Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 43: 649–663, 2016
doi: 10.1111/sjos.12196

A Semiparametrically Efficient Estimator of the Time-Varying Effects for Survival Data with Time-Dependent Treatment

HUAZHEN LIN
Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics Chengdu

ZHE FEI AND YI LI
Department of Biostatistics, University of Michigan

ABSTRACT. The timing of a time-dependent treatment—for example, when to perform a kidney transplantation—is an important factor for evaluating treatment efficacy. A naïve comparison between the treated and untreated groups, while ignoring the timing of treatment, typically yields biased results that might favour the treated group because only patients who survive long enough will get treated. On the other hand, studying the effect of a time-dependent treatment is often complex, as it involves modelling 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 a 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 US Scientific Registry of Transplant Recipients.

Key words: cox proportional hazards model, semiparametrically 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 US Scientific Registry of Transplant Recipients (SRTR), which is collected by the United Network for Organ Sharing and Organ Procurement and Transplantation Network for all wait-listed kidney transplant candidates and transplant recipients in the United States. When a donor kidney becomes available, medical judgement 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 non-transplanted patients with those of transplanted patients will yield biased results, as only those who survive long enough to receive a kidney will receive the 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. For each patient indexed by $i = 1, \ldots, n$, we define a binary time-dependent covariate $x_i(t)$, which is equal to 1 if this patient has received the 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 would never receive the treatment in his or her lifetime, $t_i = \infty$ or $x_i(t) \equiv 0$ for $t \geq 0$. Conditional on the treatment history, 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 $t = 0$ is the time when a transplant candidate became wait-listed and $\beta(s)$ is an unknown smooth 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, in that the risk of death peaks right after treatment and then gradually decreases after the kidney transplant shows protective effects.

To allow for a multi-level treatment (e.g., different dose levels or different modalities of a treatment), we consider a vectorial form of $x_i(t)$, denoted by $\mathbf{x}_i(t)$. Also, let $\mathbf{z}_i(t)$ denote possible confounders [e.g., gender, body mass index (BMI), previous malignancy and diabetes], some of which are possibly time-dependent. We consider a general partial time-varying coefficient Cox model

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

where $\boldsymbol{\alpha}$ are unknown regression coefficients of $\mathbf{z}_i(t)$ and $\boldsymbol{\beta}(\cdot)$ measure the effects corresponding to various treatment options.

Although model (1.2) resembles the time-dependent coefficient (TDC), Cox model proposed by a number of authors, including Zucker & Karr (1990), Murphy & Sen (1991), Gamerman (1991), Murphy (1993), Marzec & Marzec (1997), Martinussen *et al.* (2002), Cai & Sun (2003), Tian *et al.* (2005) and Fan *et al.* (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. The estimation of model (1.2) is more difficult than that for the traditional TDC Cox model, because the existing nonparametric technique (e.g., kernel smoothing) is not directly applicable to the model (1.2). In the context of recurrent events, Chen *et al.* (2013) proposed a similar model and used the sieve approach to draw inference. However, their estimator involves a maximization over a functional parameter space with dimension increasing with the sample size, and the computational burden limits its usage in the analysis of large-scale data such as our motivating dataset. Further, for the sieve approach, the inference is obtained only for the cumulative function $\int_0^t \boldsymbol{\beta}(u)du$ but not for the effect function $\boldsymbol{\beta}(\cdot)$ *per se*, which is of our main interest.

The local partial likelihood, which is based on observations with survival time T_i in a small neighbourhood of a given t, has been widely used to estimate the TDC Cox model (Cai & Sun, 2003; Tian *et al.*, 2005; Fan *et al.*, 2006). However, it suffers efficiency loss, as the observations outside the neighbourhood, which may carry information about $\boldsymbol{\beta}(t)$, are not used. Instead, we propose to draw inference based on a full partial likelihood function and local smoothing technique. The main intuition 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 are presented in Section 4. Our analysis of the kidney transplantation data is given in Section 5. Technical proofs are relegated to the Appendix.

2. Estimation

2.1. Global partial likelihood approach

Consider n independent patients drawn from a population of interest. 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 of the survival distribution, we study patients' survival experience over $[0, \tau]$, where τ is such that $P(\min(T_i, C_i) > \tau) > 0$ and, in practise, is often the study duration. Assume that T_i and C_i are independent given the observed covariate process $\mathbf{X}_i = \{(\mathbf{x}_i(t), \mathbf{z}_i(t)), 0 \le t \le \mathcal{T}_i\}$, where $\mathbf{x}_i(t), \mathbf{z}_i(t)$ are p-dimensional 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 $t_i = \infty$. The observed data $\{\mathcal{T}_i, \Delta_i, \mathbf{X}_i, t_i\}$ are independent samples for $i = 1, \cdots, n$.

Under model (1.2), if $\boldsymbol{\beta}(\cdot)$ had been parameterized, it could have been 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 $p-$dimensional and $q-$dimensional vectors, respectively, and $\mathcal{R}(t) = \{i : \mathcal{T}_i \ge t\}$ 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, 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 $v > 0$, we expand each component of $\boldsymbol{\beta}(v)$ around t and obtain that

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

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 the kernel function $K(x)$ is a symmetric density with support $[-1, 1]$ and h represents the size of the local neighbourhood. 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 neighbourhood 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 neighbourhood of t when $\ell \ne i$, 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 observations. Let $\overline{\psi}_i(u) = \boldsymbol{\alpha}'\mathbf{z}_i(u) + \mathbf{x}_i(u)'[\boldsymbol{\delta} + \boldsymbol{\eta} \times (u - t_i - t)]$ 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:

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

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 as global partial likelihood estimator. Notice that the proposed method (2.5) uses observations, both within and outside the neighbourhood of t, to estimate $\boldsymbol{\beta}(t)$. As a result, our estimator is shown to be semiparametrically efficient by theorem 3 in Section 3.

In a related context, Chen *et al.* (2012) estimated non-parametric functions in a varying coefficient Cox model (Fan *et al.*, 2006) by maximizing the full partial likelihood function. However, our proposed framework largely differs from that work. First, Chen *et al.* (2012) focused on the cases where covariates interact non-linearly with other exposure variables. More specifically, the effects of a medical treatment Z vary with age W (Chen *et al.*, 2012), but the treatment Z *per se* is fixed for each individual. In contrast, our model accounts for the dynamic nature of a time-varying treatment. Directly applying Chen *et al.* (2012) to compare the treated and untreated groups, while ignoring the time-varying nature of the treatment, yields biased results that favour the treatment group. Second, it is unclear whether Chen *et al.* (2012) could be readily extended to accommodate time-dependent covariates, whereas our framework accommodates time-dependent covariates with established asymptotic properties. Indeed, as mentioned by Chen *et al.* (2012; p385), theoretical justifications for the global partial likelihood approaches are often difficult and should be made on a case-by-case basis.

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.

Step 1 of iteration m. Taking derivatives of (2.5) respect to $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$, we have for every given $t \in [0, \tau]$, solving the following equations for $\boldsymbol{\xi} = (\boldsymbol{\delta}', h\boldsymbol{\eta}')'$:

$$\frac{1}{n}\sum_{i=1}^{n} \Delta_i \left\{ \mathbf{w}_i(\mathcal{T}_i) K_h(\mathcal{T}_{ii} - t) - \left[\frac{\sum_{\ell \in \mathcal{R}(\mathcal{T}_i)} \mathbf{w}_\ell(\mathcal{T}_i) K_h(\mathcal{T}_{i\ell} - t) \exp\left(\vartheta_{i\ell}(\boldsymbol{\xi}; h)\right)}{\sum_{\ell \in \mathcal{R}(\mathcal{T}_i)} \exp\left(\vartheta_{i\ell}(\boldsymbol{\xi}; h)\right)} \right] \right\} = 0,$$

where $\vartheta_{i\ell}(\boldsymbol{\xi}; h) = hK_h(\mathcal{T}_{i\ell} - t)\overline{\psi}_\ell(\mathcal{T}_i) + \{1 - hK_h(\mathcal{T}_{i\ell} - t)\}\, \psi_\ell(\mathcal{T}_i)$. Using (2.4), we have $\vartheta_{i\ell}(\boldsymbol{\xi}; h) \approx \psi_i(\mathcal{T}_i)$. Hence, $K_h(\mathcal{T}_{i\ell} - t) \exp\left(\vartheta_{i\ell}(\boldsymbol{\xi}; h)\right) \approx K_h(\mathcal{T}_{i\ell} - t) \exp\left(\psi_i(\mathcal{T}_i)\right) \approx K_h(\mathcal{T}_{i\ell} - t) \exp\left(\overline{\psi}_\ell(\mathcal{T}_i)\right)$, and we then consider the following set of equations for $\boldsymbol{\xi}$,

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

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

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

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

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

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

and

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

Without ambiguity, we let $\hat{\boldsymbol{\xi}}(t)$ and $\hat{\boldsymbol{\alpha}}$ be the solutions of (2.8) and (2.9), respectively, and $\hat{\boldsymbol{\beta}}(t) = \hat{\boldsymbol{\delta}}(t)$, the first component of $\hat{\boldsymbol{\xi}}(t)$.

3. Large sample properties

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

Theorem 1. *Under Conditions 1–8 listed in the Appendix, we have*

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

and

$$\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\| \to 0 \; in \; probability.$$

Theorem 2. *Under Conditions 1–8 listed in the Appendix and if* $nh^4 = o(1)$*, then*

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

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 non-parametric 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 analogues in the spline literature (Carroll *et al.*, 1997; Hastie & Tibshirani, 1990). 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 and for* $0 < t < \tau$*, we have the following Fredholm integral equation,*

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

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

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

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

Theorem 2 shows that $\hat{\boldsymbol{\alpha}}$ is an $n^{1/2}-$ consistent and asymptotically normal estimator of $\boldsymbol{\alpha}$. Moreover, the following Theorem shows that $\hat{\boldsymbol{\alpha}}$ 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 $\phi_1'\hat{\boldsymbol{\alpha}} + \int_0^\tau \phi_2'(t)\hat{\boldsymbol{\beta}}(t)dt$ 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 and if* $nh^4 = o(1)$*, then* $\phi_1'\hat{\boldsymbol{\alpha}} + \int_0^\tau \phi_2'(t)\hat{\boldsymbol{\beta}}(t)dt$ *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 $\hat{\boldsymbol{\alpha}}$ is an efficient estimator of $\boldsymbol{\alpha}_0$. By taking $\phi_1(t) = 0$, then $\int_0^\tau \phi_2'(t)\hat{\boldsymbol{\beta}}(t)dt$ 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. Theorem 2 implies that the bandwidth h is not crucial for the asymptotic performance of the estimates for the parameters $\boldsymbol{\alpha}$, as confirmed in our simulation studies. Hence, our estimates are not sensitive to the bandwidth h, and a roughly estimated h would be sufficiently good for the estimation of $\boldsymbol{\alpha}$.

However, the selection of h is crucial for the asymptotic performance of $\hat{\boldsymbol{\beta}}(\cdot)$. We use the $K-$fold cross-validation procedure for bandwidth selection, which is commonly used in the literature (Efron & Tibshirani, 1993; Tian *et al.*, 2005; Fan *et al.*, 2006). Tian *et al.* (2005) and Fan *et al.* (2006) have shown empirically that the choice of the smoothing parameter

can be quite flexible. Our simulations and data application also show that the cross-validation approach works well. See Section 5 for a detailed description.

It is difficult to obtain an estimate for the covariance matrices based on theorems 2 and 3 because their expressions involve complicated unknown functions. As a remedy, we propose to use a resampling scheme, First, we generate n independent positive random variables $\zeta_i, i = 1, \cdots, n$, with mean 1 and variance 1. Fixing the data at their observed values, we estimate $\boldsymbol{\beta}(t)$ and $\boldsymbol{\alpha}$ by iteratively solving the following ζ_i−weighted estimation equations for any t:

$$\frac{1}{n}\sum_{i=1}^{n}\zeta_i\int_0^\tau\left\{\mathbf{w}_i(u)K_h(u-t_i-t)-\sum_{\ell=1}^{n}\zeta_\ell\mathbf{w}_\ell(u)K_h(u-t_\ell-t)\right.$$
$$\left.\times\frac{Y_\ell(u)\exp\left(\boldsymbol{\alpha}^{[m-1]\prime}\mathbf{z}_\ell(u)+\boldsymbol{\xi}'\mathbf{w}_\ell(u)\right)}{\sum_{r=1}^{n}\zeta_r Y_r(u)\exp\left(\boldsymbol{\alpha}^{[m-1]\prime}\mathbf{z}_r(u)+\mathbf{x}_r(u)'\boldsymbol{\beta}^{[m-1]}(u-t_r)\right)}\right\}dN_i(u)=0, \tag{3.1}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\zeta_i\int_0^\tau\left\{\mathbf{z}_i(u)-\frac{\sum_{\ell=1}^{n}\zeta_\ell Y_\ell(u)\mathbf{z}_\ell(u)\exp\left[\boldsymbol{\alpha}'\mathbf{z}_\ell(u)+\mathbf{x}_\ell(u)'\boldsymbol{\beta}^{[m]}(u-t_\ell)\right]}{\sum_{\ell=1}^{n}\zeta_\ell Y_\ell(u)\exp\left[\boldsymbol{\alpha}'\mathbf{z}_\ell(u)+\mathbf{x}_\ell(u)'\boldsymbol{\beta}^{[m]}(u-t_\ell)\right]}\right\} \tag{3.2}$$
$$\times\, dN_i(u)=0.$$

The traditional bootstrap is a resampling method with ζ_i independently generated from a binomial distribution with parameters n and $p = 1/n$. The estimates $\boldsymbol{\beta}^*(\cdot)$ and $\boldsymbol{\alpha}^*$ can be obtained using the same iterative algorithm proposed in Section 2. Following the proofs of theorems 1 to 3, we can show that $\boldsymbol{\beta}^*(\cdot)$ and $\boldsymbol{\alpha}^*$ have the same asymptotic expansions as those for $\hat{\boldsymbol{\beta}}(\cdot)$ and $\hat{\boldsymbol{\alpha}}$, respectively, except that each individual i is weighted by ζ_i. Then by Jin *et al.* (2001), we establish the validity of the proposed resampling method.

Proposition 1. *Under Conditions 1–8 stated in the Appendix, the conditional distributions of* $n^{1/2}(\boldsymbol{\alpha}^*-\hat{\boldsymbol{\alpha}})$ *and* $(nh)^{1/2}\{\boldsymbol{\beta}^*(t)-\hat{\boldsymbol{\beta}}(t)\}$, *given the observed data, converge to the asymptotic distributions of* $n^{1/2}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)$ *and* $(nh)^{1/2}\{\hat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}(t)\}$, *respectively.*

Hence, by repeatedly generating $\zeta_i, i = 1, \cdots, n$ *many times, we obtain a large number of realizations of* $\boldsymbol{\beta}^*(\cdot)$ *and* $\boldsymbol{\alpha}^*$. *The variance estimates of* $\hat{\boldsymbol{\beta}}(\cdot)$ *and* $\hat{\boldsymbol{\alpha}}$ *can be approximated by the empirical variances of* $\boldsymbol{\beta}^*(\cdot)$ *and* $\boldsymbol{\alpha}^*$, *respectively. The number of replications is determined by monitoring the stability of the standard errors. Our numerical studies hint that 250–300 would be sufficient.*

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 1000 simulations for each configuration.

Simulation 1. We first consider a non-parametric model

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

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 25–35 per cent of data is censored. We simulated 1000 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 to investigate the efficiency of the proposed estimator. Denote the linear function of $\beta(\cdot)$ by $\nu = \sum_{k=1}^{n_{grid}} \beta(w_k)/n_{grid}$, where $\{w_k, k = 1, \cdots, n_{grid}\}$ are the uniformly distributed grid points in which the function $\beta(\cdot)$ is estimated and $n_{grid} = 200$. The estimator for ν is used to evaluate the semiparametric efficiency.

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 $\hat{\nu}$ via the absolute errors (AEs), $\mathrm{AE} = |\hat{\nu} - \nu|$. Figure 1A displays the averaged estimated function for $\beta(\cdot)$ and its 95 per cent empirical point-wise confidence limits using the proposed method, which shows that the proposed estimates are close to the true functions. The empirical pointwise confidence limits are calculated by $mean(\hat{\beta}(t)) \pm 1.96 \times SD(\hat{\beta}(t))$, where $mean(\hat{\beta}(t))$ and $SD(\hat{\beta}(t))$ are the average and the standard deviation of $\hat{\beta}(t)$ over 1000 replications for given t. Figure 1B displays the distribution of AE based on the 1000 simulated datasets using the proposed method and the ideal model. The proposed method failed to converge in only five of the 1000 replications. It appears that the AE of the proposed estimator is comparable with that of the ideal estimator, confirming the semiparametric efficiency of our estimator. Finally, we compare the proposed estimator with the ideal estimator by displaying the distribution of the point-wise absolute errors $PAE = \sum_k |\hat{\beta}(w_k) - \beta(w_k)|/n_k$ in Fig. 1C. The PAE of the proposed estimator is larger than that of the ideal estimator, which is hardly surprising as the ideal estimation is made under the true coefficient function.

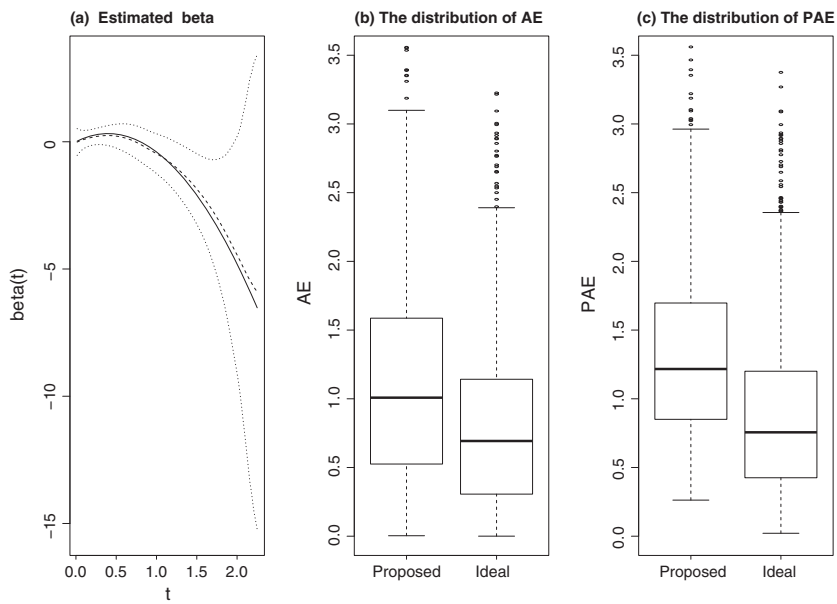

Fig. 1. (A) The averaged estimates of $\beta(t)$ for Simulation 1 (Solid lines: true functions; dashed lines: estimated; dotted lines: 95 per cent confidence limit). (B) The distribution of AE for the 1000 replications in Simulation 1. (C) The distribution of PAE for the 1000 replications in Simulation 1.

Simulation 2. Now we consider the following mixed model

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

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 $[0, 4]$, yielding a censoring proportion around 25–35 per cent.

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. The proposed method only failed to converge in one of the 1000 replications for Simulation 2. Figure 2A displays the averaged estimated function for $\beta(\cdot)$ and the 95 per cent empirical

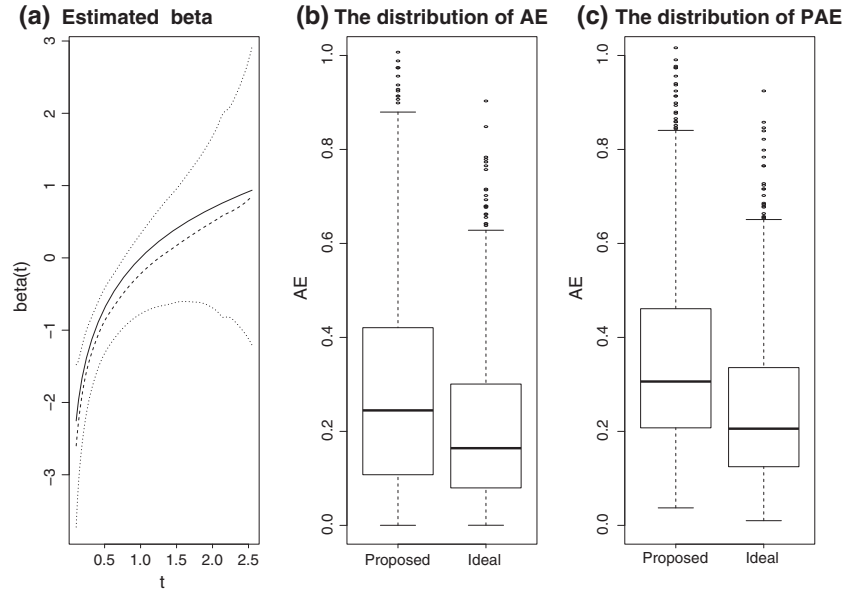

Fig. 2. Results of Simulation 2. (A) The averaged estimates of $\beta(t)$ (Solid lines: true functions; dashed lines: estimated; dotted lines: confidence limit); (B) The distribution of AE for the 1000 replications in Simulation 2. (C) The distribution of PAE for the 1000 replications in Simulation 2.

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

		Proposed	Ideal
α_1	Bias	0.0129	0.0101
	SD	0.1215	0.1205
	RMSE	0.1222	0.1209
α_2	Bias	0.0147	0.0114
	SD	0.2643	0.2618
	RMSE	0.2647	0.2620

SD, standard deviation; RMSE, root of mean squared error.

pointwise confidence limits of the proposed method, Figs 2B and C display the distributions of AE and PAE, respectively, based on the 1000 simulated datasets. Figure 2 yields similar conclusions to Fig. 1 for Simulation.

Table 1 provides the bias, empirical standard deviation (SD) and the root of mean squared error of the coefficient parameter estimators based on the 1000 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.

5. The kidney transplant programme

We study survival experience of kidney transplant patients from the U.S. Scientific Registry of Transplant Recipients , 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. As the state of Michigan has a high incident of renal failure, we focus on the renal failure patients with no history of kidney transplant who were on the waitlist between January 1, 2008 and December 31, 2011 in 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 (mean= 341.7 days; SD= 380.3 days). For each patient, the time origin is when the patient was placed on the wait list. 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}); see Table 2 The analytical goal is to investigate the effect of transplant on patient's survival and how the effect might evolve with time. Such information is much needed for post-transplantation care.

Because the treatment is time dependent, we analyse 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 patient i, $\mathbf{x}_i(t) = I(t \geq t_i)$ is the indicator for patient i having received a kidney transplant at time t. We fit the 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 non-uniformly distributed event times (number of death decreases linearly after about 3 years), we use the adaptive bandwidth (Brockmann *et al.*, 1993). We select the adaptive bandwidth for each time point so that it covered a fixed quantile, $\amalg$, of total number of events. We propose to choose $\amalg$ by a K-fold cross-validation (Cai *et al.*, 2000; Fan *et al.*, 2006), which is to minimize the prediction error

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

Table 2. *Descriptive statistics patients in the 2008-2011 scientific registry of transplant recipients 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)	Age at listing	53.16(±12.80)
PM	233 (7.5)	MACIT	34.09(±5.09)
Diabetes	1398 (44.9)		

PM, previous malignancy; BMI, body mass index; MACIT, maximum acceptable cold ischemic time.

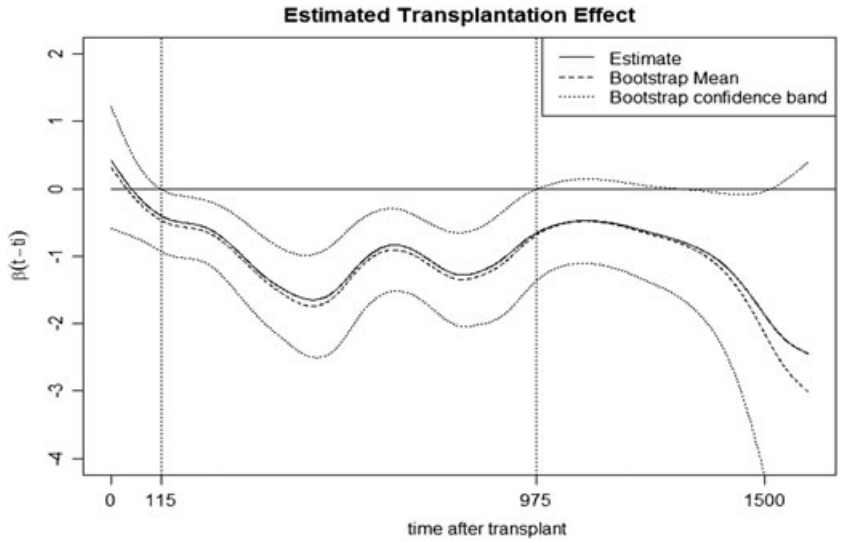

Fig. 3. Estimated effect of time on survival after transplant.

where $\hat{E}N_i(t) = \int_0^t Y_i(u)\exp\left(\mathbf{Z}_i'\hat{\boldsymbol{\alpha}} + \mathbf{x}_i(u)\hat{\beta}(u-t_i)\right) d\hat{\Lambda}_0(u)$ is the estimate of the expected failure number up to time t, $\hat{\Lambda}_0(t) = \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{dN_i(u)}{n^{-1}\sum_{j=1}^n Y_j(u)\exp\left(\mathbf{Z}_j'\hat{\boldsymbol{\alpha}}+\mathbf{x}_j(u)\hat{\beta}(u-t_j)\right)}$. For our data, we choose $K = 5$ and find that the optimal $\Pi = 0.15$. Figure 3 displays the estimated transient effect of kidney transplantation and its 95 per cent confidence limits (dotted lines) obtained from the 200 bootstrap samples. The choice of 200 is determined by monitoring the stability of the standard errors. Table 3 displays the estimated coefficients of the adjusting covariates. A lower BMI is associated with a protective effect on reducing the hazard of death, and patients without PM or diabetes experience have a better survival outcome than those with one or both of these two conditions. Gender and MACIT have no significant effects. As a comparison, we also analyse the data using the classical Cox proportional hazards model,

$$\lambda_i(t) = \lambda_0(t)\exp\{\mathbf{Z}_i'\boldsymbol{\alpha} + x_i(t)\beta\}. \tag{5.1}$$

The model yields an estimate of $\boldsymbol{\alpha}$ that is similar to those presented in Table 3.

Figure 3 shows that if the patient survives the first 115 days after transplantation, she or he would gain 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.

In practise, selection of the patients for organ transplantation is not entirely at random as it depends on a cohort of medical issues and many other logistical considerations, including patients' comorbidity conditions (e.g. diabetes, BMI). We have tried to adjust these factors as in Table 3, though the list certainly is not complete. Adjustment for unobserved confounders, using causal inference techniques, may thus be necessary. However, this may be out of the scope of this paper. Nevertheless, our results might give some useful estimates for the association between transplantation and patients' survival in an observational study setting and could give practitioners some evidence for the regulation of medical practise and the allocation of medical resources.

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

	Cox model (5.1)			Proposed		
Variable	Est.	SD	$p-$value	Est.	SD	$p-$value
Transplantation	−.914	0.13	**< 0.0001**	—	—	—
Age	.019	.0045	**< 0.0001**	.022	.0048	**< 0.0001**
BMI	−.038	.009	**< 0.0001**	−.040	.011	0.0002
Female	−.077	.10	0.46	−.024	.12	0.84
PM	.437	.16	0.005	.38	.18	0.0036
MACIT	.014	.009	0.114	.016	.010	0.11
Diabetes	.476	.11	**< 0.0001**	.55	.13	**< 0.0001**

BMI, body mass index; PM, previous malignancy; MACIT, maximum acceptable cold ischemic time.

Finally, we propose a procedure to check the validity of the assumed model (5.1), along the line of Lin *et al.* (1993) and Peng & Huang (2008). We investigate the class of stochastic processes

$$K_n(t) = \sum_{i=1}^{n} q(Z_i) M_i(t; \hat{\theta}),$$

where $q(.)$ is a known bounded function and the martingale residual

$$M_i(t; \hat{\theta}) = N_i(t) - \int_0^t Y_i(u) \exp(\mathbf{Z}_i' \hat{\alpha} + \mathbf{x}_i(u) \hat{\beta}(u - t_i)) d\hat{\Lambda}_0(u).$$

Under the null hypothesis that the proposed model is correctly specified, $K_n(t)$ would be a random Gaussian process with mean zero. Thus, a natural measure of lack-of-fit would be $\sup_t |K_n(t)|$. To calculate the p-value for the lack-of-fit test, we approximate the null distribution of $K_n(t)$ by realizations of

$$K^\star(t) = \sum_{i=1}^{n} q(Z_i) M_i(t; \hat{\theta})(1 - \zeta_i) + \sum_{i=1}^{n} q(Z_i)(M_i(t; \theta^\star) - M_i(t; \hat{\theta})),$$

where ζ_i's are weights with mean 1 and variance 1 that generated the weighted Bootstrap samples (as in Section 3) and $\theta^\star$ is the estimate from the corresponding re-weighted estimation as in (3.1) and (3.2), and the estimate

$$\Lambda_0^\star(t) = \frac{1}{n} \sum_{i=1}^{n} \zeta_i \int_0^t \frac{dN_i(u)}{n^{-1} \sum_{j=1}^{n} \zeta_j Y_j(u) \exp\left(\mathbf{Z}_j' \alpha^\star + \mathbf{x}_j(u) \beta^\star (u - t_j)\right)}$$

for the cumulative baseline hazard. We choose $q(.)$ as a quadratic function of age for our data, $((Age - mean(Age))/max(Age))^2$, so that it is smooth and bounded. Then, a p-value of 0.65 for the lack-of-fit test is calculated as the empirical proportion of $\sup_t |K^\star(t)|$ exceeding $\sup_t |K_n(t)|$ from 200 Bootstrap samples, indicating there is no evidence of lack-of-fit for our model.

6. Conclusion

To properly account for the timing of a time-dependent treatment when evaluating treatment efficacy, we propose a varying-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 confirm the

finite sample performance; we have applied the proposed method to study the efficacy of kidney transplantation among patients with end-stage renal disease, which yields some interesting results.

Moreover, there are 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 a treatment; thus, efficacy analyses should account for possible selection bias. Finally, it is worth investigating the integration of marginal structural equations or propensity matching into our framework.

Acknowledgements

Lin's research is supported by the Chinese Natural Science Foundation (nos. 11571282, 11528102) and the Fundamental Research Funds for the Central Universities. Li's research is partly supported by the Chinese Natural Science Foundation (no. 11528102). Fei and Li's research is partially supported by the U.S. National Institutes of Health. We thank the Editor, Associate Editor and two referees for their insightful suggestions, which helped significantly improve this manuscript.

Supporting information

Additional information for this article is available online including detailed proofs of theorems 1–4.

References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. & 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**, 1302–1309.

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

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

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., Lin, H. Z. & Zhou, Y. (2012). Efficient estimation for the Cox model with varying coefficients. *Biometrika* **2**, 379–392.

Chen, Q., Zeng, D., Ibrahim, J. G., Akacha, M. & Schmidli, H. (2013). Estimating time-varying effects for overdispersed recurrent events data with treatment switching. *Biometrics* **100**, 339–354.

Efron, B. & Tibshirani, R. J. (1993). *An introduction to the boostrap*, Chapman & Hall, New York.

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

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

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

Jin, Z., Ying, Z. & Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, (2), 381-390.

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

Martinussen, T., Scheike, T. H. & Skovgaard, I. M. (2002). Efficient estimation of fixed and time-varying covariates effects in multiplicative intensity models. Unpublished manuscript.
Marzec, L. & 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. & Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stoch. Proc. Appl.* **39**, 153–180.
Peng, L. & Huang, Y. (2008). Survival analysis with quantile regression models. *J. Am. Stat. Assoc.* **103**, 482.
Tian, L., Zucker, D. & Wei, L. J. (2005). On the Cox model with time-varying regression coefficients. *J. Am. Stat. Assoc.* **100**, 172–183.
Zucker, D. M. & Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Stat.* **18**, 329–353.

Received June 2014, in final form October 2015

Huazhen Lin, Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, China.
E-mail: linhz@swufe.edu.cn

Appendix A: Notations and conditions

Notations.

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

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

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

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

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

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

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

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

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

$$\mathbf{A} = \Xi_0 - \int_0^\tau G(t)\Xi_1(t)' dt, \quad \text{and}$$

$$\mathbf{B} = \int_0^\tau \left[s_{G(02)}(u) - s_{G(11)}(u) - s_{G(11)}(u)' + s_{20}(u) \right] \lambda_0(u) du$$
$$- \int_0^\tau \left[s_{G(01)}(u) - s_{10}(u) \right] \left[\frac{s_{G(01)}(u) - s_{10}(u)}{s_{00}(u)} \right]' \lambda_0(u) du.$$

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