

## Research Article

# New Inference Procedures for Semiparametric Varying-Coefficient Partially Linear Cox Models

Yunbei Ma<sup>1</sup> and Xuan Luo<sup>2</sup>

<sup>1</sup> School of Statistics and Research Center of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China

<sup>2</sup> Information Engineering University, Zhengzhou, Henan 450001, China

Correspondence should be addressed to Yunbei Ma; myb@swufe.edu.cn

Received 14 October 2013; Accepted 19 March 2014; Published 25 May 2014

Academic Editor: Jinyun Yuan

Copyright © 2014 Y. Ma and X. Luo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In biomedical research, one major objective is to identify risk factors and study their risk impacts, as this identification can help clinicians to both properly make a decision and increase efficiency of treatments and resource allocation. A two-step penalized-based procedure is proposed to select linear regression coefficients for linear components and to identify significant nonparametric varying-coefficient functions for semiparametric varying-coefficient partially linear Cox models. It is shown that the penalized-based resulting estimators of the linear regression coefficients are asymptotically normal and have oracle properties, and the resulting estimators of the varying-coefficient functions have optimal convergence rates. A simulation study and an empirical example are presented for illustration.

## 1. Introduction

To balance model flexibility and specificity, a variety of semiparametric models have been proposed in survival analysis. For example, Huang [1] studied partially linear Cox models and Cai et al. [2] and Fan et al. [3] studied the Cox proportional hazard model with varying coefficients. Tian et al. [4] proposed an estimation procedure for the Cox model with time-varying coefficients. Perhaps the most appropriate model is the semiparametric varying-coefficient model; its merit includes easy interpretation, flexible structure, potential interaction between covariates, and dimension reduction of nonparametric components models. Recently, a lot of effort has been made in this direction. Cai et al. [2] considered the Cox proportional hazard model with a semiparametric varying-coefficient structure. Yin et al. [5] proposed the semiparametric additive hazard model with varying coefficients.

In biomedical research and clinical trials, one major objective is to identify risk factors and study their risk impacts, as this identification can help clinicians to properly

make a decision and increase efficiency of treatment and resource allocation. This is essentially a variable selection procedure in spirit. When the outcomes are censored, selection of significant risk factors becomes more complicated and challenging than with the case where all outcomes are complete. However, important progress has been made in recent literature. For instance, Fan and Li [6] extended the nonconcave penalized likelihood approach proposed by Fan and Li [7] to the Cox proportional hazards model and the Cox proportional hazards frailty model. Johnson et al. [8, 9] proposed procedures for selecting variables in semiparametric linear regression models for censored data. However, these papers did not consider variable selection with functional coefficients; while Wang et al. [10] extended the application of penalized likelihood to the varying-coefficient setting and studied the asymptotic distributions of the estimators, they only focused on the SCAD penalty for completely observed data. Du et al. [11] proposed a penalized variable selection procedure for Cox models with a semiparametric relative risk.

In this paper, we study variable selection for both regression parameters and varying coefficients in the semiparametric varying-coefficient additive hazards model:

$$\lambda(t | Z, V, W) = \lambda_0(t) + \beta^T(W(t))Z(t) + \alpha^T V(t) + g(W(t)), \quad (1)$$

where  $V(t)$  is a vector of covariates with a linear effect on the logarithm of the hazard function  $\lambda(t)$ ,  $W(t)$  is the main exposure variable of interest whose effect on the logarithm of the hazard might be nonlinear, and  $Z(\cdot) = (Z_1(\cdot), \dots, Z_p(\cdot))^T$  is a vector of covariates that may interact with the exposure covariate  $W(\cdot)$ .  $\lambda_0(\cdot)$  is the baseline hazard ratio function and  $g(\cdot)$  is an unspecified smooth function.  $\alpha$  is a  $d$ -vector of constant parameters. Let  $g(0) = 0$  to assure that the model is identifiable. This model covers generally used basic semiparametric model settings. For instance, if  $\beta(\cdot) \equiv 0$  and  $g(\cdot) \equiv 0$ , model (1) reduces to Lin and Ying's additive hazard model [12] and Aalen's additive hazard model when the baseline hazard function  $\lambda_0(t)$  also equals to zero [13, 14]. Furthermore, when the exposure covariate  $W(\cdot)$  is time  $t$ , then model (1) reduces to the partly parametric additive hazard model if  $\lambda_0(t) \equiv g(t) \equiv 0$  [15, 16] and reduces to the time-varying coefficient additive hazard model [17].

To select significant elements of  $\alpha$  and coefficient functions  $\beta(\cdot)$ , we require an estimation procedure for the regression parameters  $\alpha$  and  $\beta(\cdot)$ . However, this poses a challenge because it is impossible to simultaneously get the root- $n$  consistent penalized estimators for the nonzero components of  $\alpha$  by penalizing  $\beta(\cdot)$  and  $\alpha$ . To achieve this goal, we propose the following two-step estimation procedure for selecting the relevant variables. In Step 1, we estimate  $\alpha$  by using the profile penalized least square function after locally approximating the nonparametric functions  $\beta(\cdot)$  and  $g(\cdot)$ . In Step 2, given the penalized estimator of  $\alpha$  in Step 1, we develop a penalized least square function for  $\beta(\cdot)$  by using the basis expansion. We will demonstrate that the proposed estimation procedures have oracle properties, and all penalized estimators of nonzero components can achieve their optimal convergence rate.

This paper is organized as follows. Section 2 introduces the SCAD-based variable selection procedure for the parametric components and establishes the asymptotic normality for the resulting penalized estimators. Section 3 discusses a variable selection procedure for the coefficient functions by approximating these functions by using the B-spline approach. The simulation studies and an application of the proposed methods in a real data example are included in Sections 4 and 5, respectively. The technical lemmas and proofs of the main results are given in the appendix.

## 2. Penalized Least Squares Based Variable Selection for Parametric Components

Let  $T_i$  denote the potential failure time, let  $C_i$  denote the potential censoring time, and let  $X_i = \min(T_i, C_i)$  denote the observed time for the  $i$ th individual,  $i = 1, \dots, n$ . Let  $\Delta_i$  be an indicator which equals 1 if  $X_i$  is a failure time and 0 otherwise. Let  $\mathcal{F}_{t,i}$  represent the failure, censoring, and

covariate information up to time  $t$  for the  $i$ th individual. The observed data structure is  $\{X_i, \Delta_i, Z_i(t), V_i(t), W_i(t), i = 1, \dots, n\}$ . Assume that  $T$  and  $C$  are conditionally independent given covariates and that the observation period is  $[0, \tau]$ , where  $\tau$  is a constant denoting the time for the end of the study.

Let  $N_i(t) = I(X_i \leq t, \Delta_i = 1)$  denote the counting process corresponding to  $T_i$  and let  $Y_i(t) = I(X_i \geq t)$ . Let the filtration  $\{\mathcal{F}_t : t \in [0, \tau]\}$  be the history up to time  $t$ ; that is,  $\mathcal{F}_t = \sigma\{X_i \leq u, Z_i(u), V_i(u), W_i(u), Y_i(u), \Delta_i, 0 \leq u \leq t, i = 1, \dots, n\}$ . Write  $M_i(t) = N_i(t) - \int_0^t Y_i(u) \lambda_i(u) du$ . Then  $M_i(t)$  is a martingale with respect to  $\mathcal{F}_t$ . For ease of presentation, we drop the dependence of covariates on time. The methods and proofs in this paper are applicable to external time-dependent covariates [18].

Here we use the techniques of local linear fitting [19]. Suppose that  $\beta(\cdot)$  and  $g(\cdot)$  are smooth enough to allow Taylor's expansion as follows: for each  $w_0 \in \mathcal{W}$  which is the support of  $W$  and for  $w$  in a neighborhood of  $w_0$ ,

$$\begin{aligned} \beta(w) &\approx \beta(w_0) + \beta'(w_0)(w - w_0), \\ g(w) &\approx g(w_0) + g'(w_0)(w - w_0). \end{aligned} \quad (2)$$

Then we can approximate the hazard ratio function given in (1) by

$$\lambda(t | Z_i, V_i, W_i) \approx \lambda_0^*(t, w_0) + \eta^T(w_0) Z_i^*(w_0) + \alpha^T V_i, \quad (3)$$

where  $\eta(w_0) = (\beta^T(w_0), (\beta'(w_0))^T, g'(w_0))^T$ ,  $Z_i^*(w_0) = (Z_i^T, Z_i^T(W_i - w_0), (W_i - w_0)^T)^T$ , and  $\lambda_0^*(t, w_0) = \lambda_0(t) + g(w_0)$ .

Let  $\mathbf{H}$  be a  $2p+1$  diagonal matrix with the first  $p$  diagonal elements 1 and the rest  $h$ . If  $\alpha$  is given, then by using the counting-process notation, similarly to [5], we can obtain an estimator of  $\eta(w_0)$  at each  $w_0 \in \mathcal{W}$  by solving the estimating equations whose objective function is given as follows:

$$\begin{aligned} U_n(\eta, w_0) &= \sum_{i=1}^n \int_0^\tau K_h(W_i - w_0) \mathbf{H}^{-1} \{Z_i^*(w_0) - \bar{Z}(t, w_0)\} \\ &\quad \times [dN_i(t) - Y_i(t) \\ &\quad \times \{\eta^T(w_0) Z_i^*(w_0) + \alpha^T V_i\} du], \end{aligned} \quad (4)$$

where

$$\bar{Z}(t, w_0) = \frac{\sum_{i=1}^n K_h(W_i - w_0) Y_i(t) Z_i^*(w_0)}{\sum_{i=1}^n K_h(W_i - w_0) Y_i(t)}, \quad (5)$$

$K_h(\cdot) = K(\cdot/h)/h$  with  $K(\cdot)$  being a kernel function and  $h$  being a bandwidth, and  $\tau = \infty$ . To avoid the technicality of tail problems, only data up to a finite time point  $\tau$  are frequently used.

Denote the solution of  $U_n(\boldsymbol{\eta}, w_0) = 0$  by  $\hat{\boldsymbol{\eta}}(w_0, \boldsymbol{\alpha})$ , which can be expressed as follows:

$$\begin{aligned} \hat{\boldsymbol{\eta}}(w_0, \boldsymbol{\alpha}) &= \left[ \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(W_i - w_0) \mathbf{H}^{-1} \right. \\ &\quad \times \left. \{Z_i^*(w_0) - \bar{Z}(t, w_0)\}^{\otimes 2} Y_i(t) dt \right]^{-1} \\ &\quad \times \left[ \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(W_i - w_0) \mathbf{H}^{-1} \right. \\ &\quad \times \left. \{Z_i^*(w_0) - \bar{Z}(t, w_0)\} \{dN_i(t) - Y_i(t) V_i^T \boldsymbol{\alpha} dt\} \right] \\ &\stackrel{\text{def}}{=} M_{n1}^{-1}(w_0) \{S_{n1}(w_0) - S_{n2}(w_0) \boldsymbol{\alpha}\}. \end{aligned} \quad (6)$$

Here and below,  $\mathbf{a}^{\otimes k} = 1, \mathbf{a}, \mathbf{a}\mathbf{a}^T$  for any vector  $\mathbf{a}$  and for  $k = 0, 1, 2$ .

As a consequence, we can estimate  $\boldsymbol{\beta}(\cdot)$  and  $g(\cdot)$  presuming that  $\boldsymbol{\alpha}$  is known:

$$\begin{aligned} \hat{\boldsymbol{\beta}}(W_i, \boldsymbol{\alpha}) &= (I_p, 0_{p \times p}, 0_p) M_{n1}^{-1}(W_i) \{S_{n1}(W_i) - S_{n2}(W_i) \boldsymbol{\alpha}\} \\ &\equiv T_{n1}(W_i) - T_{n2}(W_i) \boldsymbol{\alpha}, \\ \hat{g}(W_i, \boldsymbol{\alpha}) &= \int_0^{W_i} (0_{p \times p}, 0_{p \times p}, 1_p) M_{n1}^{-1}(w_0) \\ &\quad \times \{S_{n1}(w_0) - S_{n2}(w_0) \boldsymbol{\alpha}\} dw_0 \\ &\equiv T_{n3}(W_i) - T_{n4}(W_i) \boldsymbol{\alpha}. \end{aligned} \quad (7)$$

**2.1. Variable Selection.** Notice that if  $\boldsymbol{\alpha}$  is given, then from (7) we can estimate  $\Lambda_0(t, W_i) = \int_0^t \lambda_0(u, W_i) du$  by

$$\begin{aligned} \hat{\Lambda}_0(t, W_i, \boldsymbol{\alpha}) &= \sum_{i=1}^n \int_0^t \frac{[dN_i(u) - Y_i(u) \{Z_i^T T_{n1}(W_i) + T_{n3}(W_i)\} du]}{\sum_{i=1}^n Y_i(u)} \\ &\quad - \sum_{i=1}^n \int_0^t \frac{Y_i(u) [V_i^T - Z_i^T T_{n2}(W_i) - T_{n4}(W_i)] \boldsymbol{\alpha} du}{\sum_{i=1}^n Y_i(u)} \\ &\stackrel{\text{def}}{=} \sum_{i=1}^n \int_0^t \frac{\sum_{i=1}^n [dN_i(u) - Y_i(u) \{G_{ni} + V_i^*(W_i) \boldsymbol{\alpha}\} du]}{\sum_{i=1}^n Y_i(u)}. \end{aligned} \quad (8)$$

Given the observed data and based on (7) and (8),  $\boldsymbol{\beta}(\cdot)$ ,  $g(\cdot)$ , and the cumulative function  $\Lambda_0(\cdot)$  can be approximately

expressed as functions of  $\boldsymbol{\alpha}$ . These expressions inspire us to estimate  $\boldsymbol{\alpha}$  by minimizing the following objective function subject to  $\boldsymbol{\alpha}$ :

$$\begin{aligned} L_n(\boldsymbol{\alpha}) &= \frac{1}{2} \sum_{i=1}^n \left( \int_0^\tau [dN_i(t) - Y_i(t) d\hat{\Lambda}_0(t, W_i, \boldsymbol{\alpha}) \right. \\ &\quad \left. - Y_i(t) \{Z_i^T \hat{\boldsymbol{\beta}}(W_i, \boldsymbol{\alpha}) + \hat{g}(W_i, \boldsymbol{\alpha}) + V_i^T \boldsymbol{\alpha}\} dt \right]^2. \end{aligned} \quad (9)$$

Accordingly, our penalized objective function is defined as follows:

$$L_n^*(\boldsymbol{\alpha}) = L_n(\boldsymbol{\alpha}) + n \sum_{j=1}^d p_\lambda(|\alpha_j|). \quad (10)$$

Thus, minimizing  $L_n^*(\boldsymbol{\alpha})$  with respect to  $\boldsymbol{\alpha}$  results in a penalized least squares estimator of  $\boldsymbol{\alpha}$ .

Let  $a_n = \max\{p'_{\lambda_n}(|\alpha_{0j}|) : \alpha_{0j} \neq 0\}$  and  $b_n = \max\{p''_{\lambda_n}(|\alpha_{0j}|) : \alpha_{0j} \neq 0\}$ . Now we establish the oracle properties for the resulting estimators. Let  $\boldsymbol{\alpha}_0 = (\boldsymbol{\alpha}_{10}^T, \boldsymbol{\alpha}_{20}^T)^T$  be the true value of  $\boldsymbol{\alpha}$ , the first subvector  $\boldsymbol{\alpha}_{10}^T$  of length  $s$ , containing all nonzero elements of  $\boldsymbol{\alpha}$  and  $\boldsymbol{\alpha}_{20}^T = \mathbf{0}_{d-s}$ .

We introduce an additional notation as follows. Let  $\bar{Z} = (Z^T, 0_p^T, 0)^T$ . For  $k = 0, 1$ , let  $\mu_k = \int x^k K(x) dx$ ,  $\nu_k = \int x^k K^2(x) dx$ ,  $\rho_k(u, z, v, w_0) = P(X \geq u | Z = z, V = v, W = w_0) \lambda^k(t | z, v, w_0)$ . Denote

$$\begin{aligned} \varphi_{k,l}(t, w_0) &= f(w_0) E \left\{ \rho_0(t, Z, v, w_0) \bar{Z}^{\otimes k} (V^{\otimes l})^T | W = w_0 \right\}, \\ k &= 0, 1, \quad l = 0, 1, \end{aligned}$$

$$\begin{aligned} \varphi_{2,0}(t, w_0) &= f(w_0) E \left\{ \rho_0(t, Z, v, w_0) \right. \\ &\quad \times \left. \begin{pmatrix} Z^{\otimes 2} & 0_{p \times p} & 0_p \\ 0_{p \times p} & \mu_2 Z^{\otimes 2} & Z \mu_2 \\ 0_p^T & Z^T \mu_2 & \mu_2 \end{pmatrix} | W = w_0 \right\}, \end{aligned} \quad (11)$$

where  $f(\cdot)$  is the density of  $W$ . Denote  $t_1(w) = (I_p, 0_{p \times p}, 0_p) m_1^{-1}(w) m_2(w)$  and  $t_2(w) = (0_{p \times p}, 0_{p \times p}, 1_p) \int_0^w m_1^{-1}(w_0) m_2(w_0) dw_0$ , where

$$\begin{aligned} m_1(w_0) &= \int_0^\tau \left\{ \varphi_{2,0}(t, w_0) - \frac{\varphi_{1,0}^{\otimes 2}(t, w_0)}{\varphi_{0,0}(t, w_0)} \right\} dt, \\ m_2(w_0) &= \int_0^\tau \left\{ \varphi_{1,1}(t, w_0) - \frac{\varphi_{1,0}(t, w_0)}{\varphi_{0,0}(t, w_0)} \varphi_{0,1}(t, w_0) \right\} dt. \end{aligned} \quad (12)$$

For  $k = 0, 1, 2, l = 0, 1$ , let  $\kappa_{kl}(t) = E[\rho_l(t, Z, V, W)\{V - Zt_1^T(W) - t_2^T(W)\}^{\otimes k}]$ , and  $\tilde{V}_i(W_i) = V_i^T - Z_i^T t_1(W_i) - t_2(W_i)$ ,  $i = 1, \dots, n$ .  $\|\mathbf{a}\| = (\mathbf{a}^T \mathbf{a})^{1/2}$  for any vector  $\mathbf{a}$ .

**Theorem 1.** *If Assumptions (A.i)–(A.vi) are satisfied and  $b_n \rightarrow 0$  as  $n \rightarrow \infty$ , then there exist local minimizers  $\hat{\alpha}$  of  $L_n^*(\alpha)$  such that  $\|\hat{\alpha} - \alpha_0\| = O_p(n^{-1/2} + a_n)$ .*

**Remark 2.** Theorem 1 indicates that by choosing a proper  $\lambda_n$ , which leads to  $a_n = O(n^{-1/2})$ , there exists a root- $n$  consistent penalized least squares estimator of  $\alpha$ .

**Theorem 3.** *Assume that*

$$\infty > \liminf_{n \rightarrow \infty} \liminf_{\theta \rightarrow 0^+} \frac{p'_{\lambda_n}(\theta)}{\lambda_n} > 0. \quad (13)$$

*Under Assumptions (A.i)–(A.vi), if  $\lambda_n \rightarrow 0$  and  $\sqrt{n}\lambda_n \rightarrow \infty$  as  $n \rightarrow \infty$ , then the local root- $n$  consistent local minimizer  $\hat{\alpha} = (\hat{\alpha}_1^T, \hat{\alpha}_2^T)^T$  in Theorem 1 with probability tending to 1 satisfies*

- (1)  $\hat{\alpha}_2 = \mathbf{0}$ ; and
- (2) asymptotic normality

$$\sqrt{n}(\mathbf{I}_1(\tau) + \Gamma) \times \{\hat{\alpha}_1 - \alpha_{10} - (\mathbf{I}_1(\tau) + \Gamma)^{-1} \mathbf{b}\} \xrightarrow{D} N(0, \Sigma_1(\tau)), \quad (14)$$

where  $\mathbf{I}_1(\tau)$  and  $\Sigma_1(\tau)$  are the first  $s \times s$  submatrix of  $\mathbf{I}(\tau)n$  and  $\Sigma(\tau)$ , respectively, and

$$\mathbf{b} = (p'_{\lambda_n}(|\alpha_{10}|) \text{sgn}(\alpha_{10}), \dots, p'_{\lambda_n}(|\alpha_{s0}|) \text{sgn}(\alpha_{s0}))^T, \quad (15)$$

$$\Gamma = \text{diag}(p''_{\lambda_n}(|\alpha_{10}|), \dots, p''_{\lambda_n}(|\alpha_{s0}|)).$$

### 3. Variable Selection for the Varying Coefficients

When some variables of  $Z$  are not relevant in the additive hazard regression model, the corresponding functional coefficients are zero. In this section, by using basis function expansion techniques, we try to estimate those zero functional coefficients as identically zero via nonconcave penalty functions.

For a simplified expression, we denote  $g(\cdot)$  as  $\beta_0(\cdot)$  and  $Z_0 \equiv 1$ . Hence, we can rewrite the additive hazard regression model (1) as

$$\lambda(t | Z_i^*, V_i, W_i) = \lambda_0(t) + \beta^{*T}(W_i) Z_i^*(t) + \alpha^T V_i(t), \quad (16)$$

where  $Z_i^* = (1, Z_i^T)^T$  and  $\beta^*(W_i) = (\beta_0(W_i), \beta^T(W_i))^T$ . Assume  $\alpha^*$  is a given root- $n$  consistent estimator of  $\alpha$ . From the arguments in Section 2, such an  $\alpha^*$  is available.

**3.1. Penalized Function.** We approximate each element of  $\beta^*(w)$  by the basis expansion; that is,  $\beta_k(w) \approx \sum_{l=1}^{L_k} \gamma_{kl} B_{kl}(w)$  for  $k = 0, 1, \dots, p$ . This approximation indicates that the hazard function given in (16) can be approximated as follows:

$$\lambda(t | Z_i^*, V_i, W_i) \approx \lambda_0(t) + \sum_{k=0}^p Z_{ik} \left\{ \sum_{l=1}^{L_k} \gamma_{kl} B_{kl}(w) \right\} + \alpha^T V_i. \quad (17)$$

Let  $\gamma_k = (\gamma_{k1}, \dots, \gamma_{kL_k})^T$  and  $\gamma = (\gamma_0^T, \gamma_1^T, \dots, \gamma_p^T)^T$ . Write

$$\mathbf{B}(W_i) = \begin{pmatrix} B_{01}(W_i) \cdots B_{0L_0}(W_i) & 0 \cdots 0 & \cdots & 0 \cdots 0 \\ 0 \cdots 0 & B_{11}(W_i) \cdots B_{1L_1}(W_i) & \vdots & 0 \cdots 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 \cdots 0 & 0 \cdots 0 & \cdots & B_{p1}(W_i) \cdots B_{pL_p}(W_i) \end{pmatrix}, \quad (18)$$

and  $\mathbf{U}_i(W_i) = Z_i^{*T} \mathbf{B}(W_i)$ . Then, by using similar arguments as in Section 2, we estimate the basic hazard function  $\Lambda_0(t)$  by

$$\tilde{\Lambda}_0(t) = \sum_{i=1}^n \int_0^T \frac{dN_i(u) - Y_i(u) \{ \mathbf{U}_i(W_i) \gamma + \alpha^{*T} V_i \} du}{\sum_{i=1}^n Y_i(u)} \quad (19)$$

and estimate  $\gamma$  by minimizing the following least square function with respect to  $\gamma$ :

$$\begin{aligned} \tilde{L}_n(\gamma) &= \frac{1}{2} \sum_{i=1}^n \int_0^T [dN_i(t) - Y_i(t) \\ &\quad \times \{ \tilde{\lambda}_0(t) + \mathbf{U}_i(W_i) \gamma + \alpha^{*T} V_i \} dt]^2. \end{aligned} \quad (20)$$

Suppose that  $\beta_k = 0$ ,  $k = q+1, \dots, p$ , as mentioned before. We are trying to correctly identify these zero functional coefficients via nonconcave penalty functions, that is, via

minimizing the following penalized least square function with respect to  $\gamma$ :

$$PL_n(\gamma) = \tilde{L}_n(\gamma) + n \sum_{k=0}^p p_\lambda(\|\gamma_k\|_k), \quad (21)$$

where  $\|\gamma_k\|_k^2 = \gamma_k^T \mathbf{R}_k \gamma_k$  and  $\mathbf{R}_k = (r_{kij})_{L_k \times L_k}$  with  $r_{kij} = \int B_{ki}(w) B_{kj}(w) dw$ .

Let  $\hat{\gamma}$  be the minimizer of penalized least square function (21) and define  $\hat{\beta}^*(w) = \mathbf{B}(w)\hat{\gamma}$ .

*Remark 4.* Various basis systems, including Fourier basis, polynomial basis, and B-spline basis, can be used in the basis expansion. We focus on the B-spline basis and examine asymptotic properties of  $\hat{\beta}^*(w)$ .

**3.2. Asymptotic Properties.** Define  $\bar{\mathbf{U}}(t) = \sum_{i=1}^n Y_i(t) \mathbf{U}_i(W_i) / \sum_{i=1}^n Y_i(t)$  and let  $\tilde{\gamma}$  be the minimizer of  $\tilde{L}_n(\gamma)$  in (20). Hence, we can obtain  $\tilde{\gamma}$  by resolving

$$\begin{aligned} \bar{\mathbf{U}}_n(\gamma) &= \sum_{i=1}^n \int_0^\tau \{ \mathbf{U}_i(W_i) - \bar{\mathbf{U}}(t) \} \\ &\quad \times [dN_i(t) - Y_i(t) \{ \mathbf{U}_i^T(W_i) \gamma + \alpha^{*T} V_i \}] dt = 0. \end{aligned} \quad (22)$$

$\bar{\mathbf{U}}_n(\gamma)$  is the derivative of  $\tilde{L}_n(\gamma)$  with respect to  $\gamma$ , the usual score function of  $\gamma$  without the penalty.

For any square integrable functions  $b(w)$  and  $c(w)$  on  $\mathcal{W}$ , let  $\|b(w)\|_{L_2}$  denote the  $L_2$  norm of  $b(w)$  on  $\mathcal{W}$ , and define  $\|b(w) - c(w)\|_{L_\infty} = \sup_{w \in \mathcal{W}} |b(w) - c(w)|$  as the  $L_\infty$  distance between  $b(w)$  and  $c(w)$ . Let  $\mathcal{G}_k$  denote all functions that have the form  $\sum_{l=1}^{L_k} \gamma_{kl} B_{kl}(w)$ ,  $k = 0, 1, \dots, p$  for B-spline base.  $L_{\max} = \max_{0 \leq k \leq p} L_k$  and  $\rho_n = \max_{0 \leq k \leq p} \inf_{c \in \mathcal{G}_k} \|\beta_k - c\|_{L_\infty}$ . Denote  $\tilde{\beta}^*(w) = \mathbf{B}(w)\tilde{\gamma}$ .

Let  $a_n^* = \max\{p'_{\lambda_n}(\|\beta_k\|_{L_2}) : \|\beta_k\|_{L_2} \neq 0\}$  and  $b_n^* = \max\{p''_{\lambda_n}(\|\beta_k\|_{L_2}) : \|\beta_k\|_{L_2} \neq 0\}$ .

**Theorem 5.** Suppose that (13) and Assumptions (A.i) and (A.vii)–(A.iv) hold. If  $\lambda_n \rightarrow 0$  and  $\lambda_n / \max\{\rho_n, a_n^*, (L_{\max}/n)^{1/2}\} \rightarrow \infty$  as  $n \rightarrow \infty$ , then one has

- (a)  $\hat{\beta}_k = 0$ ,  $k = q + 1, \dots, p$ , with probability approaching 1; and
- (b)  $\|\hat{\beta}_k - \beta_k\|_{L_2} = O_p(\max\{\rho_n, a_n^*, (L_{\max}/n)^{1/2}\})$ ,  $k = 0, \dots, q$ .

*Remark 6.* Let  $\mathcal{G}_k$  be a space of splines with a degree no less than 1 and with  $L_k$  equally spaced interior knots, where  $L_k \asymp n^{1/5}$ ,  $k = 0, 1, \dots, p$ . Note that  $\rho_n = O(L_{\max}^{-2})$  [20, Theorem 6.21]. Hence, if  $a_n^* = o_p(\rho_n)$ , Theorem 5 implies that  $\|\hat{\beta}_k - \beta_k\|_{L_2} = O_p(n^{-2/5})$ ,  $k = 0, \dots, q$ .

*Remark 7.* Suppose that  $L_k \asymp n^{1/5}$ ,  $k = 0, 1, \dots, p$ . Because  $a_n^* = \max\{p'_{\lambda_n}(\|\beta_k\|) : \|\beta_k\| \neq 0\}$ , which implies that

for the hard thresholding and SCAD penalty function, the convergent rate of penalized least square estimator is  $n^{-2/5}$ . This is the optimal rate for nonparametric regression [21] if  $\lambda_n \rightarrow 0$ . However, for the  $L_1$  penalty,  $a_n^* = \lambda_n$ ; hence  $\lambda_n = o_p(\rho_n)$  is required to lead to the optimal rate. On the other hand, the oracle property in part (b) of Theorem 5 requires  $\lambda_n / \rho_n \rightarrow \infty$ , which contradicts  $\lambda_n = o_p(\rho_n)$ . As a consequence, for the  $L_1$  penalty, the oracle property does not hold.

Next, we demonstrate the asymptotic normality of  $\hat{\beta}^*$ . First, we analyze  $\hat{\beta}^{(1)} = (\hat{\beta}_0, \dots, \hat{\beta}_q)^T$ , that is, the penalized estimator of  $\beta^{(1)} = (\beta_0, \dots, \beta_q)^T$ . Let  $Z_i^{(1)}$ ,  $\mathbf{U}_i^{(1)}$ , and  $\bar{\mathbf{U}}^{(1)}$  denote the selected columns of  $Z_i^*$ ,  $\mathbf{U}_i$ , and  $\bar{\mathbf{U}}$ , respectively, corresponding to  $\beta^{(1)}$ . Similarly, let  $\mathbf{B}^{(1)}(W_i)$  denote selected diagonal blocks of  $\mathbf{B}(W_i)$ ,  $i = 1, \dots, n$ . Define  $\gamma^{(1)} = (\gamma_0^T, \gamma_1^T, \dots, \gamma_q^T)^T$  and  $\hat{\gamma}^{(1)} = (\hat{\gamma}_0^T, \hat{\gamma}_1^T, \dots, \hat{\gamma}_q^T)^T$ . Then by part (a) of Theorem 5 and (21),  $\hat{\gamma}^{(1)}$  is the local solution of

$$\begin{aligned} PU_n^*(\gamma^{(1)}) &= -\frac{1}{n} \sum_{i=1}^n \int_0^\tau \left( \mathbf{U}_i^{(1)}(W_i) - \bar{\mathbf{U}}^{(1)}(t) \right)^T \\ &\quad \times [dN_i(t) - Y_i(t) \\ &\quad \times \{ \mathbf{U}_i^{(1)}(W_i) \gamma^{(1)} + \alpha^{*T} V_i \}] dt \\ &\quad + \sum_{k=0}^q \frac{\partial p_\lambda(\|\gamma_k\|_k)}{\partial \gamma^{(1)}}. \end{aligned} \quad (23)$$

Recall that  $\alpha^*$  is root- $n$  consistent. It follows from (23) that

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{U}_i^{(1)}(W_i) - \bar{\mathbf{U}}^{(1)}(t) \right\}^T dM_i(t) \\ &\quad \times (1 + O_p(n^{-1/2})) - \sum_{k=0}^q \frac{\partial p_\lambda(\|\hat{\gamma}_k\|_k)}{\partial \gamma^{(1)}} \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{U}_i^{(1)}(W_i) - \bar{\mathbf{U}}^{(1)}(t) \right\}^T Y_i(t) \\ &\quad \times \{ \mathbf{U}_i^{(1)}(W_i) \hat{\gamma}^{(1)} - Z_i^{(1)T} \beta^{(1)}(W_i) \} dt \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{U}_i^{(1)}(W_i) - \bar{\mathbf{U}}^{(1)}(t) \right\}^T dM_i(t) \\ &\quad \times (1 + O_p(n^{-1/2})) - \sum_{k=0}^q \frac{\partial p_\lambda(\|\hat{\gamma}_k\|_k)}{\partial \gamma^{(1)}} \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{U}_i^{(1)}(W_i) - \bar{\mathbf{U}}^{(1)}(t) \right\}^T Z_i^{(1)T} Y_i(t) \\ &\quad \times \{ \hat{\beta}^{(1)}(W_i) - \beta^{(1)}(W_i) \} dt. \end{aligned} \quad (24)$$

Thus we can obtain the following theorem.



**Theorem 8.** Suppose Assumptions (A.i) and (A.vii)–(A.iv) hold,  $\lim_{n \rightarrow \infty} \rho_n = 0$ , if  $\lambda_n \rightarrow 0$  and  $\lambda_n / \max\{\rho_n, a_n^*, (L_{\max}/n)^{1/2}\} \rightarrow \infty$  as  $n \rightarrow \infty$ . For any  $w \in \mathcal{W}$ , let  $\bar{\gamma}^{(1)}$  and  $\bar{\beta}^{(1)}(w)$  be the conditional means of  $\hat{\gamma}^{(1)}$  and  $\hat{\beta}^{(1)}(w)$  given  $\{Z_i, V_i, W_i, i = 1, 2, \dots, n\}$ . We then have

$$\sqrt{nL_{\max}^{-1}} \left\{ \mathbf{B}^{(1)}(w) (\Xi_1 + \Gamma^*)^{-1} \Omega^* (\Xi_1 + \Gamma^*)^{-1} \left\{ \mathbf{B}^{(1)}(w) \right\}^T \right\}^{-1/2} \times \left\{ \hat{\beta}^{(1)}(w) - \bar{\beta}^{(1)}(w) \right\} \xrightarrow{L} N(0, \mathbf{I}_{q+1}), \quad (25)$$

where  $\Gamma^* = \text{diag}((L_{\max}/2)(p'_\lambda(\|\beta_k\|_{L_2})/\|\beta_k\|_{L_2})\mathbf{R}_k)_{0 \leq k \leq q}$ ,  $\Xi_1 = \lim_{n \rightarrow \infty} (L_{\max}/n) \sum_{i=1}^n \int_0^\tau \{\mathbf{U}_i^{(1)}(w) - \bar{\mathbf{U}}^{(1)}(t)\}^{\otimes 2} Y_i(t) dt$ , and  $\Omega^* = \lim_{n \rightarrow \infty} (L_{\max}/n) \sum_{i=1}^n \int_0^\tau \mathbf{U}_i^{(1)}(W_i) - \bar{\mathbf{U}}^{(1)}(t)\}^{\otimes 2} Y_i(t) \lambda_i(t) dt$ .

#### 4. Simulation Studies

We carried out two sets of simulation studies to examine the finite-sample properties of the proposed methods.  $w$  was generated from a uniform distribution over  $[0, 3]$  and  $\lambda_n$  was selected via cross-validation [7]. We set  $a = 2 + \sqrt{3}$  and reported sampling properties based on 200 replications. We consider two scenarios: (I) the performance of the regression coefficient estimates. Without loss of generality, we set  $p = 1$ . We generated failure times from the partially linear additive hazards model (1) with  $\beta(w) = 1.2 + \sin(2w)$ ,  $\alpha = (1.5, 0, 1, 0, 0)^T$ ,  $g(w) = 0.3w^2$ , and  $\lambda_0(t) = 0.5$ . Covariate  $Z$  was generated from a uniform distribution over  $[0, 1]$ ,  $V_1$  was a Bernoulli random variable taking a value of 0 or 1 with a probability of 0.5,  $V_2, V_3$  were generated from uniform distributions over  $[0, 1]$  and  $[0, 2]$ , respectively, and  $(V_4, V_5)^T$  was generated from a normal distribution with mean zero, variance 0.25, and covariance 0.25. We generated the censoring time  $C$  from a uniform distribution over  $[cc/2, 3cc/2]$ . We took  $cc = 0.86$  to yield an approximate censoring rate of  $\text{CR} = 20\%$  and  $cc = 0.35$  to yield an approximate censoring rate of  $\text{CR} = 30\%$ . We used sample size  $n = 200$ . We selected the optimal bandwidth  $\hat{h}_{\text{opt}}$  by using the method of [22] and found that the value is approximately 0.203. We present the average number of zero coefficients in Table 1, in which the column labeled “correct” presents the average restricted only to the true zero coefficients, while the column labeled “incorrect” depicts the average of coefficients erroneously set to 0. We also report the standard deviations (SD), the average standard errors (SE), and the coverage probability (CP) of the 95% confidence interval for the nonzero parameters in Table 2.

As shown in Tables 1 and 2, SCAD, HARD, and  $L_1$  perform well and select about the same correct number of significant variables. The coverage probability (CP) based on SCAD and HARD, however, is better than that based on  $L_1$ .

In scenario (II), we focused on functional coefficients  $\beta(\cdot)$ . We set  $d = 1$ ,  $g(w) = 0.3w^2$ , and  $\alpha = 1$ . We generated failure times from the partially linear additive hazards model (1) with  $\beta_1(w) = 1.2 + \sin(2w)$ ,  $\beta_3(w) = 0.5w^2$ , and  $\beta_2(w) = \beta_4(w) = \beta_5(w) \equiv 0$ . Both covariates  $V$  and  $Z_1$  were generated from a uniform distribution over

TABLE 1: Results of the simulation study for scenario (I). Here “correct” is the average true positive rate and “incorrect” is the average true negative rate.

	SCAD HARD $L_1$			SCAD HARD $L_1$		
	$n = 200, \text{CR} = 20\%$			$n = 200, \text{CR} = 30\%$		
Correct	2.73	2.76	2.67	Correct	2.85	2.85
Incorrect	0.12	0.11	0.21	Incorrect	0.13	0.15

TABLE 2: Simulation results for scenario (I): bias, standard error (SE), standard deviation (SD), and coverage probability (CP).

	$\alpha_1$			$\alpha_3$		
	SCAD	HARD	$L_1$	SCAD	HARD	$L_1$
$n = 200, \text{CR} = 20\%$						
bias	0.0738	0.0729	0.0527	0.0552	0.0475	0.0324
SE	0.0767	0.0767	0.0757	0.0567	0.0566	0.0558
SD	0.0635	0.0645	0.0631	0.0495	0.0501	0.0558
CP	0.96	0.91	0.91	0.90	0.89	0.88
$n = 200, \text{CR} = 30\%$						
bias	0.0726	0.0726	0.0402	0.0480	0.0466	0.0280
SE	0.0824	0.0824	0.0812	0.0612	0.0612	0.0602
SD	0.0701	0.0721	0.0726	0.0618	0.0602	0.0613
CP	0.93	0.89	0.92	0.94	0.90	0.91

$[0, 1]$ , respectively,  $(V_2, V_3)^T$  was generated from a normal distribution with a mean of  $(0.5, 0.5)^T$ , variance  $(0.25, 0.25)^T$  and covariance 0,  $V_4$  was a Bernoulli random variable taking a value of 0 or 1 with a probability of 0.5, and  $V_5$  was uniform over  $[0, 2]$ . In simulation II, we took  $cc = 1.24$  to yield an approximate censoring rate of  $\text{CR} = 10\%$  and  $cc = 0.45$  to yield an approximate censoring rate of  $\text{CR} = 30\%$ . We used sample size  $n = 200$ . The average number of zero functional coefficients based on SCAD and HARD is reported in Table 3, and fitted curves of nonzero functional coefficients are presented in Figures 1 and 2 based on SCAD and HARD, respectively.

Notice that penalized estimates of functional coefficients are based on the results in Step 1. Because the oracle property does not hold for the  $L_1$  penalty according to Remark 7, we did not take  $L_1$  penalty into account.

We can see from Table 3 and Figures 1 and 2 that the penalized spline estimation of the functional coefficient that we proposed performed well. Furthermore, for the varying coefficient part, the “correct” numbers of significant variables selected based on SCAD and HARD penalties are close. In comparing the results between different censoring rates, it is also shown that our method is extremely robust against the censoring rate.

#### 5. Real Data Example

In this section, we apply the proposed method to a SUPPORT (Study to Understand Prognoses Preferences Outcomes and Risks of Treatment) dataset [23]. This study was a multicenter study designed to examine outcomes and clinical decision

TABLE 3: Results from the simulation study for scenario (II). The legend is the same as in Table 1.

	SCAD	HARD		SCAD	HARD
	$n = 200, CR = 10\%$			$n = 200, CR = 30\%$	
Correct	2.61	2.79	Correct	2.625	2.745
Incorrect	0.165	0.195	Incorrect	0.48	0.48

making for seriously ill-hospitalized patients. One thousand patients suffering from one of eight different diseases, including acute respiratory failure (ARF), chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), cirrhosis, coma, colon cancer, lung cancer, and multiple organ system failure (MOSF) were followed up to 5.56 years. 672 observations were collected, including age, sex, race, coma score (scoma), number of comorbidities (num.co), charges, average of therapeutic intervention scoring system (avtiss), white blood cell count (wbc), heart rate (hrt), respiratory rate (resp), temperature (temp), serum bilirubin level (bili), creatinine (crea), serum sodium (sod), and imputed ADL calibrated to surrogate (adlsc). For more details about this study, refer to [23].

We suppose that binary and multinary risk factors have constant coefficients, which yields a 12-dimensional parameter after transforming multinary variables into binary. Conversely, other risk factors are supposed to have varying coefficients, that is, a 13-dimension functional coefficient. Here the main exposure variable  $W$  corresponds to age. Censoring rate of 672 patients is 32.7%. The model we considered is

$$\lambda(t | Z_i, V_i, \text{age}_i) = \lambda_0(t) + \sum_{j=1}^{13} \beta_j(\text{age}_i) Z_{ij} + \sum_{k=1}^{12} \alpha_k V_{ik} + g(\text{age}_i), \quad (26)$$

where  $V_{ij}$ ,  $j = 1, 2, \dots, 13$ , denote binary variables of the  $i$ th patient and  $Z_{ik}$ ,  $k = 1, 2, \dots, 12$ , denote the others of the  $i$ th patient. Here, baseline hazard function  $\lambda_0(t)$  is the failure risk of patients who are white females suffering from MOSF with other variables all equal to zero.

By fitting model (26) with a SCAD penalty, the parameter of “other race” is identified as zero, and functional coefficients of num.co, charges, wbc, hrt, temp, crea, sod, adlsc, and  $g(\cdot)$  are identified as zero functions. Identified significant risk factors and results of their parametric or functional coefficients are shown in Figure 3 and Table 4.

The identified zero parameter of “other race” shows that, besides Asian, black, and Hispanic, “other race” has no significant different impact on risk in contrast with white people. Table 4 demonstrates that Asians have the lowest failure probability, as do people who are suffering from MOSF. Conversely, coma is the most dangerous among these eight different diseases. Figure 3 shows the fitted coefficients and their 95% pointwise confidence intervals of the five risk factors which were identified as significant.

TABLE 4: Applied to SUPPORT data. Estimation results of significant binary variables using the SCAD penalty.

Risk factor	Estimator	Variance $\times 10^{-3}$
Sex	0.0006	0.0178
Disease group		
ARF	0.0007	0.0247
CHF	0.0011	0.0309
Cirrhosis	0.0018	0.0450
Colon cancer	0.0034	0.0580
Lung cancer	0.0042	0.0585
Coma	0.0074	0.1410
COPD	0.0017	0.0344
Race		
Asian	-0.0002	0.0414
Black	0.0006	0.0232
Hispanic	0.0008	0.0421

## Appendices

In this appendix, we list assumptions and outline the proofs of the main results. The following assumptions are imposed.

### A. Assumptions

- (A.i) The density  $f(w)$  of  $w$  is continuous, has compact support  $\mathcal{W}$ , and satisfies  $\inf_{w \in \mathcal{W}} f(w) > 0$ .
- (A.ii)  $\beta(\cdot)$  and  $g(\cdot)$  have continuous bounded second derivatives on  $\mathcal{W}$ .
- (A.iii) The density function  $K(\cdot)$  is bounded and symmetric and has a compact bounded support.  $\int_0^\tau \lambda_0(t) dt < \infty$ .
- (A.iv)  $h \rightarrow 0$ ,  $nh \rightarrow \infty$  and  $nh^4 \rightarrow 0$  as  $n \rightarrow \infty$ .
- (A.v) The conditional probability  $\rho_0(t, z, v, w)$  is equicontinuous in the arguments  $(t, w)$  on the product space  $[0, \tau] \times \mathcal{W}$ .
- (A.vi)  $\varphi_{0,0}(w, t)$  and  $\kappa_{0,0}(w, t)$  are bounded away from zero on the product space  $[0, \tau] \times \mathcal{W}$ ,  $m_1(\cdot)$  and  $m_2(\cdot)$  are continuous on  $\mathcal{W}$ ,  $m_1(w_0)$  is nonsingular for all  $w_0 \in \mathcal{W}$ , and

$$\begin{aligned} \mathbf{I}(\tau) &= \int_0^\tau \left\{ \kappa_{2,0}(t) - \frac{\kappa_{1,0}^{\otimes 2}(t)}{\kappa_{0,0}(t)} \right\} dt, \\ \mathbf{\Sigma}(\tau) &= \int_0^\tau \left[ \kappa_{2,1}(t) + \left\{ \frac{\kappa_{1,0}(t)}{\kappa_{0,0}(t)} \right\}^{\otimes 2} \kappa_{0,1}(t) \right. \\ &\quad \left. - 2 \frac{\kappa_{1,0}(t)}{\kappa_{0,0}(t)} \kappa_{1,1}(t) \right] dt \end{aligned} \quad (\text{A.1})$$

are positive-definite.

- (A.vii) The eigenvalues of the matrix  $E\{\int_0^\tau Z(t)Z^T(t)Y(t)dt\}$  are uniformly bounded away from 0.

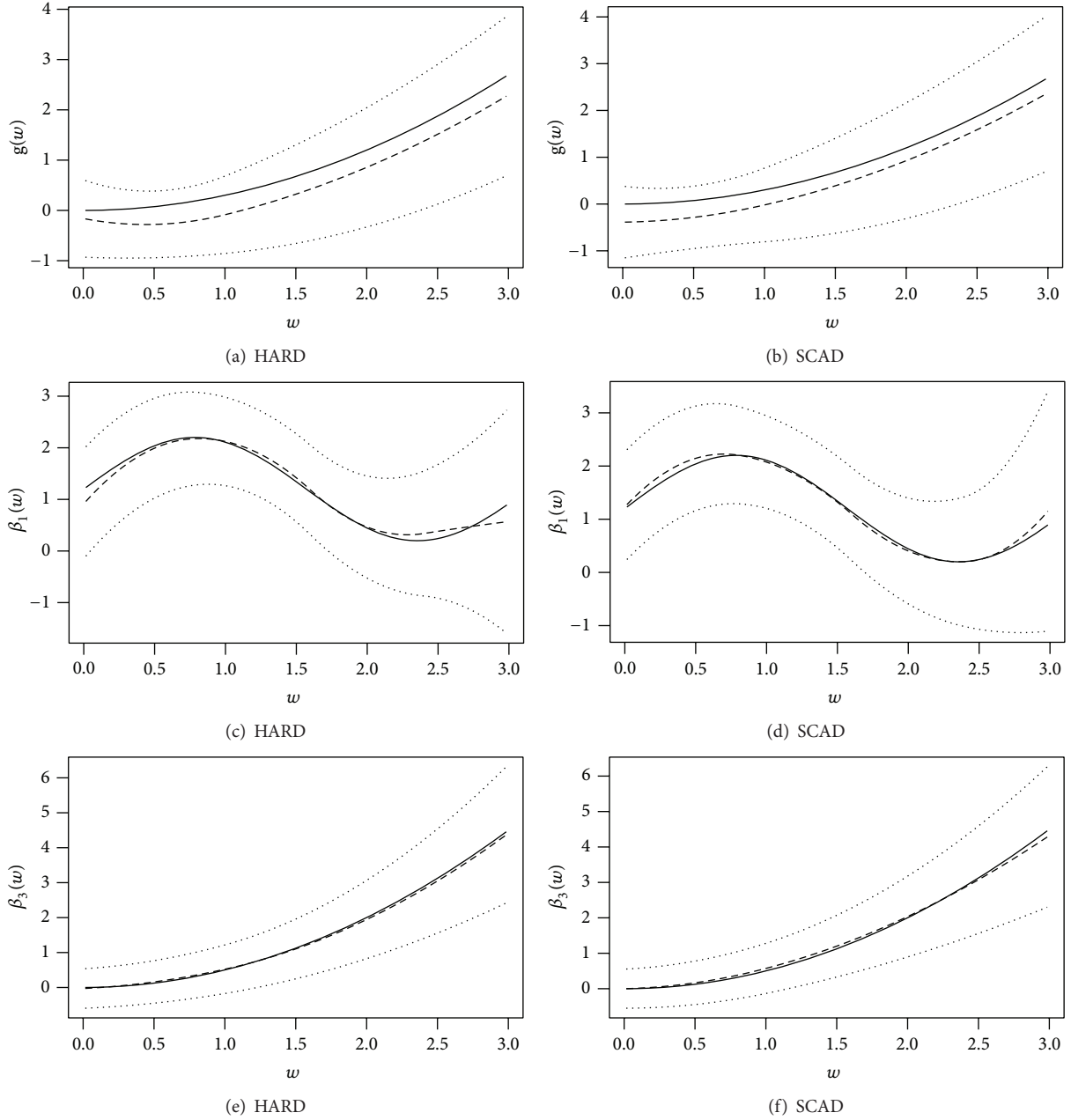


FIGURE 1: Results of the simulation study for scenario (I): the estimated curves of the nonzero functional coefficients  $\beta_1(w)$ ,  $\beta_3(w)$ , and  $g(w)$ . The solid, dashed, and dotted curves represent the true function, estimated functions, and 95% pointwise confidence intervals, respectively. Left panel—HARD penalty; right panel—SCAD penalty.

$$(A.viii) \limsup_n (\max_k L_k / \min_k L_k) < \infty.$$

$$(A.ix) \lim_{n \rightarrow \infty} n^{-1} L_{\max} \log(L_{\max}) = 0.$$

## B. Technical Lemmas

Write  $c_n = (nh)^{-1/2} + h^2$ . Before we prove the theorems, we present several lemmas for the proofs of the main results. Lemma B.1 will be used for the proofs of Theorems 1 and 3. Lemma B.5 establishes the convergence rate for the spline approximation of nonparametric functions. Its proof can be finished in a similar way as in [24].

Write

$$\begin{aligned} C_{n1}(t, w_0) &= n^{-1} \sum_{i=1}^n Y_i(t) q_1 \left( t, \frac{(W_i - w_0)}{h}, Z_i, V, w_0 \right) \\ &\quad \times K_h(W_i - w_0), \end{aligned}$$

$$C_{n2}(t, w_0) = n^{-1} \sum_{i=1}^n Y_i(t) q_2(t, W_i, Z_i, V)$$

(B.1)

for functions  $q_1(\cdot, \cdot, \cdot, \cdot, \cdot)$  and  $q_2(\cdot, \cdot, \cdot, \cdot)$ .



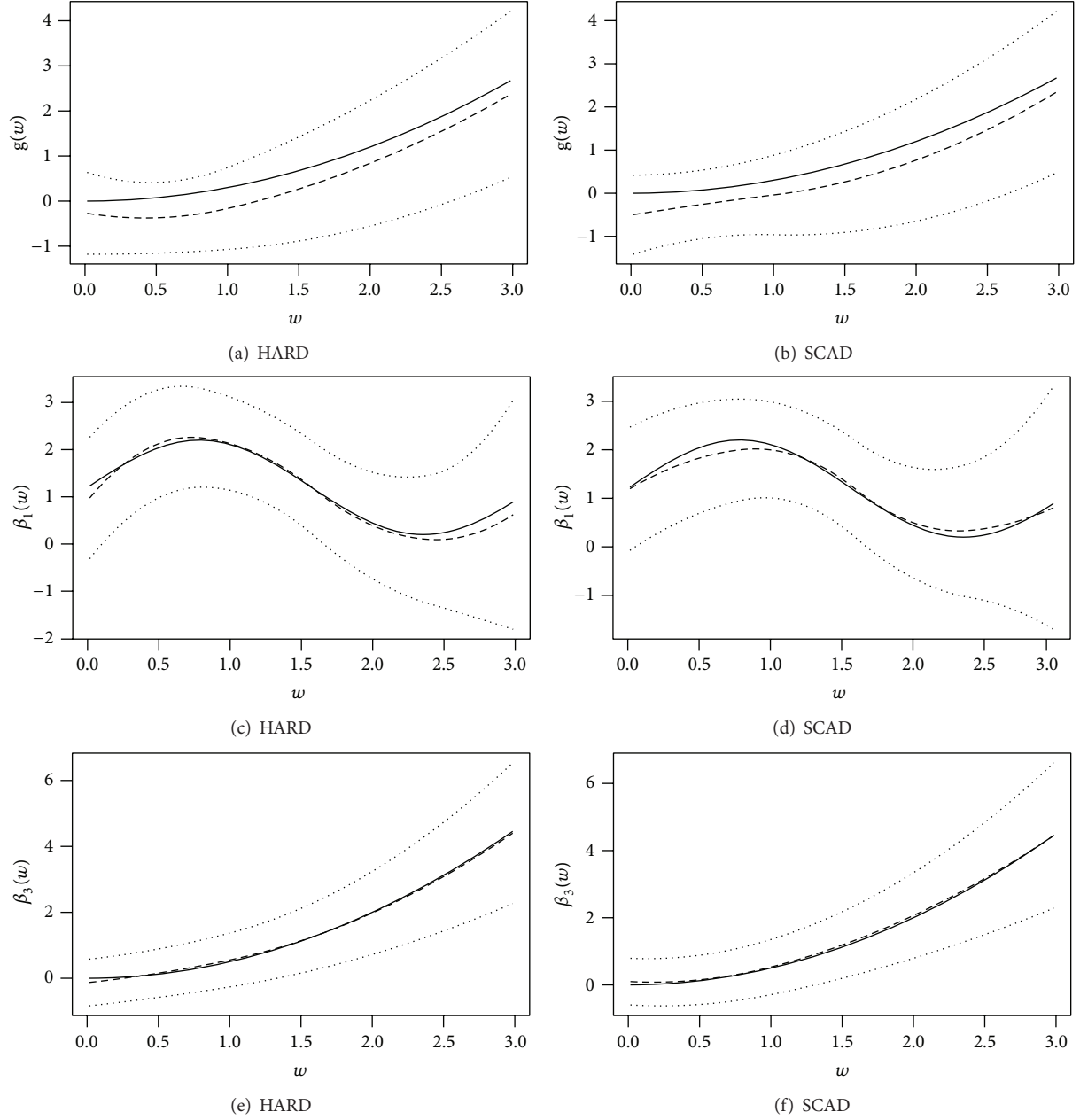


FIGURE 2: Results of the simulation study for scenario (II). The legend is the same as in Figure 1.

**Lemma B.1.** Assume that  $q_1(t, u, Z, V, w)$  is equicontinuous in its arguments  $w$  and  $t$ ,  $q_2(t, W, Z, V)$  is equicontinuous in its argument  $t$ , and that  $E(q_1(t, u, Z, V, W) \mid W = w_0)$  is equicontinuous in the argument  $w_0$ . Under Assumptions (A.i)–(A.v), one has for each  $w_0 \in \mathcal{W}$ ,

$$C_{nj}(t, w_0) = C_j(t, w_0) + O_p(c_n),$$

$$\sup_{0 \leq t \leq \tau} \sup_{w_0 \in \mathcal{W}} |C_{nj}(t, w_0) - C_j(t, w_0)| \xrightarrow{P} 0, \quad (\text{B.2})$$

where  $C_j(t, w_0) = f(w_0) \int E\{Y(t)q_j(t, u, Z, V, w_0) \mid W = w_0\}K(u)du$  for  $j = 1, 2$ .

Its proof can be finished using the similar arguments in Lemma 1 of [25].

**Lemma B.2.** Let  $U_n^*(\alpha_0) = -(\partial L_n(\alpha)/\partial \alpha)|_{\alpha=\alpha_0}$ . Assume that (13) and Assumptions (A.i)–(A.vi) hold. We then have

$$U_n^*(\alpha_0) = \sum_{i=1}^n \int_0^\tau [\bar{V}_i(W_i) - \bar{V}(t)]^\top dM_i(t) + o_p(\sqrt{n}) \quad (\text{B.3})$$

$$\equiv \sqrt{n}\tilde{U}_n(\alpha_0) + o_p(\sqrt{n}),$$

where  $\bar{V}(t) = \sum_{i=1}^n Y_i(t)\bar{V}_i(W_i)/\sum_{i=1}^n Y_i(t)$ .

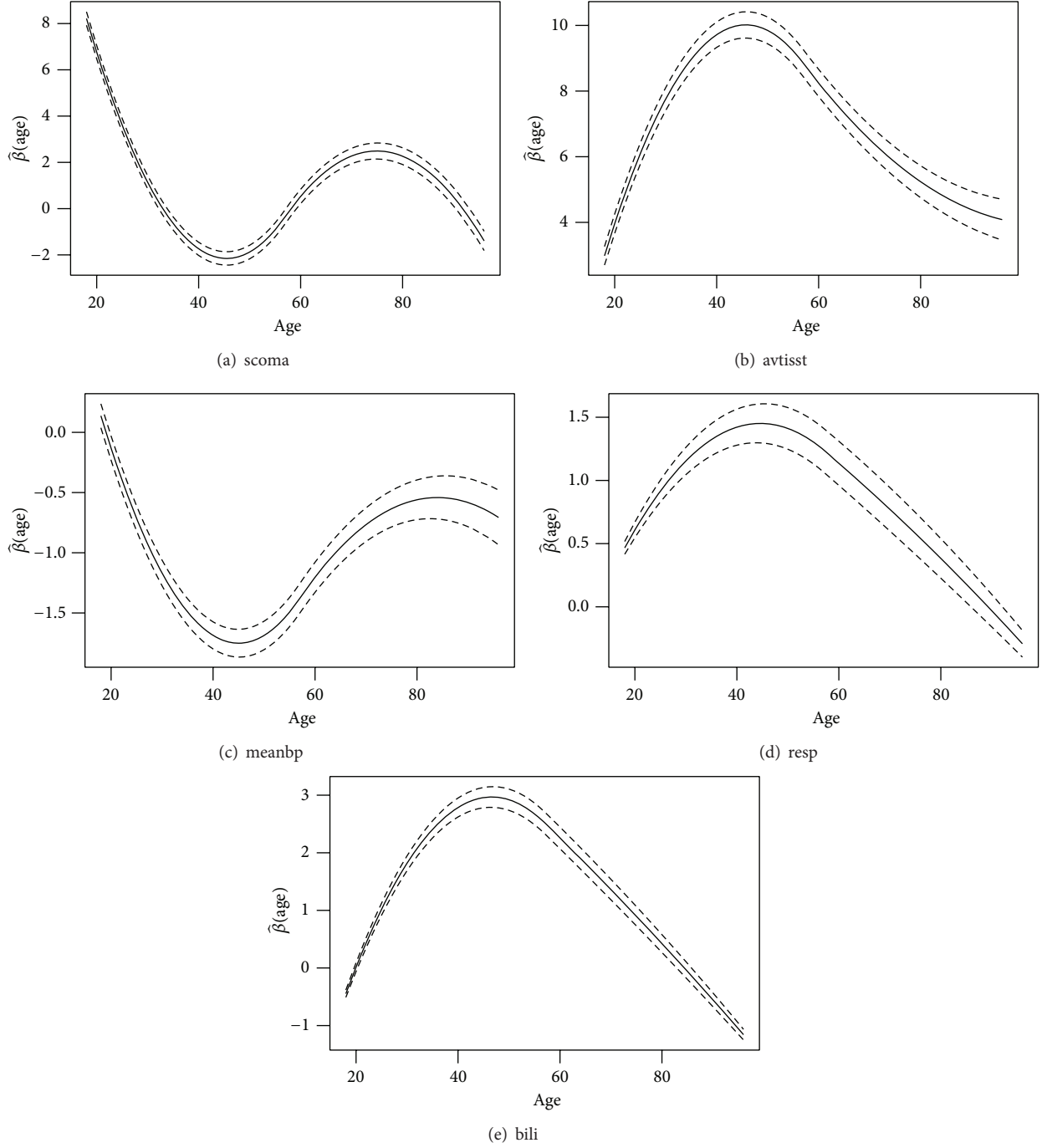


FIGURE 3: Results of the real data example. The estimated curves of those nonzero functional coefficients. The solid and dashed curves represent the estimated curve and 95% pointwise confidence intervals ( $\times 10^4$ ).

*Proof.* Because  $U_n^*(\alpha_0) = -(\partial L_n(\alpha)/\partial \alpha)|_{\alpha=\alpha_0}$ , we have

$$\begin{aligned}
 U_n^*(\alpha_0) &= \sum_{i=1}^n \int_0^\tau \{V_i^*(W_i) - \bar{V}^*(t)\}^T \\
 &\quad \times [dN_i(t) - Y_i(t) \{Z_i^T \hat{\beta}(W_i, \alpha_0) + \hat{g}(W_i, \alpha_0) + \alpha_0^T V_i\} dt] \\
 &= \sum_{i=1}^n \int_0^\tau \{V_i^*(W_i) - \bar{V}^*(t)\}^T \\
 &\quad \times [dN_i(t) - Y_i(t) \{G_{ni}(t) + V_i^*(W_i) \alpha\} dt]
 \end{aligned} \tag{B.4}$$

with  $\bar{V}^*(t) = \sum_{i=1}^n Y_i(t) V_i^*(W_i) / \sum_{i=1}^n Y_i(t)$ . By a similar discussion in [5], we can prove that the biases of  $\hat{\beta}(w_0, \alpha_0)$  and  $\hat{g}(w_0, \alpha_0)$  are the order of  $O_p(h^2)$  at each point  $w_0 \in \mathcal{W}$ , respectively.

Furthermore, recall that  $V_i^*(W_i) = V_i^T - Z_i^T T_{n2}(W_i) - T_{n4}(W_i)$  with  $T_{n2}(W_i) = (I_p, 0_{p \times p}, 0_p) M_{n1}^{-1}(W_i) S_{n2}(W_i)$  and  $T_{n4}(W_i) = \int_0^{W_i} (0_{p \times p}, 0_{p \times p}, 1_p) M_{n1}^{-1}(w_0) S_{n2}(w_0) dw_0$ , where

$$\begin{aligned} M_{n1}(w_0) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(W_i - w_0) \{Z_i^*(w_0) - \bar{Z}(t, w_0)\}^{\otimes 2} Y_i(t) dt \\ &= \int_0^\tau \left\{ \Phi_{n20}(t, w_0) - \frac{\Phi_{n10}^{\otimes 2}(t, w_0)}{\Phi_{n00}(t, w_0)} \right\} dt, \\ S_{n2}(w_0) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(W_i - w_0) \{Z_i^*(w_0) - \bar{Z}(t, w_0)\} V_i^T Y_i(t) dt \\ &= \int_0^\tau \left\{ \Phi_{n11}(t, w_0) - \frac{\Phi_{n10}(t, w_0)}{\Phi_{n00}(t, w_0)} \Phi_{n01}(t, w_0) \right\} dt, \end{aligned} \quad (B.5)$$

with  $\Phi_{nkl}(t, w_0) = (1/n) \sum_{i=1}^n K_h(W_i - w_0) \{Z_i^*(w_0)\}^{\otimes k} (V_i^{\otimes l})^T Y_i(t)$  for  $k = 0, 1, 2$  and  $l = 0, 1$ .

It follows from Lemma B.1 that

$$\begin{aligned} M_{n1}(w_0) &= m_1(w_0) + O_p(c_n), \\ S_{n2}(w_0) &= m_2(w_0) + O_p(c_n) \end{aligned} \quad (B.6)$$

for each  $w_0 \in \mathcal{W}$ . This implies that

$$\begin{aligned} T_{n2}(w) &= t_1(w) + O_p(c_n), \\ T_{n4}(w) &= t_2(w) + O_p(c_n) \end{aligned} \quad (B.7)$$

uniformly hold on  $\mathcal{W}$ . Hence it follows from Assumption (A.iv), the arguments similar to the proof of Theorem 3 in [5] and the martingale central limit that

$$U_n^*(\alpha_0) = \sum_{i=1}^n \int_0^\tau \{\tilde{V}_i(W_i) - \bar{V}(t) + O_p(c_n)\}^T dM_i(t)$$

$$\begin{aligned} & - \sum_{i=1}^n \int_0^\tau \{\tilde{V}_i(W_i) - \bar{V}(t) + O_p(c_n)\}^T Y_i(t) \\ & \times \left\{ (Z_i^T(t), 0_{p+1}^T) M_{n1}^{-1}(W_i) \right. \\ & \quad \times \frac{1}{n} \sum_{j=1}^n \int_0^\tau K_h(W_j - W_i) \mathbf{H}^{-1} \\ & \quad \times \{Z_i^*(t, W_i) - \bar{Z}(t, W_i)\} dM_j(t) \\ & \quad \left. + (0_{2p}^T, h^{-1}) \right. \\ & \quad \times \int_0^{W_i} M_{n1}^{-1}(w) \\ & \quad \times \left[ \frac{1}{n} \sum_{j=1}^n \int_0^\tau K_h(W_j - w) \mathbf{H}^{-1} \right. \\ & \quad \times \{Z_i^*(w) - \bar{Z}(t, w)\} \\ & \quad \left. \left. \times dM_j(t) \right] dw \right\} dt + o_p(\sqrt{n}). \end{aligned} \quad (B.8)$$

A further simplification indicates that  $U_n^*(\alpha_0)$  can be expressed as

$$\begin{aligned} & \sum_{i=1}^n \int_0^\tau \{\tilde{V}_i(W_i) - \bar{V}(t)\}^T dM_i(t) \\ & - \sum_{i=1}^n \int_0^\tau \{\tilde{V}_i(W_i) - \bar{V}(t) + O_p(c_n)\}^T Y_i(t) \\ & \quad \times \left[ \{Z_i^T, 0_p^T, 1\} m_1^{-1}(W_i) O_p(c_n) \right] dt \\ & \quad + o_p(\sqrt{n}) \\ & = \sum_{i=1}^n \int_0^\tau \{\tilde{V}_i(W_i) - \bar{V}(t)\}^T dM_i(t) + o_p(\sqrt{n}). \end{aligned} \quad (B.9)$$

Thus (B.3) holds.  $\square$

**Lemma B.3.** Assume that (13) and Assumptions (A.i)–(A.vi) hold. If  $\lambda_n \rightarrow 0$ ,  $\sqrt{n}\lambda_n \rightarrow \infty$  as  $n \rightarrow \infty$ , then with probability tending to 1, for any given  $\alpha_1$  satisfying  $\|\alpha_1 - \alpha_{10}\| = O_p(n^{-1/2})$ , and any given constant  $C$ ,

$$L_n^* \{(\alpha_1^T, 0_{d-s}^T)\} = \min_{\|\alpha_2\| \leq Cn^{-1/2}} L_n^* \{(\alpha_1^T, \alpha_2^T)\}. \quad (B.10)$$

*Proof.* It suffices to prove that, for any given  $\alpha_1$  satisfying  $\|\alpha_1 - \alpha_{10}\| = O_p(n^{-1/2})$ , any given constant  $C > 0$ , and each  $\alpha_j$ ,  $j = s+1, \dots, d$ ,

$$\begin{aligned} \frac{\partial L_n^*(\alpha)}{\partial \alpha_j} &< 0 \quad \text{if } -Cn^{-1/2} < \alpha_j < 0, \\ \frac{\partial L_n^*(\alpha)}{\partial \alpha_j} &> 0 \quad \text{if } 0 < \alpha_j < Cn^{-1/2} \end{aligned} \quad (\text{B.11})$$

hold with probability tending to 1. Notice that  $U_n^*(\alpha_0) = -(\partial L_n(\alpha)/\partial \alpha)|_{\alpha=\alpha_0}$ . Then for each  $\alpha$  in a neighborhood of  $\alpha_0$  from Lemma B.2 we have

$$\begin{aligned} &\frac{1}{\sqrt{n}} \{L_n(\alpha) - L_n(\alpha_0)\} \\ &= -\tilde{U}_n(\alpha_0)(\alpha - \alpha_0) \{1 + o_p(1)\} \\ &\quad + \frac{1}{2}(\alpha - \alpha_0)^T \sqrt{n}\Pi(\alpha_0)(\alpha - \alpha_0) \{1 + o_p(1)\}. \end{aligned} \quad (\text{B.12})$$

By similar algebra with the proof of Lemma B.2, we know that

$$\begin{aligned} \Pi_n(\alpha_0) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[ \{\tilde{V}_i(W_i) - \bar{V}(t)\}^T \right]^{\otimes 2} Y_i(t) \lambda_0(t) dt + O_p(c_n) \\ &= E \left( \int_0^\tau \left[ \{\tilde{V}_i(W_i) - \bar{V}(t)\}^T \right]^{\otimes 2} Y(t) \lambda_0(t) dt \right) + O_p(c_n) \\ &\equiv \Pi^*(\alpha_0) + O_p(c_n) \end{aligned} \quad (\text{B.13})$$

is positive-definite with probability tending to 1 from Assumption (A.vi) and Lemma B.1. Hence for  $\alpha - \alpha_0 = O_p(n^{-1/2})$ , by the fact of  $\tilde{U}_n = O_p(1)$ , we have

$$\begin{aligned} &\frac{1}{\sqrt{n}} \frac{\partial L_n(\alpha)}{\partial \alpha} \\ &= -\tilde{U}_n \{1 + o_p(1)\} + \sqrt{n}\Pi^*(\alpha_0)(\alpha - \alpha_0) \{1 + o_p(1)\} \\ &= \sqrt{n}\Pi^*(\alpha_0)(\alpha - \alpha_0) + O_p(1), \end{aligned} \quad (\text{B.14})$$

and then

$$\begin{aligned} &\frac{\partial L_n^*(\alpha)}{\partial \alpha_j} \\ &= \frac{\partial L_n(\alpha)}{\partial \alpha_j} + np'_\lambda(|\alpha_j|) \text{sgn}(\alpha_j) \\ &= \left\{ O_p \left( \frac{n^{-1/2}}{\lambda_n} \right) + \lambda_n^{-1} p'_\lambda(|\alpha_j|) \text{sgn}(\alpha_j) \right\} n\lambda_n. \end{aligned} \quad (\text{B.15})$$

As a result, it follows from (13) and  $\lambda_n \rightarrow 0$ ,  $\sqrt{n}\lambda_n \rightarrow \infty$  as  $n \rightarrow \infty$  that  $\text{sgn}(\partial L_n^*(\alpha)/\partial \alpha_j) = \text{sgn}(\alpha_j)$  for each  $\alpha - \alpha_0 = O_p(n^{-1/2})$ , which indicates that (B.11) holds with probability tending to 1.  $\square$

**Lemma B.4.** Suppose Assumptions (A.i) and (A.vii)–(A.ix) hold, then there are positive constants  $M_1$  and  $M_2$  such that, except in an event whose probability tends to zero, all the eigenvalues of

$$\frac{L_{\max}}{n} \sum_{i=1}^n \int_0^\tau \{U_i(W_i) - \bar{U}(t)\}^{\otimes 2} Y_i(t) dt \quad (\text{B.16})$$

between  $M_1$  and  $M_2$ , and consequently  $\sum_{i=1}^n \int_0^\tau \{U_i(W_i) - \bar{U}(t)\}^{\otimes 2} Y_i(t) dt$ , are invertible.

*Proof.* According to Lemmas A.1 and A.2 of [24], except in an event whose probability tends to zero,  $\|\sum_{k,l} \gamma_{kl} B_{kl}(w)\|_{L_2}^2 \asymp |\gamma|^2/L_{\max}$ . Note that  $\gamma^T [\sum_{i=1}^n \int_0^\tau \{U_i(W_i) - \bar{U}(t)\}^{\otimes 2} Y_i(t) dt] \gamma \asymp \|\sum_{k,l} \gamma_{kl} B_{kl}(w)\|_{L_2}^2$  by Assumption (A.ii). Hence, Lemma B.4 holds.  $\square$

In the rest of this paper, we denote  $n^{-1} \sum_{i=1}^n \int_0^\tau \{U_i(W_i) - \bar{U}(t)\}^{\otimes 2} Y_i(t) dt$  by  $\Xi_n$ . Let

$$\begin{aligned} \check{\gamma} &= \Xi_n^{-1} \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{U_i(W_i) - \bar{U}(t)\} \\ &\quad \times \{\lambda_0(t) + Y_i(t) \beta^{*\top}(W_i) Z_i^* \\ &\quad + O_p(n^{-1/2})\} dt, \end{aligned} \quad (\text{B.17})$$

and  $\check{\beta}^*(w) = B(w)\check{\gamma}$ . Then  $\check{\gamma}$  and  $\check{\beta}^*(w)$  are the mean of  $\tilde{\gamma}$  and  $\tilde{\beta}^*(w)$  conditioning on  $\{Z_i, V_i, W_i, i = 1, 2, \dots, n\}$ , respectively.

**Lemma B.5.** Suppose Assumptions (A.ii), (A.viii), and (A.ix) hold. We then have  $\rho_n \rightarrow 0$ ,  $\sqrt{n}\rho_n \rightarrow \infty$  as  $n \rightarrow \infty$ ,  $\|\tilde{\beta}^* - \check{\beta}^*\|_{L_2}^2 = O_p(L_{\max}/n)$ ,  $\|\tilde{\beta}^* - \beta^*\|_{L_2} = O_p(\rho_n)$ , and  $\|\tilde{\beta}^* - \beta^*\|_{L_2}^2 = O_p(L_{\max}/n + \rho_n^2)$ .

*Proof.* By Assumptions (A.ii) and (A.viii),  $\rho_n = O(L_{\max}^{-2})$  [20, Theorem 6.27]. Hence, it follows from Assumption (A.iv) that  $\rho_n \rightarrow 0$  and  $\sqrt{n}\rho_n \rightarrow \infty$  as  $n \rightarrow \infty$ . By the triangle inequality,  $\|\tilde{\beta}^* - \beta^*\|_{L_2} \leq \|\tilde{\beta}^* - \check{\beta}^*\|_{L_2} + \|\check{\beta}^* - \beta^*\|_{L_2}$ . Note that

$$\begin{aligned} &\tilde{\gamma} - \check{\gamma} \\ &= \Xi_n^{-1} \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{U_i(W_i) - \bar{U}(t)\} \{dM_i(t) + O_p(n^{-1/2})\}. \end{aligned} \quad (\text{B.18})$$

Thus, according to  $\Xi_n^{-1}(1/n) \sum_{i=1}^n \int_0^\tau \{U_i(W_i) - \bar{U}(t)\} dM_i(t) = O_p(L_{\max}^2/n)$ , which can be similarly proved by Lemma A.4 in [24], we have  $\|\tilde{\beta}^* - \check{\beta}^*\|_{L_2}^2 \asymp |\tilde{\gamma} - \check{\gamma}|^2/L_{\max} = O_p(L_{\max}/n)$ . On the other hand, by the properties of B-spline basis functions, we can prove that  $\|\tilde{\beta}^* - \beta^*\|_{L_2} = O_p(\rho_n)$  by the similar argument of Lemma A.7 in [24]. Thus, Lemma B.5 holds.  $\square$

**Lemma B.6.** Suppose (13) and Assumptions (A.i) and (A.vii)–(A.iv) hold. If  $\rho_n \rightarrow 0$  and  $\lambda_n/\rho_n \rightarrow \infty$ ,  $b_n^* \rightarrow 0$  as  $n \rightarrow \infty$ , we then have  $\|\hat{\beta}^* - \check{\beta}^*\|_{L_2} = O_p(a_n^* + (\lambda_n \rho_n)^{1/2} + (L_{\max}/n)^{1/2})$ .

*Proof.* Using the properties of B-spline basis functions (Section A.2 of [24]), we have

$$\|\hat{\beta}_k - \check{\beta}_k\|_{L_2}^2 = \|\hat{\gamma}_k - \check{\gamma}_k\|_k^2 \asymp L_k^{-1} \|\hat{\gamma}_k - \check{\gamma}_k\|^2, \quad (\text{B.19})$$

which we sum over  $k$  to obtain

$$\|\hat{\beta}^* - \check{\beta}^*\|_{L_2}^2 = \sum_{k=0}^p \|\hat{\gamma}_k - \check{\gamma}_k\|_k^2 \asymp L_{\max}^{-1} \|\hat{\gamma} - \check{\gamma}\|^2. \quad (\text{B.20})$$

Let  $\delta_n = a_n^* + (\lambda_n \rho_n)^{1/2} + (L_{\max}/n)^{1/2}$ . It suffices to show that, for any given  $\varepsilon > 0$ , there exists a large enough  $C$  such that

$$\Pr \left\{ \inf_{\|\mathbf{u}\|=C} PL_n(\check{\gamma} + L_{\max}^{1/2} \delta_n \mathbf{u}) > PL_n(\check{\gamma}) \right\} \geq 1 - \varepsilon, \quad (\text{B.21})$$

which implies  $PL_n(\cdot)$  can reach a local minimum in the ball  $\{\check{\gamma} + L_{\max}^{1/2} \delta_n \mathbf{u} : \|\mathbf{u}\| \leq C\}$  with a probability of at least  $1 - \varepsilon$ . Thus, there exists a local minimizer satisfying  $\|\hat{\gamma} - \check{\gamma}\| = O_p(L_{\max}^{1/2} \delta_n)$ . Let  $\hat{\gamma} = \check{\gamma} + L_{\max}^{1/2} \delta_n \mathbf{u}$ ; then by the Taylor expansion and the definition of  $\check{\gamma}$  we have

$$\begin{aligned} PL_n(\hat{\gamma}) - PL_n(\check{\gamma}) &= \frac{n}{2} (\hat{\gamma} - \check{\gamma})^T \Xi_n (\hat{\gamma} - \check{\gamma}) \{1 + o_p(1)\} \\ &\quad + \frac{n}{2} (\check{\gamma} - \check{\gamma})^T \Xi_n (\check{\gamma} - \check{\gamma}) \{1 + o_p(1)\} \\ &\quad + n \sum_{k=0}^p \{p_{\lambda_n}(\|\hat{\gamma}_k\|_k) - p_{\lambda_n}(\|\check{\gamma}_k\|_k)\}. \end{aligned} \quad (\text{B.22})$$

It follows from Lemmas B.4 and B.5 that

$$\begin{aligned} PL_n(\hat{\gamma}) - PL_n(\check{\gamma}) &= n \left( \delta_n + \left( \frac{L_{\max}}{n} \right)^{1/2} \right)^2 \|\mathbf{u}\|^2 O_p(1) \{1 + o_p(1)\} \\ &\quad + n \sum_{k=0}^p \{p_{\lambda_n}(\|\hat{\gamma}_k\|_k) - p_{\lambda_n}(\|\check{\gamma}_k\|_k)\}. \end{aligned} \quad (\text{B.23})$$

We first examine the second term on the right-hand side of (B.23). A direct calculation with the Taylor expansion and using  $p_{\lambda_n}(0) = 0$  yields that

$$\begin{aligned} &\sum_{k=0}^p \{p_{\lambda_n}(\|\hat{\gamma}_k\|_k) - p_{\lambda_n}(\|\check{\gamma}_k\|_k)\} \\ &= \sum_{k=0}^p [\{p_{\lambda_n}(\|\hat{\gamma}_k\|_k) - p_{\lambda_n}(\|\beta_k\|_{L_2})\} \\ &\quad - \{p_{\lambda_n}(\|\check{\gamma}_k\|_k) - p_{\lambda_n}(\|\beta_k\|_{L_2})\}] \\ &\geq \sum_{k=0}^q [\{p'_{\lambda_n}(\|\beta_k\|_{L_2})(\|\hat{\gamma}_k\|_k - \|\beta_k\|_{L_2}) \\ &\quad - p'_{\lambda_n}(\|\beta_k\|_{L_2})(\|\check{\gamma}_k\|_k - \|\beta_k\|_{L_2})\} \\ &\quad + \frac{1}{2} p''_{\lambda_n}(\|\beta_k\|_{L_2}) \\ &\quad \times \{(\|\hat{\gamma}_k\|_k - \|\beta_k\|_{L_2})^2 - (\|\check{\gamma}_k\|_k - \|\beta_k\|_{L_2})^2\} \\ &\quad \times (1 + o_p(1))] \\ &\quad - \sum_{l=q+1}^p p_{\lambda_n}(\|\check{\gamma}_l\|_l) \geq -a_n^* O_p(\delta_n) \|\mathbf{u}\| \\ &\quad - b_n^* O_p(\delta_n^2 + \delta_n \rho_n) \|\mathbf{u}\|^2 - p'_{\lambda_n}(0) O_p(\rho_n). \end{aligned} \quad (\text{B.24})$$

Here,  $p'_{\lambda_n}(0) = \lim_{\theta \downarrow 0} p'_{\lambda_n}(\theta)$ . Thus, combining that with (B.23) yields

$$\begin{aligned} &\frac{1}{n} \{PL_n(\hat{\gamma}) - PL_n(\check{\gamma})\} \\ &\geq \left( \delta_n + \left( \frac{L_{\max}}{n} \right)^{1/2} \right)^2 \|\mathbf{u}\|^2 O_p(1) \\ &\quad - a_n^* O_p(\delta_n) \|\mathbf{u}\| - b_n^* O_p(\delta_n^2 + \delta_n \rho_n) \|\mathbf{u}\|^2 \\ &\quad - p'_{\lambda_n}(0) O_p(\rho_n). \end{aligned} \quad (\text{B.25})$$

Notice that  $\delta_n = a_n^* + (\lambda_n \rho_n)^{1/2} + (L_{\max}/n)^{1/2}$  and  $b_n^* \rightarrow 0$  as  $n \rightarrow \infty$ . Then by choosing a sufficiently large  $C$ , the first term dominates the second and the third terms on the right-hand side of (B.25). On the other hand, according to (13), we have  $(\delta_n + (L_{\max}/n)^{1/2})^2 / p'_{\lambda_n}(0) \rho_n = O_p(1 + (L_{\max} + na_n^*)/\lambda_n \rho_n)$ . Hence, by choosing a sufficiently large  $C$ , (B.21) holds. This completes the proof of Lemma B.6.  $\square$

### C. Proof of Theorem 1

Let  $\gamma_n = n^{-1/2} + a_n$ . It is sufficient to prove that for any given  $\varepsilon > 0$ , there exists a large enough constant  $C$  such that

$$P \left\{ \inf_{\|\mathbf{u}\|=C} L_n^*(\alpha_0 + \gamma_n \mathbf{u}) > L_n^*(\alpha_0) \right\} \geq 1 - \varepsilon, \quad (\text{C.1})$$



which means  $L_n^*(\cdot)$  can reach a local minimum in the ball  $\{\alpha_0 + \gamma_n \mathbf{u} : \|\mathbf{u}\| \leq C\}$  with a probability of at least  $1 - \varepsilon$ . Thus, there exists a local minimizer satisfying  $\|\hat{\alpha} - \alpha_0\| = O_p(\gamma_n)$ .

Now we are going to prove (C.1). Note that

$$\begin{aligned} L_n^*(\alpha_0 + \gamma_n \mathbf{u}) - L_n^*(\alpha_0) \\ \geq L_n(\alpha_0 + \gamma_n \mathbf{u}) - L_n(\alpha_0) \\ + n \sum_{j=1}^s [p_{\lambda_n}(|\alpha_{0j} + \gamma_n u_j|) - p_{\lambda_n}(|\alpha_{0j}|)]. \end{aligned} \quad (\text{C.2})$$

Because  $U_n^*(\alpha_0) = -(\partial L_n(\alpha)/\partial \alpha)|_{\alpha=\alpha_0}$ , it similarly follows from (A.iv) that

$$\begin{aligned} \frac{1}{\sqrt{n}} \{L_n(\alpha_0 + \gamma_n \mathbf{u}) - L_n(\alpha_0)\} \\ \geq \widetilde{U}_n(\alpha_0) \gamma_n \mathbf{u} \{1 + o_p(1)\} \\ + \frac{1}{2} \mathbf{u}^T \sqrt{n} \Pi_n^*(\alpha_0) \mathbf{u} \{1 + o_p(1)\} \gamma_n^2. \end{aligned} \quad (\text{C.3})$$

Hence, we can obtain that

$$\begin{aligned} L_n^*(\alpha_0 + \gamma_n \mathbf{u}) - L_n^*(\alpha_0) \\ \geq -\sqrt{n} \widetilde{U}_n(\alpha_0) \gamma_n \mathbf{u} \{1 + o_p(1)\} \\ + \frac{1}{2} \mathbf{u}^T n \Pi_n^*(\alpha_0) \mathbf{u} \{1 + o_p(1)\} \gamma_n^2 \\ + n \sum_{j=1}^s \{\gamma_n p'_{\lambda_n}(|\alpha_{0j}|) \text{sgn}(\alpha_{0j}) u_j \\ + \gamma_n^2 p''_{\lambda_n}(|\alpha_{0j}|) \text{sgn}(\alpha_{0j}) u_j^2 (1 + o_p(1))\}. \end{aligned} \quad (\text{C.4})$$

Notice that  $\widetilde{U}_n(\alpha_0) = O_p(1)$ ,  $\Pi_n^*(\alpha_0) = O_p(1)$ . Hence, by choosing a large enough constant  $C$ ,  $-\sqrt{n} \widetilde{U}_n(\alpha_0) \gamma_n \mathbf{u} \{1 + o_p(1)\} + (1/2) \mathbf{u}^T n \Pi_n^*(\alpha_0) \mathbf{u} \{1 + o_p(1)\} \gamma_n^2 > 0$  uniformly in  $\|\mathbf{u}\| = C$ . Using the discussion similar to that in [7], the last two terms on the right-hand side of (C.4) are bounded by  $ns(\gamma_n a_n C + \gamma_n^2 b_n C^2)$ . It then follows from  $\lim_{n \rightarrow \infty} b_n = 0$  that when  $C$  is sufficiently large,  $L_n^*(\alpha_0 + \gamma_n \mathbf{u}) - L_n^*(\alpha_0) > 0$  uniformly holds in  $\|\mathbf{u}\| = C$ . Hence, (C.1) holds.

## D. Proof of Theorem 3

It follows from Lemma B.3 that part (1) holds. Now we are going to prove part (2). From part (1) we see that the local minimizer of  $L_n^*(\alpha)$  has the form of  $\hat{\alpha} = (\hat{\alpha}_1^T, 0_{d-s}^T)^T$  and satisfies  $\hat{\alpha} - \alpha_0 = O_p(n^{-1/2})$ . It follows from Taylor's expansion that

$$\begin{aligned} \frac{\partial L_n^*(\alpha)}{\partial \alpha} \Big|_{\hat{\alpha}} \\ = U_n^*(\hat{\alpha}) + n \text{diag}(\mathbf{b}^T, 0_{d-s}^T) \\ + n \left\{ \text{diag} \left[ \{p''_{\lambda_n}(|\alpha_{01}|)\}_{1,\dots,s}, 0_{d-s}^T \right] + o_p(1) \right\} (\hat{\alpha} - \alpha_0). \end{aligned} \quad (\text{D.1})$$

Then by some regular calculations we can obtain

$$\begin{aligned} \hat{\alpha} - \alpha_0 \\ = - \left\{ \frac{1}{n} \sum_{i=1}^n \int_0^\tau (V_i^*(W_i) - \bar{V})^{\otimes 2} Y_i(t) dt \right. \\ \left. + \text{diag} \left( \{p''_{\lambda_n}(|\alpha_{01}|)\}_{1,\dots,s}, 0_{d-s}^T \right)^T + o_p(1) \right\}^{-1} \\ \cdot \left\{ \frac{1}{\sqrt{n}} \widetilde{U}_n(\alpha_0) + \text{diag}(\mathbf{b}^T, 0_{d-s}^T) + o_p(n^{-1/2}) \right\}. \end{aligned} \quad (\text{D.2})$$

It follows from the proofs of Lemmas B.2 and B.3 that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{V_i^*(W_i) - \bar{V}\}^{\otimes 2} Y_i(t) dt \\ = \int_0^\tau \left\{ Y_{n20}(t) - \frac{Y_{n10}^{\otimes 2}(t)}{Y_{n00}(t)} \right\} dt + o_p(1), \end{aligned} \quad (\text{D.3})$$

where  $Y_{nkl}(t) = (1/n) \sum_{i=1}^n \{\widetilde{V}_i(W_i)\}^{\otimes k} \lambda^l(t | Z_i, V_i, W_i) Y_i(t)$  for  $k = 0, 1, 2$ ,  $l = 0, 1$ . Lemma B.1 and Assumptions (A.v) and (A.vi) yield

$$\frac{1}{n} \sum_{i=1}^n \int_0^\tau \{V_i^*(W_i) - \bar{V}\}^{\otimes 2} Y_i(t) dt = \mathbf{I}(\tau) + o_p(1). \quad (\text{D.4})$$

Next we consider  $\widetilde{U}_n(\alpha_0)$ . Using the martingale central limits theorem, we can prove that  $\widetilde{U}_n(\alpha_0)$  is asymptotically normally distributed with a mean of zero and covariance  $E[\widetilde{U}_n(\alpha_0)]^{\otimes 2}$ . Thus we obtain that

$$\begin{aligned} E[\widetilde{U}_n(\alpha_0)]^{\otimes 2} \\ = \lim_{n \rightarrow \infty} \int_0^\tau \left[ Y_{n21}(t) + \left\{ \frac{Y_{n10}(t)}{Y_{n00}(t)} \right\}^{\otimes 2} Y_{n01}(t) \right. \\ \left. - 2 \frac{Y_{n10}(t)}{Y_{n00}(t)} Y_{n11}(t) \right] dt = \Sigma(\tau). \end{aligned} \quad (\text{D.5})$$

Hence, by using Slutsky's Theorem, it follows from (D.2)–(D.5) that

$$\begin{aligned} \sqrt{n} \{\hat{\alpha}_1 - \alpha_{10} - (\mathbf{I}_1(\tau) + \Gamma)^{-1} \mathbf{b}\} \\ \xrightarrow{D} N(0, (\mathbf{I}_1(\tau) + \Gamma)^{-1} \Sigma_1(\tau) (\mathbf{I}_1(\tau) + \Gamma)^{-1}). \end{aligned} \quad (\text{D.6})$$

## E. Proof of Theorem 5

To prove part (a) of Theorem 5, we use proof by contradiction. Suppose that for an  $n$  sufficiently large there exists a constant  $\eta$ , such that with a probability of at least  $\eta$  there exists a  $k_0 > q$  such that  $\hat{\beta}_{k_0} \neq 0$ . Then  $\|\hat{\gamma}_{k_0}\|_{k_0} = \|\hat{\beta}_{k_0}\|_{L_2} > 0$ . Let  $\hat{\gamma}^*$  be

a vector constructed by replacing  $\hat{\gamma}_{k_0}$  with 0 in  $\hat{\gamma}$ . Then by the definition of  $\hat{\gamma}$ ,

$$\begin{aligned} PL_n(\hat{\gamma}) - PL_n(\hat{\gamma}^*) &= \{L_n(\hat{\gamma}) - L_n(\tilde{\gamma})\} - \{L_n(\hat{\gamma}^*) - L_n(\tilde{\gamma})\} \\ &\quad + np_{\lambda_n}(\|\hat{\gamma}_{k_0}\|_{k_0}) \\ &= n\{(\hat{\gamma} - \tilde{\gamma})^T \Xi_n(\hat{\gamma} - \tilde{\gamma}) - (\hat{\gamma}^* - \tilde{\gamma})^T \Xi_n(\hat{\gamma}^* - \tilde{\gamma})\} \\ &\quad \times (1 + o_p(1)) + np_{\lambda_n}(\|\hat{\gamma}_{k_0}\|_{k_0}). \end{aligned} \quad (E.1)$$

It then follows from Lemmas B.4–B.6 that the first term on the right-hand side of (E.1) is an order of  $O_p(n(\max\{(\lambda_n \rho_n)^{1/2}, a_n^*, (L_{\max}/n)^{1/2}\})^2)$ . Notice that  $p_{\lambda_n}(\|\hat{\gamma}_{k_0}\|_{k_0}) = p'_{\lambda_n}(0)\|\hat{\gamma}_{k_0}\|_{k_0}(1 + o_p(1)) = p'_{\lambda_n}(0)O_p(\max\{(\lambda_n \rho_n)^{1/2}, a_n^*, (L_{\max}/n)^{1/2}\})$ . Hence,

$$\begin{aligned} PL_n(\hat{\gamma}) - PL_n(\hat{\gamma}^*) &= nO_p\left(\left(\max\left\{(\lambda_n \rho_n)^{1/2}, a_n^*, \left(\frac{L_{\max}}{n}\right)^{1/2}\right\}\right)^2\right) \\ &\quad + np'_{\lambda_n}(0)O_p\left(\max\left\{(\lambda_n \rho_n)^{1/2}, a_n^*, \left(\frac{L_{\max}}{n}\right)^{1/2}\right\}\right) \\ &= \left\{O_p\left(\frac{\left\{(\lambda_n \rho_n)^{1/2}, a_n^*, (L_{\max}/n)^{1/2}\right\}}{\lambda_n}\right) + p'_{\lambda_n}(0)\lambda_n^{-1}\right\} \\ &\quad \times nO_p\left(\max\left\{(\lambda_n \rho_n)^{1/2}, a_n^*, \left(\frac{L_{\max}}{n}\right)^{1/2}\right\}\right)\lambda_n. \end{aligned} \quad (E.2)$$

Because  $\lambda_n \rightarrow 0$  and  $\lambda_n/\max\{\rho_n, a_n^*, (L_n/n)^{1/2}\} \rightarrow \infty$  as  $n \rightarrow \infty$ , (13) implies that  $PL_n(\hat{\gamma}) - PL_n(\hat{\gamma}^*) > 0$  with probability tending to 1. This contradicts the fact that  $PL_n(\hat{\gamma}) - PL_n(\hat{\gamma}^*) \leq 0$ . Thus, part (a) holds.

Lemmas B.5 and B.6 yield the proof of part (b).

## F. Proof of Theorem 8

Note that

$$\begin{aligned} E\left\{\sum_{k=0}^q \frac{\partial p_{\lambda}(\|\hat{\gamma}_k\|_k)}{\partial \gamma^{(1)}}\right\} &+ \frac{1}{n} \sum_{i=1}^n \int_0^{\tau} \left(\mathbf{U}_i^{(1)}(W_i) - \bar{\mathbf{U}}^{(1)}(t)\right)^T Z_i^{(1)T} Y_i(t) \\ &\cdot \left\{\bar{\beta}^{(1)}(W_i) - \beta^{(1)}(W_i)\right\} dt \end{aligned} \quad (F.1)$$

is the conditional mean of (24) given all observed covariates  $\{(X_i, Z_i, W_i)\}_{1 \leq i \leq n}$  and  $Y_i(t)$ ,  $i = 1, \dots, n$ . It follows from (24) that the expression given in (F.1) equals zero. Hence, together with (24), we have

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \int_0^{\tau} \left\{\mathbf{U}_i^{(1)}(W_i) - \bar{\mathbf{U}}^{(1)}(t)\right\}^T dM_i(t) (1 + O_p(n^{-1/2})) \\ &= \sum_{k=0}^q \left[ \frac{\partial p_{\lambda}(\|\hat{\gamma}_k\|_k)}{\partial \gamma^{(1)}} - E\left\{\frac{\partial p_{\lambda}(\|\hat{\gamma}_k\|_k)}{\partial \gamma^{(1)}}\right\} \right] \\ &\quad + \frac{1}{n} \sum_{i=1}^n \int_0^{\tau} \left\{\mathbf{U}_i^{(1)}(W_i) - \bar{\mathbf{U}}^{(1)}(t)\right\}^{\otimes 2} Y_i(t) \{\hat{\gamma}^{(1)} - \bar{\gamma}^{(1)}\} dt. \end{aligned} \quad (F.2)$$

By using similar quadratic approximation in [10], we have

$$\begin{aligned} p_{\lambda}(\|\gamma_k\|_k) &\approx p_{\lambda}(\|\beta_k\|_{L_2}) + \frac{1}{2} \frac{p'_{\lambda}(\|\beta_k\|_{L_2})}{\|\beta_k\|_{L_2}(\|\gamma_k\|_k - \|\beta_k\|_{L_2})}, \\ &\sum_{k=0}^q \left[ \frac{\partial p_{\lambda}(\|\hat{\gamma}_k\|_k)}{\partial \gamma^{(1)}} - E\left\{\frac{\partial p_{\lambda}(\|\hat{\gamma}_k\|_k)}{\partial \gamma^{(1)}}\right\} \right] \\ &= \text{diag}\left\{\frac{1}{2} \frac{p'_{\lambda}(\|\beta_k\|_{L_2})}{\|\beta_k\|_{L_2}} \mathbf{R}_k\right\}_{0 \leq k \leq q} \\ &\quad \times (\hat{\gamma}^{(1)} - \bar{\gamma}^{(1)})(1 + o_p(1)) \\ &= L_n^{-1} \Gamma^* (\hat{\gamma}^{(1)} - \bar{\gamma}^{(1)})(1 + o_p(1)). \end{aligned} \quad (F.3)$$

Finally, by using the arguments similar to the proof of Theorem 4.1 of [26] and the proof of [24], (F.2) and the martingale central theory, we know that

$$\begin{aligned} &\sqrt{nL_n^{-1}} \{d_n^T (\Xi_1 + \Gamma^*)^{-1} \Omega^* (\Xi_1 + \Gamma^*)^{-1} d_n\}^{-1/2} \\ &\quad \times d_n^T (\hat{\gamma}^{(1)} - \bar{\gamma}^{(1)}) \xrightarrow{L} N(0, 1) \end{aligned} \quad (F.4)$$

holds for any vector  $d_n$  with dimension  $\sum_{k=0}^q L_k$  and components not all 0. Then for any  $(q+1)$ -dimension vector  $d_n$  whose components are not all 0 and for any given  $w$  in  $\mathcal{W}$ , choosing  $d_n^* = (\mathbf{B}^{(1)}(w))^T d_n$  yields

$$\begin{aligned} &\sqrt{nL_n^{-1}} \left\{d_n^{*T} \mathbf{B}^{(1)}(w) (\Xi_1 + \Gamma^*)^{-1} \Omega^* (\Xi_1 + \Gamma^*)^{-1} \right. \\ &\quad \times \left. (\mathbf{B}^{(1)}(w))^T d_n^* \right\}^{-1/2} \\ &\quad \times d_n^{*T} \left\{\hat{\beta}^{(1)}(w) - \bar{\beta}^{(1)}(w)\right\} \xrightarrow{L} N(0, 1), \end{aligned} \quad (F.5)$$

which implies (25) holds.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

Ma's research was supported by the National Natural Science Foundation of China (NSFC) Grant no. 11301424 and the Fundamental Research Funds for the Central Universities Grant no. JBK120405.

## References

- [1] J. Huang, "Efficient estimation of the partly linear additive Cox model," *The Annals of Statistics*, vol. 27, no. 5, pp. 1536–1563, 1999.
- [2] J. Cai, J. Fan, J. Jiang, and H. Zhou, "Partially linear hazard regression for multivariate survival data," *Journal of the American Statistical Association*, vol. 102, no. 478, pp. 538–551, 2007.
- [3] J. Fan, H. Lin, and Y. Zhou, "Local partial-likelihood estimation for lifetime data," *The Annals of Statistics*, vol. 34, no. 1, pp. 290–325, 2006.
- [4] L. Tian, D. Zucker, and L. J. Wei, "On the Cox model with time-varying regression coefficients," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 172–183, 2005.
- [5] G. S. Yin, H. Li, and D. L. Zeng, "Partially linear additive hazards regression with varying coefficients," *Journal of the American Statistical Association*, vol. 103, no. 483, pp. 1200–1213, 2008.
- [6] J. Fan and R. Li, "Variable selection for Cox's proportional hazards model and frailty model," *The Annals of Statistics*, vol. 30, no. 1, pp. 74–99, 2002.
- [7] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [8] B. A. Johnson, D. Y. Lin, and D. Zeng, "Penalized estimating functions and variable selection in semiparametric regression models," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 672–680, 2008.
- [9] B. A. Johnson, "Variable selection in semiparametric linear regression with censored data," *Journal of the Royal Statistical Society B. Statistical Methodology*, vol. 70, no. 2, pp. 351–370, 2008.
- [10] L. Wang, H. Li, and J. Z. Huang, "Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements," *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1556–1569, 2008.
- [11] P. Du, S. Ma, and H. Liang, "Penalized variable selection procedure for Cox models with semiparametric relative risk," *The Annals of Statistics*, vol. 38, no. 4, pp. 2092–2117, 2010.
- [12] D. Y. Lin and Z. Ying, "Semiparametric analysis of the additive risk model," *Biometrika*, vol. 81, no. 1, pp. 61–71, 1994.
- [13] F. W. Huffer and I. W. McKeague, "Weighted least squares estimation for aalen's additive risk model," *Journal of the American Statistical Association*, vol. 86, no. 413, pp. 114–129, 1991.
- [14] O. O. Aalen, "A linear regression model for the analysis of life times," *Statistics in Medicine*, vol. 8, no. 8, pp. 907–925, 1989.
- [15] T. H. Scheike, "The additive nonparametric and semiparametric Aalen model as the rate function for a counting process," *Lifetime Data Analysis*, vol. 8, no. 3, pp. 247–262, 2002.
- [16] I. W. McKeague and P. D. Sasieni, "A partly parametric additive risk model," *Biometrika*, vol. 81, no. 3, pp. 501–514, 1994.
- [17] H. Li, G. Yin, and Y. Zhou, "Local likelihood with time-varying additive hazards model," *The Canadian Journal of Statistics*, vol. 35, no. 2, pp. 321–337, 2007.
- [18] J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, Hoboken, NJ, USA, 2nd edition, 2002.
- [19] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability*, Chapman & Hall, London, UK, 1996.
- [20] L. L. Schumaker, *Spline Functions: Basic Theory*, Cambridge University Press, Cambridge, UK, 3rd edition, 2007.
- [21] C. J. Stone, "Optimal global rates of convergence for nonparametric regression," *The Annals of Statistics*, vol. 10, no. 4, pp. 1040–1053, 1982.
- [22] J. Fan and T. Huang, "Profile likelihood inferences on semiparametric varying-coefficient partially linear models," *Bernoulli*, vol. 11, no. 6, pp. 1031–1057, 2005.
- [23] W. A. Knaus, F. E. Harrell Jr., J. Lynn et al., "The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults," *Annals of Internal Medicine*, vol. 122, no. 3, pp. 191–203, 1995.
- [24] J. Z. Huang, C. O. Wu, and L. Zhou, "Polynomial spline estimation and inference for varying coefficient models with longitudinal data," *Statistica Sinica*, vol. 14, no. 3, pp. 763–788, 2004.
- [25] J. Cai, J. Fan, H. Zhou, and Y. Zhou, "Hazard models with varying coefficients for multivariate failure time data," *The Annals of Statistics*, vol. 35, no. 1, pp. 324–354, 2007.
- [26] J. Z. Huang, "Local asymptotics for polynomial spline regression," *The Annals of Statistics*, vol. 31, no. 5, pp. 1600–1635, 2003.