SEMIPARAMETRIC STATISTICAL INFERENCE FOR FUNCTIONAL SURVIVAL MODELS

LIU, KIN YAT

PhD

The Hong Kong Polytechnic University

2019

The Hong Kong Polytechnic University

Department of Applied Mathematics

Semiparametric statistical inference for functional survival models

Liu, Kin Yat

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

April 2019

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________________ (Signed)

Liu, Kin Yat (Name of student)

Abstract

This thesis focuses on the development of semiparametric inference for the functional Cox proportional hazards model and the functional additive hazards model with right-censored data.

We propose a penalized partial likelihood approach and a penalized pseudo-score function approach to the estimation of the model parameters of the functional Cox proportional hazards model and that of the functional additive hazards model, respectively. We establish asymptotic properties which include the consistency, the convergence rate, and the limiting distribution of the proposed estimators. To this end, we investigate the joint Bahadur representation of finite-dimensional and infinite-dimensional estimators in the Sobolev space with proper inner products.

One major contribution made to the study of the functional Cox proportional hazards model and the functional additive hazards model is that the asymptotic joint normality of the estimators of the functional coefficient and the scalar coefficient is derived. Furthermore, the partial likelihood ratio test is developed and is shown to be optimal under the functional Cox proportional hazards model.

These two important issues are not addressed in the previous research. Our new results provide more insights and deeper understanding about the effects of functional predictors on the hazard function. The theoretical results are validated by simulation studies, and the applications of the proposed models are illustrated with a real dataset. Some discussions and closing remarks are given.

Key Words: Functional Bahadur representation; Right censored data; Partial Likelihood ratio test; semiparametric inference; Penalized likelihood; Smoothing splines.

Acknowledgements

I would like to express my sincerest gratitude to my supervisor, Dr. Zhao Xingqiu for her guidance and patience during my Ph.D studies, in addition to my M.Phil studies. I am deeply grateful to her for introducing me to this research area, and helping me to proceed. It was very fortunate to have her as my supervisor.

I also owe my deepest gratitude to my co-supervisor, Professor Chen Xiaojun, for her support and encouragement. Without her, this thesis would not even start.

I am grateful to my families for their love and support. My wife, Ruby, and my sons, Abraham, and Noah are always fuel to my life.

Contents

List of Figures

List of Tables

List of Notations

$\mathbb{R}$	Set of real numbers
T	The failure time
C	The censoring time
U	The censoring time
Y	The observed time
X	The covaraites
δ	The censoring indicator
$x^\top$	Transpose of matrix/vector x
$\|\|\|$	Euclidean norm

Chapter 1

Introduction

This chapter first introduces functional data and survival data. Reproducing kernel Hilbert space (RKHS) and the related results, which are the main tools to study functional data, are then described. Finally, the Bahadur representation of some estimators and its significance are introduced.

1.1 Functional Data

Information technology has fueled the research and developement in functional data analysis (FDA) in recently years. In the field of statistics, functional data analysis is the statistical analysis of data observed from continuous time stochastic processes. In practice, a sample function in a functional data set is recorded at some discrete time points. A data set of n sample functions observed in a time interval $[T_{\min}, T_{\max}]$ can be described mathematically as

$$X_i(t_{i,j}) \in \mathbb{R}, \quad t_{i,j} \in [T_{\min}, T_{\max}], \quad i = 1, 2, \ldots, n, \quad j = 1, \ldots, J_i. \tag{1.1}$$

Note that the values of the sample functions are available at $t_{i,j}$'s only. This leads us to focus on functions with certain properties, such as smoothness, or functions existing in certain function space, such as Sobolev space.

Although a sample function is an infinite dimensional object theoretically, many well-established statistical models are readily adapted to functional data with slight modification. For example, in classic linear regression model, one of the most fundamental tools in statistics, we have

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i, \quad \epsilon_i\ N(0, \sigma^2), \quad i = 1, \ldots, n. \tag{1.2}$$

One way to incorporate functional data to the regression model, is to replace X_{ij} by $X_i(t_j)$ where t_j's are the times of observations:

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_i(t_j) + \epsilon_i, \quad \epsilon_i\ N(0, \sigma^2), \quad i = 1, \ldots, n. \tag{1.3}$$

This method is limited to situations that all sample functions are recorded at same set of t_j. An alternative to this approach is introducing $X_i(\cdot)$ directly to the regression model:

$$Y_i = \beta_0 + \int \beta(t) X_i(t)\, dt + \epsilon_i, \quad \epsilon_i\ N(0, \sigma^2), \quad i = 1, \ldots, n. \tag{1.4}$$

The use of integral is justified by the fact that

$$\int \beta(t) X_i(t)\, dt = \lim_{\Delta t \to 0} \sum_{j=1}^{\infty} \beta_j X_i(t_j) \Delta t. \tag{1.5}$$

1.2 Survival Analysis

Survival analysis focuses on the time to an event of interest, for example, time from disease onset to death.

1.2.1 Survival function and hazard rate function

The theory of survival analysis primarily concerns two entities — the survival function and the hazard rate function. The survival function, $S(t)$, is a function of time which specifies the probability that an individual does not experience an event by time t. Formally, the survival function is defined as

$$S(t) = P(T > t). \tag{1.6}$$

where T is a random variable representing the survival time.

The hazard rate function, $h(t)$, is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T \geq t). \tag{1.7}$$

Therefore, the hazard rate function represents the probability density of T given that an individual has not experienced the event by time t.

Note that the survival function is non-increasing and nonnegative, and it satisfies $S(0) = 1$, whereas the hazard rate function is nonnegative. There are two fundamental connections between the survival function and the hazard rate function:

1.
$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{S(t) - S(t + \Delta t)}{S(t)} = -\frac{S'(t)}{S(t)}. \tag{1.8}$$

2.
$$S(t) = \exp\Big\{ - \int_0^t h(s)\, ds \Big\} \tag{1.9}$$

Berkson and Gage (1952), Cutler and Ederer (1958), and Gehan (1969) first propose to use life-time table for the estimation of survival function. One major drawback of the life-time table approach, however, is that the estimator is biased due to the arbitrary choice of groupings of event times. To circumvent this problem, Böhmer (1912) extends the life-time table approach to the product limit estimator (also known as the Kaplan-Meier estimator). The estimator, as shown by Kaplan and Meier (1958), is a nonparametric maximum likelihood estimator. Efron (1967) and Breslow and Crowley (1974) further shows that the estimator possess self-consistency and asymptotic normality properties. Dabrowska (1989), Bitouze et al. (1999), and Wellner (2007) derive the exponential bound of the Kaplan-Meier estimator. Ware and Demets (1976), Csörgö and Horváth (1980), Gomez et al. (1992) and Gómez et al. (1994) propose left Kaplan-Meier estimator (LKM) to accommodate left censored data, and establish self consistency and asymptotic properties of the proposed estimator.

Blum and Susarla (1980) and Földes et al. (1981) propose kernel methods to obtain a smoothed survival function instead of a step function. In addition to the kernel methods, Bezier curve smoothing and splines are proposed by Kim et al. (2003) and Whittemore and Keller (1986) as smoothed nonparametric maximum likelihood estimators of the survival function.

To ascertain the association of explanatory variables and time to certain event, we consider regression models for survival data. The most common regression model for survival data is the Cox regression model (Cox (1972, 1975)). This model specifies that the hazard function of an individual with explanatory variables $X_1, \ldots, X_p$ assumes the form

$$h(t|X_1, \ldots, X_p) = h_0(t) \exp(\beta_1 X_1 + \cdots + \beta_p X_p). \tag{1.10}$$

where $h_0(t)$ is called the baseline hazard function which is the hazard function of an individual when $X_1 = \cdots = X_p = 0$.

An alternative to the Cox proportional hazards model is the additive regression model due to Aalen (1989), which assumes that the hazard rate of an individual with explanatory variables $X_1, \ldots, X_p$ has the form

$$h(t|X_1, \ldots, X_p) = h_0(t) + \beta_1(t) X_1 + \cdots + \beta_p(t) X_p. \tag{1.11}$$

Similar to the baseline hazard function in the Cox model, $h_0(t)$ in the additive model is the baseline hazard which is one's hazard when $X_1 = \cdots = X_p = 0$.

Note that in Aalen's additive model, the regression coefficients are functions of time. Lin and Ying (1994) propose an alternate additive hazards regression model in which the time-varying regression coefficients are replaced by constants:

$$h(t|x_1, \ldots, x_p) = h_0(t) + \beta_1 X_1 + \cdots + \beta_p X_p. \tag{1.12}$$

1.2.2 Censoring

One distinctive feature of survival data is that they are often incomplete. For instance, to study the treatment effectiveness to inhibit cancer recurrence, one records the time to recurrence after recovery within a study period, say, 5 years. If a participant of the study has no cancer recurrence during the study period, the data is incomplete in the sense that we do not know the exact recurrence time. Instead, we know that the time is longer than 5 years.

The incomplete observations times are called censored survival times. In particular, it is called *right-censored* survival time because the exact time should appear *after* the end of the study. *Left-censored* survival times are defined similarly. In this case, the exact start times being unknown, but we only know that the start time is *before* a particular time point. For example, to study the long expectancy of HIV-infected patients, one does not know the exact time of HIV infection. Instead, one only know that the time is before the start of the study.

1.3 Reproducing Kernel Hilbert Space

In this section, we introduce Reproducing Kernel Hilbert Space (RKHS), an important functional space for functional data analysis. Essential concepts and important properties of RKHS are also described.

1.3.1 Linear Subspaces and Hilbert Spaces

Let $\mathcal{L}$ be a linear space. A functional, $L(\cdot)$, is defined as an operator of any element, $f \in \mathcal{L}$, such that $L(f) \in \mathbb{R}$. Furthermore, if $L(f+g) = L(f) + L(g)$ and $L(\alpha f) = \alpha L(f)$, for any $f, g \in \mathcal{L}$ and $\alpha \in \mathbb{R}$, L is a linear functional.

A bivariate form, $J(\cdot, \cdot)$, is defined as an operator of any two elements, $f, g \in \mathcal{L}$, such that $J(f, g) \in \mathbb{R}$. $J(\cdot, \cdot)$ is bilinear if $J(\alpha f + \beta g, h) = \alpha J(f, h) + \beta J(g, h)$ and $J(f, \alpha g + \beta h) = \alpha J(f, g) + \beta J(f, h)$, $\forall f, g, h \in \mathcal{L}$ and $\forall \alpha, \beta \in \mathbb{R}$.

A bilinear form is symmetric if it satisfies $J(f, g) = J(g, f), \forall f, g \in \mathcal{L}$. If, for any $f \in \mathcal{L}$, $J(f, f) \geq 0$, $J(\cdot, \cdot)$ is said to be non-negative definite, and it is positive definite if $J(f, f) = 0$ holds only when $f = 0$. An inner product defined on a linear space is a positive definite bilinear form, denoted by $(\cdot, \cdot)$. A norm in a linear space can be defined through an inner product by $\|f\| = \sqrt{(f, f)}$.

If $\lim_{n\to\infty} Lf_n = Lf$ whenever $\lim_{n\to\infty} f_n = f$, the functional L is said to be continuous. A Cauchy sequence, f_n, satisfies $\lim_{n,m\to\infty} \|f_m - f_n\| = 0$. If all Cauchy sequences converge in a linear space, the linear space is complete. A complete inner product linear space is called a Hilbert space.

1.3.2 Riesz Representation Theorem

Theorem 1.1. *Let L be a continuous linear functional in a Hilbert space $\mathcal{H}$. For any $f \in \mathcal{H}$, there exists a unique $g_L \in \mathcal{H}$ such that $Lf = (g_L, f)$.*

This theorem asserts that for every g in a Hilbert space $\mathcal{H}$, $L_g f = (g, f)$ defines a continuous linear functional L_g. Conversely, every continuous linear functional L in $\mathcal{H}$ has a representation $Lf = (g_L, f)$ for some $g_L \in \mathcal{H}$. g_L is called the representer of L.

1.3.3 Reproducing Kernel and Non-Negative Definite Function

Let $\mathcal{X}$ be the domain of f and $x \in \mathcal{X}$. A functional L_x defined as $L_x f = f(x)$ is called an evaluation functional. If, for all $x \in \mathcal{X}$, the evaluation function $L_x f$ is continuous in $\mathcal{H}$, then $\mathcal{H}$ is called a *reproducing kernel Hilbert space.*

By Riesz representation theorem, for every $f \in \mathcal{H}$, there exists $R_x \in \mathcal{H}$ such that $L_x f = (R_x, f)$. Define a symmetric bivariate function $R(x, y)$ as (R_x, R_y) for $x, y \in \mathcal{X}$. Note that $(R_x, R_y) = R_y(x) = R_x(y)$ and $(R_x, f) = f(x)$. In this case, $R(\cdot, \cdot)$ is called the reproducing kernel of $\mathcal{H}$. Knowing non-negative definite functions helps to construct reproducing kernel Hilbert spaces as justified by the following theorem.

Theorem 1.2. *For any reproducing kernel Hilbert space $\mathcal{H}$ of functions defined on $\mathcal{X}$, there exists a reproducing kernel $R(x, y)$, which is unique and non-negative definite. Conversely, there exists a unique reproducing kernel Hilbert space $\mathcal{H}$ for every non-negative definite function $R(x, y)$ on $\mathcal{X}$.*

1.3.4 Reproducing Kernel Hilbert Spaces and Penalized Regression

Applying penalized regression procedures to the estimation of complex functions is common (Wahba (1990); Eubank (1999); Hastie et al. (2001)). These procedures are commonly employed in functional data analysis (Ramsay and Silverman (2005)). The estimators in these procedures are defined as the solutions of optimization problems. The solution of a minimization problem is the smoothing spline in a functional space. If the minimization problem is formulated on a reproducing kernel Hilbert space, the solution exists and is guaranteed to be unique.

The central roles in penalized regression are played by the reproducing kernel Hilbert spaces (RKHS) and the corresponding reproducing kernels (RK). Wahba

(1990); Gu (2002); Pearce and Wand (2006) provide comprehensive reviews of reproducing kernel Hilbert spaces (RKHS) methods for regression analysis.

1.4 Bahadur Representation

An estimator admits a Bahadur representation if the estimator can be almost expressed a linear estimator, i.e. $\sum_{i=1}^{n} h_i(X_i) + R_n$ where R_n becomes negligible as $n \to \infty$. When a Bahadur representation of an estimator exists, the properties of the estimator can be derived easily. For example, under some regularity conditions, if $X_1, \cdots, X_n$ are independent, we may conclude that $\sum_{i=1}^{n} h_i(X_i)$ follow Normal distribution asymptotically due to Central limit theorem. Bahadur (1966) first derives the Bahadur representation of sample percentiles.

Theorem 1.3. *Let $\omega = (X_1, X_2, ...)$ be a sequence of independent random variables with each X_i distributed according to F. For each $n = 1, 2, \cdots$, let Y_n be the sample pth-percentile when the sample is $X_1, \cdots, X_n$. Let Z_n be the number of observation X_i in the sample $X_1, \cdots, X_n$ such that $X_i > \xi$. Then,*

$$Y_n = \xi + \frac{Z_n - n(1-p)}{n\, f(\xi)} + R_n \tag{1.13}$$

where R_n becomes negligible as $n \to \infty$.

Bahadur representations are also derived for more general estimators (See Carroll (1978), He and Shao (1996), Bose (1998)).

1.5 Outline

The remainder of this dissertation is organized as follows. Chapter 2 presents a penalized semiparametric maximum partial likelihood estimation and hypothesis testing for the functional Cox model in analyzing right-censored data with both functional and scalar predictors. Deriving the asymptotic joint distribution of finite-dimensional and infinite-dimensional estimators is a very challenging theoretical problem due to the complexity of semiparametric models. For the problem, we construct the Sobolev space equipped with a special inner product and discover a new joint Bahadur representation of estimators of unknown regression function and coefficients. Using this key tool, we establish the asymptotic joint normality of the proposed estimators and then construct a local confidence interval for an unknown slope function. Furthermore, we study a penalized partial likelihood ratio test, show that the test statistic enjoys the Wilks phenomenon, and also verify the optimality of the test. The theoretical results are examined through simulation studies, and a right-censored data example from the Improving Care of Acute Lung Injury Patients study is provided for illustration. In Chapter 3, the semiparametric inference of functional additive hazard model is studied. In particular, the asymptotic joint distribution of finite-dimensional and infinite-dimensional estimators is established. In addition, uniform convergence of the infinite-dimensional estimator is developed. The main tool to develop these result is the joint Bahadur representation of the estimators. The theoretical results are examined through simulation studies, and a right-censored data example from the Improving Care of Acute Lung Injury Patients study is provided for illustration. Finally, conclusion and future work are discussed in Chapter 4.

Chapter 2

Semiparametric Inference for the Functional Cox Model

2.1 Introduction

Advances in information technology enable collecting and processing of densely observed data over some temporal or spatial domains. The resulting data are coined functional data to differentiate them from the traditional, scalar data. Examples of functional data include hippocampus radial distance data Li and Luo (2017), high dimensional microarray gene expression data Chen et al. (2011), and the Sequential Organ Failure Assessment data Gellar et al. (2014, 2015).

The explosion of functional data requires the development of functional data analysis. Recently, Crambes et al. (2009), Yuan and Cai (2010), Cai and Yuan (2012), and Shang and Cheng (2015), among others, proposed roughness regularization methods to control the model complexity in a continuous manner. This overcomes the imprecise control on the model complexity due to the truncation parameter in the functional principal component analysis (FPCA)-based approaches, as pointed out by Ramsay and Silverman (2005).

When time-to-event data are available, the proportional hazards model Cox (1972) is commonly used for the analysis of such data. Under the Cox model, the hazard function of a failure time for a subject takes the form:

$$h(t|Z) = h_0(t) \exp^{\theta_0' Z},$$

where $h_0(\cdot)$ is an unspecified baseline hazard function, $Z \in \mathbb{R}^p$ is a covariate vector and $\theta_0 \in \mathbb{R}^p$ is an unknown parameter. This model was further studied by Cox (1975), Tsiatis (1981), Andersen and Gill (1982), Johansen (1983), and Jacobsen (1984), among others. When functional covariates are involved, Chen et al. (2011) proposed the following functional Cox model:

$$h(t|Z, X(\cdot)) = h_0(t) \exp\left\{\theta_0' Z + \int_{\mathbb{I}} X(s)\beta_0(s)ds\right\}, \tag{2.1}$$

where $X(\cdot)$ is a functional covariate and $\beta_0(\cdot)$ is an unknown coefficient function. Clearly, this model takes into account the effect of the entire trajectory of $X(\cdot)$ on the hazard function. Note that the Cox model with a time-dependent covariate only considers the effect of $X(t)$ on the hazards function at time t, where an overall effect of a functional covariate on the hazard function cannot be explained. Chen et al. (2011) applied the functional Cox model in studying the survival of diffuse large-B-cell lymphoma (DLBCL) patients in relation to the high-dimensional microarray gene expression of the patients, which is expressed as a functional predictor. Recently, Kong et al. (2018) established the rate of convergence of the maximum approximate partial likelihood estimator and conducted a score test for testing the nullity of the slope function related to functional predictors. Qu et al. (2016) studied the asymptotic properties of the maximum partial likelihood estimator under the framework of reproducing kernel Hilbert space and established the asymptotic normality and efficiency of the estimator of scalar covariates. However, the asymptotic distribution of the maximum partial likelihood estimator of the slope function has not been studied. Another important issue is to study the partial likelihood ratio test, which has not been addressed in the literature. Our goal is to address these challenging issues and fill the gap in the study of the functional Cox model.

Motivated by Cheng and Shang (2015), we explored a joint Bahadur representation to derive the asymptotic joint distribution of the maximum partial likelihood estimators of the slope function and coefficients in the functional Cox model. Compared to that proposed in Cheng and Shang (2015), our model is focused on the joint asymptotic study of the (generalized) partial functional survival model. Our main contribution includes the following aspects: (1)we embedded the Sobolev space with a special inner product, and deduced the joint Bahadur representation of the maximum partial likelihood estimators of finite-dimensional and infinite-dimensional parameters in the space; (2) we got the pointwise confidence interval of the functional coefficient; (3) we investigated a penalized partial likelihood ratio test for testing global effects of both scale and functional covariates on the hazard function.

The rest of this chapter is organized as follows. In Section 2.2, we construct the Sobolev space and present a penalized estimation approach for unknown regression parameters in the functional Cox model. In Section 2.3, we derive a joint Bahadur representation (FBR) of the maximum partial likelihood estimators of scalar and functional parameters in the space with a special inner product and establish the asymptotic properties of the proposed estimators. In Section 2.4, we develop a penalized likelihood ratio test for a global hypothesis. In Section 2.5, we present simulation results to evaluate the performance of the proposed asymptotic inference procedures. Section 2.6 illustrates an application of the proposed method to the data obtained from the Improving Care of Acute Lung Injury Patients (ICAP) study Needham et al. (2006). Some concluding remarks are made in Section 2.7. All technical proofs are given in the Appendix.

2.2 Estimation Method

Denote the covariates that are incorporated in the functional Cox model (1.1) by $W = (Z^\top, X(\cdot))$. Under the right censorship, let T be the survival time, C be the censoring time, $Y = \min(T, C)$ be the observed time, $\Delta = \mathbf{1}(T \le C)$ be the censoring indicator, and $N(t) = \Delta\mathbf{1}(T \le t)$ be the counting process, where $\mathbf{1}(\cdot)$ is the indicator function. For simplicity, assume $E(\Delta Z) = 0, E\{\Delta X(t)\} = 0$ for any $t \in \mathbb{I}$. Without loss of generality, we take $\mathbb{I} = [0, 1]$. As usual, assume that the survival time T and the censoring time C are conditionally independent given W. Our goal is to estimate $\alpha_0 = (\theta_0^\top, \beta_0(\cdot))$ to reveal the relationship between W and T. Suppose that $\beta_0(\cdot)$ belongs to the mth-order Sobolev space $\mathcal{H}^{(m)}(\mathbb{I})$, which is abbreviated as $\mathcal{H}^{(m)}$ for notational simplicity:

$$\mathcal{H}^{(m)}(\mathbb{I}) = \{\beta : \mathbb{I} \mapsto \mathbb{R} | \beta^{(j)} \text{ is absolutely continuous for } j = 0, 1, \dots, m-1, \beta^{(m)} \in L_2(\mathbb{I})\},$$

where the constant $m > 1/2$ is known, $\beta^{(j)}(\cdot)$ is the jth derivative of $\beta(\cdot)$ and $L_2(\mathbb{I})$ is the L_2 space defined in $\mathbb{I}$.

Define $\eta_\alpha(W) = \theta^\top Z + \int_{\mathbb{I}} X(s)\beta(s)\,ds$, and $\mathcal{Y}(t) = \mathbf{1}(Y \ge t)$. The log partial likelihood of the model (1.1) given the data $\{(Y_i, W_i, \Delta_i), i = 1, \dots, n\}$ is given by

$$l_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \Delta_i \left[\eta_\alpha(W_i) - \log \frac{1}{n}\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\right].$$

To estimate α_0, we propose to use the following penalized log partial likelihood function

$$l_{n,\lambda}(\alpha) = l_n(\alpha) - \frac{\lambda}{2} J(\beta, \beta),$$

where $J(\beta_1, \beta_2) = \int_{\mathbb{I}} \beta_1^{(m)}(s)\beta_2^{(m)}(s)\,ds$ is the penalty function, and λ is the penalty parameter which controls the balance between the bias and the smoothness of the parameter. Thus, the penalized estimator of α_0 is defined by $\hat{\alpha}_{n,\lambda} = \arg\max_{\alpha \in \mathcal{H}} l_{n,\lambda}(\alpha)$, where $\mathcal{H} = \mathbb{R}^p \times \mathcal{H}^{(m)}$.

2.3 Asymptotic Properties

Before stating the main results, we first introduce some notation and regularity conditions. For any vector z, $z^{\otimes 2} = zz^\top$, $z^{\otimes 1} = z$, and $z^{\otimes 0} = \mathbf{1}$ with all of the elements being 1. Define the semi-inner product for any $\alpha_i = (\theta_i^\top, \beta_i(\cdot)) \in \mathcal{H}, i = 1, 2$ as

$$
\begin{aligned}
&< \alpha_1, \alpha_2 >_\lambda \qquad (2.2)\\
=&E \int_0^\tau \Big[\frac{E\{\mathcal{Y}(t)\exp(\eta_{\alpha_0}(W))\eta_{\alpha_1}(W)\eta_{\alpha_2}(W)\}}{E\{\mathcal{Y}(t)\exp(\eta_{\alpha_0}(W))\}} \\
&- \frac{E\{\mathcal{Y}(t)\exp(\eta_{\alpha_0}(W))\eta_{\alpha_1}(W)\}E\{\mathcal{Y}(t)\exp(\eta_{\alpha_0}(W))\eta_{\alpha_2}(W)\}}{\big(E\{\mathcal{Y}(t)\exp(\eta_{\alpha_0}(W))\}\big)^2} \Big] \, dN(t) \\
&+ \lambda J(\beta_1, \beta_2),
\end{aligned}
$$

where τ is the end of the study. Define

$$
\begin{aligned}
S_1^{(k)}(t,\alpha) &= \frac{1}{n}\sum_{i=1}^n [\mathcal{Y}_i(t)\exp\{\eta_\alpha(W_i)\}Z_i^{\otimes k}], k = 0, 1, 2,\\
s_1^{(k)}(t,\alpha) &= E[\mathcal{Y}(t)\exp\{\eta_\alpha(W)\}Z^{\otimes k}], k = 0, 1, 2,\\
S_2^{(1)}(t,s,\alpha) &= \frac{1}{n}\sum_{i=1}^n [\mathcal{Y}_i(t)\exp\{\eta_\alpha(W_i)\}X_i(s)],\\
s_2^{(1)}(t,s,\alpha) &= E[\mathcal{Y}(t)\exp\{\eta_\alpha(W)\}X(s)],\\
S_2^{(2)}(t,s,v,\alpha) &= \frac{1}{n}\sum_{i=1}^n [\mathcal{Y}_i(t)\exp\{\eta_\alpha(W_i)\}X_i(s)X_i(v)],\\
s_2^{(2)}(t,s,v,\alpha) &= E[\mathcal{Y}(t)\exp\{\eta_\alpha(W)\}X(s)X(v)],\\
\Sigma &= E\left\{ \int_0^\tau \frac{s_1^{(2)}(t,\alpha_0)}{s_1^{(0)}(t,\alpha_0)} - \frac{s_1^{(1)}(t,\alpha_0)^{\otimes 2}}{s_1^{(0)}(t,\alpha_0)^2} \, dN(t) \right\},
\end{aligned}
$$

$$
F(s,t) = \int_0^\tau \mathrm{Cov}\{X(s), X(t) | T = v, \Delta = 1\} E[\mathcal{Y}(v)\exp\{\eta_{\alpha_0}(W)\}] h_0(v) \, dv,
$$

where

$$
\begin{aligned}
&\mathrm{Cov}\{X(s), X(t)|T=v, \Delta=1\} \\
=&E\{X(s)X(t)|T=v, \Delta=1\} - E\{X(s)|T=v, \Delta=1\}E\{X(t)|T=v, \Delta=1\} \\
=&\frac{s_2^{(2)}(v,t,s,\alpha_0)}{s_1^{(0)}(v,\alpha_0)} - \frac{s_2^{(1)}(v,s,\alpha_0)s_2^{(1)}(v,t,\alpha_0)}{s_1^{(0)}(v,\alpha_0)^2}.
\end{aligned}
$$

Define a bilinear operator $V(\cdot,\cdot)$ in $\mathcal{H}^{(m)}$ as: $V(\beta_1,\beta_2) = \int_{\mathbb{I}}\int_{\mathbb{I}} F(s,t)\beta_1(s)\beta_2(t)\,ds\,dt$, which is in fact one norm in the $\mathcal{H}^(m)$ space. Set the projection of Z on $X(\cdot)$ as $G \equiv (G_1, G_2, \cdots, G_p)^\top$ with

$$
\begin{aligned}
G_k(\cdot) = &\sum_{j=1}^{\infty}\int_{\mathbb{I}} E\Big[\int_0^\tau \frac{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}Z_k X(u)]}{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}]} \\
&- \frac{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}Z_k]}{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}]}\frac{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}X(u)]}{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}]}\,dN(t)\Big] h_j(u)\,du h_j(\cdot) \\
\equiv &\sum_{j=1}^{\infty} G_{jk}h_j(\cdot).
\end{aligned}
$$

We denote two positive sequences a_n and b_n as $a_n \asymp b_n$ if $\lim_{n\to\infty}(a_n/b_n) = c > 0$. If $c = 1$, we denote $a \sim b$. To construct a Hilbert Space and establish the theoretical properties of the proposed estimator, we need the following regularity conditions:

(C1) (i) $0 < P(Y \geq \tau) < 1$.

(ii) There exists a constant $c_1 > 0$, for any $\alpha \in \mathcal{H}$, we have

$$
\begin{aligned}
&E\int_0^\tau \Big[\frac{E\{\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}\eta_\alpha(W)^2\}}{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}]} - \frac{(E\{\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}\eta_\alpha(W)\})^2}{\big(E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}]\big)^2}\Big]\,dN(t) \\
&\geq c_1 E\{\eta_\alpha(W)\}^2.
\end{aligned}
$$

(C2) There exists a sequence of functions $\{h_j\}_{j\geq 1} \subset \mathcal{H}^{(m)}$ such that $||h_j||_{L_2} \leq c_h j^a$ for each $j \geq 1$, some constants $a \geq 0$, $c_h \geq 0$, and

$$
V(h_i,h_j) = \delta_{ij}, J(h_i,h_j) = \rho_i\delta_{ij}, \quad \text{for any} \quad i,j \geq 1,
$$

where δ_{ij} is the Kronecker's notation, and ρ_i is a nondecreasing nonnegative sequence satisfying $\rho_i \asymp i^{2k}$ for some constant $k > a + 1/2$.

(C3) $\Sigma - V(G, G^\top)$ is positive definite. There exists $b \in ((1+2a)/(2k), 1]$ such that $\sum_j |G_{jk}|^2 \rho_j^b < \infty$ for $k = 1, \ldots, p$.

(C4) There exist constants $s \in (0,1)$ and $M_0 > 0$ such that $E[\exp\{s(\|X\|_{L_2} + \|Z\|_2)\}] < \infty$, and $E\{|\eta_\alpha(W)|^4\} \leq M_0\{E|\eta_\alpha(W)|^2\}^2$ for any $\alpha \in \mathcal{H}$.

Remark 2.1. *Condition (C1)(i) is used to is very common in survival analysis, while Condition (C1)(ii) is trivial when $\beta = 0$ under Condition (C3).*

Equipped with the inner product, $\mathcal{H}$ is a Hilbert space and $\mathcal{H}^{(m)}$ is a reproducing kernel Hilbert space (RKHS) with the inner product

$$< \beta_1, \beta_2 >_m = \int_{\mathbb{I}} \int_{\mathbb{I}} F(s,t)\beta_1(s)\beta_2(t)\, ds\, dt + \lambda J(\beta_1, \beta_2). \tag{2.3}$$

Denote the reproducing kernel in $\mathcal{H}^{(m)}$ by $K(s,t)$. Define a linear nonnegative definite and self-adjoint operator W_λ as: $< W_\lambda \beta_1, \beta_2 >_m = \lambda J(\beta_1, \beta_2)$. Then, we have $< \beta_1, \beta_2 >_m = V(\beta_1, \beta_2) + < W_\lambda \beta_1, \beta_2 >_m$.

Remark 2.2. *Under Condition (C2), the eigen-system can be derived from the following integro-differential equations (Shang and Cheng (2015)):*

$$(-1)^m y_j^{(2m)}(t) = \rho_j \int_{\mathbb{I}} F(s,t) y_j(s)\, ds,$$

$$y_j^{(i)}(0) = y_j^{(i)}(1) = 0, \qquad\qquad i = m, m+1, \cdots, 2m-1.$$

Let $h_j = y_j/\sqrt{V(y_j, y_j)}$, $k = m + r + 1$ and $a = r + 1$. Then h_j and $\rho_j, j = 1, 2, \ldots$ are the eigenvectors and eigenvalues, respectively, if one of the following additional assumptions is satisfied:

(i) $r = 0$;

(ii) $r \geq 1$, *and for any* $i = 0, 1, \dots, r-1$, $F^{(i,0)}(0,t) = 0$ *for any* $t \in \mathbb{I}$, *where* $F^{(i,0)}(s,t)$ *is the ith-order partial derivative with respect to* s.

The relationships between (h_j, ρ_j), $K(\cdot,\cdot)$ *and* W_λ *are given as follows:*

$$K_t(\cdot) = \sum_{j=1}^{\infty} \frac{h_j(t)}{1+\lambda\rho_j} h_j(\cdot), (W_\lambda h_j)(\cdot) = \frac{\lambda\rho_j}{1+\lambda\rho_j} h_j(\cdot).$$

This can be referred to Shang and Cheng (2015).

Remark 2.3. *Under Condition (C3), we have that* $V(G, W_\lambda G^\top) \to 0$ *with* $\lambda \to 0$. *Furthermore, from the definition of* G, *we have* $G = \boldsymbol{0}$ *when* $X(\cdot)$ *and* Z *are independent.*

Remark 2.4. *Condition (C4) on covariates is weaker than the conditions required by Qu et al. (2016).*

In the following, we set $h = \lambda^{1/(2k)}$.

Theorem 2.1. *(Rate of Convergence) Suppose that Conditions (C1)-(C4) hold. If*

$$h = o(1), n^{-1/2} h^{-(a+1)-\frac{2k-2a-1}{4m}} \{\log(n)\}^2 \{\log\log(n)\}^{1/2} = o(1),$$

then $\hat{\alpha}_{n,\lambda}$ *is the unique estimate for* α_0 *and* $\|\hat{\alpha}_{n,\lambda} - \alpha_0\|_\lambda = O_p(r_n)$, *where* $r_n = (nh)^{-1/2} + h^k$.

This theorem shows that when we choose $\lambda = n^{-(2k)/(2k+1)}$, the estimate enjoys the same order of convergence as that in Qu et al. (2016).

For ease of interpretation, define $\mathcal{S}_n(\alpha)$ and $\mathcal{S}_{n,\lambda}(\alpha)$ be the Fréchet derivatives of $l_n(\alpha)$ and $l_{n,\lambda}(\alpha)$, respectively. Direct calculation yields that the Fréchet derivatives of $l_{n,\lambda}(\alpha)$ at the direction of α_1 is

$$\begin{aligned}\mathcal{S}_{n,\lambda}(\alpha)\alpha_1 &= \frac{1}{n}\sum_{i=1}^{n} \Delta_i \left[\eta_{\alpha_1}(W_i) - \frac{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_1}(W_j)}{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}}\right] - \lambda J(\beta,\beta_1) \\ &\equiv \mathcal{S}_n(\alpha)\alpha_1 - \lambda J(\beta,\beta_1).\end{aligned}$$

Theorem 2.2. *(Joint Bahadur Representation) Suppose that Conditions (C1)-(C4) hold. If*

$$n^{-1/2}h^{-(a+1)-\frac{2k-2a-1}{4m}}\{\log(n)\}^2\{\log\log(n)\}^{1/2} = o(1),$$

$$nh^{2k(1+b)} = o(1), \ \ and \ \sum_{j=1}^{\infty} V(\beta_0, h_j)^2\rho_j^2 < \infty,$$

then we have $\|\hat{\alpha}_{n,\lambda} - \alpha_0 - \mathcal{S}_{n,\lambda}(\alpha_0)\|_\lambda = O_p(a_n)$, *where*

$$a_n = n^{-1/2}h^{-(4ma+6m-1)/4m}r_n\{\log\log(n)\}^{1/2}\log(n)^2 + h^{-1/2}r_n^2, \ \ and \ r_n = (nh)^{-1/2} + h^k.$$

Based on the joint Bahadur representation, we can establish the asymptotic joint distribution of the proposed estimators of the slope function and the coefficients.

Theorem 2.3. *(Asymptotic Joint Distribution) Suppose that the conditions of Theorem 2.2 hold. Furthermore, assume that* $\sup_{j\geq 1}\|h_j\|_\infty \leq c_h j^a$, $n^{1/2}a_n h^{-(a+1/2)} = o(1)$, $n^{1/2}h^{k(1+b)} = o(1)$, $\sum_{j=1}^{\infty} V(\beta_0, h_j)^2\rho_j^2 < \infty$, *and* $h^{(2a+1)}\sum_{j=1}^{\infty}\frac{\|h_j(t)\|_\infty^2}{(1+\lambda\rho_j)^2} \asymp \sigma_t^2 > 0$. *Then we have*

$$\begin{bmatrix} \sqrt{n}(\hat{\theta}_{n,\lambda} - \theta_0) \\ \sqrt{nh}h^a\{\hat{\beta}_{n,\lambda}(t) - \beta_0(t)\} \end{bmatrix} \to N(0, \Phi),$$

where

$$\Phi = \begin{bmatrix} \{\Sigma - V(G, G^\top)\}^{-1} & 0 \\ 0 & \sigma_t^2 \end{bmatrix}.$$

Here, Σ can be consistently estimated by

$$\begin{aligned}\hat{\Sigma} =& \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \hat{\mathrm{Var}}(Z|T=t,\Delta=1)\mathcal{Y}_i(t)\exp\{\eta_{\hat{\alpha}}(W_i)\}\,d\hat{\Lambda}_0(t)\right] \\ =& \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \Big[\frac{S_1^{(2)}(t,\hat{\alpha})}{S_1^{(0)}(t,\hat{\alpha})} - \frac{\{S_1^{(1)}(t,\hat{\alpha})\}^{\otimes 2}}{[S_1^{(0)}(t,\hat{\alpha})]^2}\Big]\mathcal{Y}_i(t)\exp\{\eta_{\hat{\alpha}}(W_i)\}\,d\hat{\Lambda}_0(t)\right],\end{aligned}$$

where

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_{k=1}^n dN_k(s)}{\sum_{j=1}^n \mathcal{Y}_j(s)\exp\{\eta_{\hat{\alpha}}(W_j)\}}.$$

Theorem 2.3 implies that, under certain conditions, the asymptotic bias for the estimation of $\beta_0(t_0)$ vanishes. Hence, Theorem 2.3 together with the Delta-method immediately yields the point-wise confidence interval (CI) for some real-valued smooth function of $\beta_0(t)$ at any fixed point $t_0 \in \mathbb{I}$, denoted by $\rho\{\beta_0(t_0)\}$. Let $\dot{\rho}(\cdot)$ be the first derivative of $\rho(\cdot)$. By Theorem 2.3, for any fixed point $t_0 \in \mathbb{I}$ and $\dot{\rho}\{\beta_0(t_0)\} \neq 0$, we have

$$P\left(\rho\{\beta_0(t_0)\} \in \left[\rho\{\hat{\beta}(t_0)\} \pm \Phi_{\frac{\xi}{2}} \frac{\dot{\rho}\{\beta_0(t_0)\}\sqrt{\sum_{j=1}^{\infty}(|h_j(t)|^2/(1+\lambda\rho_j)^2)}}{\sqrt{n}}\right]\right)$$
$$\to 1-\xi,$$

as $n \to \infty$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function and Φ_ξ is the lower ξ-th percentile of $\Phi(\cdot)$, that is $\Phi(\Phi_\xi) = 1-\xi$.

2.4 Partial Likelihood Ratio Test

In this section, we consider testing the following "global" hypothesis:

$$H_0 : \alpha = \alpha_0 \qquad \text{versus} \qquad H_1 : \alpha \neq \alpha_0,$$

where $\alpha_0 \in \mathcal{H}$. The penalized partial likelihood ratio rest (PLRT) statistic is defined as:

$$\mathrm{PLRT}_{n,\lambda} \equiv l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\hat{\alpha}_{n,\lambda}).$$

We next derive the null limiting distribution of $\mathrm{PLRT}_{n,\lambda}$.

Theorem 2.4. *(Likelihood Ratio Test) Suppose that Conditions (C1)-(C4) hold. Assume that*

$$nh^{2k(1+b)} = O(1), nh^2 \to \infty, n^{1/2}a_n = o(1), nr_n^3 = o(1), \sum_{j=1}^{\infty} V(\beta_0, h_j)^2\rho_j^2 < \infty,$$

$$n^{1/2}h^{-\{a+1/2+(2k-2a-1)/(4m)\}}r_n^2\{\log(n)\}^2\{\log\log(n)\}^{1/2} = o(1),$$

and

$$n^{1/2}h^{-\{2a+1+(2k-2a-1)/(4m)\}}r_n^3\{\log(n)\}^3\{\log\log(n)\}^{1/2} = o(1).$$

Then under H_0, we have

$$(2\nu_\lambda)^{-1/2}(-2n\gamma_\lambda PLRT_{n,\lambda} - n\gamma_\lambda\|W_\lambda\beta_0\|_m^2 - \nu_\lambda) \xrightarrow{d} N(0,1),$$

where $\sigma_\lambda^2 \equiv \sum_{j=1}^\infty h/(1+\lambda\rho_j)$, $\rho_\lambda^2 \equiv \sum_{j=1}^\infty h/(1+\lambda\rho_j)^2$, $\gamma_\lambda \equiv \sigma_\lambda^2/\rho_\lambda^2$, and $\nu_\lambda \equiv h^{-1}\sigma_\lambda^4/\rho_\lambda^2$.

It follows from Theorem 2.3 that $n\|W_\lambda\beta_0\|_m^2 = o(n\lambda) = o(\nu_\lambda)$. Therefore, we have $2n\gamma_\lambda PLRT_{n,\lambda} \sim N(\nu_\lambda, 2\nu_\lambda)$, which is nearly $\chi^2_{\nu_\lambda}$ as $n \to \infty$. This shows that PLRT enjoys the Wilks phenomenon.

As suggested by one anonymous reviewer, our proposed method can handel some composite hypothesis testing. In fact, by examining the proof of Theorem 2.4, we find that the null limiting distribution derived therein remains the same even when the hypothesized value α_0 is unknown. An important consequence is that the proposed likelihood ratio approach can also be used to test a composite hypothesis such as

$$H_0: \theta = \theta_0,\ \beta \text{ has some linear or polynomial structure.}$$

Under H_0, the true slope function can be expressed as $\beta_0 = \sum_{j=0}^d t^j b_j^0$, where d is the order the the polynomial. Denote $\mathbf{b}_0 = (b_0^0, b_1^0, \ldots, b_d^0)^\top$. Then, $(\theta_0^\top, \mathbf{b}_0^\top)^\top$ can be estimated through the following parametric optimization:

$$\begin{aligned}(\hat\theta^0, \hat{\mathbf{b}}^0) &= \arg\max l_{n,\lambda}(\theta, \sum_{j=0}^d t^j b_j) \\ &= \arg\max l_n(\theta, \sum_{j=0}^d t^j b_j) - \frac{\lambda}{2}\mathbf{b}^\top B\mathbf{b},\end{aligned}$$

where B is a $(d+1)\times(d+1)$ matrix, with the i,j-th component being $J(t^i, t^j)$. After deriving the estimate of θ_0 and $\mathbf{b}_0$, we can obtain the estimate of β_0, denoted as $\hat{\beta}^0 = \sum_{j=0}^{d} \hat{b}_j^0 t^j$. Denote $\hat{\alpha}^0 = ((\hat{\theta}^0)^\top, \hat{\beta}^0)$, and α_0 is the unknown true parameter. Under this scenario, the logical related to the asymptotic distribution of PLRT is listed below:

$$\begin{aligned}
\mathrm{PLRT}_{n,\lambda}^{com} \equiv & l_{n,\lambda}(\hat{\alpha}^0) - l_{n,\lambda}(\hat{\alpha}_{n,\lambda}) \\
= & l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\hat{\alpha}_{n,\lambda}) + l_{n,\lambda}(\hat{\alpha}^0) - l_{n,\lambda}(\alpha_0) \\
= & l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\hat{\alpha}_{n,\lambda}) + O_p(n^{-1}).
\end{aligned}$$

The last equality holds as traditional parametric theory leads to $2n\{l_{n,\lambda}(\hat{\alpha}^0) - l_{n,\lambda}(\alpha_0)\} = O_p(1)$. As $l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\hat{\alpha}_{n,\lambda})$ is exactly the same as that proposed in Theorem 2.4, we have

$$-2n\gamma_\lambda\{l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\hat{\alpha}_{n,\lambda})\} \sim \chi^2_{\nu_\lambda}.$$

Therefore, we conclude that the null limit distribution for testing the composite hypothesis also follows $\chi^2_{\nu_\lambda}$. To conclude this section, we show that the PLRT achieves the optimal minimax rate given by Ingster (1993). To this end, we consider the alternative hypothesis $H_{1n} : \alpha = \alpha_{n_0}$, where $\alpha_{n_0} = \alpha_0 + \alpha_n$, $\alpha_0 \in \mathcal{H}$ and α_n belongs to the alternative value set $\mathcal{A} = \{\alpha \in \mathcal{H}, \|\theta\|_2 \leq \zeta, \|\beta\|_{L^2} \leq \zeta, J(\beta, \beta) \leq \zeta\}$ for some constant $\zeta > 0$.

Theorem 2.5. *Suppose that the conditions of Theorem 2.4 hold, and under H_{1n} : $\alpha = \alpha_{n_0}$, $\|\hat{\alpha}_{n,\lambda} - \alpha_{n_0}\|_\lambda = O_p\{(nh)^{-1/2} + h^k\}$ holds uniformly over $\alpha_{n_0} \in \mathcal{A}$. If $nh^{3/2+a/2} \to \infty$ as $n \to \infty$, then, for any $\delta \in (0,1)$, there exists positive constants b_0 and N such that*

$$\inf_{n \geq N} \inf_{\alpha_n \in \mathcal{A}, \|\alpha_n\|_\lambda \geq b_0 \eta_n} P(\textit{reject} \quad H_0 | H_{1n} \textit{ is } \textit{ true}) \geq 1 - \delta,$$

where $\eta_n \geq \sqrt{h^{2k} + (nh^{1/2})^{-1}}$. Moreover, the maximum lower bound of η_n is $n^{-2k/(4k+1)}$, which can be achieved when $h = h^{**} = n^{-2/(4k+1)}$.

2.5 Simulation Studies

In this section, we conduct simulation studies to assess the finite-sample performance of the estimated confidence interval given in Section 2.3 and the PLRT developed in Section 2.4.

We used a setup similar to that in Qu et al. (2016). The functional covariate X is defined as

$$X(s) = \sum_{k=1}^{50} \xi_k U_k \phi_k(s),$$

where U_k are independently sampled from the uniform distribution on $[-3, 3]$, $\xi_k = (-1)^{k+1} k^{-1/2}$, $\phi_1 = 1$, and $\phi_{k+1}(s) = \sqrt{2}\cos(k\pi s)$ for $k \geq 1$.

The functional coefficient β_0 is $\beta_0(t) = 9/(50 - 45t) - 0.9$, which is in the Sobolov space $\mathcal{H}^{(2)}(\mathbb{I})$. The penalty function is $J(\beta, \beta) = \int_{\mathbb{I}} (\beta^{(2)}(t))^2\, dt$. The scalar covariate Z is set to be univariate with distribution $N(0, 1)$ and the corresponding coefficient $\theta = 1$. The failure time T is generated from the functional Cox model:

$$h(t|W) = h_0(t) \exp\left\{\theta' Z + \int_0^1 X(s)\beta_0(s) ds\right\},$$

where $h_0(t) = t^2$. Given W, T follows a Weibull distribution. The censoring time C is generated independently, following an exponential distribution with parameter γ which controls the censoring rate. Here, $\gamma = 15$ and 3.9 lead censoring rates around 12% and 33%, respectively. We consider the sample sizes $n = 200$ and 400. We adopt the cubic spline functions for the estimation of the functional covariate. The number of knots is at the order of $q_n = [2n^{1/5}]$, and the knots are equally spaced. The smoothing parameter λ is 10^{-6} and the order m of Sobolev space is 2. For each combination of censoring rate and n, the simulation is repeated 1000 times.

Figure 2.1 displays an instance of estimated $\beta(\cdot)$ and the pointwise 95% confidence intervals among 1000 simulations. The pointwise average of the estimated $\beta(\cdot)$ and

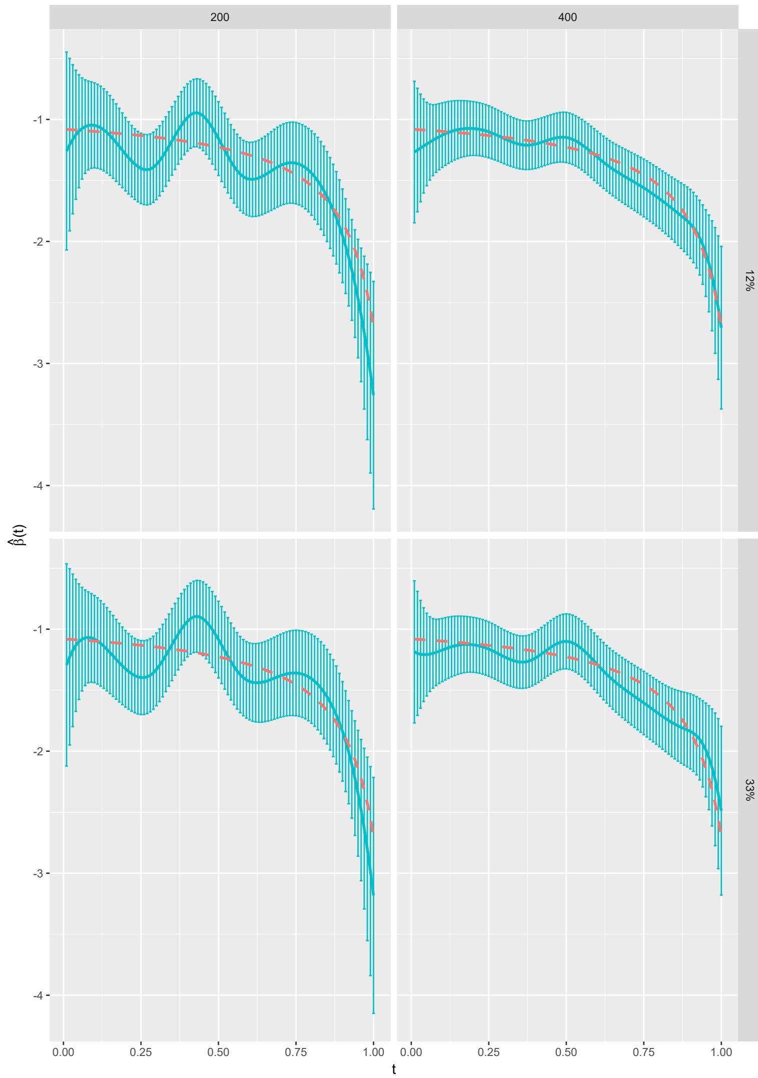

Figure 2.1: Graphical displays of $\hat{\beta}(\cdot)$ and the pointwise 95% confidence intervals of $\beta(t)$. The dashed lines represent $\beta(\cdot)$ whereas the solid lines represent $\hat{\beta}(\cdot)$.

the empirical coverage probability of the 95% pointwise confidence interval based on 1000 simulations are shown in 2.2 and 2.3, respectively. Table 2.1 reports the bias (BIAS), the sample standard error of the estimates (SSE), the average of the estimated standard errors (ESE), and the empirical coverage probability (CP) at $t = 0.1, 0.5, 0.9$. The simulation results are consistent with Theorem 2.3. In particular, these results suggest that the estimate $\hat{\beta}(\cdot)$ is consistent. In general, it is apparent that when n increases from 200 to 400 with a fixed censoring rate, the average bias

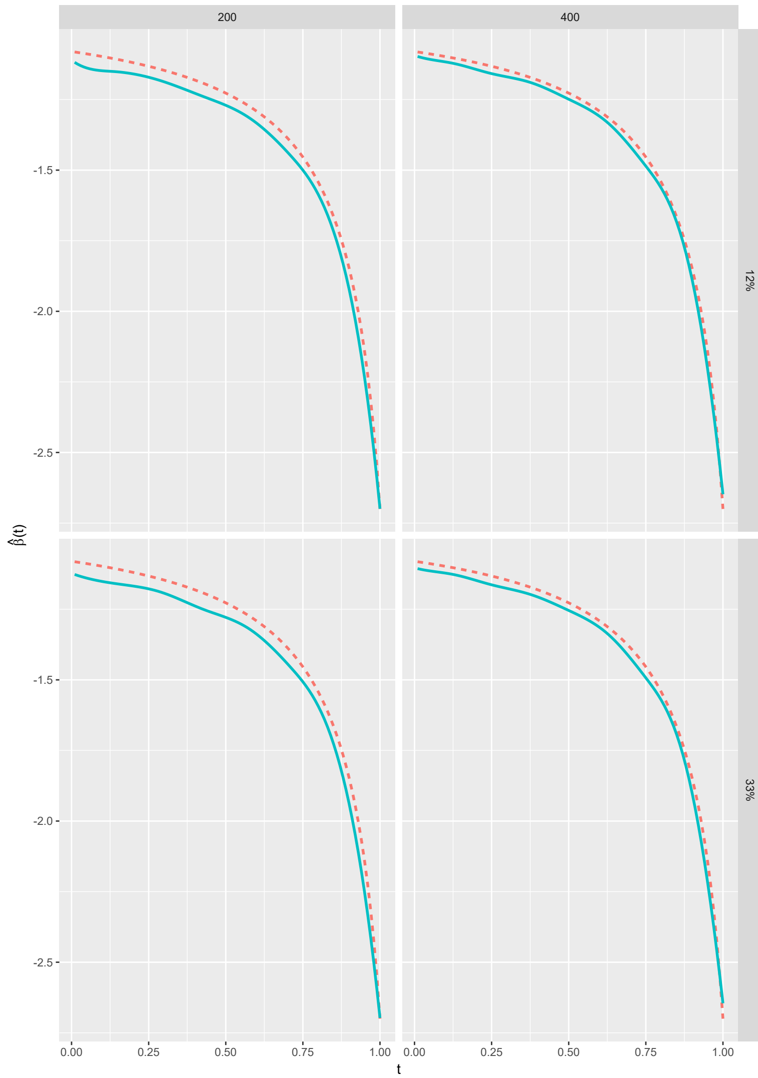

Figure 2.2: Graphical displays of the pointwise averages $\hat{\beta}(\cdot)$. The dashed lines represent $\beta(\cdot)$ whereas the solid lines represent the pointwise averages of $\hat{\beta}(\cdot)$.

and the standard error decrease steadily. Furthermore, the coverage probability also approaches the theoretical value of 95%. The average ESE at 12% censoring rate is lower in comparison to that at 33% censoring rate. This is consistent with the expectation that the lower the censoring rate is, the more accurate the estimate becomes.

For the regression coefficient of the scalar covariate, the BIAS, SSE, ESE, and CP of the estimated $\hat{\theta}$ are given in Table 2.2 for each combination of censoring

Figure 2.3: Graphical displays of the pointwise coverage probabilities (CP). The dashed lines represent 95% whereas the solid lines represent the pointwise CP of $\beta(\cdot)$.

rate and n based on 1000 simulation. As the sample size increases, the average of $\hat{\theta}$ approaches to the true value, the standard deviation reduces, and the coverage probability approaches to 95% given a fixed censoring rate. Similarly, we observe these trends when the censoring rate reduces for a given sample size.

In summary, the simulation results in Tables 2.1 and 2.2 suggest that the estimates of both scalar and functional parameters are consistent and the proposed variance estimation procedure provides reasonable estimates. Furthermore, the results on the empirical coverage probability suggest that the normal approximation is appropriate.

To study the performance of the partial likelihood ratio test, we calculate the estimated sizes and powers of the PLRT under $H_0 : \alpha = (\theta, \beta(\cdot))$, that is, the percentages of rejecting H_0. We generate α under different signal strengths. Specifically, $\alpha = (\theta + c, \beta(\cdot) + c)$, where $c = 0.0, 0.1, 0.3, 0.5$. Table 2.3 summarizes the percentages of rejecting H_0 over 1000 simulations. These results demonstrate the good performance of the PLRT. The power of the test increases as sample size n increases, and the power slightly decreases as the censoring rate increases.

Table 2.1: Simulation results for the proposed estimate of $\beta(t)$.

		$n = 200$			$n = 400$	
	0.1	0.5	0.9	0.1	0.5	0.9
12% BIAS	-0.0504	-0.0431	-0.0747	-0.0189	-0.0218	-0.0400
SSE	0.1518	0.1372	0.1751	0.1042	0.1088	0.1223
ESE	0.1927	0.1602	0.2156	0.1343	0.1117	0.1501
CP	0.9750	0.9680	0.9740	0.9840	0.9510	0.9820
33% BIAS	-0.0539	-0.0514	-0.0914	-0.0241	-0.0270	-0.0531
SSE	0.1704	0.1578	0.1919	0.1245	0.1269	0.1419
ESE	0.1999	0.1658	0.2238	0.1391	0.1158	0.1558
CP	0.9750	0.9510	0.9560	0.9690	0.9200	0.9570

Table 2.2: Simulation results for the proposed estimate of θ.

		$n = 200$	$n = 400$
12%	BIAS	0.0339	0.0154
	SSE	0.1070	0.0717
	ESE	0.1170	0.0811
	CP	0.9520	0.9690
33%	BIAS	0.0392	0.0212
	SSE	0.1261	0.0811
	ESE	0.1224	0.0849
	CP	0.9240	0.9500

Table 2.3: The simulated sizes and powers of the likelihood ratio test for $H_0 : \alpha = (\theta, \beta)$.

	c	200	400
12%	0.0	0.0510	0.0410
	0.1	0.2320	0.5680
	0.3	1.0000	1.0000
	0.5	1.0000	1.0000
33%	0.0	0.0510	0.0490
	0.1	0.1610	0.4570
	0.3	0.9950	1.0000
	0.5	1.0000	1.0000

2.6 An Application

In this section, we apply the proposed method to the Sequential Organ Failure Assessment (SOFA) data collected from the Improving Care of Acute Lung Injury Patients (ICAP) study (Gellar et al. (2014, 2015)). The primary goal of this prospective cohort study is to investigate the long-term complications of patients who suffer from acute lung injury/acute respiratory distress syndrome (ALI/ARDS).

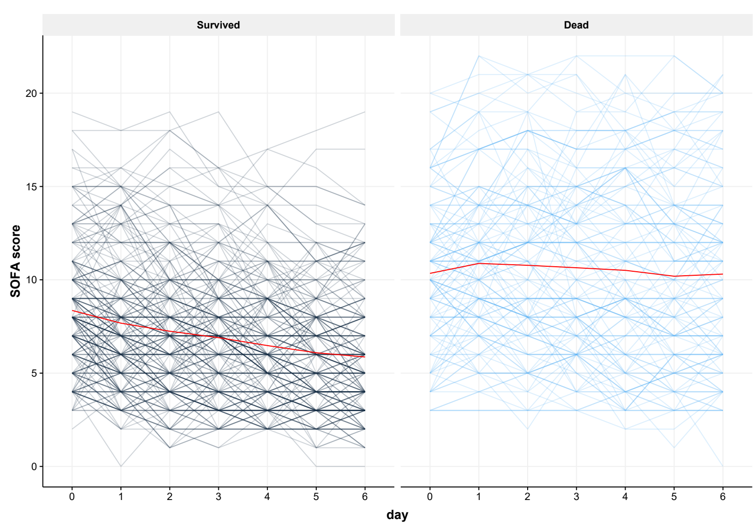

Figure 2.4: Trajectories of the SOFA score of subjects who died after the first week of the ICU hospitalization and those who survived. The red lines are the pointwise average of the SOFA score.

The ICAP study involves 520 subjects. Among them, 237(46%) subjects die in the intensive care unit (ICU). We are interested in the association between the SOFA scores and survival among the subjects who were hospitalized in ICU for more than a week. Out of the 520 subjects, 161 subjects (31.0%) died within the first week in ICU, and they are excluded from the analysis. Therefore, the proposed method is applied to the remaining 359 subjects. In the ICAP study, data were recorded once the patients were admitted in the ICU, and then daily during hospitalization. The SOFA score is one of the measurements recorded daily and it is a measure of the overall organ function status of a patient. It is composed of respiratory, cardiovascular, coagulation, liver, renal, and neurological components. Each component ranges from 0 to 4, with higher scores suggesting inferior organ function. The SOFA score, ranging from 0 to 24, is then the sum of these six scores. We treat the history of each subject's SOFA scores, in the first week, as a functional covariate, $X(s)$, where s is the number of days since the admission to the ICU. Trajectories of the SOFA score of subjects who died after the first week of ICU hospitalization and those who survived are depicted in Figure 2.4. It is apparent that among patients who manage to survive, the pointwise averages of SOFA scores are declining, whereas among patients who died after the first week of ICU hospitalization, the averages are relatively stable. Our model includes three scalar covariates as controls of a subject's baseline risk. They are age, gender, and Charlson co-morbidity index (Charlson et al. (1987)).

Our goal is to estimate the association between the trajectory of SOFA score and mortality among subjects who are hospitalized in ICU for more than a week. We adopt the cubic spline functions for the estimation of the functional covariate. The number of knots is at the order of $q_n = [2n^{1/5}] = 7$, and the knots are equally spaced. As pointed out in Verweij and Van Houwelingen (1993), typical optimization criteria, such as Mallow's C_p and Allen's PRESS (predicted residual error sum of squares) statistic, are inappropriate for the Cox model. Verweij and Van Houwelingen (1993) proposed the cross-validated log likelihood (CVL) to optimize the smoothing parameter of a penalized partial likelihood. Let $\hat{\theta}^{\lambda}_{(-i)}$ be the value of θ that maximizes $l_{\lambda,(-i)}$, the penalized log partial likelihood when observation i is omitted. Given a value of λ, the CVL is given by $CVL_{\lambda} = \sum_{i=1}^{n} l_{\lambda,i}(\theta^{\lambda}_{(-i)})$, where $l_{\lambda,i}(\cdot) = l_{\lambda,i}(\cdot) - l_{\lambda,(-i)}(\cdot)$ is the contribution of subject i to the penalized log partial likelihood. The smoothing parameter$\lambda = 10^{-3}$ leads to the optimal penalty according CVL. One may also consider less computationally intensive methods such as AIC Gellar et al. (2015) and GCV Qu et al. (2016).

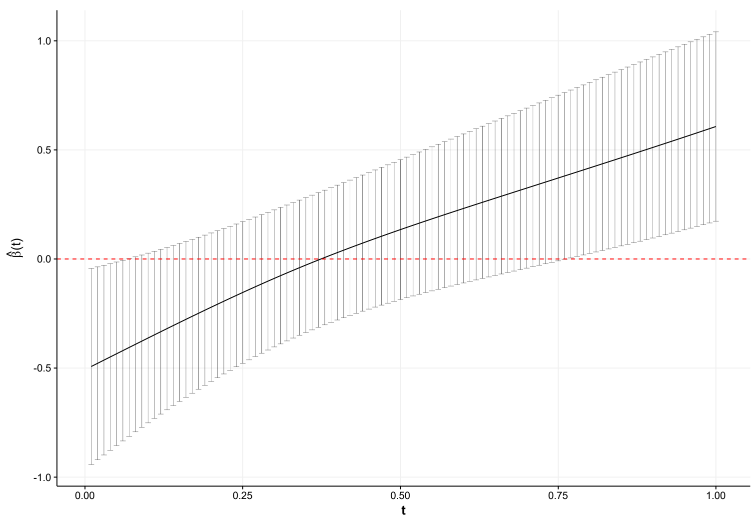

Figure 2.5: The estimated coefficient function $\hat{\beta}(\cdot)$ and the pointwise 95% confidence interval for the SOFA data analysis.

We plot the estimated coefficient function $\hat{\beta}(\cdot)$ in Figure 2.5. The result suggests that there is a functional association between time to death during the ICU stay and the SOFA score function for $t \in [0.75, 1]$, which corresponds to the sixth and the seventh day of ICU stay. This implies that the SOFA score in last two days in the first week of ICU stay may be used as an indicator of one's hazard.

Table 2.4 summarizes the estimation of the regression coefficients of the scalar covariates. In addition to the functional covariate, there seems to be a positive association with two scalar covariates: patients' age and Charlson co-morbidity index.

Table 2.4: Estimation results of regression coefficients of scalar covariates for the SOFA data analysis

	$\hat{\theta}$	$S.E.$	t-value	p-value
Age	0.0151	0.0015	10.0667	< 0.0001
Gender (male=1)	0.1640	0.1331	1.2322	0.1089
Charlson Index	-0.0348	0.0034	-10.2353	< 0.0001

On the other hand, the gender shows no significant association with the hazard of death.

2.7 Appendix

For ease of presentation, we introduce some notations related to the Fréchet derivatives. Let $\mathcal{S}_n(\alpha)$ and $\mathcal{S}_{n,\lambda}(\alpha)$ be the Fréchet derivatives of $l_n(\alpha)$ and $l_{n,\lambda}(\alpha)$, respectively. Denote the asymptotic value of $l_n(\alpha)$ as $l(\alpha)$, and $l(\alpha) - \lambda J(\beta, \beta)/2$ as $l_\lambda(\alpha)$. Similarly, let $\mathcal{S}(\alpha)$ and $\mathcal{S}_\lambda(\alpha)$ be the Fréchet derivatives of $l(\alpha)$ and $l_\lambda(\alpha)$, respectively. Let D be the Fréchet derivative operator and $\alpha_i = (\theta_i^\top, \beta_i(\cdot)), i = 1, 2, 3 \in \mathcal{H}$ be any direction. Then, we have

$$
\begin{aligned}
\mathcal{S}_{n,\lambda}(\alpha)\alpha_1 &= \frac{1}{n}\sum_{i=1}^{n} \Delta_i \left[\eta_{\alpha_1}(W_i) - \frac{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_1}(W_j)}{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}} \right] - \lambda J(\beta, \beta_1) \\
&\equiv \mathcal{S}_n(\alpha)\alpha_1 - \lambda J(\beta, \beta_1),
\end{aligned}
$$

$$
\begin{aligned}
& D\mathcal{S}_{n,\lambda}(\alpha)\alpha_1\alpha_2 \\
&= -\frac{1}{n}\sum_{i=1}^{n} \Delta_i \left[\frac{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_1}(W_j)\eta_{\alpha_2}(W_j)}{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}} \right. \\
&\left. - \frac{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_1}(W_j)\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_2}(W_j)}{[\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}]^2} \right] - \lambda J(\beta_1, \beta_2) \\
&\equiv D\mathcal{S}_n(\alpha)\alpha_1\alpha_2 - \lambda J(\beta_1, \beta_2),
\end{aligned}
$$

and

$$D^2\mathcal{S}_{n,\lambda}(\alpha)\alpha_1\alpha_2\alpha_3 = -\frac{1}{n}\sum_{i=1}^{n}\Delta_i\Big[\frac{\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_1}(W_j)\eta_{\alpha_2}(W_j)\eta_{\alpha_3}(W_j)}{\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}}$$

$$-\frac{\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_1}(W_j)\eta_{\alpha_2}(W_j)\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_3}(W_j)}{[\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}]^2}$$

$$-\frac{\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_1}(W_j)\eta_{\alpha_3}(W_j)\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_2}(W_j)}{[\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}]^2}$$

$$-\frac{\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_1}(W_j)\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_2}(W_j)\eta_{\alpha_3}(W_j)}{[\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}]^2}$$

$$+2\frac{\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_1}(W_j)\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_2}(W_j)}{[\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}]^3}$$

$$\times\sum_{j=1}^{n}\mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_3}(W_j)\Big]$$

$$\equiv D^2\mathcal{S}_n(\alpha)\alpha_1\alpha_2\alpha_3.$$

There exists a sequence of functions $\omega_k, k = 1, 2, \ldots, p$, such that $< \omega_k, \beta >_m = V(G_k, \beta)$. A direct calculation yields that $\omega_k(\cdot) \equiv \sum_{j=1}^\infty G_{jk}\frac{h_j(\cdot)}{(1+\lambda\rho_j)}$. Let ω be $(\omega_1, \omega_2, \ldots, \omega_p)^\top$. Then, $\omega = (id - W_\lambda)G$. Further, following from the Riesz representation theorem, there exists an element in $\mathcal{H}^{(m)}$, denoted as π_x, such that $< \pi_x, \beta >_m = \int_{\mathbb{I}} x(t)\beta(t)\,dt$. Through direct calculations, we have $\pi_x = \sum_{j=1}^\infty \int_{\mathbb{I}} x(t)h_j(t)\,dt/(1+\lambda\rho_j)h_j(\cdot)$. If we denote (H_w, T_w) with $w = (z, x(\cdot))$ as $\mathcal{R}_w$, where

$$H_w = \{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}^{-1}\{z - V(G, \pi_x)\},$$

$$T_w = \pi_x - \omega^\top\{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}^{-1}\{z - V(G, \pi_x)\}, \quad \text{and,}$$

then we have $< \mathcal{R}_w, \alpha >_\lambda = \theta^\top z + \int_{\mathbb{I}} x(t)\beta(t)\,dt$.

Define $\tilde{\mathcal{R}}_u$ as $\tilde{\mathcal{R}}_u : u \to (\tilde{H}_u, \tilde{T}_u) \in \mathcal{H}$, where $u = (z^\top, t)$,

$$\tilde{H}_u = \{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}^{-1}(z - \omega(t)), \quad \text{and}$$

$$\tilde{T}_u = K_t - \omega^\top\{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}^{-1}(z - \omega(t)).$$

Then we have that $< \tilde{\mathcal{R}}_u, \alpha >_\lambda = \theta^\top z + \beta(t)$.

Define $\mathcal{P}_\lambda \alpha = (\tilde{H}_\alpha, \tilde{T}_\alpha)$, where

$$\tilde{H}_\alpha = -\{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}^{-1} V(G, W_\lambda \beta), \quad \text{and}$$

$$\tilde{T}_\alpha = W_\lambda \beta - \{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}^{-1} V(G, W_\lambda \beta).$$

Then $\mathcal{P}_\lambda \alpha \in \mathcal{H}$ and $< \mathcal{P}_\lambda \alpha, \alpha_1 >_\lambda = < W_\lambda \beta, \beta_1 >_m$ for any $\alpha_1 = (\theta_1^\top, \beta_1) \in \mathcal{H}$. It follows from the Cauchy-Schwarz inequality that $\|\mathcal{P}_\lambda\|_\lambda \leq 1$ and $\mathcal{P}_\lambda$ is self-adjoint.

2.7.1 Proofs of Lemmas

Lemma 2.1. *Under Condition (C1), we have* $D\mathcal{S}_\lambda(\alpha_0) = -id$*, where* id *is the identity operator.*

This result follows directly from the definitions of the inner product and $D\mathcal{S}_\lambda(\alpha_0)$. Denote $\|\alpha\|_e = \|\theta\|_2 + \|\beta\|_{L_2}$. The following lemma provides the relationship between the general Euclidean norm $\|\cdot\|_e$ and the norm $\|\cdot\|_\lambda$.

Lemma 2.2. *There exists a constant* $\kappa > 0$ *such that, for any* $\alpha \in \mathcal{H}$, $\|\alpha\|_e \leq \kappa h^{-(2a+1)/2} \|\alpha\|_\lambda$.

Proof of Lemma 2.2 It follows from the definition of K_t that, $||K_t||_\lambda \lesssim h^{-(a+1/2)}$. Follows the line of the proof of Lemma 2.4, we have that $||\tilde{\mathcal{R}}_u||_\lambda \lesssim h^{-(a+1/2)}$.

It follows from the fact that

$$\begin{aligned}
||\alpha||_e =& ||\theta||_2 + ||\beta||_{L_2} \\
\leq& ||\theta||_2 + ||\beta||_{\sup} \\
=& \sup_{||z||_2=1, t\in\mathbb{I}} |\beta(t) + \theta^\top z| \\
=& \sup_{||z||_2=1, t\in\mathbb{I}} < \tilde{\mathcal{R}}_u, \alpha >_\lambda \\
\leq& ||\alpha||_\lambda \sup_{\|z\|_2=1, t\in\mathbb{I}} \|\tilde{\mathcal{R}}_u\|_\lambda \\
\lesssim& h^{-(a+1/2)} ||\alpha||_\lambda.
\end{aligned}$$

Lemma 2.3. *Suppose that Conditions (C1)-(C4) hold. Then, for any $\alpha \in \mathcal{H}$,*

$$E(| < \mathcal{R}_{\mathcal{W}}, \alpha >_{\lambda}^{4} |) \leq c_4 \|\alpha\|_{\lambda}^{4}.$$

Proof of Lemma 2.3 By Condition (C4) and Condition (C1)(ii), we have

$$E_W(< R_W, \alpha >_\lambda)^4 = E_W \left\{ \theta^\top Z + \int_{\mathbb{I}} X(t)\beta(t)\,dt \right\}^4$$

$$\leq M_0 \left\{ E_W |\theta^\top Z + \int_{\mathbb{I}} X(t)\beta(t)\,dt|^2 \right\}^2$$

$$\lesssim \left\{ \int_0^\tau Var[\eta_\alpha(W)|T = v, \Delta = 1] E\,dN(v) \right\}^2 \lesssim \|\alpha\|_\lambda^4.$$

Lemma 2.4. *Suppose that Conditions (C1)-(C3) hold. Then for any $x \in L_2([0,1])$, there exists a universal positive constant c_r such that*

$$< \mathcal{R}_{Z,x}, \mathcal{R}_{Z,x} >_\lambda \leq c_r(\|Z\|_2^2 + \|x\|_{L_2}^2 h^{-2a-1}) \quad \textit{and} \quad E\{\|\mathcal{R}_W\|_\lambda^2\} \leq c_r h^{-1}.$$

Proof of Lemma 2.4 A direct calculation yields that

$$< \mathcal{R}_w, \mathcal{R}_w >_\lambda = z^\top \{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}^{-1} z + < \pi_x, \pi_x >_m + V(G, W_\lambda G^\top)\}^{-1}$$
$$\times V(G, \pi_x) + V(G, \pi_x)^\top \{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}^{-1} V(G, \pi_x). \quad (2.4)$$

It follows from Condition (C3) that $\{\Sigma - V(G, G^\top)\}^{-1}$ is positive definite and $V(G, W_\lambda G^\top) \to 0$. Let c denote the minimum eigenvalue of $\{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}$. Then we have $\{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}^{-1} \leq c^{-1}\mathbf{1}$ with $\mathbf{1}$ being the identity matrix. Thus, we have $z^\top \{\Sigma - V(G, G^\top) + V(G, W_\lambda G^\top)\}^{-1} z \lesssim \|z\|_2^2$. A direct calculation yields that $< \pi_X, \pi_X >_m \lesssim \|X\|_{L^2}^2 h^{-2a-1}$ and $V(G, \pi_X) \lesssim \|X\|_{L^2} h^{-a-1/2}$. Thus, there exists a constant $c_r > 0$ such that

$$< R_W, R_W >_\lambda \leq c_r(\|Z\|_2^2 + \|X\|_{L^2}^2 h^{-2a-1}).$$

Besides, it follows from (3.17) and the proof of Lemma S.4 in Shang and Cheng (2015) that

$$E < R_W, R_W >_\lambda \leq c_r h^{-1}.$$

The proof is completed.

Define $\mathcal{F}_{p_n} = \{\alpha = (\theta^\top, \beta(\cdot)) \in \mathcal{H} : \|\theta\|_2 \le 1, \|\beta\|_{L_2} \le 1, J(\beta,\beta) \le p_n\}$,

$$H_n(\alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\phi_n(Y_i, W_i; \alpha) R_{W_i} - E\phi_n(Y_i, W_i; \alpha) R_{W_i}],$$

where $\phi_n(Y_i, W_i; \alpha)$ is a function on the data set and parameters, which might depend on n. To derive the rate of convergence, we need the following concentration inequality as a preliminary step.

Lemma 2.5. *Suppose that Conditions (C1)-(C4) hold. If $\phi_n(Y_i, W_i; \boldsymbol{0}) = 0$ a.s., and there exists a constant $C_\phi > 0$ such that*

$$|\phi_n(Y_i, W_i; \alpha_1) - \phi_n(Y_i, W_i; \alpha_2)| \le C_\phi \|\alpha_1 - \alpha_2\|_e, \quad \textit{for any} \quad \alpha_1, \alpha_2 \in \mathcal{H},$$

then we have

$$\lim_n P\Big(\sup_{\alpha \in \mathcal{F}_{p_n}} \frac{\|H_n(\alpha)\|_\lambda}{p_n^{1/(4m)} \|\alpha\|_e^\gamma + n^{-1/2}} \le \{5h^{-1} \log\log(n)\}^{1/2} \Big) = 1$$

where $\gamma = 1 - 1/(2m)$.

Proof of Lemma 2.5 Denote $N(\delta, \mathcal{F}_{p_n}, \|\cdot\|_2)$ as the $\delta-$covering number of the function class $\mathcal{F}_{p_n}$, in terms of $\|\cdot\|_2-$ norm. Then it follows from Theorem 9.20 of Kosorok (2008) that

$$\begin{aligned}
\log N(\delta, \mathcal{F}_{p_n}, \|\cdot\|_2) &\le N(\delta, p_n^{1/2} \mathcal{F}_1, \|\cdot\|_2) \\
&\le N(p_n^{-1/2}\delta, \mathcal{F}_1, \|\cdot\|_2) \\
&\lesssim \max\{(p_n^{-1/2}\delta)^{-1/m}, (p_n^{-1/2}\delta)^{-1/p}\},
\end{aligned}$$

where p is the dimension of θ. Thus, the conclusion of this lemma follows from

$$\exp(-c \max\{(p_n^{-1/2}\delta)^{-1/m}, (p_n^{-1/2}\delta)^{-1/p}\}) \le \exp(-c(p_n^{-1/2}\delta)^{-1/m})$$

and the proof of Lemma 3.4 in Shang and Cheng (2015).

2.7.2 Proofs of Theorems

Proof of Theorem 2.1

In order to prove Theorem 2.1, we need the following subset for $\mathcal{H}$. Define

$$\mathcal{F}_{p_n} = \{\alpha = (\theta^\top, \beta(\cdot)) \in \mathcal{H} : \|\theta\|_2 \leq 1, \|\beta\|_{L_2} \leq 1, J(\beta, \beta) \leq p_n\}.$$

First, we show that there exists a unique α_λ such that $\mathcal{S}_\lambda(\alpha_\lambda) = 0$. Let $r_{1n} = 2\{J(\beta_0, \beta_0) + 1\}^{1/2}h^k$, and define the operator: $T_{1h}(\alpha) = \alpha + \mathcal{S}_\lambda(\alpha_0 + \alpha), \alpha \in \mathcal{H}$. Then,

$$\|T_{1h}(\alpha)\|_\lambda = \|\alpha + \mathcal{S}_\lambda(\alpha + \alpha_0)\|_\lambda \leq \|\alpha + \mathcal{S}_\lambda(\alpha + \alpha_0) - \mathcal{S}_\lambda(\alpha_0)\|_\lambda + \|\mathcal{S}_\lambda(\alpha_0)\|_\lambda.$$

Let $\mathbb{B}(\epsilon) = \{\alpha \in \mathcal{H}, \|\alpha\|_\lambda \leq \epsilon\}$ be the ball of radius ϵ in $\mathcal{H}$. Note that $\mathcal{S}(\alpha_0) = 0$, which implies that $\mathcal{S}_\lambda(\alpha_0) = -\mathcal{P}_\lambda \alpha_0$. It follows from the Cauchy-Schwarz inequality that

$$\|\mathcal{S}_\lambda(\alpha_0)\|_\lambda = \|\mathcal{P}_\lambda \alpha_0\|_\lambda \leq \{\lambda J(\beta_0, \beta_0)\}^{1/2} \leq \{J(\beta_0, \beta_0) + 1\}^{1/2} h^k = \frac{r_{1n}}{2}. \tag{2.5}$$

By Lemma 2.1, we have

$$\begin{aligned}
\|\alpha + \mathcal{S}_\lambda(\alpha + \alpha_0) - \mathcal{S}_\lambda(\alpha_0)\|_\lambda =& \|\alpha + D\mathcal{S}_\lambda(\alpha_0)\alpha + \int_0^1 \int_0^1 sD^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha \, ds \, ds'\|_\lambda \\
=& \|\int_0^1 \int_0^1 sD^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha \, ds \, ds'\|_\lambda \\
\leq& \int_0^1 \int_0^1 s\|D^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha\|_\lambda \, ds \, ds'.
\end{aligned} \tag{2.6}$$

From the definition of $D^2\mathcal{S}_\lambda(\alpha)$, Lemmas 2.2 and 2.4, and Condition (C1), we have

$$\|D^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha\|_\lambda \lesssim \{E < \mathcal{R}_W, \alpha >_\lambda^4\}^{1/2}\{E\|\mathcal{R}_W\|_\lambda^2\}^{1/2} \lesssim \|\alpha\|_\lambda^2 c_r^{1/2} h^{-1/2}. \tag{2.7}$$

From inequalities (2.5), (2.6), and (2.7), we have

$$\|T_{1h}\|_\lambda \leq c\|\alpha\|_\lambda^2 c_r^{1/2} h^{-1/2} + \frac{r_{1n}}{2}. \tag{2.8}$$

Since $h = o(1)$ and $k > a + 1/2 \geq 1/2$, we have $r_{1n}h^{-1/2} = o(1)$. Then for any $\alpha \in \mathbb{B}(r_{1n})$, $\|T_{1h}\|_\lambda < r_{1n}$ for large n. This implies $T_{1h}(\mathbb{B}(r_{1h})) \subset \mathbb{B}(r_{1h})$. Next, we show that T_{1h} is a contraction mapping. For any $\alpha_j = (\theta_j^\top, \beta_j(\cdot)) \in \mathcal{H}, j = 1, 2$, we have

$$
\begin{aligned}
T_{1h}(\alpha_1) - T_{1h}(\alpha_2) &= \alpha_1 - \alpha_2 + \mathcal{S}_\lambda(\alpha_0 + \alpha_1) - \mathcal{S}_\lambda(\alpha_0 + \alpha_2) \\
&= \int_0^1 [D\mathcal{S}_\lambda\{\alpha_0 + \alpha_2 + s(\alpha_1 - \alpha_2)\} - D\mathcal{S}_\lambda(\alpha_0)](\alpha_1 - \alpha_2)\, ds \\
&= \int_0^1 \int_0^1 s' D^2\mathcal{S}_\lambda[\alpha_0 + s'\{\alpha_2 + s(\alpha_1 - \alpha_2)\}](\alpha_1 - \alpha_2)\{\alpha_2 + s(\alpha_1 - \alpha_2)\}\, ds\, ds'.
\end{aligned}
$$

By the similar arguments adopted in proving inequality (2.8), we have

$$
\begin{aligned}
&\|T_{1h}(\alpha_1) - T_{1h}(\alpha_2)\|_\lambda \\
&\leq \int_0^1 \int_0^1 s'\|D^2\mathcal{S}_\lambda[\alpha_0 + s'\{\alpha_2 + s(\alpha_1 - \alpha_2)\}](\alpha_1 - \alpha_2)\{\alpha_2 + s(\alpha_1 - \alpha_2)\}\|_\lambda ds ds' \\
&\lesssim \int_0^1 \int_0^1 s'\{E < \mathcal{R}_W, \alpha_1 - \alpha_2 >_\lambda^4\}^{1/4}\{E\|\mathcal{R}_W\|_\lambda^2\}^{1/2}\{E < \mathcal{R}_W, \alpha_2 + s(\alpha_1 - \alpha_2) >_\lambda^4\}^{1/4}\, ds\, ds' \\
&\lesssim \|\alpha_1 - \alpha_2\|_\lambda c_r^{1/2} h^{-1/2}(\|\alpha_1 - \alpha_2\|_\lambda + \|\alpha_2\|_\lambda) \\
&\lesssim r_{1n}\|\alpha_1 - \alpha_2\|_\lambda c_r^{1/2} h^{-1/2} \\
&\leq 1/2\|\alpha_1 - \alpha_2\|_\lambda.
\end{aligned}
$$

The last inequality follows from the fact that $r_{1n}h^{-1/2} = o(1)$. Then $T_{1h}(\alpha)$ is a contraction mapping on $\mathbb{B}(r_{1n})$. By the Banach fixed-point theorem, there exists a unique $\alpha'_\lambda \in \mathbb{B}(r_{1n})$ such that $T_{1h}(\alpha'_\lambda) = \alpha'_\lambda$. Define $\alpha_\lambda = \alpha'_\lambda + \alpha_0$. Then $\mathcal{S}_\lambda(\alpha_\lambda) = 0$ and $\|\alpha_\lambda - \alpha_0\|_\lambda \leq r_{1n}$.

Next, we show that there exists a unique $\hat{\alpha}_{n,\lambda}$ such that $\mathcal{S}_{n,\lambda}(\hat{\alpha}_{n,\lambda}) = 0$. Since $\|\alpha_\lambda - \alpha_0\|_\lambda = O(r_{1n}) = o(1)$ and $D\mathcal{S}_\lambda(\alpha_0) = -id$, it follows from the Taylor expansion and inequality (2.7) that $D\mathcal{S}_\lambda(\alpha_\lambda)$ is invertible. By applying similar arguments employed in Shang and Cheng (2015), we have $\|D\mathcal{S}_\lambda(\alpha_\lambda)\|_\lambda \in (1/2, 3/2)$. Now define the operator

$$
\begin{aligned}
&T_{2h}(\alpha) = \alpha - \{D\mathcal{S}_\lambda(\alpha_\lambda)\}^{-1}\mathcal{S}_{n,\lambda}(\alpha_\lambda + \alpha)\\
&= -\{D\mathcal{S}_\lambda(\alpha_\lambda)\}^{-1}\{D\mathcal{S}_{n,\lambda}(\alpha_\lambda)\alpha - D\mathcal{S}_\lambda(\alpha_\lambda)\alpha\}\\
&\quad - \{D\mathcal{S}_\lambda(\alpha_\lambda)\}^{-1}\{\mathcal{S}_{n,\lambda}(\alpha_\lambda+\alpha) - \mathcal{S}_{n,\lambda}(\alpha_\lambda) - D\mathcal{S}_{n,\lambda}(\alpha_\lambda)\alpha\} - \{D\mathcal{S}_\lambda(\alpha_\lambda)\}^{-1}\mathcal{S}_{n,\lambda}(\alpha_\lambda)\\
&\equiv I_1 + I_2 + I_3.
\end{aligned}
$$

It follows from the functional central limit theorem that uniformly in $t \in \mathbb{I}$

$$
\Big\|\frac{1}{n}\sum_{j=1}^{n}\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\} - s_1^{(0)}(t,\alpha_\lambda)\Big\|_\infty = O_p(n^{-1/2}). \tag{2.9}
$$

By Lemma 2.3 and the functional central limit theorem, we have

$$
\begin{aligned}
&\Big\|\frac{1}{n}\sum_{j=1}^{n}\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\mathcal{R}_{W_j} - E[\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\mathcal{R}_{W_j}]\Big\|_\lambda\\
&= \sup_{\|\alpha_1\|_\lambda=1} < \frac{1}{n}\sum_{j=1}^{n}\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\mathcal{R}_{W_j} - E[\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\mathcal{R}_{W_j}], \alpha_1 >_\lambda\\
&= O_p(n^{-1/2}h^{-a-1/2}).
\end{aligned} \tag{2.10}
$$

It follows from $\mathcal{S}_\lambda(\alpha_\lambda) = 0$, Lemma 2.2, and equations (2.9) and (2.10) that $E\|[D\mathcal{S}_\lambda(\alpha_\lambda)]I_3\|_\lambda^2 = O((hn)^{-1})$. This implies that $\|\mathcal{S}_{n,\lambda}(\alpha_\lambda)\|_\lambda = O_p((nh)^{-1/2})$. Let c be a positive constant satisfying $P(\|\mathcal{S}_{n,\lambda}(\alpha_\lambda)\|_\lambda \le c(nh)^{-1/2}) \to 1$. Define $r_{2n} = 2c(nh)^{-1/2}$ and $\mathbb{B}(r_{2n}) = \{\alpha \in \mathcal{H} : \|\alpha\|_\lambda \le r_{2n}\}$. Then we have $P(\|\mathcal{S}_{n,\lambda}(\alpha_\lambda)\|_\lambda \le r_{2n}/2) \to 1$. Define $\Gamma = \cap_{i=1}^{n} A_{ni}$, where

$$
A_{ni} = \{\|Z_i\|_2 \le c\log(n), \|X_i\|_{L^2} \le c\log(n), \exp\{\eta_{\alpha_\lambda}(W_i)\} \le c\log(n)\}
$$

for a constant c. From Condition (C4), we choose c which is large enough such that

$P(\Gamma) \to 1$ and $P(A_{ni}^c) = O(n^{-1})$. To handle I_1, we have

$$
\begin{aligned}
\|[D\mathcal{S}_\lambda(\alpha_\lambda)]I_1\|_\lambda \le & \Big\| \frac{1}{n}\sum_{i=1}^n \Delta_i \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\mathcal{R}_{W_j}}{S_0^{(1)}(Y_i,\alpha_\lambda)} \\
& - \int_0^\tau s_1^{(0)}(t,\alpha_0)\frac{E[\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\mathcal{R}_{W_j}]}{s_1^{(0)}(t,\alpha_\lambda)} h_0(t)\,dt \Big\|_\lambda \\
& + \Big\| \frac{1}{n}\sum_{i=1}^n \Delta_i \frac{\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\sum_{j=1}^n \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_\lambda}(W_j)\}\mathcal{R}_{W_j}}{[nS_0^{(1)}(Y_i,\alpha_\lambda)]^2} \\
& - \int_0^\tau s_1^{(0)}(t,\alpha_0)\frac{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_\lambda}(W)\}\eta_\alpha(W)]E[\mathcal{Y}(t)\exp\{\eta_{\alpha_\lambda}(W)\}\mathcal{R}_W]}{[s_1^{(0)}(t,\alpha_\lambda)]^2} h_0(t)\,dt \Big\|_\lambda \\
& \equiv I_{11} + I_{12}. \qquad (2.11)
\end{aligned}
$$

For I_{11}, we have

$$
\begin{aligned}
\|I_{11}\|_\lambda \le & \Big\| \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_\lambda)} \\
& - \frac{E[\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\mathcal{R}_{W_j}]}{s_1^{(0)}(t,\alpha_\lambda)}\,dN_i(t) \Big\|_\lambda \\
& + \frac{1}{n}\Big\| \sum_{i=1}^n \int_0^\tau n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\mathcal{R}_{W_j}\{\frac{1}{s_1^{(0)}(t,\alpha_\lambda)} - \frac{1}{S_1^{(0)}(t,\alpha_\lambda)}\}\,dN_i(t) \Big\|_\lambda \\
& + \Big\| \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_\lambda)}\,dN_i(t) \\
& - \int_0^\tau s_1^{(0)}(t,\alpha_0)\frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_\lambda)} h_0(t)\,dt \Big\|_\lambda \\
& \equiv I_{111} + I_{112} + I_{113}.
\end{aligned}
$$

For I_{113}, we have

$$
\begin{aligned}
I_{113} = & \Big\| \int_0^\tau \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_\lambda)} \frac{1}{n}\sum_{i=1}^n \{\,dN_i(t) - E\,dN_i(t)\} \Big\|_\lambda \\
& = O_p((nh)^{-1/2})\|\alpha\|_\lambda.
\end{aligned}
$$

To infer I_{111}, we first define

$$\phi(Y_j, W_j; \alpha) = \frac{\mathcal{Y}_j(t_0)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)}{s_1^{(0)}(t_0, \alpha_\lambda)} I_{A_{nj}}.$$

Then, for any $\alpha_1, \alpha_2 \in \mathcal{H}$, we have

$$\begin{aligned}
|\phi(Y_j, W_j; \alpha_1) - \phi(Y_j, W_j; \alpha_2)| &= \frac{1}{s_1^{(0)}(t_0, \alpha_\lambda)}\mathcal{Y}_j(t_0)\exp\{\eta_{\alpha_\lambda}(W_j)\}|\{\eta_{\alpha_1}(W_j) - \eta_{\alpha_2}(W_j)\}|I_{A_{nj}} \\
&\leq \frac{c\log(n)}{s_1^{(0)}(t_0, \alpha_\lambda)}| < \mathcal{R}_{W_j}, \alpha_1 - \alpha_2 >_\lambda |I_{A_{nj}} \\
&\leq \frac{\{c\log(n)\}^2}{s_1^{(0)}(t_0, \alpha_\lambda)}\|\alpha_1 - \alpha_2\|_e.
\end{aligned}$$

Define $\phi_n(Y_j, W_j; \alpha) = s_1^{(0)}(t_0, \alpha_\lambda)c^{-2}\{\log(n)\}^{-2}\phi(Y_j, W_j; \alpha_1)$. Then

$$|\phi_n(Y_j, W_j; \alpha_1) - \phi_n(Y_j, W_j; \alpha_1)| \leq \|\alpha_1 - \alpha_2\|_e.$$

For any $\alpha \neq 0 \in \mathcal{H}$, let $\tilde{\alpha} = \alpha/(d_n\|\alpha\|_\lambda)$, where $d_n = \kappa h^{-(2a+1)/2}$. It follows from Lemma 2.3 that $\|\tilde{\alpha}\|_e \leq d_n\|\tilde{\alpha}\|_\lambda = 1$. Then we have $\|\tilde{\theta}\|_2 + \|\tilde{\beta}\|_{L_2} \leq 1$. Meanwhile, we have $\lambda J(\tilde{\beta}, \tilde{\beta}) \leq \|\tilde{\alpha}\|_\lambda^2 = d_n^{-2}$. Then $J(\tilde{\beta}, \tilde{\beta}) \leq \lambda^{-1}d_n^{-2} \equiv p_n$. By Lemma 2.5, we obtain that for any $\alpha \in \mathbb{B}(r_{2n})$,

$$\begin{aligned}
\lim_n P(\| &\sum_{j=1}^n [\phi_n(Y_j, W_j; \tilde{\alpha})\mathcal{R}_{W_j} - E\{\phi_n(Y_j, W_j; \tilde{\alpha})\mathcal{R}_{W_j}\}]\|_\lambda \\
&\lesssim (n^{1/2}p_n^{1/(4m)} + 1)\{h^{-1}\log\log(n)\}^{1/2}) = 1.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\lim_n P(\| &\sum_{j=1}^n [\phi(Y_j, W_j; \alpha)\mathcal{R}_{W_j} - E\{\phi(Y_j, W_j; \alpha)\mathcal{R}_{W_j}\}]\|_\lambda \\
&\lesssim d_n\{\log(n)\}^2\|\alpha\|_\lambda(n^{1/2}p_n^{1/(4m)} + 1)\{h^{-1}\log\log(n)\}^{1/2}) = 1.
\end{aligned}$$

It follows from the definition of A_{ni} that

$$\begin{aligned}
\Big\|\frac{E\mathcal{Y}_j(t_0)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\mathcal{R}_{W_j}I_{A_{nj}^c}}{s_1^{(0)}(t, \alpha_\lambda)}\|_\lambda &\leq c_1 E\| < \mathcal{R}_{W_j}, \alpha > \mathcal{R}_{W_j}I_{A_{nj}^c}\Big\|_\lambda \\
&= O(P(A_{ni}^c)^{1/2}h^{-1/2})\|\alpha\|_\lambda = o(1)\|\alpha\|_\lambda.
\end{aligned}$$

Thus, we have $I_{111} = O_p(n^{-1/2}h^{-(a+1)-\frac{2k-2a-1}{4m}}\{\log(n)\}^2\{\log\log(n)\}^{1/2})\|\alpha\|_\lambda + o_p(1)\|\alpha\|_\lambda = o_p(1)\|\alpha\|_\lambda$. From Lemma 2.4, we have that

$$\frac{1}{n}\Big\|\sum_{i=1}^{n}\int_0^\tau E\big[\frac{1}{n}\sum_{j=1}^{n}\mathcal{Y}_j(t)\exp\{\eta_{\alpha_\lambda}(W_j)\}\eta_\alpha(W_j)\mathcal{R}_{W_j}\big]\,dN_i(t)\Big\|_\lambda = O_p(h^{-1/2}\|\alpha\|_\lambda).$$

Then, by equation (2.9) and $(nh)^{-1} = o(1)$, we have $I_{112} = o_p(1)\|\alpha\|_\lambda$.

Applying the approach in deriving I_{11}, we have that $I_{12} = o_p(1)\|\alpha\|_\lambda$. Therefore, for any $\alpha \in \mathbb{B}(r_{2n})$, $\|[D\mathcal{S}_\lambda(\alpha_\lambda)]I_1\| \leq r_{2n}/18$. For $\|[D\mathcal{S}_\lambda(\alpha_\lambda)]I_2\|_\lambda$, we have

$$\begin{aligned}\|[D\mathcal{S}_\lambda(\alpha_\lambda)]I_2\|_\lambda &= \|\{\mathcal{S}_{n,\lambda}(\alpha_\lambda+\alpha) - \mathcal{S}_{n,\lambda}(\alpha_\lambda) - D\mathcal{S}_{n,\lambda}(\alpha_\lambda)\alpha\}\|_\lambda \\ &= \|\int_0^1\int_0^1 sD^2\mathcal{S}_{n,\lambda}(\alpha_\lambda + ss'\alpha)\alpha\alpha\,ds\,ds'\|_\lambda.\end{aligned}$$

It follows from inequality (2.7) that

$$\begin{aligned}&\|D^2\mathcal{S}_{n,\lambda}(\alpha_\lambda+ss'\alpha)\alpha\alpha\|_\lambda \\ &\leq \|D^2\mathcal{S}_{n,\lambda}(\alpha_\lambda+ss'\alpha)\alpha\alpha - D^2\mathcal{S}_\lambda(\alpha_\lambda+ss'\alpha)\alpha\alpha\|_\lambda + \|D^2\mathcal{S}_\lambda(\alpha_\lambda+ss'\alpha)\alpha\alpha\|_\lambda \\ &= \|D^2\mathcal{S}_{n,\lambda}(\alpha_\lambda+ss'\alpha)\alpha\alpha - D^2\mathcal{S}_\lambda(\alpha_\lambda+ss'\alpha)\alpha\alpha\|_\lambda + O(h^{-1/2})\|\alpha\|_\lambda^2.\end{aligned}$$

By employing the arguments in obtaining I_{111}, we have that

$$\begin{aligned}&\|D^2\mathcal{S}_{n,\lambda}(\alpha_\lambda+ss'\alpha)\alpha\alpha - D^2\mathcal{S}_\lambda(\alpha_\lambda+ss'\alpha)\alpha\alpha\|_\lambda \\ &= O_p\Big(n^{-1}h^{-(2a+1)-\frac{2k-2a-1}{4m}}\log(n)^3\{\log\log(n)\}^{1/2}\{1+n^{-1/2}\}\Big)\|\alpha\|_\lambda \\ &\quad + O_p\Big(\frac{1}{h^{1/2}} + \frac{1}{\sqrt{n}h^{(a+1)+\frac{2k-2a-1}{4m}}}\{\log(n)\}^2\{\log\log(n)\}^{1/2}h^{-1/2}(1+\{nh^2\}^{-1/2}) \\ &\quad + n^{-1/2}\log(n)h^{-(2a+1)/2} + n^{-1/2}h^{-(a+1)-\frac{2k-2a-1}{4m}}\{\log(n)\}^2\{\log\log(n)\}^{1/2} + n^{-1/2}h^{-1}\Big)\|\alpha\|_\lambda^2.\end{aligned}$$

It follows from $\alpha \in \mathbb{B}(r_{2n})$ and the conditions in the theorem that $\|D^2\mathcal{S}_{n,\lambda}(\alpha_\lambda + ss'\alpha)\alpha\alpha - D^2\mathcal{S}_\lambda(\alpha_\lambda+ss'\alpha)\alpha\alpha\|_\lambda = o_p(1)\|\alpha\|_\lambda$. Then we have $\|[D\mathcal{S}_\lambda(\alpha_\lambda)]I_2\|_\lambda \leq 11\|\alpha\|_\lambda/18$. Therefore, for any $\alpha \in \mathbb{B}(r_{2n})$, $\|T_{2h}(\alpha)\|_\lambda \leq \|I_1\|_\lambda + \|I_2\|_\lambda + \|I_3\|_\lambda \leq$

$11r_{2n}/12$. That is, $T_{2h}(\mathbb{B}(r_{2n})) \subset \mathbb{B}(r_{2n})$. Using the same argument above, we have that T_{2h} is a contraction mapping in $\mathbb{B}(r_{2n})$. Therefore, there exists a unique $\alpha' \in \mathbb{B}(r_{2n})$ such that $T_{2h}(\alpha') = \alpha'$ and this implies $\mathcal{S}_{n,\lambda}(\alpha_\lambda + \alpha') = 0$. Let $\hat{\alpha}_{n,\lambda} = \alpha_\lambda + \alpha'$. Then $\mathcal{S}_{n,\lambda}(\hat{\alpha}_{n,\lambda}) = 0$. Therefore, with probability approaching to 1, we have

$$\|\hat{\alpha}_{n,\lambda} - \alpha_0\|_\lambda \leq r_{1n} + r_{2n} = O_P((nh)^{-1/2} + h^k).$$

Proof of Theorem 2.2

It follows from Theorem 2.1 that there exists a constant $M > 0$ such that, with probability approaching to one, $\|\hat{\alpha}_{n,\lambda} - \alpha_0\|_\lambda \leq Mr_n$. For simplicity, denote $\hat{\alpha}_{n,\lambda} - \alpha_0$ as α. We assume that $\|\alpha\|_\lambda \leq Mr_n$ since its complement is negligible in terms of probability. Let $d_n = \kappa M h^{-(2a+1)/2} r_n$, $\tilde{\alpha} = d_n^{-1}\alpha$, and $p_n = \kappa^{-2}h^{1-2k}$, where κ is a constant given in Lemma 2.3. Since $h \to 0$ with $n \to \infty$ and $1 - 2k < 0$, we have that $p_n \geq 1$ when n is large enough. It can be shown that $\|\alpha\|_\lambda \leq Mr_n$ and this implies $\tilde{\alpha} \in \mathcal{F}_{p_n}$. To see this, write $\tilde{\alpha} = (\tilde{\theta}^\top, \tilde{\beta}(\cdot))$. Then $\|\tilde{\alpha}\|_e = d_n^{-1}\|\alpha\|_e \leq d_n^{-1}\kappa h^{-(2a+1)/2}\|\alpha\|_\lambda \leq d_n^{-1}\kappa h^{-(2a+1)/2} M r_n = 1$. Thus, we get

$$J(\tilde{\beta}, \tilde{\beta}) = d_n^2 \lambda^{-1}\{\lambda J(\beta, \beta)\} \leq d_n^2 \lambda \|\alpha\|_\lambda^2 \leq d_n^{-2}\lambda^{-1}(Mr_n)^2 = \kappa^{-2}h^{1-2k} = p_n.$$

Besides, we have

$$\begin{aligned} &\|\mathcal{S}_{n,\lambda}(\alpha + \alpha_0) - \mathcal{S}_{n,\lambda}(\alpha_0) - \{\mathcal{S}_\lambda(\alpha + \alpha_0) - \mathcal{S}_\lambda(\alpha_0)\}\|_\lambda \\ =&\|\mathcal{S}_n(\alpha + \alpha_0) - \mathcal{S}_n(\alpha_0) - \{\mathcal{S}(\alpha + \alpha_0) - \mathcal{S}(\alpha_0)\}\|_\lambda. \end{aligned} \tag{2.12}$$

On the right hand side of equation (2.12), we have

$$
\begin{aligned}
&\|\mathcal{S}_n(\alpha+\alpha_0)-\mathcal{S}_n(\alpha_0)-\{\mathcal{S}(\alpha+\alpha_0)-\mathcal{S}(\alpha_0)\}\|_\lambda \\
&=\Big\|\frac{1}{n}\sum_{i=1}^n \Delta_i\left[\mathcal{R}_{W_i}-\frac{n^{-1}\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha+\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{S_1^{(0)}(Y_i,\alpha+\alpha_0)}\right] \\
&\quad -\frac{1}{n}\sum_{i=1}^n \Delta_i\left[\mathcal{R}_{W_i}-\frac{n^{-1}\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{S_1^{(0)}(Y_i,\alpha_0)}\right]-\{\mathcal{S}(\alpha+\alpha_0)-\mathcal{S}(\alpha_0)\}\Big\|_\lambda \\
&\le \Big\|\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[\left\{\frac{\sum_j \mathcal{Y}_j(t)\exp\{\eta_{\alpha+\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{nS_1^{(0)}(t,\alpha_0+\alpha)}-\frac{\sum_j \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{nS_1^{(0)}(t,\alpha_0)}\right\}\right. \\
&\qquad\left.+\left\{\frac{E\mathcal{Y}(t)\exp\{\eta_{\alpha+\alpha_0}(W)\}\mathcal{R}_W}{s_1^{(0)}(t,\alpha_0+\alpha)}-\frac{E\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}\mathcal{R}_W}{s_1^{(0)}(t,\alpha_0)}\right\}\right]dN_i(t)\Big\|_\lambda+(nh)^{-1/2} \\
&\le \Big\|\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[\left\{\frac{\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha+\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{ns_1^{(0)}(t,\alpha_0+\alpha)}-\frac{\sum_j \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{ns_1^{(0)}(t,\alpha_0)}\right\}\right. \\
&\qquad\left.+\left\{\frac{E\mathcal{Y}(t)\exp\{\eta_{\alpha+\alpha_0}(W)\}\mathcal{R}_W}{s_1^{(0)}(t,\alpha_0+\alpha)}-\frac{E\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}\mathcal{R}_W}{s_1^{(0)}(t,\alpha_0)}\right\}\right]dN_i(t)\Big\|_\lambda \\
&\quad +O_p\left(\frac{1}{(nh)^{1/2}}+\frac{1}{h^{1/2+a}n^{1/2}}\right).
\end{aligned}
$$

Define $\Gamma=\cap_{i=1}^n A_{ni}$, where

$$
A_{ni}=\{\|Z_i\|_2\le c\log(n),\|X_i\|_{L^2}\le c\log(n),\exp\{\eta_{\alpha_\lambda}(W_i)\}\le c\log(n)\}.
$$

For any t_0, define $\varphi(Y_j;\alpha)=[\mathcal{Y}_j(t_0)\exp\{\eta_{\alpha+\alpha_0}(W_j)\}-\mathcal{Y}_j(t_0)\exp\{\eta_{\alpha_0}(W_j)\}]$, $D_n=\{c\log(n)\}^2 d_n^{-1}$, and $\varphi_n(Y_j;\tilde\alpha)=D_n\varphi(Y_j;d_n\tilde\alpha)\mathbf{1}_{A_{ni}}$. Then $|\varphi_n(Y_j;\tilde\alpha_1)-\varphi_n(Y_j;\tilde\alpha_2)|\le\|\tilde\alpha_1-\tilde\alpha_2\|_e$. Since $\|\alpha\|_\lambda\le Mr_n$, $\tilde\alpha\in\mathcal{F}_{p_n}$, it follows from Lemma 2.5 that with probability approaching to one, we have

$$
\begin{aligned}
&n^{-1/2}\|\sum_{j=1}^n \varphi_n(Y_j;\tilde\alpha)\mathcal{R}_{W_j}-E\varphi_n(Y_j;\tilde\alpha)\mathcal{R}_{W_j}\|_\lambda\lesssim(p_n^{1/(4m)}\|\tilde\alpha\|_e^\gamma+n^{-1/2})\{h^{-1}\log\log(n)\}^{1/2} \\
&\lesssim(p_n^{1/(4m)}+n^{-1/2})\{h^{-1}\log\log(n)\}^{1/2}, \qquad (2.13)
\end{aligned}
$$

where $\gamma = 1 - 1/(2m)$. On the other hand, by the Taylor expansion, the Cauchy-Schwarz inequality, Lemma 2.2 and Theorem 2.1, we have

$$\|E\{\varphi(Y_j; d_n\tilde{\alpha})\mathcal{R}_{W_j} I^c_{A_{n_j}}\}\|_\lambda \le \{E|\varphi(Y_j; d_n\tilde{\alpha})I^c_{A_{nj}}|^2\}^{1/2}\{E\|\mathcal{R}_{W_j}\|^2_\lambda\}^{1/2}$$

$$\lesssim [E\{\mathcal{Y}_j(t_0)\exp(\eta_{\alpha_0}(W_j)) < \mathcal{R}_{W_j}, \tilde{\alpha} >_\lambda\}^2]^{1/2} c_r^{1/2} h^{-1/2}$$

$$\lesssim [E\{\mathcal{Y}_j(t_0)\exp\{\eta_{\alpha_0}(W_j)\}\}^4]^{1/4} P(A^c_{n_j})^{1/4}[E\{(\|Z\|_2 + \|X_j\|_{L_2})\}^4]^{1/4} c_r h^{-1/2}.$$

From Condition (C4), we choose c large enough such that

$$n^{1/2}h^{-1/2}P(A^c_{n_i})^{1/4} = o(p_n^{1/(4m)}\{h^{-1}\log\log(n)\}^{1/2}).$$

Then

$$n^{1/2}\|E\{\varphi(Y_j; d_n\tilde{\alpha})\mathcal{R}_{W_j} I^c_{A_{n_j}}\}\|_\lambda \lesssim p_n^{1/(4m)}\{h^{-1}\log\log(n)\}^{1/2}.$$

Thus, on Γ_n, as n approaches ∞, we have

$$n^{-1/2}D_n\|\mathcal{S}_n(\alpha+\alpha_0) - \mathcal{S}_n(\alpha_0) - \{\mathcal{S}(\alpha+\alpha_0) - \mathcal{S}(\alpha_0)\}\|_\lambda \lesssim p_n^{1/(4m)}\{h^{-1}\log\log(n)\}^{1/2}. \tag{2.14}$$

On the left hand side of equation (2.12), we have

$$\|\mathcal{S}_{n,\lambda}(\alpha+\alpha_0) - \mathcal{S}_{n,\lambda}(\alpha_0) - \{\mathcal{S}_\lambda(\alpha+\alpha_0) - \mathcal{S}_\lambda(\alpha_0)\}\|_\lambda$$

$$= \| - \mathcal{S}_{n,\lambda}(\alpha_0) - D\mathcal{S}_\lambda(\alpha_0)\alpha - \int_0^1\int_0^1 sD^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha\, ds\, ds'\|_\lambda$$

$$= \Big\|\alpha - \mathcal{S}_{n,\lambda}(\alpha_0) - \int_0^1\int_0^1 sD^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha\, ds\, ds'\Big\|_\lambda$$

$$\ge \Big\|\alpha - \mathcal{S}_{n,\lambda}(\alpha_0)\|_\lambda - \|\int_0^1\int_0^1 sD^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha\, ds\, ds'\Big\|_\lambda. \tag{2.15}$$

It follows from Lemma 2.3 that

$$\Big\|\int_0^1\int_0^1 sD^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha\, ds\, ds'\Big\|_\lambda \le \int_0^1\int_0^1 s\|D^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha\|_\lambda\, ds\, ds'$$

$$\lesssim \|\alpha\|^2_\lambda c_r^{1/2}h^{-1/2} \lesssim h^{-1/2}r_n^2. \tag{2.16}$$

Therefore, it follows from (2.12),(2.14), (2.15), and (2.16) that

$$\|\alpha - \mathcal{S}_{n,\lambda}(\alpha_0)\|_\lambda \le O_p(a_n).$$

The proof of Theorem 2.2 is completed.

Proof of Theorem 2.3

Define $\hat{\alpha}^h_{n,\lambda} = (\hat{\theta}_{n,\lambda}, h^{a+1/2}\hat{\beta}_{n,\lambda})$, $\alpha^*_0 = (id - \mathcal{P}_\lambda)\alpha_0$, $\alpha^{*h}_0 = (\theta^*_0, h^{a+1/2}\beta^*_0)$, $\tilde{\mathcal{R}}^h_u = (\tilde{H}_u, h^{a+1/2}\tilde{T}_u)$, and

$$Rem_n = \hat{\alpha}_{n,\lambda} - \alpha^*_0 - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[\mathcal{R}_{W_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}}\right] dN_i(t).$$

It follows from Theorem 2.2 that $\|Rem_n\|_\lambda = O_p(a_n)$. Thus, we have

$$\|\hat{\theta}_{n,\lambda} - \theta^*_0 - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[\mathcal{H}_{W_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{H}_{W_j}}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}}\right] dN_i(t)\|_2 = O_p(a_n).$$

Define

$$Rem^h_n = \hat{\alpha}^h_{n,\lambda} - \alpha^{*h}_0 - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[\mathcal{R}^h_{W_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}^h_{W_j}}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}}\right] dN_i(t).$$

Then it is easy to show that $\|Rem^h_n - h^{a+1/2}Rem_n\|_\lambda = O_p(a_n)$. It follows from $a_n = o(n^{-1/2})$ that

$$\|Rem^h_n\|_\lambda \le \|Rem^h_n - h^{a+1/2}Rem_n\|_\lambda + h^{a+1/2}\|Rem_n\|_\lambda = o_p(n^{-1/2}).$$

Next, we will use Rem^h_n to obtain the target joint limiting distribution. The idea is to employ the Cramér-Wold device. For any $u = (z^\top, t) \in \mathbb{R}^p \times \mathbb{I}$, we obtain the limiting distribution of $n^{1/2}z^\top(\hat{\theta}_{n,\lambda} - \theta^*_0) + n^{1/2}h^{a+1/2}\{\hat{\beta}_{n,\lambda}(t) - \beta^*_0(t)\}$. Note that this is equivalent to getting the asymptotic result of $n^{1/2} < \tilde{\mathcal{R}}_u, \hat{\alpha}^h_{n,\lambda} - \alpha^{*h}_0 >_\lambda$. It follows from Theorem 2.2 that $n^{1/2}| < \tilde{\mathcal{R}}_u, Rem^h_n >_\lambda | = O_p(n^{1/2}h^{-(a+1/2)}a_n)$. Thus, we need to get the limiting distribution of

$$n^{1/2} < \tilde{\mathcal{R}}_u, \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[\mathcal{R}^h_{W_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}^h_{W_j}}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}}\right] dN_i(t) >_\lambda .$$

A direct calculation yields that

$$n^{1/2} < \tilde{\mathcal{R}}_u, \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[\mathcal{R}^h_{W_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}^h_{W_j}}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}}\right] dN_i(s) >_\lambda$$

$$=n^{-1/2}\sum_{i=1}^n \int_0^\tau \left[z^\top H_{W_i} + h^{a+1/2}T_{W_i}(t)\right.$$

$$\left. - \frac{\sum_{j=1}^n \mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}\{z^\top H_{W_j} + h^{a+1/2}T_{W_j}(t)\}}{\sum_{j=1}^n \mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}}\right] dM_i(s)$$

$$\equiv \mathcal{U}_n.$$

Define $\mathcal{K}_i(u) \equiv z^\top H_{W_i} + h^{a+1/2}T_{W_i}(t)$. Therefore, we have

$$\mathcal{U}_n = n^{-1/2}\sum_{i=1}^n \int_0^\tau \{\mathcal{K}_i(u) - \frac{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}[\mathcal{K}_j(u)]}{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}}\} dM_i(s) + O_p(n^{-1/2}h^{-a-1/2})$$

$$\equiv n^{-1/2}\sum_{i=1}^n \mathcal{U}_i + o_p(1).$$

A direct calculation yields that

$$\begin{aligned}
Var(\mathcal{U}_i) &= E\int_0^\tau \left[\mathcal{K}_i(u) - \frac{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{K}_j(u)}{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}}\right]^2 dN_i(s) \\
&= h^{2a+1}E\int_0^\tau \left[\pi_{X_i}(t) - \frac{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}\pi_{X_j}(t)}{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}}\right]^2 dN_i(s) \\
&\quad +2h^{a+1/2}(z - h^{a+1/2}\omega(t))^\top\{\Sigma - V(G,G^\top) + V(G,W_\lambda G^\top)\}^{-1} \\
&\qquad \times E\int_0^\tau \left[\pi_{X_i}(t) - \frac{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}\pi_{X_j}(t)}{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}}\right] \\
&\qquad \times \left[\{Z_i - V(G,\pi_{X_i})\} - \frac{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}\{Z_j - V(G,\pi_{X_j})\}}{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}}\right] dN_i(s) \\
&\quad +(z - h^{a+1/2}\omega(t))^\top\{\Sigma - V(G,G^\top) + V(G,W_\lambda G^\top)\}^{-1} \\
&\qquad \times E\int_0^\tau \left[\{Z_i - V(G,\pi_{X_i})\} - \frac{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}\{Z_j - V(G,\pi_{X_j})\}}{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}}\right]^{\otimes 2} dN_i(s) \\
&\qquad \times\{\Sigma - V(G,G^\top) + V(G,W_\lambda G^\top)\}^{-1}(z - h^{a+1/2}\omega(t)),
\end{aligned}$$

and

$$E\int_0^\tau \left[\{Z_i - V(G,\pi_{X_i})\} - \frac{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}\{Z_j - V(G,\pi_{X_j})\}}{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}}\right]$$

$$\times\left\{\pi_{X_i}(t) - \frac{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}\pi_{X_j}(t)}{E\mathcal{Y}_j(s)\exp\{\eta_{\alpha_0}(W_j)\}}\right\} dN_i(s)$$

$$= \sum_{j=1}^{\infty}\left\{\frac{G_j}{1+\lambda\rho_j}h_j(t) - \frac{G_j}{(1+\lambda\rho_j)^2}h_j(t)\right\}$$

$$= \sum_{j=1}^{\infty}\frac{\lambda\rho_j G_j}{(1+\lambda\rho_j)^2}h_j(t) = W_\lambda\omega(t).$$

It follows from similar arguments adopted in the proof of Theorem 2.1 in Shang and Cheng (2015) that $h^{a+1/2}\omega(t)\to 0$, $h^{a+1/2}W_\lambda\omega(t)\to 0$, $\sqrt{n}\{\theta_0^* - \theta_0\}\to 0$, and $\sqrt{n}h^{a+1/2}\{\beta_0^*(t) - \beta_0(t) + \{W_\lambda(\beta_0)\}(t)\}\to 0$. Then, as $\lambda\to 0$, we have

$$\begin{aligned} Var(\mathcal{U}_i) \to & \sigma_t^2 + 2(z+\gamma_0)^\top\{\Sigma - V(G,G^\top)\}^{-1}\xi_0 + (z+\gamma_0)^\top\{\Sigma - V(G,G^\top)\}^{-1}(z+\gamma_0) \\ \equiv & (z^\top,1)\Phi(z^\top,1)^\top, \end{aligned}$$

It follows from the Lindeberg's central limit theorem that

$$\begin{bmatrix} \sqrt{n}(\hat{\theta}_{n,\lambda} - \theta_0) \\ \sqrt{nh}h^a\{\hat{\beta}_{n,\lambda}(t) - \beta_0(t) + (W_\lambda\beta_0)(t_0)\}\end{bmatrix} \to N(0,\Phi).$$

Since $n^{1/2}h^{k(1+b)} = o(1)$, we can get that $nh^{4k} = o(1)$. Then, we have

$$\begin{aligned} |(W_\lambda\beta_0)(t_0)| &= \Big|\sum_{j=1}^{n}\frac{b_j\lambda\rho_j}{1+\lambda\rho_j}h_j(t_0)\Big| \\ &\le c_h\lambda\left\{\sum_{j=1}^{\infty}b_j^2\rho_j^2\right\}^{1/2}\left\{\sum_{j=1}\frac{j^{2a}}{(1+\lambda\rho_j)^2}\right\}^{1/2} \\ &= O(\lambda h^{-a-1/2}) \\ &= o(1). \end{aligned}$$

Hence, it leads to $\sqrt{n}h^{a+1/2}\{W_\lambda(\beta_0)\}(t) = o(1)$. Thus, the conclusion follows directly.

Proof of Theorem 2.4

Define $\alpha = \hat{\alpha}_{n,\lambda} - \alpha_0$. It follows from Theorem 2.1 that for some $M > 0$, we have $\|\alpha\|_\lambda \leq Mr_n$ with probability approaching to one. Therefore, we assume $\|\alpha\|_\lambda \leq Mr_n$. Applying the Taylor expansion, we have

$$l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\hat{\alpha}_{n,\lambda}) = -\mathcal{S}_{n,\lambda}(\hat{\alpha}_{n,\lambda})\alpha + \int_0^1\int_0^1 sD\mathcal{S}_{n,\lambda}(\hat{\alpha}_{n,\lambda} - ss'\alpha)\alpha\alpha\,ds\,ds'$$

$$= \int_0^1\int_0^1 s\{D\mathcal{S}_{n,\lambda}(\hat{\alpha}_{n,\lambda} - ss'\alpha) - D\mathcal{S}_{n,\lambda}(\alpha_0)\}\alpha\alpha\,ds\,ds' + \frac{1}{2}D\mathcal{S}_{n,\lambda}(\alpha_0)\alpha\alpha. \tag{2.17}$$

It follows from Lemma 2.1 that

$$\begin{aligned} l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\hat{\alpha}_{n,\lambda}) = &\int_0^1\int_0^1 s[D\mathcal{S}_{n,\lambda}(\hat{\alpha}_{n,\lambda} - ss'\alpha) - D\mathcal{S}_{n,\lambda}(\alpha_0)]\alpha\alpha\,ds\,ds' \\ &+ \frac{1}{2}[D\mathcal{S}_{n,\lambda}(\alpha_0) - D\mathcal{S}_{\lambda}(\alpha_0)]\alpha\alpha - \frac{1}{2}\|\alpha\|_\lambda \\ \equiv & I_1 + I_2 - \frac{1}{2}\|\alpha\|_\lambda. \end{aligned}$$

To get the order of I_1, we define $\alpha' = \hat{\alpha}_{n,\lambda} - ss'\alpha - \alpha_0 = (1-ss')\alpha$, where $0 \leq s, s' \leq 1$. A direct calculation yields that

$$
\begin{aligned}
&|[D\mathcal{S}_{n,\lambda}(\hat{\alpha}_{n,\lambda} - ss'\alpha) - D\mathcal{S}_{n,\lambda}(\alpha_0)]\alpha\alpha| = |D^2\mathcal{S}_{n,\lambda}(\alpha_0 + \delta\alpha')\alpha\alpha\alpha'| \\
&\asymp \Big|\frac{1}{n}\sum_{i=1}^{n}\Delta_i\Big[\frac{\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\eta_\alpha(W_j)\eta_{\alpha'}(W_j)}{nS_1^{(0)}(Y_i,\alpha_0)} \\
&\qquad - \frac{\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\eta_\alpha(W_j)\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_0}(W_j)\}\eta_{\alpha'}(W_j)}{[nS_1^{(0)}(Y_i,\alpha_0)]^2} \\
&\qquad - 2\frac{\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\eta_{\alpha'}(W_j)\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)}{[nS_1^{(0)}(Y_i,\alpha_0)]^2} \\
&\qquad + 2\frac{\{\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\}^2\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_0}(W_j)\}\eta_{\alpha'}(W_j)}{[nS_1^{(0)}(Y_i,\alpha_0)]^3}\Big]\Big| \\
&\lesssim \sup_{t\in\mathbb{I}}\Big|\frac{\sum_j \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\eta_\alpha(W_j)\eta_{\alpha'}(W_j)}{ns_1^{(0)}(t,\alpha_0)}\Big| \\
&+ \sup_{t\in\mathbb{I}}\Big|\frac{\sum_j \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\eta_\alpha(W_j)\sum_j \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_0}(W_j)\}\eta_{\alpha'}(W_j)}{\{ns_1^{(0)}(t,\alpha_0)\}^2}\Big| \\
&+ 2\sup_{t\in\mathbb{I}}\Big|\frac{\sum_j \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\eta_{\alpha'}(W_j)\sum_j \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)}{\{ns_1^{(0)}(t,\alpha_0)\}^2}\Big| \\
&+ 2\sup_{t\in\mathbb{I}}\Big|\frac{\{\sum_j \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\}^2\sum_j \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\eta_{\alpha'}(W_j)}{[ns_1^{(0)}(t,\alpha_0)]^3}\Big| \\
&\equiv I_{11} + I_{12} + I_{13} + I_{14},
\end{aligned}
$$

where $0 \leq \delta \leq 1$. Define $\Gamma = \cap_{i=1}^n A_{ni}$, where

$$A_{ni} = \{\|Z_i\|_2 \leq c\log(n), \|X_i\|_{L^2} \leq c\log(n), \exp\{\eta_{\alpha_0}(W_i)\} \leq c\log(n)\}$$

for a constant c. For I_{11}, we have

$$I_{11} = \sup_{t\in\mathbb{I}} \Big| \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\eta_\alpha(W_j)\eta_{\alpha'}(W_j)}{s_1^{(0)}(t,\alpha_0)} \Big|$$

$$\leq \Big| \frac{n^{-1}\sum_{j=1}^n \exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\eta_\alpha(W_j)\eta_{\alpha'}(W_j)}{s_1^{(0)}(\tau,\alpha_0)} \Big|$$

$$\leq \frac{Mr_n}{ns_1^{(0)}(\tau,\alpha_0)} \Big| \sum_{j=1}^n \exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\eta_\alpha(W_j)\|\mathcal{R}_{W_j}\|_\lambda \Big|$$

$$\leq \frac{Mr_n}{ns_1^{(0)}(\tau,\alpha_0)} \Big| < \sum_{j=1}^n \exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\|\mathcal{R}_{W_j}\|_\lambda \mathcal{R}_{W_j}, \alpha >_\lambda \Big|.$$

Let $d_n = \kappa M h^{-(2a+1)} r_n$, and $\tilde{\alpha} = d_n^{-1}\alpha$, where κ is given in Lemma 2.3. Note that $\tilde{\alpha} \in \mathcal{F}_{p_n}$, where $p_n = \kappa^{-2} h^{2a+1-2k} > 1$ when n is large enough. Denote

$$\phi_n(Y_i, \Delta_i, W_i; \tilde{\alpha}) = \frac{\exp\{\eta_{\alpha_0}(W_i)\}\eta_{\tilde{\alpha}}(W_i)\|\mathcal{R}_{W_i}\|_\lambda \mathcal{R}_{W_i}}{\sqrt{2c_r}\{c\log(n)\}^3 h^{-(a+1/2)}} I_{A_{ni}}.$$

Then it can be shown that

$$|\phi_n(Y_i, \Delta_i, W_i; \tilde{\alpha}_1) - \phi_n(Y_i, \Delta_i, W_i; \tilde{\alpha}_2)| \leq \frac{\sqrt{2c_r}\{c\log(n)\}^3 h^{-a-1/2}}{\sqrt{2c_r}\{c\log(n)\}^3 h^{-(a+1/2)}} \|\tilde{\alpha}_1 - \tilde{\alpha}_2\|_e.$$

It follows from Lemma 2.5 that with probability approaching to one,

$$\Big\| n^{-1/2} \sum_{j=1}^n \big[\exp\{\eta_{\alpha_0}(W_j)\}\eta_{\tilde{\alpha}}(W_j)\|\mathcal{R}_{W_j}\|_\lambda \mathcal{R}_{W_j} I_{A_{nj}} - E\exp\{\eta_{\alpha_0}(W_j)\}\eta_{\tilde{\alpha}}(W_j)\|\mathcal{R}_{W_j}\|_\lambda \mathcal{R}_{W_j} I_{A_{nj}}\big] \Big\|_\lambda$$

$$\lesssim p_n^{1/(4m)}\{h^{-1}\log\log(n)\}^{1/2}\{c\log(n)\}^3 h^{-(a+1/2)}.$$

Therefore, we have

$$\Big\| \sum_{j=1}^n [\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\|\mathcal{R}_{W_j}\|_\lambda \mathcal{R}_{W_j} I_{A_{nj}} - E\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\|\mathcal{R}_{W_j}\|_\lambda \mathcal{R}_{W_j} I_{A_{nj}}] \Big\|_\lambda$$

$$\lesssim n^{1/2} p_n^{1/(4m)}\{h^{-1}\log\log(n)\}^{1/2}\{c\log(n)\}^3 h^{-(a+1/2)} d_n$$

$$= c^3 M r_n n^{1/2} \kappa^{1-1/(2m)} h^{-(2a+3/2)+(2a+1-2k)/(4m)}\{\log(n)\}^3\{\log\log(n)\}^{1/2}.$$

It follows from the Cauchy-Schwarz inequality, Lemma 2.2 and Lemma 2.4 that

$$
\begin{aligned}
&|E\exp\{\eta_{\alpha_0}(W_j)\}\eta_\alpha(W_j)\|\mathcal{R}_{W_j}\|_\lambda < \mathcal{R}_{W_j}, \alpha >_\lambda| \\
\le&\big(E[\exp\{\eta_{\alpha_0}(W_j)\}]^2\big)^{1/2}\{E\|\mathcal{R}_W\|_\lambda^2\}\{E < \mathcal{R}_{W_j}, \alpha >_\lambda^4\}^{1/2} \\
\lesssim&\big(E[\exp\{\eta_{\alpha_0}(W_j)\}]^2\big)^{1/2}\sqrt{c_r}h^{-1/2}\sqrt{\|\alpha\|_\lambda^4} \\
\lesssim&\big(E[\exp\{\eta_{\alpha_0}(W_j)\}]^2\big)^{1/2}h^{-1/2}M^2r_n^2.
\end{aligned}
$$

Thus, with probability going to one,

$$|I_{11}| \lesssim (r_n^3 n^{-1/2}h^{-(2a+3/2)+(2a+1-2k)/(4m)}\{\log(n)\}^3\{\log\log(n)\}^{1/2} + h^{-1/2}r_n^3).$$

Similarly, we can prove that

$I_{12} = O_p(r_n^3 n^{-1/2}h^{-(2a+3/2)+(2a+1-2k)/(4m)}\{\log(n)\}^3\{\log\log(n)\}^{1/2} + h^{-1/2}r_n^3)$,

$I_{13} = O_p(r_n^3 n^{-1/2}h^{-(2a+3/2)+(2a+1-2k)/(4m)}\{\log(n)\}^3\{\log\log(n)\}^{1/2} + h^{-1/2}r_n^3)$,

$I_{14} = O_p(r_n^3 n^{-1/2}h^{-(2a+3/2)+(2a+1-2k)/(4m)}\{\log(n)\}^3\{\log\log(n)\}^{1/2} + h^{-1/2}r_n^3)$. Therefore, we have

$I_1 = O_p(r_n^3 n^{-1/2}h^{-(2a+3/2)+(2a+1-2k)/(4m)}\{\log(n)\}^3\{\log\log(n)\}^{1/2} + h^{-1/2}r_n^3) = o_p(n^{-1}h^{-1/2})$.

It follows from equation (2.11) that

$$
\begin{aligned}
2|I_2| &= |[D\mathcal{S}_{n,\lambda}(\alpha_0) - D\mathcal{S}_\lambda(\alpha_0)]\alpha\alpha| \\
&= O_p(n^{-1/2}h^{-(a+1)-\frac{2k-2a-1}{4m}}\{\log(n)\}^2\{\log\log(n)\}^{1/2}r_n^2) = o_p(n^{-1}h^{-1/2}).
\end{aligned}
$$

Therefore, we have

$$-2n\mathrm{PLRT}_{n,\lambda} = n\|\hat{\alpha}_{n,\lambda} - \alpha_0\|_\lambda^2 + o_p(h^{-1/2}).$$

It follows from Theorem 2.2 and $n^{1/2}a_n = o(1)$ that

$$-2n\mathrm{PLRT}_{n,\lambda} = n\|\mathcal{S}_{n,\lambda}(\alpha_0)\|_\lambda^2(1 + o_p(1)) + o_p(h^{-1/2}).$$

A direct calculation yields that

$$
\begin{aligned}
& n\|\mathcal{S}_{n,\lambda}(\alpha_0)\|_\lambda^2 \\
=& \frac{1}{n}\Big\|\sum_{i=1}^n \int_0^\tau \Big[\mathcal{R}_{W_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{S_1^{(0)}(t,\alpha_0)}\Big]\, dM_i(t) - \mathcal{P}_\lambda\alpha_0\Big\|_\lambda^2 \\
=& \frac{1}{n}\Big\|\sum_{i=1}^n \int_0^\tau \Big[\mathcal{R}_{W_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{S_1^{(0)}(t,\alpha_0)}\Big]\, dM_i(t)\Big\|_\lambda^2 \\
& - 2 < \frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big[\mathcal{R}_{W_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{S_1^{(0)}(t,\alpha_0)}\Big]\, dM_i(t), \mathcal{P}_\lambda\alpha_0 >_\lambda + n\|\mathcal{P}_\lambda\alpha_0\|_\lambda^2 \\
\equiv& J_1 + J_2 + J_3.
\end{aligned}
$$

For J_1 and J_2, it follows from Condition (C1) that

$$
\begin{aligned}
J_1 =& \frac{1}{n}\Big\|\sum_{i=1}^n \int_0^\tau \Big[\mathcal{R}_{W_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{S_1^{(0)}(t,\alpha_0)}\Big]\, dM_i(t)\Big\|_\lambda^2 \\
=& \frac{1}{n}\Big\|\sum_{i=1}^n \int_0^\tau \Big[\mathcal{R}_{W_i} - \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_0)}\Big]\, dM_i(t)\Big\|_\lambda^2 + O_p(n^{-1}h^{-1-2a}) \\
=& \frac{1}{n}\Big\|\sum_{i=1}^n \int_0^\tau \Big[\mathcal{R}_{W_i} - \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_0)}\Big]\, dM_i(t)\Big\|_\lambda^2 + o_p(1),
\end{aligned}
$$

and

$$
\begin{aligned}
& \Big|\frac{J_2}{2}\Big| = \Big| < \frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big[\mathcal{R}_{W_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{S_1^{(0)}(t,\alpha_0)}\Big]\, dM_i(t), \mathcal{P}_\lambda\alpha_0 >_\lambda \Big| \\
\le& \Big| < \frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big[\mathcal{R}_{W_i} - \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_0)}\Big]\, dM_i(t), \mathcal{P}_\lambda\alpha_0 >_\lambda \Big| \\
+& \Big| < \frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big[\frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_0)} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{S_1^{(0)}(t,\alpha_0)}\Big]\, dM_i(t), \mathcal{P}_\lambda\alpha_0 >_\lambda \Big| \\
=& \Big| < \frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big[\mathcal{R}_{W_i} - \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_0)}\Big]\, dM_i(t), \mathcal{P}_\lambda\alpha_0 >_\lambda \Big| + O_p(n^{-1/2}h^{-1/2-a})\|\mathcal{P}_\lambda\alpha_0\|_\lambda.
\end{aligned}
$$

Denote $\beta_0 = \sum_j b_j h_j$. Since $J(\beta_0, \beta_0) = \sum_j b_j^2 \rho_j < \infty$, and $\lambda = o(1)$, it follows from the Lebesgue dominated convergence theorem that

$$E\left[\sum_{i=1}^n \int_0^\tau \left\{\mathcal{R}_{W_i} - \frac{E\mathcal{Y}_j(t)\exp\left(\eta_{\alpha_0}(W_j)\right)\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_0)}\right\} dM_i(t), \mathcal{P}_\lambda \alpha_0 >_\lambda^2\right]$$

$$= nE\left[\int_0^\tau \left\{\int_{\mathbb{I}} \{X(t) - E(X(t)|T = v, \Delta = 1)\} W_\lambda(\beta_0)(t)\, dt\right\}^2 \mathcal{Y}(v)\exp\{\eta_{\alpha_0}(W)\} h_0(v)\, dv\right]$$

$$= nV(W_\lambda(\beta_0), W_\lambda(\beta_0)) \le n\|W_\lambda(\beta_0)\|_m = n\lambda \sum_j b_j^2 \rho_j \frac{\lambda\rho_j}{(1+\lambda\rho_j)} = o_p(n\lambda).$$

Therefore, we have $J_2 = o_p((n\lambda)^{1/2})(1 + (nh)^{-1/2}) = o_p((n\lambda)^{1/2})$. Note that $J_3 = n\|\mathcal{P}_\lambda \alpha_0\|_\lambda^2 = n\|W_\lambda(\beta_0)\|_m^2 = o(n\lambda)$. Hence, we have

$$-2n\mathrm{PLRT}_{n,\lambda}$$

$$= \frac{1}{n}\Big\| \sum_{i=1}^n \int_0^\tau \left[\mathcal{R}_{W_i} - \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_0)}\right] dM_i(t)\Big\|_\lambda^2 + n\|W_\lambda(\beta_0)\|_m^2 + o_p(h^{-1/2}).$$

Denote $R_i(t) = \mathcal{R}_{W_i} - \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}\mathcal{R}_{W_j}}{s_1^{(0)}(t,\alpha_0)}$. To obtain the asymptotic result of $-2n\mathrm{PLRT}_{n,\lambda}$, we need to investigate the properties of

$$\frac{1}{n}\Big\|\sum_{i=1}^n \int_0^\tau R_i(t)\, dM_i(t)\Big\|_\lambda^2 = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \int_0^\tau < R_i(t), R_j(s) >_\lambda\ dM_i(t)\, dM_i(s) + \frac{1}{n}\sum_{1\le i<j\le n} W_{ij},$$

where $W_{ij} = 2\int_0^\tau \int_0^\tau < R_i(t), R_j(s) >_\lambda\ dM_i(t)\, dM_j(s)$. Write $W_n = \sum_{1\le i<j\le n} W_{ij}$. So, W_n is clean Jong (1987). Next, we aim to derive the limiting distribution of W_n. Let $\sigma_n^2 = Var(W_n)$. Then

$$\begin{aligned}
\sigma_n^2 &= \frac{n(n-1)}{2} E(W_{ij}^2) = 2n(n-1)E\left\{\int_0^\tau \int_0^\tau < R_i(t), R_j(s) >_\lambda\ dM_i(t)\, dM_j(s)\right\}^2 \\
&\asymp 2n(n-1)\{\sum_{l=1}^\infty \frac{1}{(1+\lambda\rho_l)^2} + 1\} \asymp 2n^2 h^{-1}\sigma_\lambda^4/\rho_\lambda^2.
\end{aligned}$$

Define M_1, M_2 and M_3 as follows:

$$M_1 \equiv \sum_{i<j} E(W_{ij}^4), \quad M_2 \equiv \sum_{i<j<k} \left\{E(W_{ij}^2 W_{ik}^2) + E(W_{ji}^2 W_{jk}^2) + E(W_{ki}^2 W_{kj}^2)\right\}, \quad \text{and}$$

$$M_3 \equiv \sum_{i<j<k<l} \left\{E(W_{ij}W_{ik}W_{lj}W_{lk}) + E(W_{ij}W_{il}W_{kj}W_{kl}) + E(W_{ik}W_{il}W_{jk}W_{jl})\right\}.$$

By Proposition 3.2 of Jong (1987), if M_1, M_2, M_3 are all of order lower than σ_n^4, then $\sigma_n^{-1} W_n$ converges weakly to the standard normal distribution. Now, we study the order of each $M_i, i = 1, 2, 3$. First, observe that

$$E\left(W_{ij}^4\right) = 16E\left\{\int_0^\tau \int_0^\tau < R_i(t), R_j(s) >_\lambda \; dM_i(t)\, dM_j(s)\right\}^4$$

$$= 16 \int_0^\tau \int_0^\tau \int_0^\tau \int_0^\tau \int_0^\tau \int_0^\tau \int_0^\tau \int_0^\tau E < R_i(t_1), R_j(s_1) >_\lambda < R_i(t_2), R_j(s_2) >_\lambda < R_i(t_3), R_j(s_3) >_\lambda$$

$$< R_i(t_4), R_j(s_4) >_\lambda \left\{ dM_i(t_1)\, dM_j(s_1)\, dM_i(t_2)\, dM_j(s_2)\, dM_i(t_3)\, dM_j(s_3)\, dM_i(t_4)\, dM_j(s_4)\right\}$$

$$= O(h^{-4}),$$

which implies $M_1 = O(n^2 h^{-4})$. Next, by the Cauchy-Schwarz inequity,

$$E(W_{ij}^2 W_{ik}^2) \le \{E(W_{ij}^4)\}^{1/2}\{E(W_{ik}^4)\}^{1/2} = O(h^{-4}),$$

which yields $M_2 = O(n^3 h^{-4})$. A straightforward calculation yields that

$$E(W_{ij}W_{ik}W_{lj}W_{lk}) \sim 16 \sum_{j=0}^{\infty} \frac{1}{(1+\lambda\rho_j)^4} = O(h^{-1}).$$

Therefore, $M_3 = O(n^4 h^{-1})$. Combining the fact that $\sigma_n^4 = (\sigma_n^2)^2 = O(n^4 h^{-2})$ with the assumptions that $nh^2 \to \infty$ and $h = o(1)$, we have that M_1, M_2, M_3 are of order lower than that of σ_n^4. Hence, by Jong (1987), $\sigma_n^{-1} W_n \xrightarrow{d} N(0,1)$ as $n \to \infty$. Recall that $\rho_\lambda^2 = \sum_{j=0}^{\infty} h/(1+\lambda\rho_j)^2$. We have

$$\frac{1}{\sqrt{2h^{-1}}n\rho_\lambda} W_n \xrightarrow{d} N(0,1). \tag{2.18}$$

Finally, we consider $n^{-1}\sum_{i=1}^{n}\int_0^\tau\int_0^\tau < R_i(t), R_j(s) >_\lambda \ dM_i(t)\,dM_i(s)$. Through a direct calculation, we obtain

$$E\left\{\int_0^\tau\int_0^\tau < R_i(t), R_j(s) >_\lambda \ dM_i(t)\,dM_i(s)\right\}^2 = O(\{E\|\mathcal{R}_{W_i}\|_\lambda^2\}^2) = O(h^{-2}).$$

Then,

$$\begin{aligned} & E\left\{\sum_{i=1}^{n}\int_0^\tau\int_0^\tau < R_i(t), R_j(s) >_\lambda \ dM_i(t)\,dM_i(s) - h^{-1}\sigma_\lambda^2 - 1\right\}^2 \\ \le \ & nE\left\{\int_0^\tau\int_0^\tau < R_i(t), R_j(s) >_\lambda \ dM_i(t)\,dM_i(s)\right\}^2 = O(nh^{-2}), \end{aligned}$$

where $\sigma_\lambda^2 = \sum_{j=0}^{\infty} h/1 + \lambda\rho_j$. Combining these results, we have

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\int_0^\tau < R_i(t), R_j(s) >_\lambda \ dM_i(t)\,dM_i(s) = 1 + h^{-1}\sigma_\lambda^2 + O_p\{(n^{1/2}h)^{-1}\}. \quad (2.19)$$

By (2.18) and (2.19), we have $n\|\mathcal{S}_{n,\lambda}(\alpha_0)\|_\lambda^2 = O_p(h^{-1})$ and, therefore, $n^{1/2}\|\mathcal{S}_{n,\lambda}(\alpha_0)\|_\lambda = O_p(h^{-1/2})$. As a result,

$$-2n\mathrm{PLRT}_{n,\lambda} = \{n^{1/2}\|\mathcal{S}_{n,\lambda}(\alpha_0)\|_\lambda + o_p(1)\}^2 + o_p(h^{-1/2}) \quad (2.20)$$

$$= n\|\mathcal{S}_{n,\lambda}(\alpha_0)\|_\lambda^2 + o_p(h^{-1/2}). \quad (2.21)$$

In view of (2.18), (2.19) and (2.20), we conclude that as $n \to \infty$,

$$(2h^{-1}\sigma_\lambda^4/\rho_\lambda^2)^{-1/2}\left\{-2n\gamma_\lambda PLRT_{n,\lambda} - n\gamma_\lambda\|W_\lambda\beta_0(t)\|_\lambda^2 - h^{-1}\sigma_\lambda^4/\rho_\lambda^2\right\} \xrightarrow{d} N(0,1).$$

The proof of Theorem 2.4 is completed.

Proof of Theorem 2.5

Throughout this proof, we only consider $\alpha_{n0} = \alpha_0 + \alpha_n$ for $\alpha_n \in \mathcal{A}$ in H_1. To prove the theorem, we write

$$-2n \cdot \mathrm{PLRT}_{n,\lambda} = -2n\{l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\alpha_{n0})\} - 2n\{l_{n,\lambda}(\alpha_{n0}) - l_{n,\lambda}(\hat{\alpha}_{n,\lambda})\} \equiv I_1 + I_2. \quad (2.22)$$

We first consider I_1. For simplicity, we denote

$$R_i = \Delta_i \left[\eta_{\alpha_0}(W_i) - \log \frac{1}{n}\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_0}(W_j)\}\right]$$

$$- \Delta_i \left[\eta_{\alpha_{n_0}}(W_i) - \log \frac{1}{n}\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{\eta_{\alpha_{n_0}}(W_j)\}\right]$$

$$= -\int_0^\tau \left[\eta_{\alpha_n}(W_i) - \frac{\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0+s'\alpha_n}(W_j)\}\eta_{\alpha_n}(W_j)}{\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0+s'\alpha_n}(W_j)\}}\right] dN_i(t)$$

$$= -\int_0^\tau \left[\eta_{\alpha_n}(W_i) - \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0+s'\alpha_n}(W_j)\}\eta_{\alpha_n}(W_j)}{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0+s'\alpha_n}(W_j)\}}\right] dN_i(t)$$

$$+ \int_0^\tau \left[\frac{\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0+s'\alpha_n}(W_j)\}\eta_{\alpha_n}(W_i)}{\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0+s'\alpha_n}(W_j)\}} - \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0+s'\alpha_n}(W_j)\}\eta_{\alpha_n}(W_j)}{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0+s'\alpha_n}(W_j)\}}\right] dN_i(t)$$

$$= -\int_0^\tau \left[\eta_{\alpha_n}(W_i) - \frac{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0+s'\alpha_n}(W_j)\}\eta_{\alpha_n}(W_j)}{E\mathcal{Y}_j(t)\exp\{\eta_{\alpha_0+s'\alpha_n}(W_j)\}}\right] dN_i(t) + o_p(\|\alpha_n\|_\lambda),$$

where $0 \le s' \le 1$. Then

$$E\{R_i^2\} \asymp E\int_0^\tau Var\{\eta_{\alpha_n}(W)|T=t,\Delta=1\}\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}h_0(t)\,dt = O(\|\alpha_n\|_\lambda^2).$$

Therefore, we get

$$E\left\{\left|\sum_{i=1}^n (R_i - ER_i)\right|^2\right\} \le nE\{R_i^2\} = O(n\|\alpha_n\|_\lambda^2).$$

Combining these gives

$$n\left[l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\alpha_{n_0}) - E\{l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\alpha_{n_0})\}\right] = O_p(n^{1/2}\|\alpha_n\|_\lambda).$$

On the other hand, since $D\mathcal{S}_\lambda(\alpha)\alpha_n\alpha_n < 0$ for any $\alpha \in \mathcal{H}$, there exists a constant $c > 0$ such that $\{D\mathcal{S}_\lambda(\alpha^*_{n_0})\alpha_n\alpha_n\} \le c\{D\mathcal{S}_\lambda(\alpha_{n_0})\alpha_n\alpha_n\} = -c\|\alpha_n\|^2_\lambda$. Then, we have

$$E\{l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\alpha_{n_0})\} = E\left\{\mathcal{S}_{n,\lambda}(\alpha_{n_0})(-\alpha_n) + \frac{1}{2}D\mathcal{S}_{n,\lambda}(\alpha^*_{n_0})\alpha_n\alpha_n\right\}$$

$$\le \lambda J(\alpha_{n_0}, \alpha_n) - \frac{c\|\alpha_n\|^2_\lambda}{2} \le \{J(\alpha_n, \alpha_n) + J(\alpha_0, \alpha_n)\} - \frac{c\|\alpha_n\|^2_\lambda}{2}$$

$$\le \{J(\alpha_n, \alpha_n) + J(\alpha_0, \alpha_0)^{1/2} J(\alpha_n, \alpha_n)^{1/2}\} - \frac{c\|\alpha_n\|^2_\lambda}{2}$$

$$= O(\lambda) - \frac{c\|\alpha_n\|^2_\lambda}{2}.$$

It then follows that

$$I_1 \ge n\|\alpha_n\|^2_\lambda + O_p(n\lambda + n^{1/2}\|\alpha_n\|_\lambda +) = n\|\alpha_n\|^2_\lambda\{1 + O_p(\lambda\|\alpha_n\|^{-2}_\lambda + n^{-1/2}\|\alpha_n\|^{-1}_\lambda)\}. \tag{2.23}$$

Next, we consider I_2. Under the alternative hypothesis, $\|\hat{\alpha}_{n,\lambda} - \alpha_{n_0}\| = O_p\{(nh)^{-1/2} + h^k\}$. It then follows by the joint functional Bahadur representation in Theorem 2.2 that

$$\inf_{n \ge N} \inf_{\alpha_n \in \mathcal{A}} P_{\alpha_{n_0}}\left(\|\hat{\alpha}_{n,\lambda} - \alpha_{n_0} - S_{n,\lambda}(\alpha_{n_0})\|_\lambda \le M r_n\right) \to 1, \tag{2.24}$$

where $r_n = (nh)^{-1/2} + h^k$, and $P_{\alpha_{n_0}}$ is the probability which depends on α_{n_0}. Note that, under the alternative hypothesis H_{1n}, I_2 is the same as (2.17) except one constant term $-2n$. Along the lines of Theorem 2.4, we can show that I_2 has the same limiting distribution as that in Theorem 2.4, uniformly for any $\alpha_n \in \mathcal{A}$. In other words, uniformly over all $\alpha_n \in \mathcal{A}$, we have

$$(2\nu_{n_0})^{-1/2}(I_2 - n\|W_\lambda \beta_{n_0}\|^2_m - h^{-1}\sigma^2_{n_0,\lambda}) = O_p(1), \tag{2.25}$$

where $\nu_{n_0} = h^{-1}\sigma^4_{n_0,\lambda}/\rho^2_{n_0,\lambda}$, $\sigma^2_{n_0,\lambda}$ and $\rho^2_{n_0,\lambda}$ are same as σ^2_λ and ρ^2_λ but with eigenvalues and eigenvectors obtained under α_{n_0}. Next, let $V_{n_0}(f, g) = \int_{\mathbb{I}}\int_{\mathbb{I}} F_{\alpha_{n_0}}(s, t) f(t) g(s)\, dt\, ds$ and $V_0(f, g) = \int_{\mathbb{I}}\int_{\mathbb{I}} F_{\alpha_0}(s, t) f(t) g(s)\, dt\, ds$, where $F_{\alpha_0}(s, t) = F(s, t)$, while $F_{\alpha_{n_0}}(s, t)$

has the same formula as $F_{\alpha_0}(s,t)$ with α_0 replaced by α_{n_0}. Thus, for any $f \in \mathcal{H}^{(m)}$, there exists a constant c such that

$$
\begin{aligned}
|V_{n_0}(f,f) - V_0(f,f)| &= \left| \int_{\mathbb{I}} \int_{\mathbb{I}} \left[F_{\alpha_{n_0}}(s,t) - F_{\alpha_0}(s,t)\right] f(t) f(s)\, dt\, ds \right| \\
&\le E\| \exp\{\alpha_n(W)\}\|_\infty V_0(f,f)\|\alpha_n\|_\infty = cV_0(f,f)\|\alpha_n\|_\infty.
\end{aligned}
$$

It follows from the Supplementary Material (page 56) of Shang and Cheng (2015) that

$$
\sigma^2_{n_0,\lambda} - \sigma^2_\lambda = O(h^{-(a+1)/2}\|\alpha_n\|_\lambda). \tag{2.26}
$$

Combining (2.23), (2.25) and (2.26) gives

$$
\begin{aligned}
&(2\nu_n)^{-1/2}(-2nr_\lambda \mathrm{PLRT}_{n,\lambda} - \nu_n) = (2\nu_n)^{-1/2}\{-r_\lambda(I_1 + I_2) - \nu_n\} \\
=&(2\nu_n)^{-1/2} r_\lambda (I_2 - n\|\mathcal{P}_\lambda \alpha n_0\|^2_\lambda - h^{-1}\sigma^2_{n_0,\lambda}) + (2\nu_n)^{-1/2} r_\lambda n \|\mathcal{P}_\lambda \alpha n_0\|^2_\lambda \\
&+ (2\nu_n)^{-1/2} r_\lambda I_1 + (2\nu_n)^{-1/2} r_\lambda h^{-1}(\sigma^2_{n_0,\lambda} - \sigma^2_\lambda) \\
\ge& O_p(1) + (2\nu_n)^{-1/2} r_\lambda n \|\alpha_n\|^2_\lambda \{1 + O_p(\lambda\|\alpha_n\|^{-2}_\lambda + n^{-1/2}\|\alpha_n\|^{-1}_\lambda)\} \\
&+ O(h^{-3/2-a/2}\|\alpha_n\|_\lambda),
\end{aligned}
$$

where $O_p(1)$ holds uniformly in $\mathcal{A}$, $\nu_n = h^{-1}\sigma^4_\lambda/\rho^2_\lambda$, and r_λ is defined in Theorem 2.4. Let $\lambda\|\alpha_n\|^{-2}_\lambda \le 1/c$, $n^{-1/2}\|\alpha_n\|^{-1}_\lambda \le 1/c$, $ch^{-3/2-a/2}\|\alpha_n\|_\lambda \le n\|\alpha_n\|^2_\lambda$, and $\|\alpha_n\|^2_\lambda \ge c(nh^{1/2})^{-1}$ for some sufficiently small constant c. In other words,

$$
|(2\nu_n)^{-1/2}(-2nr_\lambda \mathrm{PLRT}_{n,\lambda} - \nu_n)| \ge c_\alpha,
$$

where c_α is the critical value (based on $N(0,1)$) to H_0^{global} at nominal level α. This leads to

$$
\|\alpha_n\|^2_\lambda \gtrsim \{h^{2k} + (nh^{1/2})^{-1}\}. \tag{2.27}
$$

Combining (2.24) and (2.27), we complete the proof of Theorem 2.5.

Chapter 3

Semiparametric Statistical Inference for Functional Additive hazard Model

3.1 Introduction

When hazard differences are of focus, the additive hazards model is often preferred over the Cox proportional hazards models. O'Neill (1986) discovers that if the additive hazards model is appropriate but the proportional hazards model is assumed, there is a significant bias.

The additive hazards model attracts many research interests. The large-sample theory is developed in Lin and Ying (1994) by utilizing the martingale approach. Andersen and Gill (1982) originally develops this approach for the Cox model. Lin et al. (1998) and Martinussen and Scheike (2002) extend the additive hazards model to handle interval censored data. Kulich and Lin (2000) considers measurement error problems in the additive hazards model. Huffer and Mckeague (1991) investigates the weighted least squares estimation for a nonparametric additive risk model. A partly parametric additive hazards model incorporating time-dependent and constant regression coefficients is developed by Mckeague and Sasieni (1994).

Recently, Chen et al. (2011) proposes the functional Cox model by incorporating functional predictors and scalar predictors. The goal is to study the survival of diffuse large-B-cell lymphoma (DLBCL) patients. The study considers the gene expression as a functional parameter to handle the high-dimensional problem. It is of interest to consider the additive hazards model which is compatible with functional data in order to avoid potential serious bias resulting from the use of the Cox proportional hazards model.

Introducing functional data to additive hazards costs more difficult theoretical investigation. One major contribution of this work is the derivation of the Bahadur representation of the estimators in the additive hazards model. When an estimator can be *almost* expressed as a sum of identical and independent variables, the estimator admits a *Bahadur representation.* Bahadur (1966) first establishes an asymptotic almost sure representation of a sample quartile for independent and identically distributed random variables. The major incentive of deriving Bahadur representation is that the asymptotic normality can be immediately established with the central limit theorem under some appropriate regularity conditions. In this chapter, we develop a new technical tool, called a joint Bahadur representation (JBR), for studying the joint asymptotic results. As far as we know, our joint asymptotic theories and inference procedures are new. The only relevant references of which we are aware are Shang and Cheng (2015) and Cheng and Shang (2015), which focus on generalized functional linear models and semi-nonparametric regression models with partially linear structure, respectively.

3.2 Estimation Method

The functional additive hazards model with a p-dimensional covariate vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_p)^\top$ and a functional covariate $X(\cdot)$ defines the hazard function by

$$h(t|\mathbf{W}) = h_0(t) + \{\theta_0^\top \mathbf{Z} + \int_{\mathbb{I}} X(s)\beta_0(s)\,ds\}, \tag{3.1}$$

where $\mathbf{W} = (\mathbf{Z}^\top, X(\cdot))^\top$. Without loss of generality, we assume that $\mathbb{I} = [0, 1]$. We focus on right censored data. Let T be the survival time, C be the censoring time, $Y = \min(T, C)$ be the observation time, and $\Delta = \mathbf{I}(T \leq C)$ be the censoring indicator, where $\mathbf{I}$ is the indicator function. Let $\{(\mathbf{W}_i^\top, Y_i, \Delta_i)^\top, i = 1, 2, \cdots, n\}$ be the observations.

Our goal is to estimate $\alpha_0 = (\theta_0^\top, \beta_0(\cdot))^\top$ in order to ascertain the relationship between $\mathbf{W}$ and the survival time T. We assume that $\beta_0(\cdot)$ belongs to the mth-order Sobolev space $\mathcal{H}^{(m)}(\mathbb{I})$ which is abbreviated as $\mathcal{H}^{(m)}$:

$$\mathcal{H}^{(m)} = \{\beta : \mathbb{I} \mapsto \mathbb{R} | \beta^{(j)} \text{ is absolutely continuous for } j = 0, 1, \ldots, m-1, \\ \beta^{(m)} \in L_2(\mathbb{I})\},$$

where m is a known constant which is $> 1/2$, $\beta^{(j)}$ is the jth derivative of β, and $L_2(\mathbb{I})$ is the L_2 space defined in $\mathbb{I}$. Then, we have $\alpha_0 \in \mathbb{R}^p \times \mathcal{H}^m$ and denote $\mathbb{R}^p \times \mathcal{H}^m$ as $\mathcal{H}$. Define $\eta_\alpha(\mathbf{W}) = \theta^\top \mathbf{Z} + \int_{\mathbb{I}} X(s)\beta(s)\, ds$, $\mathcal{Y}(t) = \mathbf{I}(Y \geq t)$, and $J(\beta_1, \beta_2) = \int_{\mathbb{I}} \beta_1^{(m)}(t)\beta_2^{(m)}(t)\, dt$. Let $N(t)$ be the counting process $\Delta\mathbf{I}(Y \leq t)$, and $M(t)$ be the martingale process $N(t) - \int_0^t \mathcal{Y}(s)h(s|\mathbf{W})\, ds$. For any $\alpha_l = (\theta_l^\top, \beta_l(\cdot))^\top \in \mathcal{H}, l = 1, 2$, define the semi-inner product as:

$$\begin{aligned} &< \alpha_1, \alpha_2 >_\lambda \\ &= \int_0^\tau [E\{\mathcal{Y}(t)\eta_{\alpha_1}(\mathbf{W})\eta_{\alpha_2}(\mathbf{W})\} - E\{\mathcal{Y}(t)\}\tilde{\eta}_{\alpha_1}(t)\tilde{\eta}_{\alpha_2}(t)]\, dt + \lambda J(\beta_1, \beta_2), \end{aligned} \tag{3.2}$$

where τ is end of the study time, and $\tilde{\eta}_{\alpha_1}(t)$ is the asymptotic value of $\bar{\eta}_{\alpha_1}(t)$ which is defined as

$$\bar{\eta}_{\alpha_1}(t) = \frac{\sum_{j=1}^n \mathcal{Y}_j(t)\eta_{\alpha_1}(\mathbf{W}_j)}{\sum_{j=1}^n \mathcal{Y}_j(t)}.$$

For any vector $\mathbf{z}$, define $\mathbf{z}^{\otimes 0} = 1$, $\mathbf{z}^{\otimes 1} = \mathbf{z}$, $\mathbf{z}^{\otimes 2} = \mathbf{z}\mathbf{z}^\top$. Furthermore, we impose the following assumptions to construct a Hilbert space and to establish the asymptotic results.

Assumption A1.

(a) The survival time T and the censoring time C are conditionally independent given $\mathbf{W}$.

(b) $P(Y \geq \tau) > 0$.

(c) There exists a constant $c_1 > 1$ satisfying that:

$$\int_0^\tau [E\{\mathcal{Y}(t)\eta_\alpha(\mathbf{W})^2\} - E\{\mathcal{Y}(t)\}\tilde{\eta}_\alpha(t)^2]\, dt \geq c_1 E[\{\eta_\alpha(W)\}^2],$$

for any $\alpha \in \mathcal{H}$.

Assumption A1(a) is very common in right censored data to guarantee the non-informative censoring mechanism. Assumption A1(b) is used to make $\tilde{\eta}_\alpha(t)$ meaningful. Assumption A1(c) is easy to verify under the scenario that $\beta(s) = 0$ and the following Assumption (A3) holds. Define

$$S_1^{(k)}(t) = \frac{1}{n}\sum_{i=1}^{n}\{\mathcal{Y}_i(t)\mathbf{Z}_i^{\otimes k}\}, s_1^{(k)}(t) = E\{\mathcal{Y}(t)\mathbf{Z}^{\otimes k}\}, k = 0, 1, 2,$$

$$S_2^{(2)}(t, s, v) = \frac{1}{n}\sum_{i=1}^{n}\{\mathcal{Y}_i(t)X_i(s)X_i(v)\}, s_2^{(2)}(t, s, v) = E\{\mathcal{Y}(t)X(s)X(v)\},$$

$$S_3^{(1)}(t, s) = \frac{1}{n}\sum_{i=1}^{n}\{\mathcal{Y}_i(t)X_i(s)\}, s_3^{(1)}(t, s) = E\{\mathcal{Y}(t)X(s)\},$$

$$F(s, t) = \int_0^\tau \mathrm{Cov}\{X(s), X(t)|T = v, \Delta = 1\}E\{\mathcal{Y}(v)\}\, dv,$$

where

$$\begin{aligned}
&\mathrm{Cov}\{X(s), X(t)|T = v, \Delta = 1\}\\
=&E\{X(s)X(t)|T = v, \Delta = 1\} - E\{X(s)|T = v, \Delta = 1\}E\{X(t)|T = v, \Delta = 1\}\\
=&\frac{s_2^{(2)}(v, t, s)}{s_1^{(0)}(v)} - \frac{s_2^{(1)}(v, s)s_2^{(1)}(v, t)}{s_1^{(0)}(v)^2}.
\end{aligned}$$

Then, $\mathcal{H}$ is a Hilbert space and $\mathcal{H}^{(m)}$ is a Reproducing Kernel Hilbert Space (RKHS) with the inner product

$$< \beta_1, \beta_2 >_m = \int_{\mathbb{I}} \int_{\mathbb{I}} F(s,t)\beta_1(s)\beta_2(t)\, ds\, dt + \lambda J(\beta_1, \beta_2). \tag{3.3}$$

Denote the reproducing kernel in $\mathcal{H}^{(m)}$ as $K(s,t)$, and $\|\cdot\|_m$ as the norm induced by the inner product $< \cdot, \cdot >_m$. Define a bilinear operator $V(\cdot, \cdot)$ in $\mathcal{H}^{(m)}$ as:

$$V(\beta_1, \beta_2) = \int_{\mathbb{I}} \int_{\mathbb{I}} F(s,t)\beta_1(s)\beta_2(t)\, ds\, dt \tag{3.4}$$

and a linear nonnegative definite and self-adjoint operator W_λ as:

$$< W_\lambda \beta_1, \beta_2 >_m = \lambda J(\beta_1, \beta_2). \tag{3.5}$$

Define a linear operator $\mathcal{S}_n$ as:

$$\mathcal{S}_n(\alpha)\alpha_1 = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \{\eta_{\alpha_1}(\mathbf{W}_i) - \bar{\eta}_{\alpha_1}(t)\}\{\, dN_i(t) - \mathcal{Y}_i(t)\eta_\alpha(\mathbf{W}_i)\, dt\}.$$

for any $(\alpha, \alpha_1) \in \mathcal{H} \times \mathcal{H} \to \mathbb{R}$. It follows from Lin and Ying (1994) that $\mathcal{S}_n(\alpha)$ is the pseudo-score function for α. Define a least square-type loss function $l_n(\alpha)$ as:

$$\begin{aligned} & l_n(\alpha) \\ = & -\frac{1}{2n} \sum_{i=1}^{n} \int_0^\tau \{\eta_\alpha(\mathbf{W}_i) - \bar{\eta}_\alpha(t)\}^2 Y_i(t)\, dt + \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \{\eta_\alpha(\mathbf{W}_i) - \bar{\eta}_\alpha(t)\}\, dN_i(t). \end{aligned}$$

Then the first Fréchet derivative of $l_n(\alpha)$ with respect to α at α_1 is $\mathcal{S}_n(\alpha)\alpha_1$. Thus, to maximize $l_n(\alpha)$ is equivalent to solve $\mathcal{S}_n(\alpha) = 0$. To obtain a smoothed estimate for β, we introduce a penalty term to the loss function. The objective function is then defined as:

$$l_{n\lambda}(\alpha) \equiv l_n(\alpha) - \frac{\lambda}{2} J(\beta, \beta),$$

where $J(\beta, \beta)$ is the penalty function and λ is the smoothing parameter which controls the balance between the smoothness of β and the fit to data. The estimate for α is defined as:

$$\hat{\alpha}_{n\lambda} = \arg\max l_{n\lambda}(\alpha).$$

Let D be the Fréchet derivative operator. Then the first Fréchet derivative of $l_{n\lambda}(\alpha)$ with respect to α at any direction $\alpha_1 = (\theta_1^\top, \beta_1(\cdot))^\top \in \mathcal{H}$ is:

$$\begin{aligned} &\mathcal{S}_{n\lambda}(\alpha)\alpha_1 \\ =&\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{\eta_{\alpha_1}(\mathbf{W}_i) - \bar{\eta}_{\alpha_1}(t)\}\{\, dN_i(t) - \mathcal{Y}_i(t)\eta_\alpha(\mathbf{W}_i)\, dt\} - \lambda J(\beta, \beta_1). \end{aligned} \tag{3.6}$$

The first and the second Fréchet derivatives of $\mathcal{S}_{n\lambda}(\alpha)$ are:

$$D\mathcal{S}_{n\lambda}(\alpha)\alpha_1\alpha_2 = -\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathcal{Y}_i(t)\{\eta_{\alpha_1}(\mathbf{W}_i) - \bar{\eta}_{\alpha_1}(t)\}\eta_{\alpha_2}(\mathbf{W}_i)\, dt - \lambda J(\beta_1, \beta_2),$$

and

$$D^2\mathcal{S}_{n\lambda}(\alpha)\alpha_1\alpha_2\alpha_3 = 0,$$

respectively. Denote the asymptotic value of $\mathcal{S}_{n\lambda}(\alpha)\alpha_1$ and that of $D\mathcal{S}_{n\lambda}(\alpha)\alpha_1\alpha_2$ as $\mathcal{S}_\lambda(\alpha)\alpha_1$ and $D\mathcal{S}_\lambda(\alpha)\alpha_1\alpha_2$, respectively.

Then we have

$$\mathcal{S}_\lambda(\alpha)\alpha_1 = E\int_0^\tau \{\eta_{\alpha_1}(\mathbf{W}) - \tilde{\eta}_{\alpha_1}(t)\}\{\, dN(t) - \mathcal{Y}(t)\eta_\alpha(\mathbf{W})\, dt\} - \lambda J(\beta, \beta_1),$$

$$\begin{aligned} &D\mathcal{S}_\lambda(\alpha)\alpha_1\alpha_2 \\ =&-E\int_0^\tau [\mathcal{Y}(t)\{\eta_{\alpha_1}(\mathbf{W}) - \tilde{\eta}_{\alpha_1}(t)\}\eta_{\alpha_2}(\mathbf{W})]\, dt - \lambda J(\beta_1, \beta_2) \\ =&-\int_0^\tau [E\{\eta_{\alpha_1}(\mathbf{W})\eta_{\alpha_2}(\mathbf{W})\mathcal{Y}(t)\} - E\{\mathcal{Y}(t)\eta_{\alpha_2}(\mathbf{W})\tilde{\eta}_{\alpha_1}(t)\}]\, dt - \lambda J(\beta_1, \beta_2) \\ =&-\int_0^\tau \left[\frac{E\{\eta_{\alpha_1}(\mathbf{W})\eta_{\alpha_2}(\mathbf{W})\mathcal{Y}(t)\}}{E\{\mathcal{Y}(t)\}} - \tilde{\eta}_{\alpha_2}(t)\tilde{\eta}_{\alpha_1}(t)\right] E\{\mathcal{Y}(t)\}\, dt - \lambda J(\beta_1, \beta_2). \end{aligned}$$

Proposition 3.1. *Under Assumption A1 and the definition of the inner product, we have that for any $\alpha \in \mathcal{H}$, $D\mathcal{S}_\lambda(\alpha) = -id$, where id is the identity operator.*

This result directly follows from the definition of the inner product and the definition of $D\mathcal{S}_\lambda(\alpha)$.

We denote two positive sequences a_n and b_n as $a_n \asymp b_n$ if $\lim_{n\to\infty}(a_n/b_n) = c > 0$. If $c = 1$, we have $a_n \sim b_n$.

Assumption A2.

There exists a sequence of functions $\{h_j\}_{j\geq 1} \subset \mathcal{H}^{(m)}$ such that $\|h_j\|_{L_2} \leq c_h j^a$ for each $j \geq 1$, some constants $a \geq 0$, $c_h \geq 0$ and

$$V(h_i, h_j) = \delta_{ij}, J(h_i, h_j) = \rho_i \delta_{ij}, \quad \text{for any} \quad i, j \geq 1, \tag{3.7}$$

where δ_{ij} is the Kronecker's notation, and ρ_i is a nondecreasing nonnegative sequence satisfying $\rho_i \asymp i^{2k}$ for some constant $k > a + 1/2$. Furthermore, for any $\beta \in \mathcal{H}^{(m)}$, β admits the Fourier expansion $\beta = \sum_{i=1}^{\infty} V(\beta, h_i) h_i$.

Following the ideas from Shang and Cheng (2015), we derive the eigen-system with the following integro-differential equations:

$$\begin{aligned} (-1)^m y_j^{(2m)}(t) &= \rho_j \int_0^1 F(s,t) y_j(s)\, ds, \\ y_j^{(i)}(0) = y_j^{(i)}(1) &= 0, \qquad\qquad i = m, m+1, \cdots, 2m-1. \end{aligned} \tag{3.8}$$

Let $h_j = y_j / \sqrt{V(y_j, y_j)}$, with $k = m + r + 1$ and $a = r + 1$. We have that h_j and ρ_j are the eigenvector and eigenvalue, respectively, if one of the following additional assumptions is satisfied:

1. $r = 0$;

2. $r \geq 1$, and for any $i = 0, 1, \ldots, r-1$, $F^{(i,0)}(0,t) = 0$ for any $t \in \mathbb{I}$, where $F^{(i,0)}(s,t)$ is the ith-order partial derivative with respect to s.

The relationships between (h_j, ρ_j) and $K(\cdot,\cdot)$ or W_λ are given as follows:

$$K_t(\cdot) = \sum_{j=1}^{\infty} \frac{h_j(t)}{1+\lambda\rho_j} h_j(\cdot), (W_\lambda h_j)(\cdot) = \frac{\lambda\rho_j}{1+\lambda\rho_j} h_j(\cdot).$$

This can be referred to Shang and Cheng (2015).

On the basis of the Riesz representation theorem, there exists an element $\pi_x \in \mathcal{H}^{(m)}$ such that $< \pi_x, \beta >_m = \int_0^1 x(t)\beta(t)\, dt$. Through a direct calculation, we have $\pi_x = \sum_{j=1}^{\infty} \int_{\mathbb{I}} x(t) h_j(t)\, dt/(1+\lambda\rho_j) h_j$. In addition, there exist $\omega_k, G_k \in \mathcal{H}^{(m)}$ such that the following relationship holds:

$$\begin{aligned} V(G_k, \beta) &= < \omega_k, \beta >_m \\ &= \int_{\mathbb{I}} \int_0^\tau \left[\frac{E\{\mathcal{Y}(t) Z_k X(s)\}}{s_1^{(0)}(t)} - \frac{s_{1,k}^{(1)}(t)}{s_1^{(0)}(t)} \frac{s_3^{(1)}(t,s)}{s_1^{(0)}(t)} \right] s_1^{(0)}(t)\, dt \beta(s)\, ds, \end{aligned}$$

where $s_{1,k}^{(1)}$ is the kth element of $s_1^{(1)}$. Denote $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_p)^\top$, $\mathbf{G} = (G_1, G_2, \ldots, G_p)^\top$. A direct calculation yields

$$\begin{aligned} \mathbf{G} &= \sum_{j=1}^{\infty} \int_{\mathbb{I}} \int_0^\tau \left[\frac{E\{\mathcal{Y}(t) \mathbf{Z} X(s)\}}{s_1^{(0)}(t)} - \frac{s_1^{(1)}(t)}{s_1^{(0)}(t)} \frac{s_3^{(1)}(t,s)}{s_1^{(0)}(t)} \right] s_1^{(0)}(t)\, dt h_j(s)\, ds h_j \\ &= \sum_{j=1}^{\infty} < \int_0^\tau Cov(\mathbf{Z}\pi_X | T=t, \Delta=1) s_1^{(0)}(t)\, dt, h_j >_m h_j \\ &\equiv \sum_{j=1}^{\infty} V(\mathbf{G}, h_j) h_j. \end{aligned}$$

$$\begin{aligned} \boldsymbol{\omega} &= \sum_{j=1}^{\infty} \int_{\mathbb{I}} \int_0^\tau \left[\frac{E\{\mathcal{Y}(t) \mathbf{Z} X(s)\}}{s_1^{(0)}(t)} - \frac{s_1^{(1)}(t)}{s_1^{(0)}(t)} \frac{s_3^{(1)}(t,s)}{s_1^{(0)}(t)} \right] s_1^{(0)}(t)\, dt h_j(s)\, ds \frac{h_j}{1+\lambda\rho_j} \\ &= \sum_{j=1}^{\infty} < \int_0^\tau Cov(\mathbf{Z}\pi_X | T=t, \Delta=1) s_1^{(0)}(t)\, dt, h_j >_m \frac{h_j}{1+\lambda\rho_j} \\ &\equiv \sum_{j=1}^{\infty} V(\mathbf{G}, h_j) \frac{h_j}{1+\lambda\rho_j}. \end{aligned}$$

Thus, we have $\boldsymbol{\omega} = (id - W_\lambda)\mathbf{G}$. It follows from the definition of $\mathbf{G}$ and $\boldsymbol{\omega}$ that if $X(\cdot)$ and $\mathbf{Z}$ are independent, we have $\mathbf{G} = \boldsymbol{\omega} = \mathbf{0}$. Define $\Sigma = \int_0^\tau \mathrm{Var}[\mathbf{Z}|T = t, \Delta = 1] s_1^{(0)}(t)\, dt$, and $\Omega = V(\mathbf{G}, \mathbf{G}^\top)$. Let $h = \lambda^{1/(2k)}$.

Assumption A3.

$\Sigma - \Omega$ is positive definite. There exists $b \in ((1 + 2a)/(2k), 1]$ such that $V(\mathbf{G}, h_j)$ satisfies that:

$$\sum_{j=1}^{\infty} \|V(\mathbf{G}, h_j)\|_2^2 \rho_j^b < \infty.$$

It follows from Assumption A3 that $V(\mathbf{G}, W_\lambda \mathbf{G}^\top) \to \mathbf{0}$ with $\lambda \to 0$.

Proposition 3.2. *For any $\mathbf{w} = (\mathbf{z}^\top, x(\cdot))^\top$, define $\mathcal{R}_\mathbf{w} : \mathbf{w} \to (H_\mathbf{w}, T_\mathbf{w}) \in \mathcal{H}$, where*

$$H_\mathbf{w} = \{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1}\{\mathbf{z} - V(\mathbf{G}, \pi_x)\} \quad \textit{and}$$

$$T_\mathbf{w} = \pi_x - \boldsymbol{\omega}^\top \{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1}\{\mathbf{z} - V(\mathbf{G}, \pi_x)\}.$$

Then, we have $< \mathcal{R}_\mathbf{w}, \alpha >_\lambda = \theta^\top \mathbf{z} + \int_0^1 x(t)\beta(t)\, dt$.

It follows from the fact $V(\mathbf{G}, W_\lambda \mathbf{G}^\top) \to \mathbf{0}$ that the definition of $\{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1}$ is meaningful. In fact, $\{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\} = \{\Sigma - < \boldsymbol{\omega}, \boldsymbol{\omega}^\top >_m \}$ is positive definite with $\lambda \to 0$.

Denote $\mathbf{u} = (\mathbf{z}^\top, t)^\top$.

Proposition 3.3. *For any $\mathbf{u} \in \mathbb{R}^p \times \mathbb{I}$, define $\tilde{\mathcal{R}}_\mathbf{u} : \mathbf{u} \to (\tilde{H}_\mathbf{u}, \tilde{T}_\mathbf{u}) \in \mathcal{H}$, where*

$$\tilde{H}_\mathbf{u} = \{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1}\{\mathbf{z} - \boldsymbol{\omega}(t)\} \quad \textit{and}$$

$$\tilde{T}_\mathbf{u} = K_t - \boldsymbol{\omega}^\top \{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1}\{\mathbf{z} - \boldsymbol{\omega}(t)\}.$$

Then, we have $< \tilde{\mathcal{R}}_\mathbf{u}, \alpha >_\lambda = \theta^\top \mathbf{z} + \beta(t)$.

Proposition 3.4. *For any $\alpha \in \mathcal{H}$, define $\mathcal{P}_\alpha : \alpha \to (H_\alpha^*, T_\alpha^*) \in \mathcal{H}$, where*

$$H_\alpha^* = -\{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1} V(\mathbf{G}, W_\lambda \beta) \quad \textit{and}$$

$$T_\alpha^* = W_\lambda \beta + \boldsymbol{\omega}^\top \{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1} V(\mathbf{G}, W_\lambda \beta).$$

Then, $\mathcal{P}_\lambda \alpha \in \mathcal{H}$ and $< \mathcal{P}_\lambda \alpha, \alpha_1 >_\lambda = < W_\lambda \beta, \beta_1 >_m$ for any $\alpha_1 = (\theta_1^\top, \beta_1)^\top \in \mathcal{H}$.

It follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
\|\mathcal{P}_\lambda \alpha\|_\lambda &= \sup_{\|\alpha_1\|_\lambda=1} < \mathcal{P}_\lambda \alpha, \alpha_1 >_\lambda \\
&= \sup_{\|\alpha_1\|_\lambda=1} \lambda |J(\beta, \beta_1)| \\
&\leq \sup_{\|\alpha_1\|_\lambda=1} \sqrt{\lambda J(\beta_1, \beta_1)} \sqrt{\lambda J(\beta, \beta)} \leq \|\alpha\|_\lambda.
\end{aligned}$$

Thus, we have that $\|\mathcal{P}_\lambda\|_\lambda \leq 1$ and $\mathcal{P}_\lambda$ is self-adjoint.

Lemma 3.1. *Suppose Assumptions A1-A3 are satisfied. Then, for any $x \in L_2([0,1])$, we have*

$$\begin{aligned}
&< \mathcal{R}_{\mathbf{w}}, \mathcal{R}_{\mathbf{w}} >_\lambda \\
=&\mathbf{z}^\top \{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1} \mathbf{z} + < \pi_x, \pi_x >_m - 2\mathbf{z}^\top \{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1} \\
&\times V(\mathbf{G}, \pi_x) + V(\mathbf{G}, \pi_x)^\top \{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1} V(\mathbf{G}, \pi_x).
\end{aligned}$$

Furthermore, there exists a universal positive constant c_r such that

$$< \mathcal{R}_{\mathbf{w}}, \mathcal{R}_{\mathbf{w}} >_\lambda \leq c_r(\|\mathbf{z}\|_2^2 + \|x\|_{L_2}^2 h^{-2a-1}), E\{\|\mathcal{R}_{\mathbf{W}}\|_\lambda^2\} \leq c_r h^{-1}.$$

3.3 Joint Bahadur Representation

Let $\|\alpha\|_e = \|\theta\|_2 + \|\beta\|_{L_2}$. The following lemma states the relationship between the general Euclidean norm $\|\cdot\|_e$ and $\|\cdot\|_\lambda$.

Lemma 3.2. *The exists a constant $\kappa > 0$ such that for any $\alpha \in \mathcal{H}$, $\|\alpha\|_e \leq \kappa h^{-(2a+1)/2} \|\alpha\|_\lambda$.*

Assumption A4.

There exists a constant $s \in (0,1)$ such that

$$E[\exp\{s(\|X\|_{L_2} + \|\mathbf{Z}\|_2)\}] < \infty. \tag{3.9}$$

Moreover, there exists a constant $M_0 > 0$ such that for any $\alpha \in \mathcal{H}$,

$$E\left\{\left|\eta_\alpha(\mathbf{W})\right|^4\right\} \leq M_0\left[E\left|\eta_\alpha(\mathbf{W})\right|^2\right]^2.$$

Assumption A4 allows more relax conditions on the covariates than that in Qu et al. (2016).

Lemma 3.3. *Under Assumptions A1-A4, we have that for any* $\alpha \in \mathcal{H}$, $E\{| < \mathcal{R}_{\mathbf{W}}, \alpha >^4 |\} \leq c_2 \|\alpha\|_\lambda^4$, *where* c_2 *is a positive constant that is independent of* α.

Theorem 3.1. *(Rate of Convergence) Under Assumptions A1-A4,*

$$h = o(1), \quad \textit{and} \quad n^{-1/2} h^{-(a+1)-\frac{2k-2a-1}{4m}} \{\log(n)\}\{\log\log(n)\}^{1/2} = o(1),$$

we have that $\hat{\alpha}_{n\lambda}$ *is the unique estimate for* α_0 *and* $\|\hat{\alpha}_{n\lambda} - \alpha_0\|_\lambda = O_p(r_n)$, *with* $r_n = (nh)^{-1/2} + h^k$.

This theorem shows that when we choose $\lambda = n^{-(2k)/(2k+1)}$, the estimate enjoys the same order of convergence as that in Qu et al. (2016).

Theorem 3.2. *(Joint Functional Bahadur Representation) Suppose that Assumptions A1-A4 hold. If* $n \to \infty$, $n^{-1/2} h^{-(a+1)-\frac{2k-2a-1}{4m}} \{\log(n)\}\{\log\log(n)\}^{1/2} = o(1)$, $h = o(1)$, *and* $nh^2 \to \infty$, *we have that*

$$\|\hat{\alpha}_{n\lambda} - \alpha_0 - \mathcal{S}_{n\lambda}(\alpha_0)\|_\lambda = O_p(a_n)$$

with

$$a_n = n^{-1/2} h^{-(4ma+6m-1)/4m} r_n \{\log\log(n)\}^{1/2} \log(n)$$

and $r_n = (nh)^{-1/2} + h^k$.

On the basis of the joint functional Bahadur Representation, we derive the asymptotic properties of the estimators of the functional coefficient and the scalar coefficient.

Theorem 3.3. *(Joint Asymptotic Normality) Suppose that Assumptions A1-A4 hold. Furthermore, suppose*

$$h^{2a+1}E\left(\int_0^\tau \left[\pi_X(t) - \frac{E\{\mathcal{Y}(s)\pi_X(t)\}}{E\{\mathcal{Y}(s)\}}\right]^2 dN(s)\right) \to \sigma_t^2 > 0,$$

$$h^{a+1/2}E\left\{\int_0^\tau \left(\{\mathbf{Z} - V(\mathbf{G}, \pi_X)\} - \frac{E[\mathcal{Y}(s)\{\mathbf{Z} - V(\mathbf{G}, \pi_X)\}]}{E\{\mathcal{Y}(s)\}}\right)\right.$$
$$\left.\times \left(\pi_X(t) - \frac{E\{\mathcal{Y}(s)\pi_X(t)\}}{E\{\mathcal{Y}(s)\}}\right) dN(s)\right\} \to \xi_0, \quad \textit{and}$$

$$E\left\{\int_0^\tau \left(\{\mathbf{Z} - V(\mathbf{G}, \pi_X)\} - \frac{E[\mathcal{Y}(s)\{\mathbf{Z}_j - V(\mathbf{G}, \pi_X)\}]}{E\{\mathcal{Y}(s)\}}\right)^{\otimes 2} dN(s)\right\} \to B_0,$$

where B_0 is positive definite. In addition, if $n \to \infty$,$nh^2 \to \infty$, $n^{1/2}h^{k(1+b)} = o(1)$, $\sum_{j=1}^\infty V(\beta_0, h_j)^2\rho_j^2 < \infty$, $n^{1/2}a_n h^{-(a+1/2)} = o(1)$, $h = o(1)$, with

$$a_n = n^{-1/2}h^{-(4ma+6m-1)/4m}r_n\{\log\log(n)\}^{1/2}\log(n),$$

and $r_n = (nh)^{-1/2} + h^k$, we have that, for any fixed t,

$$\begin{bmatrix} \sqrt{n}(\hat{\theta}_{n\lambda} - \theta_0) \\ \sqrt{nh}h^a\{\hat{\beta}_{n\lambda}(t) - \beta_0(t)\} \end{bmatrix} \to N(0, \Phi),$$

with

$$\Phi = \begin{bmatrix} (\Sigma - \Omega)^{-1}B_0(\Sigma - \Omega)^{-1} & (\Sigma - \Omega)^{-1}\xi_0 \\ \xi_0^\top(\Sigma - \Omega)^{-1} & \sigma_t^2 \end{bmatrix}.$$

Next, we can derive the uniform convergence result about $\hat{\beta}_{n\lambda}(s)$ in $\mathbb{I}$.

Theorem 3.4. *Assume that the conditions in Theorem 3 are satisfied, we have that*

$$\sqrt{nh}\{\hat{\beta}_{n\lambda}(s) - \beta_0(s)\}$$

converges to a mean zero Gaussian process $\mathcal{G}(s)$ in the Hilbert space $\mathcal{H}^{(m)}(\mathbb{I})$ with the inner product $V(\cdot,\cdot)$. The covariance for $\mathcal{G}(s)$ at s_1, s_2 is

$$\Gamma(s_1, s_2) = h^{2a+1} E\left(\int_0^\tau \left[\pi_X(s_2) - \frac{E\{\mathcal{Y}(s)\pi_X(s_2)\}}{E\{\mathcal{Y}(s)\}}\right] \left[\pi_X(s_1) - \frac{E\{\mathcal{Y}(s)\pi_X(s_1)\}}{E\{\mathcal{Y}(s)\}}\right] dN(s) \right).$$

In practice, to construct the simultaneous confidence band in a closed subset $[\zeta, 1-\zeta]$, we employ the resampling method of Lin et al. (1993) for distributional approximation. For illustration, let $(\epsilon_1, \ldots, \epsilon_n)$ be independent standard normal random variables, independent of the data $(Y_i, \Delta_i, \mathbf{Z}_i^\top, X_i(\cdot)), i = 1, \ldots, n$. It can be shown that the distribution of the limiting process $\mathcal{G}(s)$ can be approximated by the distribution of the following zero-mean Gaussian process

$$\hat{\mathcal{G}}(s) \equiv \frac{1}{\sqrt{n}h^{-a-1/2}} \sum_{i=1}^n \int_{\mathbb{I}} K_t(s)\, d\tilde{W}_i(t)\epsilon_i, \tag{3.10}$$

with

$$\tilde{W}_i(s) = \int_0^\tau \left\{ X_i(s) - \frac{\sum_{j=1}^n \mathcal{Y}_j(t) X_j(s)}{\sum_{j=1}^n \mathcal{Y}_j(t)} \right\} dM_i(t).$$

Specifically, we obtain a large number of realizations of $\hat{\mathcal{G}}(s)$ by repeatedly generating the standard normal random samples $(\epsilon_1, \ldots, \epsilon_n)$ while fixing the data. One may use the empirical distribution of these random samples to approximate the distribution of $\mathcal{G}(s)$. In particular, the α-percentile of $\sup_{\zeta \le s \le 1-\zeta} |\mathcal{G}(s)|$ can be approximated by the empirical percentile of a large number of realizations of $\sup_{\zeta \le s \le 1-\zeta} |\hat{\mathcal{G}}(s)|$, denoted by $\hat{\mathcal{G}}_\alpha$. Finally, we can construct the global confidence band of $\beta_0(\cdot)$ as follows:

$$\left(\hat{\beta}_{n\lambda}(\cdot) - \frac{1}{\sqrt{nh}h^a}\hat{\mathcal{G}}_\alpha, \quad \hat{\beta}_{n\lambda}(\cdot) + \frac{1}{\sqrt{nh}h^a}\hat{\mathcal{G}}_\alpha \right).$$

3.4 Simulation Studies

In this section, we conduct simulation studies to assess the finite-sample performance of the estimated confidence interval given in Section 3.2 and the uniform convergence result developed in Section 3.3.

We used a similar setup as that in Qu et al. (2016). The functional covariate X is defined as

$$X(s) = \sum_{k=1}^{50} \xi_k U_k \phi_k(s),$$

where U_k are independently sampled from the uniform distribution on $[-1, 1]$, $\xi_k = (-1)^{k+1} k^{-1/2}$, $\phi_1 = 1$, and $\phi_{k+1}(s) = \sqrt{2}\cos(k\pi s)$ for $k \geq 1$.

The functional coefficient β_0 is $\beta_0(t) = \sin(3\pi t) + 2t + 1.5$, which is from a Sobolov space $\mathcal{H}^{(2)}(\mathbb{I})$. The penalty function is $J(\beta, \beta) = \int_{\mathbb{I}} (\beta^{(2)}(t))^2 \, dt$. The scalar covariate Z is set to be univariate with distribution $N(0, 1)$ and the corresponding coefficient $\theta = 1$. The failure time T is generated from the functional Cox model:

$$h(t|W) = h_0(t) + \theta' Z + \int_0^1 X(s)\beta_0(s) \, ds,$$

where $h_0(t) = t + 5$. The censoring times, τ, are 0.4 and 0.2 which lead censoring rates around 10% and 30%, respectively. We consider the sample sizes $n = 250$, 500 and 1000. We adopt the cubic spline functions for the estimation of the functional coefficient. The number of knots is at the order of $q_n = [2n^{1/5}]$, and the knots are equally spaced. The smoothing parameter λ is 10^{-6} and the order m of Sobolev space is 2. For each combination of censoring rate and n, the simulation is repeated 1000 times.

Figure 3.1 displays an instance of estimated β and that of the pointwise 95% confidence intervals. The pointwise average of the estimated $\beta(\cdot)$ and the empirical coverage probability of the 95% pointwise confidence interval based on 1000 simu-

Table 3.1: Simulation results for the proposed estimate of θ.

		10%	30%
n=250	BIAS	0.0817	0.0787
	SSE	0.5030	0.5680
	ESE	0.7146	0.6224
	CP	0.9520	0.9360
500	BIAS	0.0484	0.0418
	SSE	0.3351	0.3881
	ESE	0.3650	0.3721
	CP	0.9570	0.9350
1000	BIAS	0.0126	0.0182
	SSE	0.2456	0.2720
	ESE	0.2513	0.3126
	CP	0.9430	0.9340

lations are shown in Figures 3.2 and 3.3, respectively. The simulation results are consistent with Theorem 3.3. In particular, these results suggest that the estimate $\hat{\beta}(t)$ is consistent. In general, it is apparent that when n increases from 250 to 1000 with the censoring rate fixed, the average bias and the standard error decrease steadily. Furthermore, the coverage probability also approaches the theoretical value of 95%. The average ESE at 10% censoring rate is lower in comparison to that at 30% censoring rate. This is consistent with the expectation that the lower the censoring rate is, the more accurate the estimate becomes.

For the regression coefficient of the scalar covariate, the BIAS, SSE, ESE, and CP of the estimated $\hat{\theta}$ are given in Table 3.1 for each setting of censoring rate and n over 1000 repetitions. As the sample size increases, the average of $\hat{\theta}$ approaches to the true value, the standard deviation reduces, and the coverage probability approaches to 95% given a fixed censoring rate. Similarly, we observe these trends as the censoring rate reduces for a given sample size.

Table 3.2 reports the coverage probability and the average width of the global confidence band derived from the uniform convergence result. Figure 3.4 displays an instance of estimated β and that of the global 95% confidence bands. The simulation

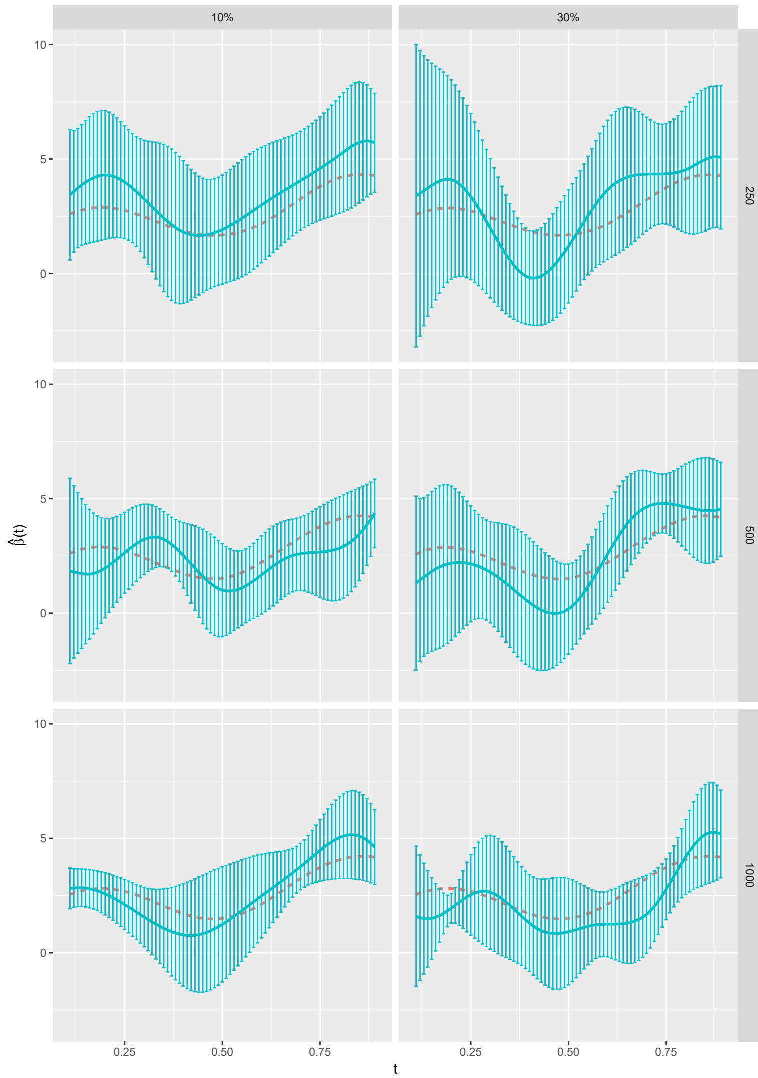

Figure 3.1: Graphical displays of $\hat{\beta}(\cdot)$ and the pointwise 95% confidence intervals of $\beta(t)$. The dashed lines represent $\beta(\cdot)$ whereas the solid lines represent $\hat{\beta}(\cdot)$.

results suggest that the coverage probability approaches the theoretical value of 95%.

In summary, the simulation results suggest that the estimates of both scalar and functional parameters are consistent and the proposed variance estimation procedure provides reasonable estimates. Also the results on the empirical coverage probability suggest that the normal approximation seems to be appropriate.

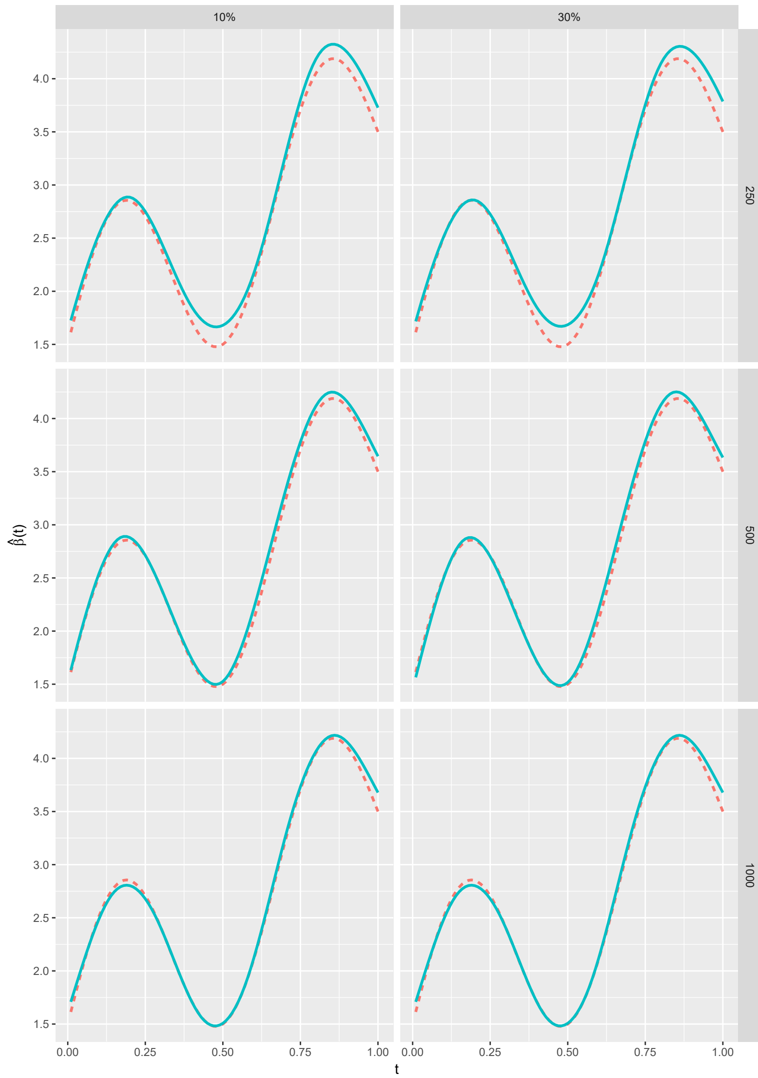

Figure 3.2: Graphical displays of the pointwise averages $\hat{\beta}(\cdot)$. The dashed lines represent $\beta(\cdot)$ whereas the solid lines represent the pointwise averages of $\hat{\beta}(\cdot)$.

3.5 An Application

In this section, we apply the proposed method to the Sequential Organ Failure Assessment (SOFA) data obtained from the Improving Care of Acute Lung Injury Patients (ICAP) study Gellar et al. (2014, 2015). The primary goal of this prospective cohort study is to investigate the long-term complications of patients who suffer from acute lung injury/acute respiratory distress syndrome (ALI/ARDS).

Figure 3.3: Graphical displays of the pointwise coverage probabilities (CP). The dashed lines represent 95% whereas the solid lines represent the pointwise CP of $\beta(\cdot)$.

The ICAP study involves 520 subjects, with 237 (46%) dying in the intensive care unit (ICU). We are interested in the association between the SOFA scores and survival among the subjects who were hospitalized in ICU for more than a week. Out of the 520 subjects, 161 subjects (31.0%) died within the first week in ICU, and they are excluded from the analysis. Therefore, the proposed method is applied to the remaining 359 subjects. In the ICAP study, data were recorded once the patients were admitted in the ICU, and then daily during hospitalization. The SOFA

Table 3.2: Simulation results for the proposed estimate of $\beta(t)$.

		10%		30%	
	$(\xi, 1-\xi)$	CP	$Width$	CP	$Width$
n=250	$(0.00, 1.00)$	0.998	49.1589	1.000	59.468
	$(0.05, 0.95)$	0.999	27.038	1.000	32.309
	$(0.10, 0.90)$	0.992	13.688	0.991	16.078
	$(0.15, 0.85)$	0.978	11.309	0.975	13.155
	$(0.20, 0.80)$	0.969	10.576	0.963	12.233
	$(0.25, 0.75)$	0.959	9.864	0.961	11.269
	$(0.30, 0.70)$	0.951	9.398	0.962	10.631
	$(0.35, 0.65)$	0.952	9.059	0.957	10.228
	$(0.40, 0.60)$	0.942	8.487	0.945	9.589
	$(0.45, 0.55)$	0.948	7.499	0.949	8.452
500	$(0.00, 1.00)$	1.000	47.884	1.000	36.882
	$(0.05, 0.95)$	1.000	21.563	0.999	21.053
	$(0.10, 0.90)$	0.989	10.873	0.992	10.749
	$(0.15, 0.85)$	0.970	9.072	0.969	8.646
	$(0.20, 0.80)$	0.956	7.870	0.958	8.124
	$(0.25, 0.75)$	0.959	6.932	0.959	7.660
	$(0.30, 0.70)$	0.953	6.560	0.950	7.331
	$(0.35, 0.65)$	0.941	6.308	0.944	7.064
	$(0.40, 0.60)$	0.949	5.888	0.952	6.627
	$(0.45, 0.55)$	0.947	5.188	0.939	5.881
1000	$(0.00, 1.00)$	1.000	22.632	1.000	28.385
	$(0.05, 0.95)$	1.000	12.816	0.998	15.912
	$(0.10, 0.90)$	0.986	6.438	0.982	8.037
	$(0.15, 0.85)$	0.959	5.271	0.954	6.477
	$(0.20, 0.80)$	0.946	4.997	0.934	6.079
	$(0.25, 0.75)$	0.933	4.714	0.919	5.689
	$(0.30, 0.70)$	0.927	4.517	0.913	5.421
	$(0.35, 0.65)$	0.938	4.349	0.926	5.225
	$(0.40, 0.60)$	0.927	4.087	0.927	4.883
	$(0.45, 0.55)$	0.916	3.617	0.917	4.300

score is one of the measurements recorded daily. SOFA is a measure of the overall organ function status of a patient. It is composed of respiratory, cardiovascular, coagulation, liver, renal, and neurological components. Each component ranges from 0 to 4, with higher scores suggesting inferior organ function. The SOFA score, ranging from 0 to 24, is then the sum of these six scores. We treat the history of

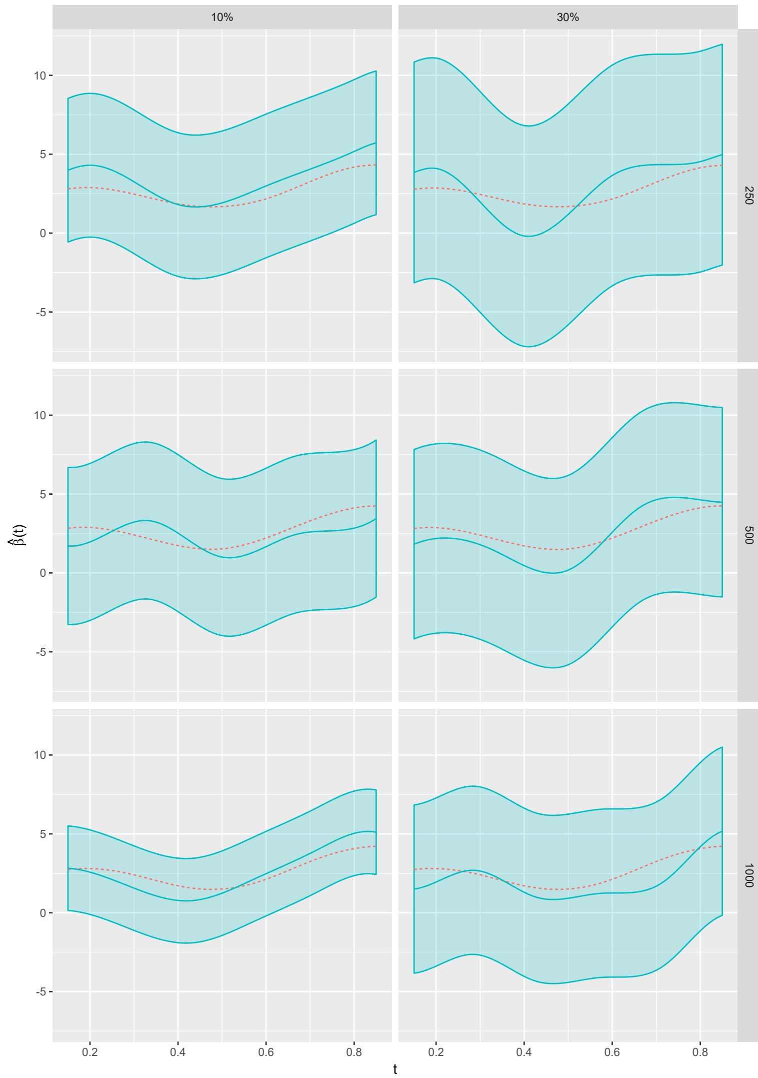

Figure 3.4: Graphical displays of $\hat{\beta}(\cdot)$ and the global 95% confidence band of $\beta(\cdot)$. The dashed lines represent $\beta(\cdot)$ whereas the solid lines represent $\hat{\beta}(\cdot)$.

each subject's SOFA scores, in the first week, as a functional covariate, $X(s)$, where s is the number of days since the admission to the ICU. Trajectories of the SOFA score of subjects who died after the first week of ICU hospitalization and those who survived are depicted in Figure 4. It is apparent that among patients who manage to survive, the pointwise averages of SOFA scores are declining, whereas among patients who died after the first week of ICU hospitalization, the averages are relatively stable. Our model includes three scalar covariates as controls of a subject's baseline risk.

They are age, gender, and Charlson co-morbidity index (Charlson et al. (1987)).

Our goal is to estimate the association between the trajectory of SOFA scores and mortality among subjects who are hospitalized in ICU for more than a week. We adopt the cubic spline functions for the estimation of the functional coefficient. The number of knots is at the order of $q_n = [2n^{1/5}] = 7$, and the knots are equally spaced. We apply 5-fold cross validation to optimize the smoothing parameter of a penalized pseudo-score function.

We plot the estimated functional coefficient $\hat{\beta}(\cdot)$ in Figure 3.6. The result suggests that there is a functional association between time to death during the ICU stay and the SOFA score function for $t \in [0, 0.1] \cup [0.75, 1]$, which corresponds to the first two days and the last two days of ICU stay. This implies that the SOFA score in the these days in the first week of ICU stay may be used as an indicator of one's hazard.

Table 3.3 summarizes the estimation of the regression coefficients of the scalar covariates. In addition to the functional covariate, there seems to be a positive association with two scalar covariates, patients' age gender, and a negative association with Charlson co-morbidity index.

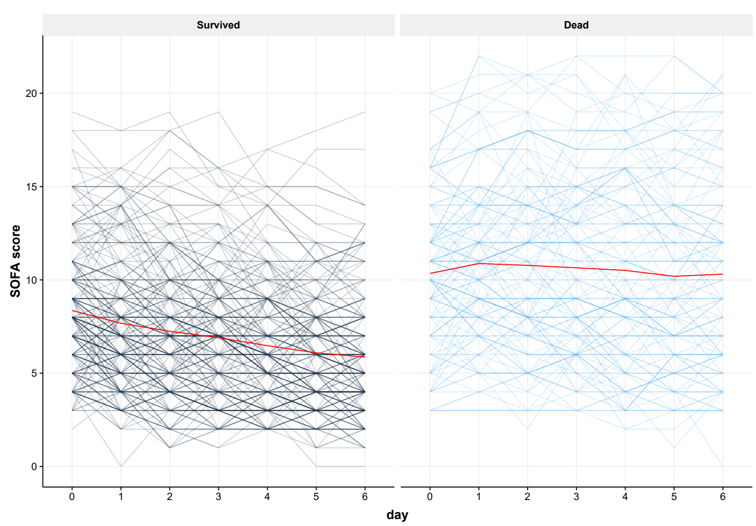

Figure 3.5: Trajectories of the SOFA score of subjects who died after the first week of the ICU hospitalization and those who survived. The red lines are the pointwise average of the SOFA score.

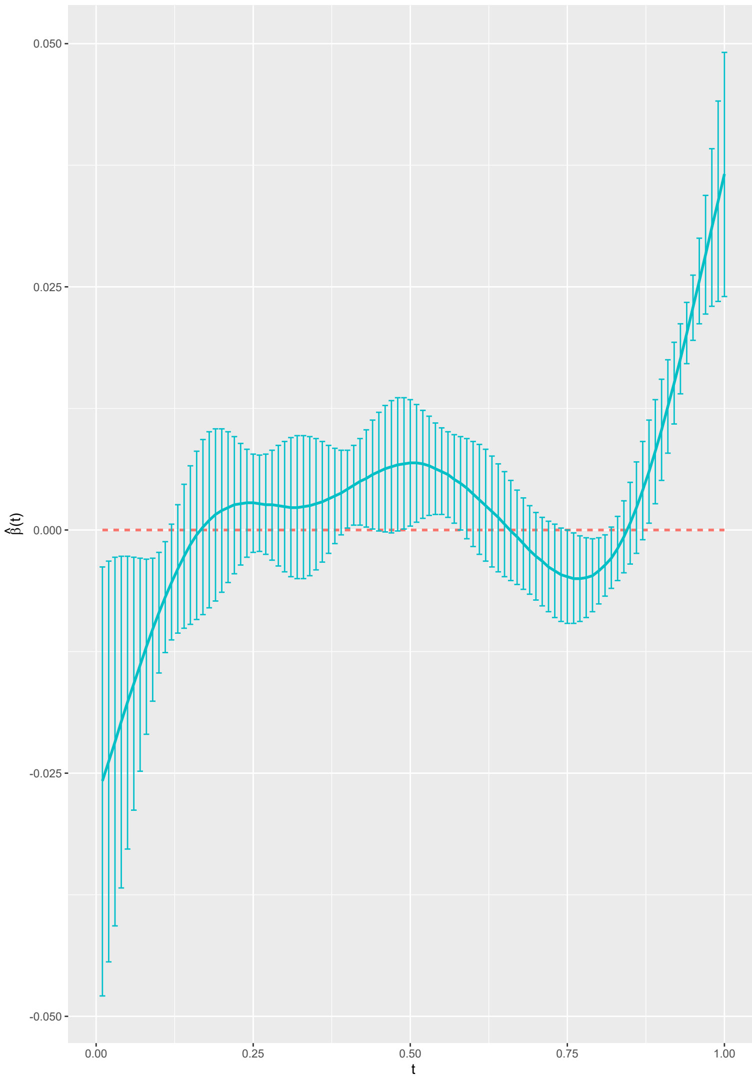

Figure 3.6: The estimated functional coefficient $\hat{\beta}(\cdot)$ and the pointwise 95% confidence interval for the SOFA data analysis.

Table 3.3: Estimation results of regression coefficients of scalar covariates for the SOFA data analysis

	$\hat{\theta}$	$S.E.$	t-value	p-value
Age	0.0003	0.0252×10^{-3}	10.6148	< 0.0001
Gender (male=1)	0.0023	0.7685×10^{-3}	2.9308	0.0017
Charlson Index	-0.0007	0.0887×10^{-3}	-8.4146	< 0.0001

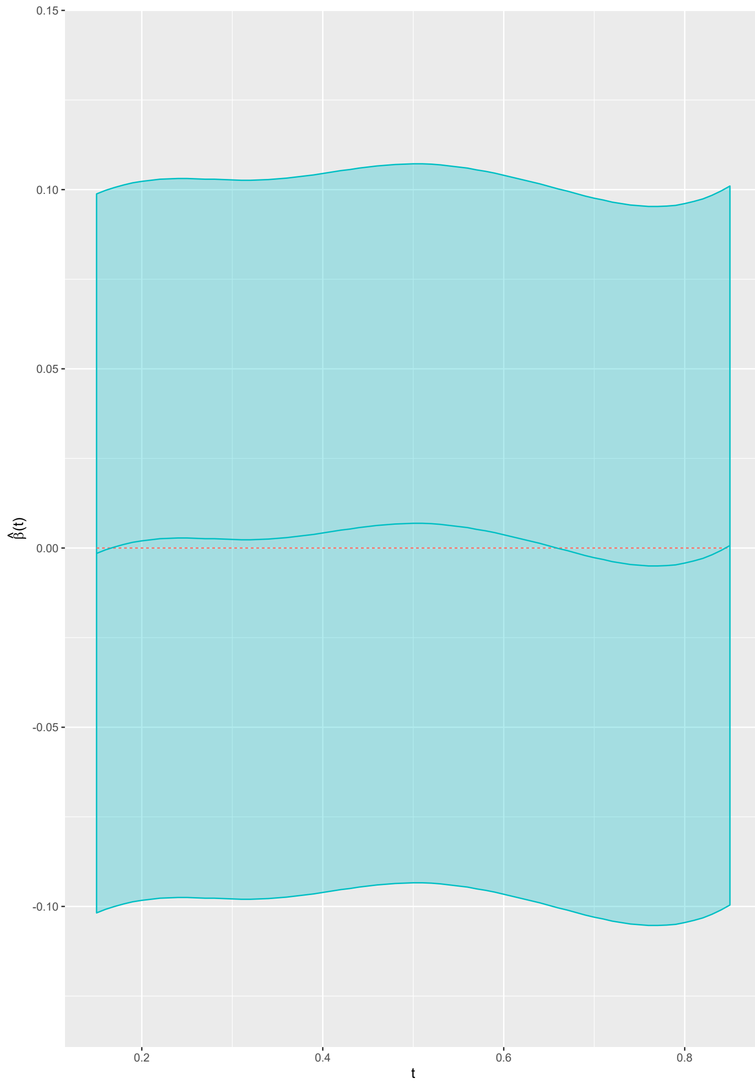

Figure 3.7: The estimated functional coefficient $\hat{\beta}(\cdot)$ and the global 95% confidence band for the SOFA data analysis.

3.6 Appendix

In the following, we use c to denote different positive constants in different places. In addition, $a \lesssim b$ means $a \leq cb$ and $a \gtrsim b$ means $a \geq cb$.

3.6.1 Proofs of Lemmas

Proof of Lemma 3.1

The first part of Lemma 3.1 follows from a direct calculation. It follows from

Assumption A3 that $\{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1}$ is asymptotic positive definite. Denote c as the minimum eigenvalue of it. Then we have $\{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1} \leq c^{-1}\mathbf{1}$ with $\mathbf{1}$ being the identity matrix. Thus, we have

$$\mathbf{z}^\top \{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1} \mathbf{z} \lesssim \|\mathbf{z}\|_2^2. \tag{3.11}$$

A direct calculation yields that

$$< \pi_x, \pi_x > \lesssim \|x\|_{L^2}^2 h^{-2a-1}.$$

Besides, it follows from Cauchy-Schwarz inequality that

$$V(G_k, \pi_x) \leq \|G_k\|_m \|\pi_x\|_m \lesssim \|x\|_{L^2} h^{-a-1/2},$$

and hence, there exits a constant $c_r > 0$ such that

$$< \mathcal{R}_{\mathbf{W}}, \mathcal{R}_{\mathbf{W}} >_\lambda \leq c_r(\|\mathbf{Z}\|_2^2 + \|X\|_{L^2}^2 h^{-2a-1}).$$

Finally, it follows from (3.11) and the proof of Lemma S.4 in Cheng and Shang (2015) that

$$E < \mathcal{R}_{\mathbf{W}}, \mathcal{R}_{\mathbf{W}} >_\lambda \leq c_r h^{-1}.$$

Proof of Lemma 3.2 By the definition of K_t, we have $\|K_t\|_\lambda \lesssim h^{-(a+1/2)}$. Following the idea in the proof of Lemma 3.1, we have that $\|\tilde{\mathcal{R}}_{\mathbf{u}}\|_\lambda \lesssim h^{-(a+1/2)}$.

It follows from the fact that

$$\begin{aligned}
\|\alpha\|_e =& \|\theta\|_2 + \|\beta\|_{L_2} \\
\leq& \|\theta\|_2 + \|\beta\|_{\sup} \\
=& \sup_{\|z\|_2=1, t\in\mathbb{I}} |\beta(t) + \theta^\top z| \\
=& \sup_{\|z\|_2=1, t\in\mathbb{I}} < \tilde{\mathcal{R}}_{\mathbf{u}}, \alpha >_\lambda \\
\leq& \|\alpha\|_\lambda \sup_{\|z\|_2=1, t\in\mathbb{I}} \|\tilde{\mathcal{R}}_{\mathbf{u}}\|_\lambda \\
\lesssim& h^{-(a+1/2)} \|\alpha\|_\lambda.
\end{aligned}$$

Proof of Lemma 3.3 It follows from Assumptions A1 and A4 that

$$\begin{aligned} E_{\mathbf{W}}\{< \mathcal{R}_{\mathbf{W}}, \alpha >_\lambda\}^4 &= E_{\mathbf{W}}[\{\theta^\top \mathbf{Z} + \int_{\mathbb{I}} X(t)\beta(t)\, dt\}^4] \\ &\leq M_0[E_{\mathbf{W}}\{|\theta^\top \mathbf{Z} + \int_{\mathbb{I}} X(t)\beta(t)\, dt|^2\}]^2 \\ &\leq c_1 \|\alpha\|_\lambda^4. \end{aligned}$$

3.6.2 Proofs of Theorems

Proof of Theorem 3.1 In order to prove Theorem 3.1, we need the following subset of $\mathcal{H}$:

$$\mathcal{F}_{p_n} = \{\alpha = (\theta^\top, \beta(\cdot))^\top \in \mathcal{H} : \|\theta\|_2 \leq 1, \|\beta\|_{L_2} \leq 1, J(\beta,\beta) \leq p_n\}.$$

It follows directly from $D^2\mathcal{S}_{n\lambda}(\alpha) = 0 = D^2\mathcal{S}_\lambda(\alpha)$ that there exists a unique value α_λ which satisfies $\mathcal{S}_\lambda(\alpha_\lambda) = 0$, and a unique value $\hat{\alpha}_{n\lambda}$ which satisfies $\mathcal{S}_{n\lambda}(\hat{\alpha}_{n\lambda}) = 0$. It follows from Proposition 3.1 and Assumption A1(c) that $\hat{\alpha}_{n\lambda}$ is the global maximum of $l_{n\lambda}(\alpha)$ asymptotically. In the following, we show the uniqueness of the estimate and derive the order of convergence of the estimate.

Let $r_{1n} = 2\{J(\beta_0,\beta_0)+1\}^{1/2}h^k$, and define the operator:

$$T_{1h}(\alpha) = \alpha + \mathcal{S}_\lambda(\alpha_0 + \alpha), \alpha \in \mathcal{H}.$$

Then,

$$\begin{aligned} \|T_{1h}(\alpha)\|_\lambda =& \|\alpha + \mathcal{S}_\lambda(\alpha + \alpha_0)\|_\lambda \\ \leq& \|\alpha + \mathcal{S}_\lambda(\alpha + \alpha_0) - \mathcal{S}_\lambda(\alpha_0)\|_\lambda + \|\mathcal{S}_\lambda(\alpha_0)\|_\lambda. \end{aligned}$$

Let $\mathbb{B}(\epsilon) = \{\alpha \in \mathcal{H}, \|\alpha\|_\lambda \leq \epsilon\}$ be a ball of radius ϵ in $\mathcal{H}$. Note that $\mathcal{S}(\alpha_0) = 0$. This implies that $\mathcal{S}_\lambda(\alpha_0) = -\mathcal{P}_\lambda\alpha_0$. Then

$$\|\mathcal{S}_\lambda(\alpha_0)\|_\lambda = \|\mathcal{P}_\lambda(\alpha_0)\|_\lambda \leq \{\lambda J(\beta_0,\beta_0)\}^{1/2} \leq \{J(\beta_0,\beta_0)+1\}^{1/2}h^k = \frac{r_{1n}}{2}. \quad (3.12)$$

Following Proposition 3.1, we have that

$$\begin{aligned}\|\alpha+\mathcal{S}_\lambda(\alpha+\alpha_0)-\mathcal{S}_\lambda(\alpha_0)\|_\lambda =&\|\alpha+D\mathcal{S}_\lambda(\alpha_0)\alpha+\int_0^1\int_0^1 sD^2\mathcal{S}_\lambda(\alpha_0+ss'\alpha)\alpha\alpha\,ds\,ds'\|_\lambda\\ =&\|\int_0^1\int_0^1 sD^2\mathcal{S}_\lambda(\alpha_0+ss'\alpha)\alpha\alpha\,ds\,ds'\|_\lambda\\ =&0. \end{aligned} \tag{3.13}$$

From (3.12) and (3.13), we have that

$$\|T_{1h}\|_\lambda \le \frac{r_{1n}}{2}. \tag{3.14}$$

This implies $T_{1h}(\mathbb{B}(r_{1n}))\subset\mathbb{B}(r_{1n})$. Next, we show that T_{1h} is a contraction mapping. For any $\alpha_j=(\theta_j^\top,\beta_j(\cdot))^\top\in\mathcal{H}, j=1,2$, we have

$$\begin{aligned}T_{1h}(\alpha_1)-T_{1h}(\alpha_2)&=\alpha_1-\alpha_2+\mathcal{S}_\lambda(\alpha_0+\alpha_1)-\mathcal{S}_\lambda(\alpha_0+\alpha_2)\\ &=\int_0^1[D\mathcal{S}_\lambda\{\alpha_0+\alpha_2+s(\alpha_1-\alpha_2)\}-D\mathcal{S}_\lambda(\alpha_0)](\alpha_1-\alpha_2)\,ds\\ &=0.\end{aligned}$$

Therefore, $T_{1h}(\alpha)$ is a contraction mapping on $\mathbb{B}(r_{1n})$. By the Banach fixed-point theorem, there exists a unique element $\alpha'_\lambda\in\mathbb{B}(r_{1n})$ such that $T_{1h}(\alpha'_\lambda)=\alpha'_\lambda$. Define $\alpha_\lambda=\alpha'_\lambda+\alpha_0$. We have $\mathcal{S}_\lambda(\alpha_\lambda)=0$ and $\|\alpha_\lambda-\alpha_0\|_\lambda\le r_{1n}$.

Now, we show that there exists a unique value $\hat{\alpha}_{n\lambda}$ which satisfies $\mathcal{S}_{n\lambda}(\hat{\alpha}_{n\lambda})=0$. As $\|\alpha_\lambda-\alpha_0\|=O(r_{1n})=o(1)$ and $D\mathcal{S}_\lambda(\alpha_0)=-id$, from Proposition 3.1, we have that $D\mathcal{S}_\lambda(\alpha_\lambda)=-id$ is invertible. Next, define the operator

$$\begin{aligned}T_{2h}(\alpha)=&\alpha-[D\mathcal{S}_\lambda(\alpha_\lambda)]^{-1}\mathcal{S}_{n\lambda}(\alpha_\lambda+\alpha)\\ =&\{D\mathcal{S}_{n\lambda}(\alpha_\lambda)\alpha-D\mathcal{S}_\lambda(\alpha_\lambda)\alpha\}\\ &+\{\mathcal{S}_{n\lambda}(\alpha_\lambda+\alpha)-\mathcal{S}_{n\lambda}(\alpha_\lambda)\\ &-D\mathcal{S}_{n\lambda}(\alpha_\lambda)\alpha\}+\mathcal{S}_{n\lambda}(\alpha_\lambda)\\ \equiv&I_1+I_2+I_3.\end{aligned}$$

It follows from the functional central limit theorem that uniformly in $t \in \mathbb{I}$, we have

$$\Big\|\frac{1}{n}\sum_{j=1}^{n}\mathcal{Y}_j(t) - s_1^{(0)}(t)\Big\|_\infty = O_p(n^{-1/2}), \tag{3.15}$$

It follows from Lemma 3.2 and the functional central limit theorem that

$$\begin{aligned}
&\|\frac{1}{n}\sum_{j=1}^{n}\mathcal{Y}_j(t)\mathcal{R}_{\mathbf{W}_j} - E[\mathcal{Y}_j(t)\mathcal{R}_{\mathbf{W}_j}]\big\|_\lambda \\
=&\sup_{\|\alpha_1\|_\lambda=1} < \frac{1}{n}\sum_{j=1}^{n}\mathcal{Y}_j(t)\mathcal{R}_{\mathbf{W}_j} - E[\mathcal{Y}_j(t)\mathcal{R}_{\mathbf{W}_j}], \alpha_1 >_\lambda \\
=&O_p(n^{-1/2}h^{-a-1/2}).
\end{aligned} \tag{3.16}$$

It follows from the Taylor expansion that

$$\begin{aligned}
&\|\frac{1}{n}\sum_{j=1}^{n}\frac{\mathcal{Y}_j(t)\mathcal{R}_{\mathbf{W}_j}}{S_1^{(0)}(t)} - \frac{E[\mathcal{Y}_j(t)\mathcal{R}_{\mathbf{W}_j}]}{s_1^{(0)}(t)}\|_\lambda \\
=&O_p(n^{-1/2}h^{-a-1/2}).
\end{aligned} \tag{3.17}$$

It follows from $\mathcal{S}_\lambda(\alpha_\lambda) = 0$, Lemma 3.1 and formula (3.15 – 3.17) that

$$
\begin{aligned}
&E\|I_3\|_\lambda^2 \\
&= E\{\|\mathcal{S}_{n\lambda}(\alpha_\lambda) - \mathcal{S}_\lambda(\alpha_\lambda)\|_\lambda\}^2 \\
&= E\Big\{\Big\|\frac{1}{n}\sum_{i=1}^n \int_0^1 \big[\mathcal{R}_{\mathbf{W}_i} - \frac{E\{\mathcal{Y}(t)\mathcal{R}_{\mathbf{W}}\}}{s_1^{(0)}(t)}\big]\, dM_i(t) \\
&- E\Big(\int_0^1 [\mathcal{R}_{\mathbf{W}_i} - \frac{E\{\mathcal{Y}(t)\mathcal{R}_{\mathbf{W}}\}}{s_1^{(0)}(t)}]\, dM_i(t)\Big)\Big\|_\lambda^2\Big\} \\
&+ O(n^{-1}h^{-2a-1}(n^{-1} + h^{2k})) \\
&\le 2E\Big\{\|\frac{1}{n}\sum_{i=1}^n \big[\Delta_i \mathcal{R}_{\mathbf{W}_i} - E\Delta_i\mathcal{R}_{\mathbf{W}_i}]\|_\lambda^2 \\
&+ \|\frac{1}{n}\sum_{i=1}^n \int_0^1 \frac{E\{\mathcal{Y}(t)\mathcal{R}_{\mathbf{W}}\}}{s_1^{(0)}(t)}\{\, dN_i(t) - E\, dN_i(t)\}\|_\lambda^2 \\
&+ \|[\frac{1}{n}\sum_{i=1}^n \int_0^1 \frac{E\{\mathcal{Y}(t)\mathcal{R}_{\mathbf{W}}\}}{s_1^{(0)}(t)}[\mathcal{Y}_i(t)\eta_{\alpha_0}(\mathbf{W}_i) - E\{\mathcal{Y}_i(t)\eta_{\alpha_0}(W_i)\}]\, dt\|_\lambda^2 \\
&+ \|[\frac{1}{n}\sum_{i=1}^n \int_0^1 [\mathcal{R}_{\mathbf{W}_i}\mathcal{Y}_i(t)\eta_{\alpha_0}(\mathbf{W}_i) - E\{\mathcal{R}_{\mathbf{W}_i}\mathcal{Y}_i(t)\eta_{\alpha_0}(\mathbf{W}_i)\}]\, dt\|_\lambda^2\Big\} \\
&+ \|[\frac{1}{n}\sum_{i=1}^n \int_0^1 \frac{E[\mathcal{Y}_j(t)\mathcal{R}_{\mathbf{W}_j}]}{s_1^{(0)}(t)}[\mathcal{Y}_i(t) - E\{\mathcal{Y}_i(t)\}]h_0(t)\, dt\|_\lambda^2 \\
&+ \|[\frac{1}{n}\sum_{i=1}^n \int_0^1 [\mathcal{R}_{\mathbf{W}_i}\mathcal{Y}_i(t)\eta_{\alpha_0}(\mathbf{W}_i) - E\{\mathcal{R}_{\mathbf{W}_i}\mathcal{Y}_i(t)h_0(t)\}]\, dt\|_\lambda^2\Big\} + o(n^{-1}h^{-1}) \\
&= O((hn)^{-1}).
\end{aligned}
$$

This implies that $\|\mathcal{S}_{n\lambda}(\alpha_\lambda)\|_\lambda = O_p((nh)^{-1/2})$. Let c be a positive constant such that $P(\|\mathcal{S}_{n\lambda}(\alpha_\lambda)\|_\lambda \le c(nh)^{-1/2}) \to 1$. Define $r_{2n} = 2c(nh)^{-1/2}$ and $\mathbb{B}(r_{2n}) = \{\alpha \in \mathcal{H} : \|\alpha\|_\lambda \le r_{2n}\}$. Then we have $P(\|\mathcal{S}_{n\lambda}(\alpha_\lambda)\|_\lambda \le r_{2n}/2) \to 1$. Define $\Gamma = \cap_{i=1}^n A_{ni}$, where

$$A_{ni} = \{\|\mathbf{Z}_i\|_2 \le c\log(n), \|X_i\|_{L^2} \le c\log(n)\},$$

and c is positive definite. Under Assumption A4, we choose c large enough such that $P(\Gamma) \to 1$, and $P(A_{ni}^c) = O(n^{-1})$. To handle I_1, we have that

$$
\begin{aligned}
&\|I_1\|_\lambda \\
&= \|D\mathcal{S}_{n\lambda}(\alpha_\lambda)\alpha - D\mathcal{S}_\lambda(\alpha_\lambda)\alpha\|_\lambda \\
&\le \Big\|\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau [\mathcal{Y}_i(t)\eta_\alpha(\mathbf{W}_i)\mathcal{R}_{\mathbf{W}_i} - E\{\mathcal{Y}_i(t)\eta_\alpha(\mathbf{W}_i)\mathcal{R}_{\mathbf{W}_i}\}]\,dt\Big\|_\lambda \\
&+ \Big\|\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau [\mathcal{Y}_i(t)\mathcal{R}_{\mathbf{W}_i} - E\{\mathcal{Y}_i(t)\mathcal{R}_{\mathbf{W}_i}\}]\frac{E\{\mathcal{Y}(t)\eta_\alpha(\mathbf{W})\}}{s_1^{(0)}(t)}\,dt\Big\|_\lambda \\
&+ \Big\|\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{\mathcal{Y}_i(t)\mathcal{R}_{\mathbf{W}_i}\}\Big[\frac{E\{\mathcal{Y}(t)\eta_\alpha(\mathbf{W})\}}{s_1^{(0)}(t)} - \frac{n^{-1}\sum_{j=1}^n\{\mathcal{Y}_j(t)\eta_\alpha(\mathbf{W}_j)\}}{S_1^{(0)}(t)}\Big]\,dt\Big\|_\lambda \\
&\equiv I_{11} + I_{12} + I_{13}.
\end{aligned}
\tag{3.18}
$$

For I_{11}, we have that

$$
I_{11} = \Big\|\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau [\mathcal{Y}_i(t)\eta_\alpha(\mathbf{W}_i)\mathcal{R}_{\mathbf{W}_i} - E\{\mathcal{Y}_i(t)\eta_\alpha(\mathbf{W}_i)\mathcal{R}_{\mathbf{W}_i}\}]\,dt\Big\|_\lambda.
$$

To infer I_{11}, we define

$$
\phi(Y_j, \mathbf{W}_j; \alpha) = \frac{\mathcal{Y}_j(t_0)\eta_\alpha(\mathbf{W}_j)}{s_1^{(0)}(t_0)} I_{A_{nj}}.
$$

Then for any $\alpha_1, \alpha_2 \in \mathcal{H}$, we have that

$$
\begin{aligned}
&|\phi(Y_j, \mathbf{W}_j; \alpha_1) - \phi(Y_j, \mathbf{W}_j; \alpha_2)| \\
&= \frac{1}{s_1^{(0)}(t_0)}\mathcal{Y}_j(t_0)|\{\eta_{\alpha_1}(\mathbf{W}_j) - \eta_{\alpha_2}(\mathbf{W}_j)\}|I_{A_{nj}} \\
&\le \frac{1}{s_1^{(0)}(t_0)}| < \mathcal{R}_{\mathbf{W}_j}, \alpha_1 - \alpha_2 >_\lambda |I_{A_{nj}} \\
&\le \frac{c\log(n)}{s_1^{(0)}(t_0)}\|\alpha_1 - \alpha_2\|_e.
\end{aligned}
$$

Let $\phi_n(Y_j, \mathbf{W}_j; \alpha) = s_1^{(0)}(t_0)c^{-1}\{\log(n)\}^{-1}\phi(Y_j, \mathbf{W}_j; \alpha_1)$. Then $|\phi_n(Y_j, \mathbf{W}_j; \alpha_1) - \phi_n(Y_j, \mathbf{W}_j; \alpha_1)| \leq \|\alpha_1 - \alpha_2\|_e$. For any $\alpha \neq 0 \in \mathcal{H}$, let $\tilde{\alpha} = \alpha/(d_n\|\alpha\|_\lambda)$, where $d_n = \kappa h^{-(2a+1)/2}$. It follows from Lemma 3.2 that $\|\tilde{\alpha}\|_e \leq d_n\|\tilde{\alpha}\|_\lambda = 1$. Then we have $\|\tilde{\theta}\|_2 + \|\tilde{\beta}\|_{L_2} \leq 1$. Meanwhile, we have $\lambda J(\tilde{\beta}, \tilde{\beta}) \leq \|\tilde{\alpha}\|_\lambda^2 = d_n^{-2}$. Then $J(\tilde{\beta}, \tilde{\beta}) \leq \lambda^{-1}d_n^{-2} \equiv p_n$. Then, we have that for any $\alpha \in \mathbb{B}(r_{2n})$,

$$\lim_n P(\| \sum_{j=1}^n [\phi_n(Y_j, \mathbf{W}_j; \tilde{\alpha})\mathcal{R}_{\mathbf{W}_j} - E\{\phi_n(Y_j, \mathbf{W}_j; \tilde{\alpha})\mathcal{R}_{\mathbf{W}_j}\}]\|_\lambda$$
$$\lesssim (n^{1/2}p_n^{1/(4m)} + 1)\{h^{-1}\log\log(n)\}^{1/2}) = 1.$$

Therefore, we have

$$\lim_n P(\| \sum_{j=1}^n [\phi(Y_j, \mathbf{W}_j; \alpha)\mathcal{R}_{\mathbf{W}_j} - E\{\phi(Y_j, \mathbf{W}_j; \alpha)\mathcal{R}_{\mathbf{W}_j}\}]\|_\lambda$$
$$\lesssim d_n\{\log(n)\}\|\alpha\|_\lambda(n^{1/2}p_n^{1/(4m)} + 1)\{h^{-1}\log\log(n)\}^{1/2}) = 1.$$

It follows from the definition of A_{ni} that

$$\|\frac{E\{\mathcal{Y}_j(t_0)\eta_\alpha(\mathbf{W}_j)\mathcal{R}_{\mathbf{W}_j}I_{A_{nj}^c}\}}{s_1^{(0)}(t)}\|_\lambda$$
$$\leq c_1 E\| < \mathcal{R}_{\mathbf{W}_j}, \alpha > \mathcal{R}_{\mathbf{W}_j}I_{A_{nj}^c}\|_\lambda = O(P(A_{ni}^c)^{1/2}h^{-1/2})\|\alpha\|_\lambda = o(1)\|\alpha\|_\lambda.$$

Thus, we have $I_{11} = O_p(n^{-1/2}h^{-(a+1)-\frac{2k-2a-1}{4m}}\{\log(n)\}\{\log\log(n)\}^{1/2})\|\alpha\|_\lambda + o_p(1)\|\alpha\|_\lambda = o_p(1)\|\alpha\|_\lambda$.

Similarly to I_{11}, we obtain

$$I_{12}$$
$$= \Big\|\frac{1}{n}\sum_{i=1}^n \int_0^\tau [\mathcal{Y}_i(t)\mathcal{R}_{\mathbf{W}_i} - E\{\mathcal{Y}_i(t)\mathcal{R}_{\mathbf{W}_i})\}]\frac{E\{\mathcal{Y}(t)\eta_\alpha(\mathbf{W})\}}{s_1^{(0)}(t)}\Big\|_\lambda$$
$$= O_p(n^{-1/2}h^{-a-1/2})\|\alpha\|_\lambda.$$

For I_{13}, we have

$$I_{13} = \Big\| \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \{\mathcal{Y}_i(t)\mathcal{R}_{\mathbf{W}_i}\} \Big[\frac{E\{\mathcal{Y}(t)\eta_\alpha(\mathbf{W})\}}{s_1^{(0)}(t)} - \frac{n^{-1}\sum_{j=1}^n \{\mathcal{Y}_j(t)\eta_\alpha(\mathbf{W}_j)\}}{S_1^{(0)}(t)} \Big] \, dt \Big\|_\lambda$$

$$= O_p(n^{-1/2}h^{-a-1})\|\alpha\|_\lambda.$$

Therefore, for any $\alpha \in \mathbb{B}(r_{2n})$, we have $\|I_1\|_\lambda \leq r_{2n}/18$ and For $\|I_2\|_\lambda$, we have

$$\begin{aligned}
\|I_2\|_\lambda &= \|\{\mathcal{S}_{n\lambda}(\alpha_\lambda + \alpha) - \mathcal{S}_{n\lambda}(\alpha_\lambda) - D\mathcal{S}_{n\lambda}(\alpha_\lambda)\alpha\}\|_\lambda \\
&= \| \int_0^1 \int_0^1 sD^2\mathcal{S}_{n\lambda}(\alpha_\lambda + ss'\alpha)\alpha\alpha \, ds \, ds'\|_\lambda \\
&= 0.
\end{aligned}$$

Therefore, for any $\alpha \in \mathbb{B}(r_{2n})$,

$$\|T_{2h}(\alpha)\|_\lambda \leq \frac{r_{2n}}{2}.$$

That is, $T_{2h}(\mathbb{B}(r_{2n})) \subset \mathbb{B}(r_{2n})$. Using the arguments above, we show that T_{2h} is a contraction mapping in $\mathbb{B}(r_{2n})$. Therefore, there exists a unique element $\alpha' \in \mathbb{B}(r_{2n})$, such that $T_{2h}(\alpha') = \alpha'$. This implies $\mathcal{S}_{n\lambda}(\alpha_\lambda + \alpha') = 0$. Let $\hat{\alpha}_{n\lambda} = \alpha_\lambda + \alpha'$. Then $\mathcal{S}_{n\lambda}(\hat{\alpha}_{n\lambda}) = 0$. Finally, with probability going to 1, we have

$$\|\hat{\alpha}_{n\lambda} - \alpha_0\| \leq r_{1n} + r_{2n} = O_P((nh)^{-1/2} + h^k).$$

Proof of Theorem 3.2

It follows from Theorem 3.1 that there exists a constant $M > 0$ such that, with probability approaching to one, $\|\hat{\alpha}_{n\lambda} - \alpha_0\|_\lambda \leq Mr_n$. For simplicity, denote $\hat{\alpha}_{n\lambda} - \alpha_0$ as α. We assume that $\|\alpha\|_\lambda \leq Mr_n$ as its complement is negligible in terms of probability. Let $d_n = \kappa M h^{-(2a+1)/2} r_n$ and $\tilde{\alpha} = d_n^{-1}\alpha$. Let $p_n = \kappa^{-2}h^{1-2m}$, where κ is the constant given in Lemma 3.2. When n is large, $h \to 0$ and $1 - 2k < 0$. Hence, $p_n \geq 1$ as $n \to \infty$. It can be shown that $\|\alpha\|_\lambda \leq Mr_n$ implies $\tilde{\alpha} \in \mathcal{F}_{p_n}$. To see this, write $\tilde{\alpha} = (\tilde{\theta}^\top, \tilde{\beta})^\top$. Then $\|\tilde{\alpha}\|_e = d_n^{-1}\|\alpha\|_e \leq d_n^{-1}\kappa h^{-(2a+1)/2}\|\alpha\|_\lambda \leq$

$d_n^{-1}\kappa h^{-(2a+1)/2} M r_n = 1$. Thus, it follows from

$$J(\tilde{\beta},\tilde{\beta}) = d_n^{-2}\lambda^{-1}\{\lambda J(\beta,\beta)\}$$
$$\leq d_n^{-2}\lambda^{-1}\|\alpha\|_\lambda^2 \leq d_n^{-2}\lambda^{-1}(Mr_n)^2 = \kappa^{-2}h^{1-2m} = p_n.$$

Thus we have that

$$\begin{aligned}&\|\mathcal{S}_{n\lambda}(\alpha+\alpha_0) - \mathcal{S}_{n\lambda}(\alpha_0) - \{\mathcal{S}_\lambda(\alpha+\alpha_0) - \mathcal{S}_\lambda(\alpha_0)\}\|_\lambda \\ =&\|\mathcal{S}_n(\alpha+\alpha_0) - \mathcal{S}_n(\alpha_0) - \{\mathcal{S}(\alpha+\alpha_0) - \mathcal{S}(\alpha_0)\}\|_\lambda. \end{aligned} \tag{3.19}$$

On the left hand side of equation (3.19), we have that

$$\begin{aligned}&\|\mathcal{S}_{n\lambda}(\alpha+\alpha_0) - \mathcal{S}_{n\lambda}(\alpha_0) - \{\mathcal{S}_\lambda(\alpha+\alpha_0) - \mathcal{S}_\lambda(\alpha_0)\}\|_\lambda \\ =&\| - \mathcal{S}_{n\lambda}(\alpha_0) - D\mathcal{S}_\lambda(\alpha_0)\alpha - \int_0^1\int_0^1 sD^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha\, ds\, ds'\|_\lambda \\ =&\|\alpha - \mathcal{S}_{n\lambda}(\alpha_0) - \int_0^1\int_0^1 sD^2\mathcal{S}_\lambda(\alpha_0 + ss'\alpha)\alpha\alpha\, ds\, ds'\|_\lambda \\ =&\|\alpha - \mathcal{S}_{n\lambda}(\alpha_0)\|_\lambda. \end{aligned} \tag{3.20}$$

It follows from $D^2\mathcal{S}_n(\alpha) = 0$ and $D^2\mathcal{S}(\alpha) = 0$ that, on the right hand side of equation (3.19), we have

$$\begin{aligned}\|\mathcal{S}_n(\alpha+\alpha_0) - \mathcal{S}_n(\alpha_0) - \{\mathcal{S}(\alpha+\alpha_0) - \mathcal{S}(\alpha_0)\}\|_\lambda =&\|D\mathcal{S}_n(s'\alpha+\alpha_0)\alpha - D\mathcal{S}(s''\alpha+\alpha_0)\alpha\|_\lambda \\ =&\|D\mathcal{S}_n(\alpha_\lambda)\alpha - D\mathcal{S}(\alpha_\lambda)\alpha\|_\lambda \\ =&\|D\mathcal{S}_{n\lambda}(\alpha_\lambda)\alpha - D\mathcal{S}_\lambda(\alpha_\lambda)\alpha\|_\lambda \end{aligned}$$

Finally, it follows from equation (3.19) and the lines of the proof of Theorem 3.1 that

$$\|\alpha - \mathcal{S}_{n\lambda}(\alpha_0)\|_\lambda \leq O_p(a_n).$$

The proof of Theorem 3.2 is completed.

Proof of Theorem 3.3

Define $\hat{\alpha}_{n\lambda}^h = (\hat{\theta}_{n\lambda}^\top, h^{a+1/2}\hat{\beta}_{n\lambda})^\top$, $\alpha_0^* = (id - \mathcal{P}_\lambda)\alpha_0 \equiv (\theta_0^{*\top}, \beta_0^*)^\top$, $\alpha_0^{*h} = (\theta_0^{*\top}, h^{a+1/2}\beta_0^*)^\top$, and $\mathcal{R}_{\mathbf{W_i}}^h = (H_{\mathbf{W_i}}, h^{a+1/2}T_{\mathbf{W_i}})$. Denote

$$
\begin{aligned}
&Rem_n\\
=&\hat{\alpha}_{n\lambda} - \alpha_0^* - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{\mathcal{R}_{\mathbf{W}_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\mathcal{R}_{\mathbf{W}_j}}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)}\}\{\, dN_i(t) - \mathcal{Y}_i(t)\eta_{\alpha_0}(\mathbf{W}_i)\, dt\}.
\end{aligned}
$$

It follows from Theorem 3.2 that $\|Rem_n\|_\lambda = O_p(a_n)$. Thus, we have

$$
\begin{aligned}
&\|\hat{\theta}_{n\lambda} - \theta_0^* - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{\mathcal{H}_{\mathbf{W}_i} - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\mathcal{H}_{\mathbf{W}_j}}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)}\}\{\, dN_i(t) - \mathcal{Y}_i(t)\eta_{\alpha_0}(\mathbf{W}_i)\, dt\}\|_2\\
=&O_p(a_n).
\end{aligned}
$$

Define

$$
\begin{aligned}
&Rem_n^h\\
=&\hat{\alpha}_{n\lambda}^h - \alpha_0^{*h} - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{\mathcal{R}_{\mathbf{W}_i}^h - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\mathcal{R}_{\mathbf{W}_j}^h}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)}\}\{\, dN_i(t) - \mathcal{Y}_i(t)\eta_{\alpha_0}(\mathbf{W}_i)\, dt\}.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
&\|Rem_n^h - h^{a+1/2}Rem_n\|_\lambda\\
\le&(1 - h^{a+1/2})\|\hat{\theta}_{n\lambda} - \theta_0^* - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{\mathcal{H}_{\mathbf{W}_i}\\
&- \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\mathcal{H}_{\mathbf{W}_j}}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)}\}\{\, dN_i(t) - \mathcal{Y}_i(t)\eta_{\alpha_0}(\mathbf{W}_i)\, dt\}\|_2\\
=&O_p(a_n).
\end{aligned}
$$

It follows from $a_n = o(n^{-1/2})$ that

$$
\begin{aligned}
\|Rem_n^h\|_\lambda \le& \|Rem_n^h - h^{a+1/2}Rem_n\|_\lambda + h^{a+1/2}\|Rem_n\|_\lambda\\
=& o_p(n^{-1/2}).
\end{aligned}
$$

Next, we use Rem_n^h to obtain the target joint limiting distribution. The idea is to employ the Cramér-Wold device. For any $\mathbf{u} = (\mathbf{z}^\top, t)^\top \in \mathbb{R}^p \times \mathbb{I}$, we obtain the limiting distribution of $n^{1/2}\mathbf{z}^\top(\hat{\theta}_{n\lambda} - \theta_0^*) + n^{1/2}h^{a+1/2}\{\hat{\beta}_{n\lambda}(t) - \beta_0^*(t)\}$. Note that this is equivalent to obtaining the asymptotic result of $n^{1/2} < \tilde{\mathcal{R}}_{\mathbf{u}}, \hat{\alpha}_{n\lambda}^h - \alpha_0^{*h} >_\lambda$. It follows from Theorem 3.2 that

$$n^{1/2}| < \tilde{\mathcal{R}}_{\mathbf{u}}, Rem_n^h >_\lambda | = O_p(n^{1/2}h^{-(a+1/2)}a_n).$$

Thus, we need to derive the limiting distribution of

$$n^{1/2} < \tilde{\mathcal{R}}_{\mathbf{u}}, \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{\mathcal{R}_{\mathbf{W}_i}^h - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)\mathcal{R}_{\mathbf{W}_j}^h}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(t)}\}\{ dN_i(t) - \mathcal{Y}_i(t)\eta_{\alpha_0}(\mathbf{W}_i)\, dt\} >_\lambda .$$

A direct calculation yields that

$$\begin{aligned}
& n^{1/2} < \tilde{\mathcal{R}}_{\mathbf{u}}, \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{\mathcal{R}_{\mathbf{W}_i}^h - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(s)\mathcal{R}_{\mathbf{W}_j}^h}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(s)}\}\{ dN_i(s) - \mathcal{Y}_i(s)\eta_{\alpha_0}(\mathbf{W}_i)\, ds\} >_\lambda \\
=& n^{-1/2}\sum_{i=1}^n \int_0^\tau \{z^\top H_{\mathbf{W}_i} + h^{a+1/2}T_{\mathbf{W}_i}(t) \\
& - \frac{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(s)[z^\top H_{\mathbf{W}_j} + h^{a+1/2}T_{\mathbf{W}_j}(t)]}{n^{-1}\sum_{j=1}^n \mathcal{Y}_j(s)}\}\, dM_i(s) \\
\equiv& \mathcal{U}_n.
\end{aligned}$$

A direct calculation yields that

$$\begin{aligned}
& \mathcal{K}(\mathbf{W}_i) \\
\equiv& \mathbf{z}^\top H_{\mathbf{W}_i} + h^{a+1/2}T_{\mathbf{W}_i}(t) \\
=& \mathbf{z}^\top\{\Sigma - \Omega + V(\mathbf{G}, W_\lambda\mathbf{G}^\top)\}^{-1}\{\mathbf{Z}_i - V(\mathbf{G}, \pi_{X_i})\} \\
& + h^{a+1/2}[\pi_{X_i}(t) - \boldsymbol{\omega}^\top(t)\{\Sigma - \Omega + V(\mathbf{G}, W_\lambda\mathbf{G}^\top)\}^{-1}\{\mathbf{Z}_i - V(\mathbf{G}, \pi_{X_i})\}] \\
=& h^{a+1/2}\pi_{X_i}(t) + \{\mathbf{z} - h^{a+1/2}\boldsymbol{\omega}(t)\}^\top\{\Sigma - \Omega + V(\mathbf{G}, W_\lambda\mathbf{G}^\top)\}^{-1}\{\mathbf{Z}_i - V(\mathbf{G}, \pi_{X_i})\}.
\end{aligned}$$

It follows from $\|K_{\mathbf{z}}\|_m \le h^{-a-1/2}$ that $\|\pi_X\|_m^2 \le \|X\|_{L^2}^2 h^{-2a-1}$. And it follows from

$V(\mathbf{G}, W_\lambda \mathbf{G}^\top) \to \mathbf{0}$ that $W_\lambda \mathbf{G} \to \mathbf{0}$. Thus, we have

$$
\begin{aligned}
|V(G_k, \pi_x)| =& | < G_k, \pi_x >_m - < (W_\lambda G_k), \pi_x >_m | \\
=& | \int_{\mathbb{I}} x(s) G_k(s)\, ds - \int_{\mathbb{I}} x(s)(W_\lambda G_k)(s)\, ds| \\
\asymp & | \int_{\mathbb{I}} x(s) G_k(s)\, ds| \\
\lesssim & \|x\|_{L_2}.
\end{aligned}
$$

Thus, we have $|\mathcal{K}(\mathbf{W}_i)| \lesssim \{\|X_i\|_{L_2} + \|\mathbf{Z}_i\|_2\}$ a.s. Therefore, we have

$$
\begin{aligned}
& \mathcal{U}_n \\
=& n^{-1/2} \sum_{i=1}^{n} \int_0^\tau \Big[\mathcal{K}(\mathbf{W}_i) - \frac{E\{\mathcal{Y}_j(s) \mathcal{K}(\mathbf{W}_j)\}}{E\{\mathcal{Y}_j(s)\}} \Big] \, dM_i(s) \\
& + O_p(n^{-1/2}) \\
=& n^{-1/2} \sum_{i=1}^{n} \int_0^\tau \Big[\mathcal{K}(\mathbf{W}_i) - \frac{E\{\mathcal{Y}_j(s) \mathcal{K}(\mathbf{W}_j)\}}{E\{\mathcal{Y}_j(s)\}} \Big] \, dM_i(s) + o_p(1) \\
\equiv & n^{-1/2} \sum_{i=1}^{n} \mathcal{U}_i + o_p(1).
\end{aligned}
$$

A direct calculation yields that

$$
\begin{aligned}
&Var[\mathcal{U}_i]\\
=&E\int_0^\tau \left[\mathcal{K}(\mathbf{W}_i) - \frac{E\{\mathcal{Y}_j(s)\mathcal{K}(\mathbf{W}_j)\}}{E\{\mathcal{Y}_j(s)\}}\right]^2 dN_i(s)\\
=&h^{2a+1}E\int_0^\tau \left[\pi_{X_i}(t) - \frac{E\{\mathcal{Y}_j(s)\pi_{X_j}(t)\}}{E\{\mathcal{Y}_j(s)\}}\right]^2 dN_i(s)\\
&+2h^{a+1/2}(\mathbf{z}-h^{a+1/2}\boldsymbol{\omega}(t))^\top\\
&\times\{\Sigma-\Omega+V(\mathbf{G},W_\lambda\mathbf{G}^\top)\}^{-1}E\int_0^\tau \left[\pi_{X_i}(t) - \frac{E\{\mathcal{Y}_j(s)\pi_{X_j}(t)\}}{E\{\mathcal{Y}_j(s)\}}\right]\\
&\times\left(\{\mathbf{Z}_i - V(\mathbf{G},\pi_{X_i})\} - \frac{E[\mathcal{Y}_j(s)\{Z_j - V(\mathbf{G},\pi_{X_j})\}]}{E\{\mathcal{Y}_j(s)\}}\right) dN_i(s)\\
&+\{\mathbf{z}-h^{a+1/2}\boldsymbol{\omega}(t)\}^\top\{\Sigma-\Omega+V(\mathbf{G},W_\lambda\mathbf{G}^\top)\}^{-1}E\int_0^\tau \Bigg(\{\mathbf{Z}_i - V(\mathbf{G},\pi_{X_i})\}\\
&-\frac{E[\mathcal{Y}_j(s)\{\mathbf{Z}_j - V(\mathbf{G},\pi_{X_j})\}]}{E\{\mathcal{Y}_j(s)\}}\Bigg)^{\otimes 2} dN_i(s)\{\Sigma-\Omega+V(\mathbf{G},W_\lambda\mathbf{G}^\top)\}^{-1}(\mathbf{z}-h^{a+1/2}\boldsymbol{\omega}(t)).
\end{aligned}
$$

Assume $h^{a+1/2}\boldsymbol{\omega}(t)\to-\boldsymbol{\gamma}_0$. Thus, as $\lambda\to 0$, we have

$$
\begin{aligned}
&Var[\mathcal{U}_i]\to\sigma_t^2+2(\mathbf{z}+\boldsymbol{\gamma}_0)^\top(\Sigma-\Omega)^{-1}\xi_0+(\mathbf{z}+\boldsymbol{\gamma}_0)^\top(\Sigma-\Omega)^{-1}B_0(\Sigma-\Omega)^{-1}(\mathbf{z}+\boldsymbol{\gamma}_0)\\
\equiv&(\mathbf{z}^\top,1)\Phi^*(\mathbf{z}^\top,1)^\top,
\end{aligned}
$$

with

$$
\Phi^*=\begin{bmatrix}(\Sigma-\Omega)^{-1}B_0(\Sigma-\Omega)^{-1} & (\Sigma-\Omega)^{-1}(\boldsymbol{\gamma}_0+\boldsymbol{\xi}_0)\\ (\boldsymbol{\gamma}_0+\boldsymbol{\xi}_0)^\top(\Sigma-\Omega)^{-1} & \sigma_t^2+2\boldsymbol{\gamma}_0^\top(\Sigma-\Omega)^{-1}\boldsymbol{\xi}_0+\boldsymbol{\gamma}_0^\top(\Sigma-\Omega)^{-1}\boldsymbol{\gamma}_0\end{bmatrix}.
$$

It follows from the Lindeberg's CLT that

$$
\begin{bmatrix}\sqrt{n}(\hat{\theta}_{n\lambda}-\theta_0^*)\\ \sqrt{nh}h^a\{\hat{\beta}_{n\lambda}(t)-\beta_0^*(t)\}\end{bmatrix}\to N(0,\Phi^*).
$$

From the fact that there exists $b\in((1+2a)/(2k),1]$ such that $V(G,h_j)$ satisfies that

$$
\sum_j \|V(G,h_j)\|_2^2\rho_j^b<\infty,
$$

we have that

$$
\begin{aligned}
&\|h^{a+1/2}\boldsymbol{\omega}(t)\|_{L_2}\\
=&h^{a+1/2}\|\sum_{j}^{\infty}V(\mathbf{G},h_j)/(1+\lambda\rho_j)h_j(t)\|_{L_2}\\
\leq&h^{a+1/2}\{\sum_{j}^{\infty}\|V(\mathbf{G},h_j)\|_2^2(1+\rho_j)^b\}^{1/2}\{\sum_{j=1}^{\infty}\{\frac{j^{2a}}{(1+\lambda\rho_j)^2(1+\rho_j)^b}\}\}^{1/2}\\
=&O(h^{a+1/2}).
\end{aligned}
$$

This implies that $\boldsymbol{\gamma}_0 = 0$. Thus, we have $\Phi^* \to \Phi$ as n goes to infinity. Furthermore, we have that

$$
\begin{aligned}
\| < \boldsymbol{\omega}, W_\lambda\beta_0 >_m \|_2 =& |\sum_{j=1}^{\infty} G_j V(\beta_0, h_j)\frac{\lambda\rho_j}{1+\lambda\rho_j}|\\
\leq& \sum_{j=1}^{\infty}\|G_j\|_2\frac{\lambda\rho_j}{1+\lambda\rho_j}\sum_{j=1}^{\infty}V(\beta_0,h_j)^2\frac{\lambda\rho_j}{1+\lambda\rho_j}\\
\lesssim& \lambda\sum_{j=1}^{\infty}\|G_j\|_2\rho_j^b\frac{\lambda\rho_j^{1-b}}{1+\lambda\rho_j}\\
\lesssim& \lambda^{1+b}.
\end{aligned}
$$

Thus, we have $\| < \boldsymbol{\omega}, W_\lambda\beta_0 >_m \|_2 = O_p(\lambda^{1+b}) = o_p(n^{-1/2})$, and so, $\sqrt{n}\{\theta_0^* - \theta_0\} \to 0$ and $\sqrt{n}h^{a+1/2}[\beta_0^*(t) - \beta_0(t) + \{W_\lambda(\beta_0)\}(t)] \to 0$. As $n^{1/2}h^{k(1+b)} = o(1)$, we have $nh^{4k} = o(1)$. Thus, if we define $\beta_0 = \sum_{j=1}^{\infty} b_j h_j$ with $\sum_{j=1}^{\infty} b_j^2\rho_j^2 < \infty$, we have

$$
\begin{aligned}
&|(W_\lambda\beta_0)(t)|\\
=&|\sum_{j=1}^{n}\frac{b_j\lambda\rho_j}{1+\lambda\rho_j}h_j(t)|\\
\leq&c_h\lambda\{\sum_{j=1}^{\infty}b_j^2\rho_j^2\}^{1/2}\{\sum_{j=1}\frac{j^{2a}}{(1+\lambda\rho_j)^2}\}^{1/2}\\
=&O(\lambda h^{-a-1/2}).
\end{aligned}
$$

Hence, this leads to

$$\sqrt{n}h^{a+1/2}\{W_\lambda(\beta_0)\}(t) = O(\sqrt{n}h^{2k}) = o(1).$$

Thus, the conclusion follows directly.

Proof of Theorem 3.4

By Theorem 3.2 and the proof of Theorem 3.3, we have

$$n^{1/2}h^{a+1/2}\sup_{s\in\mathbb{I}}|\hat{\beta}_{n,\lambda}(s) - \beta_0(s) - \tilde{\mathcal{S}}_n(\alpha_0)(s)| = o_p(1), \tag{3.21}$$

where

$$\begin{aligned}
\tilde{\mathcal{S}}_n(\alpha_0)(s) \equiv& \frac{1}{n}\int_0^\tau \sum_{i=1}^n \Big(T_{W_i}(s) - \frac{E[\mathcal{Y}(t)T_W(s)]}{E[\mathcal{Y}(t)]}\Big)\,dM_i(t) \\
=& \frac{1}{n}\int_{\mathbb{I}} K(s,u)\int_0^\tau \sum_{i=1}^n \Big(X_i(u) - \frac{E[\mathcal{Y}(t)X(u)]}{E[\mathcal{Y}(t)]}\Big)\,dM_i(t)\,du \\
&- \boldsymbol{\omega}(s)\{\Sigma - \Omega + V(\mathbf{G}, W_\lambda\mathbf{G}^\top)\}^{-1}\frac{1}{n}\sum_{i=1}^n\int_0^\tau \{Z_i - V(\mathbf{G}, \pi_{X_i})\}\,dM_i(t) \\
\equiv& \tilde{\mathcal{S}}_{n1}(\alpha_0)(s) - \tilde{\mathcal{S}}_{n2}(\alpha_0)(s).
\end{aligned}$$

(i) Denote $H_n(s) = \sqrt{nh}h^a\tilde{\mathcal{S}}_{n1}(\alpha_0)(s)$. Our first step is to show that $H_n(s)$ converges to the Gaussian process $\mathcal{G}(s)$ in the Hilbert Space $\mathcal{H}^{(m)}$ with the inner product $V(\cdot,\cdot)$, where $h_j, j = 1, 2, \cdots$, are the orthonormal basis. Direct calculation yields

$$\begin{aligned}
H_n(s) =& \frac{h^{1/2+a}}{\sqrt{n}}\int_{\mathbb{I}} K(s,u)\int_0^\tau \sum_{i=1}^n \Big(X_i(u) - \frac{E[\mathcal{Y}(t)X(u)]}{E[\mathcal{Y}(t)]}\Big)\,dM_i(t)\,du \\
=& \frac{1}{\sqrt{n}}\sum_{j=1}^\infty \int_{\mathbb{I}} \frac{h_j(u)h_j(s)h^{1/2+a}}{1+\lambda\rho_j}\int_0^\tau \sum_{i=1}^n \Big(X_i(u) - \frac{E[\mathcal{Y}(t)X(u)]}{E[\mathcal{Y}(t)]}\Big)\,dM_i(t)\,du.
\end{aligned}$$

It follows from Theorem 1.8.4 in van der Vaart and Wellner (1996) that to prove $H_n(t)$ converges to the Gaussian process $\mathcal{G}(t)$ in the Hilbert Space $\mathcal{H}^{(m)}$, we just need to prove $H_n(\cdot)$ is asymptotic finite-dimensional and $V(H_n, h_j)$ converges in distribution of $V(\mathcal{G}, h_j)$. It follows from the definition of $H_n(\cdot)$ that

$$\sum_{j=1}^{\infty} V(H_n, h_j)^2 = \sum_{j=1}^{\infty} \frac{h^{1+2a}}{(1+\lambda\rho_j)^2} \Bigg\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{\mathbb{I}} h_j(u) \int_0^{\tau} \Big(X_i(u) - \frac{E[\mathcal{Y}(t)X(u)]}{E[\mathcal{Y}(t)]} \Big) \, dM_i(t) \, du \Bigg\}^2 .$$

It is easy to verify that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{\mathbb{I}} h_j(u) \int_0^{\tau} \Big(X_i(u) - \frac{E[\mathcal{Y}(t)X(u)]}{E[\mathcal{Y}(t)]} \Big) \, dM_i(t) \, du$$

is asymptotic tight and bounded by cj^a. Besides, we have $\sum_j \frac{h^{1+2a} j^{2a}}{(1+\lambda\rho_j)^2} \asymp \int_0^{\infty} \frac{x^{2a}}{(1+x^{2k})^2} \, dx < \infty$. Then for every $\varepsilon > 0$, there exits J_0, which satisfies that $\sum_{j \geq J_0} \frac{h^{1+2a} j^{2a}}{(1+\lambda\rho_j)^2} < \varepsilon$. Thus, we have for any $\varepsilon > 0$,

$$\limsup_n P\Big(\sum_{j \geq J_0} V(H_n, h_j)^2 > \varepsilon\Big) \to 0.$$

Namely, H_n is asymptotic finite-dimensional.

Furthermore, it follows from the definitions of h_j and $V(\cdot, \cdot)$ that

$$\begin{aligned} V(H_n, h_j) &= \frac{h^{1/2+a}}{(1+\lambda\rho_j)} \Bigg\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{\mathbb{I}} h_j(u) \int_0^{\tau} \Big(X_i(u) - \frac{E[\mathcal{Y}(t)X(u)]}{E[\mathcal{Y}(t)]} \Big) \, dM_i(t) \, du \Bigg\} \\ &\xrightarrow{d} N\Bigg(0, \frac{h^{1+2a}}{(1+\lambda\rho_j)^2} E\Big(\int_0^{\tau} [\int_{\mathbb{I}} X(u) h_j(u) du - \frac{E\{\mathcal{Y}(t) \int_{\mathbb{I}} X(u) h_j(u) du\}}{E[\mathcal{Y}(t)]}]^2 \, dN(s) \Big) \Bigg) . \end{aligned} \tag{3.22}$$

Following Karhunen-Loéve theorem (Alexanderian (2015)), we can get that:

$$\mathcal{G}(t) = \sum_j a_j \eta_j \phi_j(t),$$

where $\eta_j, j = 1, 2, \cdots$ are i.i.d standard normal random variables, $\phi_j(t)$ are the standard orthogonal basis in the Hilbert space $L_2(\mathbb{I})$. It follows from the proof of

Karhunen-Loéve theorem that under our case, $\phi_j(t) \in \mathcal{H}^{(m)}(\mathbb{I})$. Direct calculations yield that

$$V(\mathcal{G}, h_j) = \sum_k a_k \eta_k V(\phi_k, h_j).$$

Then, $EV(\mathcal{G}, h_j) = 0$. It follows from Fubini's theorem that

$$EV(\mathcal{G}, h_j)^2 = \frac{h^{1+2a}}{(1+\lambda\rho_j)^2} E\Big(\int_0^\tau [\int_{\mathbb{I}} X(u)h_j(u)du - \frac{E\{\mathcal{Y}(t) \int_{\mathbb{I}} X(u)h_j(u)du\}}{E[\mathcal{Y}(t)]}]^2 \, dN(s)\Big).$$

Thus, we have $V(\mathcal{G}, h_j)$ follows

$$N\left(0, \frac{h^{1+2a}}{(1+\lambda\rho_j)^2} E\Big(\int_0^\tau [\int_{\mathbb{I}} X(u)h_j(u)du - \frac{E\{\mathcal{Y}(t) \int_{\mathbb{I}} X(u)h_j(u)du\}}{E[\mathcal{Y}(t)]}]^2 \, dN(s)\Big)\right).$$

It follows from equation (3.22) that $V(H_n, h_j)$ converges to $V(\mathcal{G}, h_j)$. The proof of the first step is completed.

(ii)The next step is to prove that $\sqrt{n}h^{1/2+a}\tilde{\mathcal{S}}_{n2}(\alpha_0)(s)$ uniformly in s converges to zero. It follows from the arguments in Cheng and Shang (2015) that $\sup_s |\boldsymbol{\omega}(s)| = O(1)$. Besides, since $h \to 0$, $\{\Sigma - \Omega + V(\mathbf{G}, W_\lambda \mathbf{G}^\top)\}^{-1} \to \{\Sigma - \Omega\}^{-1}$, $n^{-1/2}\sum_{i=1}^n \int_0^\tau \{Z_i - V(\mathbf{G}, \pi_{X_i})\} \, dM_i(t)$ is asymptotic to a normal distribution, we have $h^{a+1/2}\sqrt{n}\tilde{\mathcal{S}}_{n2}(\alpha_0)(s) \xrightarrow{d} 0$.

Theorem 3.4 directly follows from Slutsky's theorem, step (i) and step (ii).

Chapter 4

Conclusion and future work

This chapter draws conclusion on the thesis, and suggest some possible future research directions.

4.1 Conclusion

This thesis focuses on the development of semiparametric inference for the functional Cox model and the functional additive hazards model with right-censored data.

We propose a penalized partial likelihood approach for the estimation of model parameters for the functional Cox proportional hazards model and penalized pseudo-score function approach for the functional additive hazards model. We establish the asymptotic properties including the consistency, the convergence rate, and the limiting distribution of the proposed estimators. To this end, we investigate the joint Bahadur representation of finite-dimensional and infinite-dimensional estimators in the Sobolev space with a proper inner product.

One major contribution made to the study of the functional Cox model and the functional hazards model is that the asymptotic joint normality of the estimators of the functional and scalar coefficients is derived. Furthermore, the partial likelihood ratio test are developed and its optimality is shown under the functional Cox model. We also discover the Wilks phenomenon.

These two important issues are addressed in the previous research. Our new results will provide more insights and deeper understanding about effects of functional predictors on the hazard function of failure time. Simulation studies demonstrate that the proposed estimators perform well and the penalized partial likelihood ratio test has good power.

4.2 Future Work

A further interesting research is to explore other useful functional survival models such as the functional accelerated failure time models where a partial likelihood is unavailable. In addition, it is of interest to extend the analysis for other censoring schemes which often occur in medical studies. For instance, we can explore the functional Cox model, the functional additive hazards model, and the functional linear model in the presence of interval censoring.

It is also of interest to consider the semiparametric additive transformation model with current status data Cheng and Wang (2011):

$$H(U) = Z'\beta + \sum_{j=1}^{d} h_j(W_j) + \epsilon,$$

where $H(\cdot)$ is a monotone transformation function, $h_j(\cdot)$'s are smooth regression functions, and ϵ is a random variable with a known distribution $F(\cdot)$ with support $\mathbb{R}$. This model is a general transformation framework which covers a wide range of survival models and econometric models.

Bibliography

Aalen, O. O. (1989), "A linear regression model for the analysis of life times," *Statistics in Medicine*, 8, 907–925.

Alexanderian, A. (2015), "A Brief Note on the Karhunen-Loève Expansion," *ArXiv.org*.

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

Bahadur, R. R. (1966), "A Note on Quantiles in Large Samples," *The annals of mathematical statistics*, 37, 577–580.

Berkson, J. and Gage, R. P. (1952), "Survival curve for cancer patients following treatment," *Journal of the American Statistical Association*, 47, 501–515.

Bitouze, D., Laurent, B., and Massart, P. (1999), "A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator," *In Annales de l'IHP Probabilites et statistiques*, 35, 735–763.

Blum, J. and Susarla, V. (1980), "Maximal deviation theory of density and failure rate function estimates based on censored data," *Multivariate analysis*, 5, 213–222.

Böhmer, P. E. (1912), "Theorie der unabhängigen Wahrscheinlichkeiten," *In Rapports Memoires et Proces verbaux de Septieme Congres International dActuaires Amsterdam*, 2, 327–343.

Bose, A. (1998), "Bahadur Representation of M_m Estimates," *The Annals of Statistics*, 26, 771—777.

Breslow, N. and Crowley, J. (1974), "A large sample study of the life table and product limit estimates under random censorship," *The Annals of Statistics*, 2, 437–453.

Cai, T. and Yuan, M. (2012), "Minimax and adaptive prediction for functional linear regression," *Journal of the American Statistical Association*, 107, 1201–1216.

Carroll, R. J. (1978), "On Almost Sure Expansions for M-Estimates," *The Annals of Statistics*, 6, 314–318.

Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. R. (1987), "A new method of classifying prognostic comorbidity in longitudinal studies: development and validation," *Journal of Chronic Diseases and Management*, 40, 373–83.

Chen, K., Chen, K., Müller, H., and Wang, J. L. (2011), "Stringing high-dimensional data for functional analysis," *Journal of the American Statistical Association*, 106, 275–284.

Cheng, G. and Shang, Z. (2015), "Joint asymptotics for semi-nonparametric regression models with partially linear structure," *The Annals of Statistics*, 43, 1351–1390.

Cheng, G. and Wang, X. (2011), "Semiparametric Additive Transformation Model under Current Status Data," *Electronic Journal of Statistics*, 5, 1735–1764.

Cox, D. R. (1972), "Regression models and life-tables," *Journal of the Royal Statistical Society, Series B (Methodology)*, 34, 187–220.

Cox, D. R. (1975), "Partial likelihood," *Biometrika*, 62, 269–276.

Crambes, C., Kneip, A., and Sarda, P. (2009), "Smoothing splines estimators for functional linear regression," *The Annals of Statistics*, 37, 35–72.

Csörgö, S. and Horváth, L. (1980), "Random censorship from the left," *Studia Scientiarum Mathematicarum Hungarica*, 15, 397–491.

Cutler, S. J. and Ederer, F. (1958), "Maximum utilization of the life table method in analyzing survival," *Journal of chronic diseases*, 8, 699–712.

Dabrowska, D. M. (1989), "Uniform consistency of the kernel conditional Kaplan-Meier estimate," *The Annals of Statistics*, 17, 1157–1167.

Efron, B. (1967), "The two sample problem with censored data," *In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 4, 831–853.

Eubank, R. (1999), *Nonparametric Regression and Spline Smoothing*, vol. 157, New York; Basel: Marcel Dekker.

Földes, A., Rejtö, L., and Winter, B. (1981), "Strong consistency properties of nonparametric estimators for randomly censored data, II: Estimation of density and failure rate," *Periodica Mathematica Hungarica*, 12, 15–29.

Gehan, E. A. (1969), "Estimating survival functions from the life table," *Journal of chronic diseases*, 21, 629–644.

Gellar, J., Colantuoni, E., Needham, D., and Crainiceanu, C. (2014), "Variable-domain functional regression for modeling ICU data," *Journal of the American Statistical Association*, 109, 1425–1439.

Gellar, J., Colantuoni, E., Needham, D., and Crainiceanu, C. (2015), "Cox regression models with functional covariates for survival data," *Statistical Modelling*, 15, 256–278.

Gomez, G., Julià, O., Utzet, F., and Moeschberger, M. L. (1992), "Survival analysis for left censored data," in *Survival Analysis: State of the Art*, vol. 211, pp. 269–288, Springer.

Gómez, G., Julià, O., and Utzet, F. (1994), "Asymptotic properties of the left Kaplan-Meier estimator," *Communications in Statistics-Theory and Methods*, 23, 123–135.

Gu, C. (2002), *Smoothing Spline ANOVA Models*, New York: Springer-Verlag.

Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer-Verlag.

He, X. and Shao, Q.-M. (1996), "A General Bahadur Representation of M-Estimators and its Application to Linear Regression with Nonstochastic Designs," *The Annals of Statistics*, 24, 2608–2630.

Huffer, F. W. and Mckeague, I. W. (1991), "Weighted Least Squares Estimation for Aalen's Additive Risk Model," *Journal of the American Statistical Association*, 86, 114–129.

Ingster, Y. I. (1993), "Asymptotically minimax hypothesis testing for nonparametric alternatives II," *Mathematical Methods of Statistics*, 2, 249–268.

Jacobsen, M. (1984), "Maximum likelihood estimation in the multiplicative intensity model: a survey," *International Statistical Review*, 52, 193–207.

Johansen, S. (1983), "An extension of Cox's regression model," *International Statistical Review*, 51, 165–174.

Jong, D. P. (1987), "A central limit theorem for generalized quadratic forms," *Probability Theory and Related Fields*, 75, 261–277.

Kaplan, E. L. and Meier, P. (1958), "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, 53, 457–481.

Kim, C., Park, B. U., Kim, W., and Lim, C. (2003), "Bezier curve smoothing of the Kaplan-Meier estimator," *Annals of the Institute of Statistical Mathematics*, 55, 359–367.

Kong, D., Ibrahim, J. G., Lee, E., and Zhu, H. (2018), "FLCRM: functional linear cox regression model," *Biometrics*, 74, 109–117.

Kosorok, M. (2008), “Introduction to Empirical Processes and Semiparametric Inference Springer: New York,” .

Kulich, M. and Lin, D. (2000), “Additive hazards regression for case-cohort studies,” *Biometrika*, 87, 73–87.

Li, K. and Luo, S. (2017), “Functional joint model for longitudinal and time-to-event data: an application to alzheimer’s disease,” *Statistics in Medicine*, 36, 3560–3572.

Lin, D. and Ying, Z. (1994), “Semiparametric analysis of the additive risk model,” *Biometrika*, 81, 61–71.

Lin, D., Oakes, D., and Ying, Z. (1998), “Additive hazards regression with current status data,” *Biometrika*, 85, 289–298.

Lin, D. Y., Wei, L.-J., and Ying, Z. (1993), “Checking the Cox model with cumulative sums of martingale-based residuals,” *Biometrika*, 80, 557–572.

Martinussen, T. and Scheike, T. H. (2002), “Efficient estimation in additive hazards regression with current status data,” *Biometrika*, 89, 649–658.

Mckeague, I. W. and Sasieni, P. D. (1994), “A Partly Parametric Additive Risk Model,” *Biometrika*, 81, 501–514.

Needham, D. M., Dennison, C. R., Dowdy, D. W.and Mendez-Tellez, P. A., Ciesla, N., Desai, S. V., Sevransky, J., Shanholtz, C., Scharfstein, D., Herridge, M. S., and Pronovost, P. J. (2006), “Study protocol: The Improving Care of Acute Lung Injury Patients (ICAP) study,” *Critical Care*, 10, R9–R9.

O’Neill, T. J. (1986), “Inconsistency of the misspecified proportional hazards model,” *Statistics and Probability Letters*, 4, 219–222.

Pearce, N. and Wand, M. (2006), “Penalised Splines and Reproducing Kernel Methods,” *The American Statistician*, 60, 233–240.

Qu, S., Wang, J. L., and Wang, X. (2016), “Optimal estimation for the functional Cox model,” *The Annals of Statistics*, 44, 1708–1738.

Ramsay, S. and Silverman, B. W. (2005), *Functional data analysis*, Springer Series in Statistics, Springer, New York, NY, 2 edn.

Shang, Z. and Cheng, G. (2015), “Nonparametric inference in generalized functional linear models,” *The Annals of Statistics*, 43, 1742–1773.

Tsiatis, A. (1981), “A large sample study of Cox’s regression model,” *The Annals of Statistics*, 9, 93–108.

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak convergence and empirical processes : with applications to statistics*, Springer.

Verweij, P. and Van Houwelingen, H. (1993), "Cross-validation in survival analysis," *Statistics in Medicine*, 12, 2305–2314.

Wahba, G. (1990), *Spline Models for Observational Data*, vol. 59.

Ware, J. H. and Demets, D. L. (1976), "Reanalysis of some baboon descent data," *Biometrics*, 32, 459–463.

Wellner, J. A. (2007), "On an exponential bound for the Kaplan-Meier estimator," *Lifetime data analysis*, 13, 481–496.

Whittemore, A. S. and Keller, J. B. (1986), "Survivals estimation using splines," *Biometrics*, 42, 495–506.

Yuan, M. T. and Cai, T. (2010), "A reproducing kernel Hilbert space approach to functional linear regression," *The Annals of Statistics*, 38, 3412–3444.