An anti-confounding method for estimating optimal regime in a survival context using instrumental variable

Junwen Xia, Zishu Zhan and Jingxiao Zhang *

School of Statistics, Renmin University of China, Beijing, China

Abstract

There is extensive literature on the estimation of the optimal individualized treatment regime in a survival context, which dictates treatment to maximize the expected survival probability. Those methods are based on the key assumption that we can collect all the confoundings in the observational studies or in the randomized trials with noncompliance. However, the assumption sometimes is too restrictive to be applied and the violation would yield bias on the estimation of the optimal regime. In the article, we propose a method to learn the optimal regime when some of the confoundings are not observed and a valid binary instrumental variable is available. Specifically, we establish the estimator for the potential survival function under any given treatment regime and for the optimal regime by maximizing the potential survival function under a prespecified class of regimes. We also propose the doubly robust estimator to avoid possibly wrong assignment of the nuisance model. Since the estimators of the potential survival function is jagged, we utilize the kernel smoothed technique to relieve the burden of the optimization. Asymptotic properties of the proposed estimators are provided, moreover, simulation results confirm the finite sample performance when unmeasured confounding exists. Our methods are also examined and illustrated by a real-world example to dictate personalized colorectal cancer screening.

Keywords: Time-to-event data, Unmeasured confounding, Instrumental variable, Optimal individualized treatment regime.

*The corresponding author. E-mail address: zhjxiao@ruc.edu.cn.
1 Introduction

The optimal regime refers to assigning different treatments to patients according to their basic characteristics in order to achieve a better value. The definition of the value function can be varied, including cure rate, $t$-year survival probability or any other clinical benefit or efficacy. There have been lots of literature on estimating the optimal regime, where Q-learning (Qian and Murphy, 2011; Laber et al., 2014; Murphy, 2005) and A-learning (Murphy, 2003; Robins, 2004) are two ubiquitous methods based on regression-type modelling. An alternative class of methods known as value search methods is popular, which is based on maximizing a consistent estimator of value function under a prespecified class of regimes. Zhang et al. (2012) proposed inverse propensity score weighted and augmented inverse propensity score weighted estimators for the normal outcome and Zhao et al. (2015a) extended the results to time-to-event data. In recent years, more developments and studies along this research line have been published, referring to Qi and Liu (2018); Zhang and Zhang (2018); Pan and Zhao (2021), Blümlein et al. (2022) for more details.

A commonly used assumption to posit model or construct value function is that of no unmeasured confounding (Jiang et al., 2017a,b; Wang et al., 2018), which indicates we can collect all the variables having causal effect to both treatment and outcome. However, the assumption is sometimes too restrictive to be realistic for the noncompliance in randomized trials (Dawid, 2003) or the impossible ability to collect all confoundings in observational studies. The violation of the assumption would necessarily yield bias in the estimation of the optimal regime and further wrong assignment of clinical treatment. To tackle the problem it brings, as mentioned in Wang and Tchetgen Tchetgen (2018); Cui and Tchetgen Tchetgen (2021b); Wang et al. (2022), one ubiquitous method in the economy and epidemiology called instrumental variable (IV) can be utilized. An instrumental variable is defined as a pretreatment variable that is independent of all unmeasured confoundings and does not have a direct causal effect on the outcome other than through the treatment. In randomized trials with noncompliance, the assignment of the treatment is a valid instrumental variable for the real treatment the participants receive. In observational studies, if they meet the Mendelian randomization study, leveraging genetic variants known to be associated with the phenotype as an instrument is commonly accepted. A well-known illustration of the approach takes 39 independent single nucleotide polymorphisms as the genetic IV to estimate the causal association between diabetes and mortality (Mailman et al., 2007; Wang et al., 2022).

Instrumental variable was introduced to construct the value function and estimate the optimal regime by Cui and Tchetgen Tchetgen (2021b) for a normal outcome. Qiu et al. (2021) extended it to the framework of multi-objective optimization and applied it to study the effect of the combat deployment on suicide among U.S. Army soldiers. Chen and Zhang (2021) considered the multiple-stage setting and proposed a time-varying in-
instrumental variable to estimate optimal regime for each stage. Some discussions on the assumption made in Cui and Tchetgen Tchetgen (2021b); Qiu et al. (2021) and Chen and Zhang (2021) can be founded in Zhang and Pu (2021) and Cui and Tchetgen (2021). When the assumptions fail to establish, alternative assumptions such as Assumption A in Sukjin and Han (2020) and assumptions in Theorem 3.1 Cui and Tchetgen Tchetgen (2021a) or identifying the lower bounds of the value function proposed in Pu and Zhang (2020) can be utilized. In clinical observational studies, time-to-event data is common and enlarging survival probability is of interest, but the above methods cannot be directly applied to the right censoring data. Thus, in the article via the aid of a valid binary instrumental variable, we advance semiparametric estimators for the potential survival function under any given treatment regime and construct the optimal regime by maximizing the estimators from the value search perspective.

The article makes a number of contributions to both the survival analysis and the precision medicine literature. Firstly, we establish the semiparametric estimators of the potential survival probability under any treatment regime and of the optimal regime by maximizing it. To provide additional protection against model mis-specification, we further establish the doubly robust version. As far as we know, our semiparametric estimator of the optimal regime and its doubly robust version are the first work to handle unmeasured confoundings in a survival context. During the estimation procedure, we solve the following difficulties: (1) the survival time is subject to right censoring and how to incorporate the censoring time into the estimation is challenging when unmeasured confoundings exist; (2) our proposed estimators of the potential survival function are non-smoothed and non-concave subjective to the parameters of treatment regimes, which yields computational challenge in the optimization. To overcome the first difficulty, we propose the so-called IV-based Kaplan-Meier estimators to incorporate the censoring time into the estimation, which puts no model constrain on the potential survival function. For the second difficulty, the kernel smoothed technique is utilized to smooth the IV-based Kaplan-Meier estimators. Secondly, we establish the asymptotic properties for the aforementioned estimators and the properties show the merit of the doubly robust estimator on both model mis-specification and convergence rate. Moreover, simulation results confirm the finite sample performance of our proposed estimators when unmeasured confounding exists.

It is worth noting that we further construct a multi-objective optimization framework to apply our methods to the cancer screening data in order to trade off the survival probability and medical consumption. Our analysis shows that, when survival probability is our only consideration, the optimal regime recommends almost all of the people to do the screening. If we want to achieve a trade-off between survival probability and medical consumption, people who are at risks (male, family history of any cancer and age between 65 to 70) with no diseases (diabetes and colorectal polyps) are the most recommended subgroup.
The remainder of the article is organized as follows. In Section 2, we present the mathematical framework and estimators for the use of instrumental variables in estimating the optimal individualized treatment regime. The doubly robust estimator and the kernel smoothed estimators are also introduced in Section 2. Asymptotic properties of the estimators and the large-sample properties of the optimal regime are provided in Section 3. Finite sample performance is studied via simulation in Section 4. In Section 5, we illustrate the application of the analysis of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer trial data set.

2 Methodology

We first introduce some general notations used throughout the article. Let \( A \in \{0,1\} \) be a binary treatment indicator, where \( A = 1 \) represents receiving treatment and \( A = 0 \) is otherwise. We denote \( U \) as an unmeasured confounding (possibly vector-value) that affects both the treatment individual receives and the outcome. A binary instrumental variable \( Z \in \{0,1\} \) is utilized to debias the unmeasured confounding effect. \( L \in L \) are the fully observed covariates and are used to customize individual treatment regime \( d_\eta(L) \), where \( L \) is a \( p \)-dimensional vector space and \( d_\eta \in \mathcal{D} \) is a mapping from covariates \( L \) to treatment \( A \) with parameters \( \eta \). In the following, we would assume \( \mathcal{D} = \{d_\eta : d_\eta(L) = I(\tilde{L}^T \eta \geq 0); ||\eta||^2 = 1\} \), where \( \tilde{L} = (1, L^T)^T \). For the limited time of the study and the refusal of the response, we cannot observe survival time \( T \) all the time, instead, censoring time \( C \) sometimes replaces \( T \). Thus, what we observe is \( \tilde{T} = \min\{T,C\} \) and the indicator of observing status \( \delta = I\{T \leq C\} \). We assume that for the observed data, \( \{(U_i, L_i, A_i, Z_i, \tilde{T}_i, \delta_i), i = 1, \ldots, n\} \) is independent and identically distributed across \( i \).

We also define the counting process \( N_t(t) = I\{\tilde{T}_i \leq t, \delta_i = 1\} \) and the at-risk process \( Y_t(t) = I\{\tilde{T}_i \geq t\} \). For better clarification, we unify potential results with a superscript star. Accordingly, we define potential survival time \( T^*(A) \), potential counting process \( N^*(A; t) \) and the potential at-risk process \( Y^*(A; t) \) under treatment \( A \), where \( N^*(A; t) = I[\min \{T^*(A), C\} \leq t, T^*(A) \leq C] \) and \( Y^*(A; t) = I[\min \{T^*(A), C\} \geq t] \). Under regime \( d_\eta \), the corresponding potential survival time is \( T^*(d_\eta(L)) = T^*(1)I\{d_\eta(L) = 1\} + T^*(0)I\{d_\eta(L) = 0\} \), as well as the potential counting process \( N^*(d_\eta(L); t) = N^*(1; t)I\{d_\eta(L) = 1\} + N^*(0; t)I\{d_\eta(L) = 0\} \) and potential at-risk process \( Y^*(d_\eta(L); t) = Y^*(1; t)I\{d_\eta(L) = 1\} + Y^*(0; t)I\{d_\eta(L) = 0\} \).

We then introduce some notations for the development of asymptotic properties. For two random variable sequences \( \{X_n\}_{n=1}^\infty \) and \( \{Y_n\}_{n=1}^\infty \), notations \( X_n \preceq_p Y_n \) or \( X_n = O_p(Y_n) \) mean that for any \( \epsilon > 0 \), there exists a positive constant \( M \) such that \( \sup_n P(\{|X_n/Y_n| > M\} \leq \epsilon. And X_n = o_p(Y_n) \) means \( |X_n/Y_n| \to^p 0 \) as \( n \to \infty \) or in other words, \( |X_n/Y_n| \) converges to 0 in probability. \( \mathbb{P}_n \) means the expectation of random variable \( X \) and \( \mathbb{P}_n X \) means the
sample mean of $X$ with $n$ samples. We also denote $\|X\| = \sqrt{\mathbb{E}X^2}$ as $L^2(\mathbb{P})$-norm.

For classic time-to-event data, survival probability $S(t) = P(T > t)$ is of the interest, while in the context of the precision medicine, the potential survival probability $S^*(t; \eta) = P(T^*(d_\eta(L)) > t)$ under regime $d_\eta(L)$ is primarily interested, where $T^*(d_\eta(L))$ means potential survival time under regime $d_\eta$ with covariates $L$. For any predetermined $t$, $S^*(t; \eta)$ provides a basis for the definition of an optimal regime as

$$d_\eta^{opt} = \arg \max_D S^*(t; \eta) = I(L^T \eta^{opt} \geq 0),$$

where

$$\eta^{opt} = \arg \max_{||\eta||^2 = 1} S^*(t; \eta).$$

In the article, we define optimality of a regime with respect to the $t$-year potential survival probability, though, other estimators, such as restricted mean (Geng et al., 2015) and the quantile (Zhou et al., 2022) of the potential survival probability, could also be used. We also suppose the predetermined $t$ satisfies $0 < S^*(t; \eta) < 1$ through the article.

### 2.1 Estimator of the potential survival probability with unmeasured confoundings

To find an optimal regime, the consistent estimator of $S^*(t; \eta)$ for a predetermined $t$ is required. We make the uninformative censoring assumption as follows:

**Assumption 1:** $\{T^*(1), T^*(0)\} \perp C|Z, L, A$.

Let $S_C\{t|Z, L, A\} = P(C \geq t|Z, L, A)$ denote the survival function of the censoring time given $Z, L, A$. When the $d_\eta$-specific potential counting process $N_i^\ast(d_\eta(L_i); s)$ and the at-risk process $Y_i^\ast(d_\eta(L_i); s)$ for all $0 \leq s \leq t$ are observed, an intuitive estimator for $S^*(t; \eta)$ is the inverse probability of censoring weighted Kaplan-Meier estimator as illustrated in (Jiang et al., 2017a).

$$\hat{S}^*(t; \eta) = \prod_{s \leq t} \left\{ 1 - \frac{\sum_{i=1}^n dN_i^\ast(d_\eta(L_i); s) / S_C\{s|Z_i, L_i, d_\eta(L_i)\}}{\sum_{i=1}^n Y_i^\ast(d_\eta(L_i); s) / S_C\{s|Z_i, L_i, d_\eta(L_i)\}} \right\}$$

A key estimand in estimating $S^*(t; \eta)$ is the conditional survival function of the censoring time $S_C\{s|Z, L, A\}$. There are several approaches for estimating it such as the non-parametric estimator, the local Kaplan-Meier (Gonzalez-Manteiga and Cadarso-Suarez, 1994) and random forests (Lee et al., 2020), or a semi-parametric estimator using a working model, the Cox proportional hazard regression model (Zhao et al., 2015b) and additive proportional hazard regression model (Lee et al., 2020). In the clinical study when ad-
ministrative censoring occurs, it is often reasonable to assume

**Assumption 1**: \( C \perp \{T^*(0), T^*(1), Z, L, A\} \).

For simplicity as Jiang et al. (2017a) and Zhou et al. (2022), we would utilize Assumption 1 replacing Assumption 1 in the following derivation and theoretical development. Thus, \( S_C\{t|Z, L, A\} \) can be eliminated in the estimator \( \hat{S}^*(t; \eta) \) as follows:

\[
\hat{S}^*(t; \eta) = \prod_{s \leq t} \left\{ 1 - \frac{\sum_{i=1}^n dN_i^*(d\eta(L_i); s)}{\sum_{i=1}^n Y_i^*(d\eta(L_i); s)} \right\}.
\] (4)

Since \( N_i^*(d\eta(L_i); s) \) and \( Y_i^*(d\eta(L_i); s) \) for both \( d\eta(L_i) = 0, 1 \) are impossibly observed at the same time in the clinical study, \( \hat{S}^*(t; \eta) \) is not computable on the basis of the observed data. To obtain proper estimators that are computable from the observed data, tremendous literature are based on the assumption of no unmeasured confounding (Jiang et al., 2017a,b; Wang et al., 2018).

**Assumption 2**: (No unmeasured confounding). \( T^*(a) \perp A|L \) for \( a = 0, 1 \).

However, Assumption 2 sometimes is too restrictive to be applied for the noncompliance in randomized trials (Dawid, 2003) or the impossible ability to collect all confoundings in the observational studies. The violation of the assumption would necessarily yield bias on the estimation of optimal regime and further wrong assignment of clinical treatment. Let \( T^*(z,a) \) be the potential survival time with the instrument \( z \) and treatment \( a \). To tackle the problem it brings, an ubiquitous method in the economy and epidemiology called instrumental variable (Wang and Tchetgen Tchetgen, 2018; Cui and Tchetgen Tchetgen, 2021b; Wang et al., 2022) can be utilized to construct an estimator of \( S^*(t; \eta) \).

**Assumption 3**: (IV relevance). \( Z \not\perp A|L \).

**Assumption 4**: (Exclusion restriction). \( T^*(z,a) = T^*(a) \) for \( z, a = \pm 1 \) almost surely.

**Assumption 5**: (IV independence). \( Z \perp U|L \).

**Assumption 6**: (IV positivity). \( 0 < f(Z = 1|L) < 1 \) almost surely.

**Assumption 7** (Independent compliance type).

\[
\delta(L) \equiv f(A = 1|Z = 1, L) - f(A = 1|Z = 0, L) = \tilde{\delta}(L, U) \text{ almost surely.}
\]

Assumptions 3-5 are the core assumptions for all IV-based methods, while assumptions 6 and 7 support a nonparametric estimator of the potential survival function. Assumption 3 requires that the instrumental variable is associated with the treatment conditional on \( L \). Assumption 4 states that the causal effect from \( Z \) to \( T \) is all mediated by treatment
A and changes in Z only cannot affect T. Assumption 5 ensures that the causal effect of Z on T is unconfounded given L. The above three assumptions are often satisfied in randomized trials or mendelian randomized studies. Assumption 7 is no more restrictive than the assumption positing linear model for f(A|Z, L, U) as discussed in Cui and Tchetgen Tchetgen (2021b) and Wang and Tchetgen Tchetgen (2018). With these assumptions, a nonparametric estimator can be proposed as follows:

$$\hat{S}^*_t(t; \eta) = \prod_{s \leq t} \left\{ 1 - \frac{\sum_{i=1}^{n} \hat{w}_i dN_i(s)}{\sum_{i=1}^{n} \hat{w}_i Y_i(s)} \right\},$$

(5)

where

$$\hat{w}_i = \frac{(2Z_i - 1)(2A_i - 1)I\{A_i = d\eta(L_i)}{\hat{\delta}(L_i) \hat{f}(Z_i|L_i)},$$

Here, $$\hat{\delta}(L_i) = \hat{f}(A = 1|Z = 1, L_i) - \hat{f}(A = 1|Z = 0, L_i)$$ and $$\hat{f}(Z_i|L_i)$$ is an estimate of $$\delta(L_i)$$ and $$f(Z_i|L_i)$$. In general, we might posit a parametric model, like logistic regression, to estimate $$f(A|Z_i, L_i)$$ and $$f(Z_i|L_i)$$ separately.

### 2.2 Doubly robust estimator of the potential survival probability

As the instrumental variable method is inconsistent when the model for $$f(Z|L)$$ is mis-specified, we prefer an estimator with the doubly robust property by incorporating assumed model information. For example, we can assume a cox proportional hazard model for the conditional cumulative hazard function of T by

$$\Lambda_T(t|Z, L, A) = \Lambda_0(t) \exp \left\{ \left( Z, L^T, A, AL^T \right) \beta \right\},$$

where $$\Lambda_0(t)$$ is the baseline cumulative hazard function, and $$\beta$$ is a $$(2p + 2)$$-dimensional parameter vector. In order to find a doubly robust estimator for the potential survival function $$S^*(t; \eta)$$, we first think of the doubly robust estimators for $$E[dN^*(d\eta(L); s)]$$ and $$E[Y^*(d\eta(L); s)]$$, then incorporate them into the estimator 5 to achieve the doubly robustness for the potential survival function.

A commonly used method to derive a doubly robust estimator is based on the efficient influence function which provides an optimal estimator, with knowledge on asymptotic behavior, and promises both flexibility and efficiency (Lee et al., 2020). The derivation of the efficient influence function is often challenging and sometimes regarded as a dark art, but a relative intuitive and simple efficient influence function can be derived by reformulating the target estimate as a series of conditional expectation and applying the chain derivation rule (Tchetgen Tchetgen et al., 2015; Wang and Tchetgen Tchetgen, 2018; Tsiatis and A, 2006). We propose the efficient-influence-function-based estimators for $$E[dN^*(d\eta(L); s)]$$
and $E[Y^*(d_{\eta}(L); s)]$ as $n^{-1} \sum_{i=1}^{n} \hat{I}_E[d_{\eta}(d_{\eta}(L_i); s)]$ and $n^{-1} \sum_{i=1}^{n} \hat{I}_E[Y^*(d_{\eta}(L_i); s)]$, where

$$\hat{I}_E[d_{\eta}(d_{\eta}(L_i); s)] = \frac{2Z_i - 1}{\hat{\delta}(L_i)} \{(2A_i - 1)dN_i(s)I\{A_i = d_{\eta}(L_i)\}$$

$$- \gamma_1'(L_i; s) - (A_i - \hat{E}(A|Z = 0, L_i))\gamma_1(L_i; s)\} + \hat{\gamma}_1(L_i; s),$$

and

$$\hat{I}_E[Y^*(d_{\eta}(L_i); s)] = \frac{2Z_i - 1}{\hat{\delta}(L_i)} \{(2A_i - 1)Y_i(s)I\{A_i = d_{\eta}(L_i)\}$$

$$- \gamma_2'(L_i; s) - (A_i - \hat{E}(A|Z = 0, L_i))\gamma_2(L_i; s)\} + \hat{\gamma}_2(L_i; s),$$

where

$$\hat{\gamma}_1(L_i; s) = \sum_z \frac{(2z - 1)(2d_{\eta}(L_i) - 1)\hat{f}(A = d_{\eta}(L_i)|Z = z, L_i)}{\hat{\delta}(L_i)}$$

$$\times \hat{S}_T \{s|z, L_i, d_{\eta}(L_i)\} \hat{S}_C(s) d\hat{\Lambda}_T \{s|z, L_i, d_{\eta}(L_i)\},$$

$$\hat{\gamma}_1'(L_i; s) = (2d_{\eta}(L_i) - 1)\hat{f}(A = d_{\eta}(L_i)|Z = 0, L_i)$$

$$\times \hat{S}_T \{s|Z = 0, L_i, d_{\eta}(L_i)\} \hat{S}_C(s) d\hat{\Lambda}_T \{s|Z = 0, L_i, d_{\eta}(L_i)\},$$

$$\hat{\gamma}_2(L_i; s) = \sum_z \frac{(2z - 1)(2d_{\eta}(L_i) - 1)\hat{f}(A = d_{\eta}(L_i)|z, L_i)}{\hat{\delta}(L_i)}$$

$$\times \hat{S}_T \{s|z, L_i, d_{\eta}(L_i)\} \hat{S}_C(s),$$

and

$$\hat{\gamma}_2'(L_i; s) = (2d_{\eta}(L_i) - 1)\hat{f}(A = d_{\eta}(L_i)|Z = 0, L_i)$$

$$\times \hat{S}_T \{s|Z = 0, L_i, d_{\eta}(L_i)\} \hat{S}_C(s).$$

The survival function of the censoring time $S_C(s)$ can be estimated by Kaplan-Meier estimator as the survival function. $\hat{S}_T \{s|Z, L, A\}$ can be calculated via the formula $\hat{S}_T \{s|Z, L, A\} = \exp\{-\hat{\Lambda}_T(s|Z, L, A)\}$. Then the doubly robust estimator for potential
survival function $S^*(t; \eta)$ is as follows:

$$\hat{S}^*_{D}(t; \eta) = \prod_{s \leq t} \left\{ 1 - \frac{\sum_{i=1}^{n} \hat{F}E[dN^*_i(d_{\eta}(L_i); s)]}{\sum_{i=1}^{n} \hat{F}E[Y^*_i(d_{\eta}(L_i); s)]} \right\},$$

which is consistent when one of the following models is correct.

$\mathcal{M}_1'$: models for $f(Z|L)$ and $f(A|Z, L)$ are correct;

$\mathcal{M}_2'$: models for $\Lambda_T(s|Z, L, A)$ and $f(A|Z, L)$ are correct.

### 2.3 Computational aspect

$\hat{S}^*_I(t; \eta)$ and $\hat{S}^*_D(t; \eta)$ are non-smoothed functions of $\eta$. To illustrate the point, we plot $\hat{S}^*_I(t; \eta)$ and $\hat{S}^*_D(t; \eta)$ as functions of $\eta_1$ in Figure 1 for an example from the simulation setting (a) in Example 4.1 with censoring rate 15% and the optimal value for $\eta_1$ is 0.707.

The estimates are very jagged, thus a direct maximization based on the gradient is impossible for the gradient explosion, while other heuristic algorithms such as Nelder-Mead are highly consumptive.

![Figure 1](image-url)

Figure 1: Plots for the original (solid line) and smoothed (dashed line) estimates: (a) $\hat{S}^*_I(t; \eta)$; (b) $\hat{S}^*_D(t; \eta)$

To relieve the burden of the optimization and save computational resources, we propose to smooth the estimators $\hat{S}^*_I(t; \eta)$ and $\hat{S}^*_D(t; \eta)$ by using kernel smoothers. Specifically, we replace $d_{\eta}(L) = I(L^T \eta \geq 0)$ in $\hat{S}^*_I(t; \eta)$ and $\hat{S}^*_D(t; \eta)$ with $\tilde{d}_{\eta}(L) = \Phi(L^T \eta / h)$ to obtain the smoothed version $\hat{S}^*_{SI}(t; \eta)$ and $\hat{S}^*_{SD}(t; \eta)$, where $\Phi(s)$ is the cumulative distribution function for the standard normal distribution, and $h$ is a bandwidth parameter that goes to 0 as $n \to \infty$. It should be noticed that the part with $d_{\eta}(L_i)$ in $\hat{S}^*_I(t; \eta)$ and $\hat{S}^*_D(t; \eta)$
can be rewrite to \( I\{d_\eta(L_i) = a\} \) and the replacement of \( I\{d_\eta(L_i) = a\} \) is as follows:

\[
I\{d_\eta(L_i) = a\} = a d_\eta(L_i) + (1 - a)(1 - d_\eta(L_i)) = a \hat{d}_\eta(L_i) + (1 - a)(1 - \hat{d}_\eta(L_i)).
\]

For example, \( \hat{f}(A = d_\eta(L_i)|Z = z, L_i) \) in \( \hat{S}^*_D(t; \eta) \) can be rewritten and replaced as follows:

\[
\hat{f}(A = d_\eta(L_i)|Z = z, L_i) = \sum_a \hat{f}(A = a|Z = z, L_i) I\{d_\eta(L_i) = a\}
\]

\[= \sum_a \hat{f}(A = a|Z = z, L_i) \{a \hat{d}_\eta(L_i) + (1 - a)(1 - \hat{d}_\eta(L_i))\}.\]

For bandwidth selection, we set \( h = c_0 n^{-1/3} \text{sd}(L^T \eta) \), where \( c_0 \) is a constant and \( \text{sd}(X) \) is the sample standard deviation of \( X \). In the following simulations and application, we choose \( c_0 = 4^{1/3} \) as Jiang et al. (2017a) suggested. We also plot in Figure 1 the smoothed estimates and it shows that the smoothed estimates approximate the original estimates well and are not jagged, which indicates the gradient would not be too large to make the optimization method based on the gradient fail to work.

As the estimates can still be non-concave, in order to evaluate the performance of the estimators, we choose to utilize the genetic algorithm in R package \texttt{rgenoud} (Mebane and Sekhon, 2011). The genetic algorithm (a type of evolutionary algorithm) is inspired by the process of natural selection. Evolution usually starts from the population of randomly generated individuals and the optimization process is an iterative process. The population in each iteration is called a generation. In each generation, the fitness of each individual in the population is evaluated; The fitness is usually the value of the objective function in the optimization problem left to be solved. More suitable individuals are randomly selected from the current population, and the genome of each individual is modified (mutation and crossover) to form a new generation. Then a new generation of candidate solutions is used in the next iteration of the algorithm. We refer Mitchell (1998) to more details and references.

3 Asymptotic properties

The next two theorems address the large-sample behavior of our two proposed estimators. First, consider the following conditions.

**Condition 1**: The nuisance models for each estimator are in the Donsker class.

**Condition 2**: For some constant \( \epsilon > 0 \), \( P \left( \epsilon < \hat{\delta}(L) < \infty \right) = 1 \) and \( P \left( \epsilon < \hat{f}(z|L) < \infty \right) = 1 \) for all \( z \in \{0, 1\} \).

**Condition 2**: For some constant \( \epsilon > 0 \), \( P \left( \epsilon < \hat{\delta}(L) < \infty \right) = 1 \) and \( P \left( \epsilon < \hat{f}(z|L) < \infty \right) = 1 \) for all \( z \in \{0, 1\} \).
Under assumption (Kennedy, 2016).

and Cox proportional hazard model, which only requires certain smoothness or boundedness (Kennedy, 2016).

Donsker class includes usual parametric classes such as the linear model and the generalized linear model and infinite-dimensional classes such as the Kaplan-Meier estimator and Cox proportional hazard model, which only requires certain smoothness or boundedness (Kennedy, 2016).

**Theorem 3.1** Under assumption A1∗, A3-A7 and C1-C2, we have the following asymptotic properties of \( \hat{S}_t^*(t; \eta) \) for any given \( t \) and \( \eta \).

\[
\hat{S}_t^*(t; \eta) - S^*(t; \eta) = O_p\left\{ \sum_{z} ||\hat{\delta}(L)\hat{f}(z|L) - \delta(L)f(z|L)|| \right\} + S^*(t; \eta) \cdot \mathbb{P}_n \int_0^t dN^*(d\eta(L); s) - Y^*(d\eta(L); s) d\Lambda^*(s; \eta) + o_p(n^{-1/2}).
\]

where the term with \( \mathbb{P}_n \) is asymptotic normal with mean 0.

When model for \( \hat{\delta}(L) \) and \( \hat{f}(Z|L) \) both achieves \( n^{1/2} \) convergence rate such that

\[
||\hat{\delta}(L)\hat{f}(z|L) - \delta(L)f(z|L)|| = O(n^{-1/2}),
\]

which is often satisfied for a parametric model under mild condition, the estimator \( \hat{S}_t^*(t; \eta) \) achieves the \( n^{1/2} \) convergence rate and

\[
\sqrt{n}(\hat{S}_t^*(t; \eta) - S^*(t; \eta)) \text{ converges to a mean-zero normal distribution by CLT.}
\]

If we utilize nonparametric model to regress \( \delta(L) \) and \( f(Z|L) \) such as the kernel regression and random forests, the convergence rate is less than \( n^{1/2} \), say \( k \), and alternatively, \( \sqrt{k}(\hat{S}_t^*(t; \eta) - S^*(t; \eta)) \) converges to a mean-zero normal distribution as \( k \to \infty \). Moreover, both parametric model and nonparametric model consistent to \( \delta(L) \) and \( f(Z|L) \) can guarantee the consistency of \( \hat{S}_t^*(t; \eta) \).

For estimator \( \hat{S}_{t}^*(t; \eta) \), we also provide its asymptotic property and show the doubly robustness.

**Theorem 3.2** Under assumption A1∗, A3-7, C1 and C2∗, we have the following asymptotic properties of \( \hat{S}_{t}^*(t; \eta) \) for any given \( t \) and \( \eta \).

\[
\hat{S}_{t}^*(t; \eta) - S^*(t; \eta) = O_p\left\{ \sup_{0 \leq s \leq t} \sum_{z} ||\delta(L)f(z|L) - \hat{\delta}(L)f(z|L)|| \cdot ||E(A|z, L) - \hat{E}(A|z, L)|| \right\} + ||f(A = d\eta(L)|z, L)S_{T,s,z}SC,s - \hat{f}(A = d\eta(L)|z, L)\hat{S}_{T,s,z}\hat{SC,s}|| + ||f(A = d\eta(L)|z, L)S_{T,s,z}SC,s\lambda_{T,s,z} - \hat{f}(A = d\eta(L)|z, L)\hat{S}_{T,s,z}\hat{SC,s}\hat{\lambda}_{T,s,z}|| \right\} + S^*(t; \eta) \cdot \mathbb{P}_n \int_0^t dN^*(d\eta(L); s) - Y^*(d\eta(L); s) d\Lambda^*(s; \eta) + o_p(n^{-1/2}),
\]
where \( S_{T,s,z} \) is the shorthand of \( S_T\{s|z,L,d_\eta(L)\} = \exp\{-\Lambda_T(s|z,L,d_\eta(L))\} \), \( S_{C,s} = S_C(s) \) and \( \lambda_{T,s,z} = \lambda_T\{s|z,L,d_\eta(L)\} = \Lambda_T\{s|z,L,d_\eta(L)\} \). Moreover, \( \Lambda^*(s;\eta) \) is the potential cumulative hazard function under the regime \( d_\eta \). The term with \( \mathbb{P}_n \) is asymptotic normal with mean zero.

We assume that \( \hat{S}_C(s) \) is the Kaplan-Meier estimator for \( S_C(s) \) which is a nonparametric estimator and the uniform convergence rate is \( \sup_{0 \leq s \leq 1} \|\hat{S}_C(s) - S_C(s)\| = O(\sqrt{\log n/n}) \) under a mild condition (Csörg˝O and Horváth, 1983). If we want to achieve a \( \sqrt{n} \)-convergence rate, for any \( k_1 \cdot k_2 = \sqrt{n} \) where \( k_1 \geq \sqrt{\log n} \), \( k_1 \)-convergence rate on \( \mathcal{M}'_1 \) and uniform \( k_2 \)-convergence rate on \( \mathcal{M}'_2 \) is sufficient, which indicates that the convergence rate of \( \hat{S}_D^*(t;\eta) \) is no worse than \( \hat{S}_I^*(t;\eta) \) and semi-parametric model or nonparametric model to posit \( \mathcal{M}'_1 \) or \( \mathcal{M}'_2 \) is possibly to achieve \( \sqrt{n} \)-convergence rate as well. In addition to that, the above theorem also implies double robustness between model \( \mathcal{M}'_1 \) and \( \mathcal{M}'_2 \), thus when one of the model \( \mathcal{M}'_1 \) or \( \mathcal{M}'_2 \) is mis-specified, a consistent estimator on the other model is sufficient to make \( \hat{S}_D^*(t;\eta) \) consistent.

The following two theorems show the large-sample properties for \( \hat{\eta}_{opt} \) including the non-smoothed and smoothed version. We show the conditions first.

**Condition 3:** \( \sum_z \|\hat{\delta}(L)f(z|L - \delta(L)f(z|L))\| = o(1) \)

**Condition 3*:**

\[
\sup_{0 \leq s \leq 1} \sum_z \|\delta(L)f(z|L) - \hat{\delta}(L)f(z|L)\| \cdot \{\|E(A|z,L) - \hat{E}(A|z,L)\|
+\|f(A = d_\eta(L)|z,L)S_{T,s,z}S_{C,s} - \hat{f}(A = d_\eta(L)|z,L)\hat{S}_{T,s,z}\hat{S}_{C,s}\|
+\|f(A = d_\eta(L)|z,L)S_{T,s,z}\Lambda_{T,s,z} - \hat{f}(A = d_\eta(L)|z,L)\hat{S}_{T,s,z}\hat{\Lambda}_{T,s,z}\|\}
=o(1)
\]

Condition 3* is satisfied as long as condition 3 is satisfied and they are satisfied with the properly specified models or with the non-parametric models.

**Theorem 3.3** Under assumption A1*, A3-7 and C1-C3, we have the following asymptotic properties.

1. \( \sup_{|\eta|^2 = 1} \|\hat{S}_I^*(t;\eta) - \hat{S}_I^*(t;\eta)\| = o_p(1) \) for any given \( t \)
2. \( \hat{\eta}_{Iopt} \rightarrow_p \eta_{opt} \)

If we further have C4-5 in the supplementary material.

3. \( \sup_{|\eta|^2 = 1} \sqrt{n} \|\hat{S}_I^*(t;\eta) - \hat{S}_I^*(t;\eta)\| = o_p(1) \) for any given \( t \)
4. \( \hat{\eta}_{Iopt}^* \rightarrow_p \eta_{opt} \)

Here \( \hat{\eta}_{opt}^* \) is the estimate of \( \eta_{opt} \) by maximizing \( \hat{S}_I^*(t;\eta) \).

**Theorem 3.4** Under assumption A1*, A3-7, C1, C2* and C3*, we have the following asymptotic properties.
(1) : sup\(|\eta|^{2} = 1 (\hat{S}_{D}^{*}(t; \eta) - \hat{S}^{*}(t; \eta)) = o_{p}(1) \text{ for any given } t
(2) : \hat{\eta}_{D}^{opt} \rightarrow p \eta^{opt}
If we further have C4-5 in the supplementary material.
(3) : sup\(|\eta|^{2} = 1 \sqrt{n}(\hat{S}_{D}^{*}(t; \eta) - \hat{S}_{SD}^{*}(t; \eta)) = o_{p}(1) \text{ for any given } t
(4) : \hat{\eta}_{SD}^{opt} \rightarrow p \eta^{opt}
Here \( \hat{\eta}^{opt} \) is the estimate of \( \eta^{opt} \) by maximizing \( \hat{S}^{*}(t, \eta) \).

4 Simulation

We report simulation results for two different examples. In the first example, we simulate the scenario of the randomized trail with noncompliance; in the second example, we simulate the scenario of the observational study with unmeasured confounding. The major difference between these two settings is the different generation of the instrumental variable \( Z \), which would be illustrated detailedly as follows.

**Example 4.1** Noncompliance in randomized trails indicates that, though individuals are randomly assigned to treatment \( Z \), there exist individuals who self-select his or her treatment \( A \), thereby disturbing the covariate balance. In such scenario, we can choose randomized assignment \( Z \) as an instrumental variable and it is a Bernoulli event with probability \( \frac{1}{2} \). For the reason that \( Z \) is independent of all other variables, the model mis-specification problem for \( f(Z|L) \) does not exist.

We simulate the example with an instrumental variable \( Z \) generated from a Bernoulli event with probability \( \frac{1}{2} \). Baseline variables \( L \) is from a uniform distribution on \([-2, 2]^{2} \); \( U \) is from a bridge distribution with parameter \( \phi = \frac{1}{2} \) and treatment \( A \) was generated under a logistic regression with success probability (Cui and Tchetgen Tchetgen, 2021b),

\[
f(A = 1|Z, L, U) = \exp\{−2.5 + L_{1} + 5Z − 0.5U\}.
\]

Via the theorem in Wang and Louis (2003), the above data generating mechanism ensures that there exists a vector \( \alpha \) such that logit\{Pr\((A = 1|Z, L)\)\} = \((1, Z, L^{T})\alpha\), so that upon marginalizing over \( U \) the model for \( f(A|Z, L) \) remains a logistic regression and thus, the Assumption 7 is valid. When Assumption 7 fails to establish, we present some simulations in the supplementary material S2.4. Surprisingly, from the perspective of both parameters’ estimation and Misclassification Rate (MR), violation of Assumption 7 seems to have a little adverse effect on them, which suggests there might be alternative assumptions guaranteeing the estimation of parameters consistent as Assumption 7 in Cui and Tchetgen Tchetgen (2021b) and Assumption A in Sukjin and Han (2020). But the violation indeed enlarges the bias of the estimation of potential survival function as the
supplementary material suggests. How the strength of the instrument affects the variance of the estimated survival functions is also simulated in the supplementary materials S2.2.

The survival time $T$ is generated from a linear transformation model by adding an unmeasured confounding term $U$ (Jiang et al., 2017a). We vary the coefficient of $U$ to simulate the strength of the unmeasured confounding.

(a) $h(T) = -0.5L_1 + A(L_1 - L_2) + 0.5U + \varepsilon_1,$

(b) $h(T) = -0.5L_1 + A(L_1 - L_2) + U + \varepsilon_1,$

(c) $h(T) = -0.5L_1 + A(L_1 - L_2) + 0.5U + \varepsilon_2,$

(d) $h(T) = -0.5L_1 + A(L_1 - L_2) + U + \varepsilon_2,$

where $h(s) = \log\{\exp(s) - 1\} - 2$ is an increasing function. The error term $\varepsilon_1$ follows the extreme value distribution and $\varepsilon_2$ follows the logistic distribution, which corresponds to a proportional hazard model and a proportional odds model, separately. The covariate-independent censoring time $C$ is uniformly distributed on $[0, C_0]$, where $C_0$ is chosen to achieve a censoring rate of 15% and 30%. The optimal regime is $d^{opt}(L_1, L_2) = I\{L_1 - L_2 \geq 0\}$ and we search the optimal regime in the class of regimes given by $\mathcal{D} = \{d_\eta: d_\eta(L_1, L_2) = I\{\eta_0 + \eta_1L_1 + \eta_2L_2 \geq 0\}, \eta = (\eta_0, \eta_1, \eta_2)^T, ||\eta||^2 = 1\}$ by maximizing the survival probability $\hat{S}^*(t; \eta)$ at $t = 2$, which contains the true optimal regime with $\eta^{opt} = (0, 0.707, -0.707)$.

Our proposed methods are implemented with $\hat{S}(L) = \hat{f}(A = 1|Z = 1, L) - \hat{f}(A = 1|Z = 0, L)$ and $\hat{f}(Z|L)$, where $\hat{f}(A|L, Z)$ and $\hat{f}(Z|L)$ were estimated from logistic regression models. For doubly robust estimator, we posit a proportional hazard model to estimate $\Lambda_T\{s|Z, L, A\}$ and calculate $\hat{S}_T\{s|Z, L, A\}$ by $\hat{S}_T\{s|Z, L, A\} = \exp\{-\hat{\Lambda}_T(s|Z, L, A)\}$. The survival function of the censoring time $S_C(t)$ is nonparametrically estimated by Kaplan-Meier estimator. Because we posit a proportional hazard model for all four settings, in setting (c) and (d), the regression model of $\Lambda_T\{s|Z, L, A\}$ is mis-specified, thus the simulation result in setting (c) and (d) can test the double robustness of $\hat{S}^*_{SD}(t, \eta)$.

With sample size 500, we repeat the simulation 500 times, and a large independent test set with 10000 subjects is used to evaluate the performance. Additional simulation results with sample sizes 250 and 1000 are shown in the supplementary materials S2.3.

We present the results in Table 1 and denote MR as the mean of the misclassification rate by comparing the true and estimated optimal regimes. Only the smoothed versions, $\hat{S}^*_{SI}(t; \eta)$ and $\hat{S}^*_{SD}(t; \eta)$, are exhibited and the comparison between smoothed version and non-smoothed version is placed in the supplementary material S2.1. The conclusion of the simulation in supplementary material agrees with the argument that smoothed version is not much different from the non-smoothed version and even a little better than the non-smoothed version. We also present other estimators, “SIPS” and “SAIPS” in Jiang et al. (2017a), with no consideration of unmeasured confoundings to compare the performance with our proposed methods. Here, “SIPS” means the smoothed version of the inverse
propensity score method and “SAIPS” means the smoothed version of the augmented inverse propensity score method. We sometimes call these two methods together as the inverse propensity score methods for simplicity.

From Table 1, the methods we propose share a lower misclassification rate and a more accurate estimation on parameters compared to the inverse propensity score methods. Especially, the doubly robust estimator \( \hat{S}^{*}_{SD}(t; \eta) \) achieves the smallest MR in all settings. In settings (c) and (d), though, the regression model for \( \Lambda_T\{s|Z, L, A\} \) is mis-specified, the doubly robust estimator is consistent as expected. With the unmeasured confounding becoming stronger, the inverse propensity score methods deteriorate fast while our proposed methods do not. It should be notice that, though, the methods we propose share a higher standard deviation on parameters’ estimation, the standard deviation of our final target, misclassification rate, is almost the same as the inverse propensity score methods.

**Example 4.2** In Example 4.1, we assume the instrumental variable \( Z \) is generated from a Bernoulli event with probability \( \frac{1}{2} \), however, it is always not satisfied in the observational study. In the contrast, \( Z \) is dependent on some baseline covariates \( L \), thus, we assume \( Z \) is generated as follows:

\[
f(Z = 1|L) = \text{expit}\{L_1 + L_2\},
\]

where \( L \) is uniformly generated from \([-2, 2]^2\).

The survival time \( T \) is generated as the Example 4.1 as follows. Compared to Example 4.1, we choose only two settings out of the four to evaluate the performance, where setting (a) is corresponding to the proportional hazard model and setting (c) is corresponding to the proportional odds model. \( C \) is generated uniformly on \([0, C_0]\), where \( C_0 \) is chosen with 15% censoring rate.

(a) \( h(T) = -0.5L_1 + A(L_1 - L_2) + 0.5U + \varepsilon_1 \)

(c) \( h(T) = -0.5L_1 + A(L_1 - L_2) + 0.5U + \varepsilon_2 \)

The estimation procedures are the same as the Example 4.1 unless the estimation of \( f(Z|L) \). We propose both correctly specified model, \( f(Z|L) = \text{expit}\{\theta_0 + \theta_1L_1 + \theta_2L_2\} \), and mis-specified model, \( f(Z|L) = \text{expit}\{\theta_0\} \), to estimate it and test the doubly robustness of \( \hat{S}^{*}_{SD}(t; \eta) \). With sample size 500, we repeat the simulation 500 times, and a large independent test set with 10000 subjects is used to evaluate the performance.

We present the results in Table 2. When the posited model for \( f(Z|L) \) is mis-specified, the instrumental variable method by maximizing \( \hat{S}^{*}_{SI}(t; \eta) \) generally has a relatively large bias as expected, whereas the doubly robust instrumental variable method greatly reduces the bias and gives a much smaller MR. In particular, the model for \( \Lambda_T\{s|Z, L, A\} \) is mis-specified under the logistic error distribution in setting (c), but it still gives a small bias with low MR, which might owe to the good prediction ability of the Cox proportional hazard model. We also notice that the doubly robust estimator with the mis-specified
Table 1: Simulation results about the absolute bias (standard deviation of the estimates given in the parentheses) of $\hat{\eta}^{opt}$ relative to $\eta^{opt}$ and of $\hat{S}^* (t; \hat{\eta}^{opt})$ relative to $S^* (t; \eta^{opt})$. “SIVE” means the method by maximizing the smoothed instrumental variable estimation $\hat{S}^*_I (t; \eta)$; “SIVE-DR” means the method by maximizing the smoothed doubly robust instrumental variable estimation $\hat{S}^*_{SD} (t; \eta)$; “SIPS” means the smoothed inverse propensity score method and “SAIPS” means the augmented inverse propensity score method.
Table 2: Simulation results about the absolute bias (standard deviation of the estimates given in the parentheses) of \( \hat{\eta}_{\text{opt}} \) relative to \( \eta_{\text{opt}} \) and of \( \hat{S}^*_t(t; \hat{\eta}_{\text{opt}}) \) relative to \( S^*_t(t; \eta_{\text{opt}}) \). The column of the PS is valued whether T (True) or F (False), which indicates posited model for \( f(Z|L) \) is correct or mis-specified. “SIVE” means the method by maximizing the smoothed instrumental variable estimation \( \hat{S}^*_t(t; \eta) \); “SIVE-DR” means the method by maximizing the smoothed doubly robust instrumental variable estimation \( \hat{S}^*_{SD}(t; \eta) \). “SIPS” means the smoothed inverse propensity score method and “SAIPS” means the augmented inverse propensity score method.

5 Application to the cancer screening

The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial (Team et al., 2000) is a two-armed randomized trial of screening tests for prostate, lung, colorectal and ovarian cancers. Ten centers across the U.S. recruited approximately 155,000 participants between November 1993 and July 2001. The data were collected up to December 31, 2009, and the mortality data was collected through 2015. We focus our attention on the effect of colorectal cancer screening on the mortality caused by colorectal cancer and construct the optimal regime to assign the screening, where the survival time is defined as that from the trial entry to the mortality caused by colorectal cancer.

The colorectal cancer screen exams used in the trial is the flexible sigmoidoscopy (FSG) and the intervention group is assigned one baseline colorectal cancer screen and one follow-up screen. Of the \( n = 142509 \) eligible subjects (those with complete information on the baseline questionnaire, with no history of any cancer including colorectal cancer prior to entry, and age no less than 55 at trial entry), 70622(49.556\%) participants
were assigned to the control arm and 71887 (50.444%) participants were assigned to the intervention arm. Even though the intervention was randomly assigned, noncompliance was observed among the participants. Of the 71887 participants randomized to the intervention arm, 19619 (27.291%) did not comply with the intervention, meaning that they did not comply to have two colorectal screening exams. With non-compliance in the randomized experiment, direct comparison between control arm and intervention arm might yields bias (Lee et al., 2020; Kianian et al., 2021).

To address the problem of non-compliance, the instrumental variable can be utilized. Here, we define the instrumental variable $Z$ as a binary indicator of whether the participant was randomized to the intervention arm: $Z = 1$ or control arm: $Z = 0$, and $A$ is a binary indicator of the actual treatment the participant received, where $A = 0$ means participants didn’t receive two screens and $A = 1$ means participants receive two screens. Because our instrument is the randomization procedure itself, the underlying assumptions of the IV being associated with treatment (Assumption 3), the IV not being associated with unmeasured confoundings (A4), the exclusion restriction (A5) and the IV positivity (A6) are all obviously held.

Preliminary analysis results in Kianian et al. (2021) showed that sex (1 for female), Family History of Any Cancer ($fh_{cancer}$), Family History of Colorectal Cancer ($colo_fh$), Colorectal Polyps ($polyps_f$), Diabetes ($diabetes_f$) and age ($0= \leq 59$, $1= 60-64$, $2= 65-69$, $3= \geq 70$) are six important risk predictors and the male elderly who has a family history of colorectal cancer, colorectal polyps and diabetes is the riskiest. It should be noticed that the family history of colorectal cancer would indicate the family history of any cancer. Thus, we include the above six baseline covariates in constructing an optimal regime.

It is intuitive that screening would bring about a higher survival probability and people are suggested to have regular cancer screen. However, with consideration of limited medical resources, it is unrealistic to encourage everyone to do the screening. Thus, for a given $t$, our optimization objective is defined as follows to balance these two considerations.

$$\hat{\eta}^{opt} = \arg \max_{||\eta||=1} \hat{S}^* (t; \eta) - \lambda \mathbb{P}_{\eta} d_{\eta}(L),$$

where $\lambda$ is a hyperparameter used to trade off the potential survival probability and the medical consumption and $t$ is predetermined. In the following, we would vary $\lambda = 0, 0.003, 0.01$ and $t = 6000, 7000$ (days) to find the optimal regime. We also equally partition the data into the training set and testing set to estimate the optimal regime in the training set and to implement bootstrap 500 times in the testing set to give a 95% confidence interval. Because $Z$ is randomly assigned, the model for $f(Z|L)$ is always correct, we don’t present the results with doubly robust estimation for the similar results it would give.
We present our results in Table 3. For $\lambda = 0$, we only consider survival probability and the optimal regime recommends almost all of the people (98.82% for 6000 days and 99.25% for 7000 days) to do the screen. For $\lambda = 0.01$, we take medical consumption as important, and then the optimal regime recommends almost all of the people (99.39% for 6000 days and 98.79% for 7000 days) not to screen. Since the optimal regime leans to one side when $\lambda = 0$ and 0.01, the optimal parameters $\hat{\eta}_{opt}$ are not unique and thus have no explainability. If we want to achieve a trade-off between survival probability and medical consumption, we should choose $\lambda$ belonging to $CI_0$ of that $\lambda = 0$, which would not overwhelm the survival probability. Here we choose $\lambda = 0.003$ and a trade-off between survival probability and medical consumption achieves indeed. For people aged from 65 to 70, it is recommended to have a screen if they have a family history of any cancer from the perspective of 6000 days. From the perspective of 7000 days, people aged 65 to 70 without diseases of diabetes and colorectal polyps are recommended. The underlying reason might be that people with colorectal polyps and diabetes benefits less than the people without them. In the aspect of family history of any cancer (including colorectal cancer) and gender, male or people with a family history of any cancer are more recommended for their higher risks on colorectal cancer. In summary, with the consideration on both survival probability and medical consumption, people who are at risks (male, family history of any cancer and age between 65 to 70) with no diseases (diabetes and colorectal polyps) are the most recommended subgroup. More specific individual treatment regime can be calculated directly from the Table 3.

We also present the optimal regime based on “SIPS” (Jiang et al., 2017a) in Table 4. When $\lambda = 0$, the only consideration is survival probability and the optimal regime optimized when $t = 6000$ days recommends all of the people to have a screening, but the optimal regime when $t = 7000$ days only recommends 89.53 percent of individuals to screen, which is less reasonable than our proposed methods. It is also noticed that $CI_1 = [-0.1329, -0.04429] \times 10^{-2}$ when $t = 7000$ days, which suggests the weak validity of the optimal regime for the poor performance in the testing dataset. When $\lambda = 0.01$, we take medical consumption as important and the optimal regime recom-

| t     | $\lambda$ | Intercept | sex | fh_cancer | colo_fh | polyps_f | diabetes_f | age1 | age2 | age3 | $CI_1 \times 10^2$ | $CI_0 \times 10^2$ |
|-------|-----------|-----------|-----|-----------|---------|----------|------------|------|------|------|-----------------|-----------------|
| 6000  | 0         | 50.56     | 51.91 | -24.28    | -7.032  | 16.17    | -19.24     | -17.54| -38.56| -38.66| $[-0.03599, 0.02248]$ | $[0.00556, 0.4954]$ |
| 7000  | 0.003     | 74.47     | -4.636| -0.0067   | 14.92   | -59.64   | 22.24      | 4.538 | 6.706 | -6.621| $[-0.03879, 0.02152]$ | $[0.1049, 0.4891]$ |
| 6000  | 0.003     | 19.96     | 53.39 | 28.51     | -4.931  | -38.26   | -25.63     | -13.56| 45.98 | -38.67| $[-0.1166, 0.1568]$ | $[-0.99587, 0.1668]$ |
| 6000  | 0.01      | -61.83    | 27.55 | -29.05    | -26.11  | 53.46    | 1.932      | -28.92| -6.682| -12.27| $[0.5991, 0.9527]$ | $[0.00516, 0.02732]$ |
| 7000  | 0.01      | -57.71    | 31.96 | -10.08    | -24.41  | 49.16    | 1.620      | -45.96| -7.133| -18.98| $[0.5304, 0.9029]$ | $[0.00620, 0.04841]$ |

Table 3: The estimates of $\eta_{opt} \times 10^{-2}$ in the real application based on the ‘SIVE’. $CI_1$ and $CI_0$ denote the 95% confidence intervals for the difference of the objective functions defined in 7 obtained under the estimated optimal regime and the simple treatment regime assigning all to treatment 1 and 0, respectively.
mends almost all people not to screen as expected. When \( \lambda = 0.03 \), a trade-off between survival probability and medical consumption should achieve. However, when \( t = 6000 \) days, \( CI_0 = [-0.1021, 0.0476] \times 10^{-2} \) suggests the performance of the optimal regime in the testing set is less powerful than simple regime assigning all to 0. When \( t = 7000 \) days, \( CI_1 = [-0.1034, 0.0951] \times 10^{-2} \) also indicates the performance of the optimal regime in the testing set is less powerful than the simple regime assigning all to 1. Thus, the validity of the optimal regime based on "SIPS" is questionable.

| \( t \) | \( \lambda \) | intercept | sex | fh | cancer | colo_fh | polys_f | diabetes_f | age1 | age2 | age3 | \( CI_1 \times 10^{-2} \) | \( CI_0 \times 10^{-2} \) |
|-------|----------|-----------|-----|----|--------|---------|---------|-----------|------|------|------|----------------|----------------|
| 6000  | 0        | 33.33     | 33.33| 33.33| 33.33  | 33.33   | 33.33   | 33.33     | 33.33| 33.33| 33.33| [-0.1329, -0.0442] | [0.1329, 0.3566] |
| 6000  | 0.003    | 9.294     | -51.40| 29.00| -2.2704| -44.57  | -30.14  | -23.16    | -0.06749| 0.1247 | 0.3751| [-0.1021, 0.0476] | [0.1021, 0.0476] |
| 7000  | 0.003    | 10.56     | -2.090| 44.98| 22.66  | -17.30  | -20.59  | -37.04    | 60.24| 40.29| 0.09517| [-0.1034, 0.0951] | [-0.05803, 0.1152] |
| 6000  | 0.01     | -56.94    | -8.662| -53.49| 12.57  | -18.66  | 21.46   | -8.061    | 16.57| -50.15| 29.54| [-0.1034, 0.0951] | [-0.01490, 0.01195] |
| 7000  | 0.01     | -60.06    | -7.745| -5.319| 50.92  | -48.86  | -25.39  | -37.04    | 60.24| 50.92| 29.54| [-0.10674, 0.0979] | [-0.01654, 0.009309] |

Table 4: The estimates of \( \eta^{opt} \times 10^{-2} \) in the real application based on the 'SIPS'. \( CI_1 \) and \( CI_0 \) denote the 95% confidence intervals for the difference of the objective functions defined in 7 obtained under the estimated optimal regime and the simple treatment regime assigning all to treatment 1 and 0, respectively.

6 Discussion

In the article, we establish the identification of the potential survival probability under any regime for a binary treatment subject to unmeasured confoundings. In addition, we establish doubly robust estimator to identify the potential survival probability under any regime which is consistent when one of the models is correct. Our optimization objective is the \( t \)-year potential survival probability, which can be generalized to the restricted mean (Geng et al., 2015) and the quantile (Zhou et al., 2022) of the potential survival probability.

The asymptotic properties for the estimators are provided to analyze the convergence rate and show the advantage of the doubly robust estimator on both convergence rate and model mis-specification. Consistency for the kernel smoothed estimators and the optimal regime are also provided. However, the convergence rate for \( \hat{\eta}^{opt} \) is left to be solved for the very challenging it would be. The rate of the convergence rate for \( \hat{\eta}^{opt} \) is slower than the classical \( \sqrt{n} \)-rate due to the indicator function \( I(\tilde{L}^T \eta \geq 0) \). As Zhou et al. (2022) illustrated, the convergence rate might be \( n^{1/3} \)-rate under a mild condition.

Acknowledgement

The authors thank the National Cancer Institute for access to NCI’s data collected by the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial.
References

Blümllein, T., Persson, J., and Feuerriegel, S. (2022). Learning optimal dynamic treatment regimes using causal tree methods in medicine. *arXiv preprint arXiv:2204.07124*.

Chen, S. and Zhang, B. (2021). Estimating and improving dynamic treatment regimes with a time-varying instrumental variable. *arXiv preprint arXiv:2104.07822*.

Csörgő, S. and Horváth, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 62(3):411–426.

Cui, Y. and Tchetgen, E. T. (2021). Machine intelligence for individualized decision making under a counterfactual world: A rejoinder. *Journal of the American Statistical Association*, 116(533):200–206.

Cui, Y. and Tchetgen Tchetgen, E. (2021a). On a necessary and sufficient identification condition of optimal treatment regimes with an instrumental variable. *Statistics & Probability Letters*, 178:109180.

Cui, Y. and Tchetgen Tchetgen, E. (2021b). A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, 116(533):162–173.

Dawid, A. P. (2003). Causal inference using influence diagrams: The problem of partial compliance. *OXFORD STATISTICAL SCIENCE SERIES*, pages 45–65.

Delsol, L. and Van Keilegom, I. (2020). Semiparametric m-estimation with non-smooth criterion functions. *Annals of the Institute of Statistical Mathematics*, 72(2):577–605.

Geng, Y., Zhang, H. H., and Lu, W. (2015). On optimal treatment regimes selection for mean survival time. *Statistics in Medicine*, 34(7):1169–1184.

Gonzalez-Manteiga, W. and Cadarso-Suarez, C. (1994). Asymptotic properties of a generalized kaplan-meier estimator with some applications. *Communications in Statistics-Theory and Methods*, 4(1):65–78.

Jiang, R., Lu, W., Song, R., and Davidian, M. (2017a). On estimation of optimal treatment regimes for maximizing t-year survival probability. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1165–1185.

Jiang, R., Lu, W., Song, R., Hudgens, M. G., and Napravnik, S. (2017b). Doubly robust estimation of optimal treatment regimes for survival data—with application to an hiv/aids study. *The Annals of Applied Statistics*, 11(3).
Kennedy, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In Statistical Causal Inferences and Their Applications in Public Health Research, pages 141–167. Springer.

Kianian, B., Kim, J. I., Fine, J. P., and Peng, L. (2021). Causal proportional hazards estimation with a binary instrumental variable. Statistica Sinica, 31(2):673.

Laber, E. B., Linn, K. A., and Stefanski, L. A. (2014). Interactive model building for q-learning. Biometrika, 101(4):831–847.

Lee, Y., Kennedy, E. H., and Mitra, N. (2020). Doubly robust nonparametric instrumental variable estimators for survival outcomes.

Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., and Phan, L. (2007). The ncbi dbgap database of genotypes and phenotypes. Nature genetics, 39(10):1181–1186.

Mebane, W. R. and Sekhon, J. S. (2011). Genetic optimization using derivatives: The rgenoud package for r. Journal of Statistical Software, 42(11).

Mitchell, M. (1998). An Introduction to Genetic Algorithms. MIT press.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65(2):331–355.

Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. Statistics in medicine, 24(10):1455–1481.

Pan, Y. and Zhao, Y.-Q. (2021). Improved doubly robust estimation in learning optimal individualized treatment rules. Journal of the American Statistical Association, 116(533):283–294.

Pu, H. and Zhang, B. (2020). Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach. arXiv preprint arXiv:2002.02579.

Qi, Z. and Liu, Y. (2018). D-learning to estimate optimal individual treatment rules. Electronic Journal of Statistics, 12(2):3601–3638.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. The Annals of Statistics, 39(2).

Qiu, H., Carone, M., Sadikova, E., Petukhova, M., Kessler, R. C., and Luedtke, A. (2021). Optimal individualized decision rules using instrumental variable methods. Journal of the American Statistical Association, 116(533):174–191.
Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In Proceedings of the Second Seattle Symposium in Biostatistics, pages 189–326. Springer.

Sukjin and Han (2020). Comment: Individualized treatment rules under endogeneity.

Tchetgen Tchetgen, E. J., Walter, S., Vansteelandt, S., Martinussen, T., and Glymour, M. (2015). Instrumental variable estimation in a survival context. Epidemiology, 26(3):402–410.

Team, P. P., Gohagan, J. K., Prorok, P. C., Hayes, R. B., and Kramer, B.-S. (2000). The prostate, lung, colorectal and ovarian (plco) cancer screening trial of the national cancer institute: History, organization, and status. Controlled clinical trials, 21(6):251S–272S.

Tsiatis and A, A. (2006). Semiparametric Theory and Missing Data. Springer.

van der Vaart, A. W. (1998). Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK; New York, NY, USA.

Wang, L., Tchetgen, E. T., Martinussen, T., and Vansteelandt, S. (2022). Iv estimation of causal hazard ratio.

Wang, L. and Tchetgen Tchetgen, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 80(3):531–550.

Wang, Y., Fu, H., and Zeng, D. (2018). Learning optimal personalized treatment rules in consideration of benefit and risk: With an application to treating type 2 diabetes patients with insulin therapies. Journal of the American Statistical Association, 113(521):1–13.

Wang, Z. and Louis, T. A. (2003). Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. Biometrika, 90(4):765–775.

Zhang, B. and Pu, H. (2021). Discussion of cui and tchetgen tchetgen (2020) and qiu et al. (2020). Journal of the American Statistical Association, 116(533):196–199.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. Biometrics, 68(4):1010–1018.

Zhang, B. and Zhang, M. (2018). C-learning: A new classification framework to estimate optimal dynamic treatment regimes. Biometrics, 74(3):891–899.
S1 Proof from the main manuscript

Proof of Theorem 3.1: For any given regime \( d_\eta \), we first derive the asymptotic properties for the corresponding weighted Nelson-Aalen estimator.

\[
\hat{\Lambda}_I^*(t; \eta) = \int_0^t \frac{\sum_{i=1}^n \hat{w}_i dN_i(s)}{\sum_{i=1}^n \hat{w}_i Y_i(s)} \cdot \frac{A}{\delta(L)f(Z|L)}
\]

It is easy to show that \( \hat{S}_I^*(t; \eta) \) and \( \exp\left\{-\hat{\Lambda}_I^*(t; \eta)\right\} \) are asymptotically equivalent for any given \( t \) and \( \eta \). Therefore, the asymptotic properties of \( \hat{S}_I^*(t; \eta) \) easily follows those of \( \hat{\Lambda}_I^*(t; \eta) \)

Similar to the proof of Theorem 1 in Cui and Tchetgen Tchetgen (2021b), we have

\[
E[dN^*(d_\eta(L); s)] = E\left[\frac{(2Z - 1)(2A - 1)dN(s)I\{A = d_\eta(L)\}}{\delta(L)f(Z|L)}\right],
\]

and

\[
E[Y^*(d_\eta(L); s)] = E\left[\frac{(2Z - 1)(2A - 1)Y(s)I\{A = d_\eta(L)\}}{\delta(L)f(Z|L)}\right].
\]

Therefore,

\[
\hat{\Lambda}_I^*(t; \eta) - \Lambda^*(t; \eta) = \int_0^t \frac{P_n \hat{w}dN(s)}{P_n \hat{w}Y(s)} - \int_0^t \frac{PdN^*(d_\eta(L); s)}{PY^*(d_\eta(L); s)P_n \hat{w}Y(s)}
\]

\[
= \int_0^t \frac{P_n \hat{w}dN(s)PY^*(d_\eta(L); s) - PdN^*(d_\eta(L); s)P_n \hat{w}Y(s)}{PY^*(d_\eta(L); s)P_n \hat{w}Y(s)}
\]

\[
= \int_0^t \frac{P_n \hat{w}dN(s) - PdN^*(d_\eta(L); s)}{P_n \hat{w}Y(s)} \lambda^*(s; \eta) \, ds,
\]

where \( \lambda^*(s; \eta) \) is the potential hazard function under regime \( d_\eta \). We can treat the
ratio as a constant under (C2) and break down the other components into smaller pieces, for example:

\[ \mathbb{P}_{n} \hat{\omega} Y (s) - \mathbb{P} Y^* (d_{\eta}(L); s) = (\mathbb{P}_{n} - \mathbb{P}) (\hat{\omega} Y (s) - Y^* (d_{\eta}(L); s)) + (\mathbb{P}_{n} - \mathbb{P}) Y^* (d_{\eta}(L); s) + \mathbb{P} (\hat{\omega} Y (s) - Y^* (d_{\eta}(L); s)). \]

The first term on the right is \( o_{p}(1/\sqrt{n}) \) by van der Vaart (1998) as long as assume that the nuisance model belonging to Donsker class as Condition 1. The second term is normal by the CLT. We denote \( h_Y(Z, L) = E[(2A - 1)Y(s)I\{A = d(L)\}|Z, L] \), then the third term can be bounded as follows:

\[
\mathbb{P} (\hat{\omega} Y (s) - Y^* (d_{\eta}(L); s)) = \mathbb{P} [ (2Z - 1)(2A - 1)Y(s)I \{A = d_{\eta}(L)\} \{ \frac{1}{\delta(L)} f(Z|L) - \frac{1}{\delta(L)} f(z|L) \} ]
\]

\[
= \mathbb{P} [ \sum_{z} h_Y(z, L) f(z|L) (\frac{1}{\delta(L)} f(z|L) - \frac{1}{\delta(L)} f(z|L)) ]
\]

\[
\leq \mathbb{P} \sum_{z} || \hat{\delta}(L) f(z|L) - \delta(L) f(z|L)||
\]

Thus

\[
\hat{\Lambda}^* (t; \eta) - \Lambda^* (t; \eta) = O_{p}(\sum_{z} || (\hat{\delta}(L) f(z|L) - \delta(L) f(z|L)||) + \int_{0}^{t} (\mathbb{P}_{n} - \mathbb{P}) \{ dN^* (d_{\eta}(L); s) - Y^* (d_{\eta}(L); s) d\Lambda^* (s; \eta) \} \mathbb{P} Y^* (d_{\eta}(L); s) + o_{p}(n^{-1/2}).
\]

By delta method, we have \( \hat{S}^* (t; \eta) - S^* (t; \eta) = S^* (t; \eta) (\hat{\Lambda}^* (t; \eta) - \Lambda^* (t; \eta)) + o \left( \hat{\Lambda}^* (t; \eta) - \Lambda^* (t; \eta) \right) \), thus

\[
\hat{S}^* (t; \eta) - S^* (t; \eta) = O_{p}(\sum_{z} || (\hat{\delta}(L) f(z|L) - \delta(L) f(z|L)||) + S^* (t; \eta) \int_{0}^{t} (\mathbb{P}_{n} - \mathbb{P}) \{ dN^* (d_{\eta}(L); s) - Y^* (d_{\eta}(L); s) d\Lambda^* (s; \eta) \} \mathbb{P} Y^* (d_{\eta}(L); s) + o_{p}(n^{-1/2}).
\]

We also notice that \( \mathbb{P} \{ dN^* (d_{\eta}(L); s) - Y^* (d_{\eta}(L); s) d\Lambda^* (s; \eta) \} = 0 \), which completes our proof.

**Proof of Theorem 3.2:** We First introduce the following two lemmas about the efficient influence function. Lemma 1 is a corollary of Theorem S2 in Cui and Tchetgen Tchetgen
Lemma 6.2 provides an upper bound for the efficient-influence-function-based estimator. It should be noticed that \( \hat{d}n^*(d_\eta(L); s) \) and \( Y^*(d_\eta(L); s) \) are measurable functions of \( T^*(d_\eta(L)) \) and \( C \), thus we unify them by \( g(T^*, C) \), where \( g \) is an measurable function of \( T^*(d_\eta(L)) \) and \( C \).

**Lemma 6.1** Under A1* and A3-7, for any measurable function \( g \) of \( T^*(d_\eta(L)) \) and \( C \), the efficient influence function of \( E[g(T^*, C)] \) is given by

\[
EIF_{E[g(T^*, C)]} = \frac{(2Z - 1)(2A - 1)g(T, C)|\{A = d(L)\}}{f(\hat{g}(L))\hat{\delta}(L)} - \left\{ \frac{(2Z - 1)E[(2A - 1)g(T, C)|\{A = d(L)\}]}{f(\hat{g}(L))\hat{\delta}(L)} \right. \\
+ \frac{(2Z - 1)[A - E(A|Z, L)]}{f(\hat{g}(L))\hat{\delta}(L)} \sum_z \frac{E[(2A - 1)g(T, C)|\{A = d(L)\}]}{\hat{\delta}(L)}|Z = z, L|(2z - 1) \\
+ \sum_z \frac{E[(2A - 1)g(T, C)|\{A = d(L)\}]}{\hat{\delta}(L)}|Z = z, L|(2z - 1) - E[g(T^*, C)]
\]

Via semiparametric theory and the property of efficient influence function, a doubly robust estimator of \( E[g(T^*, C)] \) is as follows:

\[
P_n\hat{IF}_{E[g(T^*, C)]} = \frac{1}{n} \sum_{i=1}^{n} \frac{2Z_i - 1}{\hat{\delta}(L_i)f(Z_i|L_i)} \{(2A_i - 1)g(T, C)|\{A_i = d_\eta(L_i)\} \\
- \hat{\gamma}_g(L_i) - (A_i - \hat{E}(A|Z = 0, L_i))\hat{\gamma}_g(L_i)\} + \hat{\gamma}_g(L_i),
\]

where \( \hat{\gamma}_g(L) = \hat{h}_g(Z = 0, L), \hat{\gamma}_g(L) = \sum_z \hat{h}_g(z, L)(2z - 1)/\hat{\delta}(L) \), and \( \hat{h}_g(Z, L) = \hat{E}[(2A - 1)g(T, C)|\{A = d(L)\}]|Z, L|\).

**Lemma 6.2** Under A1*, A3-7, C1 and C2*, for any measurable function \( g \) of \( T^*(d_\eta(L)) \) and \( C \),

\[
P_n\hat{IF}_{E[g(T^*, C)]} - Pg(T^*, C) = O_p(\sum_z ||h_g(z, L) - \hat{h}_g(z, L)|| \cdot ||\delta(L)f(z|L) - \hat{\delta}(L)f(z|L)||) \\
+ ||E(A|z, L) - \hat{E}(A|z, L)|| \cdot ||\delta(L)f(z|L) - \hat{\delta}(L)f(z|L)|| \\
+ (P_n - \mathbb{P})g(T^*, C^*) + o_p(n^{-1/2})
\]

**Proof of Lemma 6.2:**

\[
P_n\hat{IF} - \mathbb{P}g = (P_n - \mathbb{P})(\hat{IF} - g) + (P_n - \mathbb{P})g + \mathbb{P}(\hat{IF} - g)
\]

The first term on the right is \( o_p(1/\sqrt{n}) \) by van der Vaart (1998) as long as assume that

26
the nuisance model belonging to Donsker class as Condition 1. The second term is normal
by the CLT. The third term can be bounded as follows.

\[
P(\tilde{I}F - g) = P[(2Z - 1)(2A - 1)g(T, C)I\{A = d_\eta(L)\} \left(\frac{1}{\delta(L)f(Z|L)} - \frac{1}{\delta(L)f(Z|L)}\right)
- \frac{2Z - 1}{\delta(L)f(Z|L)} \{\gamma_g(L) + (A - \hat{E}(A|Z = 0, L))\hat{\gamma}_g(L)\} + \gamma_g(L)\]
= P[(2Z - 1)h_g(Z, L)\left(\frac{1}{\delta(L)f(Z|L)} - \frac{1}{\delta(L)f(Z|L)}\right)
- \frac{2Z - 1}{\delta(L)f(Z|L)} \{\gamma_g(L) + (E(A|Z, L) - \hat{E}(A|Z = 0, L))\hat{\gamma}_g(L)\} + \gamma_g(L)\]
= P \sum_z [(2Z - 1)h_g(z, L)\left(\frac{1}{\delta(L)f(z|L)} - \frac{1}{\delta(L)f(z|L)}\right) f(z|L)
- \frac{2Z - 1}{\delta(L)f(z|L)} \{\gamma(z)\delta(L) + \gamma(z)\hat{\gamma}(L) + (E(A|z, L) - \hat{E}(A|Z = 0, L) - z\delta(L))\hat{\gamma}_g(L)\} f(z|L)
+ \gamma_g(L)f(z|L)\]
= P \sum_z [(2Z - 1)h_g(z, L)\left(\frac{1}{\delta(L)f(z|L)} - \frac{1}{\delta(L)f(z|L)}\right) f(z|L)
- \frac{2Z - 1}{\delta(L)f(z|L)} \{\tilde{h}(z, L) + (E(A|z, L) - \hat{E}(A|z, L))\hat{\gamma}_g(L)\} f(z|L) + \gamma_g(L)f(z|L)\]
= P \sum_z [(2Z - 1)h_g(z, L)\left(\frac{1}{\delta(L)f(z|L)} - \frac{1}{\delta(L)f(z|L)}\right) f(z|L)
- \frac{2Z - 1}{\delta(L)f(z|L)} \{\tilde{h}(z, L) + (E(A|z, L) - \hat{E}(A|z, L))\hat{\gamma}_g(L)\} f(z|L) + \gamma_g(L)f(z|L)\]
= P \sum_z [(2Z - 1)(h_g(z, L) - \hat{h}(z, L))\left(\frac{1}{\delta(L)f(z|L)} - \frac{1}{\delta(L)f(z|L)}\right) f(z|L)
- \frac{2Z - 1}{\delta(L)f(z|L)} \{\tilde{h}(z, L) + (E(A|z, L) - \hat{E}(A|z, L))\hat{\gamma}_g(L)\} f(z|L)\]
= P \sum_z [(2Z - 1)(h_g(z, L) - \hat{h}(z, L))\left(\frac{1}{\delta(L)f(z|L)} - \frac{1}{\delta(L)f(z|L)}\right) f(z|L)
+ (2Z - 1)(E(A|z, L) - \hat{E}(A|z, L))\hat{\gamma}_g(L)\left(\frac{1}{\delta(L)f(z|L)} - \frac{1}{\delta(L)f(z|L)}\right) f(z|L)]
\[ \leq p \sum_z ||h_{g_\gamma}(z, L) - \hat{h}_{g_\gamma}(z, L)|| \cdot ||\delta(L)f(z|L) - \hat{\delta}(L)\hat{f}(z|L)|| \\
+ ||E(A|z, L) - \hat{E}(A|z, L)|| \cdot ||\delta(L)f(z|L) - \hat{\delta}(L)\hat{f}(z|L)|| \]

The second equality follows from the law of iterated expectation conditioning on \( L, Z \). The fifth and the sixth equality follows from the definition of \( \hat{\gamma}_g(L) \) and \( \hat{\gamma}_g(L) \). The eighth equality follows from the following equality:

\[ \mathbb{P} \sum_z \frac{(2z - 1)\hat{h}(z, L)}{\hat{\delta}(L)} = \hat{\gamma}_g(L) = \sum_z (2z - 1)\hat{\gamma}_g(L) \]

The last inequality follows from Cauchy Schwartz’s inequality with Condition 2*. Thus we complete the proof of Lemma 2.

Now, we start the proof of Theorem 3.2. We first notice that

\[ E[(2A - 1)I\{A = d_\eta(L)\}dN^* \{d_\eta(L); s\} | Z, L] = (2d_\eta(L) - 1) \times \\
f(A = d_\eta(L)|Z, L)S_T \{s\} | Z, L, d_\eta(L) \} S_C(s)dA_T \{s\} | Z, L, d_\eta(L) \} , \]

and

\[ E[(2A - 1)I\{A = d_\eta(L)\}Y^* \{d_\eta(L); s\} | Z, L] = (2d_\eta(L) - 1) \times \\
f(A = d_\eta(L)|Z, L)S_T \{s\} | Z, L, d_\eta(L) \} S_C(s) , \]

where \( f(A = d_\eta(L)|Z, L) = \sum_a I\{d_\eta(L) = a\}f(A = a|Z, L) \). Then, with an application of Lemma 2, we have

\[ \mathbb{P}_n \hat{F} E[dN^*(d_\eta(L); s)] - \mathbb{P} dN^*(d_\eta(L); s) \]

\[ = O_p \{ \sum_z ||\delta(L)f(z|L) - \hat{\delta}(L)\hat{f}(z|L)|| \cdot (||E(A|z, L) - \hat{E}(A|z, L)||) \\
+ ||f(A = d_\eta(L)|z, L)S_{T,s,z}S_{C,s}dA_{T,s,z} - \hat{f}(A = d_\eta(L)|z, L)\hat{S}_{T,s,z}\hat{S}_{C,s}d\hat{A}_{T,s,z}||) \}
+ (\mathbb{P}_n - \mathbb{P})dN^*(d_\eta(L); s) + o_p(n^{-1/2}) , \]

and

\[ \mathbb{P}_n \hat{F} E[Y^*(d_\eta(L); s)] - \mathbb{P} Y^*(d_\eta(L); s) \]

\[ = O_p \{ \sum_z ||\delta(L)f(z|L) - \hat{\delta}(L)\hat{f}(z|L)|| \cdot (||E(A|z, L) - \hat{E}(A|z, L)||) \\
+ ||f(A = d_\eta(L)|z, L)S_{T,s,z}S_{C,s} - \hat{f}(A = d_\eta(L)|z, L)\hat{S}_{T,s,z}\hat{S}_{C,s}||) \}
+ (\mathbb{P}_n - \mathbb{P})Y^*(d_\eta(L); s) + o_p(n^{-1/2}) , \]
where $S_{T,s,z}$ is the shorthand of $S_T \{ s | z, L, d_\eta(L) \}$, $S_{C,s} = S_C(s)$ and $\Lambda_{T,s,z} = \Lambda_T \{ s | z, L, d_\eta(L) \}$. Thus,

\[
\hat{\Lambda}_D^* (t; \eta) - \Lambda^* (t; \eta) = \int_0^t \frac{\mathbb{P}_n \hat{F}_E [ dN^* (d_\eta(L); s) ]}{\mathbb{P}_n \hat{F}_E [ Y^* (d_\eta(L); s) ]} ds - \int_0^t \frac{\mathbb{P}_n \hat{F}_E [ Y^* (d_\eta(L); s) ]}{\mathbb{P}_n \hat{F}_E [ Y^* (d_\eta(L); s) ]} ds
\]

\[
= \int_0^t \frac{\mathbb{P}_n \hat{F}_E [ dN^* (d_\eta(L); s) ]}{\mathbb{P}_n \hat{F}_E [ Y^* (d_\eta(L); s) ]} df (d_\eta(L); s) - \mathbb{P}_n \hat{F}_E [ Y^* (d_\eta(L); s) ] ds
\]

\[
= \int_0^t \frac{\mathbb{P}_n \hat{F}_E [ dN^* (d_\eta(L); s) ]}{\mathbb{P}_n \hat{F}_E [ Y^* (d_\eta(L); s) ]} df (d_\eta(L); s) - ( \mathbb{P}_n \hat{F}_E [ Y^* (d_\eta(L); s) ] - \mathbb{P}_n \hat{F}_E [ Y^* (d_\eta(L); s) ] ) ds + o_p (n^{-1/2}),
\]

By delta method, we have $\hat{S}_D^* (t; \eta) = S^* (t; \eta) (\hat{\Lambda}_D^* (t; \eta) - \Lambda^* (t; \eta)) + o \left( \hat{A}_f^* (t; \eta) - A^* (t; \eta) \right)$, thus

\[
\hat{S}_D^* (t; \eta) = S^* (t; \eta) (\hat{\Lambda}_D^* (t; \eta) - \Lambda^* (t; \eta)) + o \left( \hat{A}_f^* (t; \eta) - A^* (t; \eta) \right),
\]

which completes the proof.

**Proof of Theorem 3.3 and 3.4:** Define

\[
K_1^\prime (Z, L, A, \hat{T}, \hat{\delta}, d_\eta(L)) = \int_0^t \frac{(2Z - 1) dN(s)}{\delta(L) f(Z | L) E \{ Y^* (d_\eta(L); s) \}},
\]

\[
K_2^\prime (Z, L, A, \hat{T}, \hat{\delta}, d_\eta(L)) = \int_0^t \frac{(2Z - 1) Y(s) E \{ dN^* (d_\eta(L); s) \}}{\delta(L) f(Z | L) E \{ Y^* (d_\eta(L); s) \}^2}.
\]

In addition, define

\[
K_D^\prime (Z, L, A, \hat{T}, \hat{\delta}; d_\eta(L)) = \int_0^t \frac{J(s)}{E \{ Y^* (d_\eta(L); s) \}}.
\]
\[ \begin{align*}
K_2^P(Z, L, A, \tilde{T}, \delta; d_\eta(L)) &= \int_0^t L(s) \left\{ \frac{dN^*(d_\eta(L); s)}{\left[ E \left\{ Y^*(d_\eta(L); s) \right\} \right]^2} \right\},
\end{align*} \]

where

\[ J(s) = \frac{(2Z - 1)dN(s)}{\delta(L) f(Z|L)} - \sum_a \frac{(2Z - 1)f(a|Z = 0, L) S_{T,s,Z=0,a} S_{C,s} d\Lambda_{T,s,Z=0,a}}{\delta(L) f(Z|L)} \]

\[ - \frac{(2Z - 1)(A - E(A|Z = 0, L))}{\delta^2(L) f(Z|L)} \sum_{z,a} (2Z - 1) f(a|z, L) S_{T,s,z,a} S_{C,s} d\Lambda_{T,s,z,a} \]

\[ + \sum_{z,a} (2Z - 1) f(a|z, L) S_{T,s,z,a} S_{C,s} d\Lambda_{T,s,z,a}, \]

\[ L(s) = \frac{(2Z - 1)Y(s)}{\delta(L) f(Z|L)} - \sum_a \frac{(2Z - 1)f(a|Z = 0, L) S_{T,s,Z=0,a} S_{C,s}}{\delta(L) f(Z|L)} \]

\[ - \frac{(2Z - 1)(A - E(A|Z = 0, L))}{\delta^2(L) f(Z|L)} \sum_{z,a} (2Z - 1) f(a|z, L) S_{T,s,z,a} S_{C,s} \]

\[ + \sum_{z,a} (2Z - 1) f(a|z, L) S_{T,s,z,a} S_{C,s}, \]

where \( S_{T,s,z,a} \) is the shorthand of \( S_{T}(s|z, L, a) \), \( S_{C,s} = S_{C}(s) \) and \( \Lambda_{T,s,z,a} = \Lambda_{T}(s|z, L, a) \).

We assume the following conditions:

**Condition 4:** \( \sup_{\|\eta\|=1} E \left\{ K_j^I(Z, L, A, \tilde{T}, \delta, d_\eta(L)) \right\} < \infty \) and \( \sup_{\|\eta\|=1} E \left\{ K_j^P(Z, L, A, \tilde{T}, \delta, d_\eta(L)) \right\} < \infty \) for \( j = 1, 2 \).

**Condition 5:** \( nh \to \infty \) and \( nh^4 \to 0 \) as \( n \to \infty \).

Since C3 or C3* is satisfied, (1) in Theorem 3.3 and 3.4 can be reasoned directly by that, in donsker class,

\[ \sup_{\|\eta\|=1} (\mathbb{P}_n - \mathbb{P})(\hat{g} - g) + (\mathbb{P}_n - \mathbb{P})g = O_p(n^{-1/2}), \]

where \( \hat{g} \) is an estimator of \( g \).

(2) in Theorem 3.3 and 3.4 is a corollary of (1) as the Theorem 1 in (Delsol and Van Keilegom, 2020).

Now, we show that \( \hat{S}^*_T(t; \eta) \) and \( \hat{S}^*_S(t; \eta) \) are asymptotically equivalent. For any given
\( \eta \) and \( t \), we have

\[
\sqrt{n} \left\{ \hat{\Lambda}^*_S(t; \eta) - \hat{\Lambda}'_I(t; \eta) \right\} \\
= \sqrt{n} \times \frac{1}{n} \sum_{i=1}^{n} \Phi \left( \frac{\eta^T \hat{L}_i}{h} \right) - I \left( \eta^T \hat{L}_i \geq 0 \right) \times K_1^n(Z_i, L_i, A_i, \tilde{T}_i, \delta_i, d_{\eta}(L_i)) \\
- \sqrt{n} \times \frac{1}{n} \sum_{i=1}^{n} \Phi \left( \frac{\eta^T \hat{L}_i}{h} \right) - I \left( \eta^T \hat{L}_i \geq 0 \right) \times K_2^n(Z_i, L_i, A_i, \tilde{T}_i, \delta_i, d_{\eta}(L_i)) \\
+ o_p(1)
\]

and

\[
\sqrt{n} \left\{ \hat{\Lambda}^*_S(t; \eta) - \hat{\Lambda}'_D(t; \eta) \right\} \\
= \sqrt{n} \times \frac{1}{n} \sum_{i=1}^{n} \Phi \left( \frac{\eta^T \hat{L}_i}{h} \right) - I \left( \eta^T \hat{L}_i \geq 0 \right) \times K_1^n(Z_i, L_i, A_i, \tilde{T}_i, \delta_i, d_{\eta}(L_i)) \\
- \sqrt{n} \times \frac{1}{n} \sum_{i=1}^{n} \Phi \left( \frac{\eta^T \hat{L}_i}{h} \right) - I \left( \eta^T \hat{L}_i \geq 0 \right) \times K_2^n(Z_i, L_i, A_i, \tilde{T}_i, \delta_i, d_{\eta}(L_i)) \\
+ o_p(1)
\]

Then the following proof is the same as the proof of Theorem 1(d) and 2(d) in (Jiang et al., 2017a).

(4) in Theorem 3.3 and 3.4 is also a corollary of (3) as the Theorem 1 in (Delsol and Van Keilegom, 2020).

**S2 Additional simulation**

**S2.1 Comparison of smoothed estimation and non-smoothed version**

With the same data generation way as Example 4.1, we compare the estimation of the smoothed instrumental variable method and the non-smoothed version. The results are presented in Table 5. As we state before, with less computational resources, the two are similar in parametric estimation and the MR is similar as well.

**S2.2 Sensitivity analysis on the strength of the IV**

In this supplementary section, we conduct the sensitivity analysis on the strength of the IV. Treatment \( A \) is generated under a logistic regression with success probabilities,

\[
f(A = 1|Z, L, U) = \expit \left\{ -2.5 + L^{(1)} + 4Z - 0.5U \right\}
\]

and

\[
f(A = 1|Z, L, U) = \expit \left\{ -2.5 + L^{(1)} + 3Z - 0.5U \right\},
\]
| Case | method | C% | Bias in $\hat{\eta}_1^{opt}$ | Bias in $\hat{\eta}_2^{opt}$ | Bias in $S^*(t; \eta^{opt})$ | MR |
|------|--------|----|-----------------------------|-----------------------------|-----------------------------|----|
|      | SIVE   | 30 | 0.0294 (0.348)             | 0.0737 (0.207)              | 0.0808 (0.207)              | 0.0289 (0.0445)  |
|      | IVE    | 30 | 0.0972 (0.363)             | 0.0609 (0.207)              | 0.101 (0.203)               | 0.0472 (0.0444)  |
|      | SIVE-DR| 30 | 0.0160 (0.344)             | 0.0895 (0.227)              | 0.0658 (0.194)              | 0.0294 (0.0400)  |
|      | IVE-DR | 30 | 0.0726 (0.351)             | 0.0872 (0.215)              | 0.0685 (0.185)              | 0.0472 (0.0428)  |
|      | SIVE   | 15 | 0.0291 (0.346)             | 0.0853 (0.215)              | 0.0696 (0.202)              | 0.0311 (0.0428)  |
|      | IVE    | 15 | 0.00224 (0.353)            | 0.0686 (0.219)              | 0.0910 (0.202)              | 0.0532 (0.0421)  |
|      | SIVE-DR| 15 | 0.0112 (0.338)             | 0.0640 (0.216)              | 0.0858 (0.200)              | 0.0240 (0.0402)  |
|      | IVE-DR | 15 | 0.0449 (0.338)             | 0.0866 (0.221)              | 0.0660 (0.198)              | 0.0432 (0.0422)  |

Table 5: Simulation results about the absolute bias (standard deviation of the estimates given in the parentheses) of $\hat{\eta}_1^{opt}$ relative to $\eta^{opt}$ and of $S^*(t; \hat{\eta}_1^{opt})$ relative to $S^*(t; \eta^{opt})$. 'SIVE' means the method by maximizing the smoothed instrumental variable estimation $\hat{S}_I^*(t; \eta)$; 'SIVE-DR' means the method by maximizing the smoothed doubly robust instrumental variable estimation $\hat{S}_D^*(t; \eta)$; 'SIPS' means the smoothed inverse propensity score method and 'SAIPS' means the augmented inverse propensity score method.

32
respectively. The remaining data generation way is the same as settings (a) and (c) in Example 4.1 and the censoring rate is chosen to 15%. We define the additive associations between \(A\) and \(Z\) as \(f(A = 1 \mid Z = 1) - f(A = 1 \mid Z = 0)\), which is approximately equal to 0.6708, 0.5439 and 0.4150 for coefficient of \(Z\) equaling 5, 4 and 3, respectively. As the simulation presented in Table 6, higher additive associations generally lead to a lower standard deviation of the estimated regime in terms of the optimal survival probability and misclassification rate.

| Case | method | Coef of \(Z\) | Bias in \(\hat{\eta}_1\) | Bias in \(\hat{\eta}_2\) | Bias in \(\hat{\eta}_3\) | Bias in \(\hat{\eta}_4\) | Bias in \(S^*(t; \eta^{opt})\) | MR |
|------|--------|---------------|----------------|----------------|----------------|----------------|----------------|-----|
| (a)  | SIVE   | 5             | 0.0291 (0.346) | 0.0853 (0.215) | 0.0696 (0.202) | 0.0311 (0.0428) | 0.138 (0.0639) |     |
|      | SIVE-DR| 5             | 0.0112 (0.338) | 0.0640 (0.216) | 0.0858 (0.200) | 0.0240 (0.0402) | 0.134 (0.0645) |     |
|      | SIVE   | 4             | 0.0377 (0.380) | 0.0787 (0.240) | 0.115 (0.228)  | 0.00850 (0.0516) | 0.156 (0.0753) |     |
|      | SIVE-DR| 4             | 0.0150 (0.353) | 0.105 (0.243)  | 0.0708 (0.224) | 0.00250 (0.0506) | 0.150 (0.0735) |     |
|      | SIVE   | 3             | 0.0632 (0.333) | 0.163 (0.328)  | 0.136 (0.282)  | 0.0170 (0.0753) | 0.191 (0.107)  |     |
|      | SIVE-DR| 3             | 0.0102 (0.341) | 0.180 (0.336)  | 0.123 (0.284)  | 0.0120 (0.0745) | 0.194 (0.107)  |     |
| (c)  | SIVE   | 5             | 0.0189 (0.389) | 0.0931 (0.257) | 0.111 (0.224)  | 0.0532 (0.0485) | 0.157 (0.0732) |     |
|      | SIVE-DR| 5             | 0.0221 (0.390) | 0.101 (0.241)  | 0.0984 (0.228) | 0.0342 (0.0434) | 0.157 (0.0762) |     |
|      | SIVE   | 4             | 0.0644 (0.394) | 0.120 (0.286)  | 0.114 (0.260)  | 0.00410 (0.0572) | 0.169 (0.0883) |     |
|      | SIVE-DR| 4             | 0.00353 (0.417) | 0.123 (0.274)  | 0.115 (0.247)  | 0.00190 (0.0574) | 0.175 (0.0813) |     |
|      | SIVE   | 3             | 0.0969 (0.478) | 0.206 (0.372)  | 0.204 (0.347)  | 0.0192 (0.0825) | 0.227 (0.119)  |     |
|      | SIVE-DR| 3             | 0.0365 (0.476) | 0.214 (0.377)  | 0.181 (0.334)  | 0.0112 (0.0849) | 0.226 (0.115)  |     |

Table 6: Simulation results about the absolute bias (standard deviation of the estimates given in the parentheses) of \(\hat{\eta}_1\) relative to \(\eta^{opt}\) and of \(\hat{S}_n(t; \eta^{opt})\) relative to \(S^*(t; \eta^{opt})\). The column of the 'Coef of \(Z\)' indicates the strength of the instrumental variable. For example, '4' means \(\Pr(A = 1 | Z, L, U) = \expit \{-2.5 + L^{(1)} + 4Z - 0.5U\}\). 'SIVE' means the method by maximizing the smoothed instrumental variable estimation \(\hat{S}_n(t; \eta)\); 'SIVE-DR' means the method by maximizing the smoothed doubly robust instrumental variable estimation \(\hat{S}_n(t; \eta)\).

S2.3 Simulations with different sample sizes

Utilizing the same data generation way as Example 4.1 of setting (a) and (c) with the 15% censoring rate, we present the simulation results for different sample sizes in Table 7. As the sample size enlarges, the estimation of the optimal survival probability and the optimal regime get better and the MR gets smaller.

S2.4 Simulations with different distributions of \(U\)

In Example 4.1, we generate \(U\) from a bridge distribution so as to satisfy Assumption 7. In this supplementary document, we vary the generation way of \(U\) to explore the impact of the violation on Assumption 7. In the following, \(U\) is generated from a standard normal distribution or a uniform distribution on \([-2, 2]\). The remaining generation way is the same as the setting (a) and (c) in the Example 4.1 and the censoring rate is fixed as 15%. We present the simulation result in Table 8. Surprisingly, from the perspective of both parameters’ estimation and MR, violation of Assumption 7 seems to have little
the uniform distribution. ‘ridge’ is from ridge distribution, ‘normal’ is from the normal distribution and ‘uniform’ is from variable estimation ˆ.

\[ SIVE \quad \text{uniform} \quad 0.0137 \quad (0.0189) \quad 0.0154 \quad (0.0190) \quad 0.0214 \quad (0.0251) \quad 0.0291 \quad (0.0302) \quad 0.0352 \quad (0.0359) \quad 0.0444 \quad (0.0466) \quad 0.0538 \quad (0.0578) \quad 0.0611 \quad (0.0618) \quad 0.190 \quad (0.100) \]

Table 7: Simulation results about the absolute bias (standard deviation of the estimates given in the parentheses) of \( \hat{\eta}_{opt} \) relative to \( \eta_{opt} \) and of \( \hat{S}^*(t; \hat{\eta}_{opt}) \) relative to \( S^*(t; \eta_{opt}) \). ‘SIVE’ means the method by maximizing the smoothed instrumental variable estimation \( \hat{S}_{SI}^*(t; \eta) \); ‘SIVE-DR’ means the method by maximizing the smoothed doubly robust instrumental variable estimation \( \hat{S}_{SI}^*(t; \eta) \)

adverse effect on them, which suggests there might be an alternative assumption making the estimation of parameters consistent. But the violation indeed enlarges the bias of the estimation of the optimal survival function as the setting (c) in the table shows.

\[ SIVE \quad 0.0112 \quad (0.0138) \quad 0.0984 \quad (0.0853) \quad 0.105 \quad (0.082) \quad 0.134 \quad (0.0984) \quad 0.159 \quad (0.0732) \quad 0.100 \quad (0.0618) \quad 0.190 \quad (0.100) \]

Table 8: Simulation results about the absolute bias (standard deviation of the estimates given in the parentheses) of \( \hat{\eta}_{opt} \) relative to \( \eta_{opt} \) and of \( \hat{S}^*(t; \hat{\eta}_{opt}) \) relative to \( S^*(t; \eta_{opt}) \). ‘SIVE’ means the method by maximizing the smoothed instrumental variable estimation \( \hat{S}_{SI}^*(t; \eta) \); ‘SIVE-DR’ means the method by maximizing the smoothed doubly robust instrumental variable estimation \( \hat{S}_{SI}^*(t; \eta) \). The column of the ‘generation’ means how to generate \( U \), where ‘ridge’ is from ridge distribution, ‘normal’ is from the normal distribution and ‘uniform’ is from the uniform distribution.