A SEMIPARAMETRICALLY EFFICIENT ESTIMATOR OF SINGLE-INDEX VARYING COEFFICIENT COX PROPORTIONAL HAZARDS MODELS

Huazhen Lin, Ming T. Tan and Yi Li

Southwestern University of Finance and Economics,
Georgetown University and University of Michigan

Abstract: This paper proposes a single-index varying coefficient hazards model to identify biomarkers for risk stratification and treatment selection for individual patients. Our model accommodates multiple predictive biomarkers and allows for flexible nonlinear interactions between the multiple biomarkers and the treatment. We propose a global partial likelihood to estimate the varying-coefficient functions and the regression coefficients. The proposed estimators are shown to be consistent, asymptotically normal and semiparametrically efficient. The proposed approach is applied to a clinical trial on multiple myeloma patients for risk stratification and to investigate whether biomarkers would interact with treatment for each individual patient.

Key words and phrases: Cox proportional hazards model, global partial likelihood, semiparametric efficiency, single-index, varying coefficients.

1. Introduction

Biomarkers have emerged as promising tools for risk assessment and stratification of patients (e.g. stage or subtype) with chronic diseases, such as cancer and cardiovascular disorders, which is essential in guiding the management and treatment of disease in order to achieve the optimal clinical outcome. For example, Sargent et al. (2005) found that the effect of treatment on patient's survival may depend on the level of individual biomarkers, which cannot directly be detected with the routinely used Cox proportional hazards model that ignores the inherent nonlinear heterogeneities defined by biomarkers on the effect of treatment. In a broader context, identifying nonlinear interactions between biomarkers and treatment has become a topical area with the recent precision medicine initiative (`http://www.nih.gov/precisionmedicine/`). Precision medicine seeks effective data-driven approaches for disease treatment and prevention that takes into account individual variability in personal characteristics, including biomarkers. There is an urgent need for statistical models that can facilitate the identification

of biomarkers that affect patients response to treatment with an unknown form, and lead to results that will be useful for further validations.

Such analyses were needed for a clinical trial conducted by the University of Arkansas, where newly diagnosed multiple myeloma patients were randomized to receive two treatments, total therapy II and total therapy III (Shaughnessy et al. (2007)). Several biomarkers, including patients' serum beta2-microglobulin and albumin levels were collected to predict the disease status and response to treatment. The marginal effect of the treatment was not significant as reported by previous studies; however it was of substantial interest to investigate whether the treatment would be effective for some subset of patients as defined by these biomarkers, termed predictive biomarkers. Several existing methods are available to assess the clinical utility of predictive biomarkers. An ad hoc approach is to conduct survival comparisons across two biomarker-based groups. As the method requires dichotomization of predictive biomarkers, it does not adequately quantify the clinical utility of predictive biomarkers. Freidlin and Simon (2005) proposed a design that combines prospective development of a gene expression-based classifier to select sensitive patients with a properly powered test for overall effect, but again requires dichotomization of the biomarker. Jiang, Freidlin, and Simon (2007) extended Freidlin and Simon's design to allow a continuous-scale biomarker by introducing a biomarker-adaptive threshold.

There are two limitations of Freidlin and Simon (2005) and Jiang, Freidlin, and Simon (2007). First, both methods assume that the effect of the interaction between the biomarker and the treatment is a step-function with only one jump. However, when the biomarker is continuous, it is most likely that the interaction continuously varies with the value of the biomarker. Second, both methods assume that the effect of treatment depends on only one specific biomarker. In practice, as in our motivating dataset, there may exist more than one biomarker that provides information for risk stratification and treatment selection. If the effect of the treatment does continuously depend on multiple biomarkers, then it is crucial to know how to combine the multiple biomarkers and how the effect of treatment varies with the multiple biomarkers in order to predict the treatment effect of a new subject with high accuracy.

To address these issues, we propose a single-index varying coefficients Cox model,

$$\lambda(t) = \lambda_0(t) \exp \Big\{ \sum_{k=1}^{d} \beta_k(\alpha' \mathbf{X}) Z_k \Big\}, \tag{1.1}$$

where $\mathbf{X}$ denotes the multiple biomarker vector, $\mathbf{Z} = (Z_1, \ldots, Z_d)'$ is the exposure variable, for example, treatment group indicator, $\beta(\cdot) = (\beta_1(\cdot), \ldots, \beta_d(\cdot))'$ is the unknown varying-coefficient function and characterizes how the effect varies with the biomarkers, and α is an unknown regression coefficient vector used to

combine the multiple biomarkers that potentially define pathophysiology of a chronic disease, for example, disease staging and subtyping. If $\beta(\cdot)$ is discrete, the proposal models how each stage or subtype of disease interacts with the treatment. When $\beta(\cdot)$ is continuous, as in our formulation, it evaluates the impact of the disease stage or subtype on the efficacy of treatment on a continuous scale, an extension from discretization to a continuous spectrum.   There are several additional advantages. First, α and $\beta(\cdot)$ can identify patients who are more likely to benefit from a given treatment and hence provide a personalized treatment strategy. Second, as the unknown functions $\lambda_0(\cdot)$ and $\beta(\cdot)$ are one-dimensional, the so-called curse of dimensionality is avoided. Finally, our model is general, encompassing many well-known models as special cases, as follows.

(i) When β is a constant or linear, (1.1) is the Cox proportional hazards model (Cox (1972)).

(ii) When $\mathbf{Z} = 1$ and $\mathbf{X}$ is one-dimensional, the model is a nonparametric Cox model. Statistical methods, such as nearest neighbor, spline, and local polynomial smoothing methods have been developed for the nonparametric Cox model; see Tibshirani and Hastie (1987), O'Sullivan (1988), Hastie and Tibshirani (1990), Gentleman and Crowley (1991), Kooperberg, Stone, and Truong (1995), Fan, Gijbels, and King (1997), Chen and Zhou (2007), and Chen et al. (2010), among others.

(iii) When $\mathbf{X}$ is time, (1.1) is the time-varying Cox model. The penalized method, sieve method, and local linear method have been proposed to estimate the time-varying Cox model; see Zucker and Karr (1990), Murphy and Sen (1991), Gamerman (1991), Murphy (1993), Marzec and Marzec (1997), Martinussen et al. (2002), Valsecchi, Silvestri, and Sasieni (1996), Cai and Sun (2003), and Tian, Zucker, and Wei (2005).

(iv) When $\mathbf{X}$ is a one-dimensional covariate, (1.1) reduces to the model studied by Fan, Lin, and Zhou (2006), and Chen, Lin, and Zhou (2012).

(v) When $Z_1 = 1$ and $\beta_2, \cdots, \beta_d$ are constants, the model reduces to the partially linear Cox model. Spline and local linear smoothing have been proposed to estimate the partially linear Cox model; see Huang (1999) and Chen et al. (2010), among others.

As far as we know, there is no successful extension to the single-index varying coefficients Cox model (1.1). The existing methods estimate either the unknown $\beta(\cdot)$ or the index-vector α with the assumption that α or $\beta(\cdot)$ is known, respectively. Particular, if $\beta(\cdot)$ is known, the index-vector α can be estimated by maximizing the partial likelihood function (2.1) (Cox (1972)). If α is known, the unknown $\beta(\cdot)$ can generally be estimated by using regression splines, penalized

and local polynomial methods. The spline and penalized methods simultane-
ously estimate the unknown functions. This optimization problem can be rather
complex due to a large number of parameters, especially when the number of the
nonparametric functions, d, is medium or large. In addition, sampling properties
of the spline and penalized methods remain elusive. A useful alternative is the
local polynomial technique. However, a local partial likelihood method (Fan, Lin,
and Zhou (2006)) uses only local observations, incurring efficiency loss. Here, we
propose a global partial likelihood method to estimate the unknown functions
and the regression coefficients. The proposed estimators are shown to be consis-
tent and asymptotically normal. The utility of the proposed method is further
enhanced by its semiparametric efficiency for α and $\beta(\cdot)$.

The paper is organized as follows. Section 2 introduces the estimators of $\beta(\cdot)$
and α, and Section 3 establishes the uniform consistency, asymptotic normality,
and semiparametric efficiency. The procedure is extended to a mixed model with
fixed and varying coefficients in Section 4. Numerical simulation and examples
are given in Sections 5 and 6. Technical proofs are relegated to the Appendix.

2. Partial Likelihood Estimation

Suppose that there is a random sample of size n from an underlying pop-
ulation. For the ith individual, let T_i denote the failure time, C_i the censoring
time, and $\mathcal{T}_i = \min(T_i, C_i)$ the observed time, Δ_i be an indicator that equals 1
if $\mathcal{T}_i$ is a failure time and 0 otherwise. Assume that T_i and C_i are independent
of each other given covariates $\mathbf{Z}_i$ and $\mathbf{X}_i$. The covariates $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{id})'$ and
$\mathbf{X}_i$ are allowed to be time dependent. Following Fan, Lin, and Zhou (2006), we
drop the dependence of covariates on time t, with the understanding that the
methods and proofs in this paper are applicable to time dependent covariates.
The observed data structure is $\{\mathcal{T}_i, \Delta_i, \mathbf{Z}_i, \mathbf{X}_i\}$ for $i = 1, \cdots, n$.

When all the observations are independent, the partial likelihood for (1.1) is

$$L(\beta, \alpha) = \prod_{i=1}^{n} \Big[\frac{\exp\{\beta(\mathbf{X}_i'\alpha)'\mathbf{Z}_i\}}{\sum_{k \in \mathcal{R}(\mathcal{T}_i)} \exp\{\beta(\mathbf{X}_k'\alpha)'\mathbf{Z}_k\}} \Big]^{\Delta_i}, \qquad (2.1)$$

where $R(t) = \{i : \mathcal{T}_i \geq t\}$ is the set of the individuals at risk just prior to time t. If
the unknown functions $\beta(\cdot)$ are parameterized, the parameters can be estimated
by maximizing (2.1).

Since the forms of $\beta(\cdot)$ are not specified, we first consider the estimator of
$\beta(\cdot)$ given α. Suppose that each component of $\beta(\cdot)$ is smooth and admits a Taylor
expansions: for each given v and w around v,

$$\beta(w) \approx \beta(v) + \dot{\beta}(v)'(w - v) \equiv \delta_1 + \delta_2'(w - v), \qquad (2.2)$$

where $\dot{f}(v) = df(v)/dv$ for any function or vector of functions $f(\cdot)$. Given α, (1.1) reduces to the model proposed by Fan, Lin, and Zhou (2006). Hence, we can estimate $\beta(\cdot)$ by their local partial likelihood method: estimate $\delta = (\delta_1', \delta_2')'$ by solving the local partial score function,

$$\sum_{i=1}^{n} \Delta_i K_i(v) \left\{ \mathbf{V}_i(v) - \frac{\sum_{k \in \mathcal{R}(\mathcal{T}_i)} K_k(v) \exp\{\mathbf{V}_k(v)'\delta\} \mathbf{V}_k(v)}{\sum_{k \in \mathcal{R}(\mathcal{T}_i)} K_k(v) \exp\{\mathbf{V}_k(v)'\delta\}} \right\} = 0, \qquad (2.3)$$

where $\mathbf{V}_i(v) = (\mathbf{Z}_i', \mathbf{Z}_i'(\mathbf{X}_i'\alpha - v))'$, $K_i(v) = K_h(\mathbf{X}_i'\alpha - v)$, $K_h(x) = K(x/h)/h$, K is a one-dimensional kernel density function, and h represents the size of the local neighborhood. The local partial score function (2.3) is a partial score based on observations with $\mathbf{X}_i'\alpha$ in a small neighborhood of v. The method is properly motivated and rather simple to implement and analyze. However, the localization of (2.3) suffers a loss of efficiency since the observations outside of the neighborhood of v, which carry information about $\beta(v)$, are not used. Moreover, the intercept of $\beta(v)$ cannot be directly estimated as it is cancelled out of the local partial likelihood. Although the intercept can be estimated by subsequently integrating the estimate of its derivative, the large sample property of this estimate is not formally established and statistical inference is not immediately available. Finally, the local partial likelihood approach cannot handle discrete $\mathbf{X}_i$.

We propose a global partial likelihood method to estimate $\beta(\cdot)$. The motivation of global partial likelihood is quite straightforward. For every fixed v in the range of $\mathbf{X}'\alpha$, suppose $\beta(\cdot)$ is known outside a neighborhood of v, denoted by $B_n(v)$. Then, the partial likelihood function (2.1) can be written as

$$\prod_{i=1}^{n} \left\{ \frac{I_i \exp(\overline{\psi}_i) + (1 - I_i) \exp(\psi_i)}{\sum_{j \in \mathcal{R}(\mathcal{T}_i)} \{I_j \exp(\overline{\psi}_j) + (1 - I_j) \exp(\psi_j)\}} \right\}^{\Delta_i}, \qquad (2.4)$$

where $\overline{\psi}_i = \{\delta_1 + \delta_2(\mathbf{X}_i'\alpha - v)\}' \mathbf{Z}_i$, $\psi_i = \beta(\mathbf{X}_i'\alpha)'\mathbf{Z}_i$ and I_i equals 1 if $\mathbf{X}_i'\alpha \in B_n(v)$ and equals 0 otherwise. Since the true $\beta(\cdot)$ outside of a neighborhood of v is unknown, (2.4) is not directly useful. A key idea of the proposed method is replacing $\beta(\cdot)$ outside of a neighborhood of v by the estimators from the previous step using a iteration algorithm. With the refinement of local linear smoothing and some slight variation for computational convenience, we derive an iteration algorithm. As $\beta_1(\cdot)$ is identifiable up to a location shift and α is identifiable up to a scale shift, we set $\beta_1^{(m)}(\mathbf{X}_n'\alpha) = 0$ and $\alpha_1^{(m)} = 1$, the first element of α, for all $m \geq 0$ for identification as well as notational and computational convenience.

Step 0. Choose initial values of $\alpha^{(0)}$ and function $\beta^{(0)}(v)$ for $v = \mathbf{X}_1'\alpha^{(0)}, \ldots,$ $\mathbf{X}_{n-1}'\alpha^{(0)}$, for example, with the iterative method between the local partial likelihood estimate of $\beta(\cdot)$ (Fan, Lin, and Zhou (2006)) and the partial likelihood

estimate of α. Using the proof of Fan, Lin, and Zhou (2006) and the uniform law of large numbers (Pollard (1990)), it can be shown that these initial estimators are consistent. More details are given in the Appendix.

The first step of Step m. For every given $v = \mathbf{X}_1'\alpha^{(m-1)}, \ldots, \mathbf{X}_{n-1}'\alpha^{(m-1)}$, maximizing (2.4) with respect to $\delta = (\delta_1', \delta_2')'$ and replacing I_i with a kernel function K_i that decreases smoothly to zero, obtain the equations for δ:

$$
\sum_{i=1}^{n} \Delta_i \left[\mathbf{V}_i^{(m-1)}(v) K_i^{(m-1)}(v) \right.
$$
$$
\left. - \frac{\sum_{j=1}^{n} Y_j(\mathcal{T}_i) \mathbf{V}_j^{(m-1)}(v) K_j^{(m-1)}(v) \exp\{\mathbf{V}_j^{(m-1)}(v)'\delta\}}{\sum_{j=1}^{n} Y_j(\mathcal{T}_i) \exp\{\beta^{(m-1)}(\mathbf{X}_j'\alpha^{(m-1)})'\mathbf{Z}_j\}} \right] = 0, \qquad (2.5)
$$

where $\mathbf{V}_i^{(m)}(v) = (\mathbf{Z}_i', \mathbf{Z}_i'(\mathbf{X}_i'\alpha^{(m)} - v))'$, $K_i^{(m)}(v) = K_h(\mathbf{X}_i'\alpha^{(m)} - v)$, $Y_i(t) = I(\mathcal{T}_i \geq t)$. Using the counting process notation, (2.5) can be expressed as

$$
\sum_{i=1}^{n} \int_0^{\tau} \left[\mathbf{V}_i^{(m-1)}(v) K_i^{(m-1)}(v) \right.
$$
$$
\left. - \frac{\sum_{j=1}^{n} Y_j(t) \mathbf{V}_j^{(m-1)}(v) K_j^{(m-1)}(v) \exp\{\mathbf{V}_j^{(m-1)}(v)'\delta\}}{\sum_{j=1}^{n} Y_j(t) \exp\{\beta^{(m-1)}(\mathbf{X}_j'\alpha^{(m-1)})'\mathbf{Z}_j\}} \right] dN_i(t) = 0, \qquad (2.6)
$$

where $N_i(t) = I(T_i \leq t, \Delta_i = 1)$. In this, τ is often assumed to be finite in the literature to avoid the tail problem. Let $\widehat{\delta}_1(v)$ and $\widehat{\delta}_2(v)$ be the solutions in δ_1 and δ_2. Then $\beta^{(m)}(v) = \widehat{\delta}_1(v)$ and $\dot{\beta}^{(m)}(v) = \widehat{\delta}_2(v)$ for $v = \mathbf{X}_1'\alpha^{(m-1)}, \ldots, \mathbf{X}_{n-1}'\alpha^{(m-1)}$.

The second step of Step m. For given $\beta^{(m)}$ and $\dot{\beta}^{(m)}$, estimate α by solving the partial score function,

$$
\sum_{i=1}^{n} \Delta_i \left[\mathbf{X}_i \dot{\beta}^{(m)}(\mathbf{X}_i'\alpha^{(m-1)})'\mathbf{Z}_i \right.
$$
$$
\left. - \frac{\sum_{k \in \mathcal{R}(\mathcal{T}_i)} \mathbf{X}_k \dot{\beta}^{(m)}(\mathbf{X}_k'\alpha^{(m-1)})'\mathbf{Z}_k \exp\{\beta^{(m)}(\mathbf{X}_k'\alpha)'\mathbf{Z}_k\}}{\sum_{k \in \mathcal{R}(\mathcal{T}_i)} \exp\{\beta^{(m)}(\mathbf{X}_k'\alpha)'\mathbf{Z}_k\}} \right] = 0. \qquad (2.7)
$$

Repeat this iteration procedure until convergence. Comparing the proposed partial likelihood score (2.5) and the local partial likelihood function (2.3), we can see that the local partial likelihood method is based on the estimating equation of $(\delta_1 + \delta_2(\mathbf{X}_i'\alpha - v))' \mathbf{Z}_i$ with local data, while the partial likelihood method is based on the estimating equation of $\mathbf{V}_i(v)K_i(v)$ using all the data. Moreover, the denominator in (2.5) utilizes all the data. So does the estimate of $\beta(v)$. Hence, this is literally a global estimation compared to the local estimation in (2.3). The semiparametric efficiency of the proposal is presented in Theorem 4 in Section 3.

3. Large Sample Properties

In this Section, we establish the uniform consistency and asymptotic normality of the proposed estimators. Without loss of generality, we fix $\widehat{\beta}_1(0) = \beta_1(0) = 0$, where β_1 is the first element of $\beta(\cdot)$, and assume that the support of $\mathbf{X}'\alpha$ is $[0, 1]$. Additional regularity conditions and proofs of Theorems 1–4 are given in the Appendix.

Theorem 1. *Under Conditions* 1−7 *stated in the Appendix, we have*

$$\sup_{v \in [0,1]} \|\widehat{\beta}(v) - \beta(v)\| \to 0 \ \ and \ \ \|\widehat{\alpha} - \alpha_0\| \to 0 \ \ in \ probability,$$

where α_0 is the true value of α.

Theorem 2. *Under Conditions* 1−7 *stated in the Appendix, if $nh^4 \to 0$, we have*

$$\sqrt{n}\,(\widehat{\alpha} - \alpha_0) \to N(0, \Delta), \tag{3.1}$$

where

$$\Delta = \mathbf{D}_1^{-1} \int_0^\tau E\left[\xi_i^2(t)P(t|\mathbf{Z}_i, \mathbf{X}_i)\exp\{\beta(\mathbf{X}_i'\alpha_0)'\mathbf{Z}_i\}\right]\lambda_0(t)dt(\mathbf{D}_1^{-1})',$$

$$\xi_i(t) = \int_0^1 g(v)\frac{s_{10}(t;v)}{s_{00}(t)}dv + \left\{\int_0^1 g(v)\mathbf{D}_2(v)\mathbf{D}_1^{-1}dv - I\right\}\left\{\mathbf{X}_iW_i - \frac{r_0(t)}{s_{00}(t)}\right\}$$
$$-\mathbf{Z}_ig(\mathbf{X}_i'\alpha_0),$$

$\mathbf{D}_1$, $\mathbf{D}_2$, $P(t|\mathbf{z}, \mathbf{x})$, $g(\cdot)$, $s_{10}(\cdot)$, $s_{00}(\cdot)$, $r_0(\cdot)$ *are defined in the Appendix, and* $W_i = \dot{\beta}(\mathbf{X}_i'\alpha_0)'\mathbf{Z}_i$.

To estimate a parameter at the rate $n^{-1/2}$, one must undersmooth the nonparametric part. Undersmoothing to obtain usual parametric rates of convergences is standard in the kernel literature and has analogs in the spline literature (Carroll et al. (1997)); Hastie and Tibshirani (1990)). This is achieved by $nh^4 = o(1)$, and is required by Theorem 2 to estimate parameters α at the rate $n^{-1/2}$.

Theorem 3. *Under Conditions* 1−7 *stated in the Appendix, if $nh^5 = O(1)$, for $v \in (0, 1)$,*

$$(nh)^{1/2}\left\{\widehat{\beta}(v) - \beta(v) - \frac{1}{2}h^2\mu_2(I - \mathcal{A})^{-1}(\ddot{\beta})(v)\right\} \xrightarrow{\mathcal{D}} N(0, \mathbf{V}(v)), \tag{3.2}$$

where I is the identity operator and $\mathcal{A}$ is the linear operator satisfying $\mathcal{A}(\phi)(v) = \mathbf{\Sigma}^{-1}(v)\int_w \Psi(w; v)\phi(w)dw$ for any function ϕ, $\mathbf{\Sigma}(v)$ and $\Psi(w; v)$ are defined in the Appendix, $\ddot{\beta}(v) = d^2\beta(v)/dv^2$, $\mathbf{V}(v) = \nu_0[(I - \mathcal{A})^{-1}(\mathbf{\Sigma}^{-1/2})(v)][(I - \mathcal{A})^{-1}(\mathbf{\Sigma}^{-1/2})(v)]'$, and $\nu_0 = \int_v K^2(v)dv$.

Theorem 3 states that $\widehat{\beta}(v) - \beta(v)$ is asymptotically normal, the order of the asymptotic bias of $\widehat{\beta}(v) - \beta(v)$ is h^2, and the order of the asymptotic covariance is $(nh)^{-1}$. As a consequence, the theoretically optimal bandwidth $O(n^{-1/5})$ in the nonparametric method can be taken.

When α is known, our model is the varying-coefficient Cox model considered by Fan, Lin, and Zhou (2006) and Chen, Lin, and Zhou (2012), and Theorem 3 is Theorem 2 of Chen, Lin, and Zhou (2012). The difference between the proposed single-index varying coefficient Cox proportional model and the varying coefficient Cox model of Chen, Lin, and Zhou (2012) is the same as that between the single-index model and the simple nonparametric regression model. The proposed single-index varying coefficient Cox model is a key tool to handle multiple predictive biomarkers and allows the exposure variable to interact non-linearly with multiple covariates. However, due to the presence of α, we need to investigate the asymptotic relationship between $\widehat{\alpha}$ and $\widehat{\beta}(\cdot)$ when establishing asymptotic properties; this is not trivial when the asymptotic expansion of $\widehat{\beta}(\cdot)$ is expressed as an integral equation and not in a closed form.

Theorem 2 shows that $\widehat{\alpha}$ is a $n^{1/2}-$consistent and asymptotically normal estimator of α, while the next result shows that $\widehat{\alpha}$ also is an efficient estimator of α. For any function $\phi(w) = (\phi'_1, \phi_2(w)')'$, with a continuous second derivative on $(0, 1)$, let $\phi'_1 \widehat{\alpha} + \int_0^1 \phi'_2(w) \widehat{\beta}(w) dw$ be an estimator of $\phi'_1 \alpha_0 + \int_0^1 \phi'_2(w) \beta(w) dw$.

Theorem 4. *Under the conditions of Theorem* 1, *if* $nh^4 \to 0$ *and* $nh^2 \to \infty$, *then*

$$\phi'_1 \widehat{\alpha} + \int_0^1 \phi'_2(w) \widehat{\beta}(w) dw \ \ is \ an \ efficient \ estimator \ of \ \phi'_1 \alpha_0 + \int_0^1 \phi'_2(w) \beta(w) dw.$$

By taking $\phi_2(t) = 0$, we have that $\widehat{\alpha}$ is an efficient estimator of α_0; taking $\phi_1(t) = 0$, $\int_0^\tau \phi'_2(t) \widehat{\beta}(t) dt$ is an efficient estimator of $\int_0^\tau \phi'_2(t) \beta(t) dt$.

With estimators of β and α, we can use kernel smoothing (Fan, Lin, and Zhou (2006)) to estimate the baseline hazard function by $\widehat{\lambda}_0(t) = \int K_b(t - u) d\widehat{\Lambda}_0(u)$, where b is a given bandwidth and

$$\widehat{\Lambda}_0(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{dN_i(u)}{n^{-1} \sum_{j=1}^n Y_j(u) \exp\{\widehat{\beta}(\widehat{\alpha}' \mathbf{X}_j)' \mathbf{Z}_j\}}. \tag{3.3}$$

Using Theorem 2 and Corollary 1 and the proof of Fan, Lin, and Zhou (2006), we can show that $\widehat{\lambda}_0(t)$ and $\widehat{\Lambda}_0(t)$ are uniformly consistent on $(0, \tau)$.

Though the estimators of the variances of $\widehat{\alpha}$ and $\widehat{\beta}(v)$ are available, the computation involves the unknown functions $s_{00}(t), s_{10}(t; v), s_{20}(t; v), r_0(t)$, and $g(\cdot)$, that are related to the unknown function $\beta(\cdot)$ and its derivative; theses estimates are difficult to obtain. Since $g(\cdot)$ is defined by the integral equation:

$\mathbf{D}_3(w) = g(w)\mathbf{\Sigma}(w) - \int_0^1 g(v)\Psi(w;v)dv$, its computation even more difficult. An effective remedy is to leverage a resampling scheme to approximate the variances or covariance matrices. Simulations have indicated good performance of a bootstrap approach.

4. Estimation of the Fixed and Varying Regression Coefficients

As the effects of some covariates may be constant, a mixed model with both fixed and varying coefficients is desirable. The mixed model is less restrictive than the standard proportional hazards model and simpler than model (1.1). Moreover, it is possible to obtain a $n^{1/2}$-consistent estimator for the fixed coefficient in the mixed model. Martinussen et al. (2002), Tian, Zucker, and Wei (2005) and Chen, Lin, and Zhou (2012) examined a mixed model in the setting of the traditional varying coefficient Cox model where $\beta(\cdot)$ is a function of the time t or a exposure variable, and McKeague and Sasieni (1994) studied a similar mixed model in the setting of an additive hazard model. We consider the mixed model

$$\lambda(t) = \lambda_0(t)\exp\{\theta'\mathbf{Z}_1 + \beta(\mathbf{X}'\alpha)'\mathbf{Z}_2\}. \tag{4.1}$$

For identifiability, we set $\beta_1^{(m)}(\mathbf{X}_n'\alpha) = 0$ and $\alpha_1^{(m)} = 1$ for all $m \geq 0$. Based on the idea of the global partial likelihood, we estimate θ, α, and $\beta(\cdot)$ using an iterative algorithm.

Step 0. Choose initial values of $\theta^{(0)}$, $\alpha^{(0)}$, and $\beta^{(0)}(w)$ for $w = \mathbf{X}_1'\alpha^{(0)}, \ldots, \mathbf{X}_{n-1}'\alpha^{(0)}$.

The first step of Step m. For every given $v = \mathbf{X}_1'\alpha^{(m-1)}, \ldots, \mathbf{X}_{n-1}'\alpha^{(m-1)}$, obtain the equations for $\delta = (\delta_1', \delta_2')'$:

$$\sum_{i=1}^{n} \Delta_i \left[\mathbf{V}_i^{(m-1)}(v)K_i^{(m-1)}(v) \right.$$

$$\left. - \frac{\sum_{j=1}^{n} Y_j(\mathcal{T}_i)\mathbf{V}_j^{(m-1)}(v)K_j^{(m-1)}(v)\exp\{\theta^{(m-1)'}\mathbf{Z}_{1j} + \mathbf{V}_j^{(m-1)}(v)'\delta\}}{\sum_{j=1}^{n} Y_j(\mathcal{T}_i)\exp\{\theta^{(m-1)'}\mathbf{Z}_{1j} + \beta^{(m-1)}(\mathbf{X}_j'\alpha^{(m-1)})'\mathbf{Z}_{2j}\}} \right] = 0, (4.2)$$

where $\mathbf{V}_i^{(m)}(v) = (\mathbf{Z}_{2i}', \mathbf{Z}_{2i}'(\mathbf{X}_i'\alpha^{(m)} - v))'$, $K_i^{(m)}(v) = K_h(\mathbf{X}_i'\alpha^{(m)} - v)$, $Y_i(t) = I(\mathcal{T}_i \geq t)$. Let $\widehat{\delta}_1(v)$ and $\widehat{\delta}_2(v)$ be the solutions of δ_1 and δ_2. Then $\beta^{(m)}(v) = \widehat{\delta}_1(v)$ and $\dot{\beta}^{(m)}(v) = \widehat{\delta}_2(v)$ for $v = \mathbf{X}_1'\alpha^{(m-1)}, \ldots, \mathbf{X}_{n-1}'\alpha^{(m-1)}$.

The second step of Step m. For given $\beta^{(m)}$ and $\dot{\beta}^{(m)}$, estimate θ and α by solving the partial score function

$$\sum_{i=1}^{n} \Delta_i \left[\mathbf{X}_i\dot{\beta}^{(m)}(\mathbf{X}_i'\alpha^{(m-1)})'\mathbf{Z}_{2i} \right.$$

$$\left. - \frac{\sum_{k \in \mathcal{R}(\mathcal{T}_i)} \mathbf{X}_k\dot{\beta}^{(m)}(\mathbf{X}_k'\alpha^{(m-1)})'\mathbf{Z}_{2k}\exp\{\theta^{(m-1)'}\mathbf{Z}_{1k} + \beta^{(m)}(\mathbf{X}_k'\alpha)'\mathbf{Z}_{2k}\}}{\sum_{k \in \mathcal{R}(\mathcal{T}_i)} \exp\{\theta^{(m-1)'}\mathbf{Z}_{1k} + \beta^{(m)}(\mathbf{X}_k'\alpha)'\mathbf{Z}_{2k}\}} \right] = 0.$$

$$\sum_{i=1}^{n} \Delta_i \left[\mathbf{Z}_{1i} - \frac{\sum_{k \in \mathcal{R}(\mathcal{T}_i)} \mathbf{Z}_{1k} \exp\{\theta' \mathbf{Z}_{1k} + \beta^{(m)} (\mathbf{X}_k' \alpha^{(m-1)})' \mathbf{Z}_{2k}\}}{\sum_{k \in \mathcal{R}(\mathcal{T}_i)} \exp\{\theta' \mathbf{Z}_{1k} + \beta^{(m)} (\mathbf{X}_k' \alpha^{(m-1)})' \mathbf{Z}_{2k}\}} \right] = 0. \qquad (4.3)$$

Repeat the above steps for $m = 1, 2, \ldots$ until $\theta^{(m)}$, $\alpha^{(m)}$ and $\beta^{(m)}(\mathbf{X}_i' \alpha^{(m)})$ $(i = 1, \ldots, n)$ converge. Following arguments similar to those for Theorems 1–4, we can establish the uniform consistency, asymptotic normality and semiparametric efficiency of the resulting estimators for θ, α, and $\beta(\cdot)$; these are displayed in Theorems 5–8 of the Supplementary Material. The proofs of Theorem 5–8 are analogous to those of Theorems 1–4 and are omitted here.

5. Simulation Studies

In this section, we investigate the performance of the proposed global partial likelihood estimator (GPLE). The performance of the estimator $\widehat{\beta}(\cdot)$ is assessed via the weighted mean squared errors, WMSE $= n_g^{-1} \sum_{j=1}^{p} \sum_{k=1}^{n_g} a_j \{\widehat{\beta}_j(w_k) - \beta_j(w_k)\}^2$, where a_j is reciprocal to the sample variance of $\beta_j(w_k)$, w_k ($k = 1, \ldots, n_g$) are the grid points at which the functions $\beta(\cdot)$ are estimated. We assess $\widehat{\alpha}$ by bias, standard deviation (SD) and the root of mean square errors (RMSE). In the following examples, the Epanechnikov kernel is used, $n_g = 60$.

We adopt a similar setting as Fan, Lin, and Zhou (2006) and consider a varying-coefficient model, $\lambda(t) = 4t^3 \exp[b\{Z_1(t), Z_2, W\}]$, with $b\{Z_1(t), Z_2, W\} = 0.5W (1.5 - W) Z_1(t) + \sin(2W) Z_2 + 0.5\{\exp(W - 1.5) - \exp(-1.5)\}$, where the covariate $Z_1(t) = Z_1 I(t \leq 1)/4 + Z_1 I(t > 1)$ is time-dependent, and Z_1 and Z_2 are jointly normal with correlation 0.5, each with mean 0 and standard deviation 5; $W = \mathbf{X}' \alpha$, $\mathbf{X} = (X_1, X_2, X_3)'$ and $\alpha = (1, 1, 1)'$, X_1, X_2, and X_3 are independent, X_1 and X_2 are binary covariates that take the value 1 for one half of the subjects and 0 for the other half, and X_3 is uniformly distributed on $[0, 1]$. The censoring variable C given (Z_1, Z_2, W) is distributed uniformly on $[0, a(Z_1, Z_2, W)]$, where $a(Z_1, Z_2, W) = c_1 I(b(Z_1, Z_2, W) > b_0) + c_2 I(b(Z_1, Z_2, W) \leq b_0)$ with b_0 the mean function of $b(Z_1, Z_2, W)$. The constants $c_1 = 0.8$ and $c_2 = 20$ were chosen so that about 30%-40% of the data are censored in each region of the function $a(\cdot)$. We conducted 200 simulations with a sample size of 400. To investigate the efficiency of the proposed method, we also examined its performance in comparison with the local partial likelihood estimator (LPLE), in which the varying-coefficient functions are estimated by maximizing the local partial likelihood function (2.3). Figure 5.1 depicts the distribution of the weighted mean squared errors over the 200 replications.

The minimum weighted mean squared error of the GPLE likelihood estimator is smaller than that of the LPLE. The optimal bandwidth for the GPLE is smaller than that for the LPLE, because the amount of data used by the GPLE is more than that used by the LPLE with the same bandwith, the local partial
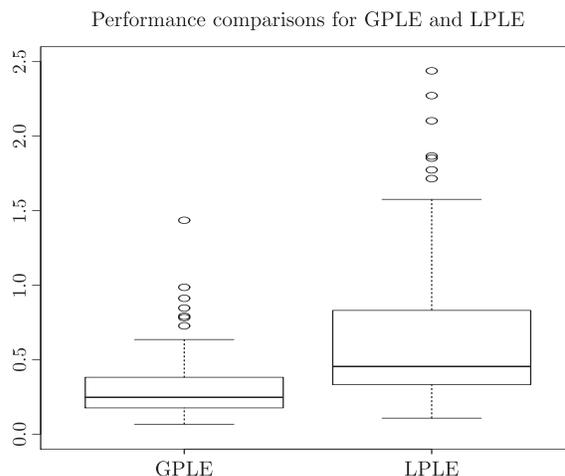
Figure 1. Boxplots of the weighted mean squared errors over 200 replications using the global partial likelihood estimator (GPLE) with the optimal bandwidth $h = 0.3$, and the local partial likelihood estimator (LPLE) with the optimal bandwidth $h = 0.6$.

Table 1. Standard deviations of the global partial likelihood estimator (GPLE) and the local partial likelihood estimator (LPLE) with bandwidth 0.5.

Function	Method	$w = 0.3$	0.75	1.5	2.25	2.70
β_1	GPLE	0.0854	0.1030	0.1481	0.1934	0.2347
	LPLE	0.1462	0.1557	0.4348	0.6590	0.7331
β_2	GPLE	0.0530	0.0478	0.0310	0.1147	0.1958
	LPLE	0.0825	0.0969	0.0773	0.2856	0.5375
β_3	GPLE	0.0639	0.0942	0.0377	0.0926	0.0895
	LPLE	0.1821	0.2427	0.2085	0.2233	0.2487

likelihood estimator needing to compensate for its lower data usage by enlarging the included range of data. In support of this conclusion, Table 5.1 presents the empirical standard deviations of 200 estimated values of $\widehat{\beta}_1(w)$, $\widehat{\beta}_2(w)$, and $\widehat{\beta}_3(w)$ using the global partial likelihood estimator and the local partial likelihood estimator with $h = 0.5$. We took $w = 0.3, 0.75, 1.5, 2.25$, and 2.7, corresponding to the 10th, 25th, 50th, 75th, and 90th percentiles of the distribution of W. Table 5.1 reveals that the estimated variance of the GPLE is smaller than that of the LPLE in all cases.

To estimate the regression coefficient α, we took the bandwidth $0.8h$, h the bandwidth to estimate the coefficient functions. Table 5.2 presents bias, SD, and RMSE over the 200 replications, using the proposed global partial likelihood estimator with bandwidth $h = 0.2, 0.3, 0.5$ and the local partial likelihood estimator

Table 2. The GPLE and LPLE for the regression coefficients.

		GPLE			LPLE			
		$h = 0.2$	$h = 0.3$	$h = 0.5$	$h = 0.5$	$h = 0.6$	$h = 0.7$	$h = 0.9$
α_2	Bias	-0.0047	-0.0027	-0.0026	-0.0035	-0.0064	-0.0119	-0.0239
	SD	0.0224	0.0198	0.0204	0.0458	0.0423	0.0392	0.0417
	RMSE	0.0229	0.0200	0.0205	0.0459	0.0428	0.0410	0.0481
α_3	Bias	0.0281	0.0285	0.0387	-0.0015	-0.0057	-0.0268	-0.0849
	SD	0.0577	0.0478	0.0410	0.1250	0.0729	0.0937	0.0839
	RMSE	0.0641	0.0556	0.0564	0.1251	0.0731	0.0974	0.1194

with bandwidth $h = 0.5, 0.6, 0.7, 0.9$. Table 5.2 shows that the GPLE for the regression coefficient is not sensitive to the selection of the bandwidth, while the LPLE of the regression coefficient is moderately sensitive to the selection of the bandwidth. Hence the LPLE may require an accurate estimator of the bandwidth to estimate the regression coefficients. In addition, both the RMSE and the SD of the global partial likelihood estimator are smaller than those of the local partial likelihood estimator in all of the presented cases, suggesting that the global partial likelihood estimator for the regression coefficient is indeed better. We also noted that the GPLE method has a discrepancy between SD and RMSE, especially when h is large, suggesting that the bandwidth to estimate α_3 at rate $n^{-1/2}$ may be less than 0.3.

6. Analysis of A Multiple Myeloma Trial

We applied the proposed method to analyze a clinical trial on newly diagnosed multiple myeloma patients who were randomized to receive two treatments, total therapy II and total therapy III, respectively (Shaughnessy et al. (2007)). Survival times were collected for 307 patients in the total therapy II (tt2) arm ($Z = 1$), where 189 deaths were observed (38.4% censoring) with a median follow-up time of 56 months, and for 170 patients in the total therapy III (tt3) arm ($Z = 0$), where 136 deaths were observed (20% censoring) with a median follow-up time of 37 months. A number of clinical and laboratory features that may provide prognostic information, including beta2-microglobulin (b2m), albumin (alb), hemoglobin level (hgb), antigen-presenting cells (apcs), bone marrow plasma cells (bmpc), magnetic resonance imaging (mri), and cytokines (cyto), were collected in the study. The goals of the analysis were to investigate whether these biomarkers would be predictive of patients' survival and how the effect of the treatment varies with the biomarkers if it does. The results may provide some clinical guidance for treatment selection as in general tt3 is more intensive than tt2 and could incur more toxicities.

In order to evaluate the prognostic significance of biomarkers, we set $\mathbf{X} =$ (b2m, alb, hgb, aspc, bmpc, mri, cyto) in patients treated with tt2 and tt3, and

Table 3. The estimators of α_1 and α_2 in the model (6.2) when $\mathbf{X}$ is the first three principle components (PC) for Multiple Myeloma data. $\|\alpha_1\| = 1$ and $\|\alpha_2\| = 1$ for the identification of the model (6.2).

	$\mathbf{X}$ = first three PCs		
α_1	0.1847	-0.9705	-0.1550
α_2	0.2577	-0.9554	-0.1445

fit the model

$$\lambda(t) = \lambda_0(t) \exp\{\beta_1(\alpha'\mathbf{X}) + \beta_2(\alpha'\mathbf{X})Z\}. \tag{6.1}$$

We assumed the same index $\alpha'\mathbf{X}$ for different therapies in the model (6.1). To check the reasonableness of the assumption, we considered the model

$$\lambda(t) = \lambda_0(t) \exp\{\beta_1(\alpha_1'\mathbf{X}) + \beta_2(\alpha_2'\mathbf{X})Z\}. \tag{6.2}$$

We approximated $\beta_k(\cdot), k = 1, 2$ by $B-$splines and then estimated α_k and $\beta_k(\cdot)$ by maximizing the partial likelihood function. If the given model (6.1) fits the data, α_1 and α_2 should agree well. Perhaps due to the relatively large number of parameters to estimate, we did not obtain convergent estimates based on the model (6.2) even when $\beta_1(\cdot)$ and $\beta_2(\cdot)$ were approximated by B-splines with few knots. As a remedy, we reduced the dimension of $\mathbf{X}$ using the principle components. As the standard deviations of the first seven principle components for $\mathbf{X}$ were $32.87, 14.54, 13.50, 4.80, 1.53, 0.52, 0.46$, we took the first three as the covariates in the model (6.2). We then used the B-spline of order 3 and 8 equally spaced knots along the direction of $\alpha_1'\mathbf{X}$ for estimating $\beta_1(\cdot)$ and $\alpha_2'\mathbf{X}$ for $\beta_2(\cdot)$. The resulting estimates for α_1 and α_2 are displayed in Table 6.3. Table 6.3 shows that α_1 and α_2 basically agree, suggesting that model (6.1). Developing a formal test is out of the scope of this project; we plan to report on it elsewhere.

We estimated the regression coefficients and functions using the proposed method with bandwidth $h = 3.5$, the bandwidth $h = 3.5$ chosen by K-fold cross-validation (Tian, Zucker, and Wei (2005); Fan, Lin, and Zhou (2006)) to minimize the prediction error $PE(h) = \int_0^\tau \left[N_i(t) - \widehat{E}\{N_i(t)\} \right]^2 d\{\sum_{k=1}^n N_k(t)\}$, where $\widehat{E}\{N_i(t)\} = \int_0^t Y_i(u) \exp\{\widehat{\beta}(\widehat{\alpha}'\mathbf{X}_i)'\mathbf{Z}_i\} d\widehat{\Lambda}_0(u)$ is the estimate of the expected number of failures up to time t. We used $K = 30$. To find the optimal bandwidth, we first specified a sequence of points of h, and at each point we computed $PE(h)$. Figure 6.2(a) shows the plot of PE vs h. With $h = 3.5$, we then obtained regression coefficients, coefficient functions, and their 95% confidence bands as shown in Table 6.4 and Figure 6.2, respectively. The calculation of the standard errors was carried out using 300 bootstrap resampled datasets, in which each subject was treated as a resampling unit in order to preserve the

Table 4. The regression coefficients estimators for Multiple Myeloma data.

	Estimated	SD	p−value
b2m	0.1143	0.0299	0.0001
alb	-0.4866	0.1591	0.0022
hgb	0.0228	0.0564	0.6860
aspc	0.0020	0.0065	0.7583
bmpc	0.0014	0.0056	0.8026
mri	0.0167	0.0081	0.0392
cyto	0.8657	0.1015	0.0000
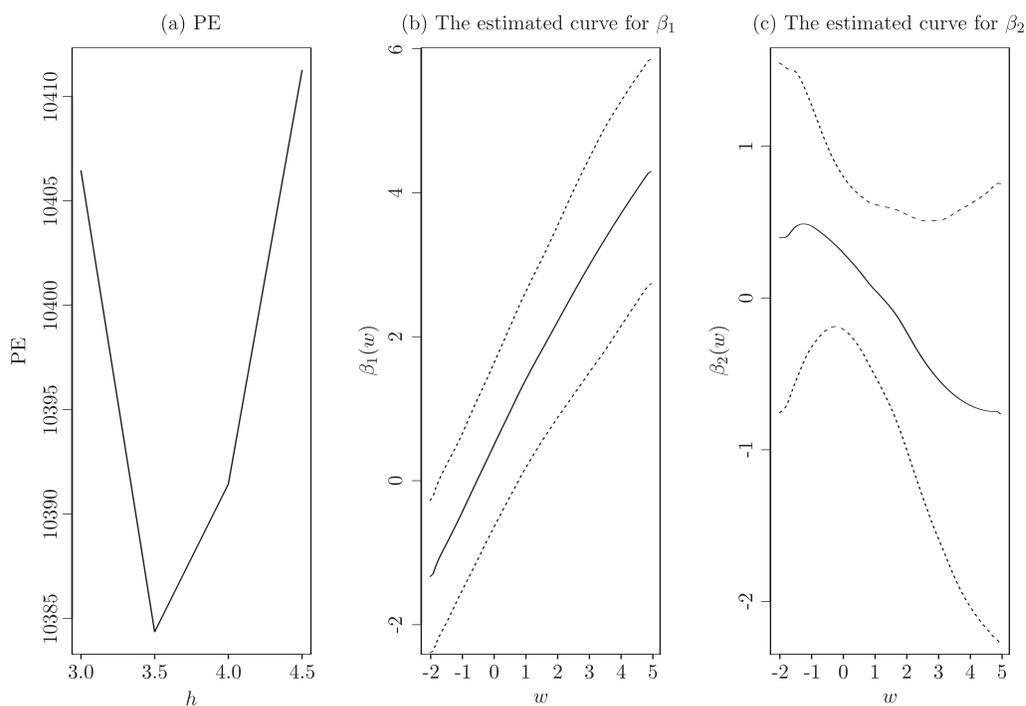
Figure 2. The estimated curves and their 95% pointwise confidence bands with $h = 3.5$.

inherent features of the data. The choice of 300 was determined by monitoring the stability of the standard errors.

The results can be summarized as follows. First, Figure 6.2(c) shows that $\beta_2(\cdot)$ is not significantly different from the zero function, suggesting that the treatments, tt2 and tt3, are not significantly different, and hence, the selected biomarkers do not interact with the treatment significantly. This is an important finding as it reveals that, with the given biomarkers, the more intensive and potentially more toxic therapy (tt3) may not offer survival advantages compared to the less intensive therapy (tt2). To confirm the conclusion, we also analyzed

the data using the standard Cox proportional hazard model with only treatment covariate and without any interaction; the resulting coefficient estimate was 0.241 with SE= 0.204 and $p-$value 0.24.

The estimates of hgb, aspc, and bmpc are not significant, suggesting that they may not be predictive biomarkers. The positive estimates of the coefficients of b2m, mri, and cyto, and the negative estimate of the coefficient of alb in Table 6.4, along with the monotone increasing function of $\beta_1(\cdot)$ across zero at around $W = 0$ as depicted in Figure 6.2(b), show that the biomarkers b2m, alb, mri, and cyto are significantly related to patients' survival. Specifically, larger values of b2m, mri, and cyto, and lower values of alb increase the risk of death, while lower values of b2m, mri, and cyto, and larger values of alb reduce the risk of death. The results have confirmed the hypotheses proposed in Shaughnessy et al. (2007), and could lead to more precise risk classifications.

7. Final Remarks

To properly identify biomarkers for risk stratification and treatment selection for individual patients, we have proposed a single-index varying coefficients hazards model. Our model accommodates multiple predictive biomarkers and allows nonparametric interactions between the multiple biomarkers and the treatment.

To increase efficiency, we have proposed to apply a global partial likelihood for inference, and have obtained appealing statistical properties, including consistency, asymptotic normality, and semiparametric efficiency. Simulation studies have verified the finite sample performance; we have applied the proposed method to study a clinical trial on multiple myeloma and have gained some biological insight. In our numerical experiments, we used the LPLE estimates as the initial values for the Newton-Raphson algorithm and the convergence was often achieved within $3-5$ step. The added computational burden is relatively small.

There are several opportunities for future research. We have focused on only a small number of biomarkers and did not discuss the case of thousands of genomic expression profiles in the dataset. Including genomic information in the construct of predictive biomarkers could potentially be useful for personalized medicine and efficient for risk stratification. However, major effort is needed to extend our proposed work to high dimensional settings with treatment and biomarker interactions. The proposed model requires the same single index to be included in all of the regression coefficient function. It is necessary to extend our work to accommodate multiple indices, especially in the presence of high-dimensional markers. Sliced inverse regression (SIR) could potentially be a useful approach and we are currently investigating its usage in our setting.

Supplementary Materials

The online supplementary material includes Theorems 5-8 and related notations.

Acknowledgement

Appendix: Notations, Conditions and Proofs for Theorems 1-4

Let $\mathcal{C}_0 = \{\delta(v) = (\delta_1(v), \ldots, \delta_p(v))$: $v \in [0,1]$, $\delta_1(0) = 0$, $\delta(v)$ is continuous on $[0,1]\}$, $\boldsymbol{\Theta}$ be the support of α, and $f(v; \alpha)$ be the conditional density function of $\mathbf{X}'\alpha$. Let

$$W(\alpha, \delta_2) = \delta_2(\mathbf{X}'\alpha)'\mathbf{Z}, \mu_i = \int x^i K(x)dx, \ \nu_i = \int x^i K^2(x)dx,$$

$$P(t \mid \mathbf{z}, \mathbf{x}) = \Pr(\mathcal{T} \geq t \mid \mathbf{Z} = \mathbf{z}, \mathbf{X} = \mathbf{x}),$$

$$\Gamma(\mathbf{z}, \mathbf{x}) = \int_0^\tau P(t \mid \mathbf{z}, \mathbf{x})\lambda_0(t)dt,$$

$$s_{00}(t; \alpha, \delta_1) = E\left[P(t \mid \mathbf{Z}, \mathbf{X})\exp\{\delta_1(\mathbf{X}'\alpha)'\mathbf{Z}\}\right],$$

$$s_{10}(t; \alpha, \delta_1, v) = E\left[\mathbf{Z}P(t \mid \mathbf{Z}, \mathbf{X})\exp\{\delta_1(v)'\mathbf{Z}\} \mid \mathbf{X}'\alpha = v\right] f(v; \alpha),$$

$$s_{20}(t; \alpha, \delta_1, v) = E\left[\mathbf{Z}\mathbf{Z}'P(t \mid \mathbf{Z}, \mathbf{X})\exp\{\delta_1(v)'\mathbf{Z}\} \mid \mathbf{X}'\alpha = v\right] f(v; \alpha),$$

$$s_{11}(t; \alpha, \delta_1, v) = E\left[\mathbf{Z}\mathbf{X}'P(t \mid \mathbf{Z}, \mathbf{X})\exp\{\delta_1(v)'\mathbf{Z}\} \mid \mathbf{X}'\alpha = v\right] f(v; \alpha),$$

$$r_0(t; \alpha_1, \alpha_2, \delta_1, \delta_2) = E\left[P(t \mid \mathbf{Z}, \mathbf{X})\mathbf{X}W(\alpha_1, \delta_2)\exp\{\delta_1(\mathbf{X}'\alpha_2)'\mathbf{Z}\}\right],$$

$$r_1(t; \alpha_1, \alpha_2, \delta_1, \delta_2, v) = E\left[P(t \mid \mathbf{Z}, \mathbf{X})\mathbf{X}W(\alpha_1, \delta_2)\exp\{\delta_1(v)'\mathbf{Z}\}\mathbf{Z}' \mid \mathbf{X}'\alpha_2 = v\right]$$
$$\times f(v; \alpha_2),$$

$$m_1(t) = E\left[P(t \mid \mathbf{Z}, \mathbf{X})\exp\{\beta(\mathbf{X}'\alpha_0)'\mathbf{Z}\}\dot{\beta}(\mathbf{X}'\alpha_0)'\mathbf{Z}\mathbf{X}\right],$$

$$m_2(t) = E\left[P(t \mid \mathbf{Z}, \mathbf{X})\mathbf{X}\mathbf{X}'\exp\{\beta(\mathbf{X}'\alpha_0)'\mathbf{Z}\}\{\dot{\beta}(\mathbf{X}'\alpha_0)'\mathbf{Z}\}^2\right],$$

$$\kappa(t, v) = E\left[P(t \mid \mathbf{Z}, \mathbf{X})\mathbf{Z}\mathbf{X}'\exp\{\mathbf{Z}'\beta(v)\}\mathbf{Z}'\dot{\beta}(v) \mid \mathbf{X}'\alpha_0 = v\right] f(v; \alpha_0),$$

$$\mathbf{D}_1 = \int_0^\tau \left\{\frac{r_0(t)m_1(t)'}{s_{00}(t)} - m_2(t)\right\} \lambda_0(t)dt,$$

$$\mathbf{D}_2(v) = \int_0^\tau \left\{\frac{s_{10}(t; v)m_1(t)'}{s_{00}(t)} - \kappa(t, v)\right\} \lambda_0(t)dt,$$

$$\mathbf{D}_3(v) = \int_0^\tau \left\{ \frac{r_0(t)s_{10}(t;v)'}{s_{00}(t)} - r_1(t;v) \right\} \lambda_0(t)dt,$$

$$\mathbf{\Sigma}(v) = \int_0^\tau s_{20}(t;v)\lambda_0(t)dt,$$

$$\Psi(w;v) = \int_0^\tau \left\{ \frac{s_{10}(t;v)s_{10}(t;w)'}{s_{00}(t)} - \mathbf{D}_2(v)\mathbf{D}_1^{-1}\left[\frac{r_0(t)s_{10}(t;w)'}{s_{00}(t)} \right. \right.$$
$$\left. \left. -r_1(t;w) \right] \right\} \lambda_0(t)dt,$$

$$r_0(t) = r_0(t;\alpha_0,\alpha_0,\beta,\dot{\beta}), \quad r_1(t;v) = r_1(t;\alpha_0,\alpha_0,\beta,\dot{\beta},v), \quad s_{00}(t) = s_{00}(t;\alpha_0,\beta),$$

$s_{10}(t;v) = s_{10}(t;\alpha_0,\beta,v)$, $s_{20}(t;v) = s_{20}(t;\alpha_0,\beta,v)$, and $s_{11}(t;v) = s_{11}(t;\alpha_0,\beta,v)$.

Let g satisfy the integral equation in $\mathcal{C}_0$: $\mathbf{D}_3(w) = g(w)\mathbf{\Sigma}(w) - \int_0^1 g(v)\Psi(w;v) \, dv$.

Conditions:

(C1) The kernel function $K(x)$ is a symmetric density function with compact support $[-1,1]$ and continuous derivative.

(C2) τ is finite, $\Pr(T > \tau) > 0$, and $\Pr(C = \tau) > 0$.

(C3) $\mathbf{Z}$ and $\mathbf{X}$ are bounded with compact support, and $P(C = 0 \mid \mathbf{Z} = \mathbf{z}, \mathbf{X} = \mathbf{x}) < 1$.

(C4) $\alpha \in \mathbf{\Theta}$, where $\mathbf{\Theta}$ is a bounded compact set.

(C5) The density function $f(x;\alpha)$ of $\mathbf{X}'\alpha$ is bounded away from zero and has a continuous second-order derivative for any $\alpha \in \mathbf{\Theta}$. The function $\beta(v)$, $s_{00}(t;\alpha,\delta_1)$, $s_{10}(t;\alpha,\delta_1,v)$, $s_{11}(t;\alpha,\delta_1,v)$, $s_{20}(t;\alpha,\delta_1,v)$, $r_0(t;\alpha,\alpha,\delta_1,\delta_2)$, $r_1(t;\alpha,\alpha,\delta_1,\delta_2,v)$, and $\kappa(t,v)$ are twice continuously differentiable for $v \in [0,\tau]$ for any $t \in [0,\tau]$, $\alpha \in \mathbf{\Theta}$, $\delta_1 \in \mathcal{C}_0$, and bounded δ_2.

(C6) There exists a unique root (α, δ_1) to

$$\int_0^\tau r_0(t;\alpha,\alpha_0,\beta,\delta_2)\lambda_0(t)dt - \int_0^\tau r_0(t;\alpha,\alpha,\delta_1,\delta_2)\frac{s_{00}(t)}{s_{00}(t;\alpha,\delta_1)}\lambda_0(t)dt = 0,$$

$$\int_0^\tau s_{10}(t;v)\lambda_0(t)dt - \int_0^\tau s_{10}(t;\alpha,\delta_1,v)\frac{s_{00}(t)}{s_{00}(t;\alpha,\delta_1)}\lambda_0(t)dt = 0,$$

in $\delta_1 \in \mathcal{C}_0$ and $\alpha \in \mathbf{\Theta}$ for any bounded δ_2.

(C7) $h^2 \log(n) \to 0$ and $nh^3 \to \infty$.

The regularity condition (C5) requires the density function of $\mathbf{X}'\alpha$ to be bounded away from zero and have a continuous second-order derivative for any α. The proposed framework does not deal with discrete components of $\mathbf{X}$.

Proof of Theorem 1. For any vector functions $\delta_1(v)$ and $\delta_2(v)$, set

$$U_n(\alpha, \delta_1, \delta_2; v) = \left\{ U_{n1}(\alpha, \delta_1, \delta_2)', U_{n2}(\alpha, \delta_1, \delta_2, v)' \right\}',$$

where

$$U_{n1}(\alpha, \delta_1, \delta_2) = n^{-1} \sum_{i=1}^{n} \int_0^\tau \left[\mathbf{X}_i W_i(\alpha, \delta_2) - \frac{R_{n0}(t; \alpha, \delta_1, \delta_2)}{S_{n0}(t; \alpha, \delta_1)} \right] dN_i(t),$$

$$U_{n2}(\alpha, \delta_1, \delta_2, v) = n^{-1} \sum_{i=1}^{n} \int_0^\tau \left[\mathbf{V}_i(v; \alpha) K_i(v; \alpha) - \frac{S_{n1}(t; \alpha, \delta, v)}{S_{n0}(t; \alpha, \delta_1)} \right] dN_i(t),$$

$$W_i(\alpha, \delta_2) = \delta_2(\mathbf{X}_i'\alpha)'\mathbf{Z}_i, \qquad \mathbf{V}_i(v; \alpha) = \left(\mathbf{Z}_i', \frac{\mathbf{Z}_i'(\mathbf{X}_i'\alpha - v)}{h} \right)',$$

$$K_i(v; \alpha) = K_h(\mathbf{X}_i'\alpha - v), \qquad \delta = (\delta_1', h\delta_2')',$$

$$S_{n0}(t; \alpha, \delta_1) = n^{-1} \sum_{j=1}^{n} Y_j(t) \exp\{\delta_1(\mathbf{X}_j'\alpha)'\mathbf{Z}_j\},$$

$$S_{n1}(t; \alpha, \delta, v) = n^{-1} \sum_{j=1}^{n} Y_j(t) \mathbf{V}_j(v; \alpha) K_j(v; \alpha) \exp\{\mathbf{V}_j(v; \alpha)'\delta(v)\},$$

$$R_{n0}(t; \alpha, \delta_1, \delta_2) = n^{-1} \sum_{j=1}^{n} Y_j(t) \mathbf{X}_j W_j(\alpha, \delta_2) \exp\{\delta_1(\mathbf{X}_j'\alpha)'\mathbf{Z}_j\}.$$

Under the varying coefficient Cox model (1.1), we have

$$U_n(\alpha, \delta_1, \delta_2; v) = u(\alpha, \delta_1, \delta_2; v) + o_p(1) \equiv \left\{ u_1(\alpha, \delta_1, \delta_2)', u_2(\alpha, \delta_1, v)', 0 \right\}' + o_p(1),$$

where
$$u_1(\alpha, \delta_1, \delta_2) = \int_0^\tau r_0(t; \alpha, \alpha_0, \beta, \delta_2) \lambda_0(t) dt - \int_0^\tau r_0(t; \alpha, \alpha, \delta_1, \delta_2) \frac{s_{00}(t)}{s_{00}(t; \alpha, \delta_1)} \lambda_0(t) dt,$$

$$u_2(\alpha, \delta_1, v) = \int_0^\tau s_{10}(t; v) \lambda_0(t) dt - \int_0^\tau s_{10}(t; \alpha, \delta_1, v) \frac{s_{00}(t)}{s_{00}(t; \alpha, \delta_1)} \lambda_0(t) dt.$$

It follows that $U_n(\widehat{\alpha}, \widehat{\beta}, \widehat{\widetilde{\beta}}; v) = 0$ and $u(\alpha_0, \beta, \delta_2; v) = 0$ for any bounded function δ_2. By Condition (C6), (α_0, β) is a unique root to $u(\alpha, \delta_1, \delta_2; v) = 0$ in $\delta_1 \in \mathcal{C}_0$ and $\alpha \in \boldsymbol{\Theta}$ for any bounded δ_2.

Define $\mathcal{B}_n = \{\delta_1 : \|\delta_1\| \leq D, \|\delta_1(v_1) - \delta_1(v_2)\| \leq d[|v_1 - v_2| + b_n], v_1, v_2 \in [0,1]\}$ for some constants $D > 0$ and $d > 0$, where $b_n = h^2 + (nh)^{-1/2}(\log n)^{1/2}$. To show the uniform consistency of $\widehat{\beta}$ and $\widehat{\alpha}$, it suffices to prove the following:

(i) for each continuous function vector δ_1 and any bounded function vector δ_2, $\sup_{0 \leq v \leq 1} \|U_n(\alpha, \delta_1, \delta_2; v) - u(\alpha, \delta_1, \delta_2; v)\| = o_p(1)$;

(ii) $\sup_{0 \leq v \leq 1, \alpha \in \boldsymbol{\Theta}, \delta_1 \in \mathcal{B}_n, \delta_2 \in \mathcal{R}} \|U_n(\alpha, \delta_1, \delta_2; v) - u(\alpha, \delta_1, \delta_2; v)\| = o_p(1)$, where $\mathcal{R}$ is the set of functions on $[0,1]$, which are bounded uniformly;

(iii) $P\{\widehat{\beta} \in \mathcal{B}_n\} \to 1$.

Once (i)$-$(iii) are established, similar to the Arzela-Ascoli Theorem, we can show that for any subsequence of $\{(\widehat{\alpha}, \widehat{\beta})\}$, there exists a further convergent subsequence $\{(\widehat{\alpha}_n, \widehat{\beta}_n)\}$ such that $\widehat{\beta}_n(v) \to \beta^*(v)$ in probability uniformly over $[0, 1]$ and $\widehat{\alpha}_n \to \alpha^*$. It is seen that $\beta^* \in \mathcal{C}_0$. Observe that $u(\alpha^*, \beta^*, \dot{\widehat{\beta}}; v) = U_n(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}; v) - \{U_n(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}; v) - u(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}; v)\} - \{u(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}; v) - u(\alpha^*, \beta^*, \dot{\widehat{\beta}}; v)\}$ and $U_n(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}; v) = 0$, where $\dot{\widehat{\beta}}(\cdot)$ is the estimator of $\dot{\beta}(\cdot)$, the derivative of $\beta(\cdot)$. It follows from (ii) and (iii) that $u(\alpha^*, \beta^*, \dot{\widehat{\beta}}; v) = 0$. Since for any bounded function δ_2, $u(\alpha, \delta_1, \delta_2; v) = 0$ has a unique root at (α_0, β), we have $\alpha^* = \alpha_0$ and $\beta^* = \beta$. The uniform consistency of $(\widehat{\alpha}, \widehat{\beta})$ is proved.

Proofs of (i) and (ii). Observe that Z and $\mathbf{X}$ are bounded. The proofs of (i) and (ii) can be obtained by using kernel theory (Fan and Gijbels (1996); Fan, Lin, and Zhou (2006)) and following the arguments in Chen et al. (2010).

Proof of (iii). Let p be the dimension of $\mathbf{Z}$. Denote $\widehat{S}_{n1}(t; v) = S_{n1}(t; \widehat{\alpha}, \widehat{\beta}, h\dot{\widehat{\beta}}, v)$, $\widehat{S}_{n1,1}$, and $U_{n2,1}$ to be the first p-elements of $\widehat{S}_{n1}$ and U_{n2}, respectively. Given any $v_1 \in [0, 1]$ and $v_2 \in [0, 1]$ with $|v_1 - v_2| < h$, since $U_{n2}(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}; v_1) = 0$ and $U_{n2}(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}; v_2) = 0$, we have

$$0 = U_{n2,1}(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}; v_1) - U_{n2,1}(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}; v_2)$$
$$= d_1(v_1) - d_1(v_2) - \int_0^\tau \frac{\widehat{S}_{n1,1}(t; v_1) - \widehat{S}_{n1,1}(t; v_2)}{S_{n0}(t; \widehat{\alpha}, \widehat{\beta})} d\overline{N}(t) + O_p(b_n), \quad (A.1)$$

where $d_1(v) = E\left[\mathbf{Z}\Gamma(\mathbf{Z}, \mathbf{X}) \exp\{\beta(\mathbf{X}'\alpha_0)'\mathbf{Z}\} \mid \mathbf{X}'\widehat{\alpha} = v\right] f(v; \widehat{\alpha})$ and $\overline{N}(t) = n^{-1} \sum_{i=1}^n N_i(t)$.

Using a Taylor expansion and kernel theory (Horowitz (1996)), we have

$$\widehat{S}_{n1,1}(t; v_1) - \widehat{S}_{n1,1}(t; v_2)$$
$$= n^{-1} \sum_{i=1}^n \mathbf{Z}_i K_i(v_1; \widehat{\alpha}) Y_i(t) \exp\{\widehat{\beta}(v_1)'\mathbf{Z}_i\} \mathbf{Z}_i' \left[\widehat{\beta}(v_1) - \widehat{\beta}(v_2)\right]$$
$$+ n^{-1} \sum_{i=1}^n \mathbf{Z}_i \frac{\dot{K}\{(\mathbf{X}_i'\widehat{\alpha} - v_1)/h\}}{h^2} Y_i(t) \exp\{\widehat{\beta}(v_1)'\mathbf{Z}_i\}(v_2 - v_1)$$
$$+ O_p(b_n + (v_2 - v_1)^2)$$
$$= s_{20}(t; \widehat{\alpha}, \widehat{\beta}, v_1) \left\{\widehat{\beta}(v_1) - \widehat{\beta}(v_2)\right\} - \frac{\partial s_{10}(t; \widehat{\alpha}, \widehat{\beta}, v_1)}{\partial v}(v_2 - v_1)$$
$$+ O_p(b_n + (v_2 - v_1)^2), \quad (A.2)$$

uniformly in $t \in [0, \tau]$ and $v_1, v_2 \in [0, 1]$ such that $|v_1 - v_2| < h$. Substituting (A.2) into (A.1) and using $U_{n2}(\widehat{\alpha}, \widehat{\beta}, h\widehat{\dot{\beta}}; v_1) - U_{n2}(\widehat{\alpha}, \widehat{\beta}, h\widehat{\dot{\beta}}; v_2) = 0$, we have

$$\int_0^\tau s_{20}(t; \widehat{\alpha}, \widehat{\beta}, v_1) \frac{s_{00}(t)}{s_{00}(t; \widehat{\alpha}, \widehat{\beta})} \lambda_0(t) dt \left\{ \widehat{\beta}(v_1) - \widehat{\beta}(v_2) \right\}$$

$$= d_1(v_1) - d_1(v_2) + \int_0^\tau \frac{\partial s_{10}(t; \widehat{\alpha}, \widehat{\beta}, v_1)}{\partial v} \frac{s_{00}(t)}{s_{00}(t; \widehat{\alpha}, \widehat{\beta})} \lambda_0(t) dt (v_2 - v_1)$$

$$+ O_p(b_n + (v_2 - v_1)^2).$$

Then (iii) holds by the conditions on $\mathbf{Z}$, $\mathbf{X}$, $s_{10}(t; \alpha, \delta_1, v)$, and $f(v; \alpha)$.

Proofs of Theorems 2 and 3. Let $a_n = \|\widehat{\alpha} - \alpha_0\|$, $c_n = \sup_{v \in [0,1]} \|\widehat{\beta}(v) - \beta(v)\|$, $d_n = \sup_{v \in [0,1]} \|h\widehat{\dot{\beta}}(v) - h\dot{\beta}(v)\|$, and $e_n = h^2 + (nh)^{-1/2}$. Let $M_i(t) = N_i(t) - \int_0^t P(t \mid \mathbf{Z}_i, \mathbf{X}_i) \exp\{\beta(\mathbf{X}_i'\alpha_0)'\mathbf{Z}_i\}\lambda_0(t)dt$. The proof consists of four steps.

Step (i). Consider the asymptotic expression

$$U_{n1}(\widehat{\alpha}, \widehat{\beta}, \widehat{\dot{\beta}}) - U_{n1}(\alpha_0, \beta, \dot{\beta}) = \mathbf{D}_1(\widehat{\alpha} - \alpha_0) + \int_0^1 \mathbf{D}_3(w) \left\{ \widehat{\beta}(w) - \beta(w) \right\} dw$$

$$+ O_p((a_n + c_n + \frac{d_n}{h})(a_n + c_n)) + o_p(n^{-1/2}). \quad \text{(A.3)}$$

If $\widehat{W}_j = W_j(\widehat{\alpha}, \widehat{\dot{\beta}})$, $W_j = W_j(\alpha_0, \dot{\beta})$, $\widehat{R}_{n0}(t; \alpha, \delta_1) = (1/n) \sum_{j=1}^n Y_j(t)\mathbf{X}_j\widehat{W}_j \exp\{\delta_1(\mathbf{X}_j'\alpha)'\mathbf{Z}_j\}$,
and $R_{n0}(t; \alpha, \delta_1) = n^{-1} \sum_{j=1}^n Y_j(t)\mathbf{X}_j W_j \exp\{\delta_1(\mathbf{X}_j'\alpha)'\mathbf{Z}_j\}$, we have

$$U_{n1}(\widehat{\alpha}, \widehat{\beta}, \widehat{\dot{\beta}}) - U_{n1}(\alpha_0, \beta, \dot{\beta})$$

$$= n^{-1} \sum_{i=1}^n \int_0^\tau \left[\mathbf{X}_i\widehat{W}_i - \mathbf{X}_i W_i - \frac{\widehat{R}_{n0}(t; \alpha_0, \beta)}{S_{n0}(t; \alpha_0, \beta)} + \frac{R_{n0}(t; \alpha_0, \beta)}{S_{n0}(t; \alpha_0, \beta)} \right] dM_i(t)$$

$$+ \int_0^\tau \frac{r_0(t)}{s_{00}(t)} \left\{ S_{n0}(t; \widehat{\alpha}, \widehat{\beta}) - S_{n0}(t; \alpha_0, \beta) \right\} \lambda_0(t) dt$$

$$- \int_0^\tau \left\{ \widehat{R}_{n0}(t; \widehat{\alpha}, \widehat{\beta}) - \widehat{R}_{n0}(t; \alpha_0, \beta) \right\} \lambda_0(t) dt + O_p((a_n + c_n)^2). \quad \text{(A.4)}$$

By large number theory, kernel theory, and Taylor series expansions,

$$S_{n0}(t; \widehat{\alpha}, \widehat{\beta}) - S_{n0}(t; \alpha_0, \beta)$$

$$= m_1(t)'(\widehat{\alpha} - \alpha_0) + \int_0^1 s_{10}(t; w)' \left[\widehat{\beta}(w) - \beta(w) \right] dw + O_p((a_n + c_n)^2),$$

$$\widehat{R}_{n0}(t; \widehat{\alpha}, \widehat{\beta}) - \widehat{R}_{n0}(t; \alpha_0, \beta)$$

$$= m_2(t)\,(\widehat{\alpha} - \alpha_0) + \int_0^1 r_1(t; w)' \left[\widehat{\beta}(w) - \beta(w)\right] dw$$

$$+ O_p\{(a_n + c_n)(a_n + c_n + \frac{d_n}{h})\}. \tag{A.5}$$

Similar to the proof of Theorem 1, we can show that $\sup_{w \in (0,1)} \|\dot{\widehat{\beta}}(w) - \dot{\beta}(w)\| \to 0$. Together with $\|\widehat{\alpha} - \alpha_0\| \to 0$, (A.5), and (A.4), we get (A.3). vskip .1cm

Step (ii). Denote $\zeta_n(v) = [\widehat{\beta}(v)' - \beta(v)', h\{\dot{\widehat{\beta}}(v) - \dot{\beta}(v)\}']'$. Consider the expression
$U_{n2}(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}, v) - U_{n2}(\alpha_0, \beta, \dot{\beta}, v)$

$$= \mathbf{D}_2(v)\,(\widehat{\alpha} - \alpha_0)\,(1,0)' - \begin{pmatrix} \boldsymbol{\Sigma}(v)\ 0 \\ 0\ \ \mu_2\boldsymbol{\Sigma}(v) \end{pmatrix} \zeta_n(v)$$

$$+ \int_0^1 \int_0^\tau \frac{s_{10}(t; v)s_{10}(t; w)'}{s_{00}(t)} \lambda_0(t)dt \left[\widehat{\beta}(w) - \beta(w)\right] dw(1,0)'$$

$$+ O_p\!\left(\frac{a_n^2}{h} + c_n^2 + d_n^2 + \frac{a_n c_n}{h} + \frac{a_n d_n}{h} + c_n d_n + \frac{a_n e_n}{h} + c_n b_n + d_n b_n\right). \tag{A.6}$$

By large number theory, kernel theory, and a Taylor series expansion, we get

$$U_{n2}(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}, v) - U_{n2}(\alpha_0, \beta, \dot{\beta}, v)$$

$$= -\int_0^\tau \frac{\partial s_{11}(t; v)}{\partial v} \lambda_0(t)dt\,(\widehat{\alpha} - \alpha_0)\,(1,0)'$$

$$+ \int_0^\tau s_{10}(t; v) \left[\frac{S_{n0}(t; \widehat{\alpha}, \widehat{\beta}) - S_{n0}(t; \alpha_0, \beta)}{s_{00}(t)}\right] \lambda_0(t)dt(1,0)'$$

$$- \int_0^\tau \left[S_{n1}(t; \widehat{\alpha}, \widehat{\beta}, h\dot{\widehat{\beta}}, v) - S_{n1}(t; \alpha_0, \beta, h\dot{\beta}, v)\right] \lambda_0(t)dt$$

$$+ O_p\left[(a_n + h^2)\left\{a_n + c_n + d_n + h + (nh^3)^{-1/2}\right\}\right]$$

$$+ O_p\left\{(a_n + c_n + d_n)n^{-1/2}\right\}, \tag{A.7}$$

where
$S_{n1}(t; \widehat{\alpha}, \widehat{\beta}, h\dot{\widehat{\beta}}, v) - S_{n1}(t; \alpha_0, \beta, h\dot{\beta}, v)$

$$= \left\{\kappa(t, v) - \frac{\partial s_{11}(t; v)}{\partial v}\right\}(\widehat{\alpha} - \alpha_0)(1,0)' + \begin{pmatrix} s_{20}(t; v)\ 0 \\ 0\ \ \mu_2 s_{20}(t; v) \end{pmatrix} \zeta_n(v)$$

$$+ O_p\left\{\frac{a_n(a_n + c_n + d_n + e_n)}{h}\right\} + O_p\left\{(c_n + d_n)(b_n + c_n + d_n)\right\}. \tag{A.8}$$

Substituting (A.8) and (A.5) into (A.7), we get (A.6).

Step (iii). Consider asymptotic expressions of

$$U_{n1}(\alpha_0, \beta, \dot{\beta}) = n^{-1} \sum_{i=1}^n \int_0^\tau \left[\mathbf{X}_i W_i - \frac{r_0(t)}{s_{00}(t)}\right] dM_i(t) + o_p(n^{-1/2}), \tag{A.9}$$

$$U_{n2,1}(\alpha_0, \beta, \dot{\beta}, v) = A_n(v) + V_n(v), \tag{A.10}$$

where

$$A_n(v) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\mathbf{Z}_i K_i(v; \alpha_0) - \frac{S_{n1,1}(t; \alpha_0, \beta, h\dot{\beta}, v)}{S_{n0}(t; \alpha_0, \beta)} \right] dM_i(t),$$

$$V_n(v) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\mathbf{Z}_i K_i(v; \alpha_0) - \frac{S_{n1,1}(t; \alpha_0, \beta, h\dot{\beta}, v)}{S_{n0}(t; \alpha_0, \beta)} \right]$$
$$\times Y_i(t) \exp\{\mathbf{Z}'_i \beta(\mathbf{X}'_i \alpha_0)\} \lambda_0(t) dt.$$

With

$$V_n(v) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \mathbf{Z}_i K_i(v; \alpha_0) Y_i(t) \left[\exp\{\mathbf{Z}'_i \beta(\mathbf{X}'_i \alpha_0)\} \right.$$
$$\left. - \exp\{\mathbf{Z}'_i \beta(v) + \mathbf{Z}'_i \dot{\beta}(v)(\mathbf{X}'_i \alpha_0 - v)\} \right] \lambda_0(t) dt,$$

it can be shown that,

$$V_n(v) = \frac{1}{2} h^2 \mu_2 \mathbf{\Sigma}(v) \ddot{\beta}(v) + o_p(h^2), \tag{A.11}$$

$$(nh)^{1/2} A_n(v) = n^{-1/2} h^{1/2} \sum_{i=1}^{n} \int_0^{\tau} \left[\mathbf{Z}_i K_i(v; \alpha_0) - \frac{s_{10}(t; v)}{s_{00}(t)} \right] dM_i(t) + o_p(e_n). \tag{A.12}$$

The Martingale Central Limit Theorem implies that $(nh)^{1/2} A_n(w)$ is asymptotically normal with mean zero and covariance matrix $\nu_0 \mathbf{\Sigma}(v)$.

Step (iv). Consider the asymptotic expression

$$\mathbf{D}_1 (\widehat{\alpha} - \alpha_0) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \xi_i(t) dM_i(t) + o_p(n^{-1/2}), \tag{A.13}$$

where

$$\xi_i(t) = \int_0^1 g(v) \frac{s_{10}(t; v)}{s_{00}(t)} dv + \left\{ \int_0^1 g(v) \mathbf{D}_2(v) \mathbf{D}_1^{-1} dv - I \right\} \left\{ \mathbf{X}_i W_i - \frac{r_0(t)}{s_{00}(t)} \right\}$$
$$- \mathbf{Z}_i g(\mathbf{X}'_i \alpha_0).$$

By (A.3), (A.9), and $U_{n1}(\widehat{\alpha}, \widehat{\beta}, \widehat{\dot{\beta}}) = 0$, we get

$$\widehat{\alpha} - \alpha_0 = -\mathbf{D}_1^{-1} n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\mathbf{X}_i W_i - \frac{r_0(t)}{s_{00}(t)} \right] dM_i(t)$$
$$- \mathbf{D}_1^{-1} \int_0^1 \mathbf{D}_3(w) \left\{ \widehat{\beta}(w) - \beta(w) \right\} dw$$
$$+ O_p((a_n + c_n + \frac{d_n}{h})(a_n + c_n)) + o_p(n^{-1/2}). \tag{A.14}$$

Substituting this into (A.6), with $U_{n2,1}(\widehat{\alpha}, \widehat{\beta}, \widehat{\dot{\beta}}; v) = 0$ and (A.10), we have

$$n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\left\{ \mathbf{Z}_i K_i(v; \alpha_0) - \frac{s_{10}(t; v)}{s_{00}(t)} \right\} - \mathbf{D}_2(v) \mathbf{D}_1^{-1} \left\{ \mathbf{X}_i W_i - \frac{r_0(t)}{s_{00}(t)} \right\} \right] dM_i(t)$$

$$= -\int_0^1 \Psi(w; v) \left\{ \widehat{\beta}(w) - \beta(w) \right\} dw + \mathbf{\Sigma}(v) \left\{ \widehat{\beta}(v) - \beta(v) \right\} - \frac{1}{2} h^2 \mu_2 \mathbf{\Sigma}(v) \ddot{\beta}(v)$$

$$+ o_p(n^{-1/2} + h^2)$$

$$+ O_p\left(\frac{a_n^2}{h} + c_n^2 + d_n^2 + \frac{a_n c_n}{h} + \frac{a_n d_n}{h} + c_n d_n + \frac{a_n e_n}{h} + c_n b_n + d_n b_n \right). \qquad \text{(A.15)}$$

By (A.15), we have,

$$c_n = \sup_{v \in [0,1]} \|\widehat{\beta}(v) - \beta(v)\|$$

$$= O_p\left\{ (nh)^{-1/2} + h^2 + \frac{a_n^2}{h} + \frac{a_n c_n}{h} + \frac{a_n d_n}{h} + \frac{a_n e_n}{h} \right\}. \qquad \text{(A.16)}$$

Similarly,

$$\sup_{v \in [0,1]} h \|\widehat{\dot{\beta}}(v) - \dot{\beta}(v)\|$$

$$= O_p\{ (nh)^{-1/2} + \frac{a_n^2}{h} + \frac{a_n c_n}{h} + \frac{a_n d_n}{h} + \frac{a_n e_n}{h} \} + o_p(h^2). \qquad \text{(A.17)}$$

Further using (A.15), we get

$$\int_0^1 \mathbf{D}_3(w) \left\{ \widehat{\beta}(w) - \beta(w) \right\} dw$$

$$= \frac{1}{2} h^2 \mu_2 \int_0^1 g(v) \mathbf{\Sigma}(v) \ddot{\beta}(v) dv + n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \int_0^1 g(v) \left[\left\{ \mathbf{Z}_i K_i(v; \alpha_0) - \frac{s_{10}(t; v)}{s_{00}(t)} \right\} \right.$$

$$\left. - \mathbf{D}_2(v) \mathbf{D}_1^{-1} \left\{ \mathbf{X}_i W_i - \frac{r_0(t)}{s_{00}(t)} \right\} \right] dv dM_i(t) + o_p(n^{-1/2} + h^2)$$

$$+ O_p\left(\frac{a_n^2}{h} + c_n^2 + d_n^2 + \frac{a_n c_n}{h} + \frac{a_n d_n}{h} + c_n d_n + \frac{a_n e_n}{h} + c_n b_n + d_n b_n \right). \qquad \text{(A.18)}$$

Substituting (A.18) into (A.14), noting that $\int_0^1 g(v) \mathbf{Z}_i K_i(v; \alpha_0) dv = \mathbf{Z}_i g(\mathbf{X}_i' \alpha_0) + O_p(h^2)$, the condition $nh^4 \to 0$, (A.16) and (A.17), we obtain (A.13). Hence, the proof of Theorem 2 is completed. Noting that

$$n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\left\{ -\frac{s_{10}(t; v)}{s_{00}(t)} \right\} - \mathbf{D}_2(v) \mathbf{D}_1^{-1} \left\{ \mathbf{X}_i W_i - \frac{r_0(t)}{s_{00}(t)} \right\} \right] dM_i(t) = O_p(n^{-1/2}),$$

the proof of Theorem 3 is finished by (A.15), (A.16), and (A.17).

Proof of Theorem 4. Let $K_i(v) = K_i(v; \alpha_0)$ and $\eta(v) = (\eta_1', \eta_2(v)')'$ satisfy the following integral equation in $\eta_2 \in \mathcal{C}_0$:

$$\phi_1' = \eta_1' \mathbf{D}_1 + \int_0^1 \eta_2(w)' \mathbf{D}_2(w) dw, \tag{A.19}$$

$$\phi_2(w)' = \eta_1' \mathbf{D}_3(w) + \int_0^\tau \frac{\int_0^1 \eta_2(v)' s_{10}(t; v) dv s_{10}(t; w)'}{s_{00}(t)} \lambda_0(t) dt - \eta_2(w)' \mathbf{\Sigma}(w).$$

Obviously,

$$\eta_1' U_{n1}(\alpha_0, \beta, \dot{\beta}) = n^{-1} \sum_{i=1}^n \int_0^\tau \eta_1' \left[\mathbf{X}_i W_i - \frac{r_0(t)}{s_{00}(t)} \right] dM_i(t) + o_p(n^{-1/2}),$$

and

$$\int_0^1 \eta_2(w)' U_{n2,1}(\alpha_0, \beta, \dot{\beta}, w) dw$$

$$= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[\eta_2(\mathbf{X}_i' \alpha_0)' \mathbf{Z}_i - \frac{S_{\eta_2}(t)}{S_{n0}(t; \alpha_0, \beta)} \right] dM_i(t) + O_p(h^2),$$

where $S_{\eta_2}(t) = (1/n) \sum_{i=1}^n \eta_2(\mathbf{X}_i' \alpha_0)' \mathbf{Z}_i Y_i(t) \exp\{\beta(\mathbf{X}_i' \alpha_0)' \mathbf{Z}_i\}$. Denote $\Upsilon(\alpha, \delta_1, \delta_2, v) = (U_{n1}(\alpha, \delta_1, \delta_2)', U_{n2,1}(\alpha, \delta_1, \delta_2, v)')'$. Using the Martingale Central Limit Theorem we have that

$$n^{1/2} \int_0^1 \eta(w)' \Upsilon(\alpha_0, \beta, \dot{\beta}, w) dw \to N(0, \sigma^2), \tag{A.20}$$

where

$$\sigma^2 = E \left\{ \int_0^\tau \left[\eta_1' \left\{ \mathbf{X} W - \frac{r_0(t)}{s_{00}(t)} \right\} + \{\eta_2(\mathbf{X}' \alpha_0)' \mathbf{Z} - \bar{\eta}_2(t)\} \right]^2 P(t|\mathbf{X}, \mathbf{Z}) \right.$$
$$\left. \times \exp(\beta(\mathbf{X}' \alpha_0)' \mathbf{Z}) \lambda_0(t) dt \right\}$$

and

$$\bar{\eta}_2(t) = \frac{E\{P(t|\mathbf{X}, \mathbf{Z}) \eta_2(\mathbf{X}' \alpha_0)' \mathbf{Z} \exp\{\beta(\mathbf{X}' \alpha_0)' \mathbf{Z}\}\}}{s_{00}(t)}.$$

On the other hand, by (A.3), (A.6) and the condition $nh^4 \to 0$, we have

$$\eta_1' \left(U_{n1}(\widehat{\alpha}, \widehat{\beta}, \widehat{\dot{\beta}}) - U_{n1}(\alpha_0, \beta, \dot{\beta}) \right)$$

$$= \eta_1' \mathbf{D}_1 (\widehat{\alpha} - \alpha_0) + \int_0^1 \eta_1' \mathbf{D}_3(w) \left\{ \widehat{\beta}(w) - \beta(w) \right\} dw + o_p(n^{-1/2}),$$

$$\int_0^1 \eta_2(w)' \left(U_{n2,1}(\widehat{\alpha}, \widehat{\beta}, \widehat{\dot{\beta}}, w) - U_{n2,1}(\alpha_0, \beta, \dot{\beta}, w) \right) dw$$

$$= \int_0^1 \left\{ \int_0^\tau \frac{\int_0^1 \eta_2(v)' s_{10}(t; v) dv s_{10}(t; w)'}{s_{00}(t)} \lambda_0(t) dt - \eta_2(w)' \mathbf{\Sigma}(w) \right\}$$

$$\times \left[\widehat{\beta}(w) - \beta(w)\right] dw + \int_0^1 \eta_2(w)' \mathbf{D}_2(w) dw \, (\widehat{\alpha} - \alpha_0) + o_p(n^{-1/2}).$$

Then by (A.19), we have,

$$\int_0^1 \eta(w)' \left[\Upsilon(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}, w) - \Upsilon(\alpha_0, \beta, \dot{\beta}, w)\right] dw$$

$$= \phi_1'(\widehat{\alpha} - \alpha_0) + \int_0^1 \phi_2(w)' \left\{\widehat{\beta}(w) - \beta(w)\right\} dw + o_p(n^{-1/2}).$$

Note that

$$-\int_0^1 \eta(w)' \Upsilon(\alpha_0, \beta, \dot{\beta}, w) dw = \int_0^1 \eta(w)' \left[\Upsilon(\widehat{\alpha}, \widehat{\beta}, \dot{\widehat{\beta}}, w) - \Upsilon(\alpha_0, \beta, \dot{\beta}, w)\right] dw,$$

by (A.20), we have

$$n^{1/2} \left[\phi_1'(\widehat{\alpha} - \alpha_0) + \int_0^1 \phi_2(w)' \left\{\widehat{\beta}(w) - \beta(w)\right\} dw\right] \to N(0, \sigma^2). \qquad \text{(A.21)}$$

Consider the parametric sub-model

$$\alpha = \alpha_0 + \varrho \eta_1 \quad \text{and} \quad \beta(w; \varrho) = \beta(w) + \varrho \eta_2(w), \qquad \text{(A.22)}$$

where ϱ is an unknown parameter, and α_0, $\beta(w)$, η_1, and $\eta_2(w)$ are fixed parameters or functions. The parameter ϱ can be consistently estimated by the solution $\widehat{\varrho}$ to Cox's partial likelihood score function

$$n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\varsigma_i(\varrho) - \frac{\sum_{j=1}^{n} Y_j(t) \varsigma_j(\varrho) \exp\{\varpi_j(\varrho)\}}{\sum_{j=1}^{n} Y_j(t) \exp\{\varpi_j(\varrho)\}}\right] dN_i(t) = 0,$$

where $\varpi_j(\varrho) = \beta(\mathbf{X}_j'\alpha_0 + \varrho \mathbf{X}_j'\eta_1)' \mathbf{Z}_j + \varrho \eta_2(\mathbf{X}_j'\alpha_0 + \varrho \mathbf{X}_j'\eta_1)' \mathbf{Z}_j$, and

$$\varsigma_i(\varrho) = \mathbf{X}_i'\eta_1 \left\{\dot{\beta}(\mathbf{X}_i'\alpha_0 + \varrho \mathbf{X}_i'\eta_1) + \varrho \dot{\eta}_2(\mathbf{X}_i'\alpha_0 + \varrho \mathbf{X}_i'\eta_1)\right\}' \mathbf{Z}_i + \eta_2(\mathbf{X}_i'\alpha_0 + \varrho \mathbf{X}_i'\eta_1)' \mathbf{Z}_i.$$

Obviously, $\varrho_0 = 0$ is the true value of ϱ. It follows from Andersen and Gill (1982) that

$$\widehat{\varrho} - \varrho_0 = \sigma^{-2} n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\mathbf{X}_i'\eta_1 \dot{\beta}(\mathbf{X}_i'\alpha_0)' \mathbf{Z}_i + \eta_2(\mathbf{X}_i'\alpha_0)' \mathbf{Z}_i\right.$$

$$\left. - \frac{\sum_{j=1}^{n} Y_j(t) \left\{\mathbf{X}_j'\eta_1 \dot{\beta}(\mathbf{X}_j'\alpha_0)' \mathbf{Z}_j + \eta_2(\mathbf{X}_j'\alpha_0)' \mathbf{Z}_j\right\} \exp\{\beta(\mathbf{X}_j'\alpha_0)' \mathbf{Z}_j\}}{\sum_{j=1}^{n} Y_j(t) \exp\{\beta(\mathbf{X}_j'\alpha_0)' \mathbf{Z}_j\}}\right] dM_i(t)$$

$$+ o_p(n^{-1/2}).$$

Under the model (A.22), we have that

$$\phi_1' \left(\widehat{\alpha} - \alpha_0 \right) + \int_0^1 \phi_2(w)' \left\{ \beta(w; \widehat{\varrho}) - \beta(w; \varrho_0) \right\} dw$$

$$= \left(\widehat{\varrho} - \varrho_0 \right) \left[\phi_1' \eta_1 + \int_0^1 \phi_2(w)' \eta_2(w) dw \right],$$

and, using (A.19), we get

$$\phi_1' \eta_1 + \int_0^1 \phi_2(w)' \eta_2(w) dw$$

$$= \int_0^\tau \left\{ \frac{\left(E \left[P(t \mid \mathbf{Z}, \mathbf{X}) \exp\{\beta(\mathbf{X}'\alpha_0)'\mathbf{Z}\} \left\{ W \eta_1' \mathbf{X} + \eta_2(\mathbf{X}'\alpha_0)'\mathbf{Z}\} \right] \right)^2}{s_{00}(t; \alpha_0, \beta)} \right.$$

$$\left. - E \left[P(t \mid \mathbf{Z}, \mathbf{X}) \exp\{\beta(\mathbf{X}'\alpha_0)'\mathbf{Z}\} \left(\eta_1' \mathbf{X} W + \eta_2(\mathbf{X}'\alpha_0)'\mathbf{Z} \right)^2 \right] \right\} \lambda_0(t) dt = -\sigma^2.$$

Thus

$$\phi_1' \left(\widehat{\alpha} - \alpha_0 \right) + \int_0^1 \phi_2(w)' \left\{ \beta(w; \widehat{\varrho}) - \beta(w; \varrho_0) \right\} dw$$

$$= -n^{-1} \sum_{i=1}^n \int_0^\tau \left[\mathbf{X}_i' \eta_1 \dot{\beta}(\mathbf{X}_i'\alpha_0)' \mathbf{Z}_i + \eta_2(\mathbf{X}_i'\alpha_0)' \mathbf{Z}_i \right.$$

$$\left. - \frac{\sum_{j=1}^n Y_j(t) \left\{ \mathbf{X}_j' \eta_1 \dot{\beta}(\mathbf{X}_j'\alpha_0)' \mathbf{Z}_j + \eta_2(\mathbf{X}_j'\alpha_0)' \mathbf{Z}_j \right\} \exp\{\beta(\mathbf{X}_j'\alpha_0)'\mathbf{Z}_j\}}{\sum_{j=1}^n Y_j(t) \exp\{\beta(\mathbf{X}_j'\alpha_0)'\mathbf{Z}_j\}} \right] dM_i(t)$$

$$+ o_p(n^{-1/2}).$$

This gives

$$n^{1/2} \phi_1' \left(\widehat{\alpha} - \alpha_0 \right) + \int_0^1 \phi_2(w)' \left\{ \beta(w; \widehat{\varrho}) - \beta(w; \varrho_0) \right\} dw \to N(0, \sigma^2),$$

which is the same as that of $n^{1/2} \left[\phi_1' \left(\widehat{\alpha} - \alpha_0 \right) + \int_0^1 \phi_2(w)' \left\{ \widehat{\beta}(w) - \beta(w) \right\} dw \right]$ by (A.21). As explained by Bickel et al. (1993), $\phi_1' \widehat{\alpha} + \int_0^1 \phi_2(w)' \widehat{\beta}(w) dw$ is an efficient estimator of $\phi_1' \alpha + \int_0^1 \phi_2(w)' \beta(w) dw$. The proof of Theorem 4 is completed.

An outline for proving the consistency of the initial estimators for α and $\beta(\cdot)$.

Firstly, for given $\alpha \in \Theta$, the model (1.2) reduces to the model proposed by Fan, Lin, and Zhou (2006). Hence we can estimate $\beta(\cdot)$ and $\dot{\beta}(\cdot)$ as does Fan, Lin, and Zhou (2006), denoted $\widehat{\beta}(\cdot|\alpha)$ and $\widehat{\dot{\beta}}(\cdot|\alpha)$, respectively. To estimate α, we use the partial score $U_n(\alpha)$, which is (2.7) with $\beta(\cdot)$ and $\dot{\beta}(\cdot)$ replaced by

their estimators $\widehat{\beta}(\cdot|\alpha)$ and $\widehat{\dot{\beta}}(\cdot|\alpha)$. Denote the solution of $U_n(\alpha) = 0$ as $\widehat{\alpha}$. By Fan, Lin, and Zhou (2006) and the Uniform Law of Large Numbers (Pollard (1990)), for any given $\alpha \in \boldsymbol{\Theta}$ we can show that $\widehat{\beta}(v|\alpha)$ and $\widehat{\dot{\beta}}(v|\alpha)$ converge in probability to nonrandom functions $\beta(v|\alpha)$ and $\dot{\beta}(v|\alpha)$ with $\beta(v|\alpha_0) = \beta(v)$ and $\dot{\beta}(v|\alpha_0) = \dot{\beta}(v)$, uniformly in $v \in [0,1]$ and $\alpha \in \boldsymbol{\Theta}$. Furthermore, we also can show that $\Sigma_n(\alpha) = \partial U_n(\alpha)/\partial \alpha$ converges in probability to a nonrandom function $\Sigma(\alpha)$ uniformly in $\alpha \in \boldsymbol{\Theta}$ and obtain $\Sigma(\alpha)$ is continuous at $\alpha \in \boldsymbol{\Theta}$. Coupled with the assumption of a positive definite matrix $\Sigma(\alpha_0)$, we can conclude that there exists a small neighborhood of α_0 inside of which the eigenvalues of $\Sigma_n(\alpha)$ are bounded away from zero for large n and $\alpha \in \boldsymbol{\Theta}$. By the Uniform Law of Large Numbers, $U_n(\alpha_0) \to 0$ in probability. Then by the Inverse Function Theorem, we have that inside a small neighborhood of α_0, there exists a unique solution $\widehat{\alpha}$ to $U_n(\alpha) = 0$ for all sufficiently large n. This also implies that $\widehat{\alpha}$ is consistent and $\widehat{\beta}(v|\widehat{\alpha}) \to \beta(v|\alpha_0) = \beta(v)$.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer-Verlag, New York.

Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in Cox's regression models. *Scand. J. Statist.* **30**, 93-111.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477-489.

Chen, K., Guo, S., Sun, L. and Wang, J. (2010). Global partial likelihood for nonparametric proportional hazards models. *J. Amer. Statist. Assoc.* **105**, 750-760.

Chen, K., Lin, H. and Zhou, Y. (2012). Efficient estimation for the Cox model with varying coefficients. *Biometrika* **2**, 379-392.

Chen, S. and Zhou, L. (2007). Local partial likelihood estimation in proportional hazards regression. *Ann. Statist.* **35**, 888-916.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B.* **4**, 187-220.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications.* Chapman and Hall, London.

Fan, J., Gijbels, I. and King, M. (1997). Local likelihood and local partial likelihood in hazard regression. *Ann. Statist.* **25**, 1661-1690.

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

Freidlin, B. and Simon, R. (2005). Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clinical Cancer Research* **11**, 7872-7878.

Gamerman, D. (1991). Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika* **85**, 215-227.

Gentleman, R. and Crowley, J. (1991). Local full likelihood estimation for the proportional hazards model. *Biometrics* **47**, 1283-1296.

Hastie, T. and Tibshirani, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics* **46**, 1005-1016.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, **64**, 103-137.

Huang, J. (1999). Efficient estimation of the partly linear additive cox model. *Ann. Statist.* **27**, 1536-1563.

Jiang, W., Freidlin, B. and Simon, R. (2007). Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J. National Cancer Institute* **99**, 1036-1043.

Kooperberg, C., Stone, C. J. and Truong, Y. (1995). Hazard regression. *J. Amer. Statist. Assoc.* **90**, 78-94.

Martinussen, T., Scheike, T.H., Skovgaard, I. M. and Matinerssen, T. (2002). Efficient estimation of fixed and time-varying covariates effects in multiplicative intensity models. *Scand. J. Statist.* **29**, 57-74.

Marzec, L. and Marzec, P. (1997). On fitting Cox's regression model with time-dependent coefficients. *Biometrika* **84**, 901-908.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501-514.

Murphy, S. A. (1993). Testing for a time dependent coefficient in Cox's regression model. *Scand. J. Statist.* **20**, 35-50.

Murphy, S. A. and Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Process. Appl.* **39**, 153-180.

O'Sullivan, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation. *SIAM J. Sci. Statist. Comput.* **9**, 531-542.

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

Sargent, D. J. et al. (2005). Clinical trial designs for predictive marker validation in cancer treatment trials. *J. Clin. Oncology* **23**, 2020-2027.

Shaughnessy, J., Zhan, F., Burington, B. E., Huang, Y., Colla, S., Hanamura, I., Stewart, J. P., Kordsmeier, B., Randolph, C., Williams, D. R., et al. (2007). A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood* **109**, 2276-2284.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *J. Amer. Statist. Assoc.* **82**, 559-567.

Tian, L. Zucker, D. and Wei, L. J. (2005). On the cox model with time-varying regression coefficients. *J. Amer. Statist. Assoc.* **100**, 172-183.

Valsecchi, M. G., Silvestri, D. and Sasieni, P. (1996). Evaluation of long-term survival: use of diagnostics and robust estimators with Cox proportional hazards model. *Statist. Medicine* **15**, 2763-2780.

Zucker, D. M. and Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Statist.* **18**, 329-353.

Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, China

E-mail: linhz@swufe.edu.cn

Department of Biostatistics, Bioinformatics & Biomathematics, Georgetown University, Washington D.C. 20057, Ann Arbor, MI 48109, USA.

E-mail: mtt34@georgetown.edu

Department of Biostatistics, University of Michigan, USA.

E-mail: yili@med.umich.edu