THE SEMIPARAMETRIC MARK-SPECIFIC PROPORTIONAL HAZARDS MODEL FOR MULTIVARIATE MARKS VIA A SINGLE-INDEX

by

Yuehan Shao

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2020

Approved by:

____

Dr. Yanqing Sun

____

Dr. Qingning Zhou

____

Dr. Yinghao Pan

____

Dr. Weidong Tian

ABSTRACT

YUEHAN SHAO. THE SEMIPARAMETRIC MARK-SPECIFIC PROPORTIONAL HAZARDS MODEL FOR MULTIVARIATE MARKS VIA A SINGLE-INDEX. (Under the direction of DR. YANQING SUN)

Competing risk analysis is commonly applied to time-to-event data with finitely many causes of failure. It alters the probability of the occurrence of an event of interest broken down by a specific cause. Motivated by the HIV vaccine efficacy trials, continuous causes-of-failure (marks) have been discussed in the literature. Methodologies have been developed to model for a continuous univariate mark or to study a parametric structure to relate multiple marks with covariates. In this dissertation, we extend the scope of the previous research and explore a semiparametric mark-specific proportional hazards model accommodating a multivariate continuum of marks via a single-index.

In our model, we allow flexible nonlinear interactions between covariates and multiple marks. To avoid the curse of dimensionality, we incorporated multiple marks into a single-index. A profile estimation procedure is introduced. We adopt the local linear smoothing technique for approximating the unknown functions and then utilize the maximum partial likelihood to estimate the unknown parameters. A detailed computational algorithm is derived. The uniform consistency and asymptotic normality of the proposed estimators are established.

We conduct two simulation studies to evaluate the finite-sample performance of the proposed estimation procedure. Besides, the proposed model and methods are applied to the datasets from two HIV vaccine efficacy trials to access the HIV vaccine efficacy taking account of various protein sequence distances (marks) between the infecting HIV and the HIV strain inside the vaccine.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my greatest appreciation to my advisor, Dr. Yanqing Sun, for her unparalleled support and enthusiastic encouragement during my graduate journey at UNC Charlotte. Her passion for research and extensive knowledge in biostatistics has profoundly impressed and inspired me. Her patient guidance and motivation help me all the time in my research and make this dissertation possible.

I would like to extend my sincere thanks to the rest of my committee members, Dr. Qingning Zhou, Dr. Yinghao Pan, and Dr. Weidong Tian, for their insightful comments and valuable instructions on my research work.

I wish to offer my special appreciation to Dr. Peter Gilbert at Fred Hutchinson Cancer Research Center and the University of Washington for motivating discussions on the data applications and for providing the datasets. This research was partially supported by the National Science Foundation grant DMS1915829 and NIAID NIH award number R37AI054165.

I am also grateful to Drs. Yuanan Diao, Mohammad A. Kazemi, and Shaozhong Deng for their supervisory positions throughout my Ph.D. study and teaching assistant role at the Department of Mathematics and Statistics.

Last but not least, I am incredibly grateful to my parents: Yangdong Shao and Yuling Ma, who instilled in me a love of learning and the value of kindness, honesty, and perseverance. I would also like to extend my gratitude to my family members, Yanghong Shao, Yuhui Wu, Jeffrey Shao, and Alan Shao, for their continuous caring and constructive suggestions in my study and career path. Special thanks to Haoyang Yu. His company, encouragement, and support help me get through the trying times.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Motivation and Literature Review

1.1.1 Proportional Hazards Model

In finitely many competing risks analysis for time-to-event data, researchers wish to develop tools for analyzing the failure time data in the presence of different causes-of-failure (marks). For example, in HIV vaccine efficacy trials, the protein sequence distance between an infecting HIV sequence and an HIV sequence presented in the vaccine is often measured in each trial and treated as a variable mark. The utmost genetic diversity of HIV is one of the most significant obstacles in producing an efficacious vaccine. The genetic heterogeneity of HIV can be measured as the weighted percentage of mismatching in protein sequence distances. This implies that one of the essential objectives of each efficacy trial is to access if and how the vaccine to reduce the risk of HIV infection depends on different protein sequence distances (marks). Mark variable is different from covariate. Mark is assumed to be observed whenever the failure time is uncensored, while covariate is observed for any individuals no matter the failure time is censored or not.

The analysis of competing risks originates from the Cox model. The proportional hazards model (Cox (1972)) is widely used in the medical study for exploring the associations between one or more covariates and survival time. Let T be the failure time, $Z(t)$ be the covariate vector. The Cox model can be written as

$$\lambda(t|z) = \lambda_0(t)\exp\left\{\beta^T z\right\}, \tag{1.1}$$

where $\lambda_0(t)$ is an unspecified baseline function, β is the unknown parameter. One of the most attractive properties of this model is that the parameters, β, can be nicely interpreted as the expected change in the logarithm of the hazard ratio when one covariate changes in one unit, given that all other covariates fixed. Prentice et al. (1978) initially utilized the Cox regression structure to model the effects of variables on conditional cause-specific hazard function,

$$\lambda_j(t|z) = \lim_{h\to 0} P\{T \in [t, t+h), J = j|T \geq t,\ Z(t) = z\}/h, \tag{1.2}$$

for $j = 1, ..., m$, where J describe causes of failure. This model takes the form

$$\lambda_j(t|z) = \lambda_{0j}(t)\exp\left\{\beta_j^T z\right\}, \tag{1.3}$$

for $j = 1, ..., m$. Thenceforth, a large deal of work extended the studies on discrete marks in failure time data. See Kalbfeisch and Prentice (1980), Sun (2001), Scheike et al. (2008) for further details.

The marks in the above model are considered to be discrete. However, in many data applications, it is better to account for continuous marks. For example, in the aforementioned HIV vaccine efficacy trials, the variety of HIV strains from the HIV sequence contained inside the vaccine reveals that discrete marks may not be a valid model assumption. Statistical analysis for quality of life data (score) in cancer clinical trials (OLSCHEWSKI and SCHUMACHER (1990)) is another example of continuous mark variables.

For this reason, researches on the mark-specific hazard function for continuous marks have been considered. This hazard function has the form

$$\lambda(t, v|z) = \lim_{h_1, h_2\to 0} P\{T \in [t, t+h_1), V \in [v, v+h_2)|T \geq t,\ Z(t) = z\}/h_1 h_2, \tag{1.4}$$

where V denotes a vector of continuous marks. Huang and Louis (1998) developed the nonparametric maximum likelihood estimator of the joint distribution of T and a single continuous mark variable V through the unconditional cumulative mark-specific hazard function. In this paper, they also derived the asymptotic properties of the estimators. Motivated by the HIV vaccine efficacy trials, Gilbert et al. (2004) developed nonparametric tests for investigating the relationship between the mark-specific hazard rate function and a single continuous mark variable. Later, Gilbert et al. (2008) expanded the scope of their work and defined the vaccine efficacy (VE) as $\mathrm{VE}(t,v) = 1 - \lambda(t,v|Z=1)/\lambda(t,v|Z=0)$, where Z is the vaccine group indicator. They applied a nonparametric technique for estimating $\mathrm{VE}(t,v)$, and proposed several semiparametric and nonparametric testing procedures for the vaccine efficacy. Besides, the large-sample results for the procedures were established. Sun et al. (2009) developed inferences for the mark-specific proportional hazards model with a univariate continuous mark. This involves the model

$$\lambda(t,v|z(t)) = \lambda_0(t,v)\exp\left\{\beta(v)^T z(t)\right\}, \tag{1.5}$$

where the baseline hazard function $\lambda_0(\cdot,v)$ depends nonparametrically on t and v, and the p-dimensional regression parameter $\beta(v)$ is the unknown continuous function of v. In this model, given any two individuals, the ratio of their hazard functions does not depend on time. In practice, however, this assumption may not always be valid. Instead, models, including stratification, were studied. In particular, a stratified mark-specific proportional hazards model with a single continuous missing mark variable was studied by Sun and Gilbert (2012). The model takes the form

$$\lambda_k(t,v|z(t)) = \lambda_{0k}(t,v)\exp\left\{\beta(v)^T z(t)\right\}, \tag{1.6}$$

for $k = 1, 2, ..., K$, where K is the number of baseline strata. They investigated two estimation procedures based upon the inverse probability weighted (IPW) complete-case (CC) method and upon the augmented inverse probability weighted (AIPW) complete-case (CC) method. Besides, the vaccine efficacy (VE(t, v)) was accessed under the model framework. A Goodness-of-fit Test for this model was conducted in Sun et al. (2014).

There is a limitation when models only include one continuous mark variable. For instance, multiple HIV sequences are contained in HIV vaccines. If more types of HIV viruses are recognized, studied, and blocked, HIV vaccines will potentially be more efficacious. In this sense, it is essential to develop a mark-specific proportional hazards model with a multivariate continuum of marks. Sun et al. (2013) studied the stratified mark-specific PH model with multivariate marks

$$\lambda_k(t, v|z(t)) = \lambda_{0k}(t, v)\exp\left\{\beta(v, \theta)^T z(t)\right\}, \tag{1.7}$$

for $k = 1, 2, ..., K$, where $\lambda_{0k}(\cdot, v)$ is the unknown baseline hazard function for kth stratum depending nonparametrically on t and v. $\beta(v, \theta)$ is a known parametric p-dimensional function with unknown parameters which takes the form

$$\beta(v, \theta) = \theta_0 + \theta_1 v_1 + \theta_2 v_2 + \theta_{12} v_1 v_2.$$

The regression parameters depend parametrically on multiple marks to avoid curse of dimensionality problem. Although polynomial approximations for more complex functions are generally used in real data analysis, it is desirable to develop a more flexible semiparametric mark-specific proportional hazards model for multivariate marks. In the meantime, the model can break the so-called "curse of dimensionality".

1.1.2 Single-index Model

The single-index model is one of the commonly used models in biostatistics and econometrics to maintain latent nonlinear features for the data without complications of high dimensionality (Hardle and Stoker (1989)). The generic formulation of this model is

$$Y = g(\alpha^T X) + \epsilon, \tag{1.8}$$

where Y is the response, X is the multivariate covariate vector, $E(\epsilon|X) = 0$ almost surely, $g(\cdot)$ is the unknown univariate smooth function, and α is an unknown unit vector with one component positive for identification purpose. The dimension reduction framework of the single-index model is particularly popular since the high-dimensional covariate X is reduced to a scalar, the linear combination $\alpha^T X$. The nonlinear function $g(\cdot)$ is utilized to preserve most of the modeling flexibility. The interpretability of the regression parameters α is another attractive feature of the single-index model. The first derivative of $E(Y|X)$ with respect to X is proportional to the coefficient α. In other words, α indicates an instantaneous rate of change in $E(Y|X)$ as X changes.

In the survival analysis, Wang (2004) proposed a two stage approach to take account of the potentially time-dependence and missingness of covariates in the single-index model with the conditional hazard function. This model can be written as

$$\lambda(t|Z) = \lambda_0(t)\text{exp}\left\{\phi(\beta^T Z)\right\}, \tag{1.9}$$

where $\lambda_0(\cdot)$, $\phi(\cdot)$ and β are unknown. Huang and Liu (2006) studied the same model. They adopted spline smoothing method for approximating the unknown link function $\phi(\cdot)$ and estimated the regression parameter β based on maximum partial likelihood. In model (1.9), all components of covariate vector X are treated equally. However, in real data application, the covariate vector X may often be divided into two parts, including principal interest covariates and "nuisance" covariates. Let U be a p-dimensional vector of principal interest covariates and Z be a q-dimensional vector of "nuisance" covariates. Lu et al. (2006) studied the partially linear single-index survival model

$$\lambda(t|U,Z) = \lambda_0(t)\exp\left\{\beta^T U + \phi(\gamma^T Z)\right\}, \tag{1.10}$$

where $\phi(\cdot)$ is unspecified and $\lambda_0(\cdot)$ is known up to a parameter θ. Sun et al. (2008) relaxed the constraint on baseline function $\lambda_0(t)$ and estimated unknown function $\phi(\cdot)$ using polynomial spline smoothing technique. Lin et al. (2016) studied a single-index varying coefficients Cox model, and proposed a global partial likelihood method to estimate $\beta(\cdot)$. The model takes the form

$$\lambda(t) = \lambda_0(t)\exp\left\{\beta(\alpha^T X)^T Z\right\}, \tag{1.11}$$

where X is the multiple biomarker vector (a vector of covariates), Z is the exposure variable, for example, treatment group indicator, $\beta(\cdot)$ is a d -dimensional vector of unknown varying-coefficient functions, and α is an unknown regression coefficient vector. The uniform consistency, asymptotic normality, and semiparametrically efficiency of the estimators were shown in the paper.

As we mentioned at the beginning of this chapter, unlike covariate, a variable mark is only observed in individuals who fail. To the best of our knowledge, the single-index model has not yet been introduced to explore the effects of marks in the competing risk context. This dissertation aims to propose a semiparametric mark-specific proportional hazards model for multivariate marks, which are incorporated into a single-index.

1.2 Dissertation Compendium

To elucidate the objective, the remainder of this dissertation is organized as follows. In Chapter 2, we focus on the proposed model and develop a profiled estimation procedure for unknown parameters and functions in the model. An explicit computational algorithm is given for implementation. In Chapter 3, we establish that, under certain conditions, the proposed estimators for parameters and functions are uniformly consistent and asymptotically normal. Chapter 4 examines our proposed model and methods for finite-sample performance through two simulation studies. One considers bivariate marks, and the other contains multiple marks. In Chapter 5, we illustrate the proposed method with real data applications to two HIV vaccine efficacy trials. The model performance and findings on the applications are discussed.

CHAPTER 2: MARK-SPECIFIC PROPORTIONAL HAZARDS MODEL FOR MULTIVARIATE MARKS VIA A SINGLE-INDEX

In this chapter, we develop the semiparametric mark-specific proportional hazards (PH) model for multivariate marks with a single-index. In Section 2.1, we introduce the proposed model with related notations and assumptions that are used throughout the dissertation. The profile estimation procedure is established for model parameters in Section 2.2. For implementation purposes, in Section 2.3, we derive a computation algorithm for the methodology. In Section 2.4, we discuss the variance estimations for covariance matrices of $\widehat{\widehat{\beta}}(\cdot)$ and $\widehat{\theta}$.

2.1 Model Descriptions

Suppose that n independent and identically distributed observations are sampled from the underlying population. Denote V to be the d-dimensional mark variable. For the purpose of setting up of the identifiability condition, we let V_1 be the first component of the mark variable V, and V_2 be other components of V. Then the d-dimensional mark variable $V = (V_1, V_2^T)^T$ is assumed to be continuous with a known and bounded support. Without loss of generality, we assume the support of V to be $[0, 1]^d$ and rescale if needed.

For i^{th} observation, let T_i be the failure time, C_{0i} be the censoring time, and τ be the end of follow-up time. The right-censored failure time X_i is defined as $min\{T_i, (C_{0i} \wedge \tau)\}$. δ_i is the indicator of non-censorship. It takes value 1 if X_i is the failure time. The possibly time-dependent covariate Z_i is a p-dimensional vector. The mark is assumed to be observed whenever $\delta_i = 1$; V_i is not meaningful and undefined when the corresponding failure time is censored. The censoring time is assumed to be conditionally independent of (T, V) given Z.

We propose the semiparametric mark-specific proportional hazards (PH) model for multivariate marks via a single-index:

$$\lambda(t, v|z(t)) = \lambda_0(t, v)\exp\left\{(\beta(\theta^T v))^T z(t)\right\}, \qquad (2.1)$$

where $\lambda_0(\cdot, v)$ is the baseline hazard function, $\beta(u)$ is a p-dimensional vector of unspecified continuous functions of $u \in R$, and θ is a d-dimensional vector of parameters. $\theta^T v$ is the index used to combine multivariate marks. For identification purpose, we impose the restriction on the first parameter of θ, $\theta_1 = 1$.

2.2 Profile Estimation Procedure

When all the observations are i.i.d., the partial likelihood for (2.1) can be expressed as

$$L(\beta, \theta) = \prod_{i=1}^{n} \left[\frac{\exp\{\beta(\theta^T V_i)^T Z_i(t)\}}{\sum_{k=1}^{n} Y_k(X_i) \exp\{\beta(\theta^T V_i)^T Z_k(t)\}} \right]^{\delta_i}, \qquad (2.2)$$

where $Y_i(t) = I(X_i \geq t)$ is an indicator that equals 1 if the i^{th} subject is at risk just before time t.

Since $\beta(\cdot)$ is an unspecified p-dimensional vector, for a fixed u, we approximate $\beta(\theta^T v)$ for $\theta^T v$ in a neighborhood of u using a Taylor expansion,

$$\beta_l(\theta^T v) \approx \beta_l(u) + \beta_l'(u)(\theta^T v - u), \quad l = 1, 2, \ldots, p. \qquad (2.3)$$

Let $\widetilde{\beta}(u) = \left(\beta_1(u), \ldots, \beta_p(u), \beta_1'(u), \ldots, \beta_p'(u)\right)^T$ and $\overline{\beta}(u, \theta^T v) = \beta(u) + \beta'(u)(\theta^T v - u)$. Denote $\widetilde{Z}_i(t, u, \theta^T v) = (1, \theta^T v - u)^T \otimes Z_i(t)$. where $\otimes$ is the Kronecker product. Specifically, $\widetilde{Z}_i(t, u, \theta^T v) = (Z_i(t)^T, Z_i(t)^T(\theta^T v - u))^T$.

Then, for the i^{th} observation,

$$\begin{aligned}
&\overline{\beta}(u,\theta^T v)^T Z_i(t)\\
&= \beta(u)^T Z_i(t) + \beta'(u)^T Z_i(t)(\theta^T v - u)\\
&= \widetilde{\beta}(u)^T \big(Z_{1i}(t), \ldots, Z_{pi}(t), Z_{1i}(t)(\theta^T v - u), \ldots, Z_{pi}(t)(\theta^T v - u)\big)^T\\
&= \widetilde{\beta}(u)^T \widetilde{Z}_i(t, u, \theta^T v). \qquad (2.4)
\end{aligned}$$

To estimate model parameters $\beta(\cdot)$ and θ, we use the convenient profile estimation approach. Given θ, we estimate $\beta(u)$ first.

Let $N_i(t,v) = I(X_i \le t, \delta_i = 1, V_i \le v)$ be the marked point counting process for subject i. It jumps when i^th observation has an uncensored failure time with the corresponding mark V_i. Denote $N_i^*(t) = I(X_i \le t, \delta_i = 1)$. Then, the localized version of the log partial likelihood function for $\beta(u)$ at a given u is

$$\begin{aligned}
\ell(\widetilde{\beta}, u, \theta) = \sum_{i=1}^{n} \int_0^{\tau} K_h(\theta^T V_i - u) \Big\{ &(\widetilde{\beta}(u))^T \widetilde{Z}_i(t, u, \theta^T V_i)\\
&- \log \left[\sum_{k=1}^{n} Y_k(t) \exp\left((\widetilde{\beta}(u))^T \widetilde{Z}_k(t, u, \theta^T V_i) \right) \right] \Big\} N_i^*(dt),
\end{aligned} \qquad (2.5)$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a symmetric kernel density function with support $[-1, 1]$, and h is the bandwidth.

Define

$$S_n^{(j)}(t, v; \widetilde{\beta}, u, \theta) = \frac{1}{n} \sum_{k=1}^{n} Y_k(t) \exp\left((\widetilde{\beta}(u))^T \widetilde{Z}_k(t, u, \theta^T v) \right) \left(\widetilde{Z}_k(t, u, \theta^T v) \right)^{\otimes j},$$

for $j = 0, 1$ and 2. Here $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$ for a column vector a.

Taking the derivative of (2.5) with respect to $\widetilde{\beta}(u)$, we obtain the score function:

$$U(\widetilde{\beta}, u, \theta) = \sum_{i=1}^{n} \int_{0}^{\tau} K_h(\theta^T V_i - u)\Big(\widetilde{Z}_i(t, u, \theta^T V_i) - \overline{Z}(t, \widetilde{\beta}, u, \theta^T V_i)\Big)\, N_i^*(dt), \quad (2.6)$$

where $\overline{Z}(t, \widetilde{\beta}, u, \theta^T v) = S_n^{(1)}(t, v; \widetilde{\beta}, u, \theta)/S_n^{(0)}(t, v; \widetilde{\beta}, u, \theta)$. The local linear maximum partial likelihood estimator $\widehat{\widetilde{\beta}}(u, \theta)$ is a solution to $U(\widetilde{\beta}, u, \theta) = 0$, and can be computed using a Newton-Raphson algorithm. The second derivative of $\ell(\widetilde{\beta}, u, \theta)$ with respect to $\widetilde{\beta}(u)$ yields

$$I(\widetilde{\beta}, u, \theta) = -\sum_{i=1}^{n} \int_{0}^{\tau} K_h(\theta^T V_i - u) \times \left\{ \frac{S_n^{(2)}(t, V_i; \widetilde{\beta}, u, \theta)}{S_n^{(0)}(t, V_i; \widetilde{\beta}, u, \theta)} - \Big(\overline{Z}(t, \widetilde{\beta}, u, \theta^T V_i)\Big)^{\otimes 2} \right\} N_i^*(dt). \quad (2.7)$$

The estimator $\widehat{\beta}(u, \theta)$ of $\beta(u)$ for the fixed θ is the vector consisting of the first p components of $\widehat{\widetilde{\beta}}(u, \theta)$. In the following, we use $\widehat{\beta}(u)$ for $\widehat{\beta}(u, \theta)$ for simplicity. After incorporating the solutions $\widehat{\beta}(u)$ into the partial likelihood function (2.2), we can update θ by maximizing the following log partial likelihood function using Newton-Raphson algorithm:

$$l_p(\theta) = \sum_{i=1}^{n} \int_{0}^{\tau} \left[(\widehat{\beta}(\theta^T V_i))^T Z_i(t) - \log\left(\sum_{k=1}^{n} Y_k(t) \exp\Big(\widehat{\beta}(\theta^T V_i)^T Z_k(t)\Big) \right) \right] N_i^*(dt). \quad (2.8)$$

Define

$$S_n^{*(j)}(t, v; \widehat{\beta}, \widehat{\beta}', \theta) = \frac{1}{n} \sum_{k=1}^{n} Y_k(t) \exp\Big(\widehat{\beta}(\theta^T v)^T Z_k(t)\Big) \Big(v_2 \widehat{\beta}'(\theta^T v)^T Z_k(t)\Big)^{\otimes j},$$

for $j = 0, 1$ and 2.

Taking derivative of (2.8) with respect to θ, the corresponding profile partial estimating function $U_p(\theta)$ for θ is

$$U_p(\theta) = \sum_{i=1}^{n} \int_0^\tau \left\{ V_{2i}\widehat{\beta}'(\theta^T V_i)^T Z_i(t) - \frac{S_n^{*(1)}(t, V_i; \widehat{\beta}, \widehat{\beta}', \theta)}{S_n^{*(0)}(t, V_i; \widehat{\beta}, \widehat{\beta}', \theta)} \right\} N_i^*(dt). \tag{2.9}$$

We estimate θ by $\widehat{\theta}$ that is the root of the score equation $U_p(\theta) = 0$. The regression function $\beta(u)$ is estimated by $\widehat{\beta}(u, \widehat{\theta})$.

2.3 Computational Algorithm

In this section, we derive an iteration algorithm for implementing the profiled estimation procedure introduced in Section 2.2.

Let $\widehat{\theta}_{(s)}$, $\widehat{\beta}_{(s)}(\cdot)$ and $\widehat{\beta}'_{(s)}(\cdot)$ be the estimators of θ, $\beta(\cdot)$ and $\beta'(\cdot)$ for the sth iteration, respectively. Particularly, we denote $\theta_{(0)}$ to be the initial value of θ, and $\beta_{(0)}(\cdot)$, $\beta'_{(0)}(\cdot)$ to be the initial values of $\beta(\cdot)$ and $\beta'(\cdot)$.

The detailed computational algorithm is provided as follows.

(1) Initialize $\theta_{(0)}$. Let $u^0_{min} = min(\theta^T_{(0)} V)$ and $u^0_{max} = max(\theta^T_{(0)} V)$. Then, we take the grid of n_0 evenly spaced points in $[u^0_{min}, u^0_{max}]$ and choose the initial values of functions $\beta_{(0)}(u)$ and $\beta'_{(0)}(u)$ for $u = u^0_{min}, ..., u^0_{max}$. Here, n_0 is the number of grid points.

(2) Given $\widehat{\theta}_{(s-1)}$, we estimate $\beta(\cdot)$ in the following.

Let $u^s_{min} = min(\widehat{\theta}^T_{(s-1)} V)$ and $u^s_{max} = max(\widehat{\theta}^T_{(s-1)} V)$. n_0 equally spaced grid points are taken in the interval $[u^s_{min}, u^s_{max}]$. For every fixed grid point $u = u^s_{min}, ..., u^s_{max}$, maximize (2.5) with respect to $\widetilde{\beta}$ and solve the following local partial score equation for $\widetilde{\beta}$:

$$\sum_{i=1}^{n} \int_0^\tau K_h(\widehat{\theta}^T_{(s-1)} V_i - u) \times \left\{ (1, \widehat{\theta}^T_{(s-1)} V_i - u)^T \otimes Z_i(t) - \overline{Z}(t, \widetilde{\beta}, u, \widehat{\theta}^T_{(s-1)} V_i) \right\} N_i^*(dt) = 0. \tag{2.10}$$

Let $\widehat{\widehat{\beta}}_{(s)}(u) = \{\widehat{\beta}_{(s)}(u)^T, \widehat{\beta}'_{(s)}(u)^T\}^T$ be the solution. For each $\widehat{\theta}^T_{(s-1)}V_i$, $i = 1, ..., n$, search for the corresponding closest grid point u^*_i. Then, $\widehat{\beta}_{(s)}(\widehat{\theta}^T_{(s-1)}V_i) = \widehat{\beta}(u^*_i)$ and $\widehat{\beta}'_{(s)}(\widehat{\theta}^T_{(s-1)}V_i) = \widehat{\beta}'(u^*_i)$ for $i = 1, 2, ..., n$.

(3) Given constants r_1 and r_2, let $a_s = u^s_{min} + r_1 h$ and $b_s = u^s_{max} - r_2 h$. For given $\widehat{\beta}_{(s)}$ and $\widehat{\beta}'_{(s)}$, the partial score equation for θ can be expressed as

$$\sum_{i=1}^{n} \int_0^{\tau} I(\theta^T V_i \in [a_s, b_s]) \Bigg\{ V_{2i} \widehat{\beta}'_{(s)}(\widehat{\theta}^T_{(s-1)} V_i)^T Z_i(t) - \frac{\sum_{k=1}^n Y_k(t) \exp\left(\widehat{\beta}_{(s)}(\theta^T V_i)^T Z_k(t)\right)\left(V_{2i}\widehat{\beta}'_{(s)}(\widehat{\theta}^T_{(s-1)}V_i)^T Z_k(t)\right)}{\sum_{k=1}^n Y_k(t) \exp\left(\widehat{\beta}_{(s)}(\theta^T V_i)^T Z_k(t)\right)} \Bigg\} \times N^*_i(dt) = 0. \tag{2.11}$$

The sth estimate of θ is $\widehat{\theta}_{(s)}$, which can be solved using the Newton-Raphson algorithm.

We repeat the steps (2) and (3) of the above iteration procedure until the maximum of the absolute differences of the estimates between two successive steps meets the convergence criteria. Let s_* be the number of iterations. After convergence, we obtain the final results $\widehat{\theta}_{(s_*)}$ and $\widehat{\widehat{\beta}}_{(s_*)}(u)$ for $u = min(\widehat{\theta}^T_{(s_*)}V), ..., max(\widehat{\theta}^T_{(s_*)}V)$. The estimator $\widehat{\beta}(u)$ of $\beta(u)$ consists of the first p elements of $\widehat{\widehat{\beta}}_{(s_*)}(u)$. The final solutions of (θ, β) are denoted by ($\widehat{\theta}$, $\widehat{\beta}$).

2.4 Variance Estimation

In this section, we discuss the variance estimation procedure for estimators, $\widehat{\theta}$ and $\widehat{\beta}(\cdot)$.

We propose to estimate the covariance matrix of $\widehat{\widehat{\beta}}(\cdot)$ similar to Sun et al. (2009).

Specifically, we let

$$\widetilde{\Sigma}_n(\widehat{\widehat{\beta}}(u)) = \frac{h}{n}\sum_{i=1}^{n}\int_0^{\tau}(K_h(\widehat{\theta}^T V_i - u))^2 \times \left\{ \frac{S_n^{(2)}(t, V_i; \widehat{\widehat{\beta}}, u, \widehat{\theta})}{S_n^{(0)}(t, V_i; \widehat{\widehat{\beta}}, u, \widehat{\theta})} - \left(\overline{Z}(t, \widehat{\widehat{\beta}}, u, \widehat{\theta}^T V_i)\right)^{\otimes 2} \right\} N_i^*(dt), \quad (2.12)$$

and

$$\mathrm{I}(\widehat{\widehat{\beta}}(u), u, \widehat{\theta}) = -\sum_{i=1}^{n}\int_0^{\tau} K_h(\widehat{\theta}^T V_i - u) \times \left\{ \frac{S_n^{(2)}(t, V_i; \widehat{\widehat{\beta}}, u, \widehat{\theta})}{S_n^{(0)}(t, V_i; \widehat{\widehat{\beta}}, u, \widehat{\theta})} - \left(\overline{Z}(t, \widehat{\widehat{\beta}}, u, \widehat{\theta}^T V_i)\right)^{\otimes 2} \right\} N_i^*(dt). \quad (2.13)$$

The asymptotic covariance matrix of $\widehat{\widehat{\beta}}(u)$ can be estimated by

$$\frac{1}{nh}(\mathrm{I}(\widehat{\widehat{\beta}}(u), u, \widehat{\theta})/n)^{-1}\widetilde{\Sigma}_n(\widehat{\widehat{\beta}}(u))(\mathrm{I}(\widehat{\widehat{\beta}}(u), u, \widehat{\theta})/n)^{-1}. \quad (2.14)$$

Then, the estimator for the variance of $\widehat{\beta}(u)$ is the first element on the diagonal of (2.14).

To estimate the covariate matrix of $\widehat{\theta}$, we apply the second-order finite difference method with profile partial likelihood for $\widehat{\theta}$ (Zeng et al. (2016)).

From the estimation procedure for θ in Section 2.2, the profile log partial likelihood function for θ can be written as

$$\ell_p(\theta) = \max_{\beta} \log L(\beta, \theta). \quad (2.15)$$

Specifically, to calculate it, we maximize again the loglikelihood $\log L(\beta, \theta)$ with θ held fixed. In other words, we follow the steps (1) and (2) in the computational algorithm to evaluate $\widehat{\beta}$, and then plug in back to the partial loglikelihood function (2.15).

Then, we can estimate the covariate matrix of $\widehat{\theta}$ by the negative inverse of a $(d-1) \times (d-1)$ matrix whose (m, n)th element is

$$\frac{\ell_p(\widehat{\theta}) + \ell_p(\widehat{\theta} + b\mathrm{e}_m + b\mathrm{e}_n) - \ell_p(\widehat{\theta} + b\mathrm{e}_m) - \ell_p(\widehat{\theta} + b\mathrm{e}_n)}{b^2}, \tag{2.16}$$

where e_m is the mth canonical basis and b is a constant of order $n^{-1/2}$.

CHAPTER 3: ASYMPTOTIC RESULTS

In this chapter, we explore the uniform consistency and asymptotic normality of the proposed estimators, $\widehat{\beta}$ and $\widehat{\theta}$. In Section 1, we introduce the notations that are used when stating the theorems. Section 2 lists all asymptotic properties of the proposed estimators. Further notations and detailed proofs can be found in Appendix A.

3.1 Notations and Conditions

We assume that the bounded support of $\theta^T v$ is $[\iota_1, \iota_2]$, where ι_1 and ι_2 are constants. The parameter θ is identifiable up to a scale shift, and we impose the restriction on the first element of θ, setting $\theta_1 = 1$. To facilitate notations, we denote $\theta = (\theta_1, \theta_2^T)^T$, where θ_1 is the first component of θ and is equal to 1, and θ_2 consists of other components of θ. Correspondingly, we set $v = (v_1, v_2^T)^T$, where v_1 contains the first element of v, and v_2 contains other elements of v. The covariate $Z(t)$ is possibly time-dependent. The proofs work for time-dependent $Z(t)$, but we may drop the dependence of $Z(t)$ on t when it does not cause confusion for simplicity.

To state the theorems, we introduce the following notations. We adopt $\oint$ to represent the integration of multidimensional v. Let Θ be the support of θ and θ_0 be the true value of θ. Under the restriction on θ, specifically, $\theta_0 = (1, \theta_{20}^T)^T$, where θ_{20} is the true value of θ_2. Set $\mathbb{S} = \{w(u) = (w_1(u), ..., w_p(u)) : u \in [\iota_1, \iota_2], w(u)$ is continuous on $[\iota_1, \iota_2]\}$, $\mathbf{H} = diag\{I_p, hI_p\}$, $\widetilde{w} = \mathbf{H}(w_1^T, w_2^T)^T$ and $\widetilde{Z}_i(u, \theta^T v) = \mathbf{H}^{-1}(Z_i^T, Z_i^T(\theta^T v - u))^T$.

Define

- $\mu_j = \int u^j K(u) du$, $\nu_j = \int u^j K^2(u) du$, for $j = 0, 1$ and 2,
- $P(t|z) = P(X \geq t | Z = z)$,

- $s^{(j)}(t,v;w_1,w_2,u,\theta) = E[P(t|Z)exp\{\widetilde{w}(u)^T\widetilde{Z}(u,\theta^Tv)\}\widetilde{Z}(u,\theta^Tv)^{\otimes j}]$, for $j = 0$ and 1,
- $\widetilde{s}^{(j)}(t,w_1,u) = E[P(t|Z)exp\{w_1(u)^TZ\}Z^{\otimes j}]$, for $j = 0, 1$ and 2,
- $s^{*(j)}(t,u) = E\Big[P(t|Z)exp\{\beta(u)^TZ\}\Big(\beta'(u)^TZ\Big)^{\otimes j}\Big]$, for $j = 0, 1$ and 2,
- $\eta_1(t,v;\theta_a,\theta_b,w_1,w_2) = E\Big[P(t|Z)exp\{w_1(\theta_a^Tv)^TZ\}\Big(w_2(\theta_b^Tv)^TZ\Big)\Big]$,
- $\phi(t,u) = E\Big[P(t|Z)exp\{\beta(u)^TZ\}\Big(\beta'(u)^TZ\Big)Z\Big]$.

Here $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$ for a column vector a.

Condition A

(A.1) The possibly time-dependent covariate Z is bounded with a compact set.

(A.2) τ is finite, $P(X > \tau) > 0$, $P(C = \tau) > 0$ and $P(C = 0|Z = z) \neq 1$.

(A.3) The unknown parameter θ is bounded with a compact support Θ.

(A.4) Each component of $\beta(u)$ has a continuous second derivative on $u \in [\iota_1, \iota_2]$. The continuous second-order partial derivative of the baseline function $\lambda_0(t,v)$ with respect to v exists on $[0,\tau] \times [0,1]^d$.

(A.6) The kernel function $K(\cdot)$ is bounded, symmetric density function with continuous derivative and compact support $[-1,1]$. The bandwidth satisfies $nh^3 \to \infty$, $nh^4 \to 0$ and $h^2 log(n) \to 0$ as $n \to \infty$.

(A.7) Define $\mathbb{M} = \{v|v \in [0,1]^d, \theta^Tv = u\}$ and adopt $\oint$ to represent the integration of multidimensional v. Functions

$$\oint_{\mathbb{M}}\int_0^{\tau}\left\{\widetilde{s}^{(1)}(t,\beta,\theta_0^Tv) - \frac{\widetilde{s}^{(1)}(t,w_1,u)}{\widetilde{s}^{(0)}(t,w_1,u)}\widetilde{s}^{(0)}(t,\beta,\theta_0^Tv)\right\}\lambda_0(t,v)dtdv = 0,$$

and

$$\oint_0^1 \int_0^\tau v_2 \left\{ \eta_1(t, v; \theta_0, \theta, \beta, w_2) - \frac{\eta_1(t, v; \theta, \theta, w_1, w_2)}{\widetilde{s}^{(0)}(t, w_1, \theta^T v)} \widetilde{s}^{(0)}(t, \beta, \theta_0^T v) \right\}$$
$$\times \lambda_0(t, v) dt dv = 0,$$

exist a unique solution (w_1, θ) for $\theta \in \Theta$, $w_1 \in \mathbb{S}$ and any bounded function w_2.

(A.8) For $j = 0, 1$ and 2, each component of functions $s^{(j)}(t, v; w_1, w_2, u, \theta)$, $\widetilde{s}^{(j)}(t, w_1, u)$, $s^{*(j)}(t, u)$, $\eta_1(t, v; \theta_a, \theta_b, w_1, w_2)$ and $\phi(t, u)$ has second continuously derivative for $t \in [0, \tau]$, $v \in [0, 1]$, $w_1 \in \mathbb{S}$, bounded w_2, $\theta \in \Theta$ and $u \in [\iota_1, \iota_2]$.

(A.9) The matrix

$$\begin{bmatrix} \widetilde{s}^{(2)}(t, \beta, u) & \widetilde{s}^{(1)}(t, \beta, u) \\ \widetilde{s}^{(1)}(t, \beta, u)^T & \widetilde{s}^{(0)}(t, \beta, u) \end{bmatrix}$$

is nonsingular at $u \in [\iota_1, \iota_2]$.

Define $\mathbb{M}_0 = \left\{ v \middle| v \in [0, 1]^d, \theta_0^T v = u \right\}$. The matrix,

$$\oint_{\mathbb{M}_0} \int_0^\tau \left\{ \widetilde{s}^{(2)}(t, \beta, u) - \frac{\widetilde{s}^{(1)}(t, \beta, u) \widetilde{s}^{(1)}(t, \beta, u)^T}{\widetilde{s}^{(0)}(t, \beta, u)} \right\} \lambda_0(t, v) dt dv,$$

is positive definite at $u \in [\iota_1, \iota_2]$.

3.2 Asymptotic Properties of the Proposed Estimators

Theorem 1. *Under Condition A, we have*

(a) $\widehat{\theta} \xrightarrow{p} \theta_0$ *as* $n \to \infty$.

(b) $\widehat{\beta}(u) \xrightarrow{p} \beta(u)$ *uniformly over* $u \in [\iota_1, \iota_2]$ *as* $n \to \infty$.

Theorem 2. *Under Condition A, if* $nh^4 \to 0$*, then*

$$\sqrt{n}(\widehat{\theta}_2 - \theta_{20}) \to N(0, A_\theta^{-1}\Sigma_\theta(A_\theta^{-1})^T),$$

where

$$A_\theta = \oint_0^1 \int_0^\tau v_2 \left\{ \frac{s^{*(1)}(t, \theta_0^T v)^{\otimes 2}}{s^{*(0)}(t, \theta_0^T v)} - s^{*(2)}(t, \theta_0^T v) \right\} v_2^T \lambda_0(t, v) dt dv,$$

$$\Sigma_\theta = \oint_0^1 \int_0^\tau E[\varphi_i^2(t, v_2) P(t|Z_i) \exp\{\beta(\theta_0^T v)^T Z_i\}] \lambda_0(t, v) dt dv,$$

$$\varphi_i(t, v_2) = \left\{ \left(\int_{\iota_1}^{\iota_2} \rho(u)\zeta(u) du A_\theta^{-1} \right) - I \right\} v_2 \left\{ \beta'(\theta_0^T v)^T Z_i - \frac{s^{*(1)}(t, \theta_0^T v)}{s^{*(0)}(t, \theta_0^T v)} \right\}$$
$$- \rho(\theta_0^T v) \left\{ Z_i - \frac{\widetilde{s}^{(1)}(t, \beta, \theta_0^T v)}{\widetilde{s}^{(0)}(t, \beta, \theta_0^T v)} \right\},$$
$$\zeta(u) = \oint_{\mathbb{M}_0} \int_0^\tau \left\{ \frac{\widetilde{s}^{(1)}(t, \beta, u) s^{*(1)}(t, u)}{\widetilde{s}^{(0)}(t, \beta, u)} - \phi(t, u) - \frac{\partial \widetilde{s}^{(1)}(t, \beta, u)}{\partial u} \right\} v_2^T$$
$$\times \lambda_0(t, v) dt dv,$$

$\mathbb{M}_0$ *is defined as* $\mathbb{M}_0 = \left\{ v | v \in [0,1]^d, \theta_0^T v = u \right\}$ *in Lemma 3 and* $\rho(\cdot)$ *is defined in (A.61) in Appendix A.*

Theorem 3. *Under Condition A, if $nh^4 \to 0$, then*

$$\sqrt{nh}\Big[\widehat{\beta}(u) - \beta(u) - \frac{1}{2}h^2\mu_2(\mathcal{I} - \mathcal{L})^{-1}\beta''(u)\Big] \to N(0, \nu_0\Pi(u)\Pi(u)^T),$$

where

$\mathcal{L}$ is the linear operator that satisfies for any function g,

$\mathcal{L}(g)(u) = \Omega^{-1}(u)\oint_0^1 \Upsilon(v;u)g(v)dv$, $\mathcal{I}$ is the identity operator,

$$\begin{aligned}
\nu_0 &= \int K^2(u)du, \\
\Pi(u) = & \quad (\mathcal{I} - \mathcal{L})^{-1}(\Omega^{-1/2})(u), \\
\Omega(u) = & \oint_{\mathbb{M}_0}\int_0^\tau \Big\{\widetilde{s}^{(2)}(t,\beta,u) - \frac{\widetilde{s}^{(1)}(t,\beta,u)\widetilde{s}^{(1)}(t,\beta,u)^T}{\widetilde{s}^{(0)}(t,\beta,u)}\Big\}\lambda_0(t,v)dtdv, \\
\Upsilon(v;u) = & -\int_0^\tau \zeta(u)A_\theta^{-1}v_2\Big\{\frac{s^{*(1)}(t,\theta_0^Tv)\widetilde{s}^{(1)}(t,\beta,\theta_0^Tv)^T}{s^{*(0)}(t,\theta_0^Tv)} - \phi(t,\theta_0^Tv)^T\Big\}\lambda_0(t,v)dt.
\end{aligned}$$

CHAPTER 4: SIMULATION STUDY

In this chapter, we conduct two simulation studies to evaluate the finite-sample performance of the proposed model and methodology.

Basically, we implement the proposed computational algorithm introduced in Section 2.3. Notice that in each replicate, after convergence, we obtain the final results $\widehat{\theta}_{(s_*)}$ and $\widehat{\widehat{\beta}}_{(s_*)}(u)$ for $u = min(\widehat{\theta}^T_{(s_*)}V), ..., max(\widehat{\theta}^T_{(s_*)}V)$, where s_* is the number of iterations. Since the estimates of $\widehat{\theta}$ are different between any two simulations, the resulting grid points u are different. For the purpose of performance evaluations, we introduce an additional step as the last step (Step L) in each simulation.

Specifically, we set a certain range $[a, b]$ to be studied and take the grid of n_0 evenly spaced points in $[a, b]$: $u = a, a + \frac{b-a}{n_0}, ..., b$. In each simulation, after convergence, we find the corresponding closest grid point $u_i^{L^*}$ for each u_i. Thus, $\widehat{\beta}(u_i) = \widehat{\beta}(u_i^{L^*})$ and $\widehat{\beta}'(u_i) = \widehat{\beta}'(u_i^{L^*})$ for $i = 1, 2, ..., n_0$.

This chapter is organized as follows. In Section 4.1, we conduct a simulation study on the proposed model with bivariate marks. We illustrate the methodologies on the case with multiple marks in Section 4.2.

4.1 Example 1: A Mark PH model with Bivariate Marks

In this section, we examine the performance of the proposed local partial likelihood estimators on the case with bivariate marks $V = (V_1, V_2)^T$. The constraint on the first component of θ is set to be $\theta_1 = 1$ for identifiability purpose. In this case, only the second component of θ, θ_2, is estimated.

Let z be a binary covariate taking value 0 or 1 with a given probability 0.5 for each subject. The variables (T, V) are generated from the following multivariate mark-specific proportional hazards model:

$$\lambda(t, v|z) = \exp\{\gamma^T v + \beta(\theta^T v)^T z\}, \tag{4.1}$$

where $0 \le t \le \tau$ with $\tau = 2$, and $v = (v_1, v_2)^T$ with $0 \le v_i \le 1$ for $i = 1, 2$.

We set $\gamma = (\gamma_1, \gamma_2)^T = (0.6, 0.4)^T$ and $\theta = (\theta_1, \theta_2)^T = (1, 1.5)^T$. Under model (4.1), the mark-specific baseline function is $\lambda_0(t, v) = \exp\{\gamma^T v\} = \exp\{0.6v_1 + 0.4v_2\}$. The unknown function $\beta(\cdot)$ is set to be a linear function with the form $\beta(u) = -1.65 + 1.0(u)$. We use the Epanechnikov kernel $K(x) = 0.75(1 - x^2)I\{|x| \le 1\}$.

We generate the censoring time (C_{0i}) from a exponential distribution with mean $\lambda = 2$. Failure time (T_i) beyond τ are considered censored. The indicator of non-censorship δ_i takes the value 1 if X_i is failure time, where $X_i = min(T_i, \tau \wedge C_{0i})$. The censoring rates range from 20% to 30%.

The bandwidths are selected as $h = 0.25, 0.35$ or 0.45. Sample sizes of $n = 800, 1000$ and 1200 are studied. We used 400 replicates for each combination of sample size and bandwidth.

We use $n_0 = n$, and choose $r_1 = r_2 = 1$ when estimating θ. In Step L, we set the interval $[a, b]$ to be $[0, 2.5]$. For the variance estimation of $\widehat{\theta}$, we use $b = 9n^{-1/2}$.

Figures 4.1 and 4.2 depict the biases, the standard errors of estimates (SEE), the means of the estimated standard error (ESE), and the coverage probabilities (CP) of the unknown function $\beta(\cdot)$ under different settings on sample sizes and bandwidths. They show that the bias of the estimator decreases as the sample size increases. The estimated standard error approximates the sample standard error pretty well, and the coverage probability is close to the nominal level (95%). The estimation results for θ_2 are summarized in Table 4.1. The parameter estimator has a smaller bias for a larger sample size. ESE and SEE are close to each other, and the coverage probability approaches their nominal level of 0.95 as the sample size increases.

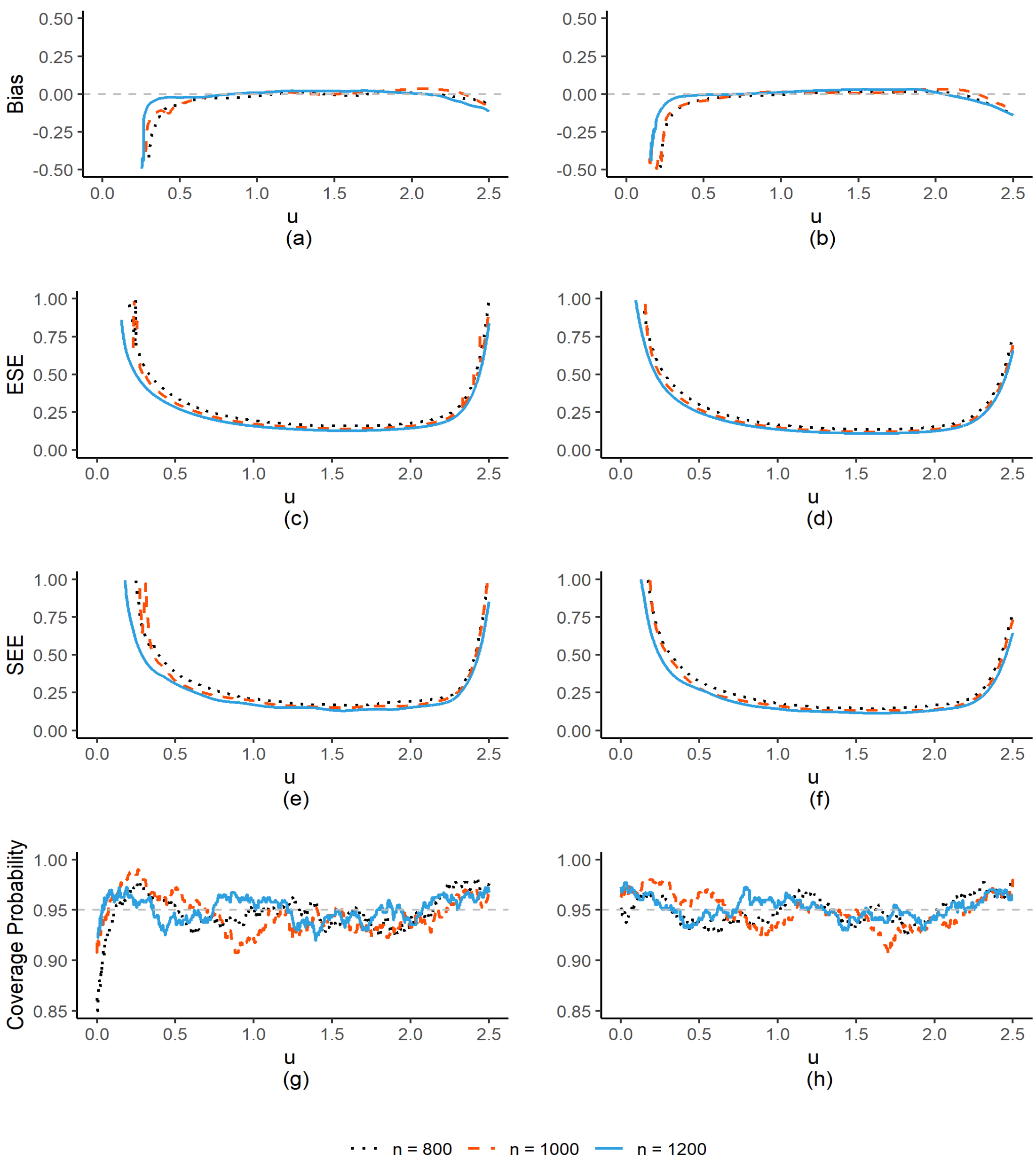


Figure 4.1: Plots of bias, SEE, ESE and CP for $\beta(u)$ with sample size of $n = 800$, $n = 1000$ and $n = 1200$ for Example 1. The number of simulations is 400. The yellow dotted lines are for sample size 800. The red dashed lines are for sample size 1000. The blue solid lines are for sample size 1200. Left panel is for bandwidth $h = 0.25$. Right panel is for bandwidth $h = 0.35$.

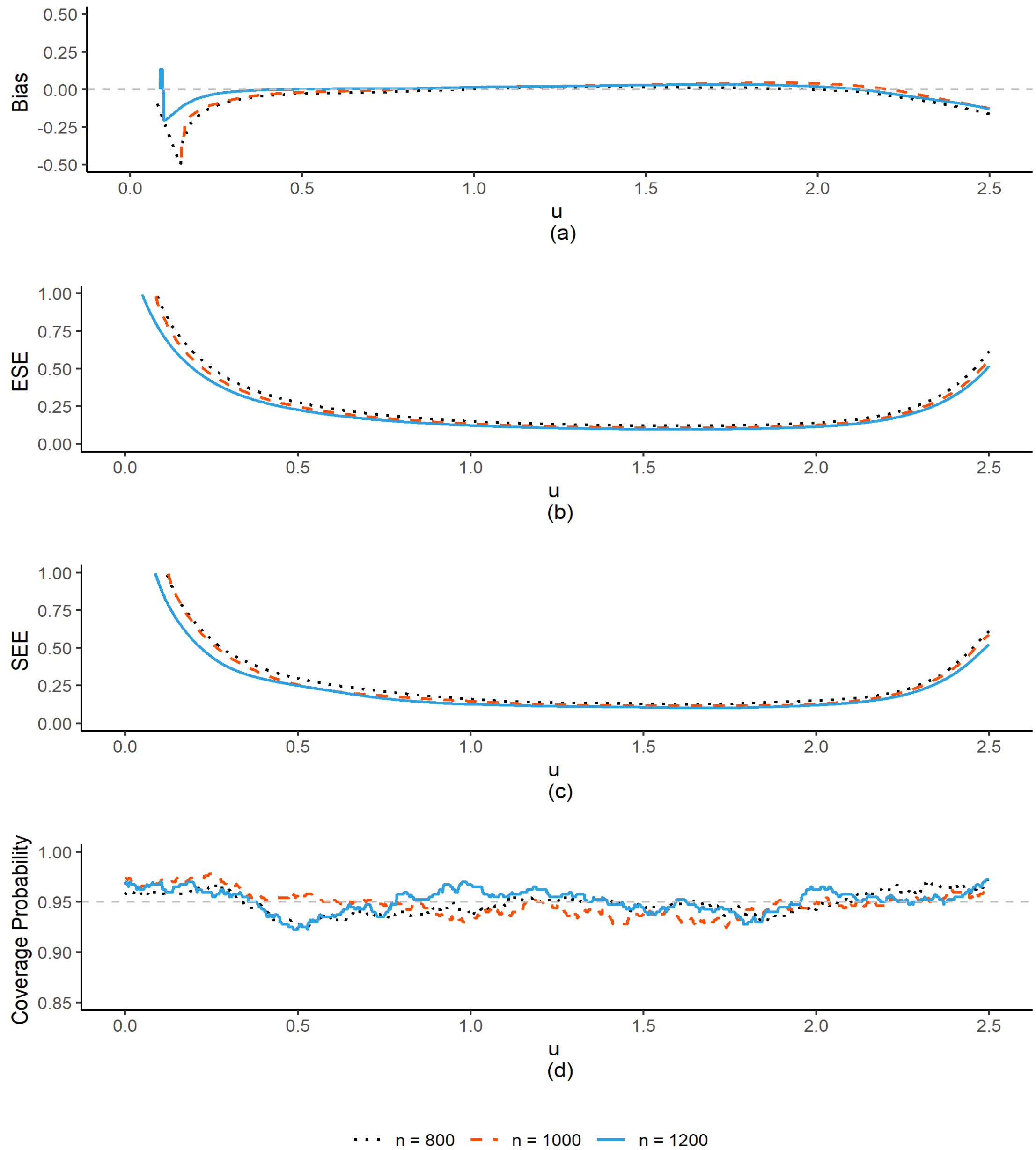


Figure 4.2: Plots of bias, SEE, ESE and CP for $\beta(u)$ with sample size of $n = 800$, $n = 1000$ and $n = 1200$ for Example 1. The number of simulations is 400. The yellow dotted lines are for sample size 800. The red dashed lines are for sample size 1000. The blue solid lines are for sample size 1200. The bandwidth is $h = 0.45$.

Table 4.1: Summary of bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error) and CP (Coverage Probability) for θ_2 for Example 1. Each scenario is based on 400 simulations.

n	h	Bias	SEE	ESE	CP
800	0.25	0.0439	0.3417	0.3812	0.92
	0.35	0.0464	0.4329	0.4599	0.92
	0.45	0.0632	0.4686	0.4875	0.92
1000	0.25	0.0272	0.3037	0.3559	0.93
	0.35	0.0313	0.3676	0.4043	0.94
	0.45	0.0189	0.4358	0.4884	0.94
1200	0.25	0.0119	0.3209	0.3691	0.95
	0.35	0.0109	0.3757	0.3695	0.94
	0.45	0.0280	0.4223	0.4648	0.95

4.2 Example 2: A Mark PH model with Three Marks

In this section, we extend the simulation settings and consider the following multivariate mark-specific proportional hazards model with three marks $V = (V_1, V_2, V_3)^T$ and the same constraint applied to θ.

Let z be a binary covariate taking value 0 or 1 with a given probability 0.5 for each observation and $v = (v_1, v_2, v_3)^T$ with $0 \le v_i \le 1$ for $i = 1, 2$ and 3. The variables (T, V) are generated from the following multivariate mark-specific proportional hazard model:

$$\lambda(t, v|z) = \exp\{\gamma^T v + \beta(\theta^T v)^T z\}, \tag{4.2}$$

where $0 \le t \le \tau$ with $\tau = 2$.

We set $\gamma = (\gamma_1, \gamma_2, \gamma_3)^T = (0.6, 0.4, 0.2)^T$ and $\theta = (\theta_1, \theta_2, \theta_3)^T = (1, 1.5, 2.5)^T$. Under model (4.1), the mark-specific baseline function is $\lambda_0(t, v) = \exp\{\gamma^T v\} = \exp\{0.6v_1 + 0.4v_2 + 0.2v_3\}$. The unknown function $\beta(u)$ is chosen to be a linear function with the form $\beta(u) = -1.65 + 1.0(u)$. We adopt the Epanechnikov kernel $K(x) = 0.75(1 - x^2)I\{|x| \le 1\}$.

We generate the censoring times from an exponential distribution with mean $\lambda = 1$, which yields the censoring rates ranging from 20% to 30%.

The bandwidths are selected as $h = 0.35, 0.45$ or 0.60. Sample sizes of $n = 800, 1000$ or 1200 are studied. We used 400 replicates for each combination of sample size and bandwidth.

We use $n_0 = n$, and choose $r_1 = 2$ and $r_2 = 1.5$ when estimating θ. In the last step (Step L), we set the interval $[a, b]$ to be $[0, 5]$. For the variance estimation of $\widehat{\theta}$, we adopt $b = 10n^{-1/2}$.

From Figure 4.3, the frequency of $\theta_0^T V$ with values less than 1 is very small. Thus, the performance of the estimators is expected to be unstable when $\theta_0^T V$ takes a value less than one.

Figure 4.4 and 4.5 present the evaluation results of the estimator of $\beta(\cdot)$ under different simulation settings. The bias of the parameter estimator is close to zero, and the ESE approximates SEE better as the sample size increases. When bandwidth $h = 0.45$ or 0.60, the 95% coverage probability is about the nominal level. When $h = 0.35$, the empirical coverage probability is slightly below the nominal level of 0.95 but gets close to 0.95 as the sample size increases.

Table 4.2 summarizes the estimation results for both θ_2 and θ_3. The biases of parameter estimators decrease with increasing sample sizes. The means of the estimated standard errors are close to the sample standard errors, and the 95% empirical coverage probabilities are closer to the nominal level at 0.95 as the sample size increases.

We notice that when $n = 800$ and $h = 0.35$, in some replicates, there are not enough data points within the neighborhood of some grid points u, which leads to the large biases of estimators θ and significant differences between SEE and ESE. The coverage probabilities, in this case, are close to 0.95 since this situation only happens a couple of times. The problem goes away when h increases to 0.45 or n increases to 1000.

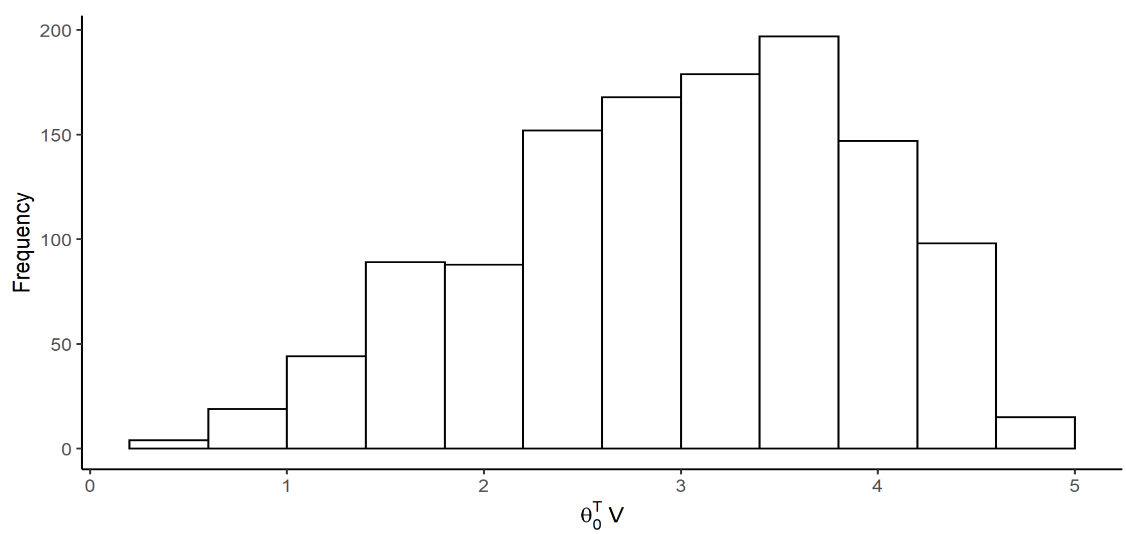


Figure 4.3: Histogram for $\theta_0^T V$ for a single simulation for sample size $n = 1200$.

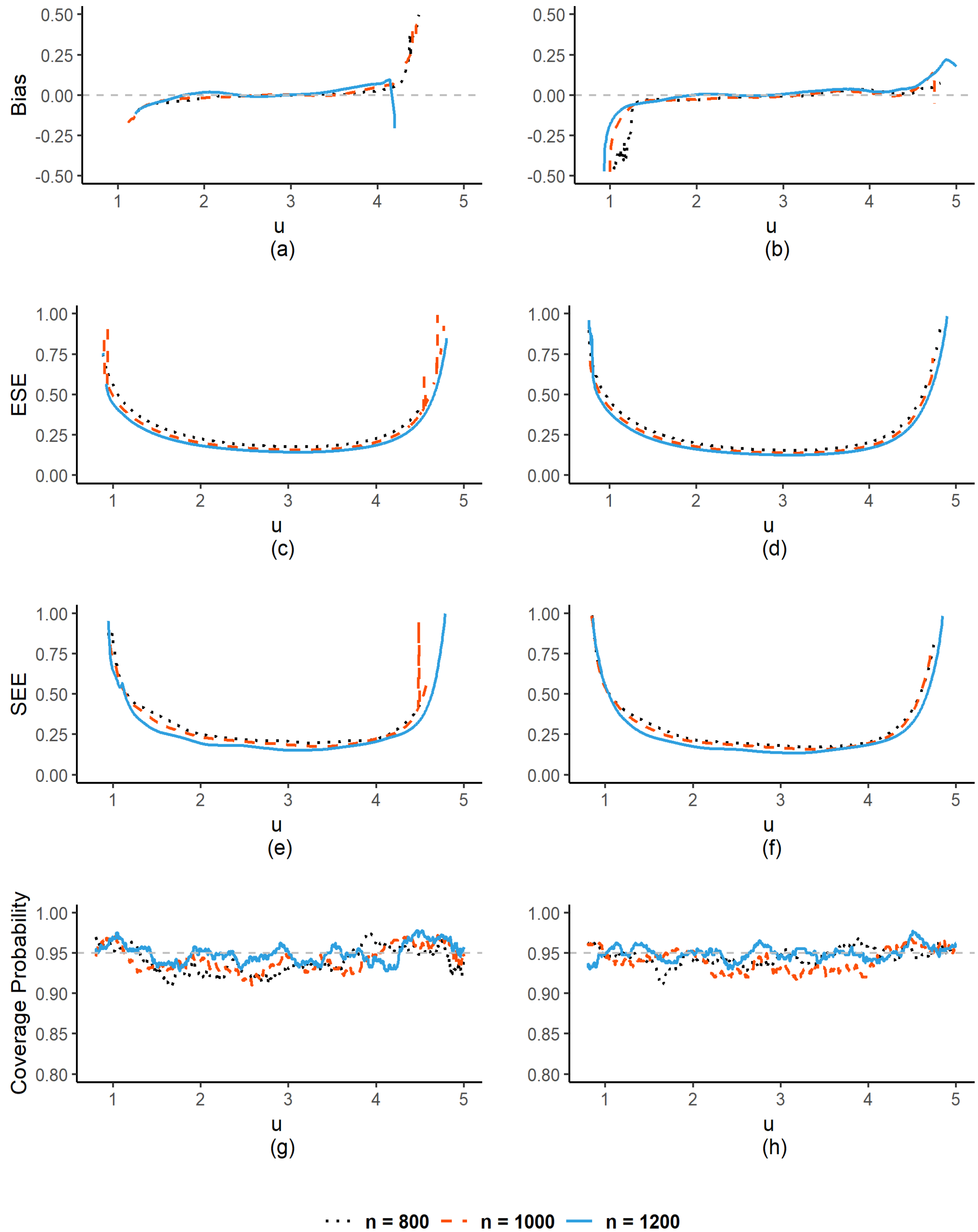


Figure 4.4: Plots of bias, SEE, ESE and CP for $\beta(u)$ with sample size of $n = 800$, $n = 1000$ and $n = 1200$ for Example 2. The number of simulations is 400. The yellow dotted lines are for sample size 800. The red dashed lines are for sample size 1000. The blue solid lines are for sample size 1200. Left panel is for bandwidth $h = 0.45$. Right panel is for bandwidth $h = 0.60$.

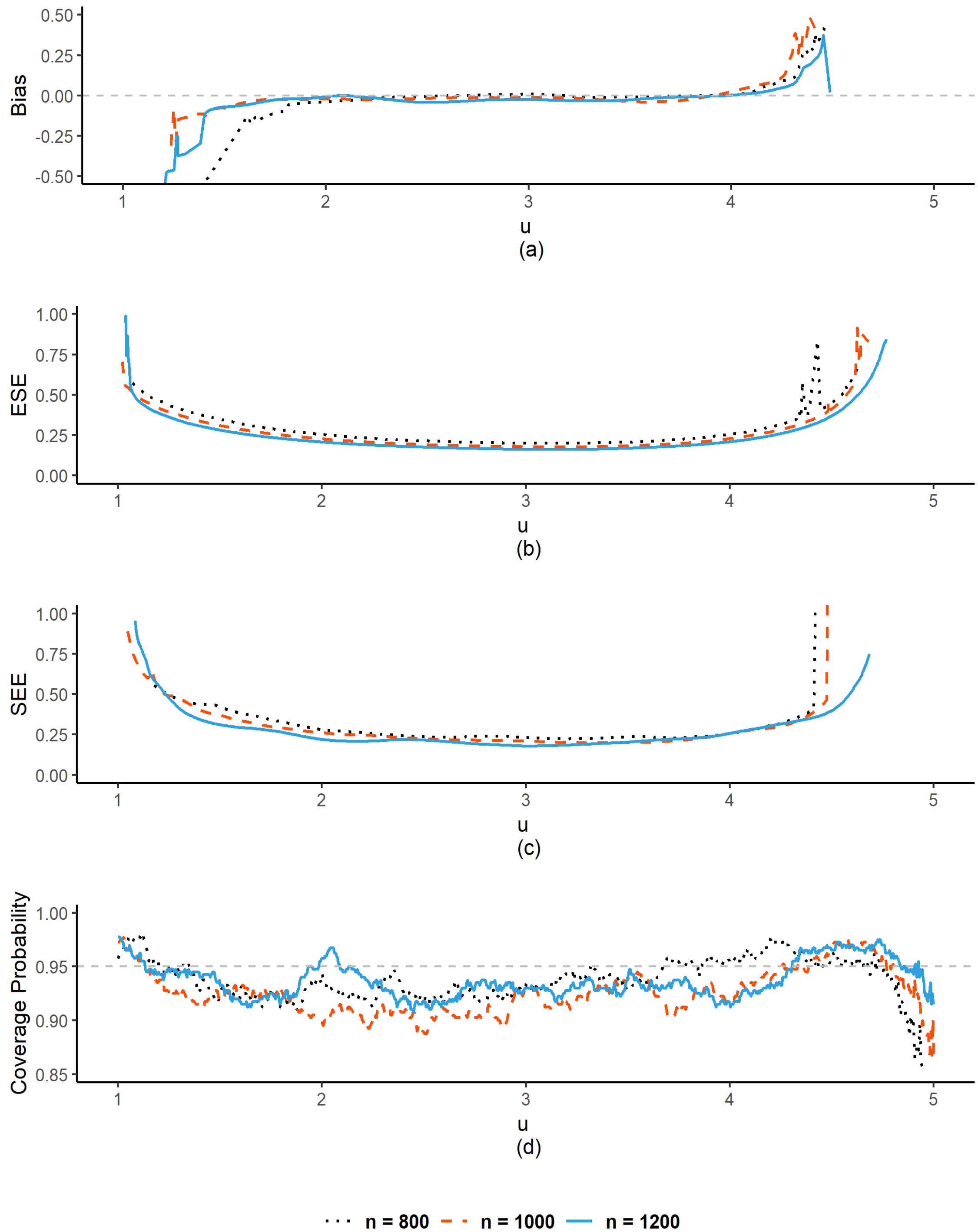


Figure 4.5: Plots of bias, SEE, ESE and CP for $\beta(u)$ with sample size of $n = 800$, $n = 1000$ and $n = 1200$ for Example 2. The number of simulations is 400. The yellow dotted lines are for sample size 800. The red dashed lines are for sample size 1000. The blue solid lines are for sample size 1200. The bandwidth is $h = 0.35$.

Table 4.2: Summary of bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error) and CP (Coverage Probability) for θ_2 and θ_3 for Example 2. Each scenario is based on 400 simulations.

n	h	Parameter	Bias	SEE	ESE	CP
800	0.35	θ_2	-0.0563	1.3894	0.8012	0.96
		θ_3	0.0573	0.8047	0.8039	0.96
	0.45	θ_2	0.0102	0.3881	0.3837	0.92
		θ_3	0.0212	0.4176	0.4254	0.93
	0.60	θ_2	0.0191	0.3806	0.4083	0.94
		θ_3	0.0380	0.4029	0.4831	0.95
1000	0.35	θ_2	-0.0165	0.3536	0.3618	0.95
		θ_3	0.0445	0.4374	0.4787	0.96
	0.45	θ_2	-0.0052	0.3016	0.3542	0.95
		θ_3	0.0255	0.3255	0.4038	0.95
	0.60	θ_2	0.0024	0.3346	0.3534	0.94
		θ_3	0.0403	0.3714	0.4307	0.95
1200	0.35	θ_2	-0.0016	0.2736	0.2843	0.95
		θ_3	0.0447	0.2715	0.3308	0.95
	0.45	θ_2	-0.0073	0.2912	0.3412	0.94
		θ_3	0.0175	0.2988	0.3899	0.95
	0.60	θ_2	-0.0032	0.3085	0.3418	0.95
		θ_3	0.0317	0.3495	0.4130	0.95

CHAPTER 5: DATA APPLICATIONS

In this chapter, we illustrate the proposed model and methods with the applications to two datasets from two HIV vaccine efficacy trials, namely, HVTN 505 DNA/Recombinant adenovirus type 5 (rAd5) vector HIV-1 vaccine trial and STEP/HVTN 502 trial, respectively.

The global priority and urgency of searching for a safe and effective preventive HIV vaccine stem from the fact that approximately 75 million people have been infected by HIV and around 32 million cases have died due to HIV at the end of 2018 since 1981, the beginning of AIDS epidemic (World-Health-Organization (2018)). Although improvements in treatment, care, and prevention methods have been progressing, HIV infection rates and mortality rates remain high. HIV-1 and HIV-2 are two types of HIV that have been characterized. Relatively, the HIV-1 is fatal and more infective than the HIV-2 (Gilbert et al. (2003)). HIV-1 is categorized into three groups, M, N, and O. The majority of HIV-1 is in group M, which consist of different subtypes, such as A, B, and C. Trials have been conducted to date showing no efficacy except for one HIV-1 vaccine regimen, RV 144. It has been indicated to have some effects in preventing HIV infections in Thailand (Rerks-Ngarm et al. (2009)). Two of the primary barriers to producing an effective HIV vaccine include the large number of mutations involved and the high degree of genetic divergence of HIV. As the aforementioned motivating example in Chapter 1, genetic heterogeneity can be measured by mark variables defined as the percentage of mismatching between two aligned amino acid sequences (one infecting HIV sequence and one HIV sequence represented in the vaccine) in different subregions of a protein, or multiple protein sequences. Mark variables are considered to be continuous for the highly mutated property of

HIV.

In both data applications, for each mark, we standardize the mark variable by recalculating it as

$$\frac{V_i - min(V)}{max(V) - min(V)}, \tag{5.1}$$

where V_i represents the value of the mark variable for i^{th} subject in the original dataset. In this case, there is at least one observed value of each mark variable at the endpoints 0 and 1.

HVTN 505 trial is studied in Section 5.1, while STEP trial is introduced in Section 5.2. Each section is organized as follows. First, we introduce the background of the dataset. Second, we conduct a preliminary analysis on the dataset with each single mark variable. In this part, we adopt the mark-specific proportional hazards model with univariate continuous mark proposed in Sun et al. (2009). This model takes the form

$$\lambda(t, v|z(t)) = \lambda_0(t, v)\exp\left\{\beta(v)^T z(t)\right\}, \tag{5.2}$$

where the baseline hazard function $\lambda_0(\cdot, v)$ depends nonparametrically on t and v, and the p-dimensional regression parameter $\beta(v)$ is unknown continuous function of v.

Then, various models with multiple marks are studied. Finally, we interpret and discuss our findings on the application.

5.1 HVTN 505 Trial Analysis

HVTN 505 DNA/Recombinant adenovirus type 5 (rAd5) vector HIV-1 vaccine efficacy trial was carried out at 21 sites in 19 cities in the United States and enrolled 2504 HIV-negative, fully circumcised men or male-to-female (MTF) transgender people who have a male sexual partner(s). Other criteria see (deCamp et al. (2017)) for details. A total of 2496 volunteers were randomized to receive either the DNA/Recombinant adenovirus type 5 (rAd5) vaccine or placebo on Days 0, 28, 56,

Table 5.1: Summary of three selected marks defined by using three sets of Env-gp120 sites for HVTN 505 trial.

Mark	number of amino acid positions	Site
V_1	93	CD4bs antibody contract site
V_2	54	CD4bs k-mer site
V_3	432	Env-gp 120 site

and 168. Among 2496 participants, 47 acquired HIV infection. Specifically, there were 27 out of 1251 vaccine recipients and 20 out of 1245 placebo recipients with the annual HIV incidences of 2.2% for the vaccine group and 1.6% for the placebo group.

In the HVTN 505 trial, the vaccine contained three HIV-1 gp120 strains, labeled VRC-A, VRC-B, and VRC-C, from a subtype A, B, and C strain, respectively. Since this trial was conducted in the U.S., where subtype B viruses circulate, it is believed that subtype B is the closest to the infecting strains. Hence, we focus on the HIV-1 gp120 strain from subtype B in the study. Motivated by the sieve analysis in (deCamp et al. (2017)), we consider distances defined by using three sets of Env-gp120 sites. Let M be the number of amino acid positions in the portion of the gp120 protein used in each set. Three sets consist of CD4bs antibody contract site ($M = 93$), CD4bs k-mer site ($M = 54$) and all Env-gp 120 sites ($M = 432$). Table 5.1 lists the details of three marks. Our study begins with an analysis of each variable mark. Then, we move on to the multiple marks analysis with one or two covariates.

5.1.1 Univariate Mark Analysis

In this section, we analyze each mark variable introduced in Table 5.1 under model (5.2), where the covariate z is the treatment indicator taking value zero for the placebo group and one for the vaccine group.

Sun et al. (2020) examined a bandwidth selection model using

$$h = C\widehat{\sigma}_v n_0^{-1/3}, \tag{5.3}$$

where C is a constant between 2 and 5, $\widehat{\sigma}_v$ is the estimated standard deviation of the observed marks, and n_0 is the number of the observed failures. Using $C = 5$ yields $h = 0.41$. Our analysis fits the model with bandwidth $h = 0.40$, and the results are shown in Figure 5.1. The left panel of Figure 5.1 displays the boxplots of each mark variable grouped by the treatment indicator. The plots of the estimated $\beta(u)$ for each mark are shown in the right panel of Figure 5.1. Overall, the logarithm of the hazard ratio ($\widehat{\beta}(u)$) increases as the distance between HIV sequences and the subtype B insert for each set of Env-gp120 site increases. The logarithm of the hazard ratio is negative (positive vaccine efficacy) when distances between the infecting HIV sequences and the subtype B vaccine insert are small. However, the evidence does not seem to be very strong. The logarithm of the hazard ratio is approximately 0 (zero vaccine efficacy) when the distance is around 0.22 for mark v_1, 0.19 for mark v_2 and around 0.20 for mark v_3, and is significantly positive (taking risks from vaccination) when the infecting HIV sequences are highly varying from the vaccine. The results are consistent with the sieve analysis on the relationship between Env-gp120 vaccine similarity and genotype-specific vaccine efficacy (deCamp et al. (2017)).

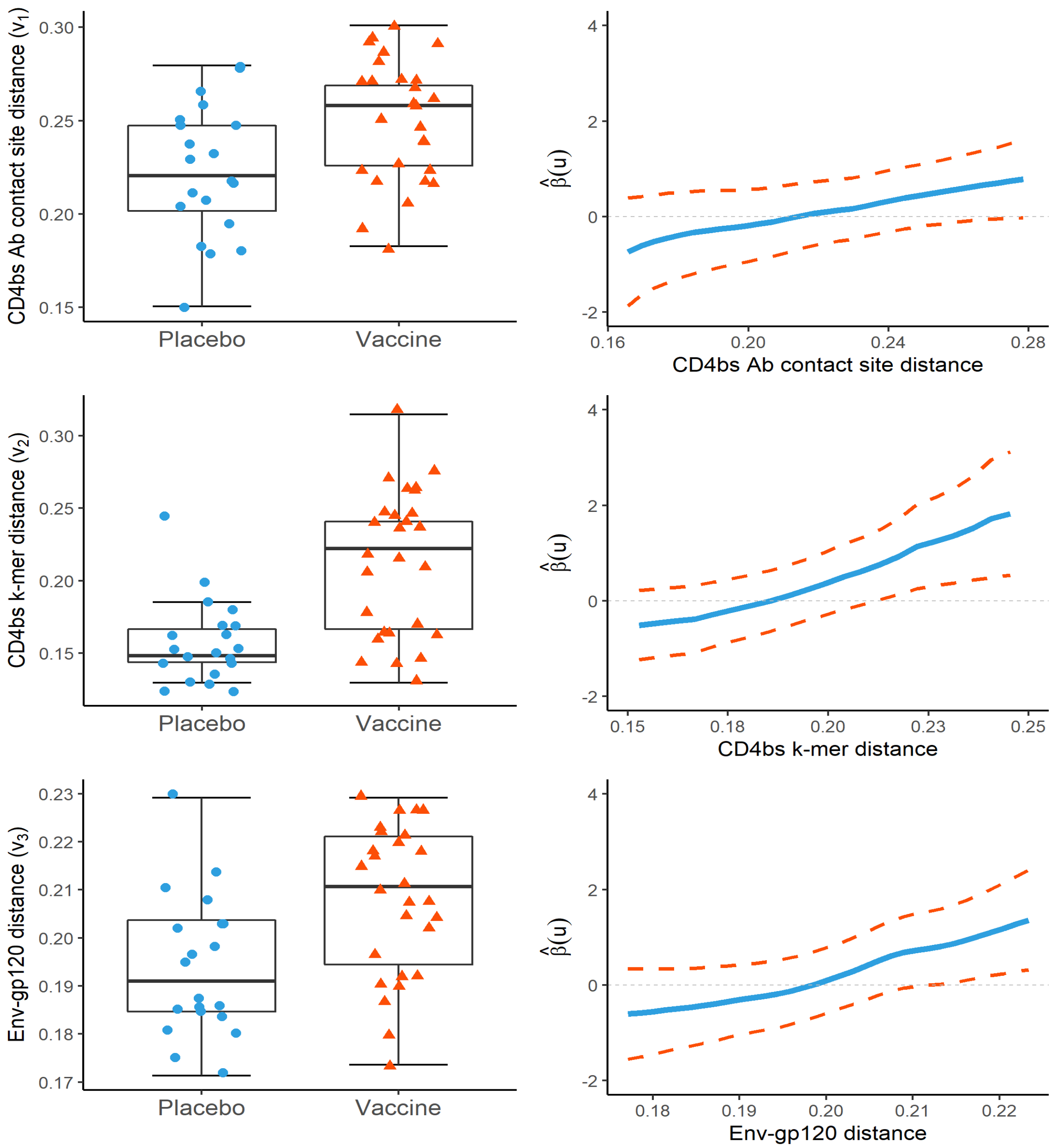


Figure 5.1: Plots of the analysis on each mark variable. Left panel is for the boxplots of single marks grouped by treatment indicator. Right panel shows the plots of the estimated regression coefficient $\beta(u)$ for each mark and its corresponding 95% confidence band with bandwidth $h = 0.40$.

5.1.2 Multivariate Marks Analysis with One Covariate

In this section, we perform the data analysis under the proposed mark-specific proportional hazards model (2.1) with all three marks, and the treatment indicator z as the covariate. The bandwidth is selected as $h = 0.40$.

In this model, we force the constraint on the first parameter of θ to be 1. Hence, it is natural to order the mark variable with the decreasing effects on the hazard ratios. From the right panel of Figure 5.1, we can see that as the value of mark variable v_2 changes in one unit, the logarithm of the hazard ratio changes most significantly (which corresponds to the largest slope). Thus, we set v_2 to be the first component of mark variable v. Specifically, $\theta^T v$ takes the form of $\theta_1 v_2 + \theta_2 v_3 + \theta_3 v_1$. The estimates $(\widehat{\theta}_2, \widehat{\theta}_3)$ are $(0.6775, 0.1431)$ with standard errors $(0.3335, 0.1505)$. This suggests that mark v_1 have no significant effects on infection against HIV-1.

The top plot in Figure 5.2 describes the distributions of the single-index $(\widehat{\theta}^T v)$ for both the placebo group and vaccine group through their quartiles. The plot of the estimated regression coefficients $\beta(\cdot)$ in Figure 5.2 shows that the logarithm of the hazard ratio is negative (positive vaccine efficacy) when u is less than 0.75 and is significantly positive (taking risks from vaccination) when u is greater than 1. This reveals that the larger value of the combination of marks v_2 and v_3 may make the vaccine effects go in the opposite direction. In other words, the further genetic distance increases the risk of taking the HIV vaccine.

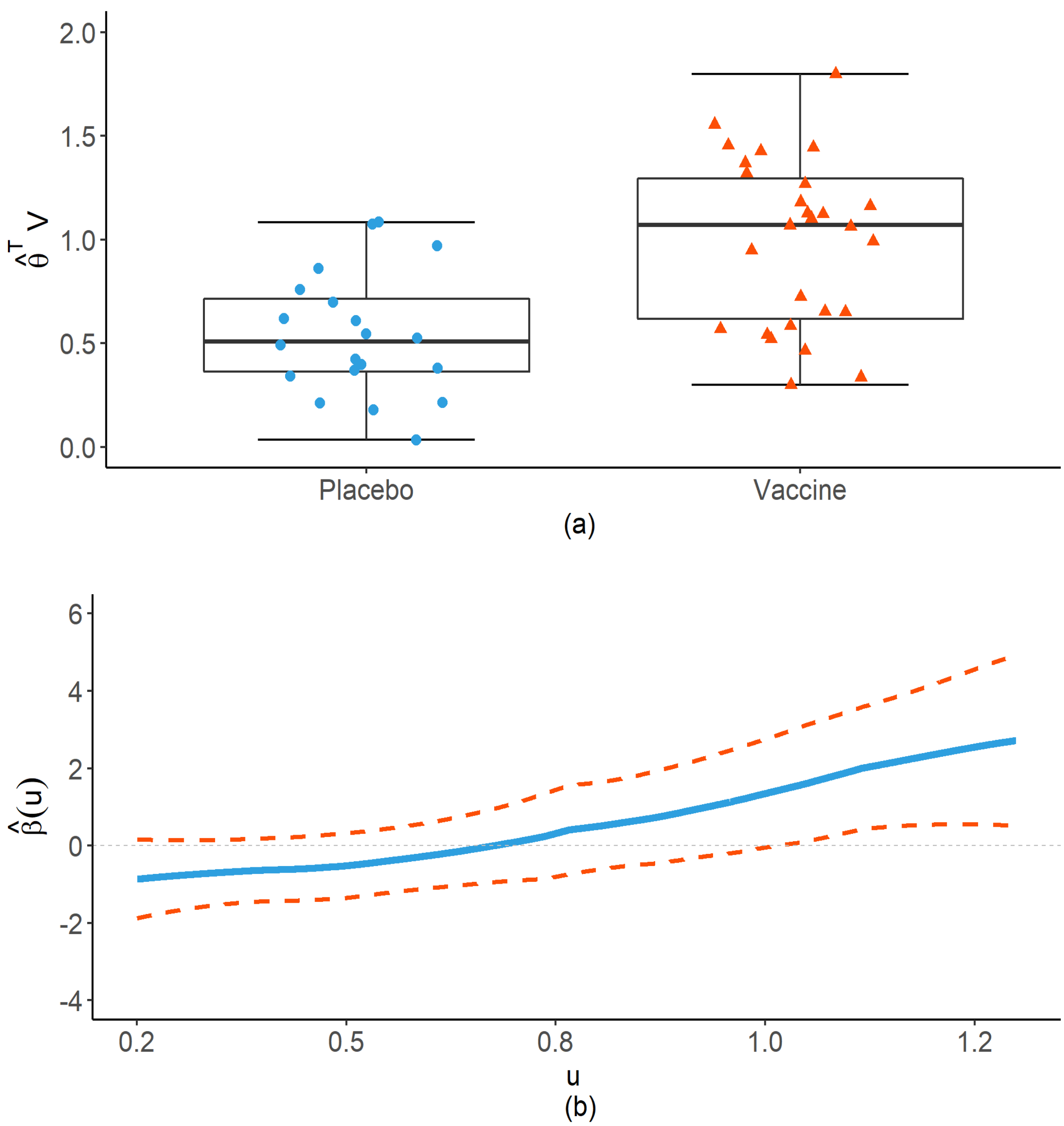


Figure 5.2: Plots of analysis on three marks (v_2, v_3, v_1). Top graph shows the boxplots of $\widehat{\theta}^T v$ grouped by treatment indicator. Bottom graph shows the plot of the estimated regression coefficient $\beta(u)$ and its corresponding 95% pointwise confidence band with bandwidth $h = 0.40$.

5.1.3 Multivariate Marks Analysis with Two Covariates

Inspired by the data application in (Sun et al. (2009)), we extend the model in the previous section by introducing an additional continuous covariate, behavior risk score. It is expected that different individuals with different behavior risk scores may be HIV exposure related to various distributions of $\widehat{\theta}^T v$. In this dataset, behavior risk takes values 0, 0.46, 0.54 and 1. We treat it as a continuous variable as recommended by the experts since each value has its meaning. We set $z = (z_1, z_2)^T$, where z_1 is the treatment indicator and z_2 is the behavioral risk score. Then, $\beta(\theta^T v)^T z = \beta_1(\theta^T v)^T z_1 + \beta_2(\theta^T v)^T z_2$, where $\theta^T v = \theta_1 v_2 + \theta_2 v_3 + \theta_3 v_1$. We use bandwidth $h = 0.40$. The estimates $(\widehat{\theta}_2, \widehat{\theta}_3)$ are $(0.7527, 0.2678)$ with standard errors $(0.2887, 1.1539)$. We can see that the coefficient estimates of mark v_1 is not significant.

The left plot of the bottom panel ($\beta_1(\cdot)$) suggests that after adjusting for covariate behavioral risk score, the vaccine has a positive effect against HIV infection for $u \leq 0.50$ and has no impact with $u > 0.90$. This phenomenon has been observed in Sun et al. (2020).

The plot of estimates for $\beta_2(\cdot)$ in Figure 5.3 is significantly above zero when u is between 0.30 and 1.20, supporting that higher behavioral risk score increases the risk of HIV infection.

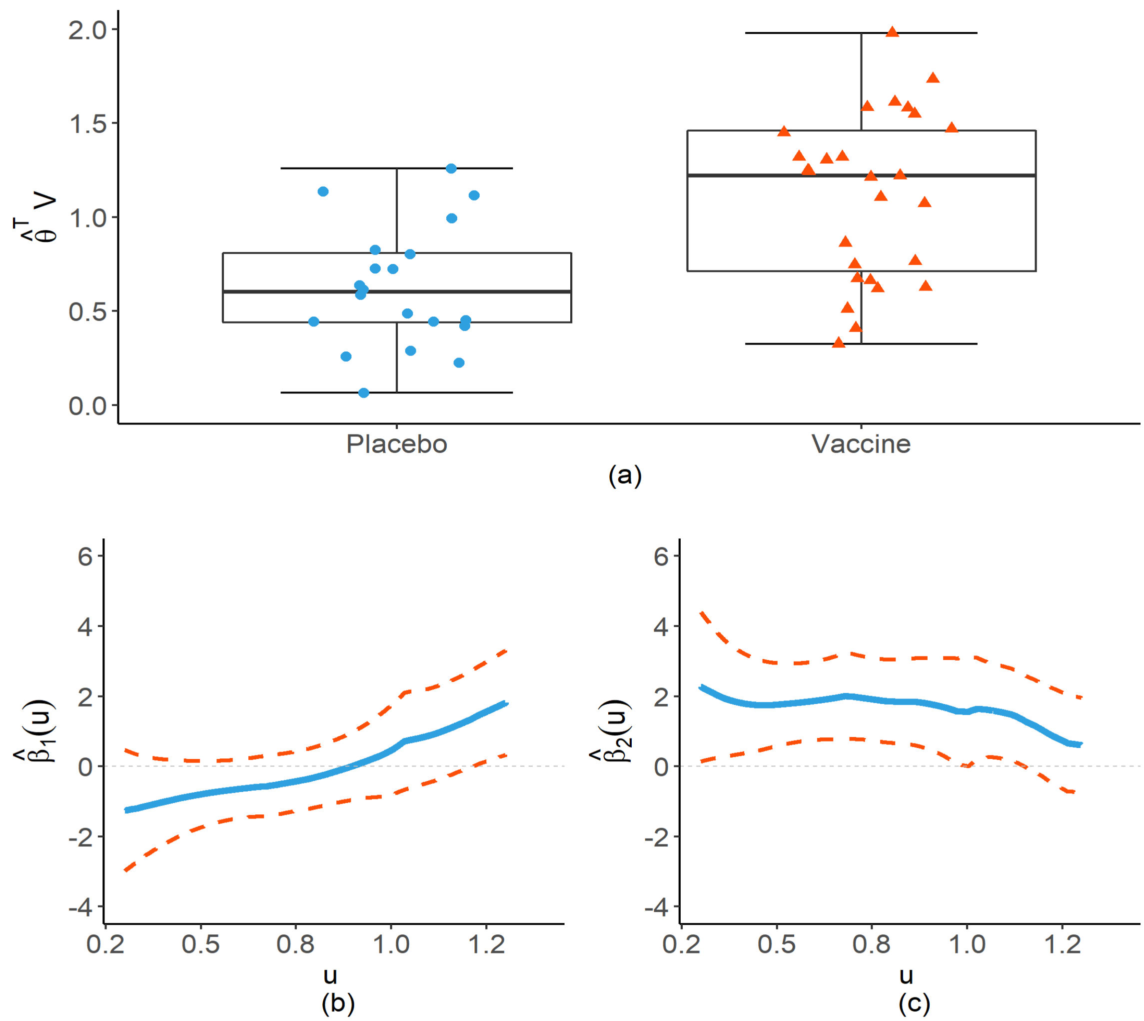


Figure 5.3: Plots of analysis on three marks (v_2, v_3, v_1) with two covariates (z_1, z_2). Top graph shows the boxplots of $\widehat{\theta}^T v$ grouped by treatment indicator. Bottom graphs depict the plots of the estimated regression coefficients $\beta_1(u)$ and $\beta_2(u)$, and their corresponding 95% confidence band with bandwidth $h = 0.40$.

5.2 STEP Trial Analysis

STEP trial was the second preventive HIV vaccine efficacy trial conducted in the Americas, Australia, and the Caribbean. 1836 HIV seronegative men were randomized to receive the MRK Ad5 gag/pol/nef vaccine or placebo. There were 87 out of 1836 HIV infections: 53 out of 914 vaccine recipients (5.8% annual HIV incidence) and 34 out of 922 placebo recipients (3.7% annual HIV incidence). The sequencing lab tried to derive HIV sequences from all 87 infected subjects through single-genome-amplification, but only succeed in 65 of them. We excluded the 22 men without sequence data.

MRK Ad5 gag/pol/nef vaccine contained three HIV-1 genes: Gag, Pol, and Nef. For each protein, it is believed that different genetic distances to the protein could make various vaccine effects on HIV infections. For control purposes, we use the central HXB2 as the reference strain for calculating the genetic distances to Env-Rev-Tat-Vif-Vpr-Vpu that is not contained in the vaccine, called control marks. Since the protein Env-Rev-Tat-Vif-Vpr-Vpu should not be able to trigger immune reactions, it is anticipated that the effect on HIV infection should not relate to the genetic distance to the control protein. Besides, two different bioinformatics methods, NetMHC (Buus et al. (2003)) and Epipred (Heckerman et al. (2007)), were used to evaluate the genetic distances for each infected individual. See (Sun et al. (2013)) for detailed introductions to these two methods.

In our study, we include the genetic distances to the proteins Gag, Pol, Nef, and Env-Rev-Tat-Vif-Vpr-Vpu based on two bioinformatics methods. Table 5.2 lists the details of the five selected marks. The covariate z is the treatment group indicator with $z = 1$ for the vaccine group and $z = 0$ for the placebo group.

Table 5.2: Summary of five selected marks for STEP trial.

Mark	Method	Reference	Protein
V_1	Epipred	HXB2	Env-Rev-Tat-Vif-Vpr-Vpu
V_2	NetMHC	HXB2	Env-Rev-Tat-Vif-Vpr-Vpu
V_3	Epipred	vaccine	Gag
V_4	Epipred	vaccine	Pol
V_5	Epipred	vaccine	Nef

5.2.1 Univariate Mark Analysis

We study all five marks described in Table 5.2 individually based on the model (5.2), where covariate z is the indicator of the treatment group taking value 1 for the vaccine group and 0 for the placebo group. We adopt the bandwidth selection formula (5.3) with constant $C = 5$, yielding $h = 0.40$. The results are given in Figures 5.4 and 5.5.

Figure 5.4 demonstrates that the vaccine efficacy would not be divergent with the control genetic distances. Figure 5.5 depicts that vaccine efficacy would not be divergent with the genetic distances to protein, Pol. The function $\beta(\cdot)$ is significantly positive when the distance is between 0.40 and 0.53. For marks variable v_3 and v_5, the functions $\beta(\cdot)$ are monotone increasing and are significantly positive (taking risks from vaccination) when the genetic distance to protein Gap is greater than 0.47, and the genetic distance to Nef is greater than 0.72.

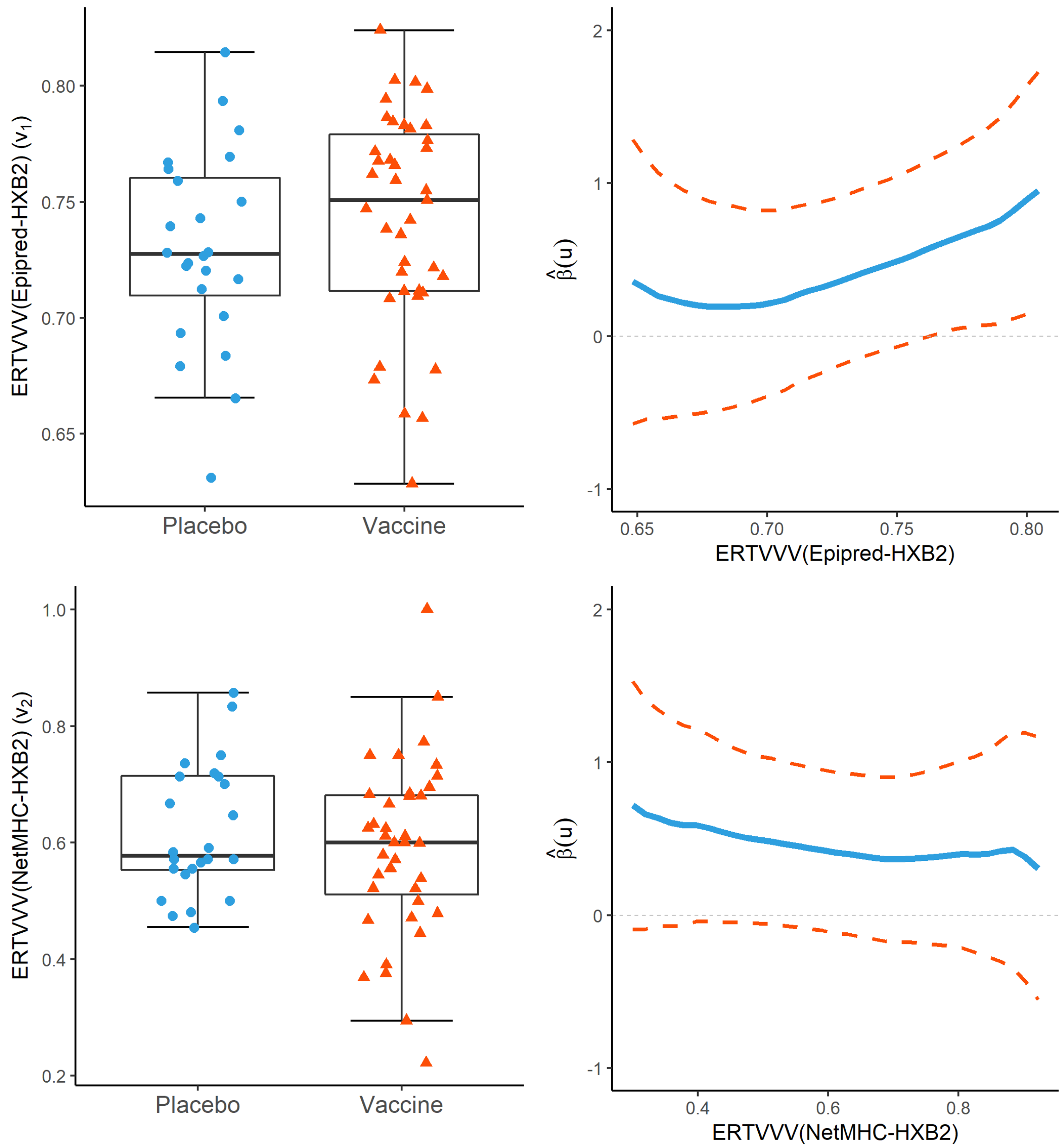


Figure 5.4: Plots for the analysis on each control mark (v_1 and v_2). Left panel shows the boxplots of the control marks grouped by treatment indicator. Right panel shows the plots of the estimated regression coefficient $\beta(u)$ for each control mark and its corresponding 95% confidence band with bandwidth $h = 0.40$.

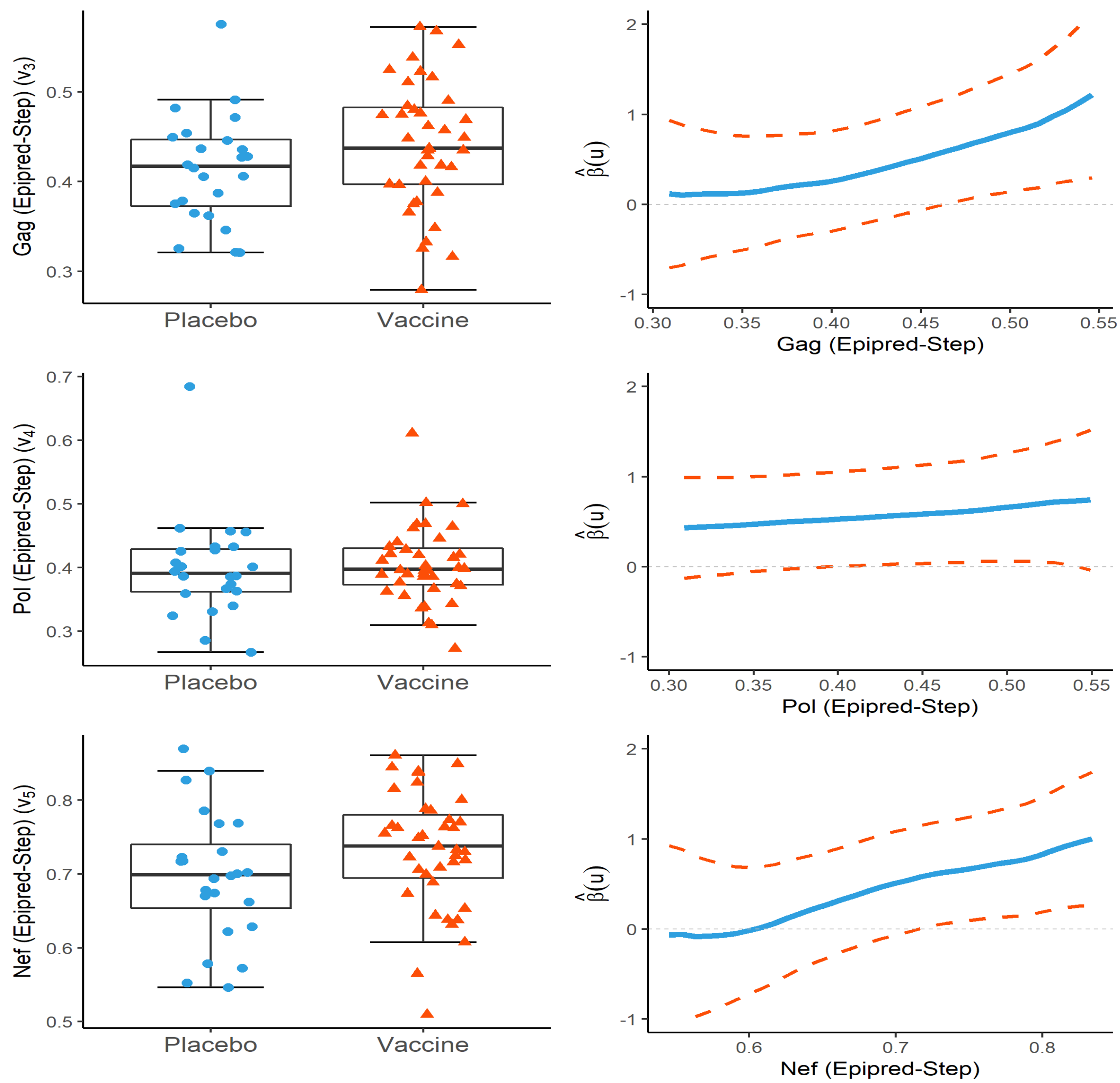


Figure 5.5: Plots for the analysis on single marks (v_3, v_4 and v_5). Left panel shows the boxplots of single marks grouped by treatment indicator. Right panel shows the plots of the estimated regression coefficient $\beta(u)$ for each mark and its corresponding 95% confidence band with bandwidth $h = 0.40$.

5.2.2 Multivariate Marks Analysis

In the analysis of multiple marks in STEP trial, we select the treatment indicator z as the covariate and design the study as follows. First, we divide five marks into two groups, A and B, and conduct the analysis on each group. Group A contains control marks v_1 and v_2, while group B contains marks v_3, v_4 and v_5. Second, we fit the mark-specific PH model with all five marks together. To determine the order of marks in each study, we apply the same rule as we introduced in Section 5.1.2.

First, for group A analysis, we fit the model with $\theta^T v = \theta_1 v_1 + \theta_2 v_2$, and bandwidth is selected as $h = 0.40$. Setting $\theta_1 = 1$, the estimate $\widehat{\theta_2}$ is -0.4890, and the corresponding standard error is 0.1962. The boxplot in Figure 5.6 describes the distributions of the single-index $(\widehat{\theta}^T v)$ for both the placebo group and the vaccine group are close except for the slightly wider range of the distribution for placebo group. $\beta(\cdot)$ function shows the vaccine effect on HIV infection does not vary with the combination of the control marks.

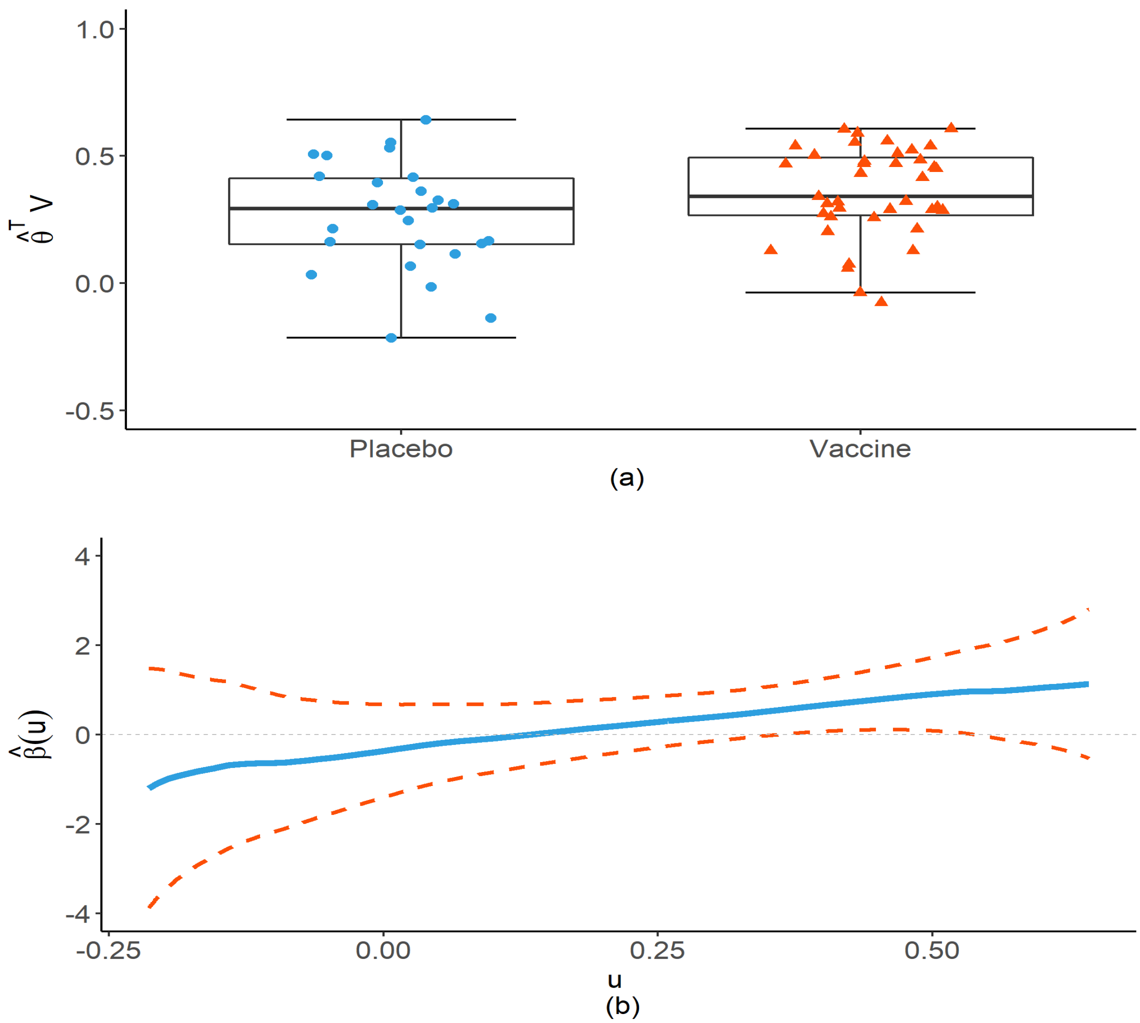


Figure 5.6: Plots for the analysis on bivariate marks (v_1, v_2). Top graph shows the boxplots of $\widehat{\theta}^T v$ grouped by treatment indicator. Bottom graph shows the plot of the estimated regression coefficient $\beta(u)$ and its corresponding 95% pointwise confidence band with bandwidth $h = 0.40$.

Second, to analyze the marks in group B, we apply the mark-specific PH model with bandwidth $h = 0.60$. $\theta^T v$ takes the form of $\theta_1^T v_3 + \theta_2^T v_5 + \theta_3^T v_4$. Setting $\theta_1 = 1$, the estimates $(\widehat{\theta}_2, \widehat{\theta}_3)$ are $(0.9430, 0.3088)$ and the standard errors are $(0.3900, 3.2917)$, implying that v_4 is not significantly related to the hazard ratio. The positive coefficients give that larger values of the combination of marks in group B increase the risk of infection. The 95% confidence band of the monotone increasing function $\beta(\cdot)$ in Figure 5.7 displays that $\beta(u)$ is significantly positive when u is between 1.30 and 1.70.

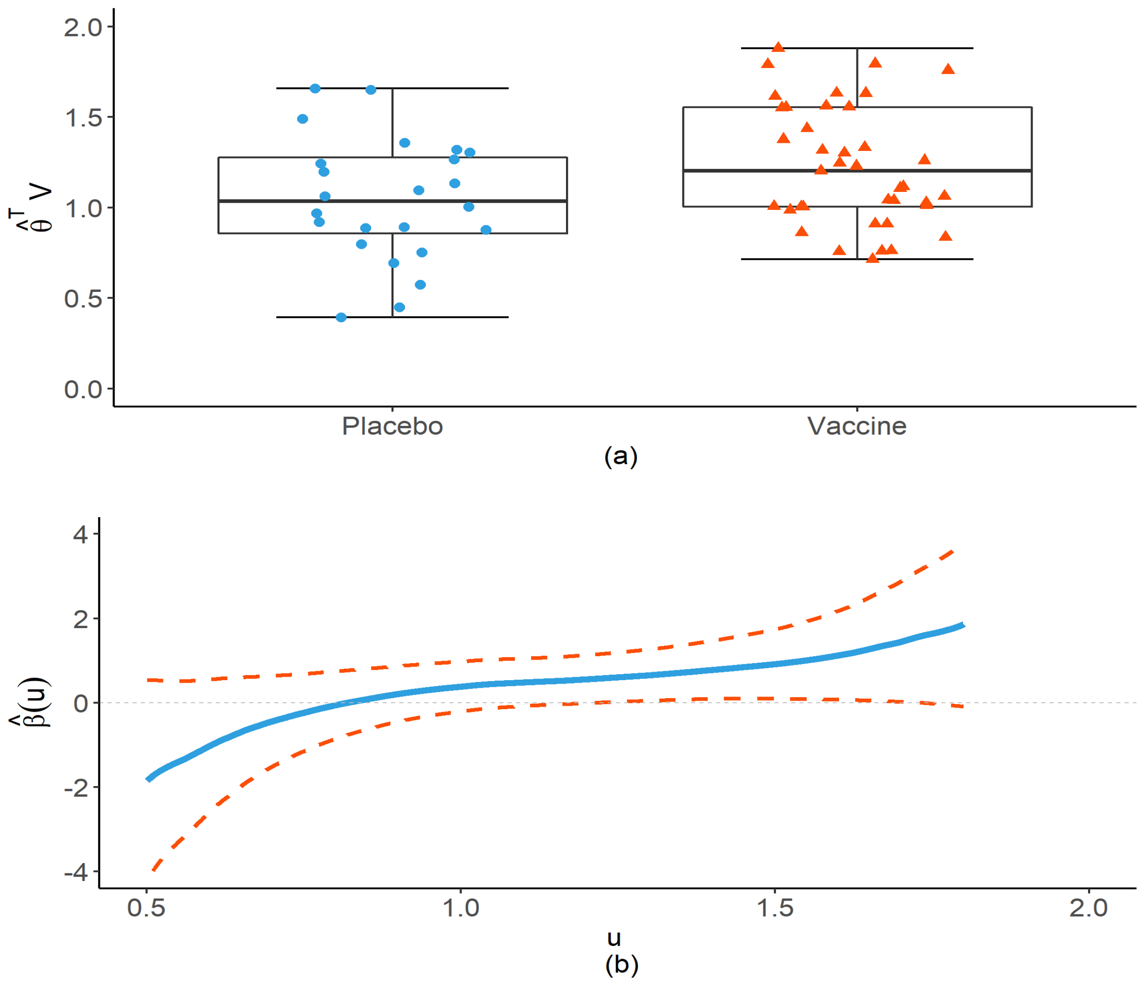
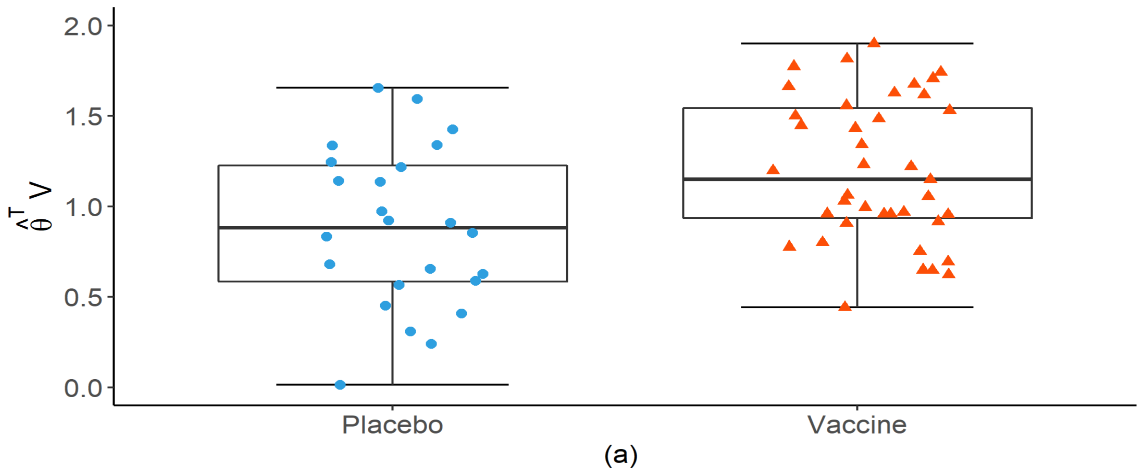


Figure 5.7: Plots for the analysis on three marks (v_3, v_5, v_4). Top graph shows the boxplots of $\widehat{\theta}^T v$ grouped by treatment indicator. Bottom graph shows the plot of the estimated regression coefficient $\beta(u)$ and its corresponding 95% pointwise confidence band with bandwidth $h = 0.60$.

Finally, we combine all five marks listed in Table 5.2 and adopt the model with $\theta^T v = \theta_1^T v_3 + \theta_2^T v_1 + \theta_3^T v_5 + \theta_4^T v_2 + \theta_5^T v_4$. The bandwidth is selected as $h = 0.70$. Setting $\theta_1 = 1$, the estimates $(\widehat{\theta}_2, \widehat{\theta}_3, \widehat{\theta}_4, \widehat{\theta}_5)$ are $(0.9840, 1.0513, -1.4232, 0.3077)$, and corresponding standard errors are $(0.3797, 0.3530, 0.1498, 1.3688)$. The monotone increasing function $\beta(\cdot)$ in Figure 5.8 shows that individuals take risks from vaccination when u is greater than 1.20. Specifically, $\theta^T v$ with larger values of mark variables v_3, v_1 and v_5, and smaller values of mark variable v_2 have higher risks of HIV infection than that with smaller values of mark variables v_3, v_1 and v_5, and larger values of mark v_2.

Our analysis shows that the Step vaccine does not provide protection against HIV infection. Perhaps, it makes the vaccinees more susceptible to infection. This is consistent with the finding of the trial in Sekaly (2008).



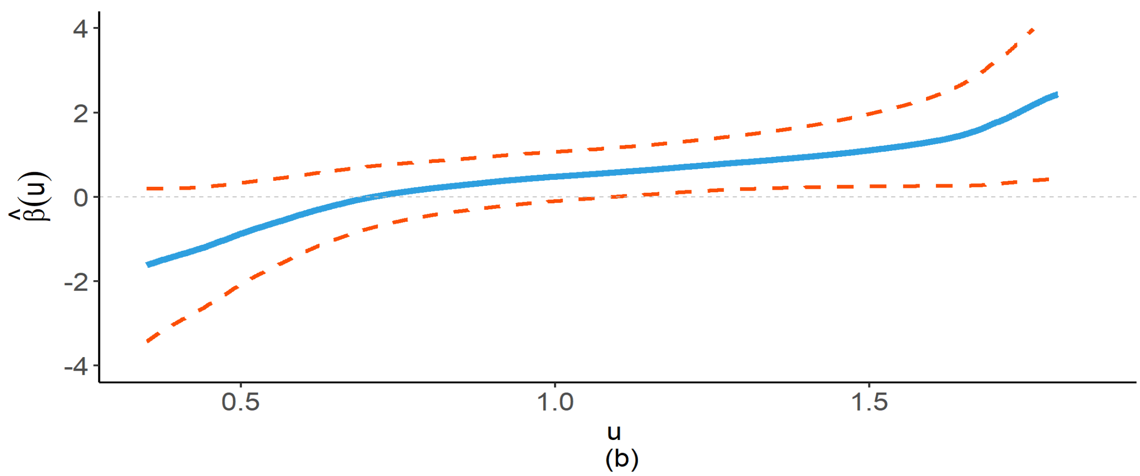


Figure 5.8: Plots for the analysis on all five marks $(v_3, v_1, v_5, v_2, v_4)$. Top graph shows the boxplots of $\widehat{\theta}^T v$ grouped by treatment indicator. Bottom graph shows the plot of the estimated regression coefficient $\beta(u)$ and its corresponding 95% pointwise confidence band with bandwidth $h = 0.70$.

REFERENCES

S. Buus, S. L. Lauemoller, P. Worning, C. Kesmir, T. Frimurer, S. Corbet, A. Fomsgaard, J. Hilden, A. Holm, and S. Brunak. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue Antigens*, 62(5):378–384, 2003.

D. R. Cox. Regression Models and Life-Tables. *Journal of the Royal Statistical Society*, 34(2):187–220, 1972.

A. C. deCamp, M. Rolland, P. T. Edlefsen, E. Sanders-Buell, B. Hall, C. A. Magaret, and et al. Sieve Analysis of Breakthrough HIV-1 Sequences in HVTN 505 Identifies Vaccine Pressure Targeting the CD4 Binding Site of Env-gp120. *PLoS ONE*, 12 (11), 2017.

J. Fan, H. Lin, and Y. Zhou. Local partial-likelihood estimation for lifetime data. *Ann. Statist*, 34(1):290–325, 2006.

P. B. Gilbert, I. W. McKeague, G. Eisen, C. Mullins, A. Gueye-NDiaye, S. Mboup, and P. Kanku. Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegai. *Statistics in Medicine*, 22(4):573–93, 2003.

P. B. Gilbert, I. W. McKeague, and Y. Sun. Tests for comparing mark-specific hazards and cumulative incidence functions. *Lifetime Data Analysis*, 10:5–28, 2004.

P. B. Gilbert, I. W. McKeague, and Y. Sun. The two-sample problem for failure rates depending on a continuous mark: An application to vaccine efficacy. *Biostatistics*, 9(2):263–76, 2008.

B. E. Hansen. Uniform Convergence Rates for Kernel Estimation with Dependent Data. *Econometric Theory*, 24:726–748, 2008.

W. Hardle and T. M. Stoker. Investigating Smooth Multiple Regression by the Method of Average Derivatives. *Journal of the American Statistical Association*, 84(408):986–995, 1989.

D. Heckerman, C. Kadie, and J. Listgarten. Leveraging information across HLA alleles/supertypes improves epitope prediction. *Journal of Computational Biology*, 14(6):736–746, 2007.

J. Z. Huang and L. Liu. Polynomial Spline Estimation and Inference of Proportional Hazards Regression Models with Flexible Relative Risk Form. *Biometrics*, 62(3): 793–802, 2006.

Y. Huang and T. Louis. Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika*, 85(4):785–798, 1998.

J. D. Kalbfeisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, 1980.

H. Lin, M. T. Tan, and Y. Li. A semiparametrically efficient estimator of single-index varying coefficient cox proportional hazards models. *Statistics Sinica*, 26:779–807, 2016.

X. Lu, G. Chen, X. K. Song, and R. Singh. A Class of Partially Linear Single-Index Survival Models. *The Canadian Journal of Statistics*, 34(1):97–112, 2006.

M. OLSCHEWSKI and M. SCHUMACHER. Statistical analysis of quality of life data in cancer clinical trials. *Statistics in Medicine*, 9(7):749–63, 1990.

D. Pollard. *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics, 1990.

R. L. Prentice, J. D. Kalbfeisch, A. V. Peterson, N. Fluornoy, V. T. Farewell, and

N. E. Breslow. The analysis of failure time in the presence of competing risks. *Biometrics*, 34:541–554, 1978.

S. Rerks-Ngarm, P. Pitisuttithum, S. Nitayaphan, J. Kaewkungwal, J. Chiu, R. Paris, and et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N Engl J Med.*, 361(23):2209–20, 2009.

H. Royden and P. Fitzpatrick. *Real Analysis*. Pearson Education, Inc, 2010.

T. Scheike, M. Zhang, and T. A. Gerds. Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95(1):205–220, 2008.

R. P. Sekaly. The Failed HIV Merck Vaccine Study: A Step Back or a Launching Point for Future Vaccine Development? *The Journal of Experimental Medicine*, 205(1):7–12, 2008.

J. Sun, K. A. Kopciuk, and X. Lu. Polynomial Spine Estimation of Partially Linear Single-Index Proportional Hazards Regression Models. *Computational Statistics and Data Analysis*, 53(1):176–188, 2008.

Y. Sun. Generalized nonparametric test procedures for comparing multiple cause-specific hazard rates. *Journal of Nonparametric Statistics*, 13(13):171–207, 2001.

Y. Sun and P. B. Gilbert. Estimation of Stratified Mark-Specific Proportional Hazards Models with Missing Marks. *Scandinavian Journal of Statistics*, 39(1):34–52, 2012.

Y. Sun, P. B. Gilbert, and I. W. Mckeague. Proportional hazards models with continuous marks. *The Annuals of Statistics*, 37(1):394–426, 2009.

Y. Sun, M. Li, and P. B. Gilbert. Mark-specific proportional hazards model with multivariate continuous marks and its application to HIV vaccine efficacay trials. *Biostatistics*, 14(1):60–74, 2013.

Y. Sun, M. Li, and P. B. Gilbert. Goodness-of-fit Test of the stratified Mark-Specific Proportional Hazards Model With Continuous Mark. *Comput Stat Data Anal.*, 93: 348–358, 2014.

Y. Sun, L. Qi, F. Heng, and P. B. Gilbert. A hybrid approach for the stratified mark-specific proportional hazards model with missing covariates and missing marks, with application to vaccine efficacy trials. *Journal of the Royal Statistical Society*, 2020.

W. Wang. Proportional Hazards Regression with Unknown Link Function and Applications To Time-to-event Longitudinal Data. *Statistica Sinica*, 14:885–905, 2004.

World-Health-Organization. Global Health Observatory(GHO) data - HIV/AIDS. 2018. URL `https://www.who.int/gho/hiv/en/`.

D. Zeng, L. Mao, and D. Lin. Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika*, 103(2):253–271, 2016.

APPENDIX A: PROOFS OF THE THEOREMS

In addition to the notations defined in Section 3.1, for $j = 0, 1$ and 2, we also denote

$$S_n^{(j)}(t, v; w_1, w_2, u, \theta) = \frac{1}{n}\sum_{k=1}^{n} Y_k(t)\exp\{\widetilde{w}(u)^T \widetilde{Z}_k(u, \theta^T v)\}\widetilde{Z}_k(u, \theta^T v)^{\otimes j},$$

$$S_n^{*(j)}(t, v; w_1, w_2, \theta_a, \theta_b) = \frac{1}{n}\sum_{k=1}^{n} Y_k(t)\exp\{w_1(\theta_a^T v)^T Z_k\}(w_2(\theta_b^T v)^T Z_k)^{\otimes j}.$$

To facilitate notations, we omit arguments w_2 and θ_b in $S_n^{*(0)}(t, v; w_1, w_2, \theta_a, \theta_b)$ whenever there is no ambiguity. Specifically, $S_n^{*(0)}(t, v; w_1, \theta_a) = S_n^{*(0)}(t, v; w_1, w_2, \theta_a, \theta_b)$.

Before proving theorem 1, we first state a lemma applied later. It is a direct application of Lemma A.1 in Fan et al. (2006).

Lemma 1. *Under Condition A, assume that $m(t, \cdot, \cdot, \cdot)$ is a continuous function at its four arguments. Let $c_n(t, u, \theta) = n^{-1}\sum_{i=1}^{n} K_h(\theta^T V_i - u)m(t, \theta^T V_i, (\theta^T V_i - u)/h, Z_i)\delta_i$ and $c(t, u, \theta) = \left[\int E\{m(t, u, y, Z_i)\delta_i | \theta^T V_i = u\}K(y)dy\right] f_{\theta^T V_i}(u)$.*
Suppose that $E[m(t, u, y, Z_i)|\theta^T V_i = u]$ is continuous at u. If $h \to 0$ and $nh/log(n) \to \infty$, then

$$\sup_{t\in[0,\tau]} |c_n(t, u, \theta) - c(t, u, \theta)| \xrightarrow{p} 0.$$

Proof of Theorem 1.

For any vector functions $w_1(u)$ and $w_2(u)$, define

$$Q_n(w_1, w_2, \theta; u) = \left[Q_{n1}(w_1, w_2, u, \theta)^T, Q_{n2}(w_1, w_2, \theta)^T\right]^T,$$

where

$$Q_{n1}(w_1, w_2, u, \theta) = \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau K_h(\theta^T v - u)\Bigg[\widetilde{Z}_i(u, \theta^T v) - \frac{S_n^{(1)}(t, v; w_1, w_2, u, \theta)}{S_n^{(0)}(t, v; w_1, w_2, u, \theta)}\Bigg] N_i(dt, dv),$$

$$Q_{n2}(w_1, w_2, \theta) = \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau v_2 \Bigg[w_2(\theta^T v)^T Z_i - \frac{S_n^{*(1)}(t, v; w_1, w_2, \theta, \theta)}{S_n^{*(0)}(t, v; w_1, \theta)}\Bigg] N_i(dt, dv),$$

and v_2 is a vector that contains all components of v expect for the first component. Recall the definition of $\mathbb{M} = \{v|v \in [0,1]^d, \theta^T v = u\}$ and let

$$q(w_1, w_2, \theta; u) = \Big[q_1(w_1, u, \theta)^T, 0, q_2(w_1, w_2, \theta)^T\Big]^T,$$

where

$$q_1(w_1, u, \theta) = \oint_{\mathbb{M}} \int_0^\tau \Bigg\{\widetilde{s}^{(1)}(t, \beta, \theta_0^T v) - \frac{\widetilde{s}^{(1)}(t, w_1, u)}{\widetilde{s}^{(0)}(t, w_1, u)}\widetilde{s}^{(0)}(t, \beta, \theta_0^T v)\Bigg\}\lambda_0(t, v)dtdv,$$

$$q_2(w_1, w_2, \theta) = \oint_0^1 \int_0^\tau v_2 \Bigg\{\eta_1(t, v; \theta_0, \theta, \beta, w_2) - \frac{\eta_1(t, v; \theta, \theta, w_1, w_2)}{\widetilde{s}^{(0)}(t, w_1, \theta^T v)}\widetilde{s}^{(0)}(t, \beta, \theta_0^T v)\Bigg\}\lambda_0(t, v)dtdv,$$

and $\widetilde{s}^{(1)}(t, \beta, \theta_0^T v)$, $\widetilde{s}^{(1)}(t, w_1, u)$, $\widetilde{s}^{(0)}(t, w_1, u)$, $\widetilde{s}^{(0)}(t, \beta, \theta_0^T v)$, $\eta_1(t, v; \theta_0, \theta, \beta, w_2)$, $\eta_1(t, v; \theta, \theta, w_1, w_2)$ and $\widetilde{s}^{(0)}(t, w_1, \theta^T v)$ are defined in Section 3.1.

We prove the uniform consistency of $\widehat{\beta}(\cdot)$ and $\widehat{\theta}$ based on the proof of the following three parts.

(1) Under the mark-specific proportional hazards model (2.1), we show that

$$Q_n(w_1, w_2, \theta; u) = q(w_1, w_2, \theta; u) + o_p(1).$$

It follows that $Q_n(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u) = 0$ and $q(\beta, w_2, \theta_0; u) = 0$ for any bounded w_2. Under

Condition A.7, we have that (β, θ_0) is the unique solution to $q(w_1, w_2, \theta; u) = 0$ for $w_1 \in \mathbb{S}$, $\theta \in \Theta$ and any bounded function w_2.

(2) Let $c_n = h^2 + \sqrt{log(n)/(nh)}$. Define

$$\mathcal{D}_n = \{w_1 : ||w_1|| \leq C, ||w_1(u_1) - w_1(u_2)|| \leq c(|u_1 - u_2| + c_n), u_1, u_2 \in [\iota_1, \iota_2]\},$$

$$\mathcal{D}'_n = \{w_2 : \exists M > 0, \forall u \in [\iota_1, \iota_2], ||w_2(u)|| \leq M\},$$

for some constant C, c and M.

We prove that

$$sup_{u \in [\iota_1, \iota_2], w_1 \in \mathcal{D}_n, w_2 \in \mathcal{D}'_n, \theta \in \Theta} ||Q_n(w_1, w_2, \theta; u) - q(w_1, w_2, \theta; u)|| \xrightarrow{p} 0.$$

(3) We verify that

$$P(\widehat{\beta} \in \mathcal{D}_n) \to 1, \quad as\, n \to \infty.$$

The Arzala-Ascoli Theorem (Page 208 in Royden and Fitzpatrick (2010)) tells us that if X is a compact metric space and $\{f_n\}$ is a uniformly bounded, equicontinuous sequence of real-valued functions on X, then $\{f_n\}$ has a subsequence that converges uniformly on X to a continuous function f on X.

Once the above three steps are established, since the interval $[\iota_1, \iota_2]$ with the absolute value metric is a compact metric space and $\widehat{\beta} \in \mathcal{D}_n$ with probability one as n goes to infinity, it follows that any subsequence of $\{\widehat{\beta}\}$ is uniformly bounded and equicontinuous on $[\iota_1, \iota_2]$. Thus, by the Arzela-Ascoli Theorem, any subsequence of $\{\widehat{\theta}, \widehat{\beta}\}$ has a further convergent subsequence $\{(\widehat{\theta}_m, \widehat{\beta}_m)\}$ uniformly, such that $\widehat{\theta}_m \xrightarrow{p} \theta_*$ and $\widehat{\beta}_m(u) \xrightarrow{p} \beta_*(u)$ in $u \in [\iota_1, \iota_2]$, where $\beta_* \in \mathbb{S}$. Note that

$$\begin{aligned} q(\beta_*, \widehat{\beta}', \theta_*; u) = \; & Q_n(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u) - \Big[Q_n(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u) - q(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u)\Big] \\ & - \Big[q(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u) - q(\beta_*, \widehat{\beta}', \theta_*; u)\Big]. \end{aligned} \tag{A.1}$$

By part (2), the second term in (A.1) converges to zero uniformly over $u \in [\iota_1, \iota_2]$. By part (3), the third term in (A.1) also converges to zero since $q(\beta, w_2, \theta, u)$ is continuous at (β_*, θ_*) uniformly in any bounded w_2 and $u \in [\iota_1, \iota_2]$. Combining with the fact that $Q_n(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u) = 0$, we have $q(\beta_*, \widehat{\beta}', \theta_*; u) = o_p(1)$. Hence, $q(\beta_*, \widehat{\beta}', \theta_*; u) = 0$. By Condition (A.7), since $q(w_1, w_2, \theta; u) = 0$ has unique solution (β, θ_0) for $\theta \in \Theta$, $\beta \in \mathbb{S}$ and any bounded w_2, then we have $\beta = \beta_*$ and $\theta_0 = \theta_*$. This completes the proof of Theorem 1.

In the following, we prove the parts (1), (2) and (3).

(1) We first prove that $Q_n(w_1, w_2, \theta; u) = q(w_1, w_2, \theta) + o_p(1)$.

Under mark-specific proportional hazards model (2.1), $M_i(t, v) = N_i(t, v)$ $- \int_0^t \int_0^v Y_i(s)\lambda_0(s, u) \exp\{\beta(\theta_0^T u)^T Z_i\} duds$. Under Condition A.8, by Lemma 1, the asymptotic expression of $Q_{n1}(w_1, w_2, u, \theta)$ as $n \to \infty$ can be expressed as

$$
\begin{aligned}
&Q_{n1}(w_1, w_2, u, \theta)\\
&= \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau K_h(\theta^T v - u)\left[\widetilde{Z}_i(u, \theta^T v) - \frac{S_n^{(1)}(t, v; w_1, w_2, u, \theta)}{S_n^{(0)}(t, v; w_1, w_2, u, \theta)}\right] N_i(dt, dv)\\
&= E\left\{\oint_0^1 \int_0^\tau K_h(\theta^T v - u)\left[\widetilde{Z}_i(u, \theta^T v) - \frac{s^{(1)}(t, v; w_1, w_2, u, \theta)}{s^{(0)}(t, v; w_1, w_2, u, \theta)}\right] N_i(dt, dv)\right\} + o_p(1)\\
&= \oint_0^1 \int_0^\tau E\Bigg\{K_h(\theta^T v - u)\left[\widetilde{Z}_i(u, \theta^T v) - \frac{s^{(1)}(t, v; w_1, w_2, u, \theta)}{s^{(0)}(t, v; w_1, w_2, u, \theta)}\right]\\
&\qquad\qquad \times Y_i(t) exp\{\beta(\theta_0^T v)^T Z_i\}\Bigg\}\lambda_0(t, v)dtdv + o_p(1)\\
&= \left\{\left[\oint_{\mathbb{M}} \int_0^\tau E\left\{\left[Z_i - \frac{\widetilde{s}^{(1)}(t, w_1, u)}{\widetilde{s}^{(0)}(t, w_1, u)}\right] Y_i(t) exp\{\beta(\theta_0^T v)^T Z_i\}\right\}\lambda_0(t, v)dtdv\right]^T, 0\right\}^T\\
&\quad + o_p(1). \qquad\qquad (A.2)
\end{aligned}
$$

By large number theory, the first part of $Q_{n1}(w_1, w_2, u, \theta)$ takes the form

$$\oint_{\mathbb{M}} \int_0^\tau \left\{ E\Big[Z_i Y_i(t) exp\{\beta(\theta_0^T v)^T Z_i\}\Big] - \frac{\widetilde{s}^{(1)}(t, w_1, u)}{\widetilde{s}^{(0)}(t, w_1, u)} E\Big[Y_i(t) exp\{\beta(\theta_0^T v)^T Z_i\}\Big] \right\} \times \lambda_0(t, v) dt dv$$

$$= \oint_{\mathbb{M}} \int_0^\tau \left\{ \widetilde{s}^{(1)}(t, \beta, \theta_0^T v) - \frac{\widetilde{s}^{(1)}(t, w_1, u)}{\widetilde{s}^{(0)}(t, w_1, u)} \widetilde{s}^{(0)}(t, \beta, \theta_0^T v) \right\} \lambda_0(t, v) dt dv$$

$$= q_1(w_1, u, \theta). \tag{A.3}$$

Hence, we have

$$Q_{n1}(w_1, w_2, u, \theta) = \{q_1(w_1, u, \theta)^T, 0\}^T + o_p(1). \tag{A.4}$$

Now, we derive the asymptotic expression of $Q_{n2}(w_1, w_2, \theta)$. Under Condition A, it follows from Lemma 2 (D.2) in Gilbert et al. (2008) that

$$\begin{aligned}
&Q_{n2}(w_1, w_2, \theta) \\
&\quad = \frac{1}{n} \sum_{i=1}^n \oint_0^1 \int_0^\tau v_2 \left[w_2(\theta^T v)^T Z_i - \frac{S_n^{*(1)}(t, v; w_1, w_2, \theta, \theta)}{S_n^{*(0)}(t, v; w_1, \theta)} \right] N_i(dt, dv) \\
&\quad = E\left\{ \oint_0^1 \int_0^\tau v_2 \left[w_2(\theta^T v)^T Z_i - \frac{S_n^{*(1)}(t, v; w_1, w_2, \theta, \theta)}{S_n^{*(0)}(t, v; w_1, \theta)} \right] N_i(dt, dv) \right\} + o_p(1) \\
&\quad = E\left\{ \oint_0^1 \int_0^\tau v_2 \left[w_2(\theta^T v)^T Z_i - \frac{S_n^{*(1)}(t, v; w_1, w_2, \theta, \theta)}{S_n^{*(0)}(t, v; w_1, \theta)} \right] \right. \\
&\qquad\qquad \left. \times Y_i(t) \lambda_0(t, v) exp\{\beta(\theta_0^T v)^T Z_i\} dt dv \right\} \\
&\qquad + o_p(1) \\
&\quad = \oint_0^1 \int_0^\tau v_2 \left\{ \eta_1(t, v; \theta_0, \theta, \beta, w_2) - \frac{\eta_1(t, v; \theta, \theta, w_1, w_2)}{\widetilde{s}^{(0)}(t, w_1, \theta^T v)} \widetilde{s}^{(0)}(t, \beta, \theta_0^T v) \right\} \lambda_0(t, v) dt dv \\
&\qquad + o_p(1) \\
&\quad = q_2(w_1, w_2, \theta) + o_p(1).
\end{aligned}$$

Thus, it follows that

$$Q_n(w_1, w_2, \theta; u) = q(w_1, w_2, \theta; u) + o_p(1).$$

(2) By the uniform strong law of large numbers (Theorem 8.3 in Pollard (1990)), for each continuous function w_1 and any bounded w_2,

$$sup_{u \in [\iota_1, \iota_2]} ||Q_n(w_1, w_2, \theta; u) - q(w_1, w_2, \theta; u)|| \xrightarrow{p} 0. \tag{A.5}$$

From the construction of ϵ-net of $\mathcal{D}_n$, let $D(\epsilon, \mathcal{D}_n, ||\cdot||_\infty)$ be the covering number of class $\mathcal{D}_n$, we can have

$$logD(\epsilon, \mathcal{D}_n, ||\cdot||_\infty) \le O\left(\frac{1}{\epsilon} log\left(\frac{1}{\epsilon}\right)\right) = o(n). \tag{A.6}$$

Applying theorem 8.2 in Pollard (1990), together with (A.5) and (A.6), we have

$$sup_{u \in [\iota_1, \iota_2], w_1 \in \mathcal{D}_n, w_2 \in \mathcal{D}'_n, \theta \in \Theta} ||Q_n(w_1, w_2, \theta; u) - q(w_1, w_2, \theta; u)|| \xrightarrow{p} 0.$$

(3) Denote $Q_{n1,p}(\widehat{\beta}, h\widehat{\beta}', \widehat{\theta}; u)$, $\widehat{S}_{n,p}^{(1)}(t, v; u)$ and $\widehat{S}_{n,p}^{(2)}(t, v; u)$ to be the first p elements of $Q_{n1}(\widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})$, $S_n^{(1)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})$ and $S_n^{(2)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})$, respectively.

Let $\widehat{S}_n^{(0)}(t, v; u) = S_n^{(0)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})$. Under mark-specific proportional hazards model (2.1), $M_i(t, v) = N_i(t, v) - \int_0^t \int_0^v Y_i(s)\lambda_0(s, u) \exp\{\beta(\theta_0^T u)^T Z_i\} duds$.

Then, given that $u_1, u_2 \in [\iota_1, \iota_2]$ and $|u_1 - u_2| \le h$, to prove that the probability of $\widehat{\beta} \in \mathcal{D}_n$ converges to one as $n \to \infty$, we first derive the asymptotic expression of $Q_{n1,p}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u_1) - Q_{n1,p}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u_2)$. Recall the definition of $\mathbb{M}_1 = \left\{v | v \in [0,1]^d, \widehat{\theta}^T v = u_1\right\}$ and let $\mathbb{M}_2 = \left\{v | v \in [0,1]^d, \widehat{\theta}^T v = u_2\right\}$.

Consider the following decomposition of $Q_{n1,p}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u_1) - Q_{n1,p}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u_2)$.

$$
\begin{aligned}
& Q_{n1,p}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u_1) - Q_{n1,p}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u_2) \\
= \quad & n^{-1}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau \left[K_h(\widehat{\theta}^T v - u_1) Z_i - K_h(\widehat{\theta}^T v - u_2) Z_i\right] N_i(dt, dv) \\
& - n^{-1}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau \left[K_h(\widehat{\theta}^T v - u_1) - K_h(\widehat{\theta}^T v - u_2)\right] \frac{\widehat{S}_{n,p}^{(1)}(t, v; u_2)}{\widehat{S}_n^{(0)}(t, v; u_2)} N_i(dt, dv) \\
& + n^{-1}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau K_h(\widehat{\theta}^T v - u_1) \left[\frac{\widehat{S}_{n,p}^{(1)}(t, v; u_2) - \widehat{S}_{n,p}^{(1)}(t, v; u_1)}{\widehat{S}_n^{(0)}(t, v; u_2)}\right] N_i(dt, dv) \\
& + n^{-1}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau K_h(\widehat{\theta}^T v - u_1) \widehat{S}_{n,p}^{(1)}(t, v; u_1) \\
& \qquad \times \left[\frac{\widehat{S}_n^{(0)}(t, v; u_1) - \widehat{S}_n^{(0)}(t, v; u_2)}{\widehat{S}_n^{(0)}(t, v; u_1) \widehat{S}_n^{(0)}(t, v; u_2)}\right] N_i(dt, dv) \\
\equiv \quad & I + II + III + IV. \qquad \text{(A.7)}
\end{aligned}
$$

First, we consider I. By the application of Theorem 2 in Hansen (2008), it can be shown that

$$
\begin{aligned}
I &= n^{-1}\sum_{i=1}^{n}\oint_0^1\int_0^\tau\left[K_h(\widehat{\theta}^T v-u_1)Z_i-K_h(\widehat{\theta}^T v-u_2)Z_i\right]N_i(dt,dv)\\
&= E\left\{\oint_0^1\int_0^\tau\left[K_h(\widehat{\theta}^T v-u_1)Z_i-K_h(\widehat{\theta}^T v-u_2)Z_i\right]N_i(dt,dv)\right\}\\
&\quad + O_p\Big(\sqrt{log(n)/(nh)}\Big)\\
&= \oint_0^1\int_0^\tau E\left\{\left[K_h(\widehat{\theta}^T v-u_1)Z_i-K_h(\widehat{\theta}^T v-u_2)Z_i\right]Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\right\}\\
&\qquad\qquad \times \lambda_0(t,v)dtdv\\
&\quad + O_p\Big(\sqrt{log(n)/(nh)}\Big)\\
&= \oint_{\mathbb{M}_1}\int_0^\tau E\left\{Z_iY_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\right\}\lambda_0(t,v)dtdv\\
&\qquad - \oint_{\mathbb{M}_2}\int_0^\tau E\left\{Z_iY_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\right\}\lambda_0(t,v)dtdv\\
&\quad + O_p\Big(h^2+\sqrt{log(n)/(nh)}\Big)\\
&= \oint_{\mathbb{M}_1}\int_0^\tau \widetilde{s}^{(1)}(t,\beta,\theta_0^T v)\lambda_0(t,v)dtdv - \oint_{\mathbb{M}_2}\int_0^\tau \widetilde{s}^{(1)}(t,\beta,\theta_0^T v)\lambda_0(t,v)dtdv\\
&\quad + O_p(c_n). \qquad\qquad (A.8)
\end{aligned}
$$

Next, we derive the asymptotic expression of II. Under Condition A.6, using a Taylor expansion, we have

$$
\begin{aligned}
K_h(\widehat{\theta}^T v-u_1)-K_h(\widehat{\theta}^T v-u_2) &= \left\{\frac{\partial}{\partial u}K_h\left(\widehat{\theta}^T v-u_2\right)\right\}(u_1-u_2)\\
&\quad + O_p\{(u_1-u_2)^2\}. \qquad\qquad (A.9)
\end{aligned}
$$

Then, under Condition A.8, substituting (A.9) into II, it follows from the

Theorem 2 in Hansen (2008) that

$$
\begin{aligned}
II &= -n^{-1}\sum_{i=1}^{n}\oint_0^1\int_0^\tau\left[K_h(\widehat{\theta}^T v-u_1)-K_h(\widehat{\theta}^T v-u_2)\right]\frac{\widehat{S}_{n,p}^{(1)}(t,v;u_2)}{\widehat{S}_n^{(0)}(t,v;u_2)}N_i(dt,dv)\\
&= -n^{-1}\sum_{i=1}^{n}\oint_0^1\int_0^\tau\left[\frac{\partial}{\partial u}K_h(\widehat{\theta}^T v-u_2)\frac{\widehat{S}_{n,p}^{(1)}(t,v;u_2)}{\widehat{S}_n^{(0)}(t,v;u_2)}\right]N_i(dt,dv)(u_1-u_2)\\
&\quad+O_p\{(u_1-u_2)^2\}\\
&= -E\left\{\oint_0^1\int_0^\tau\left[\frac{\partial}{\partial u}K_h(\widehat{\theta}^T v-u_2)\frac{\widehat{S}_{n,p}^{(1)}(t,v;u_2)}{\widehat{S}_n^{(0)}(t,v;u_2)}\right]N_i(dt,dv)\right\}(u_1-u_2)\\
&\quad+O_p\Big\{(u_1-u_2)^2+\sqrt{log(n)/(nh)}\Big\}\\
&= -\oint_{\mathbb{M}_2}\int_0^\tau\frac{\partial}{\partial u}E\left\{\frac{\widetilde{s}^{(1)}(t,\widehat{\beta},u_2)}{\widetilde{s}^{(0)}(t,\widehat{\beta},u_2)}Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\right\}\lambda_0(t,v)dtdv(u_1-u_2)\\
&\quad+O_p\{(u_1-u_2)^2+c_n\}. \qquad\qquad (A.10)
\end{aligned}
$$

By the law of large number, we have

$$
\begin{aligned}
II &= -\oint_{\mathbb{M}_2}\int_0^\tau\frac{\partial}{\partial u}\left\{\frac{\widetilde{s}^{(1)}(t,\widehat{\beta},u_2)}{\widetilde{s}^{(0)}(t,\widehat{\beta},u_2)}E\left\{Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\right\}\right\}\lambda_0(t,v)dtdv(u_1-u_2)\\
&\quad+O_p\{(u_1-u_2)^2+c_n\}\\
&= -\oint_{\mathbb{M}_2}\int_0^\tau\frac{\partial}{\partial u}\left\{\frac{\widetilde{s}^{(1)}(t,\widehat{\beta},u_2)}{\widetilde{s}^{(0)}(t,\widehat{\beta},u_2)}\widetilde{s}^{(0)}(t,\beta,\theta_0^T v)\right\}\lambda_0(t,v)dtdv(u_1-u_2)\\
&\quad+O_p\{(u_1-u_2)^2+c_n\}. \qquad\qquad (A.11)
\end{aligned}
$$

Before deriving the asymptotic expression of III and IV, under Condition A.8, we apply the Taylor expansions for $\widehat{S}_n^{(0)}(t,v;u_2)$, $\widehat{S}_n^{(0)}(t,v;u_1)-\widehat{S}_n^{(0)}(t,v;u_2)$ and $\widehat{S}_{n,p}^{(1)}(t,v;u_2)-\widehat{S}_{n,p}^{(1)}(t,v;u_1)$ as follows:

$$
\widehat{S}_n^{(0)}(t,v;u_2)=\widehat{S}_n^{(0)}(t,v;u_1)+O_p(u_1-u_2), \qquad\qquad (A.12)
$$

$$
\begin{aligned}
&\widehat{S}_n^{(0)}(t,v;u_1) - \widehat{S}_n^{(0)}(t,v;u_2) \\
&= \quad n^{-1}\sum_{k=1}^{n} Y_k(t)\exp\left[(\widehat{\beta}(u_1))^T Z_k + (\widehat{\beta}'(u_1))^T Z_k(\widehat{\theta}^T v - u_1)\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times Z_k\{\widehat{\beta}(u_1) - \widehat{\beta}(u_2)\} \\
&\quad + n^{-1}\sum_{k=1}^{n} Y_k(t)\exp\left[(\widehat{\beta}(u_1))^T Z_k + (\widehat{\beta}'(u_1))^T Z_k(\widehat{\theta}^T v - u_1)\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times Z_k(\widehat{\theta}^T v - u_1)\{\widehat{\beta}'(u_1) - \widehat{\beta}'(u_2)\} \\
&\quad + O_p\{(u_1 - u_2)^2\} \\
&= \quad \widehat{S}_{n,p}^{(1)}(t,v;u_1)\{\widehat{\beta}(u_1) - \widehat{\beta}(u_2)\} + \widehat{S}_{n,2p}^{(1)}(t,v;u_1)\{\widehat{\beta}'(u_1) - \widehat{\beta}'(u_2)\} \\
&\quad + O_p\{(u_1 - u_2)^2\}, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.13)}
\end{aligned}
$$

where

$$
\begin{aligned}
&\widehat{S}_{n,2p}^{(1)}(t,v;u_1) \\
&\qquad = n^{-1}\sum_{k=1}^{n} Y_k(t)\exp\left[(\widehat{\beta}(u_1))^T Z_k + (\widehat{\beta}'(u_1))^T Z_k(\widehat{\theta}^T v - u_1)\right] Z_k(\widehat{\theta}^T v - u_1),
\end{aligned}
$$

and

$$
\begin{aligned}
&\widehat{S}_{n,p}^{(1)}(t,v;u_2) - \widehat{S}_{n,p}^{(1)}(t,v;u_1) \\
&\quad = n^{-1}\sum_{k=1}^{n} Y_k(t)\exp\Big[(\widehat{\beta}(u_1))^T Z_k + (\widehat{\beta}'(u_1))^T Z_k(\widehat{\theta}^T v - u_1)\Big] \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times Z_k Z_k^T\big\{\widehat{\beta}(u_2) - \widehat{\beta}(u_1)\big\} \\
&\qquad + n^{-1}\sum_{k=1}^{n} Y_k(t)\exp\Big[(\widehat{\beta}(u_1))^T Z_k + (\widehat{\beta}'(u_1))^T Z_k(\widehat{\theta}^T v - u_1)\Big] \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times Z_k Z_k^T(\widehat{\theta}^T v - u_1)\big\{\widehat{\beta}'(u_2) - \widehat{\beta}'(u_1)\big\} \\
&\qquad + O_p\{(u_2 - u_1)^2\} \\
&\quad = \widehat{S}_{n,p}^{(2)}(t,v;u_1)\{\widehat{\beta}(u_2) - \widehat{\beta}(u_1)\} + \widehat{S}_{n,2p}^{(1,1)}(t,v;u_1)\{\widehat{\beta}'(u_2) - \widehat{\beta}'(u_1)\} \\
&\qquad + O_p\{(u_2 - u_1)^2\},
\end{aligned} \tag{A.14}
$$

where

$$
\begin{aligned}
&\widehat{S}_{n,2p}^{(1,1)}(t,v;u_1) \\
&\quad = n^{-1}\sum_{k=1}^{n} Y_k(t)\exp\Big[(\widehat{\beta}(u_1))^T Z_k + (\widehat{\beta}'(u_1))^T Z_k(\widehat{\theta}^T v - u_1)\Big] \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times Z_k Z_k^T(\widehat{\theta}^T v - u_1).
\end{aligned}
$$

Substituting (A.12) and (A.14) into III, it follows from the Theorem 2 in Hansen

(2008) that

$$
\begin{aligned}
III &= n^{-1}\sum_{i=1}^{n}\oint_0^1\int_0^\tau K_h(\widehat{\theta}^T v-u_1)\left[\frac{\widehat{S}_{n,p}^{(1)}(t,v;u_2)-\widehat{S}_{n,p}^{(1)}(t,v;u_1)}{\widehat{S}_n^{(0)}(t,v;u_2)}\right]N_i(dt,dv)\\
&= n^{-1}\sum_{i=1}^{n}\oint_0^1\int_0^\tau K_h(\widehat{\theta}^T v-u_1)\\
&\quad\times\left[\frac{\widehat{S}_{n,p}^{(2)}(t,v;u_1)\{\widehat{\beta}(u_2)-\widehat{\beta}(u_1)\}+\widehat{S}_{n,2p}^{(1,1)}(t,v;u_1)\{\widehat{\beta}'(u_2)-\widehat{\beta}'(u_1)\}}{\widehat{S}_n^{(0)}(t,v;u_1)}\right]N_i(dt,dv)\\
&\quad+O_p\{(u_2-u_1)^2\}\\
&= E\Bigg\{\oint_0^1\int_0^\tau K_h(\widehat{\theta}^T v-u_1)\\
&\quad\times\left[\frac{\widehat{S}_{n,p}^{(2)}(t,v;u_1)\{\widehat{\beta}(u_2)-\widehat{\beta}(u_1)\}+\widehat{S}_{n,2p}^{(1,1)}(t,v;u_1)\{\widehat{\beta}'(u_2)-\widehat{\beta}'(u_1)\}}{\widehat{S}_n^{(0)}(t,v;u_1)}\right]\\
&\qquad\times Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\lambda_0(t,v)dtdv\Bigg\}\\
&\quad+O_p\Big\{(u_1-u_2)^2+\sqrt{log(n)/(nh)}\Big\}\\
&= \oint_{\mathbb{M}_1}\int_0^\tau E\left\{\frac{\widetilde{s}^{(2)}(t,\widehat{\beta},u_1)}{\widetilde{s}^{(0)}(t,\widehat{\beta},u_1)}Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\right\}\lambda_0(t,v)dtdv\{\widehat{\beta}(u_2)-\widehat{\beta}(u_1)\}\\
&\quad+O_p\{(u_2-u_1)^2+c_n\}. \qquad (A.15)
\end{aligned}
$$

By large number theory, we get

$$
\begin{aligned}
III &= \oint_{\mathbb{M}_1}\int_0^\tau \frac{\widetilde{s}^{(2)}(t,\widehat{\beta},u_1)}{\widetilde{s}^{(0)}(t,\widehat{\beta},u_1)}E\Big\{Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\Big\}\lambda_0(t,v)dtdv\{\widehat{\beta}(u_2)-\widehat{\beta}(u_1)\}\\
&\quad+O_p\{(u_2-u_1)^2+c_n\}\\
&= \oint_{\mathbb{M}_1}\int_0^\tau\left\{\frac{\widetilde{s}^{(2)}(t,\widehat{\beta},u_1)}{\widetilde{s}^{(0)}(t,\widehat{\beta},u_1)}\widetilde{s}^{(0)}(t,\beta,\theta_0^T v)\right\}\lambda_0(t,v)dtdv\{\widehat{\beta}(u_2)-\widehat{\beta}(u_1)\}\\
&\quad+O_p\{(u_2-u_1)^2+c_n\}. \qquad (A.16)
\end{aligned}
$$

Similarly, substituting (A.12) and (A.13) into IV, together with the application of

Theorem 2 in Hansen (2008), we get

$$
\begin{aligned}
& IV \\
&= n^{-1}\sum_{i=1}^{n}\oint_0^1\int_0^\tau K_h(\widehat{\theta}^T v-u_1)\widehat{S}_{n,p}^{(1)}(t,v;u_1)\left[\frac{\widehat{S}_n^{(0)}(t,v;u_1)-\widehat{S}_n^{(0)}(t,v;u_2)}{\widehat{S}_n^{(0)}(t,v;u_1)\widehat{S}_n^{(0)}(t,v;u_2)}\right]N_i(dt,dv) \\
&= n^{-1}\sum_{i=1}^{n}\oint_0^1\int_0^\tau K_h(\widehat{\theta}^T v-u_1)\widehat{S}_{n,p}^{(1)}(t,v;u_1) \\
&\qquad\times\left[\frac{\widehat{S}_{n,p}^{(1)}(t,v;u_1)\{\widehat{\beta}(u_1)-\widehat{\beta}(u_2)\}+\widehat{S}_{n,2p}^{(1)}(t,v;u_1)\{\widehat{\beta}'(u_1)-\widehat{\beta}'(u_2)\}}{\widehat{S}_n^{(0)}(t,v;u_1)^{\otimes 2}}\right] \\
&\qquad\qquad\times N_i(dt,dv) \\
&\quad+O_p\{(u_2-u_1)^2\} \\
&= E\Bigg\{\oint_0^1\int_0^\tau K_h(\widehat{\theta}^T v-u_1)\widehat{S}_{n,p}^{(1)}(t,v;u_1) \\
&\qquad\times\left[\frac{\widehat{S}_{n,p}^{(1)}(t,v;u_1)\{\widehat{\beta}(u_1)-\widehat{\beta}(u_2)\}+\widehat{S}_{n,2p}^{(1)}(t,v;u_1)\{\widehat{\beta}'(u_1)-\widehat{\beta}'(u_2)\}}{\widehat{S}_n^{(0)}(t,v;u_1)^{\otimes 2}}\right] \\
&\qquad\qquad\times Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\lambda_0(t,v)dtdv\Bigg\} \\
&\quad+O_p\Big\{(u_1-u_2)^2+\sqrt{log(n)/(nh)}\Big\} \\
&= \oint_{\mathbb{M}_1}\int_0^\tau E\left\{\frac{\widetilde{s}^{(1)}(t,\widehat{\beta},u_1)\widetilde{s}^{(1)}(t,\widehat{\beta},u_1)^T}{\widetilde{s}^{(0)}(t,\widehat{\beta},u_1)^{\otimes 2}}Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\right\} \\
&\qquad\qquad\times\lambda_0(t,v)dtdv\{\widehat{\beta}(u_1)-\widehat{\beta}(u_2)\} \\
&\quad+O_p\{(u_2-u_1)^2+c_n\}. \qquad\qquad (A.17)
\end{aligned}
$$

By the law of large number, we get

$$IV = \oint_{\mathbb{M}_1} \int_0^{\tau} \frac{\widetilde{s}^{(1)}(t, \widehat{\beta}, u_1)\widetilde{s}^{(1)}(t, \widehat{\beta}, u_1)^T}{\widetilde{s}^{(0)}(t, \widehat{\beta}, u_1)^{\otimes 2}} E\Big\{Y_i(t) exp\{\beta(\theta_0^T v)^T Z_i\}\Big\}$$

$$\times \lambda_0(t, v) dt dv \{\widehat{\beta}(u_1) - \widehat{\beta}(u_2)\}$$

$$+ O_p\{(u_2 - u_1)^2 + c_n\}$$

$$= \oint_{\mathbb{M}_1} \int_0^{\tau} \left\{ \frac{\widetilde{s}^{(1)}(t, \widehat{\beta}, u_1)\widetilde{s}^{(1)}(t, \widehat{\beta}, u_1)^T}{\widetilde{s}^{(0)}(t, \widehat{\beta}, u_1)^{\otimes 2}} \widetilde{s}^{(0)}(t, \beta, \theta_0^T v) \right\} \lambda_0(t, v) dt dv \{\widehat{\beta}(u_1) - \widehat{\beta}(u_2)\}$$

$$+ O_p\{(u_2 - u_1)^2 + c_n\}. \tag{A.18}$$

Since $Q_{n1}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u_1) = Q_{n1}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}; u_2) = 0$, by (A.7), combining the asymptotic expressions of I, II, III and IV with $u_1, u_2 \in [\iota_1, \iota_2]$ such that $|u_1 - u_2| \leq h$, then we obtain

$$\oint_{\mathbb{M}_1} \int_0^{\tau} \left\{ \frac{\widetilde{s}^{(2)}(t, \widehat{\beta}, u_1)}{\widetilde{s}^{(0)}(t, \widehat{\beta}, u_1)} - \frac{\widetilde{s}^{(1)}(t, \widehat{\beta}, u_1)\widetilde{s}^{(1)}(t, \widehat{\beta}, u_1)^T}{\widetilde{s}^{(0)}(t, \widehat{\beta}, u_1)^{\otimes 2}} \right\} \widetilde{s}^{(0)}(t, \beta, \theta_0^T v)$$

$$\times \lambda_0(t, v) dt dv \{\widehat{\beta}(u_1) - \widehat{\beta}(u_2)\}$$

$$= \oint_{\mathbb{M}_1} \int_0^{\tau} \widetilde{s}^{(1)}(t, \beta, \theta_0^T v)\lambda_0(t, v) dt dv - \oint_{\mathbb{M}_2} \int_0^{\tau} \widetilde{s}^{(1)}(t, \beta, \theta_0^T v)\lambda_0(t, v) dt dv$$

$$- \oint_{\mathbb{M}_2} \int_0^{\tau} \frac{\partial}{\partial u} \left\{ \frac{\widetilde{s}^{(1)}(t, \widehat{\beta}, u_2)}{\widetilde{s}^{(0)}(t, \widehat{\beta}, u_2)} \widetilde{s}^{(0)}(t, \beta, \theta_0^T v) \right\} \lambda_0(t, v) dt dv (u_1 - u_2)$$

$$+ O_p\{c_n + (u_1 - u_2)^2\}. \tag{A.19}$$

Under Condition A.8, for the first terms on the right hand side of the equality in (A.19), we have

$$\oint_{\mathbb{M}_1} \int_0^{\tau} \widetilde{s}^{(1)}(t, \beta, \theta_0^T v)\lambda_0(t, v) dt dv - \oint_{\mathbb{M}_2} \int_0^{\tau} \widetilde{s}^{(1)}(t, \beta, \theta_0^T v)\lambda_0(t, v) dt dv = O(u_1 - u_2),$$

$$\oint_{\mathbb{M}_2} \int_0^{\tau} \widetilde{s}^{(1)}(t, \beta, \theta_0^T v)\lambda_0(t, v) dt dv = O(1).$$

Under Condition A, by (A.19), it follows that $P(\widehat{\beta} \in \mathcal{D}_n) \to 1$ as $n \to \infty$.

Before proving theorems 2 and 3, recall that we impose the restriction on the first element of $\theta = (\theta_1, \theta_2^T)^T$, setting $\theta_1 = 1$. Hence, to show the asymptotic normality of $\widehat{\theta}$ and $\widehat{\beta}(\cdot)$ is to prove that of $\widehat{\theta}_2$ and $\widehat{\beta}(\cdot)$.

We divide the proofs of theorems 2 and 3 into the following three steps. The first step is to consider the asymptotic expression of $\widehat{\theta}_2 - \theta_{20}$, which will be shown in Lemma 2. The second step is to derive the asymptotic expansion of $\widehat{\beta}(u) - \beta(u)$, which will be proved in Lemma 3. The third step is to combine the results from Lemma 2 and 3 to establish the asymptotic normality of $\sqrt{n}(\widehat{\theta}_2 - \theta_{20})$ and $\sqrt{nh}(\widehat{\beta}(u) - \beta(u))$.

Let $a_n = ||\widehat{\theta} - \theta_0||$, $b_n = \sup_{u \in [\iota_1, \iota_2]} ||\widehat{\beta}(u) - \beta(u)||$ and $b_n' = \sup_{u \in [\iota_1, \iota_2]} ||h\widehat{\beta}'(u) - h\beta'(u)||$. By Theorem 1, we have $a_n \to 0$ and $b_n \to 0$ uniformly in probability.

Lemma 2. *Under the mark-specific proportional hazards model (2.1) and Condition A, the asymptotic expression of $\widehat{\theta}_2 - \theta_{20}$ can be written as*

$$
\begin{aligned}
\widehat{\theta}_2 - \theta_{20} = & -\frac{1}{n} A_\theta^{-1} \sum_{i=1}^{n} \oint_0^1 \int_0^\tau v_2 \Big\{ \beta'(\theta_0^T v)^T Z_i - \frac{s^{*(1)}(t, \theta_0^T v)}{s^{*(0)}(t, \theta_0^T v)} \Big\} M_i(dt, dv) \\
& - A_\theta^{-1} \oint_0^1 W(v) \{ \widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v) \} dv + o_p(n^{-1/2}) \\
& + O_p((a_n + b_n)(a_n + b_n + h^{-1} b_n')),
\end{aligned}
$$

where

$$
\begin{aligned}
A_\theta = & \oint_0^1 \int_0^\tau v_2 \Big\{ \frac{s^{*(1)}(t, \theta_0^T v)^{\otimes 2}}{s^{*(0)}(t, \theta_0^T v)} - s^{*(2)}(t, \theta_0^T v) \Big\} v_2^T \lambda_0(t, v) dt dv, \\
W(v) = & v_2 \int_0^\tau B(t, \theta_0^T v) \lambda_0(t, v) dt, \\
B(t, \theta_0^T v) = & \Big\{ \frac{s^{*(1)}(t, \theta_0^T v) \widetilde{s}^{(1)}(t, \beta, \theta_0^T v)^T}{s^{*(0)}(t, \theta_0^T v)} - \phi(t, \theta_0^T v)^T \Big\}.
\end{aligned}
$$

Proof. We shall prove Lemma 2 through deriving the asymptotic expression of $Q_{n2}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}) - Q_{n2}(\beta, \beta', \theta_0)$ as well as that of $Q_{n2}(\beta, \beta', \theta_0)$.

First, we consider the following decomposition of $Q_{n2}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}) - Q_{n2}(\beta, \beta', \theta_0)$.

$$
\begin{aligned}
&Q_{n2}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}) - Q_{n2}(\beta, \beta', \theta_0) \\
&= \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau v_2 \Big\{ \widehat{\beta}'(\widehat{\theta}^T v)^T Z_i - \beta'(\theta_0^T v)^T Z_i \\
&\qquad - \frac{S_n^{*(1)}(t, v; \widehat{\beta}, \widehat{\beta}', \widehat{\theta}, \widehat{\theta})}{S_n^{*(0)}(t, v; \widehat{\beta}, \widehat{\theta})} + \frac{S_n^{*(1)}(t, v; \beta, \beta', \theta_0, \theta_0)}{S_n^{*(0)}(t, v; \beta, \theta_0)} \Big\} N_i(dt, dv) \\
&= \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau v_2 \Big\{ \widehat{\beta}'(\widehat{\theta}^T v)^T Z_i - \beta'(\theta_0^T v)^T Z_i \\
&\qquad - \frac{S_n^{*(1)}(t, v; \widehat{\beta}, \widehat{\beta}', \widehat{\theta}, \widehat{\theta})}{S_n^{*(0)}(t, v; \widehat{\beta}, \widehat{\theta})} + \frac{S_n^{*(1)}(t, v; \beta, \beta', \theta_0, \theta_0)}{S_n^{*(0)}(t, v; \beta, \theta_0)} \Big\} M_i(dt, dv) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau v_2 \Big\{ \widehat{\beta}'(\widehat{\theta}^T v)^T Z_i - \beta'(\theta_0^T v)^T Z_i \\
&\qquad - \frac{S_n^{*(1)}(t, v; \widehat{\beta}, \widehat{\beta}', \widehat{\theta}, \widehat{\theta})}{S_n^{*(0)}(t, v; \widehat{\beta}, \widehat{\theta})} + \frac{S_n^{*(1)}(t, v; \beta, \beta', \theta_0, \theta_0)}{S_n^{*(0)}(t, v; \beta, \theta_0)} \Big\} \\
&\qquad \times Y_i(t)\lambda_0(t, v) \exp\{\beta(\theta_0^T v)^T Z_i\} dt dv \\
&= \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau v_2 \Big\{ \widehat{\beta}'(\widehat{\theta}^T v)^T Z_i - \beta'(\theta_0^T v)^T Z_i \\
&\qquad - \frac{S_n^{*(1)}(t, v; \widehat{\beta}, \widehat{\beta}', \widehat{\theta}, \widehat{\theta})}{S_n^{*(0)}(t, v; \widehat{\beta}, \widehat{\theta})} + \frac{S_n^{*(1)}(t, v; \beta, \beta', \theta_0, \theta_0)}{S_n^{*(0)}(t, v; \beta, \theta_0)} \Big\} M_i(dt, dv) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau v_2 \Big\{ \widehat{\beta}'(\widehat{\theta}^T v)^T Z_i - \beta'(\theta_0^T v)^T Z_i \\
&\qquad - \frac{S_n^{*(1)}(t, v; \beta, \widehat{\beta}', \theta_0, \widehat{\theta})}{S_n^{*(0)}(t, v; \beta, \theta_0)} + \frac{S_n^{*(1)}(t, v; \beta, \beta', \theta_0, \theta_0)}{S_n^{*(0)}(t, v; \beta, \theta_0)} \Big\} \\
&\qquad \times Y_i(t)\lambda_0(t, v) \exp\{\beta(\theta_0^T v)^T Z_i\} dt dv \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau v_2 \Big\{ \frac{S_n^{*(1)}(t, v; \beta, \widehat{\beta}', \theta_0, \widehat{\theta})}{S_n^{*(0)}(t, v; \beta, \theta_0)} - \frac{S_n^{*(1)}(t, v; \widehat{\beta}, \widehat{\beta}', \widehat{\theta}, \widehat{\theta})}{S_n^{*(0)}(t, v; \widehat{\beta}, \widehat{\theta})} \Big\} \\
&\qquad \times Y_i(t)\lambda_0(t, v) \exp\{\beta(\theta_0^T v)^T Z_i\} dt dv \\
&\equiv I + II + III. \qquad \text{(A.20)}
\end{aligned}
$$

Let us deal with II first. It can be easily shown that

$$\begin{aligned}
II = \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau v_2\Big\{&\widehat{\beta}'(\widehat{\theta}^T v)^T Z_i - \beta'(\theta_0^T v)^T Z_i \\
&- \frac{S_n^{*(1)}(t,v;\beta,\widehat{\beta}',\theta_0,\widehat{\theta})}{S_n^{*(0)}(t,v;\beta,\theta_0)} + \frac{S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)}{S_n^{*(0)}(t,v;\beta,\theta_0)}\Big\} \\
&\times Y_i(t)\lambda_0(t,v)\exp\{\beta(\theta_0^T v)^T Z_i\}dtdv \\
= \oint_0^1\int_0^\tau v_2\Big\{&S_n^{*(1)}(t,v;\beta,\widehat{\beta}',\theta_0,\widehat{\theta}) - S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)\Big\}\lambda_0(t,v)dtdv \\
- \oint_0^1\int_0^\tau v_2\Big\{&\frac{S_n^{*(1)}(t,v;\beta,\widehat{\beta}',\theta_0,\widehat{\theta})}{S_n^{*(0)}(t,v;\beta,\theta_0)} - \frac{S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)}{S_n^{*(0)}(t,v;\beta,\theta_0)}\Big\} \\
&\times S_n^{*(0)}(t,v;\beta,\theta_0)\lambda_0(t,v)dtdv \\
= 0. & \qquad\qquad\qquad\qquad (A.21)
\end{aligned}$$

We now derive the asymptotic expression of III.

$$\begin{aligned}
III = \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau v_2\Big\{&\frac{S_n^{*(1)}(t,v;\beta,\widehat{\beta}',\theta_0,\widehat{\theta})}{S_n^{*(0)}(t,v;\beta,\theta_0)} - \frac{S_n^{*(1)}(t,v;\widehat{\beta},\widehat{\beta}',\widehat{\theta},\widehat{\theta})}{S_n^{*(0)}(t,v;\widehat{\beta},\widehat{\theta})}\Big\} \\
&\times Y_i(t)\lambda_0(t,v)\exp\{\beta(\theta_0^T v)^T Z_i\}dtdv \\
= \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau v_2\Big\{&\frac{S_n^{*(1)}(t,v;\beta,\widehat{\beta}',\theta_0,\widehat{\theta})}{S_n^{*(0)}(t,v;\beta,\theta_0)} - \frac{S_n^{*(1)}(t,v;\widehat{\beta},\widehat{\beta}',\widehat{\theta},\widehat{\theta})}{S_n^{*(0)}(t,v;\beta,\theta_0)} \\
&+ \frac{S_n^{*(1)}(t,v;\widehat{\beta},\widehat{\beta}',\widehat{\theta},\widehat{\theta})}{S_n^{*(0)}(t,v;\beta,\theta_0)} - \frac{S_n^{*(1)}(t,v;\widehat{\beta},\widehat{\beta}',\widehat{\theta},\widehat{\theta})}{S_n^{*(0)}(t,v;\widehat{\beta},\widehat{\theta})}\Big\} \\
&\times Y_i(t)\lambda_0(t,v)\exp\{\beta(\theta_0^T v)^T Z_i\}dtdv \\
= \oint_0^1\int_0^\tau v_2 S_n^{*(1)}(t,v;\widehat{\beta},\widehat{\beta}',\widehat{\theta},\widehat{\theta})&\Big\{\frac{S_n^{*(0)}(t,v;\widehat{\beta},\widehat{\theta}) - S_n^{*(0)}(t,v;\beta,\theta_0)}{S_n^{*(0)}(t,v;\widehat{\beta},\widehat{\theta})}\Big\} \\
&\times \lambda_0(t,v)dtdv \\
- \oint_0^1\int_0^\tau v_2\Big\{&S_n^{*(1)}(t,v;\widehat{\beta},\widehat{\beta}',\widehat{\theta},\widehat{\theta}) - S_n^{*(1)}(t,v;\beta,\widehat{\beta}',\theta_0,\widehat{\theta})\Big\}\lambda_0(t,v)dtdv. \quad (A.22)
\end{aligned}$$

Following the arguments in the proof of Theorem 1, it can also be shown that

$b'_n = \sup_{u \in [\iota_1, \iota_2]} ||h\widehat{\beta}'(u) - h\beta'(u)|| \xrightarrow{p} 0$. Together with Theorem 1, applying Taylor expansions, we get

$$\begin{aligned} S_n^{*(0)}(t, v; \widehat{\beta}, \widehat{\theta}) &= S_n^{*(0)}(t, v; \beta, \theta_0) + O_p(a_n + b_n), \\ S_n^{*(1)}(t, v; \widehat{\beta}, \widehat{\beta}', \widehat{\theta}, \widehat{\theta}) &= S_n^{*(1)}(t, v; \beta, \beta', \theta_0, \theta_0) + O_p(a_n + b_n + h^{-1}b'_n). \end{aligned} \tag{A.23}$$

Under Condition A.8, we have

$$\begin{aligned} S_n^{*(0)}(t, v; \beta, \theta_0) &= s^{*(0)}(t, \theta_0^T v) + o_p(1), \\ S_n^{*(1)}(t, v; \beta, \beta', \theta_0, \theta_0) &= s^{*(1)}(t, \theta_0^T v) + o_p(1). \end{aligned} \tag{A.24}$$

By Taylor expansions, we obtain

$$\begin{aligned} & S_n^{*(0)}(t, v; \widehat{\beta}, \widehat{\theta}) - S_n^{*(0)}(t, v; \beta, \theta_0) \\ = \quad & n^{-1} \sum_{k=1}^{n} Y_k(t) exp\{\beta(\theta_0^T v)^T Z_k\}[\beta'(\theta_0^T v)^T Z_k] v^T(\widehat{\theta} - \theta_0) \\ & + n^{-1} \sum_{k=1}^{n} Y_k(t) exp\{\beta(\theta_0^T v)^T Z_k\} Z_k \{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\} + O_p((a_n + b_n)^2) \\ = \quad & E\Big[P(t|Z) exp\{\beta(\theta_0^T v)^T Z\}[\beta'(\theta_0^T v)^T Z]\Big] v^T(\widehat{\theta} - \theta_0) \\ & + E\Big[P(t|Z) exp\{\beta(\theta_0^T v)^T Z\} Z\Big] \{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\} + O_p((a_n + b_n)^2) \\ = \; & s^{*(1)}(t, \theta_0^T v) v^T(\widehat{\theta} - \theta_0) + \widetilde{s}^{(1)}(t, \beta, \theta_0^T v)^T \{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\} + O_p((a_n + b_n)^2), \end{aligned} \tag{A.25}$$

and

$$
\begin{aligned}
& S_n^{*(1)}(t, v; \widehat{\beta}, \widehat{\beta}', \widehat{\theta}, \widehat{\theta}) - S_n^{*(1)}(t, v; \beta, \widehat{\beta}', \theta_0, \widehat{\theta}) \\
& \quad = \quad n^{-1}\sum_{k=1}^{n} Y_k(t) exp\{\beta(\theta_0^T v)^T Z_k\}[\beta'(\theta_0^T v)^T Z_k]^{\otimes 2} v^T(\widehat{\theta} - \theta_0) \\
& \qquad + n^{-1}\sum_{k=1}^{n} Y_k(t) exp\{\beta(\theta_0^T v)^T Z_k\}[\beta'(\theta_0^T v)^T Z_k] Z_k\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\} \\
& \qquad + O_p\{(a_n + b_n)(a_n + b_n + h^{-1} b_n')\} \\
& \quad = \quad E\left[P(t|Z) exp\{\beta(\theta_0^T v)^T Z\}[\beta'(\theta_0^T v)^T Z]^{\otimes 2}\right] v^T(\widehat{\theta} - \theta_0) \\
& \qquad + E\left[P(t|Z) exp\{\beta(\theta_0^T v)^T Z\}[\beta'(\theta_0^T v)^T Z] Z\right]\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\} \\
& \qquad + O_p\{(a_n + b_n)(a_n + b_n + h^{-1} b_n')\} \\
& \quad = s^{*(2)}(t, \theta_0^T v) v^T(\widehat{\theta} - \theta_0) + \phi(t, \theta_0^T v)^T\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\} \\
& \qquad + O_p\{(a_n + b_n)(a_n + b_n + h^{-1} b_n')\}.
\end{aligned} \tag{A.26}
$$

By the application of Lemma 2 (D.2) in Gilbert et al. (2008), plugging (A.23), (A.24), (A.25) and (A.26) into (A.22), simple algebra gives that

$$
\begin{aligned}
III = \quad & \int_0^{\tau} \oint_0^1 v_2 \frac{s^{*(1)}(t, \theta_0^T v)}{s^{*(0)}(t, \theta_0^T v)} \Big\{ s^{*(1)}(t, \theta_0^T v) v^T(\widehat{\theta} - \theta_0) \\
& \qquad + \widetilde{s}^{(1)}(t, \beta, \theta_0^T v)^T\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\}\Big\} \lambda_0(t, v) dv dt \\
& - \int_0^{\tau} \oint_0^1 v_2 \Big\{ s^{*(2)}(t, \theta_0^T v) v^T(\widehat{\theta} - \theta_0) \\
& \qquad + \phi(t, \theta_0^T v)^T\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\}\Big\} \lambda_0(t, v) dv dt \\
& + O_p((a_n + b_n)(a_n + b_n + h^{-1} b_n')).
\end{aligned} \tag{A.27}
$$

We now deal with I. The process I is a locally square integrable martingale. By

(A.23) and (A.24), from Lemma 2 (D.2) in Gilbert et al. (2008), it can be shown that

$$\begin{aligned} I = \quad & \frac{1}{n}\oint_0^1 \int_0^\tau v_2\Big(\widehat{\beta}'(\theta^T v)^T - \beta'(\theta_0^T v)^T\Big)\Big\{\sum_{i=1}^n Z_i M_i(dt,dv)\Big\} \\ & - \frac{1}{n}\oint_0^1 \int_0^\tau v_2\Big(\frac{S_n^{*(1)}(t,v;\widehat{\beta},\widehat{\beta}',\widehat{\theta},\widehat{\theta})}{S_n^{*(0)}(t,v;\widehat{\beta},\widehat{\theta})} - \frac{S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)}{S_n^{*(0)}(t,v;\beta,\theta_0)}\Big)\Big\{\sum_{i=1}^n M_i(dt,dv)\Big\} \\ = \quad & o_p(n^{-1/2}). \end{aligned} \tag{A.28}$$

Substituting (A.21), (A.27) and (A.28) into (A.20), we can obtain

$$\begin{aligned} & Q_{n2}(\widehat{\beta},\widehat{\beta}',\widehat{\theta}) - Q_{n2}(\beta,\beta',\theta_0) \\ = \quad & A_\theta(\widehat{\theta}_2 - \theta_{20}) + \oint_0^1 \int_0^\tau v_2 B(t,\theta_0^T v)\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\}\lambda_0(t,v)dtdv \\ & + o_p(n^{-1/2}) + O_p((a_n + b_n)(a_n + b_n + h^{-1}b_n')), \end{aligned} \tag{A.29}$$

where

$$\begin{aligned} A_\theta = \quad & \oint_0^1 \int_0^\tau v_2\Big\{\frac{s^{*(1)}(t,\theta_0^T v)^{\otimes 2}}{s^{*(0)}(t,\theta_0^T v)} - s^{*(2)}(t,\theta_0^T v)\Big\}v_2^T \lambda_0(t,v)dtdv, \\ B(t,\theta_0^T v) = & \Big\{\frac{s^{*(1)}(t,\theta_0^T v)\widetilde{s}^{(1)}(t,\beta,\theta_0^T v)^T}{s^{*(0)}(t,\theta_0^T v)} - \phi(t,\theta_0^T v)^T\Big\}. \end{aligned} \tag{A.30}$$

Now we consider the asymptotic expression of $Q_{n2}(\beta,\beta',\theta_0)$. Observe that

$$
\begin{aligned}
&Q_{n2}(\beta,\beta',\theta_0)\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau v_2\Big\{\beta'(\theta_0^T v)^T Z_i - \frac{S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)}{S_n^{*(0)}(t,v;\beta,\theta_0)}\Big\}M_i(dt,dv)\\
&\qquad+\frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau v_2\Big\{\beta'(\theta_0^T v)^T Z_i - \frac{S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)}{S_n^{*(0)}(t,v;\beta,\theta_0)}\Big\}\\
&\qquad\qquad\times Y_i(t)\lambda_0(t,v)\exp\{\beta(\theta_0^T v)^T Z_i\}dtdv\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau v_2\Big\{\beta'(\theta_0^T v)^T Z_i - \frac{S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)}{S_n^{*(0)}(t,v;\beta,\theta_0)}\Big\}M_i(dt,dv)\\
&\qquad+\oint_0^1\int_0^\tau v_2\Big\{S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)\\
&\qquad\qquad-\frac{S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)}{S_n^{*(0)}(t,v;\beta,\theta_0)}S_n^{*(0)}(t,v;\beta,\theta_0)\Big\}\lambda_0(t,v)dtdv\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau v_2\Big\{\beta'(\theta_0^T v)^T Z_i - \frac{S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)}{S_n^{*(0)}(t,v;\beta,\theta_0)}\Big\}M_i(dt,dv).
\end{aligned}
\tag{A.31}
$$

By uniform consistency of $S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)$ and $S_n^{*(0)}(t,v;\beta,\theta_0)$, (A.24), it follows from Lemma 2 (D.2) in Gilbert et al. (2008) that

$$
\begin{aligned}
Q_{n2}(\beta,\beta',\theta_0) &= \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau v_2\Big\{\beta'(\theta_0^T v)^T Z_i - \frac{S_n^{*(1)}(t,v;\beta,\beta',\theta_0,\theta_0)}{S_n^{*(0)}(t,v;\beta,\theta_0)}\Big\}M_i(dt,dv)\\
&= \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau v_2\Big\{\beta'(\theta_0^T v)^T Z_i - \frac{s^{*(1)}(t,\theta_0^T v)}{s^{*(0)}(t,\theta_0^T v)}\Big\}M_i(dt,dv)\\
&\quad+o_p(n^{-1/2}).
\end{aligned}
\tag{A.32}
$$

Under Condition A, substituting (A.32) into (A.29), with the fact that

$Q_{n2}(\widehat{\beta}, \widehat{\beta}', \widehat{\theta}) = 0$, we have

$$\begin{aligned}
\widehat{\theta}_2 - \theta_{20} =& -\frac{1}{n} A_\theta^{-1} \sum_{i=1}^{n} \oint_0^1 \int_0^\tau v_2 \Big\{ \beta'(\theta_0^T v)^T Z_i - \frac{s^{*(1)}(t, \theta_0^T v)}{s^{*(0)}(t, \theta_0^T v)} \Big\} M_i(dt, dv) \\
& - A_\theta^{-1} \oint_0^1 W(v) \{ \widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v) \} dv + o_p(n^{-1/2}) \\
& + O_p((a_n + b_n)(a_n + b_n + h^{-1} b_n')), \qquad \text{(A.33)}
\end{aligned}$$

where

$$W(v) = v_2 \int_0^\tau B(t, \theta_0^T v) \lambda_0(t, v) dt, \qquad \text{(A.34)}$$

and A_θ and $B(t, \theta_0^T v)$ are defined in (A.30) as

$$\begin{aligned}
A_\theta =& \quad \oint_0^1 \int_0^\tau v_2 \Big\{ \frac{s^{*(1)}(t, \theta_0^T v)^{\otimes 2}}{s^{*(0)}(t, \theta_0^T v)} - s^{*(2)}(t, \theta_0^T v) \Big\} v_2^T \lambda_0(t, v) dt dv, \\
B(t, \theta_0^T v) =& \Big\{ \frac{s^{*(1)}(t, \theta_0^T v) \widetilde{s}^{(1)}(t, \beta, \theta_0^T v)^T}{s^{*(0)}(t, \theta_0^T v)} - \phi(t, \theta_0^T v)^T \Big\}.
\end{aligned}$$

The proof of Lemma 2 is completed.

Before proceeding with stating and proving Lemma 3, recall the definition of $\mathbb{M}_0 = \Big\{v|v \in [0,1]^d, \theta_0^T v = u\Big\}$, which is the region of integration for d-dimensional vector v that will be used throughout the proof of Lemma 3.

Lemma 3. *Under the mark-specific proportional hazards model (2.1) and Condition A, the asymptotic approximation of* $\widehat{\beta}(u) - \beta(u)$ *over* $u \in [\iota_1, \iota_2]$ *can be written as*

$$\begin{aligned}
&\Omega(u)\{\widehat{\beta}(u) - \beta(u)\} \\
&= \zeta(u)(\widehat{\theta}_2 - \theta_{20}) + \frac{1}{2}h^2\mu_2\Omega(u)\beta''(u) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \oint_0^1 \int_0^\tau K_h(\theta_0^T v - u)\Big[Z_i - \frac{s_p^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{s^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\Big] M_i(dt,dv) \\
&\quad + O_p\Big\{\frac{a_n^2}{h} + b_n^2 + b_n'^2 + c_n\Big\} + o_p(h^2 + (nh)^{-1/2}),
\end{aligned}$$

where a_n*,* b_n *and* b_n' *are defined in Lemma 2,* c_n *is defined in Theorem 1,* $s_p^{(1)}(t,v;\beta,h\beta',u,\theta_0)$ *consists of the first* p *components of* $s^{(1)}(t,v;\beta,h\beta',u,\theta_0)$*, and*

$$\begin{aligned}
\mu_2 &= \int u^2 K(u)du, \\
\zeta(u) &= \oint_{\mathbb{M}_0}\int_0^\tau \Big\{\frac{\widetilde{s}^{(1)}(t,\beta,u)s^{*(1)}(t,u)}{\widetilde{s}^{(0)}(t,\beta,u)} - \phi(t,u) - \frac{\partial \widetilde{s}^{(1)}(t,\beta,u)}{\partial u}\Big\} v_2^T \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times \lambda_0(t,v)dtdv, \\
\Omega(u) &= \oint_{\mathbb{M}_0}\int_0^\tau \Big\{\widetilde{s}^{(2)}(t,\beta,u) - \frac{\widetilde{s}^{(1)}(t,\beta,u)\widetilde{s}^{(1)}(t,\beta,u)^T}{\widetilde{s}^{(0)}(t,\beta,u)}\Big\}\lambda_0(t,v)dtdv.
\end{aligned}$$

Proof. To prove Lemma 3, we establish the asymptotic expansions of $Q_{n1}(\widehat{\beta},\widehat{\beta}',u,\widehat{\theta}) - Q_{n1}(\beta,\beta',u,\theta_0)$ and $Q_{n1}(\beta,\beta',u,\theta_0)$.

First, we decompose $Q_{n1}(\widehat{\beta},\widehat{\beta}',u,\widehat{\theta}) - Q_{n1}(\beta,\beta',u,\theta_0)$ into the following five parts and derive the asymptotic expression for each part.

$$
\begin{aligned}
&Q_{n1}(\widehat{\beta}, \widehat{\beta}', u, \widehat{\theta}) - Q_{n1}(\beta, \beta', u, \theta_0) \\
&= \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau \Bigg[K_h(\widehat{\theta}^T v - u)\widetilde{Z}_i(u, \widehat{\theta}^T v) - K_h(\widehat{\theta}^T v - u)\frac{S_n^{(1)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})}{\widetilde{S}_n^{(0)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})} \\
&\qquad - K_h(\theta_0^T v - u)\widetilde{Z}_i(u, {\theta_0}^T v) + K_h(\theta_0^T v - u)\frac{S_n^{(1)}(t, v; \beta, h\beta', u, \theta_0)}{S_n^{(0)}(t, v; \beta, h\beta', u, \theta_0)} \Bigg] N_i(dt, dv) \\
&= \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau \Bigg[K_h(\widehat{\theta}^T v - u)\widetilde{Z}_i(u, \widehat{\theta}^T v) - K_h(\theta_0^T v - u)\widetilde{Z}_i(u, {\theta_0}^T v) \Bigg] \\
&\qquad\qquad \times Y_i(t) exp\{\beta(\theta_0^T v)^T Z_i\}\lambda_0(t, v) dt dv \\
&- \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau K_h(\theta_0^T v - u)\Bigg[\frac{S_n^{(1)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})}{S_n^{(0)}(t, v; \beta, h\beta', u, \theta_0)} - \frac{S_n^{(1)}(t, v; \beta, h\beta', u, \theta_0)}{S_n^{(0)}(t, v; \beta, h\beta', u, \theta_0)} \Bigg] \\
&\qquad\qquad \times Y_i(t) exp\{\beta(\theta_0^T v)^T Z_i\}\lambda_0(t, v) dt dv \\
&- \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau K_h(\theta_0^T v - u)\Bigg[\frac{S_n^{(1)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})}{S_n^{(0)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})} - \frac{S_n^{(1)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})}{S_n^{(0)}(t, v; \beta, h\beta', u, \theta_0)} \Bigg] \\
&\qquad\qquad \times Y_i(t) exp\{\beta(\theta_0^T v)^T Z_i\}\lambda_0(t, v) dt dv \\
&+ \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau \Bigg[K_h(\theta_0^T v - u) - K_h(\widehat{\theta}^T v - u) \Bigg] \frac{S_n^{(1)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})}{S_n^{(0)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})} \\
&\qquad\qquad \times Y_i(t) exp\{\beta(\theta_0^T v)^T Z_i\}\lambda_0(t, v) dt dv \\
&+ \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau \Bigg[K_h(\widehat{\theta}^T v - u)\widetilde{Z}_i(u, \widehat{\theta}^T v) - K_h(\widehat{\theta}^T v - u)\frac{S_n^{(1)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})}{\widetilde{S}_n^{(0)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})} \\
&\qquad - K_h(\theta_0^T v - u)\widetilde{Z}_i(u, {\theta_0}^T v) + K_h(\theta_0^T v - u)\frac{S_n^{(1)}(t, v; \beta, h\beta', u, \theta_0)}{S_n^{(0)}(t, v; \beta, h\beta', u, \theta_0)} \Bigg] M_i(dt, dv) \\
&\equiv I + II + III + IV + V. \qquad\qquad \text{(A.35)}
\end{aligned}
$$

Let us start from I. Under Condition A.1 and A.6, applying a Taylor expansion around $\widehat{\theta}^T v = \theta_0^T v$, it can be shown that

$$\begin{aligned}
&K_h(\widehat{\theta}^T v - u)\widetilde{Z}_i(u, \widehat{\theta}^T v) - K_h(\theta_0^T v - u)\widetilde{Z}_i(u, \theta_0^T v) \\
&= \frac{\partial}{\partial(\theta^T v)}\left\{K_h(\theta_0^T v - u)\widetilde{Z}_i(u, \theta_0^T v)\right\} v^T(\widehat{\theta} - \theta_0) + O_p\left\{\frac{a_n^2}{h}\right\} \\
&= \left\{\frac{\partial K_h(\theta_0^T v - u)}{\partial(\theta^T v)}\widetilde{Z}_i(u, \theta_0^T v) + K_h(\theta_0^T v - u)\frac{\partial \widetilde{Z}_i(u, \theta_0^T v)}{\partial(\theta^T v)}\right\} v^T(\widehat{\theta} - \theta_0) \\
&\quad + O_p\left\{\frac{a_n^2}{h}\right\} \\
&= -\left\{\frac{\partial K_h(\theta_0^T v - u)}{\partial u}\widetilde{Z}_i(u, \theta_0^T v) + K_h(\theta_0^T v - u)\frac{\partial \widetilde{Z}_i(u, \theta_0^T v)}{\partial u}\right\} v^T(\widehat{\theta} - \theta_0) + O_p\left\{\frac{a_n^2}{h}\right\} \\
&= -\frac{\partial}{\partial u}\left\{K_h(\theta_0^T v - u)\widetilde{Z}_i(u, \theta_0^T v)\right\} v^T(\widehat{\theta} - \theta_0) + O_p\left\{\frac{a_n^2}{h}\right\}. \qquad \text{(A.36)}
\end{aligned}$$

Let $\mathbf{I}_p$ be the a p-order diagonal matrix with p elements 1 and $\mathbf{0}$ be a p $\times$ p zero matrix. Denote $\mathbf{e} = \{\mathbf{I}_p, \mathbf{0}\}^T$. Substituting (A.36) into I, under Condition A, by the Theorem 2 in Hansen (2008), the asymptotic expression of I can be written as

$$\begin{aligned}
I &= \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau\Bigg\{\Big[K_h(\widehat{\theta}^T v - u)\widetilde{Z}_i(u, \widehat{\theta}^T v) - K_h(\theta_0^T v - u)\widetilde{Z}_i(u, \theta_0^T v)\Big] \\
&\qquad \times Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\Bigg\}\lambda_0(t, v)dtdv \\
&= -\frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau\left\{\frac{\partial}{\partial u}\Big[K_h(\theta_0^T v - u)\widetilde{Z}_i(u, \theta_0^T v)\Big]Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\right\} v^T(\widehat{\theta} - \theta_0) \\
&\qquad \times \lambda_0(t, v)dtdv + O_p\left\{\frac{a_n^2}{h}\right\} \\
&= -E\left\{\oint_0^1\int_0^\tau\frac{\partial}{\partial u}\Big[K_h(\theta_0^T v - u)\widetilde{Z}_i(u, \theta_0^T v)\Big]Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\} v^T\lambda_0(t, v)dtdv\right\} \\
&\qquad \times (\widehat{\theta} - \theta_0) + O_p\left\{\frac{a_n^2}{h} + \sqrt{log(n)/(nh)}\right\} \\
&= -\oint_{\mathbb{M}_0}\int_0^\tau \mathbf{e}\frac{\partial}{\partial u}E\left\{Z_i Y_i(t)exp\{\beta(u)^T Z_i\}\right\} v^T\lambda_0(t, v)dtdv(\widehat{\theta} - \theta_0) + O_p\left\{\frac{a_n^2}{h} + c_n\right\} \\
&= -\oint_{\mathbb{M}_0}\int_0^\tau \mathbf{e}\frac{\partial \widetilde{s}^{(1)}(t, \beta, u)}{\partial u} v^T\lambda_0(t, v)dtdv(\widehat{\theta} - \theta_0) + O_p\left\{\frac{a_n^2}{h} + c_n\right\}. \qquad \text{(A.37)}
\end{aligned}$$

Next, we deal with II. Let $\xi(u) = \{\widehat{\beta}(u)^T - \beta(u)^T, h[\widehat{\beta}'(u)^T - \beta'(u)^T]\}^T$. Using Taylor expansions with the uniform consistent properties of $\widehat{\beta}$ and $\widehat{\theta}$, we get

$$
\begin{aligned}
& S_n^{(1)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta}) - S_n^{(1)}(t, v; \beta, h\beta', u, \theta_0) \\
= \quad & \Bigg[n^{-1} \sum_{k=1}^{n} Y_k(t) \exp\{\beta(u)^T Z_k + \beta'(u)^T Z_k(\theta_0^T v - u)\} \widetilde{Z}_k(u, \theta_0^T v) \\
& \qquad \times \left\{ (\beta'(u)^T Z_k) + \frac{\partial \widetilde{Z}_k(u, \theta_0^T v)^T}{\partial(\theta^T v)} \right\} \Bigg] v^T(\widehat{\theta} - \theta_0) \\
& + S_n^{(2)}(t, v; \beta, h\beta', u, \theta_0)\xi(u) + O_p\left\{a_n^2 + b_n^2 + b_n'^2\right\} \\
= \quad & \phi^{(n)}(t, \theta_0^T v; \beta, h\beta', u) v^T(\widehat{\theta} - \theta_0) \\
& + S_n^{(2)}(t, v; \beta, h\beta', u, \theta_0)\xi(u) + O_p\left\{a_n^2 + b_n^2 + b_n'^2\right\},
\end{aligned}
\tag{A.38}
$$

where

$$
\begin{aligned}
\phi^{(n)}(t, \theta_0^T v; \beta, h\beta', u) = n^{-1} \sum_{k=1}^{n} Y_k(t) \exp\{\beta(u)^T Z_k + \beta'(u)^T Z_k(\theta_0^T v - u)\} \widetilde{Z}_k(u, \theta_0^T v) \\
\times \left\{ (\beta'(u)^T Z_k) + \frac{\partial \widetilde{Z}_k(u, \theta_0^T v)^T}{\partial(\theta^T v)} \right\}.
\end{aligned}
$$

Under Condition A.8, plugging (A.38) into II, by the application of Theorem 2 in

Hansen (2008), it can be shown that

$$
\begin{aligned}
&II \\
&= -\frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau\Bigg\{K_h(\theta_0^T v - u)\left[\frac{S_n^{(1)}(t,v;\widehat{\beta},h\widehat{\beta}',u,\widehat{\theta})}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)} - \frac{S_n^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right] \\
&\qquad\qquad \times Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\Bigg\}\lambda_0(t,v)dtdv \\
&= -\frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau\Bigg\{K_h(\theta_0^T v - u) \\
&\qquad \times\left[\frac{\phi^{(n)}(t,\theta_0^T v;\beta,h\beta',u)v^T(\widehat{\theta}-\theta_0) + S_n^{(2)}(t,v;\beta,h\beta',u,\theta_0)\xi(u)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right] \\
&\qquad\qquad \times Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\Bigg\}\lambda_0(t,v)dtdv \\
&\quad + O_p\left\{a_n^2 + b_n^2 + b_n'^2\right\} \\
&= -E\Bigg\{\oint_0^1\int_0^\tau\Bigg\{K_h(\theta_0^T v - u)\frac{1}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)} \\
&\qquad \times\left[\phi^{(n)}(t,\theta_0^T v;\beta,h\beta',u)v^T(\widehat{\theta}-\theta_0) + S_n^{(2)}(t,v;\beta,h\beta',u,\theta_0)\xi(u)\right] \\
&\qquad\qquad \times Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\Bigg\}\lambda_0(t,v)dtdv\Bigg\} \\
&\quad + O_p\left\{a_n^2 + b_n^2 + b_n'^2 + \sqrt{log(n)/(nh)}\right\} \\
&= -\oint_{\mathbb{M}_0}\int_0^\tau E\Bigg\{\frac{1}{\widetilde{s}^{(0)}(t,\beta,u)} \\
&\qquad \times\left[\mathbf{e}\phi(t,u)v^T(\widehat{\theta}-\theta_0) + \begin{pmatrix}\widetilde{s}^{(2)}(t,\beta,u) & 0 \\ 0 & \mu_2\widetilde{s}^{(2)}(t,\beta,u)\end{pmatrix}\xi(u)\right] \\
&\qquad\qquad \times Y_i(t)exp\{\beta(u)^T Z_i\}\Bigg\}\lambda_0(t,v)dtdv \\
&\quad + O_p\left\{a_n^2 + b_n^2 + b_n'^2 + c_n\right\}, \qquad (A.39)
\end{aligned}
$$

where

$$\phi(t,u) = E\Big[P(t|Z)exp\{\beta(u)^T Z\}Z\Big(\beta'(u)^T Z\Big)\Big].$$

By the law of large number, we have

$$
\begin{aligned}
&II \\
&= -\oint_{\mathbb{M}_0}\int_0^\tau \left[\mathbf{e}\phi(t,u)v^T(\widehat{\theta}-\theta_0) + \begin{pmatrix} \widetilde{s}^{(2)}(t,\beta,u) & 0 \\ 0 & \mu_2\widetilde{s}^{(2)}(t,\beta,u)\end{pmatrix}\xi(u)\right] \\
&\qquad\qquad \times \frac{1}{\widetilde{s}^{(0)}(t,\beta,u)}E\big\{Y_i(t)exp\{\beta(u)^T Z_i\}\big\}\lambda_0(t,v)dtdv \\
&\quad + O_p\left\{a_n^2 + b_n^2 + b_n'^2 + c_n\right\} \\
&= -\oint_{\mathbb{M}_0}\int_0^\tau \left[\mathbf{e}\phi(t,u)v^T(\widehat{\theta}-\theta_0) + \begin{pmatrix} \widetilde{s}^{(2)}(t,\beta,u) & 0 \\ 0 & \mu_2\widetilde{s}^{(2)}(t,\beta,u)\end{pmatrix}\xi(u)\right]\lambda_0(t,v)dtdv \\
&\quad + O_p\left\{a_n^2 + b_n^2 + b_n'^2 + c_n\right\}. \qquad (A.40)
\end{aligned}
$$

We now show the asymptotic expression of III. By Taylor expansions and Theorem 1, we have

$$
\begin{aligned}
S_n^{(0)}(t,v;\widehat{\beta},h\widehat{\beta}',u,\widehat{\theta}) &= S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0) + O_p(a_n + b_n + b_n'), \\
S_n^{(1)}(t,v;\widehat{\beta},h\widehat{\beta}',u,\widehat{\theta}) &= S_n^{(1)}(t,v;\beta,h\beta',u,\theta_0) + O_p(a_n + b_n + b_n'), \qquad (A.41)
\end{aligned}
$$

and

$$
\begin{aligned}
& S_n^{(0)}(t,v;\widehat{\beta},h\widehat{\beta}',u,\widehat{\theta}) - S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0) \\
&\quad = \quad n^{-1}\sum_{k=1}^{n} Y_k(t)\exp\{\beta(u)^T Z_k + \beta'(u)^T Z_k(\theta_0^T v - u)\}(\beta'(u)^T Z_k) v^T(\widehat{\theta}-\theta_0) \\
&\qquad + S_n^{(1)}(t,v;\beta,h\beta',u,\theta_0)^T \xi(u) + O_p\left\{a_n^2 + b_n^2 + b_n'^2\right\} \\
&\quad = \quad \phi_1^{(n)}(t,\theta_0^T v;\beta,h\beta',u) v^T(\widehat{\theta}-\theta_0) + S_n^{(1)}(t,v;\beta,h\beta',u,\theta_0)^T \xi(u) \\
&\qquad + O_p\left\{a_n^2 + b_n^2 + b_n'^2\right\}, \qquad\qquad \text{(A.42)}
\end{aligned}
$$

where

$$
\begin{aligned}
\phi_1^{(n)}(t,\theta_0^T v;\beta,h\beta',u) = n^{-1}\sum_{k=1}^{n} Y_k(t)\exp\{\beta(u)^T Z_k + \beta'(u)^T Z_k(\theta_0^T v - u)\} \\
\times (\beta'(u)^T Z_k).
\end{aligned}
$$

Under Condition A.8, substituting (A.41) and (A.42) into III, by the Theorem 2 in

Hansen (2008), it can be shown that

$$
\begin{aligned}
III = &\frac{1}{n}\sum_{i=1}^{n}\oint_{0}^{1}\int_{0}^{\tau}\Bigg\{K_h(\theta_0^T v - u)S_n^{(1)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta}) \\
&\times \left[\frac{S_n^{(0)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta}) - S_n^{(0)}(t, v; \beta, h\beta', u, \theta_0)}{S_n^{(0)}(t, v; \widehat{\beta}, h\widehat{\beta}', u, \widehat{\theta})S_n^{(0)}(t, v; \beta, h\beta', u, \theta_0)}\right] \\
&\times Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\Bigg\}\lambda_0(t, v)dtdv \\
= &\frac{1}{n}\sum_{i=1}^{n}\oint_{0}^{1}\int_{0}^{\tau}\Bigg\{K_h(\theta_0^T v - u)S_n^{(1)}(t, v; \beta, h\beta', u, \theta_0) \\
&\times \left[\frac{\phi_1^{(n)}(t, \theta_0^T v; \beta, h\beta', u)v^T(\widehat{\theta} - \theta_0) + S_n^{(1)}(t, v; \beta, h\beta', u, \theta_0)^T\xi(u)}{S_n^{(0)}(t, v; \beta, h\beta', u, \theta_0)^{\otimes 2}}\right] \\
&\times Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\Bigg\}\lambda_0(t, v)dtdv \\
&+ O_p\left\{a_n^2 + b_n^2 + b_n'^2\right\} \\
= &\oint_{\mathbb{M}_0}\int_0^{\tau} E\Bigg\{\mathbf{e}\widetilde{s}^{(1)}(t, \beta, u) \\
&\times \left[\frac{s^{*(1)}(t, u)v^T(\widehat{\theta} - \theta_0) + \{\widetilde{s}^{(1)}(t, \beta, u)^T, 0^T\}\xi(u)}{\widetilde{s}^{(0)}(t, \beta, u)^{\otimes 2}}\right] \\
&\times Y_i(t)exp\{\beta(u)^T Z_i\}\Bigg\}\lambda_0(t, v)dtdv \\
&+ O_p\left\{a_n^2 + b_n^2 + b_n'^2 + c_n\right\} \\
= &\oint_{\mathbb{M}_0}\int_0^{\tau} E\Bigg\{\mathbf{e}\widetilde{s}^{(1)}(t, \beta, u) \\
&\times \left[\frac{s^{*(1)}(t, u)v^T(\widehat{\theta} - \theta_0) + \widetilde{s}^{(1)}(t, \beta, u)^T\{\widehat{\beta}(u) - \beta(u)\}}{\widetilde{s}^{(0)}(t, \beta, u)^{\otimes 2}}\right] \\
&\times Y_i(t)exp\{\beta(u)^T Z_i\}\Bigg\}\lambda_0(t, v)dtdv \\
&+ O_p\left\{a_n^2 + b_n^2 + b_n'^2 + c_n\right\}. \qquad (A.43)
\end{aligned}
$$

By large number theory, we have

$$III = \oint_{\mathbb{M}_0} \int_0^\tau \mathbf{e}\widetilde{s}^{(1)}(t,\beta,u) \times \left[\frac{s^{*(1)}(t,u)v^T(\widehat{\theta}-\theta_0) + \widetilde{s}^{(1)}(t,\beta,u)^T\{\widehat{\beta}(u)-\beta(u)\}}{\widetilde{s}^{(0)}(t,\beta,u)^{\otimes 2}} \right] \times E\Big\{Y_i(t)exp\{\beta(u)^T Z_i\}\Big\}\lambda_0(t,v)dtdv$$

$$+ O_p\left\{a_n^2 + b_n^2 + b_n'^2 + c_n\right\}$$

$$= \oint_{\mathbb{M}_0} \int_0^\tau \mathbf{e}\widetilde{s}^{(1)}(t,\beta,u) \times \left[\frac{s^{*(1)}(t,u)v^T(\widehat{\theta}-\theta_0) + \widetilde{s}^{(1)}(t,\beta,u)^T\{\widehat{\beta}(u)-\beta(u)\}}{\widetilde{s}^{(0)}(t,\beta,u)} \right] \times \lambda_0(t,v)dtdv$$

$$+ O_p\left\{a_n^2 + b_n^2 + b_n'^2 + c_n\right\}. \tag{A.44}$$

Under Condition A, by a Taylor expansion, it can be shown that

$$IV = \frac{1}{n}\sum_{i=1}^n \oint_0^1 \int_0^\tau \Bigg\{ \left[K_h(\theta_0^T v - u) - K_h(\widehat{\theta}^T v - u)\right] \frac{S_n^{(1)}(t,v;\widehat{\beta},h\widehat{\beta}',u,\widehat{\theta})}{S_n^{(0)}(t,v;\widehat{\beta},h\widehat{\beta}',u,\widehat{\theta})} \times Y_i(t)exp\{\beta(\theta_0^T v)^T Z_i\}\Bigg\}\lambda_0(t,v)dtdv$$

$$= O_p\left\{\frac{a_n^2}{h}\right\}. \tag{A.45}$$

Finally, we decompose V into the following two parts.

$$
\begin{aligned}
V &= \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau\Bigg[K_h(\widehat{\theta}^Tv-u)\widetilde{Z}_i(v,u;\widehat{\theta})-K_h(\widehat{\theta}^Tv-u)\frac{S_n^{(1)}(t,v;\widehat{\beta},h\widehat{\beta}',u,\widehat{\theta})}{\widetilde{S}_n^{(0)}(t,v;\widehat{\beta},h\widehat{\beta}',u,\widehat{\theta})}\\
&\quad -K_h(\theta_0^Tv-u)\widetilde{Z}_i(v,u;\theta_0)+K_h(\theta_0^Tv-u)\frac{S_n^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\Bigg]M_i(dt,dv)\\
&= \frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau\Big[K_h(\widehat{\theta}^Tv-u)-K_h(\theta_0^Tv-u)\Big]\\
&\qquad\qquad \times\left[\widetilde{Z}_i(v,u;\theta_0)-\frac{S_n^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right]M_i(dt,dv)\\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau K_h(\widehat{\theta}^Tv-u)\Bigg[\Big(\widetilde{Z}_i(v,u;\widehat{\theta})-\widetilde{Z}_i(v,u;\theta_0)\Big)\\
&\qquad -\left(\frac{S_n^{(1)}(t,v;\widehat{\beta},h\widehat{\beta}',u,\widehat{\theta})}{\widetilde{S}_n^{(0)}(t,v;\widehat{\beta},h\widehat{\beta}',u,\widehat{\theta})}-\frac{S_n^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right)\Bigg]M_i(dt,dv)\\
&= \quad X_n(\tau)+Y_n(\tau). \qquad\qquad (A.46)
\end{aligned}
$$

Under Condition A, it can be shown that

$$
\begin{aligned}
X_n(\tau) &= o_p((nh)^{-1/2}),\\
Y_n(\tau) &= o_p((nh)^{-1/2}). \qquad\qquad (A.47)
\end{aligned}
$$

Substituting the asymptotic expression of I, II, III, IV and V, that is, (A.37), (A.40), (A.44), (A.45) and (A.47), respectively, into $Q_{n1}(\widehat{\beta},\widehat{\beta}',u,\widehat{\theta})-Q_{n1}(\beta,\beta',u,\theta_0)$, we obtain

$$
\begin{aligned}
&Q_{n1}(\widehat{\beta}, \widehat{\beta}', u, \widehat{\theta}) - Q_{n1}(\beta, \beta', u, \theta_0) \\
&= \mathbf{e}\zeta(u)(\widehat{\theta}_2 - \theta_{20}) - \oint_{\mathbb{M}_0} \int_0^\tau \begin{pmatrix} \widetilde{s}^{(2)}(t,\beta,u) & 0 \\ 0 & \mu_2 \widetilde{s}^{(2)}(t,\beta,u) \end{pmatrix} \xi(u)\lambda_0(t,v)dtdv \\
&\quad + \oint_{\mathbb{M}_0} \int_0^\tau \mathbf{e} \frac{\widetilde{s}^{(1)}(t,\beta,u)\widetilde{s}^{(1)}(t,\beta,u)^T}{\widetilde{s}^{(0)}(t,\beta,u)} \lambda_0(t,v)dtdv\{\widehat{\beta}(u) - \beta(u)\} \\
&\quad + O_p\left\{\frac{a_n^2}{h} + b_n^2 + b_n'^2 + c_n\right\} + o_p((nh)^{-1/2}), \qquad \text{(A.48)}
\end{aligned}
$$

where

$$
\begin{aligned}
\zeta(u) = \oint_{\mathbb{M}_0} \int_0^\tau \left\{ \frac{\widetilde{s}^{(1)}(t,\beta,u)s^{*(1)}(t,u)}{\widetilde{s}^{(0)}(t,\beta,u)} - \phi(t,u) - \frac{\partial \widetilde{s}^{(1)}(t,\beta,u)}{\partial u} \right\} v_2^T \\
\times \lambda_0(t,v)dtdv. \qquad \text{(A.49)}
\end{aligned}
$$

Let $Q_{n1,p}(\beta, \beta', u, \theta_0)$ and $S_{n,p}^{(1)}(t, v; \beta, h\beta', u, \theta_0)$ be the first p elements of $Q_{n1}(\beta, \beta', u, \theta_0)$ and $S_n^{(1)}(t, v; \beta, h\beta', u, \theta_0)$, respectively. We now derive the asymptotic expression of $Q_{n1,p}(\beta, \beta', u, \theta_0)$. Observe that

$$
\begin{aligned}
&Q_{n1,p}(\beta,\beta', u, \theta_0) \\
&\quad = \frac{1}{n}\sum_{i=1}^n \oint_0^1 \int_0^\tau K_h(\theta_0^T v - u)\left[Z_i - \frac{S_{n,p}^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right] N_i(dt,dv) \\
&\quad = \frac{1}{n}\sum_{i=1}^n \oint_0^1 \int_0^\tau K_h(\theta_0^T v - u)\left[Z_i - \frac{S_{n,p}^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right] M_i(dt,dv) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \oint_0^1 \int_0^\tau K_h(\theta_0^T v - u)\left[Z_i - \frac{S_{n,p}^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right] \\
&\qquad \times Y_i(t)\exp\{\beta(\theta_0^T v)^T Z_i\}\lambda_0(t,v)dtdv \\
&\quad \equiv R_n(u) + D_n(u). \qquad \text{(A.50)}
\end{aligned}
$$

We first consider $D_n(u)$. Note that

$$
\begin{aligned}
D_n(u) = \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau K_h(\theta_0^T v - u)\left[Z_i - \frac{S_{n,p}^{(1)}(t, v; \beta, h\beta', u, \theta_0)}{S_n^{(0)}(t, v; \beta, h\beta', u, \theta_0)}\right] Y_i(t) \\
\times \Big[\exp\{\beta(\theta_0^T v)^T Z_i\} - \exp\{\beta(u)^T Z_i + \beta'(u)^T Z_i(\theta_0^T v - u)\} \\
+ \exp\{\beta(u)^T Z_i + \beta'(u)^T Z_i(\theta_0^T v - u)\}\Big] \\
\times \lambda_0(t, v) dt dv \\
= \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau K_h(\theta_0^T v - u)\left[Z_i - \frac{S_{n,p}^{(1)}(t, v; \beta, h\beta', u, \theta_0)}{S_n^{(0)}(t, v; \beta, h\beta', u, \theta_0)}\right] Y_i(t) \\
\times \Big[\exp\{\beta(\theta_0^T v)^T Z_i\} - \exp\{\beta(u)^T Z_i + \beta'(u)^T Z_i(\theta_0^T v - u)\}\Big] \\
\times \lambda_0(t, v) dt dv.
\end{aligned} \tag{A.51}
$$

Using a Taylor expansion around $|\theta_0^T v - u| < h$, we have

$$
\begin{aligned}
&\exp\{\beta(\theta_0^T v)^T Z_i\} - \exp\{\beta(u)^T Z_i + \beta'(u)^T Z_i(\theta_0^T v - u)\} \\
&\quad = \frac{1}{2}\exp\{\beta(u)^T Z_i + \beta'(u)^T Z_i(\theta_0^T v - u)\}\beta''(u)^T Z_i(\theta_0^T v - u)^2 + o_p(h^2).
\end{aligned} \tag{A.52}
$$

Substituting (A.52) into $D_n(u)$, we get

$$\begin{aligned}
D_n(u) &= \frac{1}{2n}\sum_{i=1}^{n}\oint_0^1 \int_0^\tau K_h(\theta_0^T v - u)\left[Z_i - \frac{S_{n,p}^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right]Y_i(t) \\
&\qquad \times \left[\exp\{\beta(u)^T Z_i + \beta'(u)^T Z_i(\theta_0^T v - u)\}\beta''(u)^T Z_i(\theta_0^T v - u)^2\right] \\
&\qquad \times \lambda_0(t,v)dtdv + o_p(h^2) \\
&= \frac{1}{2}h^2 \oint_{\mathbb{M}_0}\int_0^\tau E\left\{\left[Z_i\mu_2 - \frac{\widetilde{s}^{(1)}(t,\beta,u)}{\widetilde{s}^{(0)}(t,\beta,u)}\mu_2\right]Y_i(t)\exp\{\beta(u)^T Z_i\}\beta''(u)^T Z_i\right\} \\
&\qquad \times \lambda_0(t,v)dtdv + o_p(h^2) \\
&= \frac{1}{2}h^2\mu_2 \oint_{\mathbb{M}_0}\int_0^\tau E\left\{\left[Z_i - \frac{\widetilde{s}^{(1)}(t,\beta,u)}{\widetilde{s}^{(0)}(t,\beta,u)}\right]Y_i(t)\exp\{\beta(u)^T Z_i\}Z_i\right\} \\
&\qquad \times \lambda_0(t,v)dtdv\beta''(u) + o_p(h^2) \\
&= \frac{1}{2}h^2\mu_2\Omega(u)\beta''(u) + o_p(h^2), \qquad (A.53)
\end{aligned}$$

where

$$\Omega(u) = \oint_{\mathbb{M}_0}\int_0^\tau \left\{\widetilde{s}^{(2)}(t,\beta,u) - \frac{\widetilde{s}^{(1)}(t,\beta,u)\widetilde{s}^{(1)}(t,\beta,u)^T}{\widetilde{s}^{(0)}(t,\beta,u)}\right\}\lambda_0(t,v)dtdv. \qquad (A.54)$$

Now, we want to show that $\sqrt{nh}R_n(u)$ is asymptotically normally distributed with mean zero and covariance matrix $\nu_0\Omega(u)$.

Observe that

$$\begin{aligned}
&\langle\sqrt{nh}R_n(u), \sqrt{nh}R_n(u)\rangle(\tau) \\
&\qquad = \frac{h}{n}\sum_{i=1}^{n}\oint_0^1\int_0^\tau K_h^2(\theta_0^T v - u)\left[Z_i - \frac{S_{n,p}^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right]^{\otimes 2} \\
&\qquad\qquad \times Y_i(t)\exp\{\beta(\theta_0^T v)^T Z_i\}\lambda_0(t,v)dtdv.
\end{aligned}$$

Under Condition A.6 and A.9, it follows from the application of Lemma 1 that

$$\begin{aligned}
&\lim_{n\to\infty} E\langle\sqrt{nh}R_n^*(u),\sqrt{nh}R_n^*(u)\rangle(\tau)\\
&\quad = \nu_0 \oint_{\mathbb{M}_0}\int_0^\tau E\left\{\left[Z-\frac{\widetilde{s}^{(1)}(t,\beta,u)}{\widetilde{s}^{(0)}(t,\beta,u)}\right]^{\otimes 2} P(t|Z)\exp\{\beta(u)^T Z\}\right\}\lambda_0(t,v)dtdv\\
&\quad = \nu_0\Omega(u).
\end{aligned}$$

Under Condition A, applying Lemma 2 (D.2) in Gilbert et al. (2008), it can be shown that

$$\begin{aligned}
&\sqrt{nh}R_n(u)\\
&= \sqrt{nh}\cdot\frac{1}{n}\sum_{i=1}^n\oint_0^1\int_0^\tau K_h(\theta_0^T v-u)\left[Z_i-\frac{S_{n,p}^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{S_n^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right]M_i(dt,dv)\\
&= \sqrt{\frac{h}{n}}\sum_{i=1}^n\oint_0^1\int_0^\tau K_h(\theta_0^T v-u)\left[Z_i-\frac{s_p^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{s^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right]M_i(dt,dv)\\
&\quad + o_p\{(nh)^{-1/2}\}. \qquad (A.55)
\end{aligned}$$

Using Martingale Central Limit Theorem, it follows that $\sqrt{nh}R_n(u)$ is asymptotically normally distributed with mean zero and covariance $\nu_0\Omega(u)$.

Substituting (A.50), the asymptotic expression of $Q_{n1,p}(\beta,\beta',u,\theta_0)$, into (A.48), together with the fact that $Q_{n1}(\widehat{\beta},\widehat{\beta}',u,\widehat{\theta})=0$, we can obtain

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\oint_0^1\int_0^\tau K_h(\theta_0^T v-u)\left[Z_i-\frac{s_p^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{s^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\right]M_i(dt,dv)\\
&= -\zeta(u)(\widehat{\theta}_2-\theta_{20})+\Omega(u)\{\widehat{\beta}(u)-\beta(u)\}-\frac{1}{2}h^2\mu_2\Omega(u)\beta''(u)\\
&\quad + O_p\left\{\frac{a_n^2}{h}+b_n^2+b_n'^2+c_n\right\}+o_p\{(nh)^{-1/2}+h^2\}, \qquad (A.56)
\end{aligned}$$

where $\Omega(u)$ is defined in (A.54) as

$$\Omega(u) = \oint_{\mathbb{M}_0} \int_0^{\tau} \left\{ \widetilde{s}^{(2)}(t, \beta, u) - \frac{\widetilde{s}^{(1)}(t, \beta, u) \widetilde{s}^{(1)}(t, \beta, u)^T}{\widetilde{s}^{(0)}(t, \beta, u)} \right\} \lambda_0(t, v) dt dv,$$

and $\zeta(u)$ is defined in (A.49) as

$$\zeta(u) = \oint_{\mathbb{M}_0} \int_0^{\tau} \left\{ \frac{\widetilde{s}^{(1)}(t, \beta, u) s^{*(1)}(t, u)}{\widetilde{s}^{(0)}(t, \beta, u)} - \phi(t, u) - \frac{\partial \widetilde{s}^{(1)}(t, \beta, u)}{\partial u} \right\} v_2^T \times \lambda_0(t, v) dt dv.$$

The proof of Lemma 3 is completed.

We shall finish the proofs of Theorem 2 and 3 by combining the results of Lemma 2 and 3.

Proof of Theorem 2.

Recall that the asymptotic expansion of $\widehat{\theta}_2 - \theta_{20}$ is shown in (A.33) as:

$$\begin{aligned}\widehat{\theta}_2 - \theta_{20} = & -\frac{1}{n}A_\theta^{-1}\sum_{i=1}^{n}\oint_0^1\int_0^\tau v_2\Big\{\beta'(\theta_0^T v)^T Z_i - \frac{s^{*(1)}(t,\theta_0^T v)}{s^{*(0)}(t,\theta_0^T v)}\Big\}M_i(dt,dv) \\ & - A_\theta^{-1}\oint_0^1 W(v)\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\}dv + o_p(n^{-1/2}) \\ & + O_p((a_n+b_n)(a_n+b_n+h^{-1}b_n')),\end{aligned}$$

where A_θ and $B(t,\theta_0^T v)$ are defined in (A.30), and $W(v)$ is defined in (A.34):

$$\begin{aligned}A_\theta = & \oint_0^1\int_0^\tau v_2\Big\{\frac{s^{*(1)}(t,\theta_0^T v)^{\otimes 2}}{s^{*(0)}(t,\theta_0^T v)} - s^{*(2)}(t,\theta_0^T v)\Big\}v_2^T\lambda_0(t,v)dtdv, \\ W(v) = & v_2\int_0^\tau B(t,\theta_0^T v)\lambda_0(t,v)dt, \\ B(t,\theta_0^T v) = & \left\{\frac{s^{*(1)}(t,\theta_0^T v)\widetilde{s}^{(1)}(t,\beta,\theta_0^T v)^T}{s^{*(0)}(t,\theta_0^T v)} - \phi(t,\theta_0^T v)^T\right\}.\end{aligned}$$

We first substitute the asymptotic expression of $\widehat{\theta}_2 - \theta_{20}$, (A.33), into (A.56), the asymptotic approximation of $\widehat{\beta}(u) - \beta(u)$:

$$\begin{aligned}G_n(u) = & \Omega(u)\{\widehat{\beta}(u) - \beta(u)\} - \frac{1}{2}h^2\mu_2\Omega(u)\beta''(u) \\ & + \zeta(u)A_\theta^{-1}\oint_0^1 W(v)\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\}dv \\ & + O_p\left\{\frac{a_n^2}{h} + b_n^2 + b_n'^2 + \frac{a_n b_n'}{h} + \frac{b_n b_n'}{h} + c_n\right\} + o_p\{(nh)^{-1/2} + h^2 + n^{-1/2}\},\end{aligned} \tag{A.57}$$

where

$$G_n(u) = \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau \Bigg[K_h(\theta_0^T v - u)\Bigg[Z_i - \frac{s_p^{(1)}(t,v;\beta,h\beta',u,\theta_0)}{s^{(0)}(t,v;\beta,h\beta',u,\theta_0)}\Bigg] - \zeta(u)A_\theta^{-1}v_2\Bigg\{\beta'(\theta_0^T v)^T Z_i - \frac{s^{*(1)}(t,\theta_0^T v)}{s^{*(0)}(t,\theta_0^T v)}\Bigg\}\Bigg] M_i(dt,dv),$$

and $s_p^{(1)}(t,v;\beta,h\beta',u,\theta_0)$ consists of the first p components of $s^{(1)}(t,v;\beta,h\beta',u,\theta_0)$.

Recall that the range of $\theta_0^T v$ is $[\iota_1,\iota_2]$, and we impose the restriction on the first element of θ, setting $\theta_1 = 1$. Define set $\mathbb{M}_\kappa = \Big\{v\Big|v \in [0,1]^d, \theta_0^T v = \kappa\Big\}$. Then, with Jacobian determinant $1/\|\theta_0\|$, it can be shown that

$$\oint_0^1 W(v)\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\}dv = \int_{\iota_1}^{\iota_2}\oint_{\mathbb{M}_\kappa} W(v)\{\widehat{\beta}(\kappa) - \beta(\kappa)\}\frac{1}{\|\theta_0\|}dv d\kappa. \tag{A.58}$$

Denote

$$\gamma(\kappa) = \oint_{\mathbb{M}_\kappa} W(v)\frac{1}{\|\theta_0\|}dv,$$
$$\Psi(\kappa;u) = -\zeta(u)A_\theta^{-1}\gamma(\kappa). \tag{A.59}$$

Substituting (A.58) and (A.59) into (A.57), we have

$$\begin{aligned} G_n(u) = {} & \Omega(u)\{\widehat{\beta}(u) - \beta(u)\} - \frac{1}{2}h^2\mu_2\Omega(u)\beta''(u) \\ & - \int_{\iota_1}^{\iota_2}\Psi(\kappa;u)\{\widehat{\beta}(\kappa) - \beta(\kappa)\}d\kappa \\ & + O_p\left\{\frac{a_n^2}{h} + b_n^2 + b_n'^2 + \frac{a_n b_n'}{h} + \frac{b_n b_n'}{h} + c_n\right\} + o_p\{(nh)^{-1/2} + h^2 + n^{-1/2}\}. \end{aligned} \tag{A.60}$$

Let ρ be a function in $\mathbb{S}$ that satisfies the following integral equation

$$\gamma(\kappa) = \rho(\kappa)\Omega(\kappa) - \int_{\iota_1}^{\iota_2} \rho(u)\Psi(\kappa; u)du. \tag{A.61}$$

Then it follows that

$$\begin{aligned}&\int_{\iota_1}^{\iota_2} \gamma(\kappa)\{\widehat{\beta}(\kappa) - \beta(\kappa)\}d\kappa \\ &\quad = \int_{\iota_1}^{\iota_2} \rho(\kappa)\Omega(\kappa)\{\widehat{\beta}(\kappa) - \beta(\kappa)\}d\kappa - \int_{\iota_1}^{\iota_2}\int_{\iota_1}^{\iota_2} \rho(u)\Psi(\kappa; u)du\{\widehat{\beta}(\kappa) - \beta(\kappa)\}d\kappa.\end{aligned} \tag{A.62}$$

Applying (A.62) to (A.60), we have

$$\begin{aligned}&\int_{\iota_1}^{\iota_2} \gamma(\kappa)\{\widehat{\beta}(\kappa) - \beta(\kappa)\}d\kappa \\ &\quad = \frac{1}{2}h^2\mu_2 \int_{\iota_1}^{\iota_2} \rho(u)\Omega(u)\beta''(u)du + \int_{\iota_1}^{\iota_2} \rho(u)G_n(u)du \\ &\quad + O_p\left\{\frac{a_n^2}{h} + b_n^2 + b_n'^2 + \frac{a_n b_n'}{h} + \frac{b_n b_n'}{h} + c_n\right\} + o_p\{(nh)^{-1/2} + h^2 + n^{-1/2}\},\end{aligned} \tag{A.63}$$

where

$$\begin{aligned}&\int_{\iota_1}^{\iota_2} \rho(u)G_n(u)du \\ &\quad = \frac{1}{n}\sum_{i=1}^{n} \oint_0^1 \int_0^\tau \Bigg[\int_{\iota_1}^{\iota_2} \rho(u)K_h(\theta_0^T v - u)\left[Z_i - \frac{s_p^{(1)}(t, v; \beta, h\beta', u, \theta_0)}{s^{(0)}(t, v; \beta, h\beta', u, \theta_0)}\right]du \\ &\qquad - \int_{\iota_1}^{\iota_2} \rho(u)\zeta(u)A_\theta^{-1}v_2\left\{\beta'(\theta_0^T v)^T Z_i - \frac{s^{*(1)}(t, \theta_0^T v)}{s^{*(0)}(t, \theta_0^T v)}\right\}du\Bigg] M_i(dt, dv).\end{aligned}$$

Notice that

$$\int_{\iota_1}^{\iota_2} \rho(u)K_h(\theta_0^T v - u)\left[Z_i - \frac{s_p^{(1)}(t, v; \beta, h\beta', u, \theta_0)}{s^{(0)}(t, v; \beta, h\beta', u, \theta_0)}\right] du$$
$$= \rho(\theta_0^T v)\left[Z_i - \frac{\widetilde{s}^{(1)}(t, \beta, \theta_0^T v)}{\widetilde{s}^{(0)}(t, \beta, \theta_0^T v)}\right] + O_p(h^2). \tag{A.64}$$

From (A.57), we can also obtain

$$b_n = O_p\left\{\frac{a_n^2}{h} + \frac{a_n b_n'}{h} + \frac{b_n b_n'}{h}\right\} + o_p\{(nh)^{-1/2} + h^2\}, \tag{A.65}$$

and

$$b_n' = O_p\left\{\frac{a_n^2}{h} + \frac{a_n b_n'}{h} + \frac{b_n b_n'}{h}\right\} + o_p\{(nh)^{-1/2} + h^2\}. \tag{A.66}$$

By Lemma 2, plugging (A.63) into the asymptotic expression of $\widehat{\theta}_2 - \theta_{20}$, (A.33), together with (A.64), (A.65) and (A.66) and if $nh^4 \to 0$, then

$$\widehat{\theta}_2 - \theta_{20} = A_\theta^{-1} n^{-1} \sum_{i=1}^{n} \oint_0^1 \int_0^\tau \varphi_i(t, v_2) M_i(dt, dv) + o_p(n^{-1/2}). \tag{A.67}$$

where

$$\varphi_i(t, v_2) = \left\{\left(\int_{\iota_1}^{\iota_2} \rho(u)\zeta(u) du A_\theta^{-1}\right) - I\right\} v_2 \left\{\beta'(\theta_0^T v)^T Z_i - \frac{s^{*(1)}(t, \theta_0^T v)}{s^{*(0)}(t, \theta_0^T v)}\right\}$$
$$- \rho(\theta_0^T v)\left\{Z_i - \frac{\widetilde{s}^{(1)}(t, \beta, \theta_0^T v)}{\widetilde{s}^{(0)}(t, \beta, \theta_0^T v)}\right\}. \tag{A.68}$$

It follows that

$$\sqrt{n}(\widehat{\theta}_2 - \theta_{20}) \to N(0, A_\theta^{-1} \Sigma_\theta (A_\theta^{-1})^T),$$

where

$$\Sigma_\theta = \oint_0^1 \int_0^\tau E[\varphi_i^2(t, v_2) P(t|Z_i) \exp\{\beta(\theta_0^T v)^T Z_i\}] \lambda_0(t, v) dt dv.$$

Here, we complete the proof of Theorem 2.

Proof of Theorem 3.

To prove Theorem 3, we further using (A.57). Let $\Upsilon(v;u) = -\zeta(u)A_\theta^{-1}W(v)$. Then (A.57) can be rewritten as

$$\begin{aligned} G_n(u) = {} & \Omega(u)\{\widehat{\beta}(u) - \beta(u)\} - \frac{1}{2}h^2\mu_2\Omega(u)\beta''(u) \\ & - \oint_0^1 \Upsilon(v;u)\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\}dv \\ & + O_p\left\{\frac{a_n^2}{h} + b_n^2 + b_n'^2 + \frac{a_n b_n'}{h} + \frac{b_n b_n'}{h} + c_n\right\} + o_p\{(nh)^{-1/2} + h^2 + n^{-1/2}\}. \end{aligned} \tag{A.69}$$

Denote $\mathcal{L}$ to be the linear operator that satisfies for any function g,

$$\mathcal{L}(g)(u) = \Omega^{-1}(u)\oint_0^1 \Upsilon(v;u)g(v)dv. \tag{A.70}$$

Set

$$r_n = \quad O_p\left\{\frac{a_n^2}{h} + b_n^2 + b_n'^2 + \frac{a_n b_n'}{h} + \frac{b_n b_n'}{h} + c_n\right\} + o_p\{(nh)^{-1/2} + h^2 + n^{-1/2}\}. \tag{A.71}$$

Substituting (A.70) and (A.71) into (A.69), we get

$$\begin{aligned} G_n(u) + \frac{1}{2}h^2\mu_2\Omega(u)\beta''(u) &= \Omega(u)\{\widehat{\beta}(u) - \beta(u)\} + r_n \\ &\quad - \oint_0^1 \Upsilon(v;u)\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\}dv \\ \Omega(u)^{-1}G_n(u) + \frac{1}{2}h^2\mu_2\beta''(u) &= \{\widehat{\beta}(u) - \beta(u)\} + r_n \\ &\quad - \Omega(u)^{-1}\oint_0^1 \Upsilon(v;u)\{\widehat{\beta}(\theta_0^T v) - \beta(\theta_0^T v)\}dv \\ \Omega(u)^{-1}G_n(u) + \frac{1}{2}h^2\mu_2\beta''(u) &= \{\widehat{\beta}(u) - \beta(u)\} - \mathcal{L}(\widehat{\beta} - \beta)(u) + r_n \\ \{\widehat{\beta}(u) - \beta(u)\} &= \frac{1}{2}h^2\mu_2(\mathcal{I} - \mathcal{L})^{-1}(\beta'')(u) \\ &\quad + (\mathcal{I} - \mathcal{L})^{-1}\Omega(u)^{-1}G_n(u) + r_n. \end{aligned} \tag{A.72}$$

By Martingale Central Limit Theorem, $\sqrt{nh}G_n(u)$ converges to a normal distribution. It can be shown that $\sqrt{nh}(\mathcal{I}-\mathcal{L})^{-1}\Omega(u)^{-1}G_n(u)$ is also asymptotically normal. Combining (A.65), (A.66) and (A.72), if $nh^4 \to 0$, then we have

$$\sqrt{nh}\left[\widehat{\beta}(u)-\beta(u)-\frac{1}{2}h^2\mu_2(\mathcal{I}-\mathcal{L})^{-1}\beta''(u)\right] \to N(0, \nu_0\Pi(u)\Pi(u)^T),$$

where $\Pi(u) = (\mathcal{I}-\mathcal{L})^{-1}(\Omega^{-1/2})(u)$.

The proof of Theorem 3 is finished.