University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working Paper Series

Year 2014 *Paper* 111

A global partial likelihood estimator of the time-varying effects for time-dependent treatment

Huazhen Lin* Zhe Fei† Yi Li‡

*Southwestern University of Finance and Economics, linhz@swufe.edu.cn

†University of Michigan School of Public Health

‡University of Michigan - Ann Arbor

A global partial likelihood estimator of the time-varying effects for time-dependent treatment

Huazhen Lin, Zhe Fei, and Yi Li

Abstract

The timing of time-dependent treatment - e.g., when to perform kidney transplantation - is an important factor for evaluating treatment efficacy. A naive comparison between the treatment and nontreatment groups, while ignoring the timing of treatment, typically yields results that might biasedly favor the treatment group, as only patients who survive long enough will get treated. On the other hand, studying the effect of time-dependent treatment is often complex, as it involves modeling treatment history and accounting for the possible time-varying nature of the treatment effect. We propose a varying-coefficient Cox model that investigates the efficacy of time-dependent treatment by utilizing a global partial likelihood, which renders appealing statistical properties, including consistency, asymptotic normality and semiparametric efficiency. Extensive simulations verify the finite sample performance, and we apply the proposed method to study the efficacy of kidney transplantation for end-stage renal disease patients in the U.S. Scientific Registry of Transplant Recipients (SRTR).

A global partial likelihood estimator of the time-varying effects for time-dependent treatment

Huazhen Lin

Center of Statistical Research, School of Statistics,

Southwestern University of Finance and Economics,

Chengdu, Sichuan, People's Republic of China

email: linhz@swufe.edu.cn

Zhe Fei, Yi Li

Department of Biostatistics, University of Michigan, USA

SUMMARY The timing of time-dependent treatment—e.g., when to perform kidney transplantation—is an important factor for evaluating treatment efficacy. A naïve comparison between the treatment and nontreatment groups, while ignoring the timing of treatment, typically yields results that might biasedly favor the treatment group, as only patients who survive long enough will get treated. On the other hand, studying the effect of time-dependent treatment is often complex, as it involves modeling treatment history and accounting for the possible time-varying nature of the treatment effect. We propose a varying-coefficient Cox model that investigates the efficacy of time-dependent treatment by utilizing a global partial likelihood, which renders appealing statistical properties, including consistency, asymptotic normality and semiparametric efficiency. Extensive simulations verify the finite sample performance, and we apply the proposed method to study the efficacy of kidney transplantation for end-stage renal disease patients in the U.S. Scientific Registry of Transplant Recipients (SRTR).

KEY WORDS: Cox proportional hazards model; efficient; survival data; time-dependent treatment; varying-coefficient.

1 Introduction

This paper is motivated by the study of a national cohort of kidney transplant patients from the U.S. Scientific Registry of Transplant Recipients (SRTR), which is collected by the United Network for Organ Sharing and Organ Procurement and Transplantation Network (UNOS/OPTN) for all wait-listed kidney transplant candidates and transplant recipients in the United States. When a donor kidney becomes available, medical judgment is used to select the patient who should receive it. In the state of Michigan, 1446 of the 3115 patients on the waitlist between 2008 and 2011 received a kidney transplant. A naïve comparison of the survival times of nontransplanted patients with those of transplanted patients will yield biased results, as only those who survive long enough to receive a kidney will receive treatment. Moreover, the risks associated with surgery lead to an immediate peak in a patient's death hazard following transplantation, which gradually decreases when the patient stabilizes.

To accommodate these two distinguishing features—the time dependence of the treatment and the time-varying nature of the treatment effect—we propose the following time-varying Cox model. Patient $i = 1, \ldots, n$, is denoted by a binary time-dependent covariate $x_i(t)$, which is equal to 1 if the patient has received treatment (kidney transplant) by time t and equal to 0 otherwise. If t_i is the time of treatment, then $x_i(t) = I(t \geq t_i)$ describes the treatment process of patient i. If a patient never received treatment, $t_i = \infty$ or $x_i(t) \equiv 0$ for $t \geq 0$. Conditional on the treatment history and in the absence of confounders, we model the hazard of death for patient i as follows

$$\lambda_i(t) = \lambda_0(t) \exp\{x_i(t)\beta(t - t_i)\}, \tag{1.1}$$

where $\beta(s)$ is an unknown smoothing function defined when $s \geq 0$, to explore whether and how the treatment effect varies over time since treatment. To avoid ambiguity, we define $\beta(s) = 0$ for $s < 0$. Model (1.1) reveals that patient i has the baseline hazard $\lambda_0(t)$ at time $t < t_i$ (i.e., prior to treatment). Once $t \geq t_i$, patient i enters the treatment group, with the treatment effect initiating at t_i. The size of the effect depends on $t - t_i$, as observed in transplant studies, that the risk of death peaks right after treatment and then gradually decreases until the kidney transplant shows protective effects.

To compensate for multi-level treatment (e.g., different dose levels or different modalities of treatment) and to adjust for possible confounders (e.g., gender, BMI, previous malignancy, diabetes), we consider a more general partial time-varying coefficient Cox model

$$\lambda_i(t) = \lambda_0(t) \exp\{\mathbf{z}_i(t)'\boldsymbol{\alpha} + \mathbf{x}_i(t)'\boldsymbol{\beta}(t - t_i)\}. \tag{1.2}$$

Although model (1.2) resembles the time-dependent coefficient (TDC) Cox model proposed by a number of authors, including Zucker and Karr (1990), Murphy and Sen (1991), Gamerman (1991), Murphy (1993), Marzec and Marzec (1997), Martinussen *et al.* (2002), Cai and Sun (2003), Tian, Zuker and Wei (2002) and Fan, Lin and Zhou (2006), it differs in that $\boldsymbol{\beta}(\cdot)$ in our model is a vector of functions of the gap time $t - t_i$, as opposed to the current "calendar" time t. As a result, the estimation of model (1.2) is more involved than that for the traditional TDC Cox model. The traditional nonparametric technique (e.g., kernel smoothing) is not directly applicable to the model (1.2) because of the individual-specific argument of the function $\boldsymbol{\beta}(\cdot)$.

The local partial likelihood, which is based on observations with T_i in a small neighborhood of a given t, has been widely used to estimate the TDC Cox model (Cai and Sun, 2003; Tian, Zuker and Wei, 2002; Fan, Lin and Zhou, 2006). However, it suffers efficiency loss, as the observations outside the neighborhood which carry information about $\boldsymbol{\beta}(t)$ are not used. Instead, we propose to draw inference based on a full partial likelihood function, the main intuition of which is to utilize all observations for the estimation of $\boldsymbol{\beta}(t)$. The superiority of this proposed method is reflected in its semiparametric efficiency in terms of linear functionals (Bickel *et al.*, 1993). Finally, we also show that the proposed estimator is uniformly consistent and asymptotically normal.

The remainder of the paper is organized as follows. Section 2 presents the estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}(\cdot)$. Asymptotic distribution properties, including efficiency, of the estimators are provided in Section 3. The simulation studies we used to investigate the performance of the proposed estimators are presented in Section 4. Our analysis of the kidney transplantation data is given in Section 5. Technical proofs are relegated to the Supplementary material.

2 Estimation

2.1 Global Partial Likelihood Approach

Assume a random sample of size n from a population of patients. For the ith individual, let T_i be the potential failure time, C_i the potential censoring time and $\mathcal{T}_i = \min(T_i, C_i)$ the observed failure time. To avoid the technicality at the tail, we study patients' survival experience over $[0, \tau]$, where τ is such that $P(\min(T_i, C_i) > \tau) > 0$ and, in practice, is often the study duration. Assume that T_i and C_i are independent given the covariate process $\mathbf{X}_i = \{(\mathbf{x}_i(t), \mathbf{z}_i(t)), 0 \leq t \leq \tau\}$, where $\mathbf{x}_i(t), \mathbf{z}_i(t)$ are p- and q-dimensional vector functions respectively. Let Δ_i be an indicator that equals 1 if $\mathcal{T}_i$ is a failure time and 0 otherwise. Let t_i be the treatment time; if the treatment does not occur prior to τ, we set $t_i = \infty$. Thus, the observed data structure is

$$\{\mathcal{T}_i, \Delta_i, \mathbf{X}_i, t_i\} \quad \text{for } i = 1, \cdots, n.$$

Under model (1.2), if $\boldsymbol{\beta}(\cdot)$ were parameterized it could be estimated by maximizing the partial likelihood:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ \frac{\exp\left[\boldsymbol{\alpha}'\mathbf{z}_i(\mathcal{T}_i) + \mathbf{x}_i(\mathcal{T}_i)'\boldsymbol{\beta}(\mathcal{T}_i - t_i)\right]}{\sum_{\ell \in \mathcal{R}(\mathcal{T}_i)} \exp\left[\boldsymbol{\alpha}'\mathbf{z}_\ell(\mathcal{T}_i) + \mathbf{x}_\ell(\mathcal{T}_i)'\boldsymbol{\beta}(\mathcal{T}_i - t_\ell)\right]} \right\}^{\Delta_i}, \tag{2.1}$$

where $\boldsymbol{\beta}(\cdot)$ and $\boldsymbol{\alpha}$ are $p-$ and $q-$dimensional vectors, respectively, and $\mathcal{R}(t) = \{i : \mathcal{T}_i \geq t\}$ denotes the set of the individuals at risk just prior to time t. If the functional form of $\boldsymbol{\beta}(\cdot)$ is not available, it may seem natural to take the local likelihood approach; however, as shown below, a direct application of the local likelihood approach does not work.

To be specific, we assume that each component of $\boldsymbol{\beta}(s) = (\beta_1(s), \ldots, \beta_p(s))'$ is smooth when $s > 0$ and admits a Taylor expansion. For a given $t > 0$ and $v > 0$ around t,

$$\boldsymbol{\beta}(v) \approx \boldsymbol{\beta}(t) + \dot{\boldsymbol{\beta}}(t) \times (v - t). \tag{2.2}$$

Denote $\boldsymbol{\delta} = \boldsymbol{\beta}(t)$ and $\boldsymbol{\eta} = \dot{\boldsymbol{\beta}}(t) = (d\beta_1(t)/dt, \ldots, d\beta_p(t)/dt)'$. Let $K_h(\cdot) = K(\cdot/h)/h$, where K is a kernel function and h represents the size of the local neighborhood. Substituting (2.2) into (2.1), we estimate $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$ by maximizing the following logarithm of the local partial likelihood:

$$\sum_{i=1}^{n} \Delta_i \log \left\{ \frac{\exp\left(\boldsymbol{\alpha}'\mathbf{z}_i(\mathcal{T}_i) + \mathbf{x}_i(\mathcal{T}_i)'[\boldsymbol{\delta} + \boldsymbol{\eta} \times (\mathcal{T}_i - t_i - t)]\right)}{\sum_{\ell \in \mathcal{R}(\mathcal{T}_i)} \exp\left[\boldsymbol{\alpha}'\mathbf{z}_\ell(\mathcal{T}_i) + \mathbf{x}_\ell(\mathcal{T}_i)'\boldsymbol{\beta}(\mathcal{T}_i - t_\ell)\right]} \right\} K_h(\mathcal{T}_i - t_i - t). \tag{2.3}$$

When the weight $K_h(\mathcal{T}_i - t_i - t) > 0$, it implies that $\mathcal{T}_i - t_i$ is in the neighborhood of t, and hence $\boldsymbol{\beta}(\mathcal{T}_i - t_i)$ can be replaced by $\boldsymbol{\delta} + \boldsymbol{\eta}(\mathcal{T}_i - t_i - t)$. However, the $\boldsymbol{\beta}(\mathcal{T}_i - t_\ell)$ in the denominator of (2.3) cannot be approximated by $\boldsymbol{\delta} + \boldsymbol{\eta}(\mathcal{T}_i - t_\ell - t)$ because $\mathcal{T}_i - t_\ell$ could be outside the neighborhood of t when $\ell \neq i$, nullifying the Taylor expansion. Thus, with an unknown $\boldsymbol{\beta}(\cdot)$, the local partial likelihood method (2.3) cannot estimate $\boldsymbol{\beta}(\cdot)$ in our model (1.2), as would be the case with the traditional TDC Cox model.

Our new approach stems from the following observation. Denote $\overline{\psi}_i(u) = \boldsymbol{\alpha}'\mathbf{z}_i(u) + \mathbf{x}_i(u)'[\boldsymbol{\delta} + \boldsymbol{\eta} \times (u - t_i - t)]$ and $\psi_i(u) = \boldsymbol{\alpha}'\mathbf{z}_i(u) + \mathbf{x}_i(u)'\boldsymbol{\beta}(u - t_i)$. Thus,

$$\begin{aligned} \psi_i(u) &= hK_h(u - t_i - t)\psi_i(u) + \{1 - hK_h(u - t_i - t)\}\psi_i(u) \\ &\approx hK_h(u - t_i - t)\overline{\psi}_i(u) + \{1 - hK_h(u - t_i - t)\}\,\psi_i(u). \end{aligned} \tag{2.4}$$

Substituting (2.4) into (2.1), we estimate $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$ by maximizing the following logarithm of the full partial likelihood:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \Delta_i \left\{ hK_h(\mathcal{T}_{ii} - t)\overline{\psi}_i(\mathcal{T}_i) + \{1 - hK_h(\mathcal{T}_{ii} - t)\}\, \psi_i(\mathcal{T}_i) \right.$$
$$\left. - \log\left(\sum_{\ell \in \mathcal{R}(\mathcal{T}_i)} \left[hK_h(\mathcal{T}_{i\ell} - t) \exp\left\{\overline{\psi}_\ell(\mathcal{T}_i)\right\} + \{1 - hK_h(\mathcal{T}_{i\ell} - t)\} \exp\left\{\psi_\ell(\mathcal{T}_i)\right\} \right] \right) \right\}, \tag{2.5}$$

where $\mathcal{T}_{i\ell} = \mathcal{T}_i - t_\ell$. Because the estimator based on (2.5) is a standard partial likelihood estimator rather than a local partial likelihood estimator, we term the proposed estimator a global partial likelihood estimator. Notice that the proposed method (2.5) uses observations, whether within or outside the neighborhood of t, to estimate $\boldsymbol{\beta}(t)$. As a result, our estimator is more efficient than the traditional local linear technique; this is demonstrated in Theorem 3 from Section 3, which shows that the proposed estimator is semiparametrically efficient.

2.2 An Iterative Algorithm for Estimation

Because (2.5) depends on the unknown $\boldsymbol{\beta}(\cdot)$, it is not directly useful for estimation. However, the form of (2.5) naturally leads to an iterative algorithm. Specifically, suppose we choose initial values of $\boldsymbol{\alpha}$ and function $\boldsymbol{\beta}(\cdot)$, then we perform the following iterations until convergence.

Step 1 of iteration m. For every given $t \in [0, \tau]$, solve the following equations for $\boldsymbol{\xi} = (\boldsymbol{\delta}', h\boldsymbol{\eta}')'$:

$$\ell(\boldsymbol{\delta}, \boldsymbol{\eta}) = \frac{1}{n} \sum_{i=1}^{n} \Delta_i \left\{ \mathbf{w}_i(\mathcal{T}_i) K_h(\mathcal{T}_{ii} - t) \right.$$
$$\left. - \left[\frac{\sum_{\ell \in \mathcal{R}(\mathcal{T}_i)} \mathbf{w}_\ell(\mathcal{T}_i) K_h(\mathcal{T}_{i\ell} - t) \exp\left(\boldsymbol{\alpha}^{[m-1]\prime} \mathbf{z}_\ell(\mathcal{T}_i) + \boldsymbol{\xi}' \mathbf{w}_\ell(\mathcal{T}_i)\right)}{\sum_{r \in \mathcal{R}(\mathcal{T}_i)} \exp\left(\boldsymbol{\alpha}^{[m-1]\prime} \mathbf{z}_r(\mathcal{T}_i) + \mathbf{x}_r(\mathcal{T}_i)' \boldsymbol{\beta}^{[m-1]}(\mathcal{T}_{ir})\right)} \right] \right\} = 0, \tag{2.6}$$

where $\mathbf{w}_i(u) = \begin{pmatrix} \mathbf{x}_i(u) \\ (u - t_i - t)\mathbf{x}_i(u)/h \end{pmatrix}$. Let $\widehat{\boldsymbol{\delta}}$ and $\widehat{\boldsymbol{\eta}}$ be the solutions of $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$. Thus, $\boldsymbol{\beta}^{[m]}(t) = \widehat{\boldsymbol{\delta}}$. The entire estimated function $\boldsymbol{\beta}^{[m]}(\cdot)$ is obtained by using the above procedures with t varying in $[0, \tau]$.

Step 2 of iteration m. Update $\boldsymbol{\alpha}$ by solving the following equations for $\boldsymbol{\alpha}$:

$$\sum_{i=1}^{n} \Delta_i \left\{ \mathbf{z}_i(\mathcal{T}_i) - \frac{\sum_{\ell \in \mathcal{R}(\mathcal{T}_i)} \mathbf{z}_\ell(\mathcal{T}_i) \exp\left[\boldsymbol{\alpha}'\mathbf{z}_\ell(\mathcal{T}_i) + \mathbf{x}_\ell(\mathcal{T}_i)'\boldsymbol{\beta}^{[m]}(\mathcal{T}_{i\ell})\right]}{\sum_{\ell \in \mathcal{R}(\mathcal{T}_i)} \exp\left[\boldsymbol{\alpha}'\mathbf{z}_\ell(\mathcal{T}_i) + \mathbf{x}_\ell(\mathcal{T}_i)'\boldsymbol{\beta}^{[m]}(\mathcal{T}_{i\ell})\right]} \right\} = 0. \qquad (2.7)$$

To facilitate further derivations, we express our global partial likelihood by using the counting process notation. Let $N_i(t) = I(T_i \le t, \Delta_i = 1)$ and $Y_i(t) = I(\mathcal{T}_i \ge t)$. Then, (2.6) and (2.7) can be expressed as

$$\frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau} \left\{ \mathbf{w}_i(u)K_h(u - t_i - t) - \sum_{\ell=1}^{n} \mathbf{w}_\ell(u)K_h(u - t_\ell - t) \times \frac{Y_\ell(u)\exp\left(\boldsymbol{\alpha}^{[m-1]\prime}\mathbf{z}_\ell(u) + \boldsymbol{\xi}'\mathbf{w}_\ell(u)\right)}{\sum_{r=1}^{n} Y_r(u)\exp\left(\boldsymbol{\alpha}^{[m-1]\prime}\mathbf{z}_r(u) + \mathbf{x}_r(u)'\boldsymbol{\beta}^{[m-1]}(u - t_r)\right)} \right\} dN_i(u) = 0, \quad (2.8)$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau} \left\{ \mathbf{z}_i(u) - \frac{\sum_{\ell=1}^{n} Y_\ell(u)\mathbf{z}_\ell(u)\exp\left[\boldsymbol{\alpha}'\mathbf{z}_\ell(u) + \mathbf{x}_\ell(u)'\boldsymbol{\beta}^{[m]}(u - t_\ell)\right]}{\sum_{\ell=1}^{n} Y_\ell(u)\exp\left[\boldsymbol{\alpha}'\mathbf{z}_\ell(u) + \mathbf{x}_\ell(u)'\boldsymbol{\beta}^{[m]}(u - t_\ell)\right]} \right\} dN_i(u) = 0. \qquad (2.9)$$

Without ambiguity, we let $\widehat{\boldsymbol{\xi}}(t)$ and $\widehat{\boldsymbol{\alpha}}$ be the solutions of (2.8) and (2.9), respectively.

3 Large Sample Properties

The uniform consistency, asymptotic normality and semiparametric efficiency of the proposed estimator are established. Some regularity conditions are required and presented in the Appendix. The detailed proofs are given in the Supplementary material.

Theorem 1 *Under Conditions 1-8 stated in the Appendix, we have*

$$\sup_{0<t<\tau} \|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\| \to 0 \ \textit{in probability}$$

and

$$\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\| \to 0 \ \textit{in probability.}$$

Theorem 2 *Under Conditions 1-8 stated in the Appendix, if* $nh^4 = o(1)$*, then*

$$n^{1/2}\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) \to N(0, \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{A}^{-1}\right)'),$$

where $\mathbf{A}$ *and* $\mathbf{B}$ *are defined in the Appendix.*

To estimate the parameter $\boldsymbol{\alpha}$ at the rate $n^{-1/2}$, one must undersmooth the nonparametric part, requiring $nh^4 = o(1)$. The need to undersmooth for achieving usual parametric rates of convergence is standard in the kernel literature and has analogs in the spline literature (Carroll *et al.*, 1997; Hastie and Tibshirani, 1990). Obviously, the optimal bandwidth h to estimate $\boldsymbol{\alpha}$ is not adaptive to estimate function $\boldsymbol{\beta}(\cdot)$. Hence, we need one extra step to estimate $\boldsymbol{\beta}(\cdot)$. At this final step, fixing $\boldsymbol{\alpha}$ at its estimated value, we estimate $\boldsymbol{\beta}(\cdot)$ by (2.8), while taking the bandwidth h to be the estimated optimal bandwidth for the estimation of $\boldsymbol{\beta}(\cdot)$. Our estimator for $\boldsymbol{\beta}(\cdot)$ is consistent and asymptotically normal as implied by the following theorem.

Theorem 3 *Under Conditions 1-8 stated in the Appendix, for* $0 < t < \tau$*, we have the following Fredholm integral equation,*

$$\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) = \vartheta_2^{-1}(t) \int_0^\tau \vartheta_1(t,v)\left(\widehat{\boldsymbol{\beta}}(v) - \boldsymbol{\beta}(v)\right) dv$$
$$+ (nh)^{-1/2}\Sigma_0(t)\varphi + \frac{1}{2}h^2\mu_2\ddot{\boldsymbol{\beta}}(t) + o_p(h^2 + (nh)^{-1/2}),$$

where $\Sigma_0(t)\Sigma_0'(t) = \nu_0\vartheta_2^{-1}(t) - \vartheta_2^{-1}(t)\vartheta_1(t,t)\vartheta_2^{-1}(t)$, φ *is a standard normal random vector,* $\vartheta_1(\cdot,\cdot)$ *and* $\vartheta_2(\cdot)$ *are defined in the Appendix, and* $\nu_0 = \int K^2(x)dx$.

Denote with $\mathcal{B}$ the linear operator satisfying

$$\mathcal{B}(\phi)(t) = \vartheta_2^{-1}(t)\int_0^\tau \vartheta_1(t,v)\phi(v)dv$$

for any function ϕ. Theorem 3 implies that

$$\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) = (nh)^{-1/2}(I-\mathcal{B})^{-1}(\Sigma_0)(t)\varphi + \frac{1}{2}h^2\nu_2(I-\mathcal{B})^{-1}(\ddot{\boldsymbol{\beta}})(t) + o_p(h^2 + (nh)^{-1/2}).$$

Hence, $\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)$ is asymptotically normal, the order of the asymptotic bias of $\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)$ is h^2 and the order of the asymptotic covariance is $(nh)^{-1}$. Theorem 3 also implies that the bias and variance of $\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)$ are the same as if $\boldsymbol{\alpha}$ were known. This result is due to the fact that the rate of convergence of $\widehat{\boldsymbol{\alpha}}$ is faster than that of $\widehat{\boldsymbol{\beta}}(t)$, so that the uncertainty from $\widehat{\boldsymbol{\alpha}}$ can be ignored.

Theorem 2 shows that $\widehat{\boldsymbol{\alpha}}$ is an $n^{1/2}-$ consistent and asymptotically normal estimator of $\boldsymbol{\alpha}$. Moreover, the following Theorem shows that $\widehat{\boldsymbol{\alpha}}$ is also an efficient estimator of $\boldsymbol{\alpha}$. For any vector of functions $\phi(t) = (\phi_1', \phi_2(t)')'$, which has a continuous second derivative on $[0,\tau]$, let $\phi_1'\widehat{\boldsymbol{\alpha}} + \int_0^\tau \phi_2'(t)\widehat{\boldsymbol{\beta}}(t)dt$ be an estimator of $\phi_1'\boldsymbol{\alpha}_0 + \int_0^\tau \phi_2'(t)\boldsymbol{\beta}(t)dt$, we have the following efficiency result.

Theorem 4 *Under Conditions 1-8 stated in the Appendix, if* $nh^4 = o(1)$*, then* $\phi_1'\widehat{\boldsymbol{\alpha}} + \int_0^\tau \phi_2'(t)\widehat{\boldsymbol{\beta}}(t)dt$ *is an efficient estimator of* $\phi_1'\boldsymbol{\alpha}_0 + \int_0^\tau \phi_2'(t)\boldsymbol{\beta}(t)dt$.

Hence, by taking $\phi_2(t) = 0$, we know that $\widehat{\boldsymbol{\alpha}}$ is an efficient estimator of $\boldsymbol{\alpha}_0$. By taking $\phi_1(t) = 0$, then $\int_0^\tau \phi_2'(t)\widehat{\boldsymbol{\beta}}(t)dt$ is an efficient estimator of $\int_0^\tau \phi_2'(t)\boldsymbol{\beta}(t)dt$.

To use (2.6), we need to choose the bandwidth h. Because the leading terms in Theorem 2 do not depend on h, we conclude that the bandwidth h is not crucial for the asymptotic performance of the estimates for the parameters $\boldsymbol{\alpha}$; this conclusion is also confirmed in our simulation studies. A practical implication is that our estimates are not sensitive to the bandwidth h, which makes a roughly estimated h sufficiently good for estimating the parameters $\boldsymbol{\alpha}$.

However, the selection of h is crucial for the asymptotic performance of the estimates for $\boldsymbol{\beta}(\cdot)$. We use the $K-$fold cross-validation procedure for bandwidth selection of the function, which is commonly used in the literature (Efron and Tibshirani, 1993; Tian, Zucker and Wei, 2005; Fan, Lin and Zhou, 2006). Tian, Zucker and Wei (2005) and Fan, Lin and Zhou (2006) have shown empirically that the choice of the smoothing parameter can be quite flexible. Our simulations and example also show that the cross-validation approach works well. See Section 5 for a detailed description.

Because the expression for the covariance matrices of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}(t)$ involves complicated unknown functions, it is difficult to obtain an estimate for the covariance matrices based on Theorems 2 and 3. As a remedy, we propose to use a resampling scheme—for example, a bootstrap method—to approximate the variances or covariance matrices.

4 Simulation

We conduct simulation studies to investigate the finite sample performance of the proposed method. In the following simulations and examples, we use the Epanechnikov kernel. We conduct 200 simulations for each configuration.

Simulation 1: We first consider a nonparametric model

$$\lambda_i(t) = \exp\left\{X_i(t)\beta(t - t_i)\right\}, \tag{4.1}$$

where $\beta(t) = -2t(t - 0.8)$, $X_i(t) = I(t > t_i)$ and t_i is uniformly distributed on $[0, 1]$. The censoring random variable C_i is distributed uniformly on $[0, 4]$, so that about $25\% \sim 35\%$ of data is censored. We simulated 200 datasets each consisting of

$n = 400$ subjects.

To investigate the performance of our estimator, we compare the proposed method with an ideal model, wherein $\beta(\cdot)$ is correctly specified up to a finite-dimensional parameter. In particular, we fit the data using the following ideal model: $\lambda_i(t) = \lambda_0(t)\exp\{X_i(t)\beta_0(t-t_i)\}$, where $\beta_0(t) = \theta_1 t^2 + \theta_2 t + \theta_3$, and θ_1, θ_2 and θ_3 are unknown parameters. The estimator based on the ideal model is designated as "ideal" and serves as the gold standard we use to investigate the efficiency of the proposed estimator. Denote $\nu = \sum_{k=1}^{n_{grid}} \beta(w_k)/n_{grid}$, where $\{w_k, k = 1, \cdots, n_{grid}\}$ are the uniformly distributed grid points in which the function $\beta(\cdot)$ is estimated and $n_{grid} = 200$.

For each generated dataset, we estimate $\beta(\cdot)$ and ν using the proposed method with $h = 0.4$ and the ideal method. We assess the performance of estimator $\widehat{\nu}$ via the absolute errors (AEs), AE $= |\widehat{\nu} - \nu|$. Figure 1(a) displays the averaged estimated function for $\beta(\cdot)$ and its 95% empirical point-wise confidence limits using the proposed method, which shows that the proposed estimates are close to the true functions. Figure 1(b) displays the distribution of AE based on the 200 simulated datasets using the proposed method and the ideal model. It appears that the AE of the proposed estimator is comparable to that of the ideal estimator, confirming the high efficiency of our estimator.

[Figure 1 about here.]

Simulation 2: Now we consider the following mixed model

$$\lambda_i(t) = t^{2/5} \exp\{\boldsymbol{\alpha}'\mathbf{Z}_i + X_i(t)\beta(t-t_i)\}, \qquad (4.2)$$

where $\beta(t) = \log(t)$, $\boldsymbol{\alpha} = (1/2, 1/2)'$, $\mathbf{Z}_i = (Z_{i1}, Z_{i2})'$, Z_{i1} is a Poisson variable with mean 0.2, Z_{i2} is a uniform variable on $[0, 0.7]$, $X_i(t) = I(t \geq t_i)$, and t_i is uniformly distributed on $[0, 1]$. The censoring random variable C is distributed uniformly on $[0, 4]$, so that about $25\% \sim 35\%$ of data is censored.

For each generated dataset consisting of $n = 500$ subjects, we use the proposed method and the ideal model: $\lambda_i(t) = \lambda_0(t)\exp\{\boldsymbol{\alpha}'\mathbf{Z}_i + X_i(t)\beta_0(t-t_i)\}$, where $\beta_0(t) = \theta_1 + \theta_2 log(t)$, and $\boldsymbol{\alpha}, \theta_1$ and θ_2 are unknown parameters.

We estimate $\boldsymbol{\alpha}$, $\beta(\cdot)$ and ν using the proposed method with $h = 0.5$ and the ideal method. Figure 2(a) displays the averaged estimated function for $\beta(\cdot)$ and the 95% empirical pointwise confidence limits of the proposed method, Figure 2(b) displays the distribution of AE based on the 200 simulated datasets. Figure 2 yields similar conclusions to Figure 1 for Simulation 1 so we have omitted a detailed discussion.

Table 1 provides the bias, empirical standard deviation (SD) and the root of mean squared error (RMSE) of the coefficient parameter estimators based on the 200 replications, using the proposed method and the ideal method. It is apparent that these two estimators perform similarly, indicating the high efficiency of our estimator.

[Table 1 and Figure 2 about here.]

5 The Kidney Transplant Program

We study kidney transplant patients from the U.S. Scientific Registry of Transplant Recipients (SRTR), which is collected by the United Network for Organ Sharing and Organ Procurement and Transplantation Network (UNOS/OPTN) for all wait-listed kidney transplant candidates and transplant recipients in the United States. The studied population includes adults with no history of kidney transplant who were on the waitlist between January 1, 2008 and December 31, 2011 in the state of Michigan (n=3115 patients). Of these, 1446 patients received a kidney transplant, with the waiting time for a transplant ranging from 0 to 1787 days (M=341.7 days; SD=380.3 days). The predictors used to adjust for transplant effect are gender (Z_{i1}), BMI (Z_{i2}), previous malignancy (PM; Z_{i3}), maximum acceptable cold ischemic time (MACIT;

Z_{i4}) and diabetes (Z_{i5}). Table 2 lists some descriptive statistics for these variables.

[Table 2 about here.]

Because the treatment is time dependent, we analyze the SRTR data using the proposed method. Denote $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, \cdots, Z_{i5})'$, t_i is the transplant time for ith patient, $\mathbf{x}_i(t) = I(t \geq t_i)$ is the indicator for patient i having received a kidney transplant at time t. We fit the data using the following model

$$\lambda_i(t) = \lambda_0(t) \exp\{\mathbf{Z}_i'\boldsymbol{\alpha} + \mathbf{x}_i(t)\beta(t - t_i)\}.$$

Due to the long span of follow-up time (more than 4 years) and nonuniformly distributed event times (number of death decreases linearly after about 3 years), we use the adaptive bandwidth (Brockmann *et al.*, 1993). We selected the adaptive bandwidth for each time point so that it covered a fixed quantile, q, of total number of events. The $q = 0.15$ chosen by K-fold cross-validation (Cai *et al.*, 2000; Fan, Lin and Zhou, 2006) to minimize the prediction error

$$\int_0^\tau \left(N_i(t) - \widehat{E}N_i(t)\right)^2 d\left\{\sum_{k=1}^n N_k(t)\right\},$$

where $\widehat{E}N_i(t) = \int_0^t Y_i(u) \exp\left(\mathbf{Z}_i'\widehat{\boldsymbol{\alpha}} + \mathbf{x}_i(u)\widehat{\beta}(u - t_i)\right) d\widehat{\Lambda}_0(u)$ is the estimate of the expected failure number up to time t, $\widehat{\Lambda}_0(t) = \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{dN_i(u)}{n^{-1}\sum_{j=1}^n Y_j(u)\exp\left(\mathbf{Z}_j'\widehat{\boldsymbol{\alpha}} + \mathbf{x}_j(u)\widehat{\beta}(u-t_j)\right)}$. We choose $K = 5$. Figure 3 displays the estimated transient effect of kidney transplantation and its 95% confidence limits (dotted lines) obtained from the 200 bootstrap samples. We chose a sample size of 200 by monitoring the stability of the standard errors. Table 3 displays the estimated coefficients of the adjusting covariates. The proposed coefficient estimators in Table 3 show that lower BMI has a significant protective effect on hazard of death, and patients without PM or diabetes experience better survival outcome than those with these two conditions. Gender and MACIT have no significant effect on a patient's survival status. As a comparison, we also analyze the data using the classical Cox proportional hazards model, which has the form

$$\lambda_i(t) = \lambda_0(t) \exp\{\mathbf{Z}_i'\boldsymbol{\alpha} + x_i\beta\},$$

where x_i is the indicator for patient i receiving a kidney transplant. The model yields estimates similar to those presented in Table 2, which hints at robustness of the proposed method.

Figure 3 shows that a patient's mortality hazard decreases across the first 500 days after transplantation. If the patient survives the first 240 days after transplantation, he gains a statistically significant benefit from the transplantation. This large-scale, data-based result is significant, as it has implications for optimal organ allocation and post-transplant care.

[Figure 3 and Table 3 about here.]

6 Conclusion

To properly account for the timing of a time-dependent treatment when evaluating treatment efficacy, we propose a time-dependent treatment and coefficient Cox model. To increase efficiency, we utilize a global partial likelihood, which renders appealing statistical properties, including consistency, asymptotic normality and semiparametric efficiency. Simulation studies verify the finite sample performance; we applied the proposed method to study the efficacy of kidney transplantation among patients with end-stage renal disease patients, which yields interesting results.

There are, however, several opportunities for future research. First, we have implicitly assumed that for each individual the treatment time is a scalar. Although this is a useful assumption for the current examination of kidney transplant patients, in reality, a single patient may have data available on multiple treatments and multiple treatment times. Our method can be extended to cover this case, using more involved computation. Second, in an observational setting healthier patients may be more likely to receive treatment; thus, efficacy analyses of treatment should account for possible selection bias. Finally, it is worth investigating the integration of marginal structural equations or propensity matching into our framework.

Acknowledgment

Lin's research is supported by the Chinese Natural Science Foundation (11125104) and the Fundamental Research Funds for the Central Universities. Li's research is supported by the U.S. National Institutes of Health.

References

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.

Brockmann, M., Gasser T., Herrmann E. (1993). Locally adaptive bandwidth choice for kernel regression estimators.*J Am Stat Assoc*, 88.424;1302–1309.

Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J Am Stat Assoc*, 95;888–902.

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

Cavender, J.B., Rogers, W.J., Fisher, L.D., Gersh, B.J., Coggin, J.C., Myers, W.O. (1992). Effects of smoking on survival and morbidity in patients randomized to medical or surgical therapy in the Coronary Artery Surgery Study (CASS): 10-year follow-up. *J Am Coll Cardiol*, 20;287–294.

Carroll, R.J., Fan, J., Gijbels, I., Wand, M.P. (1997). Generalized partially linear single-index models. *J Am Stat Assoc*, 92;477–489.

Chen, K., Guo, S., Sun, L. and Wang, J. (2010). Global partial likelihood for nonparametric proportional hazards models. *J Am Stat Assoc*, 105;750–760.

Cox, D.R. (1972). Regression models and life tables (with discussion). *J Royal Stat Soc B*, 4;187–220.

Cox, D.R. and Oakes, D. (1984). *Analysis of survival data.* Chapman and Hall, London, UK.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Boostrap.* New York: Chapman & Hall.

Fan, J., Lin, H.Z., and Zhou, Y. (2006). Local partial-likelihood estimation for life time data. *Annal Stat*, 34;290–325.

Fisher, L.D. and Lin, D.Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Rev Public Health*, 20;145–157.

Gail, M.H. (1972). Does cardiac transplantation prolong life? A reassessment. *Annals Intern Med*, 76;815–817.

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

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

Heinzl, H., Kaider, A. and Zlabinger, G. (1996). Assessing interactions of binary time-dependent covariates with time in Cox proportional hazards regression models using cubic spline functions. *Stat Med*, 15;2589–2601.

Hoover, D.R., Rice, J.A., Wu, C.O. and Yang, L.P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85;809–822.

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

Martinussen, T., Scheike, T.H., and Skovgaard, I.M. (2002). Efficient estimation of fixed and time-varying covariates effects in multiplicative intensity models. Unpublished manuscript.

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

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

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

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

Tian, L., Zucker, D. and Wei, L.J. (2002). *On the Cox Model with Time-Varying Regression Coefficients.* Working paper.

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

Appendix: Notations and Conditions

Notations.

To express explicitly the asymptotic expression of the estimators $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0$ and $\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)$, we introduce necessary notation. Denote $\nu_k = \int x^k K^2(x)dx$, $\mu_k = \int x^k K(x)dx$, $\mathbf{X}_i = \{(\mathbf{x}_i(u), \mathbf{z}_i(u) : u \le \tau\}$,$P(u|\mathbf{X}_i, t_i) = \Pr(\mathcal{T}_i \ge u|\mathbf{X}_i, t_i)$,

$$\Gamma_i(u, \boldsymbol{\alpha}, \boldsymbol{\delta}) = P(u|\mathbf{X}_i, t_i) \exp\left(\mathbf{z}_i(u)'\boldsymbol{\alpha} + \mathbf{x}_i(u)'\boldsymbol{\delta}(u - t_i)\right), \ \Gamma_i(u) = \Gamma_i(u, \boldsymbol{\alpha}_0, \boldsymbol{\beta})$$

$$s_{r0}(u, \boldsymbol{\alpha}, \boldsymbol{\delta}) = E\left\{\Gamma_i(u, \boldsymbol{\alpha}, \boldsymbol{\delta})\mathbf{z}_i(u)^{\otimes r}\right\}, \ \ s_{r0}(u) = s_{r0}(u, \boldsymbol{\alpha}_0, \boldsymbol{\beta}) \ \ r = 0, 1, 2,$$

$$s_{0r}(u, \boldsymbol{\alpha}, \boldsymbol{\delta}, t) = E\left\{\Gamma_i(u, \boldsymbol{\alpha}, \boldsymbol{\delta})\mathbf{x}_i(u)^{\otimes r}|t_i = t\right\} f(t), \ \ s_{0r}(u, t) = s_{0r}(u, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, t) \ \ r = 1, 2,$$

$$s_{11}(u, \boldsymbol{\alpha}, \boldsymbol{\delta}, t) = E\left\{\Gamma_i(u, \boldsymbol{\alpha}, \boldsymbol{\delta})\mathbf{z}_i(u)\mathbf{x}_i(u)'|t_i = t\right\} f(t), s_{11}(u, t) = s_{11}(u, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, t),$$

$$\vartheta_1(t, v) = \int_0^\tau \frac{s_{01}(u, u-t)s_{01}'(u, u-v)}{s_{00}(u)}\lambda_0(u)du, \ \ \vartheta_2(t) = \int_0^\tau s_{02}(u, u-t)\lambda_0(u)du,$$

$$\Xi_0 = \int_0^\tau \left(\frac{s_{10}(u)s_{10}'(u)}{s_{00}(u)} - s_{20}(u)\right)\lambda_0(u)du,$$

$$\Xi_1(t) = \int_0^\tau \left(\frac{s_{10}(u)s_{01}(u, u-t)'}{s_{00}(u)} - s_{11}(u, u-t)\right)\lambda_0(u)du.$$

Let $G(t)$ satisfy the following integral equation:

$$\Xi_1(t) = -G(t)\vartheta_2(t) + \int_0^\tau G(w)\vartheta_1(w, t)dw,$$

$s_{G(rs)}(u) = E\left\{(G(u - t_i)\mathbf{x}_i(u))^{\otimes s}\mathbf{z}_i(u)^{\otimes r}\Gamma_i(u)\right\}$. Denote

$$\begin{aligned}
\mathbf{A} &= \Xi_0 - \int_0^\tau G(t)\Xi_1(t)'dt, \ \text{and} \\
\mathbf{B} &= \int_0^\tau \left[s_{G(02)}(u) - s_{G(11)}(u) - s_{G(11)}(u)' + s_{20}(u)\right]\lambda_0(u)du \\
&\quad - \int_0^\tau \left[s_{G(01)}(u) - s_{10}(u)\right]\left[\frac{s_{G(01)}(u) - s_{10}(u)}{s_{00}(u)}\right]'\lambda_0(u)du.
\end{aligned}$$

Denote $\boldsymbol{\Theta}$ to be the bounded support of $\boldsymbol{\alpha}$, and

$$\mathcal{C}_0 = \{\boldsymbol{\delta}(t) : t \in [0, \tau], \|\boldsymbol{\delta}(t+h) - \boldsymbol{\delta}(t)\| = O(h)\ \}.$$

Conditions:

1. The kernel function $K(\cdot)$ is a symmetric density function with a compact support $[-1, 1]$ and bounded derivative.
2. $t_i, i = 1, \cdots, n$ are independent random variables of the density function $f(\cdot)$, which is positive and has a continuous second derivative on $[0, \tau]$.
3. $\mathbf{x}_i(t)$ is bounded with compact support. $P(C_i = 0|\mathbf{X}_i) < 1$.
4. The functions $\boldsymbol{\beta}(\cdot)$ have a continuous second derivative on the corresponding compact support, $\boldsymbol{\alpha} \in \boldsymbol{\Theta}$.
5. The conditional probability $P(u|\mathbf{X}_i = \mathbf{x}, t_i = t)$ is positive and has a continuous second derivative on $[0, \tau]$ for each $\mathbf{x}$ and t over the corresponding compact support.
6. Denote

$$u_1(\boldsymbol{\alpha}, \boldsymbol{\delta}) = \int_0^\tau \left\{ s_{10}(u) - \frac{s_{10}(u, \boldsymbol{\alpha}, \boldsymbol{\delta})}{s_{00}(u, \boldsymbol{\alpha}, \boldsymbol{\delta})} s_{00}(u) \right\} \lambda_0(u) du,$$
$$u_2(\boldsymbol{\alpha}, \boldsymbol{\delta}; t) = \int_0^\tau \left[s_{01}(u, u - t) - \frac{s_{01}(u, \boldsymbol{\alpha}, \boldsymbol{\delta}, u - t)}{s_{00}(u, \boldsymbol{\alpha}, \boldsymbol{\delta})} s_{00}(u) \right] \lambda_0(u) du.$$

   Then, there exists a unique root to $u(\boldsymbol{\alpha}, \boldsymbol{\delta}; t) \equiv (u_1(\boldsymbol{\alpha}, \boldsymbol{\delta})', u_2(\boldsymbol{\alpha}, \boldsymbol{\delta}; t)')' = 0$ in $\boldsymbol{\Theta} \otimes \mathcal{C}_0$.

7. $s_{r0}(u)$, $s_{0r}(u, t)$ and $s_{11}(u, t)$ have a continuous second derivative on $(u, t) \in [0, \tau] \times [0, \tau]$.
8. $h(\log n)^2 \to 0$, $(nh)/(logn)^2 \to \infty$ and $nh^3 \to \infty$.

Table 1. Simulation results of the parameters for Simulation 2.

		Proposed	Ideal
α_1	Bias	0.0243	0.0224
	SD	0.1199	0.1194
	RMSE	0.1223	0.1214
α_2	Bias	0.0191	0.0176
	SD	0.2551	0.2527
	RMSE	0.2558	0.2533

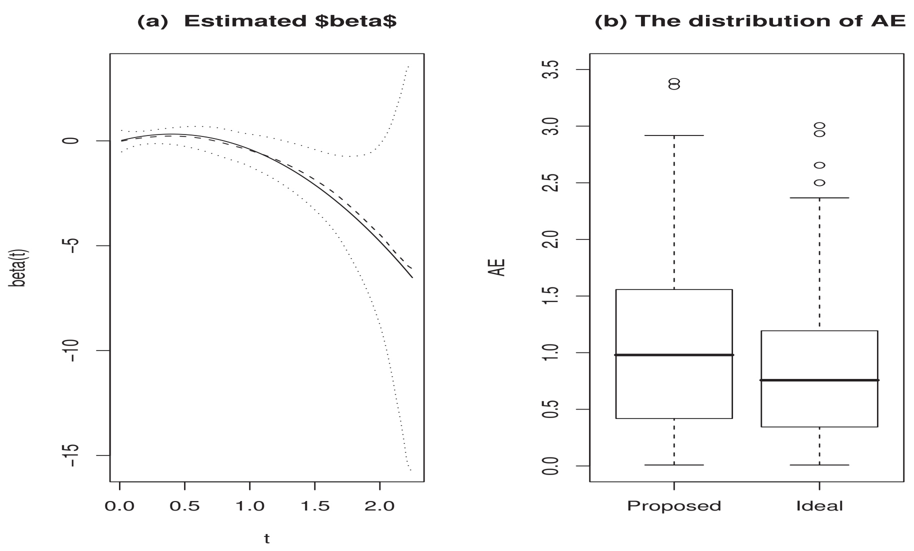

Figure 1: (a) The averaged estimates of $\beta(t)$ for simulation 1 (Solid— : true functions; dashed— : estimated; dotted— : 95% confidence limit). (b) The distribution of AE for the 200 replications in Simulation 1.

Table 2. Descriptive statistics patients in the 2008-2011 SRTR data (n=3115).

Variable	Count(%)	Variable	Mean(± SD)
Transplantation	1446(46.4%)	Waiting Time	341.7(±380.3)
Death	329(10.6%)	BMI	29.49(±5.86)
Female	1194(38.3%)	MACIT	34.09(±5.09)
PM	233(7.5%)		
Diabetes	1398(44.9%)		

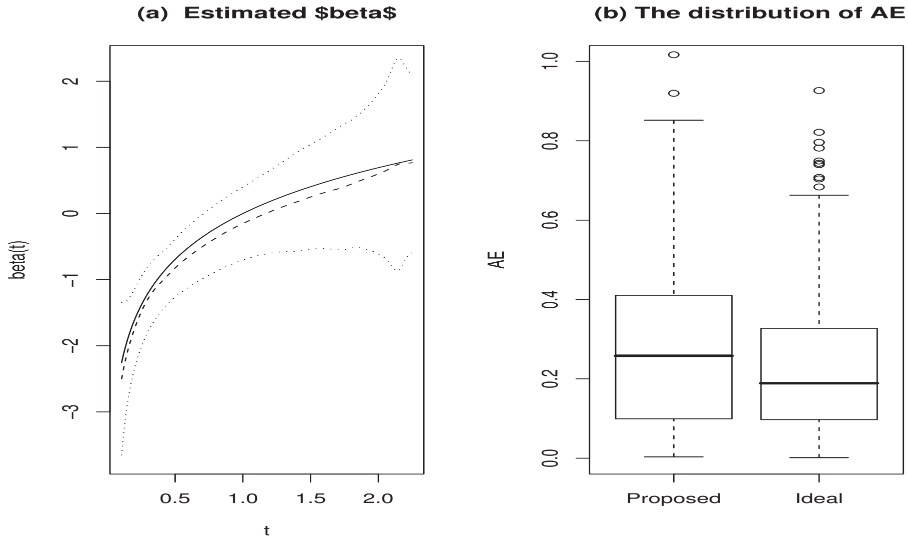

Figure 2: Results of Simulation 2. (a) The averaged estimates of $\beta_1(t)$; (b) The averaged estimates of $\beta_2(t)$ (Solid— : true functions; dashed— : estimated; dotted— : confidence limit).

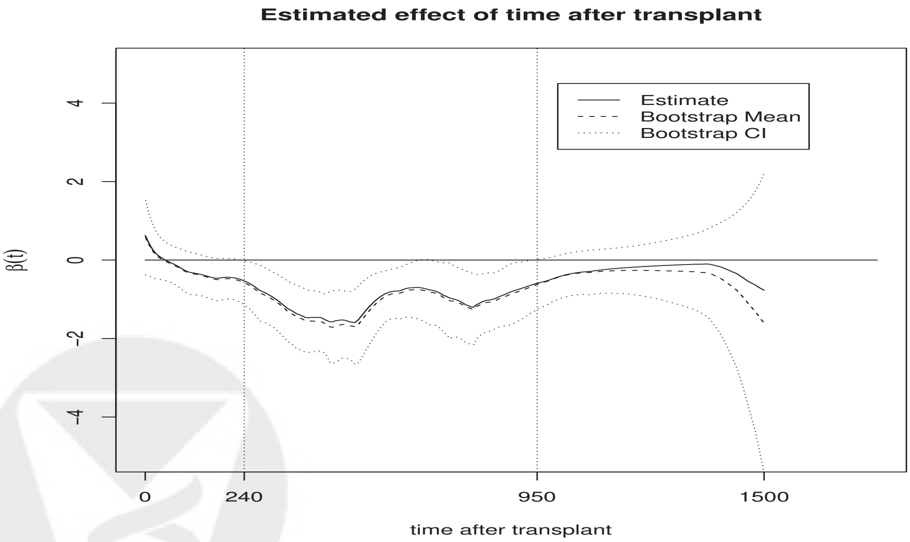

Figure 3: Estimated effect of time on survival after transplant.

Table 3. Estimated parameters from the proposed and Cox survival models.

	Cox			Proposed		
Variable	Est.	SD	$p-$value	Est.	SD	$p-$value
Transplantation	-1.24	0.13	0	-	-	-
BMI	-.041	0.010	4.13e-05	-.038	0.010	1.44e-04
Female	-.026	0.11	0.813	-.034	0.12	0.777
PM	.59	0.17	5.19e-04	.55	0.15	2.46e-04
MACIT	.012	0.010	0.230	.014	0.0092	0.128
Diabetes	.63	0.12	1.52e-07	0.64	0.12	9.64e-08

A global partial likelihood estimator of the time-varying effects for time-dependent treatment (Supplementary Material: Proofs of Theorems 1 to 4)

Huazhen Lin

Center of Statistical Research, School of Statistics,

Southwestern University of Finance and Economics,

Chengdu, Sichuan, China,

linhz@swufe.edu.cn

Zhe Fei and Yi Li

Department of Biostatistics, University of Michigan, USA

Proof of Theorem 1.

For any $\alpha \in \mathbf{\Theta}$ and any vector functions $\boldsymbol{\delta}(\cdot)$ and $\boldsymbol{\eta}(\cdot)$, set $\boldsymbol{\xi}(t) = (\boldsymbol{\delta}(t)', h\boldsymbol{\eta}(t)')'$,

$$U_1(\boldsymbol{\alpha}, \boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{z}_i(u) - \frac{S_{n1,1}(u,\boldsymbol{\alpha},\boldsymbol{\delta})}{S_{n0}(u,\boldsymbol{\alpha},\boldsymbol{\delta})}\right\}dN_i(u),$$

$$U_2(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\eta}; t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left[\mathbf{W}_i(u, u-t)K_i(u-t) - \frac{S_{n1,2}(u,\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t)}{S_{n0}(u,\boldsymbol{\alpha},\boldsymbol{\delta})}\right]dN_i(u), \tag{S.1}$$

where $\mathbf{W}_i(u,v) = \varpi_i(v)\otimes \mathbf{x}_i(u)$, $\varpi_i(v) = (1, (v-t_i)/h)'$, $K_i(u) = K_h(u-t_i)$,

$$S_{n0}(u,\boldsymbol{\alpha},\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n} Y_i(u)\exp\left\{\mathbf{z}_i(u)'\boldsymbol{\alpha} + \mathbf{x}_i(u)'\boldsymbol{\delta}(u-t_i)\right\},$$

$$S_{n1,1}(u,\boldsymbol{\alpha},\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n} Y_i(u)\mathbf{z}_i(u)\exp\left[\boldsymbol{\alpha}'\mathbf{z}_i(u) + \mathbf{x}_i(u)'\boldsymbol{\delta}(u-t_i)\right],$$

$$S_{n1,2}(u,\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(u)\exp\left[\mathbf{z}_i(u)'\boldsymbol{\alpha} + \mathbf{W}_i(u,u-t)'\boldsymbol{\xi}(t)\right]\mathbf{W}_i(u,u-t)K_i(u-t).$$

Using kernel theory, we have

$$\begin{aligned} S_{n0}(u,\boldsymbol{\alpha},\boldsymbol{\delta}) &\to s_{00}(u,\boldsymbol{\alpha},\boldsymbol{\delta}), \\ S_{n1,1}(u,\boldsymbol{\alpha},\boldsymbol{\delta}) &\to s_{10}(u,\boldsymbol{\alpha},\boldsymbol{\delta}) and \\ S_{n1,2}(u,\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t) &\to s_{01}(u,\boldsymbol{\alpha},\boldsymbol{\delta},u-t)(1,0)'. \end{aligned}$$

Hence, under model (1.2) and the regular conditions, we have

$$\begin{aligned} U_1(\boldsymbol{\alpha},\boldsymbol{\delta}) &= u_1(\boldsymbol{\alpha},\boldsymbol{\delta}) + o_p(1) and \\ U_2(\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t) &= u_2(\boldsymbol{\alpha},\boldsymbol{\delta};t)(1,0)' + o_p(1). \end{aligned} \tag{S.2}$$

Obviously $u_1(\boldsymbol{\alpha}_0,\boldsymbol{\beta}) = 0$ and $u_2(\boldsymbol{\alpha}_0,\boldsymbol{\beta};t) = 0$, by using Condition 6, $(\boldsymbol{\alpha}_0,\boldsymbol{\beta})$ is the unique root to the equation $u(\boldsymbol{\alpha},\boldsymbol{\delta};t) = 0$ in $\mathbf{\Theta}\otimes\mathcal{C}_0$.

Denote $U(\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t) = (U_1(\boldsymbol{\alpha},\boldsymbol{\delta})', U_2(\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t)')'$ and define

$$\mathcal{B}_n = \left\{\boldsymbol{\delta} : \|\boldsymbol{\delta}\| \le D, \|\boldsymbol{\delta}(t_1) - \boldsymbol{\delta}(t_2)\| \le d[|t_1 - t_2| + b_n], \text{for } t_1, t_2 \in [0,\tau]\right\},$$

for some constants $D > 0$ and $d > 0$, where $b_n = h + (nh)^{-1/2}(\log n)^{1/2}$.

To show the consistency of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$, it suffices to prove the following (i)–(iii):

(i) For each vector $\boldsymbol{\alpha} \in \boldsymbol{\Theta}$, each function vector $\boldsymbol{\delta} \in \mathcal{C}_0$ and any bounded function vector $\boldsymbol{\eta}$,

$$\sup_{0\le t\le\tau} \|U(\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t) - u(\boldsymbol{\alpha},\boldsymbol{\delta};t)\| = o_p(1).$$

(ii) $\sup_{0\le t\le\tau,\boldsymbol{\alpha}\in\boldsymbol{\Theta},\boldsymbol{\delta}\in\mathcal{B}_n,\boldsymbol{\eta}\in\mathcal{R}} \|U(\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t) - u(\boldsymbol{\alpha},\boldsymbol{\delta};t)\| = o_p(1)$, where $\mathcal{R}$ is the set of functions on $[0,\tau]$, which are bounded uniformly.

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

Once (i)–(iii) are established, using an idea similar to the Arzela-Ascoli theorem and (iii), we can show that for any subsequence of $\{\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}\}$, there exists a further convergent subsequence $\{\widehat{\boldsymbol{\alpha}}_n, \widehat{\boldsymbol{\beta}}_n\}$ such that uniformly in $t \in [0,\tau]$, $\widehat{\boldsymbol{\alpha}}_n \to \widehat{\boldsymbol{\alpha}}^*$ and $\widehat{\boldsymbol{\beta}}_n(t) \to \boldsymbol{\beta}^*(t)$ in probability. It is easily seen that $\boldsymbol{\beta}^*(\cdot) \in \mathcal{C}_0$. Note that

$$u(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*;t) = U(\widehat{\boldsymbol{\alpha}}_n,\widehat{\boldsymbol{\beta}}_n,\widehat{\dot{\boldsymbol{\beta}}}_n;t) - \left[U(\widehat{\boldsymbol{\alpha}}_n,\widehat{\boldsymbol{\beta}}_n,\widehat{\dot{\boldsymbol{\beta}}}_n;t) - u(\widehat{\boldsymbol{\alpha}}_n,\widehat{\boldsymbol{\beta}}_n;t)\right] - \left[u(\widehat{\boldsymbol{\alpha}}_n,\widehat{\boldsymbol{\beta}}_n;t) - u(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*;t)\right],$$

and $U(\widehat{\boldsymbol{\alpha}}_n,\widehat{\boldsymbol{\beta}}_n,\widehat{\dot{\boldsymbol{\beta}}}_n;t) = 0$. It follows from (ii) that $u(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*;t) = 0$. Since $u(\boldsymbol{\alpha},\boldsymbol{\delta};t) = 0$ has a unique root $(\boldsymbol{\alpha}_0,\boldsymbol{\beta})$, we have $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}$, which yields the consistency of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$.

Proof of (i). (i) follows from (S.2).

Proof of (ii). Noting that $\mathbf{x}_i$ is bounded, the arguments used to prove (ii) are essentially the same as those in Chen, *et al* (2010).

Proof of (iii). Denote $S_{n121}(u,\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t)$ and $U_{21}(\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t)$ to be the first p-elements of $S_{n12}(u,\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t)$ and $U_2(\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\eta};t)$, respectively. $\widehat{S}_{n121}(u,t) = S_{n121}(u,\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\dot{\boldsymbol{\beta}}};t)$. Given any $t_1 \in [0,\tau]$ and $t_2 \in [0,\tau]$ that $t_1 - t_2 = o_p(1)$, since $U_{21}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\dot{\boldsymbol{\beta}}};t_1) = 0$ and $U_{21}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\dot{\boldsymbol{\beta}}};t_2) = 0$, we have

$$\begin{aligned} 0 &= U_{21}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\dot{\boldsymbol{\beta}}};t_1) - U_{21}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\dot{\boldsymbol{\beta}}};t_2) \\ &= v_2(t_1) - v_2(t_2) + O_p(b_n) - \int_0^\tau \frac{\widehat{S}_{n121}(u,t_1) - \widehat{S}_{n121}(u,t_2)}{S_{n0}(u,\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}})} d\overline{N}(u), \end{aligned} \tag{S.3}$$

where $\overline{N}(u) = \frac{1}{n}\sum_{i=1}^n N_i(u)$ and $v_2(t) = \int_0^\tau s_{01}(u,u-t)\lambda_0(u)du$. Using the empirical process theory and kernel theory, it can be shown that (e.g., Fan, Lin and Zhou, 2006),

$$\begin{aligned} &\widehat{S}_{n121}(u,t_1) - \widehat{S}_{n121}(u,t_2) \\ &\quad = s_{01}^{(01)}(u,\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},u-t_1)(t_1-t_2) + s_{02}(u,\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},u-t_2)\left\{\widehat{\boldsymbol{\beta}}(t_1) - \widehat{\boldsymbol{\beta}}(t_2)\right\} \\ &\qquad - s_{02}(u,\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},u-t_1)\widehat{\dot{\boldsymbol{\beta}}}(t_1)(t_1-t_2) + O_p(b_n) + o_p(t_1-t_2), \end{aligned}$$

uniformly over $t_1 \in [0,\tau]$ and $t_2 \in [0,\tau]$, where $s_{01}^{(01)}(u,\boldsymbol{\alpha},\boldsymbol{\delta},v) = ds_{01}(u,\boldsymbol{\alpha},\boldsymbol{\delta},v)/dv$. This, coupled with (S.3), proves (iii).

Proof of Theorem 2.

Denote $a_n = |\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0|$, $c_n = \sup_{t\in[0,\tau]} |\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)|$, $d_n = \sup_{t\in[0,\tau]} |h\widehat{\dot{\boldsymbol{\beta}}}(t) - h\dot{\boldsymbol{\beta}}(t)|$. The proof consists of the following four steps.

Step 1. Obtain the asymptotic expression (S.7).

Denote:

$$\begin{aligned} U_1(\widehat{\boldsymbol{\alpha}};\widehat{\boldsymbol{\beta}}) - U_1(\boldsymbol{\alpha}_0;\boldsymbol{\beta}) \;=\;& \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\Big(S_{n0}(u,\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}) - S_{n0}(u,\boldsymbol{\alpha}_0,\boldsymbol{\beta})\Big)\frac{S_{n1,1}(u,\boldsymbol{\alpha}_0,\boldsymbol{\beta})}{S_{n0}^2(u,\boldsymbol{\alpha}_0,\boldsymbol{\beta})}dN_i(u) \\ & -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\Big(S_{n1,1}(u,\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}) - S_{n1,1}(u,\boldsymbol{\alpha}_0,\boldsymbol{\beta})\Big)\frac{1}{S_{n0}(u,\boldsymbol{\alpha}_0,\boldsymbol{\beta})}dN_i(u) \\ & + O_p(a_n^2 + c_n^2). \end{aligned}$$

Using Taylor expansion and the consistency of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$, we have

$$\begin{aligned} S_{n1,1}(u,\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}) - S_{n1,1}(u,\boldsymbol{\alpha}_0,\boldsymbol{\beta}) &= s_{20}(u)(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \\ &\quad + \int_0^{\tau} s_{11}(u,u-t)\Big(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\Big)dt + O_p(a_n^2 + c_n^2), \\ S_{n0}(u,\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}) - S_{n0}(u,\boldsymbol{\alpha}_0,\boldsymbol{\beta}) &= s_{10}'(u)(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \\ &\quad + \int_0^{\tau} s_{01}'(u,u-t)\Big(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\Big)dt + O_p(a_n^2 + c_n^2). \end{aligned} \tag{S.4}$$

Hence,

$$U_1(\widehat{\boldsymbol{\alpha}};\widehat{\boldsymbol{\beta}}) - U_1(\boldsymbol{\alpha}_0;\boldsymbol{\beta}) = \Xi_0(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + \int_0^{\tau}\Xi_1(t)\Big(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\Big)dt + O_p(a_n^2 + c_n^2). \tag{S.5}$$

Because $U_1(\widehat{\boldsymbol{\alpha}};\widehat{\boldsymbol{\beta}}) = 0$ and

$$U_1(\boldsymbol{\alpha}_0;\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{z}_i(u) - \frac{S_{n1,1}(u,\boldsymbol{\alpha}_0,\boldsymbol{\beta})}{S_{n0}(u,\boldsymbol{\alpha}_0,\boldsymbol{\beta})}\right\}dM_i(u), \tag{S.6}$$

where $M_i(u) = N_i(u) - \int_0^u Y_i(s)\exp\{\mathbf{z}_i(s)'\boldsymbol{\alpha}_0 + \mathbf{x}_i(s)'\boldsymbol{\beta}(s-t_i)\}\lambda_0(s)ds$, we see

$$\begin{aligned} \Xi_0(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) = & -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{z}_i(u) - \frac{s_{10}(u)}{s_{00}(u)}\right\}dM_i(u) \\ & - \int_0^{\tau}\Xi_1(t)\Big(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\Big)dt + o_p(n^{-1/2}) + O_p(a_n^2 + c_n^2). \end{aligned} \tag{S.7}$$

Step 2. Obtain the expression (S.10) of $\int_0^{\tau}\Xi_1(t)\Big(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\Big)dt$.

Denote $S_{n121}(u,t) = S_{n121}(u, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, t)$, $S_{n0}(u) = S_{n0}(u, \boldsymbol{\alpha}_0, \boldsymbol{\beta})$, $\widehat{S}_{n121}(u,t) = S_{n121}(u, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, t)$ and $\widehat{S}_{n0}(u) = S_{n0}(u, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$. Since

$$
\begin{aligned}
& U_{21}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\dot{\boldsymbol{\beta}}}; t) - U_{21}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \dot{\boldsymbol{\beta}}; t) \\
&= -\int_0^\tau \left[\widehat{S}_{n121}(u,t) - S_{n121}(u,t)\right] \frac{d\overline{N}(u)}{\widehat{S}_{n0}(u)} + \int_0^\tau \frac{\left[\widehat{S}_{n0}(u) - S_{n0}(u)\right] S_{n121}(u,t)}{S_{n0}(u)\widehat{S}_{n0}(u)} d\overline{N}(u).
\end{aligned} \tag{S.8}
$$

Let $\zeta_n(t) = (\widehat{\boldsymbol{\beta}}(t)^T - \boldsymbol{\beta}(t)^T, h(\widehat{\dot{\boldsymbol{\beta}}}(t) - \dot{\boldsymbol{\beta}}(t))^T)^T$. Using Taylor expansion, consistency of $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$ and the boundiness of $\widehat{\dot{\boldsymbol{\beta}}}$, we have

$$
\begin{aligned}
& \widehat{S}_{n121}(u,t) - S_{n121}(u,t) \\
&\quad = \frac{1}{n}\sum_{i=1}^n Y_i(u) \exp\left[\mathbf{z}_i(u)'\boldsymbol{\alpha}_0 + \mathbf{x}_i(u)'\boldsymbol{\beta}(u-t_i)\right] \mathbf{x}_i(u)\mathbf{z}_i(u)' K_i(u-t)\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) \\
&\qquad + \frac{1}{n}\sum_{i=1}^n Y_i(u) \exp\left[\mathbf{z}_i(u)'\boldsymbol{\alpha}_0 + \mathbf{x}_i(u)'\boldsymbol{\beta}(u-t_i)\right] \mathbf{x}_i(u)\mathbf{W}_i(u,u-t)' K_i(u-t)\zeta_n(t) \\
&\qquad + O_p(h^2(a_n + c_n + d_n) + a_n^2 + c_n^2 + d_n^2) \\
&\quad = s_{11}(u, u-t)'\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) + s_{02}(u, u-t)\left\{\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\right\} \\
&\qquad\qquad + O_p(b_n(a_n + c_n + d_n) + a_n^2 + c_n^2 + d_n^2).
\end{aligned}
$$

Substituting this and (S.4) into (S.8), we get

$$
\begin{aligned}
U_{21}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\dot{\boldsymbol{\beta}}}; t) - U_{21}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \dot{\boldsymbol{\beta}}; t) &= \Xi_1(t)'\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) \\
& - \vartheta_2(t)(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)) + \int_0^\tau \vartheta_1(t,v)(\widehat{\boldsymbol{\beta}}(v) - \boldsymbol{\beta}(v))dv.
\end{aligned} \tag{S.9}
$$

Let $G(t)$ satisfy the following integral equation in $\mathcal{C}_0$:

$$
\Xi_1(t) = -G(t)\vartheta_2(t) + \int_0^\tau G(w)\vartheta_1(w,t)dw.
$$

Integrating both sides of Equation (S.9), then multiplying by $G(t)$, we get

$$
\begin{aligned}
& \int_0^\tau \Xi_1(t)\left(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\right) dt \\
&= \int_0^\tau G(t)\left(U_{21}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\dot{\boldsymbol{\beta}}}; t) - U_{21}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \dot{\boldsymbol{\beta}}; t)\right) dt - \int_0^\tau G(t)\Xi_1(t)' dt\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) \\
&= -\int_0^\tau G(t) U_{21}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \dot{\boldsymbol{\beta}}; t)dt - \int_0^\tau G(t)\Xi_1(t)' dt\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right).
\end{aligned} \tag{S.10}
$$

Step 3. Obtain the expansion $\int_0^\tau G(t)U_{21}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \dot{\boldsymbol{\beta}}; t)dt$.

Since

$$
U_{21}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \dot{\boldsymbol{\beta}}; t) = A_n(t) + V_n(t), \tag{S.11}
$$

where

$$A_n(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left[\mathbf{x}_i(u)K_i(u-t) - \frac{S_{n121}(u,t)}{S_{n0}(u)}\right] dM_i(u),$$

$$V_n(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left[\mathbf{x}_i(u)K_i(u-t) - \frac{S_{n121}(u,t)}{S_{n0}(u)}\right] \times Y_i(u)\exp\left\{\mathbf{z}_i(u)'\boldsymbol{\alpha}_0 + \mathbf{x}_i(u)'\boldsymbol{\beta}(u-t_i)\right\}\lambda_0(u)du.$$

Note that

$$V_n(t) = \int_0^\tau \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i(u)K_i(u-t)Y_i(u)\exp\left\{\mathbf{z}_i(u)'\boldsymbol{\alpha}_0 + \mathbf{x}_i(u)'\boldsymbol{\beta}(u-t_i)\right\} - S_{n121}(u,t)\right]\lambda_0(u)du.$$

It can be shown that (e.g., Fan *et al.*, 2006),

$$V_n(t) = \frac{\mu_2 h^2}{2}\vartheta_2(t)\ddot{\boldsymbol{\beta}}(t) + o_p(h^2). \tag{S.12}$$

The martingale central limit theorem implies that

$$(nh)^{1/2}A_n(t) = n^{-1/2}h^{1/2}\sum_{i=1}^{n}\int_0^\tau \left[\mathbf{x}_i(u)K_i(u-t) - \frac{s_{01}(u,u-t)}{s_{00}(u)}\right] dM_i(u) + o_p(1), \tag{S.13}$$

is asymptotically normal with mean zero and covariance matrix

$$\begin{aligned}\Sigma(t) &= h\int_0^\tau E\left[\mathbf{x}_i(u)K_i(u-t) - \frac{s_{01}(u,u-t)}{s_{00}(u)}\right]^{\otimes 2}\Gamma_i(u)\lambda_0(u)du \\ &= \int_0^\tau \left[s_{02}(u,u-t)\nu_0 - \frac{s_{01}(u,u-t)s_{01}'(u,u-t)}{s_{00}(u)}\right]\lambda_0(u)du \\ &= \nu_0\vartheta_2(t) - \vartheta_1(t,t).\end{aligned}$$

Hence, by (S.11), (S.12) and (S.13), we get

$$U_{21}(\boldsymbol{\alpha}_0,\boldsymbol{\beta},\dot{\boldsymbol{\beta}};t) = n^{-1}\sum_{i=1}^{n}\int_0^\tau \left[\mathbf{x}_i(u)K_i(u-t) - \frac{s_{01}(u,u-t)}{s_{00}(u)}\right] dM_i(u) + \frac{\mu_2 h^2}{2}\vartheta_2(t)\ddot{\boldsymbol{\beta}}(t) + o_p(h^2 + (nh)^{-1/2}). \tag{S.14}$$

Then, we can write

$$\int_0^\tau G(t)U_{21}(\boldsymbol{\alpha}_0,\boldsymbol{\beta},\dot{\boldsymbol{\beta}};t)dt = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left[G(u-t_i)\mathbf{x}_i(u) - \frac{S_{nG}(u)}{S_{n0}(u,\boldsymbol{\alpha}_0,\boldsymbol{\beta})}\right] dM_i(u) + O(h^2), \tag{S.15}$$

where

$$S_{nG}(u) = \frac{1}{n}\sum_{i=1}^{n}Y_i(u)\exp\left(\mathbf{z}_i(u)'\boldsymbol{\alpha}_0 + \mathbf{x}_i(u)'\boldsymbol{\beta}(u-t_i)\right)G(u-t_i)\mathbf{x}_i(u) \to s_{G(01)}(u).$$

Step 4. Obtain the asymptotic expression $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0$.

By (S.10) and (S.15), using $U_{21}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\dot{\boldsymbol{\beta}}}; t) = 0$, we get

$$\int_0^\tau \Xi_1(t)\left(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\right) dt = -\int_0^\tau G(t)\Xi_1(t)' dt\,(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$$
$$-\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[G(u-t_i)\mathbf{x}_i(u) - \frac{s_{G(01)}(u)}{s_{00}(u)}\right] dM_i(u) + O(h^2). \qquad \text{(S.16)}$$

Denote $\mathbf{A} = \Xi_0 - \int_0^\tau G(t)\Xi_1(t)' dt$. By (S.7) and (S.16), we obtain

$$n^{1/2}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \approx \frac{\mathbf{A}^{-1}}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \Big[G(u-t_i)\mathbf{x}_i(u) - \mathbf{z}_i(u)$$
$$-\frac{s_{G(01)}(u) - s_{10}(u)}{s_{00}(u)}\Big] dM_i(u) + O(h^2).$$

Theorem 2 follows from the martingale central limit theorem and the conditions on h.

Proof of Theorem 3.

By (S.9), $U_{21}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\dot{\boldsymbol{\beta}}}; t) = 0$ and (S.14) we obtain,

$$\vartheta_2(t)\left(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\right) = \Xi_1(t)'(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + A_n(t)$$
$$+\int_0^\tau \vartheta_1(t,v)(\widehat{\boldsymbol{\beta}}(v) - \boldsymbol{\beta}(v))dv + \frac{\mu_2 h^2}{2}\vartheta_2(t)\ddot{\boldsymbol{\beta}}(t) + o_p(h^2). \qquad \text{(S.17)}$$

where

$$A_n(t) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[\mathbf{x}_i(u)K_i(u-t) - \frac{s_{01}(u,t)}{s_{00}(u)}\right] dM_i(u).$$

Thus, Theorem 3 follows from the martingale central limit theorem and Theorem 2.

Proof of Theorem 4.

Let $g(t) = (g_1', g_2(t)')'$ satisfy the following integral equation in $\mathcal{C}_0$:

$$\phi_1 = \tau\Xi_0' g_1 + \int_0^\tau \Xi_1(t) g_2(t) dt, \qquad \text{(S.18)}$$
$$\phi_2(t) = \tau\Xi_1'(t) g_1 - \vartheta_2(t)' g_2(t) + \int_0^\tau \vartheta_1(u,t)' g_2(u) du. \qquad \text{(S.19)}$$

Denote $\Omega(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\eta}; t) = (U_1(\boldsymbol{\alpha}, \boldsymbol{\delta}), U_{21}(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\eta}; t))$. By (S.5) and (S.9), we get

$$\int_t g(t)'\left(\Omega(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\dot{\boldsymbol{\beta}}}; t) - \Omega(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \dot{\boldsymbol{\beta}}; t)\right) dt$$
$$= \left(\tau\Xi_0' g_1 + \int_0^\tau \Xi_1(t) g_2(t) dt\right)' (\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$$
$$+\int_0^\tau \left(\tau\Xi_1'(t) g_1 - \vartheta_2(t)' g_2(t) + \int_0^\tau \vartheta_1(u,t)' g_2(u) du\right)' \left(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\right) dt$$
$$\equiv \phi_1'(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + \int_0^\tau \phi_2(t)'\left(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\right) dt. \qquad \text{(S.20)}$$

Furthermore, by (S.6) and (S.15), we obtain

$$
\begin{aligned}
&\int_t g(t)'\Omega(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \dot{\boldsymbol{\beta}}; t)dt \\
&\quad = \frac{\tau g_1'}{n} \sum_{i=1}^{n} \int_0^{\tau} \left\{ \mathbf{z}_i(u) - \frac{S_{n1,1}(u, \boldsymbol{\alpha}_0, \boldsymbol{\beta})}{S_{n0}(u, \boldsymbol{\alpha}_0, \boldsymbol{\beta})} \right\} dM_i(u) \\
&\quad\quad + \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} \left[g_2'(u - t_i)\mathbf{x}_i(u) - \frac{S_{ng_2}(u)}{S_{n0}(u, \boldsymbol{\alpha}_0, \boldsymbol{\beta})} \right] dM_i(u) + O(h^2),
\end{aligned}
$$

where $S_{ng_2}(u) = \frac{1}{n}\sum_{i=1}^{n} Y_i(u) \exp\left(\mathbf{z}_i(u)'\boldsymbol{\alpha}_0 + \mathbf{x}_i(u)'\boldsymbol{\beta}(u - t_i)\right) g_2'(u - t_i)\mathbf{x}_i(u)$. By the martingale central limit theorem, we get

$$
\sqrt{n} \int_t g(t)'\Omega(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \dot{\boldsymbol{\beta}}; t)dt \to N(0, \sigma_g^2),
$$

where

$$
\begin{aligned}
\sigma_g^2 = \int_0^{\tau} &\left\{ \tau^2 g_1' \left(s_{20}(u) - \frac{s_{10}(u)s_{10}'(u)}{s_{00}(u)} \right) g_1 \right. \\
&\left. + 2\tau g_1' \left(s_{g_2(11)}(u) - \frac{s_{10}(u)s_{g_2(01)}(u)}{s_{00}(u)} \right) + s_{g_2(02)}(u) - \frac{s_{g_2(01)}^2(u)}{s_{00}(u)} \right\} \lambda_0(u)du,
\end{aligned}
$$

and $s_{g_2(k_1,k_2)}(u) = E\left\{ \Gamma_i(u)\mathbf{z}_i(u)^{\otimes k_1} \left(g_2'(u - t_i)\mathbf{x}_i(u)\right)^{k_2} \right\}$. Noting that $\Omega(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\dot{\boldsymbol{\beta}}}; t) = 0$, it follows by (S.20),

$$
\sqrt{n} \left\{ \phi_1' \left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) + \int_0^{\tau} \phi_2(t)' \left(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) \right) dt \right\} \to N\left(0, \sigma_g^2\right). \tag{S.21}
$$

We now consider a parametric submodel

$$
\boldsymbol{\beta}(t; \theta) = \boldsymbol{\beta}(t) + \theta g_2(t),
$$

where θ is an unknown parameter, and $\boldsymbol{\beta}(t)$ and $g_2(t)$ are fixed functions. Denote $\mathbf{z}_{i,g_2}(u) = \begin{pmatrix} \mathbf{z}_i(u) \\ \mathbf{x}_i(u)'g_2(u - t_i) \end{pmatrix}$. The parameters $\boldsymbol{\alpha}$ and θ can be consistently estimated by the solution $\widehat{\boldsymbol{\alpha}}_n$ and $\widehat{\theta}$ to the following Cox partial likelihood score function

$$
\begin{aligned}
n^{-1} \sum_{i=1}^{n} \int_0^{\tau} &\left[\mathbf{z}_{i,g_2}(u) \right. \\
&\left. - \frac{\sum_{j=1}^{n} Y_j(u)\mathbf{z}_{j,g_2}(u) \exp\left\{\mathbf{z}_j(u)'\boldsymbol{\alpha} + \mathbf{x}_j(u)'\left(\boldsymbol{\beta}(u - t_j) + \theta g_2(u - t_j)\right)\right\}}{\sum_{r=1}^{n} Y_r(u) \exp\left\{\mathbf{z}_r'\boldsymbol{\alpha} + \mathbf{x}_r(u)'\left(\boldsymbol{\beta}(u - t_r) + \theta g_2(u - t_r)\right)\right\}} \right] dN_i(u) = 0.
\end{aligned}
$$

Obviously, $\theta_0 = 0$ is the true value of θ. Then it follows from Anderson and Gill (1982) that

$$
\begin{pmatrix} \widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0 \\ \widehat{\theta} - \theta_0 \end{pmatrix} = \sigma^{-2} n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\mathbf{z}_{i,g_2}(u) - \frac{\begin{pmatrix} S_{n1,1}(u) \\ S_{ng_2}(u) \end{pmatrix}}{S_{n0}(u)} \right] dM_i(u). \tag{S.22}
$$

where

$$\sigma^2 = \int_0^\tau \left\{ \begin{pmatrix} s_{20}(u) - \frac{s_{10}(u)^{\otimes 2}}{s_{00}(u)} & s_{g_2(11)}(u) - \frac{s_{10}(u)s_{g_2(01)}(u)}{s_{00}(u)} \\ s_{g_2(11)}(u)' - \frac{s_{10}(u)'s_{g_2(01)}(u)}{s_{00}(u)} & s_{g_2(02)}(u) - \frac{s^2_{g_2(01)}(u)}{s_{00}(u)} \end{pmatrix} \right\} \lambda_0(u)du.$$

Note that

$$\int_0^\tau \Xi_1(t)g_2(t)dt = \int_0^\tau \left(\frac{s_{10}(u)s_{g_2(01)}(u)}{s_{00}(u)} - s_{g_2(11)}(u)\right)\lambda_0(u)du,$$

$$\int_0^\tau g_2(t)'\vartheta_2(t)'g_2(t)dt = \int_0^\tau s_{g_2(02)}(u)\lambda_0(u)du,$$

$$\int_0^\tau \int_0^\tau g_2(t)'\vartheta_1(u,t)'g_2(u)dudt = \int_0^\tau \frac{s^2_{g_2(01)}(u)}{s_{00}(u)}\lambda_0(u)du,$$

and the expressions (S.18) and (S.19), we obtain

$$\begin{pmatrix} \phi_1 \\ \int_0^\tau \phi_2'(t)g_2(t)dt \end{pmatrix} = -\sigma^2 \begin{pmatrix} \tau g_1 \\ 1 \end{pmatrix}.$$

This together with (S.22) gives

$$\sqrt{n}\left(\phi_1'\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right) + \int_0^\tau \phi_2(t)'\left(\boldsymbol{\beta}(t;\widehat{\theta}) - \boldsymbol{\beta}(t;\theta_0)\right)dt\right) \to N\left(0,\sigma_g^2\right),$$

which is the same as that of $\sqrt{n}\left\{\phi_1'\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) + \int_0^\tau \phi_2(t)'\left(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\right)dt\right\}$ by (S.21). As explained by Bickel *et al.* (1993), $\phi_1'\widehat{\boldsymbol{\alpha}} + \int_0^\tau \phi_2(t)'\widehat{\boldsymbol{\beta}}(t)dt$ is an efficient estimator of $\phi_1'\boldsymbol{\alpha}_0 + \int_0^\tau \phi_2(t)'\boldsymbol{\beta}(t)dt$. The proof of Theorem 4 is complete.