widehat{\mu}_S - \mu_{true}) \quad \xrightarrow{d} \quad & \frac{\partial \mu}{\partial \Lambda_0}\mathcal{V}(t) + \left(\frac{\partial \mu}{\partial \beta_z} + \frac{\partial \mu}{\partial \Lambda_0}F_z(t)\right)^\top A_{zz}^{-1}J_z \\ & + (\omega+\kappa)^\top \left\{\eta - (A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\mathcal{U}\right\},\end{aligned}$$

where $\mathcal{V}(t)$ is the zero mean Gaussian process defined in Theorem 4.1, $\mathbb{H}_S = (A^{xx})^{-1/2}\pi_S^\top\left\{\pi_S(A^{xx})^{-1}\pi_S^\top\right\}^{-1}\pi_S(A^{xx})^{-1/2}$, $\kappa = \frac{\partial\mu}{\partial\Lambda_0}F_x(t) - A_{zx}^\top A_{zz}^{-1}\frac{\partial\mu}{\partial\Lambda_0}F_z(t)$, $\omega = \frac{\partial\mu}{\partial\beta_x} - A_{zx}^\top A_{zz}^{-1}\frac{\partial\mu}{\partial\beta_z}$, $\mathcal{U} = \eta + \mathcal{W}$, $\mathcal{W} = A^{xx}J_x - A^{xx}A_{xz}A_{zz}^{-1}J_z$, and J_z and J_x are random variables having the distributions $N(0_q, B_{zz})$ and $N(0_p, B_{xx})$, respectively.

A useful special case is that the focus parameter $\mu = \mu(\beta_x, \beta_z)$ is the function of parameter β alone. Then Theorem 4.2 says that as $n \to \infty$,

$$\sqrt{n}\,(\widehat{\mu}_S - \mu_{\text{true}}) \xrightarrow{d} \left(\frac{\partial\mu}{\partial\beta_z}\right)^\top A_{zz}^{-1}J_z + \omega^\top\left\{\eta - (A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\mathcal{U}\right\}. \tag{4.2}$$

We comment that Theorem 4.2 differs from the results established by [9] and [11] who considered censored data with the FIC. Their model setup and the features of data are different from what we discuss here which involves left-truncation and measurement error.

5. Focus parameter and model averaging

We examine the *focus parameter* under several useful settings and discuss the selection criterion accordingly. Furthermore, we establish large sample properties of model averaging estimators.

5.1. Useful settings and focus information criterion

In this subsection, we illustrate the choice of the focus parameters using examples which are pertinent to the hazard ratio and the survivor function. Furthermore, we employ the focus information criterion (FIC) (e.g., [6]) to conduct model selection. The key idea is to select the best candidate model using the smallest estimated mean squared errors (MSE) derived for estimators of the focus parameter under all candidate models.

Setting 1: The hazard ratio.
Under the Cox model (2.1), the hazard ratio for $V = v_0$ relative to $V = v_0 + 1_d$ is

$$\mu \triangleq \frac{\lambda(t|v_0 + 1_d)}{\lambda(t|v_0)} = \exp\left(1_d^\top\beta\right), \tag{5.1}$$

where 1_d is the $d \times 1$ unit vector, and v_0 is a value of V. In this case, the focus parameter is taken as the hazard ratio μ which is a function of β.

Theorem 4.1 shows that $\mathcal{W}$ has a normal distribution $N(0, \sigma_{xx})$ with σ_{xx} being the asymptotic covariance matrix of $\widehat{\beta}_{x,\text{full}}$ that is given in (B.6) in Appendix B.3. Therefore, for any candidate model S, by (4.2), the expectation of $\widehat{\mu}_S - \mu_{\text{true}}$ and the variance of $\widehat{\mu}_S$ are, respectively, given by

$$E(\widehat{\mu}_S - \mu_{\text{true}}) = \omega^\top \left\{ I_{p\times p} - (A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2} \right\} \eta$$

and

$$\begin{aligned} \operatorname{var}(\widehat{\mu}_S) &= \left(\frac{\partial \mu}{\partial \beta_z}\right)^\top A_{zz}^{-1} B_{zz} A_{zz}^{-1} \left(\frac{\partial \mu}{\partial \beta_z}\right) \\ &\quad + \omega^\top (A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\sigma_{xx}(A^{xx})^{-1/2}\mathbb{H}_S(A^{xx})^{1/2}\omega, \end{aligned}$$

and thus, the MSE of $\widehat{\mu}_S$ is derived as

$$\begin{aligned} E\left\{(\widehat{\mu}_S - \mu_{\text{true}})^2\right\} &= \left(\frac{\partial \mu}{\partial \beta_z}\right)^\top A_{zz}^{-1} B_{zz} A_{zz}^{-1} \left(\frac{\partial \mu}{\partial \beta_z}\right) \\ &\quad + \omega^\top \left\{ (I_{p\times p} - \Phi_S)\, \eta\eta^\top (I_{p\times p} - \Phi_S)^\top + \Phi_S \sigma_{xx} \Phi_S^\top \right\} \omega, \end{aligned} \tag{5.2}$$

where $\Phi_S = (A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}$.

The first term of (5.2) does not depend on the candidate model S, so to make the comparison among different candidate models be focused, we drop this term and simply let the second term of (5.2) reflect the MSE of $\widehat{\mu}_S$:

$$\omega^\top \left\{ (I_{p\times p} - \Phi_S)\, \eta\eta^\top (I_{p\times p} - \Phi_S)^\top + \Phi_S \sigma_{xx} \Phi_S^\top \right\} \omega. \tag{5.3}$$

To use (5.3), $\eta\eta^\top$ needs to be estimated. By Theorem 4.1 (a), $\sqrt{n}\widehat{\beta}_{x,\text{full}} = \widehat{\eta} \sim N(\eta, \sigma_{xx})$, thus $E\left(n\widehat{\beta}_{x,\text{full}}\widehat{\beta}_{x,\text{full}}^\top\right) = \sigma_{xx} + \eta\eta^\top$, so that $n\widehat{\beta}_{x,\text{full}}\widehat{\beta}_{x,\text{full}}^\top - \widehat{\sigma}_{xx}$ is an asymptotically unbiased estimator of $\eta\eta^\top$. Thus, (5.3) is estimated by

$$\begin{aligned} \widehat{\text{MSE}}_S &= \widehat{\omega}_1^\top \Big\{ \left(I_{p\times p} - \widehat{\Phi}_S\right)\left(n\widehat{\beta}_{x,\text{full}}\widehat{\beta}_{x,\text{full}}^\top - \widehat{\sigma}_{xx}\right)\left(I_{p\times p} - \widehat{\Phi}_S\right)^\top \\ &\quad + \widehat{\Phi}_S \widehat{\sigma}_{xx} \widehat{\Phi}_S^\top \Big\} \widehat{\omega}_1, \end{aligned} \tag{5.4}$$

where $\widehat{\Phi}_S = \widehat{A}_{xx}^{-1/2}\widehat{H}_S\widehat{A}_{xx}^{1/2}$, $\widehat{\omega}_1 = \frac{\partial \mu(\widehat{\beta}_S)}{\partial \beta_x} - \widehat{A}_{zx}^\top \widehat{A}_{zz}^{-1}\frac{\partial \mu(\widehat{\beta}_S)}{\partial \beta_z}$, and $\widehat{\sigma}_{xx}$ is the estimated asymptotic covariance matrix of $\widehat{\beta}_{x,\text{full}}$.

Setting 2: Linear combinations of covariate effects.
In contrast to the hazard ratio, we may be interested in a linear combination of covariate effects (e.g., [21, 22, 30]). Then the focus parameter is set as $\mu = c_x^\top \beta_x + c_z^\top \beta_z$, where c_x and c_z denote p-dimensional and q-dimensional

vectors of the coefficients, respectively. Similar to the derivations for Setting 1, $\widehat{\mathrm{MSE}}_S$ in (5.4) can be used to select the best candidate model, where $\widehat{\omega}_1$ is taken as $c_x - \widehat{A}_{zx}^{-1}\widehat{A}_{zz}c_z$.

Setting 3: The cumulative baseline hazard function.
In some applications, as discussed by [11], the cumulative baseline hazard function $\Lambda_0(\cdot)$ is of prime interest, and in this case the focus parameter μ is set as $\Lambda_0(t_0)$ for some time point, say t_0.

Applying Theorem 4.2 with $\omega = 0$ and $\kappa = F_x(t_0) - A_{zx}^\top A_{zz}^{-1} F_z(t_0)$, we can work out the MSE of $\widehat{\mu}_S$ for the candidate model S. Similar to (5.4), with the candidate model S, MSE is estimated as

$$\begin{aligned}\widehat{\mathrm{MSE}}_S \quad = \quad & \widehat{\kappa}_2^\top \Big\{ \Big(I_{p\times p} - \widehat{\Phi}_S\Big)\Big(n\widehat{\beta}_{x,\mathrm{full}}\widehat{\beta}_{x,\mathrm{full}}^\top - \widehat{\sigma}_{xx}\Big)\Big(I_{p\times p} - \widehat{\Phi}_S\Big)^\top \\ & + \widehat{\Phi}_S\widehat{\sigma}_{xx}\widehat{\Phi}_S^\top\Big\}\widehat{\kappa}_2, \end{aligned} \tag{5.5}$$

where $\widehat{\Phi}_S$ and $\widehat{\sigma}_{xx}$ are the same as described in Setting 1, $\widehat{\kappa}_2 = \frac{\partial\mu(\widehat{\Lambda}_{0,S}(t_0))}{\partial\Lambda_0}\widehat{F}_x(t_0) - \widehat{A}_{zx}^\top\widehat{A}_{zz}^{-1}\frac{\partial\mu(\widehat{\Lambda}_{0,S}(t_0))}{\partial\Lambda_0}\widehat{F}_z(t_0)$, and the terms free of the candidate model S are omitted.

Setting 4: The survivor function.
In applications, we are often interested in the survivor function

$$\mathcal{F}(t|v) = \exp\left\{-\Lambda_0(t)\exp\left(v^\top\beta\right)\right\}$$

at certain time point, say t_0. In this situation, we take the focus parameter to be

$$\mu \triangleq \mu(\beta, \Lambda_0(t_0)) = \exp\left\{-\Lambda_0(t_0)\exp\left(v_0^\top\beta\right)\right\}$$

for some given covariate value v_0.

Again, by Theorem 4.2 with the similar discussion in Setting 1, the MSE of $\widehat{\mu}_S$ is given by

$$\begin{aligned} E\left\{(\widehat{\mu}_S - \mu_{\mathrm{true}})^2\right\} \quad = \quad & \left(\frac{\partial\mu}{\partial\beta_z} + \frac{\partial\mu}{\partial\Lambda_0}F_z(t_0)\right)^\top A_{zz}^{-1}B_{zz}A_{zz}^{-1}\left(\frac{\partial\mu}{\partial\beta_z} + \frac{\partial\mu}{\partial\Lambda_0}F_z(t_0)\right) \\ & + (\omega+\kappa)^\top\Big\{(I_{p\times p} - \Phi_S)\,\eta\eta^\top(I_{p\times p} - \Phi_S)^\top \\ & + \Phi_S\sigma_{xx}\Phi_S^\top\Big\}(\omega+\kappa). \end{aligned}$$

Similar to the discussion for (5.4), we drop those quantities which are unrelated to S and replace $\eta\eta^\top$ by its asymptotically unbiased estimator, and then we obtain the estimate of MSE for the candidate model S:

$$\begin{aligned}\widehat{\mathrm{MSE}}_S \quad = \quad & (\widehat{\omega}_3 + \widehat{\kappa}_3)^\top\Big\{\Big(I_{p\times p} - \widehat{\Phi}_S\Big)\Big(n\widehat{\beta}_{x,\mathrm{full}}\widehat{\beta}_{x,\mathrm{full}}^\top - \widehat{\sigma}_{xx}\Big)\Big(I_{p\times p} - \widehat{\Phi}_S\Big)^\top \\ & + \widehat{\Phi}_S\widehat{\sigma}_{xx}\widehat{\Phi}_S^\top\Big\}(\widehat{\omega}_3 + \widehat{\kappa}_3), \end{aligned} \tag{5.6}$$

where $\widehat{\omega}_3 = \frac{\partial\mu(\widehat{\beta}_S,\widehat{\Lambda}_{0,S}(t_0))}{\partial\beta_x} - \widehat{A}_{zx}^\top \widehat{A}_{zz}^{-1} \frac{\partial\mu(\widehat{\beta}_S,\widehat{\Lambda}_{0,S}(t_0))}{\partial\beta_z}$, $\widehat{\kappa}_3 = \frac{\partial\mu(\widehat{\beta}_S,\widehat{\Lambda}_{0,S}(t_0))}{\partial\Lambda_0}\widehat{F}_x(t_0) - \widehat{A}_{zx}^\top \widehat{A}_{zz}^{-1} \frac{\partial\mu(\widehat{\beta}_S,\widehat{\Lambda}_{0,S}(t_0))}{\partial\Lambda_0}\widehat{F}_z(t_0)$, and $\widehat{\Phi}_S$ and $\widehat{\sigma}_{xx}$ are described in Setting 1.

These settings cover the scenarios we usually encounter in survival analysis. With the focus parameter specified differently, the FIC can yield different final models, which are carried out based on the smallest $\widehat{\mathrm{MSE}}_S$ for all the candidate models S.

5.2. Frequentist model averaging

As discussed by [8], conducting parameter estimation using a specifically selected model is not ideal since the associated uncertainty is ignored. To alleviate this issue, we employ the frequentist model averaging (FMA) method to construct an estimator of μ. The idea is to use the estimators derived from different candidate models to work out a suitable linear combination of them, given by

$$\widehat{\mu}_{\text{ave}} = \sum_{S\in\mathcal{S}} w(S|\widehat{\eta})\widehat{\mu}_S,$$

where $\widehat{\mu}_S$ represents the estimator of μ obtained from a method described in Section 4.2 under the candidate model S, $\widehat{\eta} = \sqrt{n}\widehat{\beta}_{x,\text{full}}$ with $\widehat{\beta}_{x,\text{full}}$ described in (3.16), and $w(S|\widehat{\eta})$ is a nonnegative weight pertinent to the candidate model S which is data-driven and constrained by $\sum\limits_{S\in\mathcal{S}} w(S|\widehat{\eta}) = 1$ ([7], p.195). Conventional weights $w(S|\widehat{\eta})$ can be constructed using the AIC, BIC, or FIC; details can be found in [10].

Theorem 5.1. *Assume that the conditions in Theorem 4.2 hold. Then as $n \to \infty$,*

$$\begin{aligned}\sqrt{n}\left(\widehat{\mu}_{ave} - \mu_{true}\right) \quad \xrightarrow{d} \quad & \frac{\partial\mu}{\partial\Lambda_0}\mathcal{V}(t) + \left\{\frac{\partial\mu}{\partial\beta_z} + \frac{\partial\mu}{\partial\Lambda_0}F_z(t)\right\}^\top A_{zz}^{-1}J_z \\ & + (\omega+\kappa)^\top \left\{\mathcal{U} - \sum_{S\in\mathcal{S}} w(S|\mathcal{U})(A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\mathcal{U}\right\},\end{aligned}$$

where $w(S|\mathcal{U})$ represents the weight to which $w(S|\widehat{\eta})$ converges in distribution.

We comment that the technical details in this theorem differ from the work concerning model averaging with FIC (e.g., [22, 30]) due to the differences in the model setup and the natural of data. However, the derivations of those results basically align with the development of [6] and [10], though different conditions may be required to reflect the involvement of different processes. Our development here typically hinges on four processes related to survival, censoring, truncation, and measurement error. Untangling the relationship among those processes requires care when deriving the results in Theorem 5.1.

6. Numerical studies

In this section, we first conduct simulation studies to assess the performance of the proposed estimators, and then implement the methods to analyze a real dataset.

6.1. Simulation studies

For each setting, we run 500 simulations, where the sample size $n = 100, 200$ or 400. For the covariates, we generate X from $N(0_6, \Sigma_X)$ and Z from $N(0_2, \Sigma_Z)$ independently, where

$$\Sigma_X = \begin{pmatrix} 1 & 0.2 & \cdots & 0.2 \\ 0.2 & 1 & \cdots & 0.2 \\ \vdots & \vdots & \ddots & \vdots \\ 0.2 & 0.2 & \cdots & 1 \end{pmatrix}_{6\times 6} \quad \text{and} \quad \Sigma_Z = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}.$$

The goal here is to select the important variables in X and always retain the covariate Z in the models. The survival time is generated using model (2.1) where the baseline hazard function is set as $\lambda_0(t) = 2t$ or $\log t$. Specifically, the failure times are determined by

$$\widetilde{T} = \sqrt{-\exp\left(X^\top \beta_{x0} + Z^\top \beta_{z0}\right) \log(1-U)} \ \text{ if } \ \lambda_0(t) = 2t,$$

or

$$\widetilde{T}\left(\log \widetilde{T} - 1\right) \exp\left(X^\top \beta_{x0} + Z^\top \beta_{z0}\right) + \log(1-U) = 0 \ \text{ if } \ \lambda_0(t) = \log t,$$

where U is simulated from the uniform distribution $U(0,1)$, and the parameter $\beta_0 = \left(\beta_{x0}^\top, \beta_{z0}^\top\right)^\top$ is set as $\beta_{x0} = \frac{\eta}{\sqrt{n}}$ and $\beta_{z0} = (0.6, 0.6)^\top$. We consider three cases with

$$(1)\ \eta = (0,0,0,0,0,0)^\top, \quad (2)\ \eta = (1,1,1,0,0,0)^\top, \ \text{ and } (3)\ \eta = (1,1,1,1,1,1)^\top.$$

Case (1) gives a *null* model, Case (2) indicates that some covariates are not included in the true model, and Case (3) says that the *full* model contains all the covariates.

Let the truncation time $\widetilde{A}$ be generated from the exponential distribution with mean 10. The observed data (A, T, V) are then obtained from $(\widetilde{A}, \widetilde{T}, V)$ using the condition $\widetilde{T} \geq \widetilde{A}$. Independently repeat this data simulation step n times to generate a sample of size n. The censoring variable C is generated from the uniform distribution $U(0, c)$ where c is a constant that is chosen to yield about 50% censored data. Consequently, Y and Δ are determined by $Y = \min\{T, A + C\}$ and $\Delta = I(T \leq A + C)$.

Consistent with Section 2.2, X is the error-prone covariates and X^* is the observed variable which is generated from

$$X^* = X + \epsilon, \tag{6.1}$$

where ϵ is independently generated from $N(0_6, \Sigma_\epsilon)$, and Σ_ϵ is a diagonal matrix whose diagonal elements are all specified as 0.1 or 0.5.

For the focus parameter, we consider three forms, given by (a) $\mu_{10}(\beta) = \exp\left(1_6^\top \beta\right)$, (b) $\mu_{20}(\Lambda_0) = \Lambda_0(1)$, and (c) $\mu_{30}(\beta, \Lambda_0) = \exp\left\{-\Lambda_0(1)\exp\left(1_6^\top \beta\right)\right\}$, respectively. We are interested in selecting variables for different focus parameters using the proposed FIC method. As comparisons, we also apply AIC or BIC to select variables.

First, we examine the selection results of candidate models. Among the 500 simulations, let p(True) denote the proportion of selecting the true model, let p(S) be the proportion of selecting additional variables, and let p(FN) represent the proportion of false exclusion of variables. We report the results for the case of $n = 400$ and $\lambda_0(t) = 2t$ in Table 1. For the sake of space constraints, we omit the results for other settings which show patterns similar to those of Table 1. Under the null model $\eta = (0, \cdots, 0)^\top$, AIC tends to select more irrelevant variables, and BIC and FIC are more frequently select the true model. On the contrary, for the full model $\eta = (1, \cdots, 1)^\top$, the best selected candidate models by BIC are smaller than the true model, while the best candidate models determined by AIC and FIC are relatively close to the true model. Regarding the true model with $\eta = (1, 1, 1, 0, 0, 0)^\top$, the proposed FIC method has larger proportions to determine the true model than AIC and BIC which are more frequently select either larger or smaller models. Moreover, we observe that without suitably adjusting the effects of measurement error, the naive method constantly fails to select the true model, regardless of the selection criteria and the form of the focus parameters. In summary, the proposed error-correction FIC method performs well in model selection.

After determining the best candidate model based a selection criterion, we then use the plug-in method to obtain estimates of the focus parameters. Let $\widehat{\mu}_j$ denote the resultant estimate of a focus parameter for simulation j, where $j = 1, \cdots, 500$. We compute the square root of the mean squared error (RMSE) as $\sqrt{500^{-1} \sum_{j=1}^{500} \left(\widehat{\mu}_j - \mu_0\right)^2}$. In contrast, we also report the results obtained from the naive method which implements (3.1) and (3.8) with X_i replaced by X_i^*. The results for $\lambda_0(t) = 2t$ and $\lambda_0(t) = \log t$ are summarized in Tables 2–4 and Tables 5–7, respectively. Furthermore, model averaging estimators with weights defined in [7] are also investigated, and the results are displayed under the headings sAIC, sBIC, and sFIC in Tables 2–7.

As expected, the RMSEs for the proposed estimators are smaller than those for the naive estimators regardless of the selection criteria; and the differences become more noticeable as measurement error is more substantial. This demonstrates the necessity of addressing measurement error effects in inferential procedures. Interestingly, for both the naive approach and the proposed method, using FIC tends to result in smaller RMSEs than using AIC or BIC under our simulation settings. While there is no rigorous theory to show this is always the case, such a phenomenon was also being observed by other authors such as [21], [22], [30], and [31]. Furthermore, the model averaging estimators, sAIC, sBIC and sFIC, are comparable to their counterparts, AIC, BIC and FIC, respectively, and sFIC outperforms both sAIC and sBIC under the settings we consider.

6.2. Analysis of Worcester Heart Attack study data

In this section, we apply the proposed methods to analyze the data arising from the Worcester Heart Attack Study (WHAS500). Data were collected over thirteen 1-year periods beginning in 1975 and extending in 2001 on all patients with acute myocardial infarction (MI) admitted to hospitals in Worcester, Massachusetts Standard Metropolitan Statistical Area. Three types of time were recorded for the study subjects: the date of the hospital admission, the date of the hospital discharge, and the time of the last follow-up (which is either the death or censoring time). We are interested in studying survival times of patients who were discharged alive from the hospital. Hence, a selection criterion was imposed that only those subjects who were discharged alive were eligible to be included in the analysis; individuals were not enrolled in the analysis if they died before discharging from the hospital, hence left-truncation occurs. Consistent with [12], we define the survival time as the time length between the hospital admission and the last follow-up, and a truncation time as the time length between the hospital admission and the hospital discharge. With such a criterion, a sample of size 461 was available, and the censoring rate was 61.8%.

The following covariates are included in our analysis: initial heart rate (X_1), initial systolic blood pressure (X_2), initial diastolic blood pressure (X_3), body mass index (X_4), age (Z_1), and gender (Z_2). Let $\beta = (\beta_{x_1}, \beta_{x_2}, \beta_{x_3}, \beta_{x_4}, \beta_{z_1}, \beta_{z_2})^\top$ denote the vector of the parameters for model (2.1). Covariates X_1, X_2, X_3 and X_4 are error-prone due to the reasons including inaccurate measurement devices and/or procedures, the biological variability, and temporal variations. Similar to the settings in Section 6.1, we discuss three focus parameters: the hazard ratio (μ_1), the cumulative baseline hazard function (μ_2) at time t_0, $\Lambda_0(t_0)$, and the survivor function (μ_3) at time t_0, $\mathcal{F}(t_0|v)$, with covariates taken as empirical means of variables, where for illustration, we take t_0 as the 50% percentiles of the observed survival times Y_i (e.g., [25]), bearing in mind that other values of interest can also be specified as t_0. Our goal is to select important variables from X_1 to X_4 for different focus parameters, with Z_1 and Z_2 always retained.

We first present the estimators of β under the full model using both the proposed approach discussed in Sections 3.1 and 3.2, and the naive approach which ignores measurement error by directly implementing (3.1) or (3.8) with X_i replaced by X_i^*. Since this dataset contains no additional information, such as repeated measurements or validation data, to characterize the degree of measurement error, we conduct sensitivity analyses by examining settings with the parameters for the measurement error model specified as different values. Let Σ_X and Σ_{X^*} denote covariance matrices of X and X^*, respectively, and let σ_{Xij}, σ_{X^*ij}, and $\sigma_{\epsilon ij}$ denote the (i,j) entry of Σ_X, Σ_{X^*}, and Σ_ϵ, respectively. Measurement error model (2.5) gives $\Sigma_{X^*} = \Sigma_X + \Sigma_\epsilon$, suggesting that σ_{Xij} is smaller than σ_{X^*ij} for all i and j. Using the reliability ratio $R_{ij} \equiv \frac{\sigma_{Xij}}{\sigma_{X^*ij}}$, or equivalently, $\sigma_{\epsilon ij} = (R_{ij}^{-1} - 1)\sigma_{Xij}$, we specify different values of $\sigma_{\epsilon ij}$ by varying R_{ij} and/or σ_{Xij}. For ease of exposition, we take R_{ij} as a constant, say R, for all i and j and consider that σ_{Xij} is given. For example, we take σ_{Xij} as $\sigma_{Xij} = c\widehat{\sigma}_{X^*ij}$ where $\widehat{\sigma}_{X^*ij}$ is the (i,j) entry of the empirical estimate of Σ_{X^*} and c is a positive constant. For illustrations, we take $c = 0.9$, bearing in mind that other values in the interval $(0,1)$ can be typically set for c as well. Here we specify R in the interval $[0.5, 0.9]$, and the estimation results are shown in Figure 2. We see that as the degree of measurement error changes, the patterns of $\widehat{\beta}_{x_j}(j = 1,2,3,4)$ are fluctuated while $\widehat{\beta}_{z_1}$ and $\widehat{\beta}_{z_2}$ are fairly stable.

To examine the proposed estimators more closely, we report the results for $R = 0.85, 0.75$ and 0.65 in Table 8 which includes the estimates, the standard errors (SE) and the p-values for the proposed and the naive methods under the full model. As expected, SEs of the naive estimator are generally smaller than those of the proposed estimator. Both the naive and the proposed methods suggest all the covariates are significant, regardless of measurement error degrees.

Next, we report the variable selection results based on AIC, BIC, and FIC for the three focus parameters discussed in Section 6.1 and present the best five candidate models in Table 9. Here we use a label, such as "134" to represent that the variables X_1, X_3, and X_4 are selected for the model. The best model selected by AIC contains more variables than those by BIC. It is interesting to see that the selection results by the naive approach and the proposed approach are similar. The FIC approach results in relatively more parsimonious models, regardless of using the proposed method or the naive method. In Table 10 we summarize the results for the estimates of the focus parameters based on the best models and derived from the model averaging estimators.

7. Discussion

Left-truncated and right-censored data arise commonly from studies of survival information. Analysis of such survival data is further complicated by the presence of error-prone covariates and unimportant covariates. In this paper, we develop estimation methods using the FIC criterion to handle such data. We implement the model averaging technique to derive estimators of the focus parameters and establish asymptotic results of the proposed estimators. Numerical studies confirm the satisfactory performance of our proposed methods.

The development here focuses on the case with continuous covariates subject to measurement error. In applications with discrete covariates subject to mismeasurement, the present development may be modified by following the discussion of [29] and [32]. When error-prone data include both discrete and continuous variables (e.g., [29], p.71), we may employ a procedure along the lines of [28] to address the mismeasurement effects. This topic warrants careful research. Further, the discussion here considers settings with covariates divided as X_i and Z_i, where X_i is error-prone and includes unimportant predictors, and Z_i is precisely measured and contains important predictors. Separating error-prone and error-free covariates offers us convenience in modeling the measurement error process, and it allows us to conduct inference by inducing minimal model assumptions with Z_i left unattended to. Such a strategy has been widely adopted in the literature of measurement error problems (e.g., [2, 29]).

While we focus on the classical measurement error model in this paper, it is possible to modify the proposed methods to accommodate other measurement error models (e.g., [29], Section 2.6). Strategies outlined in Section 2.5 of [29] may be employed to account for measurement error effects.

The development here assumes Σ_ϵ to be known. Such an assumption is typically feasible for two circumstances: (1) prior studies provide the information on the covariate mismeasurement and offer an estimate of Σ_ϵ, and (2) in conducting sensitivity analyses, different values of Σ_ϵ are specified to understand how mismeasurement effects may affect inference results about the parameters associated with the survival model. Although taking Σ_ϵ to be given gives us convenience in implementing the developed methods, it is recognized, as pointed out by a referee, that we ignore the uncertainty induced from the discrepancy between any specified or estimated value of Σ_ϵ and its true value. When Σ_ϵ is estimated from an additional data source as outlined in Section 2.3, the induced variability for the estimation procedure should be taken into account. While the ideas considered by [18], [19] and [29] can be adapted for this purpose, it is interesting to carry out careful explorations to work out the technical details. Further, in the lack of additional data for characterizing Σ_ϵ, [1] proposed a new method to estimate the variance of classical additive error of a normal distribution for settings with a univariate covariate subject to measurement error. It may be useful to generalize the setting of [1] to accommodating multiple covariates which are error-prone, and then modify the development here accordingly. Computational intensity may be an issue as the dimension of error-contaminated covariates becomes large.

As described in Section 5.1, our development is carried out for the focus parameters which mainly pertain to the covariate effects β or the cumulative baseline hazard function $\Lambda_0(\cdot)$. Other types of focus parameters, such as percentiles (e.g., median), the probability of dying in an interval, and the expected lifetime beyond a given time point, can be of interest as well. In this instance, it is generally needed to establish the process convergence (rather than at a single time point) for the estimator of the cumulative baseline hazard function $\Lambda_0(t)$, in combination with that of the estimator of β. The discussion of [14] can be adapted in conjunction with the development here to address the effects due to left-truncation and measurement error.

Table 1

Simulation results: the selection of best candidate model for the focus parameters with $n = 400$ *and* $\lambda_0(t) = 2t$.

η	σ_ϵ			Proposed			Naive		
				p(True)	p(S)	p(FN)	p(True)	p(S)	p(FN)
(1)	0.1	μ_1	AIC	0.832	0.168	0.000	0.118	0.882	0.000
			BIC	0.886	0.114	0.000	0.172	0.828	0.000
			FIC	0.904	0.096	0.000	0.226	0.774	0.000
		μ_2	AIC	0.792	0.208	0.000	0.162	0.838	0.000
			BIC	0.812	0.188	0.000	0.212	0.788	0.000
			FIC	0.914	0.086	0.000	0.306	0.694	0.000
		μ_3	AIC	0.780	0.220	0.000	0.118	0.882	0.000
			BIC	0.838	0.162	0.000	0.172	0.828	0.000
			FIC	0.928	0.072	0.000	0.332	0.668	0.000
	0.5	μ_1	AIC	0.794	0.206	0.000	0.108	0.892	0.000
			BIC	0.868	0.132	0.000	0.136	0.864	0.000
			FIC	0.901	0.099	0.000	0.205	0.765	0.000
		μ_2	AIC	0.770	0.230	0.000	0.150	0.850	0.000
			BIC	0.805	0.195	0.000	0.170	0.830	0.000
			FIC	0.908	0.092	0.000	0.298	0.602	0.000
		μ_3	AIC	0.757	0.243	0.000	0.107	0.893	0.000
			BIC	0.827	0.173	0.000	0.155	0.845	0.000
			FIC	0.919	0.081	0.000	0.253	0.747	0.000
(2)	0.1	μ_1	AIC	0.662	0.236	0.102	0.196	0.562	0.242
			BIC	0.784	0.052	0.264	0.148	0.274	0.578
			FIC	0.912	0.048	0.040	0.433	0.359	0.208
		μ_2	AIC	0.690	0.244	0.066	0.176	0.498	0.326
			BIC	0.749	0.050	0.201	0.180	0.326	0.494
			FIC	0.918	0.078	0.004	0.364	0.524	0.112
		μ_3	AIC	0.659	0.200	0.141	0.143	0.511	0.346
			BIC	0.693	0.041	0.266	0.163	0.327	0.510
			FIC	0.934	0.064	0.002	0.368	0.566	0.066
	0.5	μ_1	AIC	0.542	0.332	0.126	0.132	0.644	0.224
			BIC	0.700	0.056	0.244	0.140	0.254	0.606
			FIC	0.892	0.076	0.032	0.389	0.428	0.274
		μ_2	AIC	0.656	0.322	0.022	0.170	0.338	0.492
			BIC	0.722	0.008	0.270	0.174	0.338	0.488
			FIC	0.902	0.090	0.008	0.360	0.426	0.214
		μ_3	AIC	0.645	0.343	0.012	0.137	0.596	0.267
			BIC	0.678	0.022	0.300	0.150	0.312	0.538
			FIC	0.924	0.050	0.026	0.345	0.564	0.091
(3)	0.1	μ_1	AIC	0.886	0.000	0.114	0.366	0.000	0.634
			BIC	0.810	0.000	0.190	0.245	0.000	0.755
			FIC	0.935	0.000	0.065	0.387	0.000	0.613
		μ_2	AIC	0.874	0.000	0.153	0.345	0.000	0.655
			BIC	0.765	0.000	0.235	0.313	0.000	0.687
			FIC	0.952	0.000	0.048	0.365	0.000	0.635
		μ_3	AIC	0.889	0.000	0.111	0.322	0.000	0.678
			BIC	0.734	0.000	0.266	0.295	0.000	0.705
			FIC	0.930	0.000	0.070	0.373	0.000	0.627
	0.5	μ_1	AIC	0.863	0.000	0.137	0.329	0.000	0.671
			BIC	0.789	0.000	0.211	0.223	0.000	0.777
			FIC	0.911	0.000	0.089	0.369	0.000	0.651
		μ_2	AIC	0.866	0.000	0.134	0.320	0.000	0.680
			BIC	0.747	0.000	0.253	0.297	0.000	0.703
			FIC	0.941	0.000	0.059	0.347	0.000	0.653
		μ_3	AIC	0.873	0.000	0.127	0.309	0.000	0.691
			BIC	0.722	0.000	0.278	0.280	0.000	0.720
			FIC	0.918	0.000	0.082	0.354	0.000	0.646

μ_1 is the hazards ratio; μ_2 is the cumulative baseline hazard at time $t = 1$; μ_3 is the survivor function at time $t = 1$.

p(True) is the proportion of selecting the true model; p(S) is the proportion of selecting additional variables; p(FN) is the proportion of false exclusion of variables.

Naive: The naive method is the approach of implementing (3.1) and (3.8) with the difference in X^* and X ignored.

Proposed: The proposed method is described in Section 3.

(1): $\eta = (0,0,0,0,0,0)^\top$; (2): $\eta = (1,1,1,0,0,0)^\top$; (3): $\eta = (1,1,1,1,1,1)^\top$.

TABLE 2
Simulation results: RMSE of the estimators for the focus parameters with $n = 100$ and $\lambda_0(t) = 2t$.

Method	σ_ϵ		η	AIC	BIC	FIC	sAIC	sBIC	sFIC
Proposed	0.1	μ_1	(1)	0.358	0.266	0.264	0.357	0.262	0.261
			(2)	0.379	0.285	0.274	0.374	0.281	0.270
			(3)	0.385	0.297	0.286	0.379	0.293	0.282
		μ_2	(1)	0.287	0.267	0.268	0.280	0.263	0.262
			(2)	0.296	0.285	0.272	0.292	0.283	0.273
			(3)	0.312	0.294	0.284	0.304	0.292	0.286
		μ_3	(1)	0.156	0.145	0.148	0.152	0.144	0.141
			(2)	0.184	0.159	0.152	0.183	0.155	0.153
			(3)	0.212	0.182	0.160	0.210	0.180	0.164
	0.5	μ_1	(1)	0.370	0.283	0.282	0.367	0.282	0.279
			(2)	0.388	0.296	0.286	0.382	0.292	0.284
			(3)	0.391	0.327	0.297	0.387	0.316	0.293
		μ_2	(1)	0.304	0.297	0.290	0.297	0.295	0.288
			(2)	0.336	0.322	0.307	0.331	0.320	0.296
			(3)	0.349	0.326	0.314	0.346	0.324	0.312
		μ_3	(1)	0.252	0.234	0.232	0.247	0.050	0.045
			(2)	0.266	0.247	0.239	0.263	0.243	0.232
			(3)	0.287	0.253	0.246	0.282	0.250	0.241
Naive	0.1	μ_1	(1)	0.953	0.927	0.924	0.951	0.925	0.919
			(2)	0.960	0.943	0.930	0.958	0.938	0.927
			(3)	0.989	0.956	0.943	0.983	0.950	0.939
		μ_2	(1)	0.946	0.933	0.932	0.943	0.930	0.924
			(2)	0.953	0.943	0.939	0.950	0.937	0.934
			(3)	0.975	0.958	0.945	0.971	0.956	0.949
		μ_3	(1)	0.973	0.948	0.946	0.871	0.947	0.944
			(2)	0.986	0.962	0.955	0.981	0.956	0.953
			(3)	0.994	0.976	0.965	0.986	0.970	0.967
	0.5	μ_1	(1)	0.972	0.958	0.957	0.964	0.951	0.948
			(2)	0.997	0.977	0.968	0.992	0.971	0.962
			(3)	1.019	0.994	0.975	1.016	0.987	0.970
		μ_2	(1)	0.950	0.946	0.941	0.947	0.942	0.937
			(2)	0.987	0.966	0.953	0.981	0.962	0.952
			(3)	1.011	0.984	0.969	1.003	0.983	0.961
		μ_3	(1)	0.976	0.967	0.964	0.974	0.965	0.960
			(2)	0.998	0.981	0.978	0.996	0.977	0.974
			(3)	1.025	1.006	0.981	1.015	0.996	0.983

μ_1 is the hazards ratio; μ_2 is the cumulative baseline hazard at time $t = 1$; μ_3 is the survivor function at time $t = 1$.
Naive: The naive method is the approach of implementing (3.1) and (3.8) with the difference in X^* and X ignored.
Proposed: The proposed method is described in Section 3.
(1): $\eta = (0, 0, 0, 0, 0, 0)^\top$; (2): $\eta = (1, 1, 1, 0, 0, 0)^\top$; (3): $\eta = (1, 1, 1, 1, 1, 1)^\top$.

TABLE 3
Simulation results: RMSE of the estimators for the focus parameters with $n = 200$ and $\lambda_0(t) = 2t$.

Method	σ_ϵ		η	AIC	BIC	FIC	sAIC	sBIC	sFIC
Proposed	0.1	μ_1	(1)	0.338	0.247	0.245	0.335	0.244	0.241
			(2)	0.356	0.264	0.249	0.353	0.262	0.243
			(3)	0.379	0.286	0.263	0.374	0.282	0.260
		μ_2	(1)	0.210	0.204	0.202	0.208	0.202	0.198
			(2)	0.226	0.210	0.207	0.222	0.208	0.204
			(3)	0.239	0.224	0.211	0.235	0.223	0.209
		μ_3	(1)	0.142	0.138	0.136	0.140	0.135	0.132
			(2)	0.177	0.147	0.142	0.174	0.144	0.143
			(3)	0.195	0.156	0.149	0.192	0.155	0.147
	0.5	μ_1	(1)	0.357	0.268	0.267	0.356	0.263	0.260
			(2)	0.362	0.279	0.271	0.357	0.276	0.273
			(3)	0.385	0.293	0.276	0.380	0.292	0.278
		μ_2	(1)	0.234	0.224	0.221	0.230	0.223	0.220
			(2)	0.244	0.231	0.226	0.243	0.229	0.227
			(3)	0.251	0.244	0.231	0.248	0.240	0.227
		μ_3	(1)	0.228	0.218	0.217	0.226	0.215	0.213
			(2)	0.247	0.230	0.226	0.245	0.227	0.224
			(3)	0.253	0.242	0.231	0.250	0.239	0.228
Naive	0.1	μ_1	(1)	0.936	0.915	0.913	0.930	0.913	0.912
			(2)	0.944	0.927	0.922	0.938	0.924	0.923
			(3)	0.993	0.943	0.929	0.990	0.941	0.924
		μ_2	(1)	0.932	0.926	0.923	0.932	0.924	0.921
			(2)	0.944	0.935	0.928	0.942	0.929	0.924
			(3)	0.968	0.947	0.937	0.964	0.942	0.938
		μ_3	(1)	0.966	0.936	0.927	0.963	0.933	0.930
			(2)	0.975	0.945	0.930	0.972	0.942	0.933
			(3)	0.987	0.961	0.943	0.985	0.955	0.938
	0.5	μ_1	(1)	0.951	0.929	0.927	0.947	0.923	0.920
			(2)	0.964	0.938	0.930	0.962	0.936	0.932
			(3)	1.015	0.971	0.942	1.010	0.968	0.941
		μ_2	(1)	0.943	0.937	0.935	0.939	0.934	0.931
			(2)	0.978	0.956	0.944	0.974	0.951	0.946
			(3)	0.995	0.975	0.958	0.993	0.971	0.952
		μ_3	(1)	0.971	0.952	0.950	0.969	0.950	0.947
			(2)	0.994	0.968	0.959	0.992	0.962	0.953
			(3)	1.012	0.984	0.967	1.009	0.979	0.964

μ_1 is the hazards ratio; μ_2 is the cumulative baseline hazard at time $t = 1$; μ_3 is the survivor function at time $t = 1$.
Naive: The naive method is the approach of implementing (3.1) and (3.8) with the difference in X^* and X ignored.
Proposed: The proposed method is described in Section 3.
(1): $\eta = (0, 0, 0, 0, 0, 0)^\top$; (2): $\eta = (1, 1, 1, 0, 0, 0)^\top$; (3): $\eta = (1, 1, 1, 1, 1, 1)^\top$.

TABLE 4
Simulation results: RMSE of the estimators for the focus parameters with $n = 400$ and $\lambda_0(t) = 2t$.

Method	σ_ϵ		η	AIC	BIC	FIC	sAIC	sBIC	sFIC
Proposed	0.1	μ_1	(1)	0.324	0.224	0.222	0.320	0.223	0.221
			(2)	0.335	0.251	0.239	0.331	0.248	0.236
			(3)	0.346	0.289	0.250	0.343	0.285	0.246
		μ_2	(1)	0.175	0.171	0.168	0.172	0.163	0.160
			(2)	0.180	0.178	0.171	0.175	0.168	0.164
			(3)	0.202	0.198	0.193	0.197	0.190	0.188
		μ_3	(1)	0.136	0.130	0.129	0.134	0.128	0.127
			(2)	0.151	0.145	0.133	0.150	0.141	0.134
			(3)	0.162	0.149	0.137	0.160	0.148	0.138
	0.5	μ_1	(1)	0.337	0.245	0.243	0.335	0.243	0.241
			(2)	0.346	0.257	0.249	0.343	0.255	0.246
			(3)	0.355	0.278	0.256	0.353	0.275	0.254
		μ_2	(1)	0.184	0.176	0.171	0.182	0.172	0.173
			(2)	0.195	0.188	0.179	0.193	0.186	0.176
			(3)	0.217	0.197	0.183	0.211	0.192	0.180
		μ_3	(1)	0.145	0.137	0.139	0.144	0.133	0.130
			(2)	0.160	0.155	0.146	0.157	0.152	0.141
			(3)	0.178	0.169	0.152	0.176	0.168	0.148
Naive	0.1	μ_1	(1)	0.902	0.892	0.886	0.899	0.890	0.884
			(2)	0.913	0.905	0.894	0.907	0.901	0.896
			(3)	0.925	0.914	0.909	0.920	0.911	0.903
		μ_2	(1)	0.894	0.863	0.864	0.890	0.862	0.860
			(2)	0.905	0.879	0.871	0.902	0.876	0.868
			(3)	0.923	0.895	0.879	0.920	0.891	0.876
		μ_3	(1)	0.898	0.876	0.878	0.896	0.875	0.873
			(2)	0.912	0.898	0.882	0.909	0.895	0.879
			(3)	0.927	0.915	0.890	0.923	0.913	0.892
	0.5	μ_1	(1)	0.925	0.916	0.914	0.920	0.912	0.910
			(2)	0.934	0.927	0.918	0.929	0.922	0.916
			(3)	0.948	0.936	0.922	0.945	0.933	0.923
		μ_2	(1)	0.916	0.875	0.873	0.912	0.873	0.870
			(2)	0.924	0.898	0.879	0.920	0.896	0.876
			(3)	0.937	0.923	0.895	0.934	0.921	0.896
		μ_3	(1)	0.920	0.884	0.882	0.916	0.882	0.880
			(2)	0.936	0.915	0.894	0.933	0.912	0.896
			(3)	0.949	0.927	0.905	0.946	0.953	0.902

μ_1 is the hazards ratio; μ_2 is the cumulative baseline hazard at time $t = 1$; μ_3 is the survivor function at time $t = 1$.
Naive: The naive method is the approach of implementing (3.1) and (3.8) with the difference in X^* and X ignored.
Proposed: The proposed method is described in Section 3.
(1): $\eta = (0, 0, 0, 0, 0, 0)^\top$; (2): $\eta = (1, 1, 1, 0, 0, 0)^\top$; (3): $\eta = (1, 1, 1, 1, 1, 1)^\top$.

Table 5
Simulation results: RMSE of the estimators for the focus parameters with $n = 100$ and $\lambda_0(t) = \log t$.

Method	σ_ϵ		η	AIC	BIC	FIC	sAIC	sBIC	sFIC
Proposed	0.1	μ_1	(1)	0.391	0.283	0.284	0.388	0.280	0.278
			(2)	0.397	0.295	0.288	0.393	0.291	0.284
			(3)	0.414	0.308	0.294	0.409	0.306	0.296
		μ_2	(1)	0.166	0.140	0.141	0.162	0.138	0.136
			(2)	0.185	0.179	0.171	0.182	0.176	0.169
			(3)	0.203	0.198	0.183	0.200	0.195	0.181
		μ_3	(1)	0.146	0.140	0.142	0.144	0.139	0.138
			(2)	0.165	0.158	0.144	0.162	0.154	0.142
			(3)	0.173	0.164	0.155	0.171	0.161	0.151
	0.5	μ_1	(1)	0.411	0.322	0.324	0.408	0320	0.318
			(2)	0.437	0.339	0.328	0.423	0.331	0.329
			(3)	0.448	0.350	0.336	0.440	0.347	0.333
		μ_2	(1)	0.175	0.156	0.154	0.171	0.152	0.150
			(2)	0.194	0.186	0.178	0.190	0.188	0.179
			(3)	0.221	0.214	0.196	0.217	0.210	0.194
		μ_3	(1)	0.167	0.158	0.157	0.165	0.155	0.154
			(2)	0.174	0.168	0.161	0.170	0.166	0.159
			(3)	0.182	0.173	0.169	0.178	0.170	0.167
Naive	0.1	μ_1	(1)	0.934	0.916	0.915	0.931	0.914	0.911
			(2)	0.946	0.925	0.919	0.944	0.923	0.917
			(3)	0.957	0.948	0.928	0.953	0.947	0.926
		μ_2	(1)	0.912	0.895	0.896	0.908	0.896	0.894
			(2)	0.930	0.916	0.899	0.927	0.910	0.895
			(3)	0.957	0.939	0.925	0.956	0.935	0.921
		μ_3	(1)	0.933	0.917	0.915	0.930	0.914	0.914
			(2)	0.948	0.926	0.922	0.947	0.924	0.920
			(3)	0.955	0.934	0.928	0.952	0.933	0.926
	0.5	μ_1	(1)	0.950	0.934	0.935	0.948	0.933	0.932
			(2)	0.963	0.942	0.938	0.960	0.939	0.936
			(3)	0.978	0.966	0.951	0.975	0.964	0.949
		μ_2	(1)	0.933	0.918	0.916	0.930	0.916	0.917
			(2)	0.945	0.924	0.920	0.944	0.922	0.919
			(3)	0.966	0.942	0.934	0.963	0.940	0.932
		μ_3	(1)	0.946	0.925	0.924	0.944	0.924	0.923
			(2)	0.957	0.938	0.931	0.955	0.936	0.929
			(3)	0.967	0.945	0.934	0.965	0.943	0.935

μ_1 is the hazards ratio; μ_2 is the cumulative baseline hazard at time $t = 1$; μ_3 is the survivor function at time $t = 1$.
Naive: The naive method is the approach of implementing (3.1) and (3.8) with the difference in X^* and X ignored.
Proposed: The proposed method is described in Section 3.
(1): $\eta = (0, 0, 0, 0, 0, 0)^\top$; (2): $\eta = (1, 1, 1, 0, 0, 0)^\top$; (3): $\eta = (1, 1, 1, 1, 1, 1)^\top$.

Table 6
Simulation results: RMSE of the estimators for the focus parameters with $n = 200$ and $\lambda_0(t) = \log t$.

Method	σ_ϵ		η	AIC	BIC	FIC	sAIC	sBIC	sFIC
Proposed	0.1	μ_1	(1)	0.375	0.264	0.260	0.367	0.258	0.256
			(2)	0.379	0.278	0.266	0.370	0.274	0.268
			(3)	0.383	0.286	0.270	0.379	0.279	0.271
		μ_2	(1)	0.154	0.133	0.132	0.150	0.131	0.129
			(2)	0.179	0.172	0.163	0.175	0.170	0.164
			(3)	0.195	0.183	0.171	0.188	0.179	0.173
		μ_3	(1)	0.137	0.130	0.131	0.134	0.129	0.128
			(2)	0.150	0.145	0.138	0.146	0.142	0.139
			(3)	0.162	0.153	0.145	0.159	0.150	0.142
	0.5	μ_1	(1)	0.384	0.277	0.276	0.380	0.273	0.272
			(2)	0.391	0.285	0.279	0.383	0.280	0.281
			(3)	0.398	0.297	0.286	0.395	0.293	0.284
		μ_2	(1)	0.164	0.146	0.145	0.160	0.143	0.141
			(2)	0.185	0.178	0.169	0.181	0.175	0.166
			(3)	0.203	0.192	0.178	0.196	0.188	0.177
		μ_3	(1)	0.155	0.143	0.141	0.154	0.141	0.140
			(2)	0.167	0.158	0.149	0.166	0.156	0.147
			(3)	0.178	0.166	0.157	0.177	0.163	0.155
Naive	0.1	μ_1	(1)	0.916	0.897	0.895	0.914	0.894	0.891
			(2)	0.924	0.905	0.903	0.919	0.899	0.896
			(3)	0.933	0.920	0.911	0.930	0.917	0.907
		μ_2	(1)	0.894	0.879	0.876	0.889	0.876	0.874
			(2)	0.915	0.893	0.882	0.910	0.890	0.885
			(3)	0.933	0.917	0.895	0.928	0.910	0.891
		μ_3	(1)	0.912	0.896	0.895	0.910	0.895	0.893
			(2)	0.926	0.913	0.907	0.922	0.911	0.909
			(3)	0.934	0.922	0.915	0.929	0.920	0.910
	0.5	μ_1	(1)	0.947	0.925	0.926	0.944	0.923	0.921
			(2)	0.956	0.933	0.929	0.951	0.930	0.930
			(3)	0.962	0.941	0.934	0.959	0.938	0.935
		μ_2	(1)	0.926	0.897	0.896	0.923	0.895	0.893
			(2)	0.937	0.918	0.907	0.933	0.914	0.901
			(3)	0.949	0.926	0.914	0.945	0.922	0.910
		μ_3	(1)	0.928	0.917	0.915	0.926	0.915	0.914
			(2)	0.936	0.924	0.919	0.934	0.920	0.917
			(3)	0.947	0.933	0.926	0.945	0.931	0.928

μ_1 is the hazards ratio; μ_2 is the cumulative baseline hazard at time $t = 1$; μ_3 is the survivor function at time $t = 1$.
Naive: The naive method is the approach of implementing (3.1) and (3.8) with the difference in X^* and X ignored.
Proposed: The proposed method is described in Section 3.
(1): $\eta = (0, 0, 0, 0, 0, 0)^\top$; (2): $\eta = (1, 1, 1, 0, 0, 0)^\top$; (3): $\eta = (1, 1, 1, 1, 1, 1)^\top$.

TABLE 7
Simulation results: RMSE of the estimators for the focus parameters with $n = 400$ and $\lambda_0(t) = \log t$.

Method	σ_ϵ		η	AIC	BIC	FIC	sAIC	sBIC	sFIC
Proposed	0.1	μ_1	(1)	0.347	0.233	0.220	0.299	0.220	0.211
			(2)	0.353	0.242	0.236	0.352	0.252	0.246
			(3)	0.360	0.255	0.243	0.372	0.259	0.236
		μ_2	(1)	0.142	0.126	0.125	0.138	0.123	0.120
			(2)	0.173	0.167	0.155	0.169	0.163	0.158
			(3)	0.187	0.177	0.163	0.182	0.173	0.168
		μ_3	(1)	0.124	0.120	0.118	0.121	0.119	0.117
			(2)	0.126	0.123	0.120	0.124	0.122	0.119
			(3)	0.129	0.125	0.122	0.128	0.125	0.123
	0.5	μ_1	(1)	0.355	0.253	0.250	0.306	0.245	0.240
			(2)	0.369	0.278	0.257	0.366	0.275	0.254
			(3)	0.370	0.282	0.261	0.367	0.279	0.258
		μ_2	(1)	0.148	0.133	0.131	0.145	0.129	0.127
			(2)	0.179	0.172	0.160	0.176	0.169	0.162
			(3)	0.193	0.184	0.175	0.188	0.181	0.170
		μ_3	(1)	0.135	0.129	0.123	0.133	0.126	0.122
			(2)	0.138	0.133	0.126	0.135	0.130	0.124
			(3)	0.146	0.137	0.128	0.143	0.135	0.127
Naive	0.1	μ_1	(1)	0.908	0.889	0.877	0.903	0.887	0.872
			(2)	0.913	0.895	0.882	0.909	0.889	0.875
			(3)	0.922	0.904	0.896	0.917	0.890	0.883
		μ_2	(1)	0.883	0.875	0.869	0.877	0.860	0.858
			(2)	0.890	0.879	0.874	0.883	0.869	0.863
			(3)	0.897	0.886	0.879	0.893	0.878	0.871
		μ_3	(1)	0.894	0.882	0.875	0.887	0.876	0.860
			(2)	0.909	0.894	0.883	0.894	0.887	0.878
			(3)	0.916	0.899	0.890	0.905	0.893	0.888
	0.5	μ_1	(1)	0.935	0.901	0.897	0.924	0.896	0.893
			(2)	0.944	0.912	0.903	0.931	0.905	0.897
			(3)	0.952	0.934	0.921	0.945	0.922	0.914
		μ_2	(1)	0.902	0.883	0.883	0.895	0.874	0.871
			(2)	0.916	0.892	0.887	0.911	0.884	0.881
			(3)	0.929	0.916	0.895	0.923	0.910	0.892
		μ_3	(1)	0.913	0.892	0.889	0.908	0.883	0.883
			(2)	0.925	0.913	0.899	0.917	0.897	0.893
			(3)	0.933	0.919	0.906	0.920	0.910	0.903

μ_1 is the hazards ratio; μ_2 is the cumulative baseline hazard at time $t = 1$; μ_3 is the survivor function at time $t = 1$.
Naive: The naive method is the approach of implementing (3.1) and (3.8) with the difference in X^* and X ignored.
Proposed: The proposed method is described in Section 3.
(1): $\eta = (0, 0, 0, 0, 0, 0)^\top$; (2): $\eta = (1, 1, 1, 0, 0, 0)^\top$; (3): $\eta = (1, 1, 1, 1, 1, 1)^\top$.

Table 8. *Sensitivity analyses for Worcester Heart Attack Study data: estimation results of full models*

Method	Variable	Estimate	SE	p-value
Proposed ($R = 0.65$)	Initial Heart Rate (X_1)	0.970	0.356	0.006
	Initial Systolic Blood Pressure (X_2)	0.472	0.158	0.003
	Initial Diastolic Blood Pressure (X_3)	-0.953	0.189	4.599e-07
	Body Mass Index (X_4)	-1.510	0.238	2.231e-10
	Age (Z_1)	0.074	0.023	0.001
	Gender (Z_2)	-0.549	0.182	0.002
Proposed ($R = 0.75$)	Initial Heart Rate (X_1)	0.986	0.317	1.811e-02
	Initial Systolic Blood Pressure (X_2)	0.287	0.115	1.221e-02
	Initial Diastolic Blood Pressure (X_3)	-0.855	0.164	1.893e-07
	Body Mass Index (X_4)	-1.745	0.223	6.548e-15
	Age (Z_1)	0.055	0.018	5.265e-11
	Gender (Z_2)	-0.394	0.155	1.157e-12
Proposed ($R = 0.85$)	Initial Heart Rate (X_1)	0.994	0.234	2.167e-05
	Initial Systolic Blood Pressure (X_2)	0.396	0.079	5.247e-07
	Initial Diastolic Blood Pressure (X_3)	-0.889	0.113	1.196e-21
	Body Mass Index (X_4)	-1.762	0.100	0.000
	Age (Z_1)	0.054	0.015	2.920e-24
	Gender (Z_2)	-0.361	0.139	7.892e-20
Naive	Initial Heart Rate (X_1)	0.894	0.110	4.391e-16
	Initial Systolic Blood Pressure (X_2)	0.326	0.047	4.029e-12
	Initial Diastolic Blood Pressure (X_3)	-0.834	0.109	1.988e-14
	Body Mass Index (X_4)	-1.694	0.010	0.000
	Age (Z_1)	0.054	0.007	1.217e-14
	Gender (Z_2)	-0.379	0.116	0.001

Table 9. *Sensitivity analyses for Worcester Heart Attack Study data: variable selection results*

Method	AIC		BIC		FIC $-\mu_1$		FIC $-\mu_2$		FIC $-\mu_3$	
	Variables	Values	Variables	Values	Variables	Values	Variables	Values	Variables	Values
Proposed	1234	-7541.776	234	-7527.376	4	0.593	12	0.596	24	0.589
($R = 0.65$)	234	-7539.776	123	-7526.938	14	0.726	13	0.617	23	0.633
	123	-7537.720	23	-7526.647	134	0.857	123	0.684	123	0.674
	134	-7537.151	1234	-7525.399	13	0.868	234	0.718	134	0.705
	124	-7536.616	134	-7524.780	34	0.889	124	0.739	234	0.723
Proposed	1234	-7541.495	234	-7527.767	4	0.679	12	0.625	24	0.611
($R = 0.75$)	234	-7539.439	123	-7527.348	14	0.782	13	0.705	23	0.689
	123	-7538.764	23	-7527.039	3	0.977	123	0.740	13	0.722
	134	-7537.026	1234	-7526.363	23	1.059	124	0.756	134	0.736
	124	-7536.670	134	-7524.961	13	1.076	234	0.780	123	0.755
Proposed	1234	-7542.056	234	-7528.134	4	0.640	12	0.655	24	0.630
($R = 0.85$)	234	-7539.454	123	-7527.860	14	0.787	13	0.713	23	0.716
	123	-7537.720	23	-7527.053	134	0.819	124	0.792	13	0.754
	134	-7537.191	1234	-7525.522	13	0.941	123	0.859	123	0.820
	124	-7536.401	134	-7525.320	23	1.003	234	0.876	134	0.844
Naive	134	-7755.097	134	-7762.497	24	0.709	234	2.601	2	1.228
	123	-7755.013	13	-7764.221	1234	1.304	23	2.632	24	1.840
	13	-7755.954	14	-7766.855	123	1.330	24	2.650	1	2.004
	1234	-7756.657	123	-7767.413	234	1.332	123	2.885	1234	2.175
	14	-7758.589	1	-7768.988	23	1.342	124	3.067	4	3.042

Table 10
Sensitivity analyses for Worcester Heart Attack Study data: estimates of the focus parameters

μ	Method	AIC	sAIC	BIC	sBIC	FIC	sFIC
μ_1	Proposed $(R = 0.65)$	0.224	0.236	0.085	0.106	0.137	0.155
	Proposed $(R = 0.75)$	0.208	0.227	0.077	0.107	0.126	0.147
	Proposed $(R = 0.85)$	0.189	0.206	0.071	0.113	0.124	0.141
	Naive	0.141	0.169	0.141	0.158	0.184	0.193
μ_2	Proposed $(R = 0.65)$	0.377	0.388	0.356	0.380	0.288	0.290
	Proposed $(R = 0.75)$	0.367	0.386	0.344	0.377	0.272	0.275
	Proposed $(R = 0.85)$	0.366	0.385	0.343	0.374	0.264	0.268
	Naive	0.348	0.356	0.348	0.352	0.257	0.262
μ_3	Proposed $(R = 0.65)$	0.718	0.722	0.699	0.710	0.680	0.681
	Proposed $(R = 0.75)$	0.714	0.721	0.665	0.676	0.649	0.653
	Proposed $(R = 0.85)$	0.689	0.696	0.665	0.687	0.637	0.644
	Naive	0.603	0.612	0.603	0.608	0.461	0.507

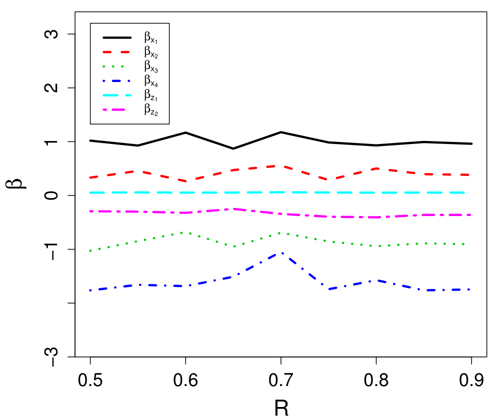

Fig 2. *Sensitivity analyses of the estimates obtained for WHAS500 data.*

Appendix A: Regularity conditions

(C1) Θ is a compact set, and the parameter value β_0 is an interior point of Θ.
(C2) $\int_0^\tau \lambda_0(t)dt < \infty$, where τ is the time that the study ends.
(C3) The $\{N_i(t), Y_i(t), Z_i, X_i^*\}$ are independent and identically distributed for $i = 1, \cdots, n$.
(C4) The covariates Z and X^* are bounded with probability one. That is, there exist finite numbers a_x, b_x, a_z, and b_z with $a_x < b_x$ and $a_z < b_z$ such that $a_x < X^* < b_x$ and $a_z < Z < b_z$ with probability one.
(C5) Conditional on V, $\left(\widetilde{T}, C, V\right)$ are independent of $\widetilde{A}$.
(C6) Censoring times are noninformative. That is, the failure time and the censoring time are independent, given the covariates.
(C7) The matrix $\mathcal{A} \triangleq E\left[-\frac{1}{n}\frac{\partial}{\partial\beta}\left\{U_{\mathrm{P}}\left(0_p, \beta_{z0}\right) + U_{\mathrm{M}}\left(0_p, \beta_{z0}\right)\right\}\right]$ is assumed to be invertible and positive definite.

Condition (C1) is used to derive the maximizer from the target function. Condition (C4) is commonly assumed in the literature (e.g., [3, 14, 25]). Other conditions are standard in survival analysis which allow us to derive the asymptotic properties of the estimators.

Appendix B: Proofs for the results in Sections 4 and 5

B.1. Proof of Lemma 4.1

For any given candidate model S, we have that

$$\begin{aligned}\Sigma_{X_S^*} &= E\left\{\left(X_S^* - \mu_{X_S^*}\right)\left(X_S^* - \mu_{X_S^*}\right)^\top\right\}\\ &= E\left\{\left(\pi_S X^* - \pi_S\mu_{X^*}\right)\left(\pi_S X^* - \pi_S\mu_{X^*}\right)^\top\right\}\\ &= \pi_S E\left\{\left(X^* - \mu_{X^*}\right)\left(X^* - \mu_{X^*}\right)^\top\right\}\pi_S^\top\\ &= \pi_S\Sigma_{X^*}\pi_S^\top,\end{aligned}$$

which yields that $I_{|S|\times|S|} = \Sigma_{X_S^*}\cdot\Sigma_{X_S^*}^{-1} = \pi_S\Sigma_{X^*}\pi_S^\top\cdot\Sigma_{X_S^*}^{-1}$, provided $\Sigma_{X_S^*}^{-1}$ exists. Multiplying π_S on both sides gives

$$\pi_S\Sigma_{X^*}\pi_S^\top\cdot\Sigma_{X_S^*}^{-1}\pi_S = \pi_S$$

or $\pi_S\left(\Sigma_{X^*}\pi_S^\top\cdot\Sigma_{X_S^*}^{-1}\pi_S - I_{|S|\times|S|}\right) = 0$, which implies that

$$\Sigma_{X^*}\pi_S^\top\cdot\Sigma_{X_S^*}^{-1}\pi_S = I_{|S|\times|S|},$$

or equivalently,

$$\pi_S^\top\cdot\Sigma_{X_S^*}^{-1}\pi_S = \Sigma_{X^*}^{-1},$$

and this proof is completed. □

B.2. Proof of Lemma 4.2

Proof of (a):
First, for any candidate model S, we denote

$$\begin{aligned} & G_S^{(1)}\left(u, \beta_{x,S}, \beta_{z,S}\right) \\ = \; & \frac{1}{n} \sum_{i=1}^{n} \Pi_S \begin{pmatrix} x_i^* \\ z_i \end{pmatrix} Y_i(u) \exp\left\{ \left((\pi_S x_i^*)^\top, z_i^\top \right) \begin{pmatrix} \beta_{x,S} \\ \beta_{z,S} \end{pmatrix} \right\}. \end{aligned} \quad \text{(B.1)}$$

Let $G^{(0)}\left(u, \beta_x, \beta_z\right)$ denote (3.7) when S is the full model, and let

$$G^{(1)}\left(u, \beta_x, \beta_z\right) = \begin{pmatrix} G_x^{(1)}(u, \beta_x, \beta_z) \\ G_z^{(1)}(u, \beta_x, \beta_z) \end{pmatrix}, \quad \text{(B.2)}$$

where

$$\begin{pmatrix} G_x^{(1)}(u, \beta_x, \beta_z) \\ G_z^{(1)}(u, \beta_x, \beta_z) \end{pmatrix} \triangleq \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} x_i^* \\ z_i \end{pmatrix} Y_i(u) \exp\left\{ \left(x_i^{*\top}, z_i^\top \right) \begin{pmatrix} \beta_x \\ \beta_z \end{pmatrix} \right\},$$

and

$$G^{(2)}\left(u, \beta_x, \beta_z\right) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} x_i^* \\ z_i \end{pmatrix}^{\otimes 2} Y_i(u) \exp\left\{ \left(x_i^{*\top}, z_i^\top \right) \begin{pmatrix} \beta_x \\ \beta_z \end{pmatrix} \right\}, \quad \text{(B.3)}$$

where $a^{\otimes 2} = aa^\top$ for any vector a.

Then setting $(\beta_x, \beta_z) = \left(0_{|S|}, \beta_z\right)$ gives

$$\begin{aligned} G_S^{(1)}\left(u, 0_{|S|}, \beta_z\right) &= \frac{1}{n} \sum_{i=1}^{n} \Pi_S \begin{pmatrix} x_i^* \\ z_i \end{pmatrix} Y_i(u) \exp\left\{ \left((\pi_S x_i^*)^\top, z_i^\top \right) \begin{pmatrix} 0_{|S|} \\ \beta_z \end{pmatrix} \right\} \\ &= \frac{1}{n} \Pi_S \sum_{i=1}^{n} \begin{pmatrix} x_i^* \\ z_i \end{pmatrix} Y_i(u) \exp\left\{ \left(x_i^{*\top}, z_i^\top \right) \begin{pmatrix} 0_p \\ \beta_z \end{pmatrix} \right\} \\ &= \Pi_S G^{(1)}\left(u, 0_p, \beta_z\right). \end{aligned}$$

Similarly, from (3.7), one has

$$\begin{aligned} G_S^{(0)}\left(u, 0_{|S|}, \beta_z\right) &= \frac{1}{n} \sum_{i=1}^{n} Y_i(u) \exp\left\{ (\pi_S x_i^*)^\top 0_{|S|} + z_i^\top \beta_z \right\} \\ &= \frac{1}{n} \sum_{i=1}^{n} Y_i(u) \exp\left(z_i^\top \beta_z \right) \\ &= \frac{1}{n} \sum_{i=1}^{n} Y_i(u) \exp\left(x_i^{*\top} 0_p + z_i^\top \beta_z \right) \\ &= G^{(0)}\left(u, 0_p, \beta_z\right). \end{aligned}$$

Therefore, for any β_z and $j = 0, 1$, we have

$$G_S^{(j)}\left(u, 0_{|S|}, \beta_z\right) = \Pi_S^{\otimes j} G^{(j)}\left(u, 0_p, \beta_z\right), \tag{B.4}$$

where $A^{\otimes 0} = I_{p\times p}$ and $A^{\otimes 1} = A$ for any matrix A.

Consequently, direct calculations show that

$$\begin{aligned} U_{\mathrm{P},S}\left(\beta_{x,S}, \beta_{z,S}\right) &= \frac{\partial}{\partial \beta_S} \ell^*_{\mathrm{P},S}\left(\beta_S\right) \\ &= \sum_{i=1}^{n} \int_0^{\tau} \left\{ \Pi_S \begin{pmatrix} x_i^* \\ z_i \end{pmatrix} + \begin{pmatrix} \pi_S \Sigma_\epsilon \pi_S^\top \beta_{x,S} \\ 0_q \end{pmatrix} \right. \\ &\qquad \left. - \frac{G_S^{(1)}\left(u, \beta_{x,S}, \beta_{z,S}\right)}{G_S^{(0)}\left(u, \beta_{x,S}, \beta_{z,S}\right)} \right\} dN_i(u), \end{aligned}$$

and

$$\begin{aligned} U_{\mathrm{P}}\left(\beta_x, \beta_z\right) &= \frac{\partial}{\partial \beta} \ell^*_{\mathrm{P}}(\beta) \\ &= \sum_{i=1}^{n} \int_0^{\tau} \left\{ \begin{pmatrix} x_i^* \\ z_i \end{pmatrix} + \begin{pmatrix} \Sigma_\epsilon \beta_x \\ 0_q \end{pmatrix} - \frac{G^{(1)}\left(u, \beta_x, \beta_z\right)}{G^{(0)}\left(u, \beta_x, \beta_z\right)} \right\} dN_i(u). \end{aligned} \tag{B.5}$$

Thus, plugging in $\left(\beta_x^\top, \beta_z^\top\right)^\top = \left(0_p^\top, \beta_{z0}^\top\right)^\top$ and $\left(\beta_{x,S}^\top, \beta_{z,S}^\top\right)^\top = \left(0_{|S|}^\top, \beta_{z0}^\top\right)^\top$ to $U_{\mathrm{P},S}\left(\beta_{x,S}, \beta_{z,S}\right)$ and $U_{\mathrm{P}}\left(\beta_x, \beta_z\right)$, respectively, gives

$$U_{\mathrm{P},S}\left(0_{|S|}, \beta_{z0}\right) = \Pi_S U_{\mathrm{P}}\left(0_p, \beta_{z0}\right).$$

Proof of (b):

Let $\widehat{x}_i$ denote (3.10) when S is the full model. We first show the relationship between $\widehat{x}_{i,S}$ defined by (3.10) and $\widehat{x}_i$. Applying $x^*_{i,S} = \pi_S x_i^*$ to (3.10), we have

$$\begin{aligned} \widehat{x}_{i,S} &= \pi_S \widehat{\mu}_{X^*} + \left(I_{|S|\times|S|} - \pi_S \Sigma_\epsilon \pi_S^\top \widehat{\Sigma}^{-1}_{X_S^*}\right)\left(x^*_{i,S} - \widehat{\mu}_{X_S^*}\right) \\ &= \pi_S \widehat{\mu}_{X^*} + \left(I_{|S|\times|S|} - \pi_S \Sigma_\epsilon \pi_S^\top \widehat{\Sigma}^{-1}_{X_S^*}\right) \pi_S \left(x_i^* - \widehat{\mu}_{X^*}\right) \\ &= \pi_S \widehat{\mu}_{X^*} + \left(\pi_S - \pi_S \Sigma_\epsilon \pi_S^\top \widehat{\Sigma}^{-1}_{X_S^*} \pi_S\right)\left(x_i^* - \widehat{\mu}_{X^*}\right) \\ &= \pi_S \left\{ \widehat{\mu}_{X^*} + \left(I_{p\times p} - \Sigma_\epsilon \pi_S^\top \widehat{\Sigma}^{-1}_{X_S^*} \pi_S\right)\left(x_i^* - \widehat{\mu}_{X^*}\right) \right\} \\ &= \pi_S \left\{ \widehat{\mu}_{X^*} + \left(I_{p\times p} - \Sigma_\epsilon \widehat{\Sigma}^{-1}_{X^*}\right)\left(x_i^* - \widehat{\mu}_{X^*}\right) \right\} \\ &= \pi_S \left\{ \widehat{\mu}_{X^*} + \left(\widehat{\Sigma}_{X^*} - \Sigma_\epsilon\right)^\top \widehat{\Sigma}^{-1}_{X^*}\left(x_i^* - \widehat{\mu}_{X^*}\right) \right\} \\ &= \pi_S \widehat{x}_i, \end{aligned}$$

where the second identity is due to $\widehat{\mu}_{X_S^*} = \pi_S \widehat{\mu}_{X^*}$, and the third last step is due to Lemma 4.1.

To prove $U_{\mathrm{M},S}\left(0_{|S|}, \beta_{z0}\right) = \Pi_S U_{\mathrm{M}}(0_p, \beta_{z0})$, we first examine the partial derivative of $\widehat{\ell}^*_{\mathrm{M},S}$. Note that we can express $\widehat{\ell}^*_{\mathrm{M},S} = \widehat{\ell}^*_{\mathrm{M}1,S} - \widehat{\ell}^*_{\mathrm{M}2,S}$, where

$$\widehat{\ell}^*_{\mathrm{M}1,S} = \sum_{i=1}^n \left[\log\left\{d\widehat{H}_S(a_i)\right\} - \widehat{\Lambda}_{0,S}(a_i)\exp\left(\widehat{x}_{i,S}^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right)\right],$$

and

$$\widehat{\ell}^*_{\mathrm{M}2,S} = \sum_{i=1}^n \log\int_0^\tau \exp\left\{-\widehat{\Lambda}_{0,S}(u)\exp\left(\widehat{x}_{i,S}^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right)\right\} d\widehat{H}_S(u).$$

Let $\Xi = \begin{pmatrix} \pi_S \Sigma_\epsilon \pi_S^\top & 0_{|S|\times q} \\ 0_{q\times|S|} & 0_{q\times q}\end{pmatrix}$ and $\beta_S = \begin{pmatrix}\beta_{x,S} \\ \beta_{z,S}\end{pmatrix}$. Then direct calculations give us

$$\begin{aligned}
& U_{\mathrm{M}1,S}\left(\beta_{x,S}, \beta_{z,S}\right) \\
&= \frac{\partial}{\partial \beta_S}\widehat{\ell}^*_{\mathrm{M}1,S}\left(\beta_S\right) \\
&= \frac{\partial}{\partial \beta_S}\left(\sum_{i=1}^n \left[\log\left\{d\widehat{H}_S(a_i)\right\} - \widehat{\Lambda}_{0,S}(a_i)\exp\left(\widehat{x}_{i,S}^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right)\right]\right) \\
&= -\sum_{i=1}^n \frac{\partial}{\partial \beta_S}\left\{\widehat{\Lambda}_{0,S}(a_i)\exp\left(\widehat{x}_{i,S}^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right)\right\} \\
&= -\sum_{i=1}^n \frac{\partial}{\partial \beta_S}\left\{\int_0^{a_i} \frac{\frac{1}{n}\sum\limits_{j=1}^n dN_j(u)}{\left\{m_S\left(\beta_{x,S}\right)\right\}^{-1} G_S^{(0)}\left(u, \beta_{x,S}, \beta_{z,S}\right)} \exp\left(\widehat{x}_{i,S}^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right)\right\} \\
&= -\sum_{i=1}^n \left[\int_0^{a_i}\left\{\frac{-\frac{1}{n}\sum\limits_{j=1}^n dN_j\left(u\right)\left\{m_S\left(\beta_{x,S}\right)\right\}^{-1} G_S^{(1)}\left(u, \beta_{x,S}, \beta_{z,S}\right)}{\left[\left\{m_S\left(\beta_{x,S}\right)\right\}^{-1} G_S^{(0)}\left(u, \beta_{x,S}, \beta_{z,S}\right)\right]^2}\right.\right. \\
&\qquad \left. + \frac{\frac{1}{n}\sum\limits_{j=1}^n dN_j\left(u\right)\Xi\beta_S\left\{m_S\left(\beta_{x,S}\right)\right\}^{-1} G_S^{(0)}\left(u, \beta_{x,S}, \beta_{z,S}\right)}{\left[\left\{m_S\left(\beta_{x,S}\right)\right\}^{-1} G_S^{(0)}\left(u, \beta_{x,S}, \beta_{z,S}\right)\right]^2}\right\} \\
&\qquad \left. \times \exp\left(\widehat{x}_{i,S}^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right) + \widehat{\Lambda}_{0,S}\left(a_i\right)\Pi_S\begin{pmatrix}\widehat{x}_i \\ z_i\end{pmatrix}\exp\left\{\widehat{x}_{i,S}^\top \beta_{x,S} + z_i^\top \beta_{z,S}\right\}\right],
\end{aligned}$$

where the fourth equality is due to the estimator (3.6), the last equality is due to that

$$\frac{\partial m_S^{-1}\left(\beta_{x,S}\right)}{\partial \beta_S} \quad = \quad \frac{\partial}{\partial \beta_S}\exp\left(\frac{-1}{2}\beta_{x,S}^\top \pi_S \Sigma_\epsilon \pi_S^\top \beta_{x,S}\right)$$

$$
\begin{aligned}
&= \frac{\partial}{\partial \beta_S} \exp\left(\frac{-1}{2}\beta_S^\top \Xi \beta_S\right) \\
&= -\Xi \beta_S m_S^{-1}\left(\beta_{x,S}\right),
\end{aligned}
$$

and $m_S\left(\beta_{x,S}\right)$ is defined by $\exp\left(\frac{1}{2}\beta_{x,S}^\top \pi_S \Sigma_\epsilon \pi_S^\top \beta_{x,S}\right)$ in Section 3.1.

Then plugging in $\beta_{x,S} = 0_{|S|}$ and $\beta_{z,S} = \beta_{z0}$ to $U_{\mathrm{M1},S}\left(\beta_{x,S}, \beta_{z,S}\right)$ gives

$$
\begin{aligned}
U_{\mathrm{M1},S}\left(0_{|S|}, \beta_{z0}\right) &= -\sum_{i=1}^{n}\left[\int_0^{a_i} \frac{\frac{1}{n}\sum_{j=1}^{n} dN_j(u)\, \Pi_S G^{(1)}\left(u, 0_p, \beta_{z0}\right)}{\left\{G^{(0)}\left(u, 0_p, \beta_{z0}\right)\right\}^2} \exp\left\{z_i^\top \beta_{z0}\right\}\right. \\
&\qquad \left. + \widehat{\Lambda}_0\left(a_i\right) \Pi_S \begin{pmatrix} \widehat{x}_i \\ z_i \end{pmatrix} \exp\left\{z_i^\top \beta_{z0}\right\}\right] \\
&= \Pi_S U_{\mathrm{M1}}\left(u, 0_p, \beta_{z0}\right),
\end{aligned}
$$

where the last step is due to (B.4) and that $m_S(0_{|S|}) = 1$.

Similarly, we examine $U_{\mathrm{M2},S}\left(\beta_{x,S}, \beta_{z,S}\right) = \frac{\partial}{\partial \beta}\widehat{\ell}^*_{\mathrm{M2},S}$ and plug in $\beta_{x,S} = 0_{|S|}$ and $\beta_{z,S} = \beta_{z0}$ to $U_{\mathrm{M2},S}$ and apply Lemma 4.1, yielding

$$
\begin{aligned}
& U_{\mathrm{M2},S}\left(0_{|S|}, \beta_{z0}\right) \\
= \; & \frac{1}{\int_0^\tau \exp\left\{-\widehat{\Lambda}_0(u) \exp\left(z_i^\top \beta_{z0}\right)\right\} d\widehat{H}_S(u)} \\
& \times \int_0^\tau \exp\left\{-\widehat{\Lambda}_0(u) \exp\left(z_i^\top \beta_{z0}\right)\right\} \left\{\Pi_S\left(\frac{\partial}{\partial \beta}\widehat{\Lambda}_0(u)\right) \exp\left(z_i^\top \beta_{z0}\right)\right. \\
& \left. + \widehat{\Lambda}_0(u)\, \Pi_S \begin{pmatrix} \widehat{x}_i \\ z_i \end{pmatrix} \exp\left(z_i^\top \beta_{z0}\right)\right\} d\widehat{H}_S(u) \\
= \; & \Pi_S \frac{\partial}{\partial \beta} \log \int_0^\tau \exp\left\{-\widehat{\Lambda}_0(u) \exp\left(\widehat{x}_i^\top \beta_x + z_i^\top \beta_z\right)\right\} d\widehat{H}_S(u)\Big|_{\beta_x = 0_p, \beta_z = \beta_{z0}} \\
= \; & \Pi_S U_{\mathrm{M2}}\left(u, 0_p, \beta_{z0}\right).
\end{aligned}
$$

Thus, we complete the proof. □

B.3. Proof of Theorem 4.1

B.3.1. Proof of (a)

To show it, we first prove the asymptotic result for the full model, and then apply it to the candidate model S to yield the desired result.

Step 1: Show that as $n \to \infty$,

$$
\sqrt{n}\begin{pmatrix} \widehat{\beta}_{x,\mathrm{full}} \\ \widehat{\beta}_{z,\mathrm{full}} - \beta_{z0} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \eta \\ 0 \end{pmatrix}, \mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1}\right). \tag{B.6}
$$

The proof consists of the following two steps.

Step 1.1:

Let $U_{\mathrm{P}}(\beta_x, \beta_z) = \frac{\partial \ell^*_{\mathrm{P}}}{\partial \beta}$, $U_{\mathrm{M}}(\beta_x, \beta_z) = \frac{\partial \widehat{\ell}^*_{\mathrm{M}}}{\partial \beta}$ and $U(\beta_x, \beta_z) = U_{\mathrm{P}}(\beta_x, \beta_z) + U_{\mathrm{M}}(\beta_x, \beta_z)$, where $\beta = \left(\beta_x^\top, \beta_z^\top\right)^\top$, and ℓ^*_{P} and $\widehat{\ell}^*_{\mathrm{M}}$ are given by (4.1) and (3.14) with S taken as the full model, respectively. By the fact that $\widehat{\ell}^*_{\mathrm{M}}$ and ℓ^*_{P}, defined in Sections 3.2 and 4.1, respectively, are twice continuously differentiable in the parameters and the covariates as well as Conditions (C1) and (C4), we conclude that $U_{\mathrm{P}}(\beta_x, \beta_z)$ and $U_{\mathrm{M}}(\beta_x, \beta_z)$ and their derivative are continuous and bounded. Applying the Taylor series expansion of $U\left(\widehat{\beta}_{x,\text{full}}, \widehat{\beta}_{z,\text{full}}\right)$ and $U\left(\frac{\eta}{\sqrt{n}}, \beta_{z0}\right)$ around $\left(\beta_x^\top, \beta_z^\top\right)^\top = \left(0_p^\top, \beta_{z0}^\top\right)^\top$, respectively, gives

$$\begin{aligned} 0 &= U\left(\widehat{\beta}_{x,\text{full}}, \widehat{\beta}_{z,\text{full}}\right) \\ &= U\left(0_p, \beta_{z0}\right) + \frac{\partial U\left(0_p, \beta_{z0}\right)}{\partial \beta^\top} \begin{pmatrix} \widehat{\beta}_{x,\text{full}} - 0_p \\ \widehat{\beta}_{z,\text{full}} - \beta_{z0} \end{pmatrix} + o_p\left(\frac{1}{\sqrt{n}}\right) \end{aligned} \tag{B.7}$$

and

$$U\left(\frac{\eta}{\sqrt{n}}, \beta_{z0}\right) = U\left(0_p, \beta_{z0}\right) + \frac{\partial U\left(0_p, \beta_{z0}\right)}{\partial \beta^\top} \begin{pmatrix} \frac{\eta}{\sqrt{n}} \\ 0_q \end{pmatrix} + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{B.8}$$

Combining (B.7) and (B.8) gives

$$\begin{aligned} 0 = \; & U\left(\frac{\eta}{\sqrt{n}}, \beta_{z0}\right) + \frac{\partial U\left(0_p, \beta_{z0}\right)}{\partial \beta^\top} \begin{pmatrix} \widehat{\beta}_{x,\text{full}} \\ \widehat{\beta}_{z,\text{full}} - \beta_{z0} \end{pmatrix} - \frac{\partial U\left(0_p, \beta_{z0}\right)}{\partial \beta^\top} \begin{pmatrix} \frac{\eta}{\sqrt{n}} \\ 0_q \end{pmatrix} \\ & + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{B.9}$$

and re-scaling (B.9) yields

$$\begin{aligned} \sqrt{n} \begin{pmatrix} \widehat{\beta}_{x,\text{full}} \\ \widehat{\beta}_{z,\text{full}} - \beta_{z0} \end{pmatrix} &= \left\{ \frac{-1}{n} \frac{\partial U\left(0_p, \beta_{z0}\right)}{\partial \beta^\top} \right\}^{-1} \frac{1}{\sqrt{n}} U\left(\frac{\eta}{\sqrt{n}}, \beta_{z0}\right) \\ &\quad + \begin{pmatrix} \eta \\ 0_q \end{pmatrix} + o_p(1). \end{aligned} \tag{B.10}$$

Let

$$\widehat{\zeta}_i^*\left(\beta_x, \beta_z\right) = \int_0^\tau \exp\left\{-\widehat{\Lambda}_0(u) \exp\left(\widehat{x}_i^\top \beta_x + z_i^\top \beta_z\right)\right\} d\widehat{H}(u). \tag{B.11}$$

Since $\widehat{\mu}_{X^*} = \mu_X + o_p(1)$ and $\widehat{\Sigma}_{X^*} = \Sigma_{X^*} + o_p(1)$ by the Law of Large Numbers, we obtain that $\widehat{X}_i = \widetilde{X}_i + o_p(1)$ by the continuous mapping theorem, where

$$\widetilde{X}_i = E\left(X_i|X_i^*\right) = \mu_X + \left(\Sigma_{X^*} - \Sigma_\epsilon\right)^\top \Sigma_{X^*}^{-1}\left(X_i^* - \mu_{X^*}\right).$$

Since the indicator functions $\{I(A \le t \le Y) : t \in [0,\tau]\}$ and $\{I(Y \le t) : t \in [0,\tau]\}$ are Glivanko-Cantelli classes ([20], Example 2.4.2), by Uniformly Strong Law of Large Numbers, we have that as $n \to \infty$,

$$G^{(k)}(u, \beta_x, \beta_z) \xrightarrow{a.s.} \mathcal{G}^{(k)}(u, \beta_x, \beta_z)$$

uniformly at u, where

$$\mathcal{G}^{(k)}(u, \beta_x, \beta_z) = E\left\{\left(\begin{array}{c} X^* \\ Z \end{array}\right)^{\otimes k} \exp({X^*}^\top \beta_x + Z^\top \beta_z) I(A \le u \le Y)\right\} \quad \text{(B.12)}$$

for $k = 0, 1, 2$. By the similar proof in [3], we have that as $n \to \infty$,

$$\sup_{\beta\in\Theta,\ t\in[0,\tau]} |\widehat{\Lambda}_0(t) - \Lambda_0^*(t)| \xrightarrow{a.s.} 0, \quad \text{(B.13)}$$

where

$$\Lambda_0^*(t) = \int_0^t \frac{dP\left(\Delta = 1, Y \le u\right)}{\left\{m\left(\beta_{x0}\right)\right\}^{-1} \mathcal{G}^{(0)}(u, \beta_{x0}, \beta_{z0})} \quad \text{(B.14)}$$

and $m(\cdot)$ is given by $m_S(\cdot)$ with S specified as the full model.

In addition, by the similar derivations of Lemma 4.2 of [23], we have that as $n \to \infty$,

$$\widehat{H}(u) \xrightarrow{a.s.} H(u) \quad \text{(B.15)}$$

uniformly, where $\widehat{H}(\cdot)$ is determined by (3.13) with S set as the full model. Applying the continuous mapping theorem and combining (B.13) and (B.15) yield that as $n \to \infty$,

$$\widehat{\zeta}_i^*\left(\beta_x, \beta_z\right) \xrightarrow{a.s.} \zeta_i^*\left(\beta_x, \beta_z\right),$$

where

$$\zeta_i^*\left(\beta_x, \beta_z\right) = \int_0^\tau \exp\left\{-\Lambda_0^*\left(u\right) \exp\left(\widetilde{x}_i^\top \beta_x + z_i^\top \beta_z\right)\right\} dH\left(u\right). \quad \text{(B.16)}$$

Noting that by (B.5) and $U_{\mathrm{M}} = \frac{\partial \widehat{\ell}_{\mathrm{M}}^*}{\partial \beta}$ together with (3.14) with S specified as the full model, we obtain that

$$\begin{aligned}
&\frac{-1}{n}\frac{\partial U\left(0_p, \beta_{z0}\right)}{\partial \beta} \\
&= \frac{-1}{n}\left(\frac{\partial U_{\mathrm{P}}\left(0_p, \beta_{z0}\right)}{\partial \beta} + \frac{\partial U_{\mathrm{M}}\left(0_p, \beta_{z0}\right)}{\partial \beta}\right) \\
&= \frac{-1}{n}\frac{\partial}{\partial \beta}\sum_{i=1}^n \int_0^\tau \left\{\left(\begin{array}{c} x_i^* \\ z_i \end{array}\right) + \left(\begin{array}{c} \Sigma_\epsilon 0_p \\ 0_q \end{array}\right) - \frac{G^{(1)}\left(u, 0_p, \beta_{z0}\right)}{G^{(0)}\left(u, 0_p, \beta_{z0}\right)}\right\} dN_i\left(u\right)
\end{aligned}$$

$$+\frac{1}{n}\frac{\partial}{\partial\beta}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\left\{\widehat{\Lambda}_0(a_i)\exp\left(\widehat{x}_i^\top 0_p+z_i^\top\beta_{z0}\right)\right\} \quad \text{(B.17a)}$$

$$+\frac{1}{n}\frac{\partial}{\partial\beta}\sum_{i=1}^{n}\frac{\frac{\partial}{\partial\beta}\left[\int_0^\tau\exp\left\{-\widehat{\Lambda}_0(u)\exp\left(\widehat{x}_i^\top 0_p+z_i^\top\beta_{z0}\right)\right\}d\widehat{H}(u)\right]}{\int_0^\tau\exp\left\{-\widehat{\Lambda}_0(u)\exp\left(\widehat{x}_i^\top 0_p+z_i^\top\beta_{z0}\right)\right\}d\widehat{H}(u)}. \quad \text{(B.17b)}$$

Then exchanging the order of differentiation and summation and plugging (B.11) to (B.17b) with (β_x,β_z) evaluated at $(0_p,\beta_{z0})$, yields

$$\begin{aligned}
&\frac{-1}{n}\frac{\partial U\left(0_p,\beta_{z0}\right)}{\partial\beta}\\
=\ &\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\left[\left\{\frac{G^{(2)}(u,0_p,\beta_{z0})}{G^{(0)}(u,0_p,\beta_{z0})}-\left(\frac{G^{(1)}(u,0_p,\beta_{z0})}{G^{(0)}(u,0_p,\beta_{z0})}\right)^{\otimes 2}\right\}\right.\\
&\left.-\begin{pmatrix}\Sigma_\epsilon & 0_{p\times q}\\ 0_{q\times p} & 0_{q\times q}\end{pmatrix}\right]dN_i(u)\\
&+\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\beta\partial\beta^\top}\left\{\widehat{\Lambda}_0(a_i)\exp\left(\widehat{x}_i^\top 0_p+z_i^\top\beta_{z0}\right)\right\}\\
&+\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{\widehat{\zeta}_i^*}\frac{\partial^2\widehat{\zeta}_i^*}{\partial\beta\partial\beta^\top}-\frac{1}{(\widehat{\zeta}_i^*)^2}\left(\frac{\partial\widehat{\zeta}_i^*}{\partial\beta}\right)^{\otimes 2}\right\},
\end{aligned}$$

where $G^{(k+1)}(u,\beta_x,\beta_z)=\frac{\partial}{\partial\beta}G^{(k)}(u,\beta_x,\beta_z)$ for $k=0,1$, and $\widehat{\zeta}_i^*=\widehat{\zeta}_i^*\left(0_p,\beta_{z0}\right)$.

We conclude that by the Law of Large Numbers, as $n\to\infty$,

$$\frac{-1}{n}\frac{\partial U\left(0_p,\beta_{z0}\right)}{\partial\beta}\xrightarrow{p}\mathcal{A}, \quad \text{(B.18)}$$

where

$$\begin{aligned}
\mathcal{A}\ =\ &\int_0^\tau\left[\left\{\frac{\mathcal{G}^{(2)}(u,0_p,\beta_{z0})}{\mathcal{G}^{(0)}(u,0_p,\beta_{z0})}-\left(\frac{\mathcal{G}^{(1)}(u,0_p,\beta_{z0})}{\mathcal{G}^{(0)}(u,0_p,\beta_{z0})}\right)^{\otimes 2}\right\}\right.\\
&\left.-\begin{pmatrix}\Sigma_\epsilon & 0_{p\times q}\\ 0_{q\times p} & 0_{q\times q}\end{pmatrix}\right]dE\left\{N_i(u)\right\}\\
&+E\left\{\frac{\partial^2}{\partial\beta\partial\beta^\top}\Lambda_0^*(A)\exp\left(\widetilde{X}^\top 0_p+Z^\top\beta_{z0}\right)\right.\\
&\left.+(\zeta^*)^{-2}\left(\zeta^*\frac{\partial(\zeta^*)^2}{\partial^2\beta}-\left(\frac{\partial\zeta^*}{\partial\beta}\right)^{\otimes 2}\right)\right\}
\end{aligned}$$

with $\mathcal{G}^{(k)}(\cdot)$ is given by (B.12) for $k=0,1,2$, and

$$\zeta^*=\int_0^\tau\exp\left\{-\Lambda_0^*(u)\exp\left(\widetilde{x}^\top 0_p+z^\top\beta_{z0}\right)\right\}dH(u).$$

Step 1.2:

Since $U\left(\frac{\eta}{\sqrt{n}},\beta_{z0}\right)$ contains the sample size n, it cannot be directly expressed as a sum of i.i.d. random functions. We now want to re-express it in order to derive a sum of i.i.d. random functions. Since $\exp\left(x^{*\top}\frac{\eta}{\sqrt{n}}\right)=\frac{x^{*\top}\eta}{\sqrt{n}}+O_p\left(1\right)$, then by (B.2) and the form of $G^{(0)}\left(u,\beta_x,\beta_z\right)$ given in Lemma 4.2,

$$G^{(j)}\left(u,\frac{\eta}{\sqrt{n}},\beta_z\right)=\frac{1}{\sqrt{n}}\widetilde{G}^{(j)}\left(u,\eta,\beta_z\right)+O_p\left(1\right)\ \text{ for } j=0,1, \tag{B.19}$$

where

$$\widetilde{G}^{(j)}\left(u,\eta,\beta_z\right)=\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} x_i^* \\ z_i \end{pmatrix}^{\otimes j} Y_i(u)\exp\left(z_i^\top\beta_z\right)x_i^{*\top} \tag{B.20}$$

which is a sum of i.i.d. random variables for $j=0,1$.

Combining (B.19) and $U\left(\frac{\eta}{\sqrt{n}},\beta_{z0}\right)$ gives

$$\frac{1}{\sqrt{n}}U\left(\frac{\eta}{\sqrt{n}},\beta_{z0}\right)=\frac{1}{\sqrt{n}}\widetilde{U}\left(\eta,\beta_{z0}\right)+o_p(1), \tag{B.21}$$

where

$$\widetilde{U}\left(\eta,\beta_{z0}\right)=\widetilde{U}_{\mathrm{P}}\left(\eta,\beta_{z0}\right)+\widetilde{U}_{\mathrm{M}}\left(\eta,\beta_{z0}\right), \tag{B.22}$$

$$\widetilde{U}_{\mathrm{P}}\left(\eta,\beta_{z0}\right)=\sum_{i=1}^{n}\int_0^\tau\left\{\begin{pmatrix} x_i^* \\ z_i \end{pmatrix}+\begin{pmatrix} \Sigma_\epsilon\eta \\ 0_q \end{pmatrix}-\frac{\widetilde{G}^{(1)}\left(u,\eta,\beta_z\right)}{\widetilde{G}^{(0)}\left(u,\eta,\beta_z\right)}\right\}dN_i\left(u\right)$$

and

$$\begin{aligned}
&\widetilde{U}_{\mathrm{M}}\left(\eta,\beta_{z0}\right)\\
=\;& -\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\left\{\int_0^{a_i}\frac{\frac{1}{n}\sum\limits_{j=1}^{n}dN_j(u)}{(\eta^\top\Sigma_\epsilon\eta)^{-1}\widetilde{G}^{(0)}\left(u,\eta,\beta_{z0}\right)}\exp\left(z_i^\top\beta_z\right)\widehat{x}_i^\top\eta\right\}\\
&-\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\log\int_0^\tau\exp\left\{-\widehat{\Lambda}_0(u)\exp\left(z_i^\top\beta_{z0}\right)\widehat{x}_i^\top\eta\right\}d\widehat{H}(u).
\end{aligned} \tag{B.23}$$

(B.21) suggests that to study the asymptotic behavior of $\frac{1}{\sqrt{n}}U\left(\frac{\eta}{\sqrt{n}},\beta_{z0}\right)$, it suffices to study $\frac{1}{\sqrt{n}}\widetilde{U}\left(\eta,\beta_{z0}\right)$ by expressing it as a sum of i.i.d. random functions. To this end, by (B.22), we separately examine $\widetilde{U}_{\mathrm{P}}\left(\eta,\beta_{z0}\right)$ and $\widetilde{U}_{\mathrm{M}}\left(\eta,\beta_{z0}\right)$. First, using the arguments similar to the derivations in Theorem 2 of [3], we derive

$$\frac{1}{\sqrt{n}}\widetilde{U}_{\mathrm{P}}\left(\eta,\beta_{z0}\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_{1i}+o_p(1), \tag{B.24}$$

where

$$\Psi_{1i} = \int_0^\tau \left\{ \begin{pmatrix} x_i^* \\ z_i \end{pmatrix} - \frac{\widetilde{\mathcal{G}}^{(1)}(u,\eta,\beta_{z0})}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})} + \begin{pmatrix} \Sigma_\epsilon \eta \\ 0_q \end{pmatrix} \right\} dN_i(u) - \int_0^\tau \frac{\exp\left(z_i^\top \beta_z\right) x_i^{*\top} \eta I(A_i \le u \le Y_i)}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})} \left\{ \begin{pmatrix} x_i^* \\ z_i \end{pmatrix} - \frac{\widetilde{\mathcal{G}}^{(1)}(u,\eta,\beta_{z0})}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})} \right\} dE\left\{N_i(u)\right\} \tag{B.25}$$

with

$$\widetilde{\mathcal{G}}^{(j)}(u,\eta,\beta_{z0}) = E\left\{ \begin{pmatrix} X^* \\ Z \end{pmatrix}^{\otimes j} Y(u) \exp\left(Z^\top \beta_z\right) X^{*\top} \eta \right\} \text{ for } j = 0, 1.$$

Next, we examine $\widetilde{U}_{\mathrm{M}}(\eta, \beta_{z0})$. Let

$$\zeta(\widetilde{x}, z) = \int_0^\tau \exp\left\{-\Lambda_0^*(u) \exp(z^\top \beta_{z0}) \widetilde{x}^\top \eta\right\} dH(u), \tag{B.26}$$

$$\widetilde{X} = \mu_X + \left(\Sigma_{X^*} - \Sigma_\epsilon\right)^\top \Sigma_{X^*}^{-1} \left(X^* - \mu_{X^*}\right),$$

and

$$\widehat{\zeta}(\widehat{x}, z) = \int_0^\tau \exp\left\{-\widehat{\Lambda}_0(u) \exp(z^\top \beta_{z0}) \widehat{x}^\top \eta\right\} d\widehat{H}(u). \tag{B.27}$$

To derive a sum of i.i.d. random functions and study the asymptotic behavior of $\widetilde{U}_{\mathrm{M}}(\eta, \beta_{z0})$, we further define

$$\widetilde{U}_{\mathrm{M}}^*(\eta, \beta_{z0}) = -\sum_{i=1}^n \frac{\partial}{\partial \beta} \Lambda_0^*(a_i) \exp\left(z_i^\top \beta_{z0}\right) \widetilde{x}_i^\top \eta - \sum_{i=1}^n \frac{1}{\zeta(\widetilde{x}_i, z_i)} \frac{\partial \zeta(\widetilde{x}_i, z_i)}{\partial \beta}. \tag{B.28}$$

Then by (B.23) and (B.28), the difference between $\widetilde{U}_{\mathrm{M}}$ and $\widetilde{U}_{\mathrm{M}}^*$ can be written as

$$\frac{1}{\sqrt{n}} \left\{ \widetilde{U}_{\mathrm{M}} - \widetilde{U}_{\mathrm{M}}^* \right\} = \widetilde{U}_1 + \widetilde{U}_2, \tag{B.29}$$

where

$$\widetilde{U}_1 = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \beta} \left\{ \widehat{\Lambda}_0(a_i) - \Lambda_0^*(a_i) \right\} \exp\left(z_i^\top \beta_{z0}\right) \widetilde{x}_i^\top \eta$$

and

$$\widetilde{U}_2 = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{\widehat{\zeta}\left(\widehat{x}_i, z_i\right)}\frac{\partial\widehat{\zeta}\left(\widehat{x}_i, z_i\right)}{\partial\beta} - \frac{1}{\zeta\left(\widetilde{x}_i, z_i\right)}\frac{\partial\zeta\left(\widetilde{x}_i, z_i\right)}{\partial\beta}\right\}. \tag{B.30}$$

To study the asymptotic behaviour of (B.29), we examine $\widetilde{U}_1$ and $\widetilde{U}_2$ individually. First, let $\mathcal{N}(t) = P\left(\Delta_i = 1, Y_i \le t\right)$ and $d\bar{N}(t) = \frac{1}{n}\sum_{i=1}^{n} dN_i(t)$. Then by (3.6) with S being the full model and (B.14),

$$\begin{aligned}
\widetilde{U}_1 &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\left\{\widehat{\Lambda}_0(a_i) - \Lambda_0^*(a_i)\right\}\exp\left(z_i^\top\beta_{z0}\right)\widetilde{x}_i^\top\eta \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau\left[\frac{\partial}{\partial\beta}\left\{\frac{d\bar{N}(u)}{\widetilde{G}^{(0)}(u,\eta,\beta_{z0})} - \frac{d\mathcal{N}(u)}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})}\right\}\right. \\
&\quad \left.\times\left(\eta^\top\Sigma_\epsilon\eta\right)\exp\left(z_i^\top\beta_{z0}\right)\widetilde{x}_i^\top\eta I(u\le a_i\le\tau)\right] \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau\left[\frac{\partial}{\partial\beta}\left\{\frac{d\bar{N}(u) - d\mathcal{N}(u)}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})}\right.\right. \\
&\quad \left. + \frac{d\mathcal{N}(u)\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0}) - d\bar{N}(u)\widetilde{G}^{(0)}(u,\eta,\beta_{z0})}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})\widetilde{G}^{(0)}(u,\eta,\beta_{z0})}\right\}\left(\eta^\top\Sigma_\epsilon\eta\right) \\
&\quad \left.\times\exp\left(z_i^\top\beta_{z0}\right)\widetilde{x}_i^\top\eta I(u\le a_i\le\tau)\right].
\end{aligned} \tag{B.31}$$

Since $\frac{1}{n}\sum_{i=1}^{n}\exp\left(z_i^\top\beta_{z0}\right)\widetilde{x}_i^\top\eta I(u \le a_i \le \tau)$ is an average of i.i.d. random variables due to Conditions (C3), (C4) and (C5). Then by the Law of Large Numbers, we have that as $n\to\infty$,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\exp\left(z_i^\top\beta_{z0}\right)\widetilde{x}_i^\top\eta I(u\le a_i\le\tau) \\
\xrightarrow{p}\;& E\left\{\exp\left(Z^\top\beta_{z0}\right)\widetilde{X}^\top\eta I(u\le A\le\tau)\right\} \\
=\;& \int_{-\infty}^{\infty}\int_0^\tau\exp\left(z^\top\beta_{z0}\right)\widetilde{x}^\top\eta I(u\le a\le\tau)dQ(a,\widehat{v}),
\end{aligned}$$

i.e.,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\exp\left(z_i^\top\beta_{z0}\right)\widetilde{x}_i^\top\eta I(u\le a_i\le\tau) \\
=& \int_{-\infty}^{\infty}\int_0^\tau\exp\left(z^\top\beta_{z0}\right)\widetilde{x}^\top\eta I(u\le a\le\tau)dQ(a,\widehat{v}) + O_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned} \tag{B.32}$$

(e.g., [13], p.61), where $Q(a,\widehat{v})$ is the joint density function of $(A,\widehat{V})$ with $\widehat{V}=\left(\widetilde{X},Z\right)$.

Then plugging (B.32) into (B.31) gives

$$\begin{aligned}\widetilde{U}_1 &= -\sqrt{n}\int_{-\infty}^{\infty}\int_0^{\tau}\frac{\partial}{\partial\beta}\left\{\frac{d\bar{N}(u)-d\mathcal{N}(u)}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})}\right.\\ &\quad\left.+\frac{d\mathcal{N}(u)\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})-d\bar{N}(u)\widetilde{G}^{(0)}(u,\eta,\beta_{z0})}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})\widetilde{G}^{(0)}(u,\eta,\beta_{z0})}\right\}\left(\eta^{\top}\Sigma_{\epsilon}\eta\right)\\ &\quad\times\exp\left(z^{\top}\beta_{z0}\right)\widetilde{x}^{\top}\eta I(u\le a\le\tau)dQ(a,\widehat{v})+o_p(1),\end{aligned}\tag{B.33}$$

where the order term is determined by $\sqrt{n}\times o_p(1)\times O_p\left(\frac{1}{\sqrt{n}}\right)=o_p(1)$.

Furthermore, noting that by the Uniformly Strong Law of Large Numbers,

$$d\bar{N}(t)\xrightarrow{a.s.}d\mathcal{N}(t)$$

and

$$\widetilde{G}^{(0)}(u,\eta,\beta_{z0})\xrightarrow{a.s.}\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})$$

uniformly as $n\to\infty$. That is,

$$d\bar{N}(t)=d\mathcal{N}(t)+o_p(1)\tag{B.34}$$

and

$$\widetilde{G}^{(0)}(u,\eta,\beta_{z0})=\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})+o_p(1).\tag{B.35}$$

Then we obtain that

$$\begin{aligned}\widetilde{U}_1 &= -\sqrt{n}\int_{-\infty}^{\infty}\int_0^{\tau}\frac{\partial}{\partial\beta}\left\{\frac{d\bar{N}(u)-d\mathcal{N}(u)}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})}\right.\\ &\quad\left.+\frac{d\mathcal{N}(u)\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})-d\bar{N}(u)\widetilde{G}^{(0)}(u,\eta,\beta_{z0})}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})\widetilde{G}^{(0)}(u,\eta,\beta_{z0})}\right\}\left(\eta^{\top}\Sigma_{\epsilon}\eta\right)\\ &\quad\times\exp\left(z^{\top}\beta_{z0}\right)\widetilde{x}^{\top}\eta I(u\le a\le\tau)dQ(a,\widehat{v})+o_p(1)\\ &= -\sqrt{n}\int_{-\infty}^{\infty}\int_0^{\tau}\frac{\partial}{\partial\beta}\left[\frac{d\bar{N}(u)-d\mathcal{N}(u)}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})}\right.\\ &\quad\left.+\frac{d\mathcal{N}(u)}{\left\{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})\right\}^2}\left\{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})-\widetilde{G}^{(0)}(u,\eta,\beta_{z0})\right\}\right]\\ &\quad\times\left(\eta^{\top}\Sigma_{\epsilon}\eta\right)\exp\left(z^{\top}\beta_{z0}\right)\widetilde{x}^{\top}\eta I(u\le a\le\tau)dQ(a,\widehat{v})+o_p(1),\end{aligned}$$

where we apply (B.34) and (B.35) to the numerator and denominator of the second term, respectively. Then by definition of $d\bar{N}(t)$ and (B.20), we obtain that

$$\widetilde{U}_1 = -\sqrt{n}\int_{-\infty}^{\infty}\int_0^{\tau}\frac{\partial}{\partial\beta}\left[\frac{d\bar{N}(u)}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})}-\frac{d\mathcal{N}(u)\widetilde{G}^{(0)}(u,\eta,\beta_{z0})}{\left\{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})\right\}^2}\right]\left(\eta^{\top}\Sigma_{\epsilon}\eta\right)$$

$$
\begin{aligned}
&\times \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta I(u \le a \le \tau) dQ(a, \widehat{v}) + o_p(1) \\
= \;& -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \beta} \Bigg[\int_{-\infty}^{\infty} \int_0^\tau \Bigg\{ \frac{dN_i(u)}{\widetilde{\mathcal{G}}^{(0)}(u, \eta, \beta_{z0})} \\
& - \frac{d\mathcal{N}(u) \exp\left(z_i^\top \beta_{z0}\right) x_i^{*\top} \eta I(a_i \le u \le y_i)}{\left\{\widetilde{\mathcal{G}}^{(0)}(u, \eta, \beta_{z0})\right\}^2} \Bigg\} \left(\eta^\top \Sigma_\epsilon \eta\right) \\
& \times \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta I(u \le a \le \tau) \Bigg] dQ(a, \widehat{v}) + o_p(1). \qquad \text{(B.36)}
\end{aligned}
$$

We next examine $\widetilde{U}_2$. To do so, we first derive the asymptotic result of $\sqrt{n}\left\{\widehat{\zeta}\left(\widehat{x}, z\right) - \zeta\left(\widetilde{x}, z\right)\right\}$. Since $\widehat{X}_i = \widetilde{X}_i + o_p(1)$ due to $\widehat{\mu}_{X^*} = \mu_X + o_p(1)$, $\widehat{\Sigma}_{X^*} = \Sigma_{X^*} + o_p(1)$ and the continuous mapping theorem, then

$$
\begin{aligned}
& -\widehat{\Lambda}_0(u) \exp(z^\top \beta_{z0}) \widehat{x}^\top \eta + \Lambda_0^*(u) \exp(z^\top \beta_{z0}) \widetilde{x}^\top \eta \\
= \;& -\left\{\widehat{\Lambda}_0(u) - \Lambda_0^*(u)\right\} \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta + o_p(1) \\
= \;& \frac{-1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ \frac{dN_i(u)}{\widetilde{\mathcal{G}}^{(0)}(u, \eta, \beta_{z0})} - \frac{d\mathcal{N}(u) \exp\left(z_i^\top \beta_{z0}\right) x_i^{*\top} \eta I(a_i \le u \le y_i)}{\left\{\widetilde{\mathcal{G}}^{(0)}(u, \eta, \beta_{z0})\right\}^2} \right\} \\
& \times \left(\eta^\top \Sigma_\epsilon \eta\right) \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta + o_p(1), \qquad \text{(B.37)}
\end{aligned}
$$

where the last equality is due to the expression of $\widehat{\Lambda}_0(u) - \Lambda_0^*(u)$ in (B.31).

Applying the Taylor series expansion to $\exp\left\{-\widehat{\Lambda}_0(u) \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta\right\}$ with respect to $\Lambda_0(\cdot)$ yields

$$
\begin{aligned}
& \exp\left\{-\widehat{\Lambda}_0(u) \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta\right\} - \exp\left\{-\Lambda_0^*(u) \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta\right\} \\
= & - \exp\left\{-\Lambda_0^*(u) \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta\right\} \left\{\widehat{\Lambda}_0(u) - \Lambda_0^*(u)\right\} \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta \\
& + o_p\left(\frac{1}{\sqrt{n}}\right) \\
= & \exp\left\{-\Lambda_0^*(u) \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta\right\} \\
& \times \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left[\left\{ \frac{dN_i(u)}{\widetilde{\mathcal{G}}^{(0)}(u, \eta, \beta_{z0})} - \frac{d\mathcal{N}(u) \exp\left(z_i^\top \beta_{z0}\right) x_i^{*\top} \eta I(a_i \le u \le y_i)}{\left\{\widetilde{\mathcal{G}}^{(0)}(u, \eta, \beta_{z0})\right\}^2} \right\} \right. \\
& \left. \times \left(\eta^\top \Sigma_\epsilon \eta\right) \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta \right] + o_p\left(\frac{1}{\sqrt{n}}\right) \\
= & S(u, \eta, \beta_{z0} | \widetilde{x}, z) \times \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left[\left\{ \frac{dN_i(u)}{\widetilde{\mathcal{G}}^{(0)}(u, \eta, \beta_{z0})} \right. \right.
\end{aligned}
$$

$$- \frac{d\mathcal{N}(u) \exp\left(z_i^\top \beta_{z0}\right) x_i^{*\top} \eta I(a_i \le u \le y_i)}{\left\{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})\right\}^2} \Bigg\} \left(\eta^\top \Sigma_\epsilon \eta\right) \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta \Bigg] + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{B.38}$$

where the second equality is due to (B.37), and

$$S(u,\eta,\beta_{z0}|\widetilde{x},z) = \exp\left\{-\Lambda_0^*(u) \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta\right\}.$$

Finally, using (B.26) and (B.27) in combination with (B.15) and (B.38), we obtain that

$$\sqrt{n}\left\{\widehat{\zeta}(\widehat{x},z) - \zeta(\widetilde{x},z)\right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\eta,\beta_{z0}|\widetilde{x},z) + o_p(1), \tag{B.39}$$

where

$$\begin{aligned} \psi_i(\eta,\beta_{z0}|\widetilde{x},z) &= \int_0^\tau \int_0^\tau \Bigg[h(\xi) S(\xi,\eta,\beta_{z0}|\widetilde{x},z) \Bigg\{ \frac{dN_i(u)}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})} \\ &\quad - \frac{d\mathcal{N}(u) \exp\left(z_i^\top \beta_{z0}\right) x_i^{*\top} \eta I(a_i \le u \le y_i)}{\left\{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})\right\}^2} \Bigg\} \\ &\quad \times \left(\eta^\top \Sigma_\epsilon \eta\right) \exp\left(z^\top \beta_{z0}\right) \widetilde{x}^\top \eta \Bigg] d\xi. \end{aligned}$$

Then by (B.30) and similar to the derivations of (B.33), we obtain that

$$\begin{aligned} \widetilde{U}_2 &= \frac{-1}{\sqrt{n}} \sum_{j=1}^n \left\{ \frac{1}{\widehat{\zeta}(\widehat{x}_j,z_j)} \frac{\partial \widehat{\zeta}(\widehat{x}_j,z_j)}{\partial \beta} - \frac{1}{\zeta(\widetilde{x}_j,z_j)} \frac{\partial \zeta(\widetilde{x}_j,z_j)}{\partial \beta} \right\} \\ &= -\sqrt{n} \times \frac{1}{n} \sum_{j=1}^n \left\{ \frac{1}{\widehat{\zeta}(\widehat{x}_j,z_j)} \frac{\partial \widehat{\zeta}(\widehat{x}_j,z_j)}{\partial \beta} - \frac{1}{\zeta(\widetilde{x}_j,z_j)} \frac{\partial \zeta(\widetilde{x}_j,z_j)}{\partial \beta} \right\} \\ &= -\sqrt{n} \int_{-\infty}^\infty \int_0^\tau \left\{ \frac{1}{\widehat{\zeta}(\widehat{x},z)} \frac{\partial \widehat{\zeta}(\widehat{x},z)}{\partial \beta} - \frac{1}{\zeta(\widetilde{x},z)} \frac{\partial \zeta(\widetilde{x},z)}{\partial \beta} \right\} dQ(a,\widehat{v}) \\ &\quad + o_p(1). \end{aligned} \tag{B.40}$$

To sort out a sum of i.i.d. random functions from (B.40), we add and subtract the term $\frac{1}{\zeta(\widetilde{x},z)} \frac{\partial \widehat{\zeta}(\widehat{x},z)}{\partial \beta}$ and then regroup the differences, yielding

$$\begin{aligned} \widetilde{U}_2 &= -\sqrt{n} \int_{-\infty}^\infty \int_0^\tau \left\{ \frac{1}{\widehat{\zeta}(\widehat{x},z)} \frac{\partial \widehat{\zeta}(\widehat{x},z)}{\partial \beta} - \frac{1}{\zeta(\widetilde{x},z)} \frac{\partial \widehat{\zeta}(\widehat{x},z)}{\partial \beta} + \frac{1}{\zeta(\widetilde{x},z)} \frac{\partial \widehat{\zeta}(\widehat{x},z)}{\partial \beta} \right. \\ &\quad \left. - \frac{1}{\zeta(\widetilde{x},z)} \frac{\partial \zeta(\widetilde{x},z)}{\partial \beta} \right\} dQ(a,\widehat{v}) + o_p(1) \end{aligned}$$

$$
\begin{aligned}
&= -\sqrt{n}\int_{-\infty}^{\infty}\int_0^{\tau}\left[\frac{1}{\zeta(\widetilde{x},z)}\left\{\frac{\partial\widehat{\zeta}(\widehat{x},z)}{\partial\beta}-\frac{\partial\zeta(\widetilde{x},z)}{\partial\beta}\right\}\right.\\
&\quad\left.-\frac{\partial\zeta(\widetilde{x},z)}{\partial\beta}\left\{\frac{\widehat{\zeta}(\widehat{x},z)-\zeta(\widetilde{x},z)}{\zeta^2(\widetilde{x},z)}\right\}\right]dQ(a,\widehat{v})+o_p(1)\\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\int_{-\infty}^{\infty}\int_0^{\tau}\left\{\frac{1}{\zeta(\widetilde{x},z)}\frac{\partial}{\partial\beta}\psi_i(\eta,\beta_{z0}|\widetilde{x},z)\right.\right.\\
&\quad\left.\left.-\frac{\partial\zeta(\widetilde{x},z)}{\partial\beta}\frac{1}{\zeta^2(\widetilde{x},z)}\psi_i(\eta,\beta_{z0}|\widetilde{x},z)\right\}\right]dQ(a,\widehat{v})+o_p(1), \qquad \text{(B.41)}
\end{aligned}
$$

where the second equality is due to $\widehat{\zeta}(\widehat{x},z)=\zeta(\widetilde{x},z)+o_p(1)$, and the last step is due to (B.39).

Combining (B.29), (B.36) and (B.41) gives

$$
\frac{1}{\sqrt{n}}\widetilde{U}_{\mathrm{M}}(\eta,\beta_{z0})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_{2i}+o_p(1), \qquad \text{(B.42)}
$$

where

$$
\begin{aligned}
\Psi_{2i} &= -\left[\int_{-\infty}^{\infty}\int_0^{\tau}\frac{\partial}{\partial\beta}\left\{\frac{dN_i(u)}{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})}\right.\right.\\
&\quad\left.-\frac{d\mathcal{N}(u)\exp\left(z_i^{\top}\beta_z\right)x_i^{*\top}\eta I(a_i\le u\le y_i)}{\left\{\widetilde{\mathcal{G}}^{(0)}(u,\eta,\beta_{z0})\right\}^2}\right\}(\eta^{\top}\Sigma_{\epsilon}\eta)\\
&\quad\left.\times\exp\left(z^{\top}\beta_z\right)\widetilde{x}^{\top}\eta I(u\le a\le\tau)\right]dQ(a,\widehat{v})\\
&\quad-\left[\int_{-\infty}^{\infty}\int_0^{\tau}\left\{\frac{1}{\zeta(\widetilde{x},z)}\frac{\partial}{\partial\beta}\psi_i(\eta,\beta_{z0}|\widetilde{x},z)\right.\right.\\
&\quad\left.\left.-\frac{\partial\zeta(\widetilde{x},z)}{\partial\beta}\frac{1}{\zeta^2(\widetilde{x},z)}\psi_i(\eta,\beta_{z0}|\widetilde{x},z)\right\}dQ(a,\widehat{v})\right]\\
&\quad-\frac{\partial}{\partial\beta}\Lambda_0^*(a_i)\exp\left(z_i^{\top}\beta_z\right)\widetilde{x}_i^{\top}\eta-\frac{1}{\zeta(\widetilde{x}_i,z_i)}\frac{\partial}{\partial\beta}\zeta(\widetilde{x}_i,z_i). \qquad \text{(B.43)}
\end{aligned}
$$

Therefore, using (B.21), (B.22), (B.24) and (B.42) and applying the Central Limit Theorem, we obtain that as $n\to\infty$,

$$
\frac{1}{\sqrt{n}}U\left(\frac{\eta}{\sqrt{n}},\beta_{z0}\right)\xrightarrow{d}N(0,\mathcal{B}), \qquad \text{(B.44)}
$$

where $\mathcal{B}=E(\Psi_i^{\otimes 2})$ with $\Psi_i=\Psi_{1i}+\Psi_{2i}$, and Ψ_{1i} and Ψ_{2i} are given by (B.25) and (B.43), respectively.

Finally, applying Slutsky's theorem in combination with (B.10), (B.18) and (B.44), we obtain (B.6) as $n \to \infty$.

Step 2: Show the result of Theorem 4.1 (a).

We first re-scale (B.8), which gives

$$\frac{1}{\sqrt{n}} U\left(0_p, \beta_{z0}\right) = \frac{1}{\sqrt{n}} U\left(\frac{\eta}{\sqrt{n}}, \beta_{z0}\right) - \frac{1}{n} \frac{\partial U\left(0_p, \beta_{z0}\right)}{\partial \beta^\top} \begin{pmatrix} \eta \\ 0_q \end{pmatrix} + o_p(1). \quad \text{(B.45)}$$

Combining (B.18), (B.44) and (B.45) and applying Slutsky's theorem, we obtain that as $n \to \infty$,

$$\frac{1}{\sqrt{n}} U\left(0_p, \beta_{z0}\right) \xrightarrow{d} N\left(\mathcal{A}\begin{pmatrix} \eta \\ 0_q \end{pmatrix}, \mathcal{B}\right). \quad \text{(B.46)}$$

Now we consider any candidate model S. Applying the Taylor series expansion to $U_S\left(\widehat{\beta}_{x,S}, \widehat{\beta}_{z,S}\right)$ around $\left(0_{|S|}, \beta_{z0}\right)$ gives

$$0 = U_S\left(\widehat{\beta}_{x,S}, \widehat{\beta}_{z,S}\right) = U_S\left(0_{|S|}, \beta_{z0}\right) + \frac{\partial U_S\left(0_{|S|}, \beta_{z0}\right)}{\partial \beta_S^\top} \begin{pmatrix} \widehat{\beta}_{x,S} \\ \widehat{\beta}_{z,S} - \beta_{z0} \end{pmatrix} + o_p\left(\frac{1}{\sqrt{n}}\right),$$

yielding that

$$\begin{aligned} \sqrt{n}\begin{pmatrix} \widehat{\beta}_{x,S} \\ \widehat{\beta}_{z,S} - \beta_{z0} \end{pmatrix} &= -\left(\frac{1}{n} \frac{\partial U_S\left(0_{|S|}, \beta_{z0}\right)}{\partial \beta_S^\top}\right)^{-1} \frac{1}{\sqrt{n}} U_S\left(0_{|S|}, \beta_{z0}\right) + o_p(1) \\ &= -\left(\frac{1}{n} \frac{\partial U_S\left(0_{|S|}, \beta_{z0}\right)}{\partial \beta_S^\top}\right)^{-1} \frac{1}{\sqrt{n}} \Pi_S U\left(0_{|S|}, \beta_{z0}\right) + o_p(1) \\ &\xrightarrow{d} \mathcal{A}_S^{-1} \Pi_S N\left(\mathcal{A}\begin{pmatrix} \eta \\ 0_q \end{pmatrix}, \mathcal{B}\right) \text{ as } n \to \infty, \end{aligned}$$

where the second identity is from Lemma 4.2 and the third step is due to (B.46). Thus,

$$\sqrt{n}\begin{pmatrix} \widehat{\beta}_{x,S} \\ \widehat{\beta}_{z,S} - \beta_{z0} \end{pmatrix} \xrightarrow{d} N\left(\mathcal{A}_S^{-1} \Pi_S \mathcal{A}\begin{pmatrix} \eta \\ 0_q \end{pmatrix}, \mathcal{A}_S^{-1} \mathcal{B}_S \mathcal{A}_S^{-1}\right) \text{ as } n \to \infty,$$

where $\mathcal{B}_S = \Pi_S \mathcal{B} \Pi_S^\top$. □

B.3.2. Proof of (b)

The proof consists of the following three steps.

Step 1:

For a given candidate model S, by (3.6), we have

$$\begin{aligned}
&\sqrt{n}\left\{\widehat{\Lambda}_{0,S}(t)-\Lambda_0(t)\right\}\\
=\ &\sqrt{n}\left\{\int_0^t \frac{\frac{1}{n}\sum_{i=1}^{n} dN_i(u)}{\left\{m_S(\widehat{\beta}_{x,S})\right\}^{-1} G_S^{(0)}\left(u,\widehat{\beta}_{x,S},\widehat{\beta}_{z,S}\right)}-\Lambda_0(t)\right\}\\
=\ &A+B,
\end{aligned}$$

where

$$\begin{aligned}
A\ =\ &\sqrt{n}\left\{\int_0^t \frac{\frac{1}{n}\sum_{i=1}^{n} dN_i(u)}{\left\{m_S(\widehat{\beta}_{x,S})\right\}^{-1} G_S^{(0)}\left(u,\widehat{\beta}_{x,S},\widehat{\beta}_{z,S}\right)}\right.\\
&\left.-\int_0^t \frac{\frac{1}{n}\sum_{i=1}^{n} dN_i(u)}{\left\{m(\frac{\eta}{\sqrt{n}})\right\}^{-1} G^{(0)}\left(u,\frac{\eta}{\sqrt{n}},\beta_{z0}\right)}\right\} \qquad \text{(B.47)}
\end{aligned}$$

and

$$B=\sqrt{n}\left\{\int_0^t \frac{\frac{1}{n}\sum_{i=1}^{n} dN_i(u)}{\left\{m(\frac{\eta}{\sqrt{n}})\right\}^{-1} G^{(0)}\left(u,\frac{\eta}{\sqrt{n}},\beta_{z0}\right)}-\Lambda_0(t)\right\}.$$

Step 2:

We first examine A. Applying the Taylor series expansion to $\frac{m_S(\widehat{\beta}_{x,S})}{G_S^{(0)}(u,\widehat{\beta}_{x,S},\widehat{\beta}_{z,S})}$ and $\frac{m(\frac{\eta}{\sqrt{n}})}{G^{(0)}\left(u,\frac{\eta}{\sqrt{n}},\beta_{z0}\right)}$, respectively, around $(0,\beta_{z0})$, we have

$$\begin{aligned}
\frac{m_S(\widehat{\beta}_{x,S})}{G_S^{(0)}\left(u,\widehat{\beta}_{x,S},\widehat{\beta}_{z,S}\right)}\ =\ &\frac{1}{G_S^{(0)}\left(u,0_{|S|},\beta_{z0}\right)}-\left[\frac{1}{\left\{G_S^{(0)}\left(u,0_{|S|},\beta_{z0}\right)\right\}^2}\right.\\
&\left.\times\begin{pmatrix} G_{x,S}^{(1)}\left(u,0_{|S|},\beta_{z0}\right)\\ G_{z,S}^{(1)}\left(u,0_{|S|},\beta_{z0}\right)\end{pmatrix}^{\top}\begin{pmatrix}\widehat{\beta}_{x,S}\\ \widehat{\beta}_{z,S}-\beta_{z0}\end{pmatrix}\right]\\
&+o_p\left(\frac{1}{\sqrt{n}}\right) \qquad \text{(B.48)}
\end{aligned}$$

and

$$\frac{m(\frac{\eta}{\sqrt{n}})}{G^{(0)}\left(u,\frac{\eta}{\sqrt{n}},\beta_{z0}\right)} = \frac{1}{G^{(0)}\left(u,0_p,\beta_{z0}\right)} - \frac{G_x^{(1)}\left(u,0_p,\beta_{z0}\right)}{\left\{G^{(0)}\left(u,0_p,\beta_{z0}\right)\right\}^2}\frac{\eta}{\sqrt{n}} + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{B.49}$$

where

$$G_{x,S}^{(1)}(u,\beta_{x,S},\beta_{z,S}) = \frac{\partial}{\partial \beta_{x,S}} G_S^{(0)}(u,\beta_{x,S},\beta_{z,S})$$

and

$$G_{z,S}^{(1)}(u,\beta_{x,S},\beta_{z,S}) = \frac{\partial}{\partial \beta_{z,S}} G_S^{(0)}(u,\beta_{x,S},\beta_{z,S}).$$

Since (B.4) with $j=0$ gives $G_S^{(0)}\left(u,0_{|S|},\beta_{z0}\right) = G^{(0)}\left(u,0_p,\beta_{z0}\right)$, so we combine (B.48) and (B.49) and obtain that

$$\begin{aligned}
&\frac{m_S(\widehat{\beta}_{x,S})}{G_S^{(0)}\left(u,\widehat{\beta}_{x,S},\widehat{\beta}_{z,S}\right)} - \frac{m(\frac{\eta}{\sqrt{n}})}{G^{(0)}\left(u,\frac{\eta}{\sqrt{n}},\beta_{z0}\right)} \\
&= \frac{-1}{\left\{G_S^{(0)}\left(u,0_{|S|},\beta_{z0}\right)\right\}^2}\begin{pmatrix} G_{x,S}^{(1)}\left(u,0_{|S|},\beta_{z0}\right) \\ G_{z,S}^{(1)}\left(u,0_{|S|},\beta_{z0}\right)\end{pmatrix}^\top \begin{pmatrix}\widehat{\beta}_{x,S} \\ \widehat{\beta}_{z,S}-\beta_{z0}\end{pmatrix} \\
&\quad + \frac{G_x^{(1)}\left(u,0_p,\beta_{z0}\right)}{\left\{G^{(0)}\left(u,0_p,\beta_{z0}\right)\right\}^2}\frac{\eta}{\sqrt{n}} + o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Hence, applying (B.47) gives that

$$\begin{aligned}
A &= -\int_0^t \frac{\frac{1}{n}\sum_{i=1}^n dN_i(u)}{\left\{G_S^{(0)}\left(u,0_{|S|},\beta_{z0}\right)\right\}^2}\begin{pmatrix} G_{x,S}^{(1)}\left(u,0_{|S|},\beta_{z0}\right) \\ G_{z,S}^{(1)}\left(u,0_{|S|},\beta_{z0}\right)\end{pmatrix}^\top \sqrt{n}\begin{pmatrix}\widehat{\beta}_{x,S} \\ \widehat{\beta}_{z,S}-\beta_{z0}\end{pmatrix} \\
&\quad + \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n dN_i(u) G_x^{(1)}\left(u,0_p,\beta_{z0}\right)}{\left\{G^{(0)}\left(u,0_p,\beta_{z0}\right)\right\}^2}\eta + o_p(1).
\end{aligned} \tag{B.50}$$

Now we examine the terms in (B.50) separately. Since $\{Y_i(t): t\in[0,\tau]\}$ and $\{N_i(t): t\in[0,\tau]\}$ are Glivenko-Cantelli class ([20], Theorems 2.4.1 and 2.7.5), then we have as that $n\to\infty$,

$$\frac{1}{n}\sum_{i=1}^n dN_i(t) \xrightarrow{a.s.} E\left\{dN_i(t)\right\},$$

and

$$G_S^{(k)}\left(t,0_{|S|},\beta_{z0}\right) \xrightarrow{a.s.} \mathcal{G}_S^{(k)}\left(t,0_{|S|},\beta_{z0}\right) \text{ for } k=0,1$$

uniformly at t, where $G_S^{(0)}(\cdot)$ and $G_S^{(1)}(\cdot)$ are given by (3.7) and (B.1), respectively, and

$$\mathcal{G}_S^{(k)}(u, \beta_{x,S}, \beta_{z,S}) = E\left\{ \begin{pmatrix} \pi_S X^* \\ Z \end{pmatrix}^{\otimes k} \exp((\pi_S X^*)^\top \beta_{x,S} + Z^\top \beta_{x,S}) I(A \le u \le Y) \right\}.$$

Therefore, by the continuous mapping theorem, as $n \to \infty$,

$$\int_0^t \frac{\frac{1}{n}\sum_{i=1}^n dN_i(u)}{\left\{G_S^{(0)}\left(u, 0_{|S|}, \beta_{z0}\right)\right\}^2} \begin{pmatrix} G_{x,S}^{(1)}\left(u, 0_{|S|}, \beta_{z0}\right) \\ G_{z,S}^{(1)}\left(u, 0_{|S|}, \beta_{z0}\right) \end{pmatrix}^\top \xrightarrow{a.s.} \begin{pmatrix} F_{x,S}(t) \\ F_z(t) \end{pmatrix}^\top \tag{B.51}$$

uniformly at t, where

$$F_{x,S}(t) = \int_0^t \frac{E\{dN_i(u)\}\mathcal{G}_{x,S}^{(1)}\left(u, 0_{|S|}, \beta_{z0}\right)}{\left\{\mathcal{G}_S^{(0)}\left(u, 0_{|S|}, \beta_{z0}\right)\right\}^2}$$

and

$$F_z(t) = \int_0^t \frac{E\{dN_i(u)\}\mathcal{G}_z^{(1)}\left(u, 0_{|S|}, \beta_{z0}\right)}{\left\{\mathcal{G}^{(0)}\left(u, 0_{|S|}, \beta_{z0}\right)\right\}^2}.$$

Regarding the term $\sqrt{n}\begin{pmatrix} \widehat{\beta}_{x,S} \\ \widehat{\beta}_{z,S} - \beta_{z0} \end{pmatrix}$ in (B.50), we apply Theorem 4.1 (a) and let W_S be a random vector whose distribution is the same as the limiting distribution of $\sqrt{n}\begin{pmatrix} \widehat{\beta}_{x,S} \\ \widehat{\beta}_{z,S} - \beta_{z0} \end{pmatrix}$, i.e.,

$$\sqrt{n}\begin{pmatrix} \widehat{\beta}_{x,S} \\ \widehat{\beta}_{z,S} - \beta_{z0} \end{pmatrix} \xrightarrow{d} W_S \text{ as } n \to \infty. \tag{B.52}$$

Then applying Slutsky's theorem to (B.50) in combination with (B.51) and (B.52), we have that as $n \to \infty$,

$$A \xrightarrow{d} -\begin{pmatrix} F_{x,S}(t) \\ F_z(t) \end{pmatrix}^\top W_S + F_x(t)^\top \eta. \tag{B.53}$$

<u>Step 3</u>:

Finally, we examine the asymptotic behavior of B. Noting that $\exp\left(\frac{\eta^\top \Sigma_\epsilon \eta}{n}\right) = \frac{\eta^\top \Sigma_\epsilon \eta}{n} + O(1)$ and $\frac{1}{n} = o\left(\frac{1}{\sqrt{n}}\right)$, by the arguments similar to Appendix A.4 of [17], we have

$$B = \sqrt{n}\int_0^t \left\{ \frac{m(\frac{\eta}{\sqrt{n}})\frac{1}{n}\sum_{i=1}^n dN_i(u)}{G^{(0)}\left(u, \frac{\eta}{\sqrt{n}}, \beta_{z0}\right)} - d\Lambda_0(u) \right\}$$

$$\begin{aligned}
&= \sqrt{n}\int_0^t \left\{ \frac{\frac{\eta^\top \Sigma_\epsilon \eta}{\sqrt{n}} \frac{1}{n}\sum\limits_{i=1}^n dN_i(u)}{\frac{1}{\sqrt{n}}\widetilde{G}^{(0)}\left(u,\eta,\beta_{z0}\right)} - d\Lambda_0(u) \right\} + o_p(1) \\
&= \sqrt{n}\int_0^t \left\{ \frac{\eta^\top \Sigma_\epsilon \eta \frac{1}{n}\sum\limits_{i=1}^n dN_i(u)}{\widetilde{G}^{(0)}\left(u,\eta,\beta_{z0}\right)} - d\Lambda_0(u) \right\} + o_p(1) \\
&= \frac{1}{\sqrt{n}}\int_0^t \left\{ \frac{\sum\limits_{i=1}^n \eta^\top \Sigma_\epsilon \eta dN_i(u) - \sum\limits_{i=1}^n Y_i(u)\exp\left(z_i^\top \beta_{z0}\right)(x_i^{*\top}\eta) d\Lambda_0(t)}{\widetilde{G}^{(0)}\left(u,\eta,\beta_{z0}\right)} \right\} \\
&\quad + o_p(1) \\
&= \frac{1}{\sqrt{n}}\int_0^t \left[\frac{\sum\limits_{i=1}^n \left\{\eta^\top \Sigma_\epsilon \eta dN_i(u) - Y_i(u)\exp\left(z_i^\top \beta_{z0}\right)(x_i^{*\top}\eta) d\Lambda_0(u)\right\}}{\widetilde{\mathcal{G}}^{(0)}\left(u,\eta,\beta_{z0}\right)} \right] \\
&\quad + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \left\{ \frac{\eta^\top \Sigma_\epsilon \eta dN_i(u) - Y_i(u)\exp\left(z_i^\top \beta_{z0}\right)(x_i^{*\top}\eta) d\Lambda_0(u)}{\widetilde{\mathcal{G}}^{(0)}\left(u,\eta,\beta_{z0}\right)} \right\} \\
&\quad + o_p(1) \\
&\triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n \Phi_i(t) + o_p(1),
\end{aligned}$$

where the second equality is due to (B.19) and $m(\frac{\eta}{\sqrt{n}}) = \frac{\eta^\top \Sigma_\epsilon \eta}{\sqrt{n}} + O(1)$, the fourth equality is due to (B.20), and the fifth equality is due to (B.35).

By (B.34) and (B.35), $E\left(\frac{1}{n}\sum\limits_{i=1}^n \Phi_i(t)\right) = 0$. Then by Condition (C3), the $\Phi_i(t)$ are i.i.d. with mean zero, and hence, by the Central Limit Theorem, we conclude that

$$B \xrightarrow{d} \mathcal{V}(t) \text{ as } n \to \infty, \tag{B.54}$$

where $\mathcal{V}(t)$ is the Gaussian process with mean zero and covariance function $E\left\{\Phi_i(t)\Phi_i(s)\right\}$.

Finally, combining (B.53) and (B.54) gives that as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\Lambda}_{0,S}(t) - \Lambda_0(t)\right) \xrightarrow{d} \mathcal{V}(t) - \begin{pmatrix} F_{x,S}(t) \\ F_z(t) \end{pmatrix}^\top W_S + F_x(t)^\top \eta,$$

which completes the proof. □

B.4. Proof of Theorem 4.2

For ease of exposition, we simply write $\frac{\partial\mu(0_{|S|},\beta_{z0})}{\partial\beta_{x,S}}$ and $\frac{\partial\mu(0_{|S|},\beta_{z0})}{\partial\beta_{z,S}}$ as $\frac{\partial\mu}{\partial\beta_{x,S}}$ and $\frac{\partial\mu}{\partial\beta_{z,S}}$, respectively.

The proof consists of the following two steps.

Step 1:

Let $\begin{pmatrix} J_x \\ J_z \end{pmatrix}$ be a random vector whose distribution is $N(0_d, \mathcal{B})$, where J_x is a $p\times 1$ random vector and J_z is a $q \times 1$ random vector. Define $\begin{pmatrix} J_{x,S} \\ J_z \end{pmatrix} = \Pi_S \begin{pmatrix} J_x \\ J_z \end{pmatrix}$. Then $\begin{pmatrix} J_{x,S} \\ J_z \end{pmatrix}$ is a random vector whose distribution is $N(0, \mathcal{B}_S)$. Let

$$W_S = \begin{pmatrix} C_S \\ D_S \end{pmatrix} = \mathcal{A}_S^{-1}\left\{\Pi_S\mathcal{A}\begin{pmatrix} \eta \\ 0 \end{pmatrix} + \begin{pmatrix} J_{x,S} \\ J_z \end{pmatrix}\right\} \tag{B.55}$$

be a random vector whose distribution is the asymptotic distribution of

$$\sqrt{n}\begin{pmatrix} \widehat{\beta}_{x,S} \\ \widehat{\beta}_{z,S} - \beta_{z0} \end{pmatrix},$$

where C_S and D_S, respectively, have the distribution identical to the asymptotic distributions of $\sqrt{n}\widehat{\beta}_{x,S}$ and $\sqrt{n}(\widehat{\beta}_{z,S} - \beta_{z0})$.

Furthermore, we express $\mathcal{A}$ as

$$\mathcal{A} = \begin{pmatrix} A_{xx} & A_{xz} \\ A_{zx} & A_{zz} \end{pmatrix}$$

by making the block matrices A_{xx}, A_{xz}, A_{zx} and A_{zz} be of dimensions $p\times p$, $p\times q$, $q\times p$ and $q\times q$, respectively. Similarly, the inverse matrix of $\mathcal{A}$, $\mathcal{A}_S$ and $\mathcal{A}_S^{-1}$ are expressed as $\mathcal{A}^{-1} = \begin{pmatrix} A^{xx} & A^{xz} \\ A^{zx} & A^{zz} \end{pmatrix}$, $\begin{pmatrix} A_{xxS} & A_{xzS} \\ A_{zxS} & A_{zzS} \end{pmatrix}$ and $\begin{pmatrix} A^{xxS} & A^{xzS} \\ A^{zxS} & A^{zzS} \end{pmatrix}$, respectively.

Consequently, by (B.55), we write

$$C_S = (A^{xxS}\pi_S A_{xx} + A^{xzS}A_{zx})\,\eta + A^{xxS}J_{x,S} + A^{xzS}J_z \tag{B.56}$$

and

$$D_S = (A^{zxS}\pi_S A_{xx} + A^{zzS}A_{zx})\,\eta + A^{zxS}\pi_S J_x + A^{zzS}J_z.$$

To continue the proof, we need the following lemma.

Lemma B.1. *Under regularity conditions in Appendix A, we have*

$$\mathcal{A}_S = \Pi_S\mathcal{A}\Pi_S^\top,$$

where $\mathcal{A}$ and $\mathcal{A}_S$ are the asymptotic covariances matrices in Theorem 4.1.

By Lemma B.1, we have

$$\begin{aligned} A^{xxS} &= \left(A_{xxS} - A_{xzS}A_{zzS}^{-1}A_{zxS}\right)^{-1} \\ &= \left(\pi_S A_{xx}\pi_S^\top - \pi_S A_{xz}A_{zz}^{-1}A_{zx}\pi_S^\top\right)^{-1} \\ &= \left\{\pi_S\left(A_{xx} - A_{xz}A_{zz}^{-1}A_{zx}\right)\pi_S^\top\right\}^{-1} \\ &= \left\{\pi_S(A^{xx})^{-1}\pi_S^\top\right\}^{-1} \end{aligned} \tag{B.57}$$

and

$$\begin{aligned} A^{xzS} &= -A^{xxS}A_{xzS}A_{zzS}^{-1} \\ &= -A^{xxS}A_{xzS}A_{zzS}^{-1}. \end{aligned} \tag{B.58}$$

Then combining (B.56), (B.57), and (B.58) gives

$$\begin{aligned} C_S &= \left(A^{xxS}\pi_S A_{xx} - A^{xxS}A_{xzS}A_{zzS}^{-1}A_{zx}\right)\eta + A^{xxS}J_{x,S} - A^{xxS}A_{xzS}A_{zzS}^{-1}J_z \\ &= A^{xxS}\pi_S\left(A_{xx} - A_{xz}A_{zzS}^{-1}A_{zx}\right)\eta + A^{xxS}J_{x,S} - A^{xxS}A_{xzS}A_{zzS}^{-1}J_z \\ &= A^{xxS}\pi_S(A^{xx})^{-1}\eta + A^{xxS}\pi_S J_x - A^{xxS}\pi_S A_{xz}A_{zz}^{-1}J_z \\ &= A^{xxS}\pi_S(A^{xx})^{-1}\left(\eta + A^{xx}J_x - A^{xx}A_{xz}A_{zz}^{-1}J_z\right) \\ &\triangleq A^{xxS}\pi_S(A^{xx})^{-1}\left(\eta + \mathcal{W}\right), \end{aligned} \tag{B.59}$$

where the third equality is due to $J_{x,S} = \pi_S J_x$ and $A^{xx} = \left(A_{xx} - A_{xz}A_{zz}^{-1}A_{zx}\right)^{-1}$,

$$\mathcal{W} = A^{xx}J_x - A^{xx}A_{xz}A_{zz}^{-1}J_z, \tag{B.60}$$

and $(A^{xx})^{-1}$ stands for the inverse of matrix A^{xx}.

Similarly,

$$\begin{aligned} A^{zzS} &= \left\{A_{zz} - A_{zx}\pi_S^\top\left(\pi_S A_{xx}\pi_S^\top\right)^{-1}\pi_S A_{xz}\right\}^{-1} \\ &= \left(A_{zz} - A_{zx}A_{xx}^{-1}A_{xz}\right)^{-1} \\ &= A^{zz}, \end{aligned} \tag{B.61}$$

and

$$\begin{aligned} A^{zxS} &= -A^{zzS}A_{zxS}A_{xxS}^{-1} \\ &= -A^{zz}A_{zx}\pi_S^\top A_{xxS}^{-1}. \end{aligned} \tag{B.62}$$

Thus, using (B.61) and (B.62), and direct calculations give

$$\begin{aligned} D_S &= A_{zz}^{-1}A_{zx}\left\{I_{p\times p} - \pi_S^\top A^{xxS}\pi_S\left(A^{xx}\right)^{-1}\right\}\eta + A_{zz}^{-1}J_z \\ &\quad -A_{zz}^{-1}A_{zx}\pi_S^\top\left(\pi_S A_{xx}\pi_S^\top\right)^{-1}\pi_S J_x + A_{zz}^{-1}A_{zx}\pi_S^\top\left(A_{xxS}\right)^{-1}\pi_S A_{xz}A_{zz}^{-1}J_z \\ &= A_{zz}^{-1}A_{zx}\Big[I_{p\times p} - (A^{xx})^{1/2}(A^{xx})^{-1/2}\pi_S^\top\left\{\pi_S(A^{xx})^{-1}\pi_S^\top\right\}^{-1}\pi_S \end{aligned}$$

$$
\begin{aligned}
&\times(A^{xx})^{-1/2}(A^{xx})^{-1/2}\Big]\eta + A_{zz}^{-1}J_z \\
&-A_{zz}^{-1}A_{zx}\pi_S^\top\left\{\pi_S(A^{xx})^{-1}\pi_S^\top\right\}^{-1}\pi_S\left(J_x - A_{xz}A_{zz}^{-1}J_z\right) \\
= \;& A_{zz}^{-1}A_{zx}\Big[I_{p\times p} - (A^{xx})^{1/2}(A^{xx})^{-1/2}\pi_S^\top\left\{\pi_S(A^{xx})^{-1}\pi_S^\top\right\}^{-1}\pi_S \\
&\times(A^{xx})^{-1/2}(A^{xx})^{-1/2}\Big]\eta + A_{zz}^{-1}J_z - \Big[A_{zz}^{-1}A_{zx}(A^{xx})^{1/2}(A^{xx})^{-1/2}\pi_S^\top \\
&\times\left\{\pi_S(A^{xx})^{-1}\pi_S^\top\right\}^{-1}\pi_S(A^{xx})^{-1/2}(A^{xx})^{-1/2}\left\{A^{xx}J_x - A^{xx}A_{xz}A_{zz}^{-1}J_z\right\}\Big] \\
\triangleq \;& A_{zz}^{-1}A_{zx}\left\{I_{p\times p} - (A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\right\}\eta \\
&+A_{zz}^{-1}J_z - A_{zz}^{-1}A_{zx}(A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\mathcal{W}, \qquad \text{(B.63)}
\end{aligned}
$$

where the second equality is due to (B.57), and

$$
\mathbb{H}_S = (A^{xx})^{-1/2}\pi_S^\top\left\{\pi_S(A^{xx})^{-1}\pi_S^\top\right\}^{-1}\pi_S(A^{xx})^{-1/2}.
$$

Step 2:

By the Taylor series expansion, we have

$$
\begin{aligned}
\widehat{\mu}_S - \mu_{\text{true}} \;=\; & \left(\frac{\partial\mu}{\partial\beta_{x,S}}\right)^\top\widehat{\beta}_{x,S} + \left(\frac{\partial\mu}{\partial\beta_{z,S}}\right)^\top\left(\widehat{\beta}_{z,S} - \beta_{z0}\right) \qquad \text{(B.64)} \\
&+\frac{\partial\mu}{\partial\Lambda_0}\left(\widehat{\Lambda}_{0,S} - \Lambda_0\right) - \left(\frac{\partial\mu}{\partial\beta_x}\right)^\top\frac{\eta}{\sqrt{n}} + o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Multiplying $\sqrt{n}$ on both sides and plugging in the results of Theorem 4.1 to (B.64) with Slutsky's theorem give that as $n \to \infty$,

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\mu}_S - \mu_{\text{true}}\right) \;\xrightarrow{d}\; & \frac{\partial\mu}{\partial\Lambda_0}\mathcal{V}(t) + \left\{\frac{\partial\mu}{\partial\beta_{x,S}} - \frac{\partial\mu}{\partial\Lambda_0}F_{x,S}(t)\right\}^\top C_S \qquad \text{(B.65)} \\
&+\left\{\frac{\partial\mu}{\partial\beta_z} - \frac{\partial\mu}{\partial\Lambda_0}F_z(t)\right\}^\top D_S - \left\{\frac{\partial\mu}{\partial\beta_x} + \frac{\partial\mu}{\partial\Lambda_0}F_x(t)\right\}^\top\eta.
\end{aligned}
$$

Then plugging in expressions (B.59) and (B.63) to (B.65) and applying Slutsky's theorem with $F_{x,S}(t) = \pi_S F_x(t)$, $\frac{\partial\mu}{\partial\beta_{x,S}} = \pi_S\frac{\partial\mu}{\partial\beta_x}$, and $\frac{\partial\mu}{\partial\beta_{z,S}} = \frac{\partial\mu}{\partial\beta_z}$ yield that as $n \to \infty$,

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{\mu}_S - \mu_{\text{true}}\right) \\
\xrightarrow{d}\; & \frac{\partial\mu}{\partial\Lambda_0}\mathcal{V}(t) + \left\{\frac{\partial\mu}{\partial\beta_{x,S}} - \frac{\partial\mu}{\partial\Lambda_0}F_{x,S}(t)\right\}^\top C_S + \left\{\frac{\partial\mu}{\partial\beta_z} - \frac{\partial\mu}{\partial\Lambda_0}F_z(t)\right\}^\top D_S \\
&-\left\{\frac{\partial\mu}{\partial\beta_x} + \frac{\partial\mu}{\partial\Lambda_0}F_x(t)\right\}^\top\eta \\
=\; & \frac{\partial\mu}{\partial\Lambda_0}\mathcal{V}(t) + \left\{\frac{\partial\mu}{\partial\beta_x} - \frac{\partial\mu}{\partial\Lambda_0}F_x(t)\right\}^\top\left\{(A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\right\}(\eta + \mathcal{W})
\end{aligned}
$$

$$+\left\{\frac{\partial\mu}{\partial\beta_z}-\frac{\partial\mu}{\partial\Lambda_0}F_z(t)\right\}^\top A_{zz}^{-1}A_{zx}\left\{I_{p\times p}-(A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\right\}\eta$$
$$+\left\{\frac{\partial\mu}{\partial\beta_z}-\frac{\partial\mu}{\partial\Lambda_0}F_z(t)\right\}^\top A_{zz}^{-1}J_z$$
$$-\left\{\frac{\partial\mu}{\partial\beta_z}-\frac{\partial\mu}{\partial\Lambda_0}F_z(t)\right\}^\top A_{zz}^{-1}A_{zx}(A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\mathcal{W}$$
$$-\left\{\frac{\partial\mu}{\partial\beta_x}+\frac{\partial\mu}{\partial\Lambda_0}F_x(t)\right\}^\top \eta, \qquad \text{(B.66)}$$

Finally, combining common terms together, we can derive that as $n\to\infty$,

$$\begin{aligned}\sqrt{n}\left(\widehat{\mu}_S-\mu_{\text{true}}\right) \quad \xrightarrow{d} \quad & \frac{\partial\mu}{\partial\Lambda_0}\mathcal{V}(t)+\left\{\frac{\partial\mu}{\partial\beta_z}-\frac{\partial\mu}{\partial\Lambda_0}F_z(t)\right\}^\top A_{zz}^{-1}J_z \\ & +(\omega+\kappa)^\top\left\{\eta-(A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\mathcal{U}\right\},\end{aligned}$$

where $\omega=\frac{\partial\mu}{\partial\beta_x}-A_{zx}^\top A_{zz}^{-1}\frac{\partial\mu}{\partial\beta_z}$ and $\kappa=\frac{\partial\mu}{\partial\Lambda_0}F_x(t)-A_{zx}^\top A_{zz}^{-1}\frac{\partial\mu}{\partial\Lambda_0}F_z(t)$, which completes the proof. □

B.5. Proof of Theorem 5.1

Recall that $\widehat{\mu}_{\text{ave}}=\sum\limits_{S\in\mathcal{S}}w(S|\widehat{\eta})\widehat{\mu}_S$ with weights $w(S|\widehat{\eta})$ satisfying conditions in Section 5.2. Since $\widehat{\eta}=\sqrt{n}\widehat{\beta}_{x,\text{full}}$, then by (B.6) with Slutsky's theorem, we have that as $n\to\infty$,

$$\begin{aligned}\widehat{\eta} &= \sqrt{n}\widehat{\beta}_{x,\text{full}} \\ &= (I_{p\times p},0_{q\times q})\sqrt{n}\begin{pmatrix}\widehat{\beta}_{x,\text{full}}\\ \widehat{\beta}_{z,\text{full}}-\beta_{z0}\end{pmatrix} \\ &\xrightarrow{d} (I_{p\times p},0_{q\times q})\left\{\begin{pmatrix}\eta\\ 0_q\end{pmatrix}+\mathcal{A}^{-1}\begin{pmatrix}J_x\\ J_z\end{pmatrix}\right\} \\ &= \eta+\mathcal{W} \\ &= \mathcal{U}. \end{aligned} \qquad \text{(B.67)}$$

Therefore, let $w(S|\mathcal{U})$ denote the weight to which $w(S|\widehat{\eta})$ converges.

Then by the continuous mapping theorem and the result of Theorem 4.2 and (B.67), we have that as $n\to\infty$,

$$\begin{aligned}\sqrt{n}\left(\widehat{\mu}_{\text{ave}}-\mu_{\text{true}}\right) &= \sum_{S\in\mathcal{S}}w(S|\widehat{\eta})\left\{\sqrt{n}\left(\widehat{\mu}_S-\mu_{\text{true}}\right)\right\} \\ &\xrightarrow{d} \sum_{S\in\mathcal{S}}w(S|\mathcal{U})\left[\frac{\partial\mu}{\partial\Lambda_0}\mathcal{V}(t)+\left\{\frac{\partial\mu}{\partial\beta_z}+\frac{\partial\mu}{\partial\Lambda_0}F_z(t)\right\}^\top A_{zz}^{-1}J_z\right.\end{aligned}$$

$$
\begin{aligned}
& + (\omega+\kappa)^\top \left\{\mathcal{U} - (A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\mathcal{U}\right\}\Big] \\
= \;& \frac{\partial\mu}{\partial\Lambda_0}\mathcal{V}(t) + \left\{\frac{\partial\mu}{\partial\beta_z} + \frac{\partial\mu}{\partial\Lambda_0}F_z(t)\right\}^\top A_{zz}^{-1}J_z \\
& + (\omega+\kappa)^\top \left\{\mathcal{U} - \sum_{S\in\mathcal{S}} w(S|\mathcal{U})(A^{xx})^{1/2}\mathbb{H}_S(A^{xx})^{-1/2}\mathcal{U}\right\}.
\end{aligned}
$$

Therefore, the proof of Theorem 5.1 is completed. □

B.6. *Proof of* $\mathcal{W} \sim N(0, \sigma_{xx})$

By an argument similar to (B.58), we have $A^{xz} = -A^{xx}A_{xz}A_{zz}^{-1}$, or equivalently,

$$(A^{xx})^{-1}A^{xz} = -A_{xz}A_{zz}^{-1}.$$

We re-write (B.60) as

$$
\begin{aligned}
\mathcal{W} &= A^{xx}J_x - A^{xx}A_{xz}A_{zz}^{-1}J_z \\
&= A^{xx}\left(J_x - A_{xz}A_{zz}^{-1}J_z\right) \\
&= A^{xx}\left\{J_x + (A^{xx})^{-1}A^{xz}J_z\right\}. \qquad \text{(B.68)}
\end{aligned}
$$

Write $\mathcal{B}$ as the block matrix $\begin{pmatrix} B_{xx} & B_{xz} \\ B_{zx} & B_{zz}\end{pmatrix}$ where B_{xx}, B_{xz}, B_{zx} and B_{zz} of dimensions $p\times p$, $p\times q$, $q\times p$ and $q\times q$, respectively. Noting that $\mathcal{A}^{-1}$ is a symmetric matrix, then $(A^{xx})^\top = A^{xx}$, $(A^{xz})^\top = A^{zx}$, $(A^{zx})^\top = A^{xz}$ and $(A^{zz})^\top = A^{zz}$. From (B.68), the variance of $\mathcal{W}$ can be expressed as

$$
\begin{aligned}
& \operatorname{var}(\mathcal{W}) \\
= \;& A^{xx}\operatorname{var}\left\{J_x + (A^{xx})^{-1}A^{xz}J_z\right\}A^{xx} \\
= \;& A^{xx}\operatorname{var}(J_x)A^{xx} + A^{xx}(A^{xx})^{-1}A^{xz}\operatorname{var}(J_z)A^{zx}(A^{xx})^{-1}A^{xx} \\
& + A^{xx}\operatorname{cov}\left\{J_x, (A^{xx})^{-1}A^{xz}J_z\right\}A^{xx} + A^{xx}\operatorname{cov}\left\{(A^{xx})^{-1}A^{xz}J_z, J_x\right\}A^{xx} \\
= \;& A^{xx}B_{xx}A^{xx} + A^{xz}B_{zz}A^{zx} \\
& + A^{xx}B_{xz}A^{zx}(A^{xx})^{-1}A^{xx} + A^{xx}(A^{xx})^{-1}A^{xz}B_{zx}A^{xx} \\
= \;& A^{xx}B_{xx}A^{xx} + A^{xz}B_{zz}A^{zx} + A^{xz}B_{zx}A^{xx} + A^{xx}B_{xz}A^{zx}.
\end{aligned}
$$

On the other hand, directly calculations give

$$
\begin{aligned}
\mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1} &= \begin{pmatrix} A^{xx} & A^{xz} \\ A^{zx} & A^{zz}\end{pmatrix}\begin{pmatrix} B_{xx} & B_{xz} \\ B_{zx} & B_{zz}\end{pmatrix}\begin{pmatrix} A^{xx} & A^{xz} \\ A^{zx} & A^{zz}\end{pmatrix} \\
&= \begin{pmatrix} A^{xx}B_{xx} + A^{xz}B_{zx} & A^{xx}B_{xz} + A^{xz}B_{zz} \\ A^{zx} & A^{zz}\end{pmatrix}\begin{pmatrix} A^{xx} & A^{xz} \\ A^{zx} & A^{zz}\end{pmatrix}
\end{aligned}
$$

leading to the upper left block matrix $\sigma_{xx} = A^{xx}B_{xx}A^{xx} + A^{xz}B_{zx}A^{xx} + A^{xx}B_{xz}A^{zx} + A^{xz}B_{zz}A^{zx}$, which is $\operatorname{var}(\mathcal{W})$; the proof is then completed. □

References

[1] Bertrand, A., Van Keilegom, I., and Legrand, C. (2019). Flexible parametric approach to classical measurement error variance estimation without auxiliary data. *Biometrics*, 75, 297–307. MR3953730

[2] Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Model*. CRC Press, New York. MR2243417

[3] Chen, L.-P. and Yi, G. Y. (2020a). Semiparametric methods for left-truncated and right-censored survival data with covariate measurement error. *Annals of the Institute of Statistical Mathematics*. To appear. DOI: 10.1007/s10463-020-00755-2. MR3015032

[4] Chen, L.-P. and Yi, G. Y. (2020b). Analysis of noisy survival data with graphical proportional hazards measurement error models. *Biometrics*. To appear. DOI: 10.1111/biom.13331.

[5] Claeskens, G. and Carroll, R. J. (2007). An asymptotic theory for model selection inference in general semiparametric problems. *Biometrika*, 94, 249–265. MR2331485

[6] Claeskens, G. and Hjort, N. L.(2003). The focused information criterion. *Journal of the American Statistical Association*, 98, 900–945. MR2041482

[7] Claeskens, G. and Hjort, N. L.(2008). *Model Selection and Model Averaging*. Cambridge University Press, New York. MR2431297

[8] Clyde, M. and George, E. I. (2004). Model uncertainty. *Statistical Science*, 19, 81–94. MR2082148

[9] Du, J., Zhang, Z., and Xie, T. (2017). Focus information criterion and model averaging in censored quantile regression. *Metrika*, 80, 547–570. MR3659964

[10] Hjort, N. L. and Claeskens, G. (2003). Frequestist model average estimators. *Journal of the American Statistical Association*, 98, 879–899. MR2041481

[11] Hjort, N. L. and Claeskens, G. (2006). Focused information criteria and model averaging for the Cox hazard regression model. *Journal of the American Statistical Association*, 101, 1449–1464. MR2279471

[12] Hosmer, D. W., Lemeshow, S., and May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data*, Second edition. John Wiley and Sons, New York. MR2383788

[13] Jiang, J. (2010). *Large Sample Techniques for Statistics*. Springer, New York. MR2675055

[14] Jullum, M. and Hjort, N. L. (2019). What price semiparametric Cox regression? *Lifetime Data Analysis*, 25, 406–438. MR3960833

[15] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley, New York. MR0640866

[16] Liang, H. and Zou, G. (2008). Improved AIC selection strategy for survival analysis. *Computational Statistics & Data Analysis*, 52, 2538–2548. MR2411957

[17] Lin, D. Y., Wei, L. J., Yang, L., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of*

Royal Statistical Society, Series B, 62, 711–730. MR1796287

[18] Ning, Y., Yi, G. Y., and Reid, N. (2018). A class of weighted estimating equations for semiparametric transformation models with missing covariates. *Scandinavian Journal of Statistics*, 45, 87–109. MR3764287

[19] Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89, 846–866. MR1294730

[20] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York. MR1385671

[21] Wang, H., Chen, X., and Flournoy, N. (2016). The focused information criterion for varying-coefficient partially linear measurement error models. *Statistical Papers*, 57, 99–113. MR3461951

[22] Wang, H., Li, Y., and Sun, J. (2015). Focused and model average estimation for regression analysis of panel count data. *Scandinavian Journal of Statistics*, 42, 732–745. MR3391689

[23] Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*, 86, 130–143. MR1137104

[24] Wang, M.-C., Brookmeyer, R., and Jewell, N. P. (1993). Statistical models for prevalent cohort data. *Biometrics*, 49, 1–11. MR1221402

[25] Wu, F., Kim, S., Qin, J., Saran, R. and Li, Y. (2018). A pairwise likelihood augmented Cox estimator for left-truncated data. *Biometrics*, 74, 100–108. MR3777930

[26] Xu, G., Wang, S. and Huang, J. Z. (2014). Focused information criterion and model averaging based on weighted composite quantile regression. *Scandinavian Journal of Statistics*, 41, 365–381. MR3207176

[27] Yi, G. Y. and Lawless, J. F. (2007). A corrected likelihood method for the proportional hazards model with covariates subject to measurement error. *Journal of Statistical Planning and Inference*, 137, 1816–1828. MR2323865

[28] Yi, G. Y., Ma, Y., Spiegelman, D., and Carroll R. J. (2015). Functional and structural methods with mixed measurement error and misclassification in covariates. *Journal of the American Statistical Association*, 110, 681–696. MR3367257

[29] Yi, G. Y. (2017). *Statistical Analysis with Measurement Error and Misclassication: Strategy, Method and Application*. Springer, New York. MR3676914

[30] Zhang, X. and Liang, H. (2011). Focused information criterion and model averaging for generalized additive partial linear models. *The Annals of Statistics*, 39, 174–200. MR2797843

[31] Zeng, J., Cheng, W., Hu, G., and Rong, Y. (2019). Model selection and model averaging for semiparametric partially linear models with missing data. *Communications in Statistics – Theory and Methods*, 48, 381–395. MR3945977

[32] Zucker, D. M. and Spiegelman, D. (2008). Corrected score estimation in the proportional hazards model with misclassified discrete covariates. *Statistics in Medicine*, 27, 1911–1933. MR2420352