ON ESTIMATION OF THE OPTIMAL TREATMENT REGIME WITH THE ADDITIVE HAZARDS MODEL

Suhyun Kang, Wenbin Lu and Jiajia Zhang

North Carolina State University and University of South Carolina

Abstract: We propose a doubly robust estimation method for the optimal treatment regime based on an additive hazards model with censored survival data. Specifically, we introduce a new semiparametric additive hazard model which allows flexible baseline covariate effects in the control group and incorporates marginal treatment effect and its linear interaction with covariates. In addition, we propose a time-dependent propensity score to construct an A-learning type of estimating equations. The resulting estimator is shown to be consistent and asymptotically normal when either the baseline effect model for covariates or the propensity score is correctly specified. The asymptotic variance of the estimator is consistently estimated using a simple resampling method. Simulation studies conducted to evaluate the finite-sample performance of the estimators are reported, and an application to AIDS clinical trial data is given to illustrate the methodology.

Key words and phrases: A-learning estimating equations, additive hazards model, doubly robust, optimal treatment regime, time-dependent propensity score.

1. Introduction

Different patients may respond differently to the same treatment due to individual heterogeneity; a new treatment may be more beneficial to some patients compared with a standard treatment, but it may have no effect or even worse effects for others. Personalized medicine, which targets tailored treatment based on patients' individual prognostic information, has recently attracted considerable attention. The main goal of personalized medicine is to find the optimal treatment regime to achieve the best expected clinical outcome of interest if the whole population is treated accordingly.

There have been extensive studies on estimating the optimal treatment regimes for uncensored data. For example, Q-learning (Watkins (1989); Watkins and Dayan (1992)) and A-learning (Murphy (2003); Robins (2004)) are commonly used methods for estimating optimal dynamic treatment regimes, where treatments may be given at multiple stages. Q-learning uses a parametric approach to model the outcome of interest given treatment and covariates and derives its associated Q-function. A-learning uses a semiparametric approach that directly models the contrast function needed for a treatment decision. A-learning has the double robustness property; the estimating equations are consistent when either the baseline effect model or the propensity score model is correctly specified. Zhang et al (2012a) proposed a doubly robust augmented inverse probability weighted estimator for the mean response given a treatment regime. Instead of directly maximizing the value function, Zhao et al (2012) proposed to estimate the optimal treatment regime by outcome weighted support vector machines in a weighted classification framework. Zhang et al (2012b) proposed a general classification framework for estimating the optimal treatment regime. These studies mainly focus on uncensored data.

For censored survival data, Goldberg and Kosorok (2012) studied Q-learning for estimating the optimal dynamic treatment regime based on the inverse probability of censoring weighted (IPCW) estimation. Zhao et al (2015) extended the outcome weighted learning approach of Zhao et al (2012) based on IPCW estimation and estimated the optimal treatment regime for the restricted mean survival time. Jiang et al (2016) proposed Kaplan-Meier type estimators for the regime-specific survival curve and estimated the optimal treatment regime by maximizing the t-year survival probability over a prespecified class of linear decision rules.

In this paper, we adapt the A-learning approach that is mainly studied for uncensored data to estimate optimal treatment regimes. A-learning is appealing due to its doubly robust property. We study the optimal treatment regime estimation for survival data based on a flexible additive hazards regression model and propose a doubly robust estimation method in the A-learning framework.

The proposed additive hazard model allows unspecified baseline covariate effects in the control group and thus has more flexibility in modeling covariate effects than the classical additive hazards model. Moreover, it gives a closed form estimator for the optimal treatment regime, which can be stably computed by the form of least squares with computational efficiency. The standard A-learning estimating equation for uncensored data, as studied in Robins (2004), cannot be used here since the corresponding estimating equations adjusted for the constant propensity score are not consistent when the baseline effect model is misspecified. To tackle this problem and obtain a doubly robust estimator, we propose using a time-dependent propensity score for constructing A-learning type estimating equations. In our method, the time-dependent propensity score is the probability that patients still at risk receive the treatment given their covariate information and is estimated nonparametrically using a kernel method. We show that after properly adjusting for the time-dependent propensity scores, the proposed estimator has the desired double robustness property as in A-learning. A simple resampling method is proposed to estimate the asymptotic variance of the estimator.

The remainder of the paper is organized as follows. In Section 2, we propose a new additive hazard model and review the estimating equation approaches of Lin and Ying (1994) for the additive hazards model and Robins (2004) for

A-learning with uncensored data. In Section 3, we propose a time-dependent propensity score, derive the doubly robust estimating equations, and establish the asymptotic properties of the resulting estimator. Section 4 is devoted to numerical studies. Some conclusions and discussions are given in Section 5. Derivations are contained in the Appendix.

2. Model and A-Learning

2.1. The proposed additive hazards model

Consider n independent subjects in a clinical trial or an observational study. For the ith subject, let Z_i be the p-dimensional vector of covariates and A_i be the observed treatment assignment. Let the A_i adopt the values 0 and 1 for control and treatment, respectively. Let T_i and C_i denote the failure time and the censoring time, respectively. Then n independently and identically distributed observations are $\{(Z_i, A_i, \tilde{T}_i, \delta_i), i = 1, \ldots, n\}$, where $\tilde{T}_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. The corresponding counting process is $N_i(t) = I(\tilde{T}_i \leq t, \delta_i = 1)$, and the at-risk process is $Y_i(t) = I(\tilde{T}_i \geq t)$.

We consider the additive hazards model

$$\lambda(t|Z_i) = \lambda(t) + \phi(Z_i) + A_i(\tilde{Z}_i^{'}\beta), \tag{2.1}$$

where $\lambda(t)$ is an unspecified baseline hazard function and $\phi(Z_i)$ is an unspecified baseline covariate effect model in the control group. For the treatment-covariate interaction effect, we consider the linear form with $\tilde{Z}_i = (1, Z_i)'$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_{p+1})'$. At (2.1), the primary interest is to estimate the interaction effect β with the corresponding optimal treatment regime given by $d^{opt}(z) = I(\tilde{z}'\beta < 0)$ where $\tilde{z} = (1, z')'$.

If $\phi(\cdot)$ were known, following Lin and Ying (1994), unadjusted estimating equations for β and λ are given by,

$$\sum_{i=1}^{n} \int_0^\infty A_i \tilde{Z}_i [dN_i(t) - Y_i(t)\{\lambda(t) + \phi(Z_i) + A_i \tilde{Z}_i' \beta\}]dt = 0, \qquad (2.2)$$

$$\sum_{i=1}^{n} [dN_i(t) - Y_i(t)\{\lambda(t) + \phi(Z_i) + A_i \tilde{Z}_i' \beta\}] = 0, \qquad (2.3)$$

respectively.

2.2. A-learning Estimating Equations

In general, the baseline covariate effect $\phi(\cdot)$ is unknown in practice. To use (2.2) and (2.3), we need to assume a parametric model for $\phi(\cdot)$, such as linear. To improve the robustness of the estimation method, one would like to derive a doubly robust estimation method incorporating the propensity score in the estimating equations.

For an uncensored response Y_i, consider the model $E(Y_i|A_i, Z_i) = \phi(Z_i) +$

$A_i(\tilde{Z}_i'\beta)$. Robins (2004) proposed an A-learning estimating equation for β,

$$\sum_{i=1}^{n} g(Z_i)\{A_i - \pi(Z_i)\}\{Y_i - h(Z_i) - A_i\tilde{Z}_i'\beta\} = 0, \tag{2.4}$$

where $\pi(Z_i) = P(A_i = 1|Z_i)$ is the propensity score, $g(Z_i)$ and $h(Z_i)$ are arbitrary functions of Z_i only, and $g(Z_i)$ is of the same dimension as β. He showed that the resulting estimator is consistent and asymptotically normal when either the posited baseline effect model $h(\cdot)$ or the propensity score $\pi(\cdot)$ is correctly specified. In addition, it was shown that if $\mathrm{Var}(Y_i|Z_i, A_i)$ is constant, choosing $g(Z_i) = \tilde{Z}_i$ and $h(Z_i) = \phi(Z_i)$ yields the most efficient estimating equation for β.

In practice, the propensity score and the baseline effect models are not known and need to be estimated. The posited models for $\pi(Z_i)$ and $\phi(Z_i)$ are denoted by $\pi(Z_i;\gamma)$ and $\phi(Z_i;\theta)$, respectively. For example, a logistic regression can be used for $\pi(Z_i;\gamma)$ and a linear model can be used for $\phi(Z_i;\theta)$. Following Robins (2004), for the proposed additive hazards model, it is natural to consider an A-learning type estimating equation for β,

$$\sum_{i=1}^{n}\int_0^{\infty} \tilde{Z}_i\{A_i - \pi(Z_i;\gamma)\}[dN_i(t) - Y_i(t)\{\lambda(t) + \phi(Z_i;\theta) + A_i\tilde{Z}_i'\beta\}]dt = 0. \tag{2.5}$$

However, this equation is generally biased when the baseline effect model is misspecified. To see this, it can be shown that the left-hand side of (2.5) multiplied by n^{-1} converges in probability to

$$E\left(\tilde{Z}_i\left\{\phi(Z_i)-\phi(Z_i;\theta)\right\}E\left[\{A_i-\pi(Z_i;\gamma)\}Y_i(t)\middle|Z_i\right]\right).$$

When the baseline effect model $\phi(Z_i;\theta)$ is misspecified, the above expectation is not zero even when the propensity score model $\pi(Z_i;\gamma)$ is correctly specified, since $E[\{A_i-\pi(Z_i;\gamma)\}Y_i(t)|Z_i]\neq 0$ due to the dependence between A_i and $Y_i(t)$ conditional on Z_i. To tackle this, we propose a new A-learning type of estimating equations by adjusting the time-dependent propensity score.

3. Proposed Estimation Method

3.1. Doubly robust estimating equations

The time-dependent propensity score is

$$\pi_Z(t)\equiv P\{A_i=1|Z,Y_i(t)=1\}=\frac{P\{Y_i(t)=1|A_i=1,Z_i\}}{P\{Y_i(t)=1|Z_i\}}\pi(Z_i). \tag{3.1}$$

Let $P(t;Z_i)=\frac{P\{Y_i(t)=1|A_i=1,Z_i\}}{P\{Y_i(t)=1|Z_i\}}$ and $\pi_Z(t;\gamma)=P(t;Z_i)\pi(Z_i;\gamma)$, where $\pi(Z_i;\gamma)$ is a posited model for $\pi(Z_i)$. Similarly, let $\phi(Z_i;\theta)$ denote the posited model for

$\phi(Z_i)$. When either $\phi(Z_i;\theta)$ or $\pi(Z_i;\gamma)$ is correctly specified,

$$E\biggl(\int_0^\infty \tilde{Z}_i\{A_i - \pi_Z(t;\gamma^*)\}\bigl[dN_i(t) - Y_i(t)\{\lambda_0(t) + \phi(Z_i;\theta^*) + A\tilde{Z}_i'\beta_0\}dt\bigr]\biggr) = 0, \tag{3.2}$$

where $\lambda_0(\cdot)$ and β_0 are the true values of $\lambda(\cdot)$ and β, respectively, and θ^* and γ^* are the corresponding population parameters for θ and γ based on the posited models $\phi(Z_i;\theta)$ and $\pi(Z_i;\gamma)$, respectively.

To prove (3.2), we first consider the case when $\phi(Z_i;\theta)$ is correctly specified but $\pi(Z_i;\gamma)$ may not be. Then, $\phi(Z_i) = \phi(Z_i;\theta^*)$ and we have

$$\begin{aligned}
&E\biggl(\int_0^\infty \tilde{Z}_i\{A_i - \pi_Z(t;\gamma^*)\}\bigl[dN_i(t) - Y_i(t)\{\lambda_0(t) + \phi(Z_i;\theta^*) + A\tilde{Z}_i'\beta_0\}dt\bigr]\biggr)\\
&= E\biggl[\int_0^\infty \tilde{Z}_i\{A_i - \pi_Z(t;\gamma^*)\}dM_i(t)\biggr] = 0,
\end{aligned}$$

where $M_i(t) = N_i(t) - \int_0^t Y_i(s)\{\lambda_0(s) + \phi(Z_i) + A_i(\tilde{Z}_i'\beta_0)\}ds$ is a mean-zero martingale process.

When $\pi(Z_i;\gamma)$ is correctly specified but $\phi(Z_i;\theta)$ is not, $\pi(Z_i) = \pi(Z_i;\gamma^*)$ and $\pi_Z(t;\gamma^*) = \pi_Z(t)$. Then, we have

$$\begin{aligned}
&E\biggl(\int_0^\infty \tilde{Z}_i\{A_i - \pi_Z(t;\gamma^*)\}\bigl[dN_i(t) - Y_i(t)\{\lambda_0(t) + \phi(Z_i;\theta^*) + A\tilde{Z}_i'\beta_0\}dt\bigr]\biggr)\\
&= E\left(\tilde{Z}_i\left\{\phi(Z_i) - \phi(Z_i;\theta)\right\} E\left[\{A_i - \pi_Z(t)\}Y_i(t)\middle| Z_i\right]\right) = 0,
\end{aligned}$$

because

$$\begin{aligned} & E[\{A_i - \pi_Z(t)\}Y_i(t)|Z] = E[\{A_i - \pi_z(t)\}|Y_i(t) = 1, Z_i]P\{Y_i(t) = 1|Z_i\} \\ = \ & [P\{A_i = 1|Y_i(t) = 1, Z_i\} - \pi_Z(t)]P\{Y_i(t) = 1|Z_i\} = 0. \end{aligned}$$

This motivates us to consider a doubly robust estimating equation for β,

$$\sum_{i=1}^{n} \int_0^{\infty} \tilde{Z}_i\{A_i - \hat{\pi}_Z(t;\gamma)\} \Big[dN_i(t) - Y_i(t)\{\lambda(t) + \phi(Z_i;\theta) + A_i\tilde{Z}_i'\beta\}dt\Big] = 0, \tag{3.3}$$

where $\hat{\pi}_Z(t;\gamma)$ is a consistent estimator of $\pi_Z(t;\gamma)$. To nonparametrically estimate $P\{Y_i(t) = 1|A_i = 1, Z_i\}$ and $P\{Y_i(t) = 1|Z_i\}$ in $\pi_Z(t;\gamma)$, we use a kernel smoothing technique. Specifically, the kernel estimators for $P\{Y_i(t) = 1|A_i = 1, Z_i\}$ and $P\{Y_i(t) = 1|Z_i\}$ are given by

$$P_{n1}(t;Z_i) = \frac{\sum_{j=1}^n Y_j(t)A_jK_h(Z_j - Z_i)}{\sum_{j=1}^n A_jK_h(Z_j - Z_i)},$$

$$P_{n2}(t;Z_i) = \frac{\sum_{j=1}^n Y_j(t)K_h(Z_j - Z_i)}{\sum_{j=1}^n K_h(Z_j - Z_i)},$$

respectively, where $K_h(\cdot)$ is a kernel function with the bandwidth h. Let $P_n(t;Z_i) = \frac{P_{n1}(t;Z_i)}{P_{n2}(t;Z_i)}$ and $\hat{\pi}_Z(t;\gamma) = P_n(t;Z_i)\pi(Z_i;\gamma)$. In the Appendix, we prove $P_n(t;Z_i) \xrightarrow{P}$

$P(t; Z_i)$ uniformly as $n \to \infty$. Accordingly, $\pi_Z(t; \gamma)$ is consistently estimated by $\hat{\pi}_Z(t; \gamma)$.

In general, the kernel function $K_h(\cdot)$ can be taken as a p-variate density function with h as a symmetric positive definite $p \times p$ matrix as discussed in Wand and Jones (1993). In practice, for simplicity, $K_h(\cdot)$ can be taken as the product of component-wise kernel functions with component-specific bandwidths. For a discrete variable, such as binary, we can set the corresponding h to be 0, and thus the kernel function reduces to an indicator function. We adopted this choice in our numerical implementations. For the derivation, following Zeng and Lin (2014), we consider a single bandwidth parameter h for notational simplicity. Specifically, we take $K_h(z) = K(||z||/h)$, where z is a p-dimensional vector with L_2-norm $||z||$ and K is a univariate density function.

The estimating equations for θ, λ, and γ are, respectively,

$$\sum_{i=1}^{n} \int_0^\infty \frac{\partial \phi(Z_i; \theta)}{\partial \theta} \Big[dN_i(t) - Y_i(t)\{\lambda(t) + \phi(Z_i; \theta) + A_i \tilde{Z}_i' \beta\} dt \Big] = 0, \tag{3.4}$$

$$\sum_{i=1}^{n} \Big[dN_i(t) - Y_i(t)\{\lambda(t) + \phi(Z_i; \theta) + A_i \tilde{Z}_i' \beta\} \Big] = 0, \tag{3.5}$$

$$\sum_{i=1}^{n} \tilde{Z}_i \{A_i - \pi(Z_i; \gamma)\} = 0. \tag{3.6}$$

In our implementation, for simplicity, we posit a logistic regression for the propensity score, $\pi(Z_i;\gamma) = \exp(\gamma'\tilde{Z}_i)/\{1+\exp(\gamma'\tilde{Z}_i)\}$, and a linear model for the baseline covariates effect, $\phi(Z_i;\theta) = Z_i^{'}\theta$. Other parametric models can be easily accommodated.

From (3.5), given β and θ, the baseline cumulative hazard function can be estimated by

$$\hat{\Lambda}(t;\beta,\theta) = \int_0^t \frac{\sum_{i=1}^n\{dN_i(u) - Y_i(u)(Z_i^{'}\theta + A_i\tilde{Z}_i^{'}\beta)du\}}{\sum_{i=1}^n Y_i(u)}.$$

Plugging this estimator into (3.3) and (3.4), we get estimating equations for β and θ, respectively, as

$$U_1(\beta,\theta,\hat{\gamma}) = \sum_{i=1}^n \int_0^\infty [\tilde{Z}_i\{A_i - \hat{\pi}_Z(t;\hat{\gamma})\} - Z^*(t;\hat{\gamma})]\{dN_i(t) - Y_i(t)(Z_i^{'}\theta + A_i\tilde{Z}_i^{'}\beta)dt\} = 0, \tag{3.7}$$

$$U_2(\beta,\theta) = \sum_{i=1}^n \int_0^\infty \{Z_i - \bar{Z}(t)\}\{dN_i(t) - Y_i(t)(Z_i^{'}\theta + A_i\tilde{Z}_i^{'}\beta)dt\} = 0, \tag{3.8}$$

where $\hat{\gamma}$ is the solution to (3.6), $Z^*(t;\hat{\gamma}) = \frac{\sum_{j=1}^n Y_j(t)\tilde{Z}_j\{A_j - \hat{\pi}_Z(t;\hat{\gamma})\}}{\sum_{j=1}^n Y_j(t)}$ and $\bar{Z}(t) = \frac{\sum_{j=1}^n Y_j(t)Z_j}{\sum_{j=1}^n Y_j(t)}$. Solving (3.7) and (3.8) jointly, we obtain the closed-form doubly robust estimator for β as

$$\hat{\beta}_D = (A - BC^{-1}D)^{-1}(h_1 - BC^{-1}h_2),$$

and the closed-form estimator for θ as

$$\hat{\theta} = (C - DA^{-1}B)^{-1}(h_2 - DA^{-1}h_1),$$

where

$$A = \sum_{i=1}^{n} \int_0^{\infty} Y_i(t)[\tilde{Z}_i\{A_i - \hat{\pi}_Z(t;\hat{\gamma})\} - Z^*(t;\hat{\gamma})]^{\otimes 2} dt,$$

$$B = \sum_{i=1}^{n} \int_0^{\infty} Y_i(t)[\tilde{Z}_i\{A_i - \hat{\pi}_Z(t;\hat{\gamma})\} - Z^*(t;\hat{\gamma})]Z_i^{'} dt,$$

$$h_1 = \sum_{i=1}^{n} \int_0^{\infty} [\tilde{Z}_i\{A_i - \hat{\pi}_Z(t;\hat{\gamma})\} - Z^*(t;\hat{\gamma})] dN_i(t).$$

Here $C = \sum_{i=1}^{n} \int_0^{\infty} Y_i(t)\{Z_i - \bar{Z}(t)\}^{\otimes 2} dt$, $D = \sum_{i=1}^{n} \int_0^{\infty} Y_i(t)\{Z_i - \bar{Z}(t)\}\tilde{Z}_i^{'} A_i dt$, $h_2 = \sum_{i=1}^{n} \int_0^{\infty} \{Z_i - \bar{Z}(t)\} dN_i(t)$, and $a^{\otimes 2} = aa^{'}$.

The estimators $\hat{\beta}_{\mathrm{D}}$ and $\hat{\theta}$ depend on the estimated baseline cumulative hazard function $\hat{\Lambda}(\cdot;\beta,\theta)$, which may not be monotonically increasing. This can affect the empirical performance of $\hat{\beta}_{\mathrm{D}}$. Based on our conducted simulations, the effect is mostly negligible.

3.2. Asymptotic properties

In this section, we establish the asymptotic properties of the estimators $\hat{\beta}_D$, $\hat{\theta}$ and $\hat{\gamma}$. Given $\beta = \beta_0$, the true value of β, consider the limiting estimating equations

$$E\left(\int_0^\infty \frac{\partial\phi(Z_i;\theta)}{\partial\theta}\left[dN_i(t) - Y_i(t)\{\lambda(t) + \phi(Z_i;\theta) + A_i\tilde{Z}_i^{'}\beta_0\}dt\right]\right) = 0,$$

$$E\left[dN_i(t) - Y_i(t)\{\lambda(t) + \phi(Z_i;\theta) + A_i\tilde{Z}_i^{'}\beta_0\}\right] = 0,$$

$$E[\tilde{Z}_i\{A_i - \pi(Z_i;\gamma)\}] = 0.$$

We assume they have unique solutions, denoted by $\lambda^*(\cdot)$, θ^*, and γ^*. These are least false parameters under possible model misspecification for $\phi(\cdot)$ and $\pi(\cdot)$. The estimation and theoretical properties of the least false parameters under model misspecification have been widely studied in the literature (e.g White (1982); Li and Duan (1989); Lin and Wei (1989)).

If we take

$$dM_i^{*0}(t;Z_i) = dN_i(t) - Y_i(t)\{\lambda^*(t) + Z_i^{'}\theta^* + A_i\tilde{Z}_i^{'}\beta_0\}dt,$$

then $E\{dM_i^{*0}(t;Z_i)|Z_i\} = 0$. In addition, let

$$q_{1i} = \int_0^\infty \left[\tilde{Z}_i\{A_i - \pi_Z(t;\gamma^*)\} - \mu_Z(t;\gamma^*)\right] dM_i^{*0}(t;Z_i) - v_{1i} + v_{2i} + v_{3i} - v_{4i},$$

$$q_{2i} = \int_0^\infty \{Z_i - \mu_Z(t;\gamma^*)\}dM_i^{*0}(t;Z_i),$$

$$q_{3i} = \tilde{Z}_i\{A_i - \pi(Z_i;\gamma^*)\},$$

$$A_{1\beta} = -E\bigg(\int_0^\infty Y_1(t)[\tilde{Z}_1\{A_1 - \pi_Z(t;\gamma^*)\} - \mu_Z(t;\gamma^*)]\tilde{Z}_1' A_1 dt\bigg),$$

$$A_{1\theta} = -E\bigg(\int_0^\infty Y_1(t)[\tilde{Z}_1\{A_1 - \pi_Z(t;\gamma^*)\} - \mu_Z(t;\gamma^*)]Z_1' dt\bigg),$$

$$A_{2\beta} = -E\bigg[\int_0^\infty Y_1(t)\{Z_1 - \mu_Z(t)\}A_1\tilde{Z}_1 dt\bigg],$$

$$A_{2\theta} = -E\bigg[\int_0^\infty Y_1(t)\{Z_1 - \mu_Z(t;\gamma^*)\}^{\otimes 2} dt\bigg],$$

$$A_{3\gamma} = -E\bigg\{\tilde{Z}_1 \frac{\partial \pi(Z_1;\gamma)}{\partial \gamma}\bigg\},$$

where $\mu_Z(t) = \frac{E\{Y_1(t)Z_1\}}{E\{Y_1(t)\}}$, $\mu_Z(t;\gamma^*) = \frac{E[Y_1(t)\tilde{Z}_1\{A_1 - \pi_Z(t;\gamma^*)\}]}{E\{Y_1(t)\}}$, and v_{1i}, v_{2i}, v_{3i}, and v_{4i} are independent mean zero random vectors with definitions given in the Appendix.

Theorem 1. Under the regularity conditions given in the Appendix, as $n \to \infty$, $h \to 0$, and $nh \to \infty$, we have that for any Z, $P_n(t;Z)$ converges uniformly to $P(t;Z)$ almost surely for $t \in [0,\tau]$, τ a fixed constant.

Theorem 2. Assume that either the propensity score or the baseline covariate effect model is correctly specified. Under the regularity conditions given in the

Appendix, as $n \to \infty$, $nh^2 \to \infty$ and $nh^4 \to 0$, we have

$$\sqrt{n}\begin{pmatrix}\hat{\beta}_D - \beta_0 \\ \hat{\theta} - \theta^* \\ \hat{\gamma} - \gamma^*\end{pmatrix} = A^{-1}\begin{pmatrix}-\sum_{i=1}^n q_{1i} \\ -\sum_{i=1}^n q_{2i} \\ -\sum_{i=1}^n q_{3i}\end{pmatrix} + o_p(1),$$

where

$$A = \begin{pmatrix}A_{1\beta} & A_{1\theta} & 0 \\ A_{2\beta} & A_{2\theta} & 0 \\ 0 & 0 & A_{3\gamma}\end{pmatrix}.$$

By the Multivariate Central Limit Theorem and Slutsky's Theorem, $\{\sqrt{n}(\hat{\beta}_D - \beta_0)', \sqrt{n}(\hat{\theta} - \theta^*)', \sqrt{n}(\hat{\gamma} - \gamma^*)'\}'$ converges in distribution to a multivariate normal with zero mean and variance-covariance matrix $A^{-1}\Sigma(A^{-1})'$, where

$$\mathbf{\Sigma} = \begin{pmatrix}E(q_1 q_1^{'}) & E(q_1 q_2^{'}) & E(q_1 q_3^{'}) \\ E(q_2 q_1^{'}) & E(q_2 q_2^{'}) & E(q_2 q_3^{'}) \\ E(q_3 q_1^{'}) & E(q_3 q_2^{'}) & E(q_3 q_3^{'})\end{pmatrix}.$$

3.3. Estimation of the asymptotic variance

We obtain a closed-form expression of the asymptotic variance. The matrix Σ has a complicated form and it may not be easy to obtain the stable variance estimator based on the usual plug-in method. Therefore, we adopt a resampling scheme here, as in Jin et al (2001), to approximate the asymptotic distribution of $\hat{\beta}_D$.

First, we generate n iid standard exponential random variables $\{G_i, i = 1, \ldots, n\}$. Then we solve the following while fixing the data at their observed values:

$$\sum_{i=1}^{n} G_i \int_0^\infty \tilde{Z}_i\{A_i - \tilde{\pi}_Z(t;\gamma)\} \left[dN_i(t) - Y_i(t)\{\lambda(t) + Z_i^{'}\theta + A_i\tilde{Z}_i^{'}\beta\}dt\right] = 0, \quad (3.9)$$

$$\sum_{i=1}^{n} G_i \int_0^\infty Z_i \left[dN_i(t) - Y_i(t)\{\lambda(t) + Z_i^{'}\theta + A_i\tilde{Z}_i^{'}\beta\}dt\right] = 0, \quad (3.10)$$

$$\sum_{i=1}^{n} G_i \left[dN_i(t) - Y_i(t)\{\lambda(t) + Z_i^{'}\theta + A_i\tilde{Z}_i^{'}\beta\}dt\right] = 0, \quad (3.11)$$

$$\sum_{i=1}^{n} G_i\tilde{Z}_i\{A_i - \pi(Z_i;\gamma)\} = 0, \quad (3.12)$$

where $\tilde{\pi}_Z(t;\gamma) = \frac{\sum_{j=1}^n G_j Y_j(t) A_j K_h(Z_j - Z_i)}{\sum_{j=1}^n G_j A_j K_h(Z_j - Z_i)} \frac{\sum_{j=1}^n G_j K_h(Z_j - Z_i)}{\sum_{j=1}^n G_j Y_j(t) K_h(Z_j - Z_i)} \pi(Z_i;\gamma)$ is the perturbed version of $\hat{\pi}_Z(t;\gamma)$. Let $(\tilde{\beta}, \tilde{\theta}, \tilde{\lambda}, \tilde{\gamma})$ be the resulting solutions. By generating $\{G_i, i = 1, \ldots, n\}$ M times, we can obtain a large set of resampled estimates, $\{\tilde{\beta}_l, l = 1, \ldots, M\}$. Following Jin et al (2001), it can be shown that, given the observed data, the conditional distribution of $\sqrt{n}(\tilde{\beta} - \hat{\beta}_D)$ is asymptotically equivalent to that of $\sqrt{n}(\hat{\beta}_D - \beta_0)$, and the variance of $\hat{\beta}_D$ can be estimated by the empirical variance of $\tilde{\beta}$.

4 Numerical studies

4.1. Simulation studies

We carried out simulation studies to assess the performance of the proposed doubly robust estimator. The failure time T was generated from the additive hazard model (2.1). Two independent covariates were considered, where Z_1 was Bernoulli with success probability of 0.5, and Z_2 was uniform on $[-2, 2]$. We chose the regression parameter $\beta = (\beta_0, \beta_1, \beta_2)' = (0, 1, 1)'$ and the baseline hazard function $\lambda_0(t) = 3$. The censoring time C was uniform on $U[0, c_0]$, where c_0 was chosen to yield 15% or 40% censoring rates.

For estimation, we considered both correctly specified and misspecified models for $\phi(Z_i;\theta)$ and $\pi(Z_i;\gamma)$. We considered three baseline effect models for $\phi(Z_i;\theta)$: $\phi_1(Z_i;\theta_1) = Z_i'\theta_1$; $\phi_2(Z_i;\theta) = 0.5(Z_i'\theta_1)(Z_i'\theta_2)$; and $\phi_3(Z_i;\theta) = \sin(\pi Z_i'\theta_1) + 0.1(1+Z_i'\theta_2)^2$. The first poisited linear model is correctly specified while the other two are misspecified. We set $\theta_1 = (0.5, 0.5)$, $\theta_2 = (1, 0.5)$. For the propensity score, we considered the $\pi_1(Z_i;\gamma_1) = 0.5$; $\pi_2(Z_i;\gamma_1) = \exp(\tilde{Z}_i'\gamma_1)/\{1+\exp(\tilde{Z}_i'\gamma_1)\}$; and $\pi_3(Z_i;\gamma) = \exp\{(\tilde{Z}_i'\gamma_1)(\tilde{Z}_i'\gamma_2)\}/[1 + \exp\{(\tilde{Z}_i'\gamma_1)(\tilde{Z}_i'\gamma_2)\}]$. For the first two, the posited logistic regression model is correctly specified while for the last, it is not. We set $\gamma_1 = (0, 0.5, 0.5)$ and $\gamma_2 = (0.6, -0.1, 0)$. We compared the proposed doubly robust estimator (denoted by DR) with the unadjusted estimator of Lin and Ying (1994) as the solutions to (2.2) and (2.3) (denoted by YL), and the adjusted estimator with the time-invariant propensity score as the solutions to (2.5) (denoted by YL(π)). For each scenario, we conducted 500 runs of sample size N=500.

For the bandwidth parameter h for the continuous covariate in the kernel estimator for our method, we took the optimal bandwidth $h = 4^{1/3}\sigma n^{-1/3}$, following Jones (1990), where σ is the standard deviation of Z_2 and n is the sample size. To estimate the asymptotic variance of the estimator, we generated $M = 500$ sets of $\{G_i, i = 1, \ldots, n\}$ for each simulated data and estimated the asymptotic variance of $\hat{\beta}_D$ using the sample variance of $\tilde{\beta}$'s.

The results for 15% and 40% censoring are summarized in Tables 4.1 and 4.2, respectively. The proposed doubly robust estimators are nearly unbiased for all scenarios, showing the double robustness as established in Theorem 2. Lin and Ying (1994)'s unadjusted estimators are biased when the baseline covariate effect model $\phi(Z_i; \theta)$ is misspecified. The time-invariant propensity score-adjusted estimators are also biased when either the baseline effect model or the propensity score model is misspecified. The doubly robust estimators lose some efficiency due to the additional estimation for the time-dependent propensity score function, but the extent of the efficiency loss is negligible. The estimated standard errors (SE) based on the resampling method are all close to the sample standard deviations of the estimates (SD). The Wald-type 95% confidence intervals of $\hat{\beta}_{\mathrm{D}}$

have proper empirical coverage probabilities.

To assess the computational cost of the proposed resampling method for variance estimation, we report the average (in seconds) and standard deviation of the computation time over 500 simulation runs for different numbers of resampling sets. We considered the simulation settings with B1 and P1. The values are given in Table 4.3. The computation time linearly increases with the number of resampling sets. For $M = 500$, it takes less than 4 minutes, and has a moderate computational cost.

Additional simulations were conducted to compare the proposed method with the methods of Goldberg and Kosorok (2012) (denoted by Q-survival) and Zhao et al (2015) (denoted by OWL). We considered the same simulation settings as before with the censoring rate of 15%. To evaluate the accuracy of the estimated optimal treatment regimes, we computed both the percentage of correct decision (PCD) and the value of the estimated treatment regimes. Here, the PCD for each simulation run was defined as $1 - \sum_{i=1}^{N} |\hat{d}^{opt}(z) - d^{opt}(z)|/N$, where $d^{opt}(z) = I(\tilde{z}'\beta < 0)$; while the value was computed as the mean survival time (MST) under the estimated optimal treatment regime obtained using 10,000 independently generated subjects. The results are given in Table 4.4. Under all scenarios, the proposed method yields higher accuracy in terms of PCD and gives larger MST than the other methods.

We studied the performance of the proposed method when the assumed additive hazards model was violated. Specifically, we conducted additional simulations under the proportional hazards model with the same combinations of the baseline effect and propensity score models as before. We compared the proposed method with the methods of Goldberg and Kosorok (2012) and Zhao et al (2015), and report both the percentage of correct decision (PCD) and the value of the estimated treatment regimes in terms of MST. The results for the 15% censoring rate are given in Table 4.5. Under all scenarios, the proposed method gives larger PCD and MST than other methods. This implies that the proposed method performs competitively for estimating the optimal treatment regime even when the assumed additive hazards model is violated.

4.2. Application to AIDS study (ACTG175)

We applied the proposed estimation method to a data set from AIDS Clinical Trials Group Protocol 173 (ACTG175). The study enrolled 2139 HIV-infected patients who were randomly assigned to four different antiretroviral treatment regimes; Zidovudine(ZDV) plus monotherapy, ZDV plus didanosine (ddI), ZDV plus zalcitabine (zal), and ddI monotherapy (Hammer et al (1996)). In our analysis, we focus on two groups: ZDV+ddI as treatment 1 and ZDV+zal as treatment 0. The treatment 1 group has $n_1 = 522$ patients and the treatment 0 group has $n_0 = 524$ patients, thus $\pi(Z_i) = 0.5$. A primary endpoint of interest is

Table 4.1: Simulation Results : Censoring rate 15%.

		DR				YL		YL(π)	
		Estimator	SD	SE	CP	Estimator	SD	Estimator	SD
B1,P1	β_0	-0.02	0.46	0.47	0.96	-0.02	0.43	-0.02	0.43
	β_1	1.03	0.40	0.41	0.96	1.01	0.38	1.01	0.39
	β_2	1.02	0.36	0.36	0.95	1.00	0.33	1.00	0.33
B1,P2	β_0	-0.01	0.47	0.48	0.96			-0.03	0.44
	β_1	0.99	0.44	0.44	0.95			0.98	0.41
	β_2	1.02	0.37	0.38	0.95			0.98	0.34
B1, P3	β_0	0.02	0.47	0.48	0.95			-0.01	0.47
	β_1	0.99	0.41	0.41	0.95			0.91	0.40
	β_2	1.00	0.36	0.36	0.94			0.92	0.35
B2, P1	β_0	-0.02	0.47	0.48	0.96	-0.09	0.46	0.09	0.47
	β_1	1.03	0.42	0.43	0.96	1.15	0.42	0.86	0.42
	β_2	1.02	0.39	0.38	0.94	1.15	0.37	0.86	0.35
B3, P1	β_0	-0.01	0.48	0.50	0.96	0.03	0.43	-0.11	0.48
	β_1	1.03	0.42	0.43	0.95	0.83	0.42	1.14	0.43
	β_2	1.02	0.38	0.38	0.96	0.88	0.35	1.05	0.38

† B, Baseline effect model; $B1 = \phi_1(Z_i;\theta)$, $B2 = \phi_2(Z_i;\theta)$, $B3 = \phi_3(Z_i;\theta)$. P, Propensity score model; $P1 = \pi_1(Z_i;\gamma)$, $P2 = \pi_2(Z_i;\gamma)$, $P3 = \pi_3(Z_i;\gamma)$. Est, mean of the estimates; SD, sample standard deviation of the estimates; SE, mean of the estimated standard errors; CP, empirical coverage probability of Wald-type 95% confidence intervals.

the time until one of the following events occur; having a larger than 50% decline in the CD4 count, progressing to AIDS, or death. Among $n = 1046$ patients, about 21% of them have experienced the outcome of interest. Based on Lu et al (2013), we include the baseline covariates age after log transformation and homosexual activity (0=no, 1=yes) in the model.

We checked the goodness-of-fit of an additive hazards model with the linear

Table 4.2: Simulation Results : Censoring rate 40%.

		DR				YL		YL(π)	
		Estimator	SD	SE	CP	Estimator	SD	Estimator	SD
B1,P1	β_0	-0.03	0.56	0.56	0.94	-0.04	0.54	-0.04	0.55
	β_1	1.03	0.49	0.48	0.95	1.03	0.47	1.04	0.47
	β_2	1.02	0.43	0.42	0.95	1.00	0.40	1.00	0.41
B1,P2	β_0	-0.01	0.57	0.56	0.96			-0.02	0.56
	β_1	0.99	0.51	0.50	0.95			0.98	0.49
	β_2	1.02	0.42	0.45	0.97			0.99	0.41
B1, P3	β_0	0.02	0.57	0.57	0.96			-0.02	0.59
	β_1	0.98	0.49	0.49	0.95			0.92	0.47
	β_2	1.00	0.42	0.43	0.95			0.94	0.44
B2, P1	β_0	-0.04	0.60	0.58	0.95	-0.15	0.59	0.05	0.58
	β_1	1.04	0.51	0.50	0.95	1.18	0.51	0.89	0.51
	β_2	1.02	0.46	0.45	0.96	1.20	0.42	0.92	0.42
B3,P1	β_0	-0.03	0.59	0.60	0.95	0.14	0.59	-0.10	0.58
	β_1	1.04	0.50	0.50	0.95	0.77	0.50	1.11	0.49
	β_2	1.02	0.45	0.44	0.95	0.86	0.44	1.05	0.44

† B, Baseline effect model; $B1 = \phi_1(Z_i; \theta)$, $B2 = \phi_2(Z_i; \theta)$, $B3 = \phi_3(Z_i; \theta)$.
P, Propensity score model; $P1 = \pi_1(Z_i; \gamma)$, $P2 = \pi_2(Z_i; \gamma)$, $P3 = \pi_3(Z_i; \gamma)$.
Est, mean of the estimates; SD, sample standard deviation of the estimates; SE, mean of the estimated standard errors; CP, empirical coverage probability of Wald-type 95% confidence intervals.

Table 4.3: Computational Times (In seconds).

M	100	250	500
Mean	39.64	109.426	227.08
SD	3.672	2.098	9.149

baseline and treatment-covariate interaction effects for the AIDS data. The martingale residual plot of the fitted model is given in Figure 4.1, which shows no systemic patterns or trends. This implies that the additive hazard model fits the

Table 4.4: Simulation results for comparisons with Goldberg and Kosorok (2012) and Zhao et al (2015) under the additive hazards model.

		DR		Q-Survival		OWL	
		Mean	SD	Mean	SD	Mean	SD
B1,P1	PCD	0.882	0.085	0.744	0.113	0.717	0.131
	MST	0.270	0.006	0.259	0.015	0.244	0.019
B1,P2	PCD	0.881	0.088	0.732	0.131	0.692	0.156
	MST	0.269	0.007	0.254	0.019	0.229	0.012
B1,P3	PCD	0.876	0.093	0.736	0.125	0.674	0.178
	MST	0.269	0.006	0.257	0.017	0.237	0.019
B2,P1	PCD	0.878	0.091	0.729	0.117	0.713	0.138
	MST	0.255	0.006	0.244	0.014	0.233	0.016
B3,P1	PCD	0.880	0.085	0.766	0.106	0.730	0.115
	MST	0.262	0.005	0.256	0.012	0.242	0.016

Table 4.5: Simulation results for comparisons with Goldberg and Kosorok (2012) and Zhao et al (2015) under the proportional hazards model.

		DR		Q-Survival		OWL	
		mean	sd	mean	sd	mean	sd
B1,P1	PCD	0.811	0.148	0.612	0.153	0.671	0.183
	MST	1.464	0.033	1.436	0.037	1.429	0.038
B1,P2	PCD	0.811	0.156	0.674	0.151	0.697	0.153
	MST	1.465	0.034	1.436	0.039	1.424	0.032
B1,P3	PCD	0.819	0.153	0.667	0.148	0.661	0.189
	MST	1.467	0.034	1.435	0.038	1.426	0.039
B2,P1	PCD	0.799	0.148	0.685	0.156	0.657	0.193
	MST	1.503	0.034	1.479	0.037	1.467	0.045
B3,P1	PCD	0.822	0.158	0.629	0.138	0.689	0.163
	MST	1.536	0.043	1.493	0.057	1.487	0.051

data reasonably well. We also considered smoothed estimates of the conditional hazard functions based on the local Nelson-Aalen estimators of the conditional cumulative hazard functions. The estimated smoothed conditional hazard functions are rather additive than multiplicative. This implies an additive hazard model may give a better fit than a proportional hazards model. The corresponding plots are not given here. The graphical evidences give some justifications for using the additive hazards model for the AIDS data. We applied the proposed method to the data and obtained the doubly robust estimator for the optimal treatment regime. For standard error estimation, we used the resampling approach with $M = 500$ sets of $\{G_i, i = 1, \ldots, n\}$. For comparison, we considered Lin and Ying (1994)'s unadjusted estimator. The estimation results are given in Table 4.6. From the results, the two meethods give comparable results. A possible explanation for this is that the linear baseline effect model may be a proper fit to the data as shown by the martingale residual plot given in Figure 4.1. In the simulation study, we observed that when the baseline model is correctly specified, both methods give similar results. The estimated optimal treatment regime is $d^{opt}(z) = I(0.341 - 0.104 \text{ age} + 0.033 \text{ homo} < 0)$. Under both methods, age and intercept are significant while homosexual activity is close to significant. In addition, treatment 1 is more beneficial than treatment 0 for older patients while treatment 0 is more favorable for younger patients with homo=1. The results agree with previous findings in Lu et al (2013).

5. Discussion

In this paper, we propose a doubly robust estimation method for the optimal treatment regime in an additive hazards model with censored survival data.

Figure 4.1: Martingale residuals of the additive hazard model for age

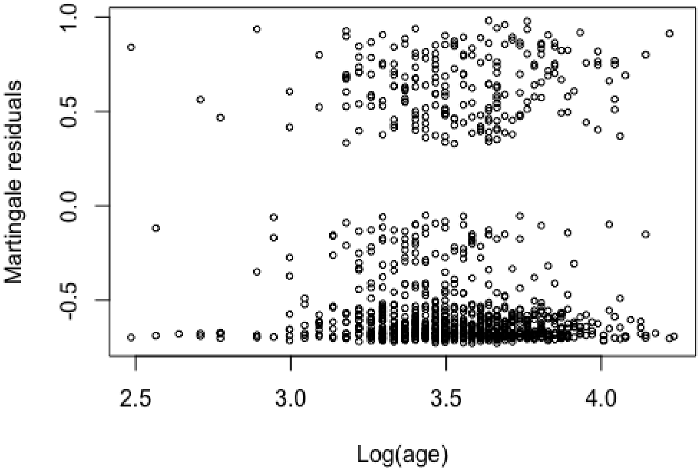

Table 4.6: Application to AIDS study

	DR		YL	
Estimator	Est	SE	Est	SE
intercept	0.341	0.164	0.338	0.178
age	-0.104	0.047	-0.103	0.051
homo	0.033	0.024	0.034	0.022

By incorporating time-dependent propensity scores, the proposed estimator has an improved robustness against misspecification of the baseline covariate effect model as in A-learning. We can extend our method to other survival models, for example, the Cox PH model. The corresponding estimation is much more complicated due to the multiplicative hazard function of the Cox model. A further investigation is warranted.

As the dimension of the covariates increases, the kernel estimation used to estimate the time-dependent propensity scores can suffer the curse of dimensionality. In addition, not all the covariates are related to the treatment decision. Variable selection can be incorporated to identify important covariates associated with treatment decision. Following Martinussen and Scheike (2009), the corresponding least-square loss can be written as

$$L(\beta) = \beta^{'}(A - BC^{-1}D)\beta - 2\beta^{'}(h_1 - BC^{-1}h_2),$$

where A, B, C, D, h_1, and h_2 are given in Section 3.1. Then, penalized estimation, such as Lasso and SCAD, can be easily incorporated.

Appendix

To establish the asymptotic results given in Theorems 1-2, we assume the following regularity conditions.

(C1) The covariate Z has bounded support; the density function of Z is continuously differentiable in the support of Z and is bounded away from 0; If $\tilde{Z}'v = 0$ for some constant vector v with probability one, then $v = 0$.

(C2) The probability $P\{Y(\tau) = 1\} > 0$, where τ is a fixed constant; the function $\Lambda_0(t)$ is continuously differentiable with $\Lambda_0(\tau) < \infty$.

(C3) The true parameter vector β_0 is an interior point of a known compact set

$\mathcal{B}$ in $\mathcal{R}^p$.

(C4) The true propensity score $\pi(Z)$ is bounded away from zero and one for all possible values of Z.

(C5) The kernel function $K_h(\cdot)$ is thrice-continuously differentiable with bounded variations.

(C6) The matrices $A_{1\beta}$, $A_{1\theta}$, $A_{2\beta}$, $A_{2\theta}$, $A_{3\gamma}$, and A are positive definite.

Conditions (C1)-(C3) are standard in survival analysis and are used to establish the consistency of the estimator of β. Conditions (C4)-(C5) are used to establish the uniform consistency and convergence rate of the kernel estimator of the time-dependent propensity score. Condition (C6) is required for establishing the asymptotic normality of the estimator of β.

Proof of Theorem 1. Under the assumed regularity conditions, by Lemma 2.4 of Schuster(1969), we have

$$\sup_{t\in[0,\tau]}\left|\frac{\frac{1}{n}\sum_{j=1}^{n}Y_j(t)A_jK_h(Z_j-Z_i)}{\frac{1}{n}\sum_{j=1}^{n}A_jK_h(Z_j-Z_i)}-\frac{P\{Y(t)=1,A=1|Z=Z_i\}f_Z(Z_i)}{P(A=1|Z=Z_i)f_Z(Z_i)}\right|\to 0,$$

$$\sup_{t\in[0,\tau]}\left|\frac{\frac{1}{n}\sum_{j=1}^{n}Y_j(t)K_h(Z_j-Z_i)}{\frac{1}{n}\sum_{j=1}^{n}K_h(Z_j-Z_i)}-\frac{P\{Y(t)=1|Z=Z_i\}f_Z(Z_i)}{f_Z(Z_i)}\right|\to 0,$$

where $f_Z(\cdot)$ is the density function of Z. Therefore,

$$\sup_{t\in[0,\tau]}\left|\frac{\frac{1}{n}\sum_{j=1}^{n}Y_j(t)A_jK_h(Z_j-Z_i)}{\frac{1}{n}\sum_{j=1}^{n}A_jK_h(Z_j-Z_i)}\frac{\frac{1}{n}\sum_{j=1}^{n}K_h(Z_j-Z_i)}{\frac{1}{n}\sum_{j=1}^{n}Y_j(t)K_h(Z_j-Z_i)}-P(A=1|Z,Y(t)=1)\right|\to 0.$$

This proves Theorem 1.

Proof for Theorem 2. By a Taylor expansion and some empirical process approximation techniques, we have

$$\begin{aligned}
0 &= \frac{1}{\sqrt{n}}U_1(\hat{\beta},\hat{\theta},\hat{\gamma}) = \frac{1}{\sqrt{n}}U_1(\beta_0,\hat{\theta},\hat{\gamma}) + A_{1\beta}\sqrt{n}(\hat{\beta}-\beta_0) + o_p(1) \\
&= \frac{1}{\sqrt{n}}U_1(\beta_0,\theta^*,\gamma^*) + A_{1\beta}\sqrt{n}(\hat{\beta}-\beta_0) + A_{1\gamma}\sqrt{n}(\hat{\gamma}-\gamma^*) + A_{1\theta}\sqrt{n}(\hat{\theta}-\theta^*) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\infty}\left[\tilde{Z}_i\{A_i-\pi(Z_i,\gamma^*)P(t;Z_i)\}-\mu_Z(t;\gamma^*)\right]dM_i^{*0}(t;Z_i) \\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\infty}\tilde{Z}_i\pi(Z_i;\gamma^*)\{P_n(t;Z_i)-P(t;Z_i)\}dM_i^{*0}(t;Z_i) \\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\infty}\{Z^*(t;\gamma^*)-\mu_Z(t;\gamma^*)\}dM_i^{*0}(t;Z_i) \\
&\quad +A_{1\beta}\sqrt{n}(\hat{\beta}-\beta_0) + A_{1\gamma}\sqrt{n}(\hat{\gamma}-\gamma^*) + A_{1\theta}\sqrt{n}(\hat{\theta}-\theta^*) + o_p(1), \qquad (A.1)
\end{aligned}$$

where

$$\begin{aligned}
A_{1\gamma} &= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\tilde{Z}_i\dot{\pi}(Z_i;\gamma^*)P_n(t;Z_i)dM_i^{*0}(t;Z_i) \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\frac{\sum_{j=1}^{n}\tilde{Z}_j\{A_j-\dot{\pi}(Z_j;\gamma^*)P_n(t;Z_i)\}}{\sum_{j=1}^{n}Y_j(t)}dM_i^{*0}(t;Z_i) = o_p(1),
\end{aligned}$$

and $\dot{\pi}(Z_i;\gamma) = \partial\pi(Z_i;\gamma)/\partial\gamma$.

For (A.1), if we write

$$\begin{aligned} Z^*(t;\gamma^*) &= \frac{\frac{1}{n}\sum_{j=1}^n Y_j(t)\tilde{Z}_j A_j - P_n(t;Z_i)\{\frac{1}{n}\sum_{j=1}^n Y_j(t)\tilde{Z}_j\pi(Z_j;\gamma^*)\}}{\frac{1}{n}\sum_{j=1}^n Y_j(t)} \\ &\equiv \frac{G_n(t)}{H_n(t)}, \end{aligned}$$

$$\mu_Z(t;\gamma^*) = \frac{E[Y_1(t)\tilde{Z}_1\{A_1 - \pi(Z_1;\gamma^*)P(t;Z_i)\}]}{E\{Y_1(t)\}} \equiv \frac{G(t)}{H(t)},$$

then

$$\begin{aligned} & \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\infty \{Z^*(t;\gamma^*) - \mu_Z(t;\gamma^*)\}dM_i^{*0}(t;Z_i) \\ = & \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\infty \left\{\frac{G_n(t)}{H_n(t)} - \frac{G(t)}{H(t)}\right\}dM_i^{*0}(t;Z_i) \\ = & \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\infty \left[\frac{G_n(t)-G(t)}{H(t)} - \frac{G(t)\{H_n(t)-H(t)\}}{H(t)^2}\right]dM_i^{*0}(t;Z_i) + o_p(1) \\ = & \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\infty \frac{1}{H(t)}\{G_n(t) - G(t)\}dM_i^{*0}(t;Z_i) + o_p(1). \end{aligned}$$

Applying kernel techniques, and after some algebra, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\infty}\{Z^*(t;\gamma^*)-\mu_Z(t;\gamma^*)\}dM_i^{*0}(t;Z_i)\\
= \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\infty}\frac{1}{H(t)}\Bigg(\frac{1}{n}\sum_{j=1}^{n}Y_j(t)\tilde{Z}_j\{A_j-\pi(Z_j;\gamma^*)P_n(t;Z_i)\}\\
&-E[Y_1(t)\tilde{Z}_1\{A_1-\pi(Z_1;\gamma^*)P(t;Z_i)\}]\Bigg)dM_i^{*0}(t;Z_i)+o_p(1)\\
= \;& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\infty}\frac{1}{H(t)}E\{Y_1(t)\tilde{Z}_1\pi(Z_1,\gamma^*)\}\{P_n(t;Z_i)-P(t;Z_i)\}dM_i^{*0}(t;Z_i)+o_p(1).
\end{aligned}$$

In combination, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\infty}\left[\tilde{Z}_i\pi(Z_i;\gamma^*)-\frac{1}{H(t)}E\{Y_1(t)\tilde{Z}_1\pi(Z_1,\gamma^*)\}\right]\{P_n(t;Z_i)-P(t;Z_i)\}dM_i^{*0}(t;Z_i). \tag{A.2}$$

To simplify the notation, write $P_n(t;Z_i)=\frac{A_n(t)}{B_n(t)}$ and $P(t;Z_i)=\frac{A(t)}{B(t)}$. Then

$$A_n(t)=\frac{\frac{1}{n}\sum_{j=1}^{n}A_jY_j(t)K_h(Z_j-Z_i)}{\frac{1}{n}\sum_{j=1}^{n}A_jK_h(Z_j-Z_i)}\xrightarrow{p}A(t)=P\{Y_1(t)=1|A_1=1,Z_1=Z_i\},$$

$$B_n(t)=\frac{\frac{1}{n}\sum_{j=1}^{n}Y_j(t)K_h(Z_j-Z_i)}{\frac{1}{n}\sum_{j=1}^{n}K_h(Z_j-Z_i)}\xrightarrow{p}B(t)=P\{Y_1(t)=1|Z_1=Z_i\}.$$

In addition, we have

$$
\begin{aligned}
C_n(t) &\equiv \frac{1}{n}\sum_{j=1}^{n} A_j Y_j(t) K_h(Z_j - Z_i) \xrightarrow{p} C(t) = P\{A_1 = 1, Y_1(t) = 1 | Z_1 = Z_i\} f_Z(Z_i),\\
D_n &= \frac{1}{n}\sum_{j=1}^{n} A_j K_h(Z_j - Z_i) \xrightarrow{p} D = P(A_1 = 1 | Z_1 = Z_i) f_Z(Z_i),\\
E_n(t) &= \frac{1}{n}\sum_{j=1}^{n} Y_j(t) K_h(Z_j - Z_i) \xrightarrow{p} E(t) = P(Y_1(t) = 1 | Z_1 = Z_i) f_z(Z_i),\\
F_n &= \frac{1}{n}\sum_{j=1}^{n} K_h(Z_j - Z_i) \xrightarrow{p} F = f_z(Z_i).
\end{aligned}
$$

Therefore, (A.2) can be written as

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty \left[\tilde{Z}_i\pi(Z_i;\gamma^*) + \frac{1}{H(t)}E\{Y_1(t)\tilde{Z}_1\pi(Z_1,\gamma^*)\}\right] \frac{A_n(t) - A(t)}{B_n(t)} dM_i^{*0}(t; Z_i)\\
& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty \left[\tilde{Z}_i\pi(Z_i;\gamma^*) - \frac{1}{H(t)}E\{Y_1(t)\tilde{Z}_1\pi(Z_1,\gamma^*)\}\right] \frac{A(t)\{B_n(t) - B(t)\}}{B_n(t)B(t)} dM_i^{*0}(t; Z_i)\\
= \quad & \frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \frac{1}{B(t)}\left[\tilde{Z}_i\pi(Z_i;\gamma^*) - \frac{1}{H(t)}E\{Y_1(t)\tilde{Z}_1\pi(Z_1,\gamma^*)\}\right] \sqrt{n}\frac{C_n(t) - C(t)}{D(t)} dM_i^{*0}(t; Z_i)\\
& -\frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \frac{1}{B(t)}\left[\tilde{Z}_i\pi(Z_i;\gamma^*) - \frac{1}{H(t)}E\{Y_1(t)\tilde{Z}_1\pi(Z_1,\gamma^*)\}\right] \sqrt{n}\frac{C(t)(D_n - D)}{D(t)^2} dM_i^{*0}(t; Z_i)\\
& -\frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \left[\tilde{Z}_i\pi(Z_i;\gamma^*) - \frac{1}{H(t)}E\{Y_1(t)\tilde{Z}_1\pi(Z_1,\gamma^*)\}\right] \frac{A(t)}{B(t)^2}\sqrt{n}\frac{E_n(t) - E(t)}{F(t)} dM_i^{*0}(t; Z_i)\\
& +\frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \left[\tilde{Z}_i\pi(Z_i;\gamma^*) - \frac{1}{H(t)}E\{Y_1(t)\tilde{Z}_1\pi(Z_1,\gamma^*)\}\right] \frac{A(t)}{B(t)^2}\sqrt{n}\frac{E(t)(F_n - F)}{F(t)^2} dM_i^{*0}(t; Z_i)\\
& +o_p(1).
\end{aligned}
$$

By some empirical process approximation and kernel estimation techniques, the first term in the last expression can be written as

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\int_0^\infty H_1(t;Z_i)\Big[A_jY_j(t)K_h(Z_j-Z_i)-E\{A_1Y_1(t)|Z_1=Z_i\}f(Z_i)\Big]dM_i^{*0}(t;Z_i)+o_p(1)\\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\int_0^\infty H_1(t;Z_i)A_jY_j(t)K_h(Z_j-Z_i)dM_i^{*0}(t;Z_i)\\
& -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{1}{n}\sum_{i=1}^{n}\int_0^\infty H_1(t;Z_i)E\{A_1Y_1(t)|Z_1=Z_i\}f(Z_i)dM_i^{*0}(t;Z_i)+o_p(1)\\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty H_1(t;Z_i)E\{A_1Y_1(t)|Z_1=Z_i\}f(Z_i)dM_i^{*0}(t;Z_i)\\
& -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}E\Big[\int_0^\infty H_1(t;Z_i)E\{A_1Y_1(t)|Z_1=Z_i\}f(Z_i)dM_i^{*0}(t;Z_i)\Big]+o_p(1)\\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty H_1(t;Z_i)E\{A_1Y_1(t)|Z_1=Z_i\}f(Z_i)dM_i^{*0}(t;Z_i)+o_p(1)\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_{1i}+o_p(1),
\end{aligned}
$$

where $H_1(t;Z_i)=\frac{1}{B(t)D(t)}\Big[\tilde{Z}_i\pi(Z_i;\gamma^*)-\frac{1}{H(t)}E\{Y_1(t)\tilde{Z}_1\pi(Z_1,\gamma^*)\}\Big]$.

Here the v_{1i}'s are i.i.d. mean-zero vectors. Similarly, after some calculations, the remaining terms can be asymptotically represented as a summation of i.i.d. mean-zero vectors, which are denoted by v_{2i}, v_{3i}, and v_{4i}, respectively. Therefore, we have

$$
\begin{aligned}
0 &= \frac{1}{\sqrt{n}}U_1(\hat{\beta},\hat{\theta},\hat{\gamma}) \qquad \text{(A.3)}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\int_0^\infty \left[\tilde{Z}_i\{A_i-\pi(Z_i;\gamma^*)P(t;Z_i)\}-\mu_Z(t;\gamma^*)\right]dM_i^{*0}(t;Z_i)-v_{1i}+v_{2i}+v_{3i}-v_{4i}\right)\\
&\quad +A_{1\beta}\sqrt{n}(\hat{\beta}-\beta_0)+A_{1\theta}\sqrt{n}(\hat{\theta}-\theta^*)+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}q_{1i}+A_{1\beta}\sqrt{n}(\hat{\beta}-\beta_0)+A_{1\theta}\sqrt{n}(\hat{\theta}-\theta^*)+o_p(1).
\end{aligned}
$$

Following similar arguments for studying the estimates of the least false parameters in misspecified models, we have

$$
\begin{aligned}
0=\frac{1}{\sqrt{n}}U_2(\hat{\beta},\hat{\theta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty\{Z_i-\mu_Z(t)\}dM_i^{*0}(t;Z_i)+A_{2\beta}\sqrt{n}(\hat{\beta}-\beta_0)+A_{2\theta}\sqrt{n}(\hat{\theta}-\theta^*)+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}q_{2i}+A_{2\beta}\sqrt{n}(\hat{\beta}-\beta_0)+A_{2\theta}\sqrt{n}(\hat{\theta}-\theta^*)+o_p(1), \qquad \text{(A.4)}
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{Z}_i\{A_i-\pi(Z_i;\gamma^*)\}+A_{3\gamma}\sqrt{n}(\hat{\gamma}-\gamma^*)+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}q_{3i}++A_{3\gamma}\sqrt{n}(\hat{\gamma}-\gamma^*)+o_p(1). \qquad \text{(A.5)}
\end{aligned}
$$

Putting (A.3), (A.4), and (A.5) together gives the representation

$$\begin{pmatrix} A_{1\beta} & A_{1\theta} & 0 \\ A_{2\beta} & A_{2\theta} & 0 \\ 0 & 0 & A_{3\gamma} \end{pmatrix} \sqrt{n} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta^* \\ \hat{\gamma} - \gamma^* \end{pmatrix} = \begin{pmatrix} -\sum_{i=1}^n q_{1i} \\ -\sum_{i=1}^n q_{2i} \\ -\sum_{i=1}^n q_{3i} \end{pmatrix} + o_p(1).$$

Theorem 2 then follows.

References

Goldberg, Y. and Kosorok, M. R. (2012). Q-learning with censored data. *Annals of Statistics Statist* **40,** 529-560.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundaker, H., Schooley, R. T., Haubrich, R. H. et al. (1996) A trial comparing HIV-infected adults with CD4 cell counts from 200 to 400 per cubic millimeter. *New England Journal of Medicine* **335,** 1081-1089.

Jiang, R., Lu, W., Song, R., and Davidian, M. (2016). On estimation of optimal treatment regimes for maximizing t-year survival probability. *Journal of the Royal Statistical Society: Series B,* **88,** 381-390.

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

Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation. *Statistics and Probability Letters* **9,** 129-132.

Li, K.-C. and Duan, N. (1989). Regression analysis under link violation. *Annals of Statistics* **17,** 1009-1052.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84,** 1074-1078.

Lu, W., Zhang, H. H. and Zeng, D. (2013). Variable selection for optimal treatment decision *Statistical methods in medical research* **22,** 493–504.

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

Martinussen, T. and Scheike, T. H. (2009). The additive hazards model with high-dimensional regressors. *Lifetime data analysis* **15,** 330–342.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society, Series B* **65,** 331–366.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. *In Proceedings of the second Seattle Symposium in Biostatistics.*

Lecture Notes in Statistics **179,** 189–326. Springer, New York.

Wand, M. P. and Jones, M. C. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation. *Journal of the American Statistical Association* **88,** 520-528.

Watkins, C. J. C. H. (1989). Learning from delayed rewards. *PhD Thesis,* Cambridge, UK.

Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning* **8,** 279-292.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50,** 1-25.

Zeng, D. and Lin, D. Y. (2014). Efficient estimation of semiparametric transformation models for two-phase cohort studies. *Journal of the American Statistical Association* **109,** 371-383.

Zhao, Y., Zeng, D., Rush, A. J. and Kosorok, M. R. (2012) Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association* **107,** 1106-1118.

Zhao, Y., Zeng, D., Laber, E., Song, R., Yuan, M. and Kosorok, M. (2015) Doubly robust learning for estimating individualized treatment with censored data. *Biometrika* **102,** 151-168.

Zhang, B., Tsiatis, A. A., Laber, E. B. and Davidian, M. (2012a). A robust method for estimating optimal treatment regimes. *Biometrics* **68,** 1010-1018.

Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M. and Laber, E. B. (2012b). Estimating optimal treatment regimes from a classification perspective. *Stat* **1,** 103-114.

Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, U.S.A.

E-mail: (skang8@ncsu.edu)

Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, U.S.A.

E-mail: (lu@stat.ncsu.edu)

Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, U.S.A.

E-mail: (jzhang@mailbox.sc.edu)