Multiply Robust Off-policy Evaluation and Learning under Truncation by Death

Jianing Chu [1] **Shu Yang** [1] **Wenbin Lu** [1]

Abstract

Typical off-policy evaluation (OPE) and off-policy learning (OPL) are not well-defined problems under "truncation by death", where the outcome of interest is not defined after some events, such as death. The standard OPE no longer yields consistent estimators, and the standard OPL results in suboptimal policies. In this paper, we formulate OPE and OPL using principal stratification under "truncation by death". We propose a *survivor value function* for a subpopulation whose outcomes are always defined regardless of treatment conditions. We establish a novel identification strategy under principal ignorability, and derive the semiparametric efficiency bound of an OPE estimator. Then, we propose multiply robust estimators for OPE and OPL. We show that the proposed estimators are consistent and asymptotically normal even with flexible semi/nonparametric models for nuisance functions approximation. Moreover, under mild rate conditions of nuisance functions approximation, the estimators achieve the semiparametric efficiency bound. Finally, we conduct experiments to demonstrate the empirical performance of the proposed estimators.

1. Introduction

In many real-world applications of personalized decision-making, experimentation and exploration can be costly, risky, or even unethical, such as healthcare (Qian & Murphy, 2011), education (Mandel et al., 2014), and e-commerce (Swaminathan et al., 2017). This motivates the study of off-policy evaluation (OPE) and off-policy learning (OPL) in contextual bandits (Dudík et al., 2014; Wang et al., 2017) and reinforcement learning (Jiang & Li, 2016; Munos et al., 2016; Fujimoto et al., 2019; Kallus & Uehara, 2020). The goal of OPE is to estimate the expected reward of a given policy using historical data generated by a potentially different policy, while OPL aims to derive an optimal policy that maximizes the expected reward based on the available historical data. In contrast, OPE and OPL have also been extensively studied under the counterfactual and potential-outcome framework (Rubin, 2005). They are closely related to a broad body of research on causal inference with observational data: OPE can be viewed as evaluating the expected outcome of patients under a given treatment policy and OPL is to identify the optimal treatment policy that yields the greatest expected outcome in populations, known as the value function (Zhang et al., 2012; Luedtke & Van Der Laan, 2016; Kitagawa & Tetenov, 2018; Athey & Wager, 2021).

A variety of studies have revealed that the treatment effects can vary significantly across different (sub)populations due to covariate shift, unmeasured confounding and other reasons. These issues result in the generalization challenges in both OPE and OPL. Standard OPE methods no longer provide consistent value estimators, and estimated optimal policies by standard OPL methods may be suboptimal to certain (sub)populations (e.g. Lipkovich et al., 2017; VanderWeele et al., 2019; Fang et al., 2022). Consequently, there is a pressing need to develop new OPE and OPL methods that target different (sub)populations. Uehara et al. (2020); Mo et al. (2021); Chu et al. (2023) focus on the generalization of policy to new populations with covariate shift and the use of weighting. Cui & Tchetgen Tchetgen (2021); Qiu et al. (2021) both utilize instrumental variables to address potential unmeasured confounding, and learn the optimal policy in subpopulations with respect to compliance. However, all of these works consider the same outcome across different (sub)populations. In many cases, decision-making goals and outcomes of interest may also be different across various (sub)populations.

For example, in critical care, the primary goal of treatment is to save the lives of seriously ill patients (i.e., decrease mortality). With advancements in medicine, short-term mortality has been reduced in many clinical situations (Wunsch et al., 2010; Guérin et al., 2013). However, surviving critical illness often comes at a cost, as many survivors experience worsened pain, cognition, physical function, and mental health during and after treatment. To account for a patient perspective, the significance of improving "patient-

[1]Department of Statistics, North Carolina State University, Raleigh, NC, USA. Correspondence to: Jianing Chu <jchu3@ncsu.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

important" outcomes, such as functional outcomes and quality of life among survivors, has gained widespread recognition and has been underlined when making treatment decisions (Fried et al., 2002; Iwashyna et al., 2010; Dinglas et al., 2018). Another example is that, in vaccine studies, the primary goal is to estimate the vaccine efficacy, which is a measure of reduction in infection risk for vaccine relative to placebo. However, in infected patients, the outcome of interest may shift to the treatment effect on symptom severity (Paltiel et al., 2021). In the aforementioned examples, the outcome of interest in subpopulations only has a meaningful definition before or after an event occurs. In the first case, once a patient dies, the quality of life is no longer defined, and thus it can only be assessed among survivors. In the second case, symptom severity can only be evaluated after an infection has occurred. The events here can be flexibly defined based on different study purposes. However, to align with the causal inference literature, we term such problems as "truncation by death" (Frangakis & Rubin, 1999). Truncation by death leads outcomes to be undefined and thus cannot be simply treated as missing data problems.

One of the proper ways to deal with truncation by death problems is principal stratification (Frangakis & Rubin, 2002; Imai, 2008; Jiang et al., 2022). Principal stratification partitions the study population into latent subpopulations, known as principal strata. The partition is based on potential values of a post-treatment intermediate variable that lies on the causal pathway between the treatment and the primary outcome (Lipkovich et al., 2022). Throughout the paper, we consider the survival indicator as the intermediate variable. A meaningful causal effect can only be defined in a subpopulation whose potential outcomes are always defined. In other words, subjects in this subpopulation would always survive regardless of the treatments received. This subpopulation is termed as the always-survivor stratum and the causal contrast of this stratum is often referred to as the survivor average causal effect (SACE). In this paper, we name the value function and the optimal policy for the always-survivor stratum as the *survivor value function* and the *survivor-optimal policy*, respectively.

Due to the fundamental problem of the potential outcome formulation, the survivor value function and the survivor-optimal policy are not identifiable without additional assumptions. Standard approaches for OPE, such as the direct method (DM), inverse propensity weighting (IPW) (Horvitz & Thompson, 1952), and doubly robust (DR) (Dudík et al., 2014; Jiang & Li, 2016) estimators, are not consistent for the survivor value function. Furthermore, it's common in medical studies that observations are subject to right censoring due to dropout or administrative censoring before the follow-up visit, which however has seldom been considered in the truncation by death literature. Censoring is a different concept from truncation by death in that the former leads to missing values in survival status and outcomes, while the latter renders outcomes undefined. Therefore, the identification of the survivor value function should be established by incorporating the censoring information.

The contribution of this paper is fourfold.

- First, we provide a novel nonparametric identification strategy for the survivor value function under principal ignorability (Follmann, 2000; Stuart & Jo, 2015), an assumption similar in spirit to treatment ignorability for estimating causal effects in observational studies (Rosenbaum & Rubin, 1983).
- Second, we derive the efficient influence function (EIF) and semiparametric efficiency bound of OPE under truncation by death. The EIF motivates novel estimators based on four models: propensity score, non-censoring probability, survival probability, and conditional mean outcome. For nuisance parameters involved in these models, we propose two different estimation strategies, by using parametric and semi/nonparametric methods, respectively.
- Third, we establish theoretical properties for the proposed estimators. We show the estimators are multiply robust to model-misspecification of nuisance functions. In addition, the estimators achieve the semiparametric efficiency bound under mild rate conditions of nuisance functions approximation.
- Fourth, an OPL method is proposed based on the multiply robust estimators. We establish theoretical properties for both the estimated optimal policy and its associated value estimator. We show the estimated optimal policy in a pre-specified class has the cubic root convergence rate and its associated value function estimator is consistent and asymptotically normal.

2. Preliminaries

In this section, we introduce the setup, formulate the problem, and review existing works.

2.1. Setup

We consider a binary treatment $A \in \{0,1\}$. Let $X \in \mathcal{X} \subseteq \mathbb{R}^p$ denote a vector of pre-treatment covariates. Let $C \in \{0,1\}$ be the censoring indicator. Let $S \in \{0,1\}$ be the survival indicator and Y the non-mortality primary outcome at the follow-up visit. We assume Y is bounded and larger values of Y are preferred by convention. We adopt the potential-outcome framework under the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 2005), and let $C(a)$, $S(a)$, and $Y(a)$ be the potential values of the censoring indicator, survival indicator, and outcome if a subject were to receive treatment condition a ($a = 0, 1$). The observed censoring indicator, survival indicator, and

Table 1. Principal stratification and survival types.

U	Survival type	Description
11	Always-survivor	Subjects who would survive regardless of the treatment conditions
10	Protectable	Subjects who would survive if treated but die otherwise
00	Never-survivor	Subjects who would die regardless of the treatment conditions
01	Defier	Subjects who would survive if untreated but die otherwise

outcome are thus $S = AS(1)+(1-A)S(0)$, $C = AC(1)+(1-A)C(0)$, and $Y = AY(1)+(1-A)Y(0)$. If censoring occurs before the follow-up visit, both S and Y are not observed. If censoring does not occur before the follow-up visit, S is observed, and in this case, Y is observed when a subject is alive ($S = 1$). When a subject is dead at the follow-up visit ($S = 0$), Y is undefined, which is also known as truncation by death. Thus, the observed data are $\{Y_iS_i(1-C_i), S_i(1-C_i), C_i, A_i, X_i, i = 1, \dots, n\}$ and we assume they are independent and identically distributed.

A policy $\pi : \mathcal{X} \to [0, 1]$ is a map from covariates to the probability of assigning treatment 1. If a policy π were implemented in the population, without censoring and truncation by death, then the population mean outcome, known as the value function, would be

$$V(\pi) = \mathbb{E}\left[Y(1)\pi(X) + Y(0)\{1-\pi(X)\}\right].$$

And the optimal policy π^* in a policy class Π is the one that maximizes the value function:

$$\pi^* = \underset{\pi\in\Pi}{\operatorname{argmax}}\, V(\pi).$$

2.2. Problem Formulation

Following Frangakis & Rubin (2002), we use the joint potential values of the survival indicator to define the principal stratification variable, $U = \{(S(1), S(0)\}$. For the ease of exposition, we simplify $\{(S(1), S(0)\}$ as $S(1)S(0)$ throughout the paper. There exists a one-to-one mapping between the survival type and the principal stratification variable, which is given in Table 1. The outcomes are only defined in the survivors, which are a mixture of different principal strata. However, a meaningful value function can only be defined in the always-survivor stratum whose potential outcomes are always defined and we term it as the survivor value function,

$$V_{11}(\pi) = \mathbb{E}\left[Y(1)\pi(X) + Y(0)\{1-\pi(X)\} \mid U = 11\right].$$

And we define the survivor-optimal policy as the one that maximizes the survivor value function:

$$\pi^*_{11} = \underset{\pi\in\Pi}{\operatorname{argmax}}\, V_{11}(\pi).$$

Unless otherwise specified, we will omit the subscript and use $V(\pi)$ and π^* to denote a survivor value function and a survivor-optimal policy, instead of a general value function and a general optimal policy, in the remaining paper. Our first goal is OPE; i.e., estimating $V(\pi)$ for a given policy π using the historical data $\{Y_iS_i(1-C_i), S_i(1-C_i), C_i, A_i, X_i, i = 1, \dots, n\}$. Our second goal is OPL; i.e., estimating the survivor-optimal policy π^*.

2.3. Standard OPE and OPL

We review three types of standard approaches to estimate $V(\pi)$ when there is no censoring and truncation by death (only one stratum, always-survivor, exists and the survivor value function is equal to the general value function).

(i) Direct method (DM) estimates the condition mean outcome functions $\mu_a(x) = \mathbb{E}[Y(a) \mid X = x]$ and $\widehat{V}^{\mathrm{DM}}(\pi) = \mathbb{P}_n[\widehat{\mu}_1(X)\pi(X) + \widehat{\mu}_0(X)\{1-\pi(X)\}]$, where $\mathbb{P}_n[h(X)] = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$ for any given function $h(X)$. DM estimators are known to be sensitive against model misspecification with regards to $\mu_a(x)$.

(ii) Inverse probability weighting (IPW) estimator $\widehat{V}^{\mathrm{IPW}}(\pi) = \mathbb{P}_n\left[\frac{\pi_A(X)}{\widehat{\varphi}_A(X)}Y\right]$, where $\pi_a(x) = a\pi(x) + (1-a)\{1-\pi(x)\}$, $\widehat{\varphi}_a(x)$ is an approximation of $\varphi_a(x) = a\varphi(x) + (1-a)\{1-\varphi(x)\}$, where $\varphi(x) = P(A = 1|X = x)$ is the propensity score, which is also known as the behavior policy that was used to generate the historical data. This estimator is unbiased when the propensity score is known but it often suffers from high variance.

(iii) Doubly robust (DB) estimator $\widehat{V}^{\mathrm{DR}}(\pi) = \widehat{V}^{\mathrm{DM}}(\pi) + \mathbb{P}_n\left[\frac{\pi_A(X)}{\widehat{\varphi}_A(X)}\{Y - \widehat{\mu}_A(X)\}\right]$. This estimator is consistent if either the conditional mean outcome model or propensity score model is correctly specified.

These estimators are also used for OPL (Zhang et al., 2013; Luckett et al., 2019; Athey & Wager, 2021). However, when utilizing these standard methods, one implicitly makes the "missing at random" assumption, which posits that the outcomes of non-survivors are missing, and their conditional distributions given covariates and treatment are equivalent to those of survivors. Nevertheless, the outcomes of non-survivors are not well-defined, and even if they were, they are likely to differ from those of survivors because death implies the deterioration of underlying health conditions. Consequently, standard OPE and OPL methods are not suitable approaches to deal with the truncation by death problems.

3. Identification, EIF, and Efficiency Bound

We first make the following identification assumptions.

Assumption 3.1. (i) $A \perp\!\!\!\perp \{C(a), S(a)\} \mid X$;
(ii) $A \perp\!\!\!\perp Y(a) \mid S(a) = 1, X$, for $a = 0, 1$.

Assumption 3.1 rules out unmeasured confounding between the treatment and potential values of the censoring indicator, survival indicator and outcome. It holds by the design of a randomized experiment. It also holds if the observed covariates include all the confounders that affect the treatment as well as the censoring indicator, survival indicator, and outcome.

Assumption 3.2. (i) $\mathbb{E}[S(a) \mid C(a) = 0, X] = \mathbb{E}[S(a) \mid X]$; (ii) $\mathbb{E}[Y(a) \mid S(a) = 1, C(a) = 0, X] = \mathbb{E}[Y(a) \mid S(a) = 1, X]$, for $a = 0, 1$.

With Assumption 3.1, Assumption 3.2 implies that the survival probability for a subject does not vary across censoring and non-censoring groups. And the conditional mean outcomes of survivors are identical in both censoring and non-censoring groups.

Assumption 3.3. $S(1) \geq S(0)$ almost surely.

Assumption 3.3 is known as the monotonicity assumption (Sommer & Zeger, 1991; Follmann, 2006), which implies that the treatment has a non-negative impact on the survival of all subjects, which rules out the defier stratum ($U = 01$). It is often plausible in observational studies since providers cannot assign inferior treatment to patients. Then, from Table 1, the observed survivors are a mixture of the always-survivor stratum ($U = 11$) and the protectable stratum ($U = 10$).

Assumption 3.4. $\mathbb{E}[Y(1) \mid U = 11, X] = \mathbb{E}[Y(1) \mid U = 10, X]$.

Assumption 3.4, known as the principal ignorability assumption, is widely used in principal causal analysis literature (Follmann, 2000; Stuart & Jo, 2015). Under Assumptions 3.1 and 3.3, Assumption 3.4 is equivalent to

$$\begin{aligned} & \mathbb{E}[Y(1)|U = 11, A = 1, S = 1, X] \\ = \; & \mathbb{E}[Y(1)|U = 10, A = 1, S = 1, X]. \end{aligned} \tag{1}$$

(1) implies that the expectations of the potential outcome $Y(1)$ conditional on covariates are identical in both the always-survivor stratum and the protectable stratum. Further with Assumption 3.2, they simplify to the observable conditional expectation $\mathbb{E}[Y \mid A = 1, C = 0, S = 1, X]$. This assumption may not be easily justified by prior knowledge and may be violated in some applications. We provide a sensitivity analysis technique for the potential violation of this assumption in Appendix C.

Define the propensity score $\varphi(x) = P(A = 1|X = x)$ and let $\varphi_a(x) = a\varphi(x) + (1 - a)\{1 - \varphi(x)\}$. Define the non-censoring probability $K_a(x) = P(C = 0 \mid A = a, X = x)$, the observed survival probability $p_a(x) = P(S = 1 \mid A = a, C = 0, X = x)$, and the observed conditional mean outcome $\mu_a(x) = \mathbb{E}[Y \mid A = a, C = 0, S = 1, X = x]$, for $a = 0, 1$. Let $p_a = \mathbb{E}[p_a(X)]$. Suppose the nuisance functions satisfy the following assumption.

Assumption 3.5. $\{\varphi_a(x), K_a(x), p_a(x)\} > 0$, $|\mu_a(x)| < L$, for some $L > 0$, and $x \in \mathcal{X}$.

The assumption of positivity for $\varphi_a(x)$ implies that every individual has a non-zero probability of receiving or not receiving the treatment. Similarly, the positivity assumption for $K_a(x)$ implies that each individual has a non-zero probability of being censored or not censored, while the positivity assumption for $p_a(x)$ implies that each individual has a non-zero probability of surviving or not surviving. In practice, these assumptions are generally expected to hold. Furthermore, the bounded conditional outcome model $\mu_a(x)$ is reasonable because outcomes, such as quality of life, are typically bounded in real-world applications.

The following theorem provides a nonparametric identification formula for $V(\pi)$.

Theorem 3.6. *Let Π be a policy class. Under Assumptions 3.1–3.5, for any given policy $\pi \in \Pi$, the survivor value function $V(\pi)$ is identified,*

$$\begin{aligned} V(\pi) = \mathbb{E}\Bigg[\frac{p_0(X)}{p_0}\Bigg\{ & \frac{A}{\varphi_1(X)}\frac{1-C}{K_1(X)}\frac{S}{p_1(X)}Y\pi(X) \\ & + \frac{1-A}{\varphi_0(X)}\frac{1-C}{K_0(X)}\frac{S}{p_0(X)}Y\{1-\pi(X)\}\Bigg\}\Bigg]. \end{aligned}$$

Based on the identification formula provided in Theorem 3.6, we derive the EIF and the semiparametric efficiency bound for $V(\pi)$.

Define

$$\psi_{S(a)} = \frac{\mathbb{I}(A = a)\mathbb{I}(C = 0)\{S - p_a(X)\}}{\varphi_a(X)K_a(X)} + p_a(X), \tag{2}$$

$$\begin{aligned} \psi_{Y(a)S(a)} = & \frac{\mathbb{I}(A = a)\mathbb{I}(C = 0)\{YS - \mu_a(X)p_a(X)\}}{\varphi_a(X)K_a(X)} \\ & + \mu_a(X)p_a(X). \end{aligned} \tag{3}$$

Theorem 3.7. *Suppose $V(\pi)$ is identified in Theorem 3.6. The EIF for $V(\pi)$ is*

$$\nu_\pi = \{\phi_\pi - V(\pi)\psi_{S(0)}\}/p_0,$$

and the semiparametric efficiency bound for $V(\pi)$ is

$$\Upsilon(\pi) = \mathbb{E}\left[\{\phi_\pi - V(\pi)\psi_{S(0)}\}/p_0\right]^2,$$

where $\phi_\pi = \left[\mu_1(X)\left\{\psi_{S(0)} - \frac{p_0(X)}{p_1(X)}\psi_{S(1)}\right\} + \frac{p_0(X)}{p_1(X)}\psi_{Y(1)S(1)}\right]\pi(X) + \psi_{Y(0)S(0)}\{1-\pi(X)\}$.

4. Multiply Robust OPE

The EIF ν_π motivates the following estimator for $V(\pi)$:

$$\widehat{V}(\pi) = \frac{\mathbb{P}_n(\widehat{\phi}_\pi)}{\mathbb{P}_n\{\widehat{\psi}_{S(0)}\}}, \tag{4}$$

where $\widehat{\phi}_\pi$ and $\widehat{\psi}_{S(0)}$ are the estimators for ϕ_π and $\psi_{S(0)}$, respectively. From (2) and (3), we can first build estimators for $\varphi_a(X)$, $K_a(X)$, and $p_a(X)$, and then combine them to get estimators for $\psi_{S(a)}$ and $\psi_{Y(a)S(a)}$. Finally, we can construct estimators for ϕ_π and $V(\pi)$.

We first consider the case when nuisance functions are estimated parametrically. Let $\varphi_a(x;\alpha)$, $K_a(x;\eta)$, $p_a(x;\gamma)$ and $\mu_a(x;\zeta)$ be working parametric models for $\varphi_a(x)$, $K_a(x)$, $p_a(x)$ and $\mu_a(x)$, respectively. Based on the maximum likelihood estimation, we obtain estimators $\widehat{\alpha}$, $\widehat{\eta}$, $\widehat{\gamma}$, and $\widehat{\zeta}$. Let α^*, η^*, γ^*, and ζ^* be the limits of $\widehat{\alpha}$, $\widehat{\eta}$, $\widehat{\gamma}$, and $\widehat{\zeta}$, respectively. We use $\mathcal{M}$ with subscripts "ps", "cs", "sp", and "om" to denote models with the correct specification of the propensity score, non-censoring probability, survival probability, and conditional mean outcome, respectively. For example, under $\mathcal{M}_{\rm ps}$, we have $\varphi_a(x;\alpha^*) = \varphi_a(x)$. In addition, we use "+" in the subscript to indicate that multiple nuisance functions are correctly specified. For example, $\mathcal{M}_{\rm ps+cs}$ denotes models with correctly specified $\varphi_a(x;\alpha)$ and $K_a(x;\eta)$. We also use the union notation to denote the correct specification of at least one nuisance function, for example, $\mathcal{M}_{\rm ps+cs} \cup \mathcal{M}_{\rm sp}$ denotes models with correctly specified $\{\varphi(x;\alpha), K_a(x;\eta)\}$ or $p_a(x;\gamma)$.

For $\psi_{S(a)}$ and $\psi_{Y(a)S(a)}$, we have the estimators $\widehat{\psi}_{S(a)} = \frac{\mathbb{I}(A=a)\mathbb{I}(C=0)\{S-p_a(X;\widehat{\gamma})\}}{\varphi_a(X;\widehat{\alpha})K_a(X;\widehat{\eta})} + p_a(X;\widehat{\gamma})$, and $\widehat{\psi}_{Y(a)S(a)} = \frac{\mathbb{I}(A=a)\mathbb{I}(C=0)\{YS-\mu_a(X;\widehat{\zeta})p_a(X;\widehat{\gamma})\}}{\varphi_a(X;\widehat{\alpha})K_a(X;\widehat{\eta})} + \mu_a(X;\widehat{\zeta})p_a(X;\widehat{\gamma})$. Then for ϕ_π, we have the estimator $\widehat{\phi}_\pi = \left[\mu_1(X;\widehat{\zeta})\left\{\widehat{\psi}_{S(0)} - \frac{p_0(X;\widehat{\gamma})}{p_1(X;\widehat{\gamma})}\widehat{\psi}_{S(1)}\right\} + \frac{p_0(X;\widehat{\gamma})}{p_1(X;\widehat{\gamma})}\widehat{\psi}_{Y(1)S(1)}\right]$ $\times\pi(X) + \widehat{\psi}_{Y(0)S(0)} \times \{1-\pi(X)\}$. Plugging $\widehat{\phi}_\pi$ and $\widehat{\psi}_{S(0)}$ into (4), we have a multiply robust (MR) estimator and we denote it as $\widehat{V}^{\rm MR}(\pi)$.

Assumption 4.1. $\{\varphi_a(x;\widehat{\alpha}), K_a(x;\widehat{\eta}), p_a(x;\widehat{\gamma})\} > 0$, $|\mu_a(x;\widehat{\zeta})| < L$, for some $L > 0$, and $x \in \mathcal{X}$.

Theorem 4.2. *Suppose Assumptions 3.1–3.5 and 4.1 hold, $\widehat{V}^{\rm MR}(\pi)$ is multiply robust in the sense that it is consistent for $V(\pi)$ under $\mathcal{M}_{\rm ps+cs+sp} \cup \mathcal{M}_{\rm ps+cs+om} \cup \mathcal{M}_{\rm sp+om}$. Moreover, under $\mathcal{M}_{\rm ps+cs+sp+om}$, $\widehat{V}^{\rm MR}(\pi)$ has the influence function ν_π and achieves the semiparametric efficiency bound $\Upsilon(\pi)$.*

Remark 4.3. Though in $\mathcal{M}_{\rm ps+cs+sp}$ and $\mathcal{M}_{\rm ps+cs+om}$, we require both of the propensity score and the non-censoring probability to be correctly specified, we actually only need the correct specification of the product of these two nuisance functions, that is $\varphi_a(x)K_a(x) = P(A=a, C=0|X=x)$, to achieve the consistency. Therefore, the estimator is triply robust in the sense it is consistent for $V(\pi)$ if any two of $P(A=a,C=0|X=x)$, $p_a(x)$, and $\mu_a(x)$ are correctly specified.

Alternatively, the nuisance functions can also be estimated using nonparametric models. We denote them as $\widehat{\varphi}_a(x)$, $\widehat{K}_a(x)$, $\widehat{p}_a(x)$, and $\widehat{\mu}_a(x)$. For a vector z, we use $\|z\|_2 = (z^T z)^{1/2}$ to denote its Euclidean norm. For a function $f(Z)$, where Z is a generic random variable, we define its L_2-norm as $\|f(Z)\| = \{\int f(z)^2 dP(z)\}^{1/2}$.

Assumption 4.4. (i) $\{\widehat{\varphi}_a(x), \{\widehat{K}_a(x), \{\widehat{p}_a(x)\} > 0$, $|\widehat{\mu}_a(x)| < L$, for some $L > 0$, and $x \in \mathcal{X}$; (ii) $\{\widehat{\varphi}_a(x), \widehat{K}_a(x), \widehat{p}_a(x), \widehat{\mu}_a(x)\}$ and $\{\varphi_a(x), K_a(x), p_a(x), \mu_a(x)\}$ are in a Donsker class, and $\{\widehat{\varphi}_a(x), \widehat{K}_a(x), \widehat{p}_a(x), \widehat{\mu}_a(x)\} \xrightarrow{p} \{\varphi_a(x), K_a(x), p_a(x), \mu_a(x)\}$ for $x \in \mathcal{X}$; (iii) $\|\widehat{g}(X) - g(X)\|\|\widehat{h}(X) - h(X)\| = o_p(n^{-1/2})$, for any $g \neq h \in (\varphi_a \times K_a, p_a, \mu_a)$.

Theorem 4.5. *Suppose that Assumptions 3.1–3.5 and 4.4 hold. $\widehat{V}^{\rm MR}(\pi)$ is asymptotically normal, has the influence function ν_π, and achieves the semiparametric efficiency bound $\Upsilon(\pi)$.*

Remark 4.6. Assumption 4.4 is analogous to those for double machine learning estimation of average causal effects (e.g. Kennedy, 2016; Farrell et al., 2021). The Donsker class assumption can be relaxed by applying the cross-fitting technique (Zheng & Laan, 2011; Chernozhukov et al., 2018) and we summarize the procedure in Algorithm 1.

5. From Robust OPE to OPL

In this section, we propose an OPL method based on the MR estimator $\widehat{V}^{\rm MR}(\pi)$ to estimate the survivor-optimal policy, which is defined as $\pi^* = \operatorname{argmax}_{\pi\in\Pi} V(\pi)$. A natural estimator for the survivor-optimal policy would be $\widehat{\pi} = \operatorname{argmax}_{\pi\in\Pi} \widehat{V}^{\rm MR}(\pi)$. In many applications such as clinical practice, it may be desirable to consider a policy class indexed by a vector of parameters β for feasibility and interpretability. We denote such a policy class as Π_β and its element as $\pi(x;\beta)$. For example, we can consider a linear policy class $\Pi_\beta = \{\pi(x;\beta) = \mathbb{I}(\beta^T\tilde{x} > 0) : \beta \in \mathbb{R}^{p+1}, \|\beta\|_2 = 1\}$, where $\tilde{x} = (1, x^T)^T$. Given a linear policy $\pi(x;\beta) \in \Pi_\beta$, we use a shorthand to write its associated survivor value function $V(\pi)$ as $V(\beta)$. Let

Algorithm 1 Multiply Robust OPE with Cross-fitting

Input: The evaluation policy π.
Take a ξ-fold random partition $(I_k)_{k=1}^{\xi}$ of observation indices $\{1, \ldots, n\}$ such that the size of each fold I_k is $n_k = n/\xi$.
For each $k \in \{1, \ldots, \xi\}$, define $I_k^c = \{1, \ldots, n\} \backslash I_k$.
Define $(\mathcal{D}_k)_{k=1}^{\xi}$ with $\mathcal{D}_k = \{Y_i S_i (1 - C_i), S_i (1 - C_i), C_i, A_i, X_i)\}_{i \in I_k^c}$.
for $k = 1$ **to** ξ **do**
  Construct estimators $\widehat{\varphi}_{a,k}(x)$, $\widehat{K}_{a,k}(x)$, $\widehat{p}_{a,k}(x)$,and $\widehat{\mu}_{a,k}(x)$ using $\mathcal{D}_k$.
  Construct an estimator $\widehat{V}_k(\pi)$ defined as (4).
end for
Construct an estimator $\widehat{V}^{\mathrm{MR}}(\pi)$ by taking the average of $\widehat{V}_k(\pi)$, for $k = 1, \ldots, \xi$, i.e., $\widehat{V}^{\mathrm{MR}}(\pi) = \frac{1}{\xi} \sum_{k=1}^{\xi} \widehat{V}_k(\pi)$.

$\beta^* = \mathrm{argmax}_{\beta : \|\beta\|_2 = 1} V(\beta)$. Then, the survivor-optimal linear policy is $\pi(x; \beta^*)$. We can establish the MR estimator for a given linear policy $\pi(x; \beta)$ and we denote it as $\widehat{V}^{\mathrm{MR}}(\beta)$. We can obtain the estimated survivor-optimal linear policy, denoted as $\pi(x; \widehat{\beta})$ by directly maximizing $\widehat{V}^{\mathrm{MR}}(\beta)$, i.e., $\widehat{\beta} = \mathrm{argmax}_{\beta : \|\beta\|_2 = 1} \widehat{V}^{\mathrm{MR}}(\beta)$.

We first impose the following regularity conditions.

Assumption 5.1. (i) The survivor value function $V(\beta)$ is twice continuously differentiable at a neighborhood of β^*; (ii) There exist some constants $\delta_0 > 0$ such that $P(|\tilde{X}^T \beta^*| \leq \delta) = O(\delta)$, where the big-O term is uniform in $0 < \delta \leq \delta_0$.

Assumption 5.1(i) is a standard regularity condition used to establish the uniform convergence results; Assumption 5.1(ii) excludes the situation with $P(\tilde{X}^T \beta^* = 0) > 0$ and ensures the true survivor-optimal linear policy is uniquely defined, known as the margin condition, which is often assumed to derive a sharp convergence rate for the value function under the estimated optimal policy (e.g. Luedtke & Van Der Laan, 2016).

We establish the following Lemma 5.2 and Theorem 5.3 when nuisance functions are estimated nonparametrically. Similar results for parametric estimation are provided in Appendix B.

Lemma 5.2. *Suppose that Assumptions 3.1–3.5, 4.4, and 5.1 hold, we have* $n^{1/3} \|\widehat{\beta} - \beta^*\|_2 = O_p(1)$.

Theorem 5.3. *Suppose that Assumptions 3.1–3.5, 4.4, and 5.1 hold, we have* $\sqrt{n} \left\{ \widehat{V}^{\mathrm{MR}}(\widehat{\beta}) - V(\beta^*) \right\} \xrightarrow{d} \mathcal{N}(0, \Upsilon(\pi(x; \beta^*)))$.

In practice, decision-makers may seek to implement the learned survivor-optimal policy in two ways: (i) across the entire population, or (ii) exclusively within the always-survivor stratum. In the case of (i), it becomes essential to strike a balance between the overall survivor probability of the entire population and the primary outcome of survivors. This balance can be attained in studies where we have access to individual survival time, denoted by T. Let $T(a)$ be the potential survival time of an individual if he/she were given treatment a. Under a given policy π, the population-level t-year survival probability (t is a user-specified time point) is given by

$$B(t; \pi) = \mathbb{E}[P(\{T(1)\pi(X) + T(0)\{1 - \pi(X)\}\} > t \mid X)].$$

The estimation of $B(t; \pi)$ has been well-studied in the existing literature (eg. Jiang et al., 2017) and we denote its estimator as $\widehat{B}(t; \pi)$. To balance the tradeoff between the survival probability of the entire population and the the primary outcome of survivors in OPL, we can target specific levels of survival probability, for example, $\widehat{B}(t, \pi) \geq \Delta$, where Δ is a user-specified probability. Then the OPL can be formulated as a constrained optimization problem:

$$\underset{\pi \in \Pi}{\mathrm{maximize}} \quad \widehat{V}^{\mathrm{MR}}(\pi), \quad \text{s.t.} \quad \widehat{B}(t; \pi) \geq \Delta.$$

Such a constrained OPL problem has been studied in the literature (e.g. Zhou et al., 2021).

In the case of (ii), it is important to recall the monotonicity assumption discussed in Section 3, which is often reasonable in practical scenarios as healthcare providers are typically able to administer superior treatment to patients. Under this assumption, we can rule out the defier stratum and are left with three strata: always-survivor, protectable, and never-survivor. In practice, for a patient with covariates x, we use the survival probability model $\hat{p}_a(x)$ to predict which stratum he/she belongs to. For instance, if the estimated survival probabilities, $\hat{p}_1(x)$ and $\hat{p}_0(x)$, are both above a specified threshold δ_1, the patient will be classified as belonging to the always-survivor stratum. Conversely, if both $\hat{p}_1(x)$ and $\hat{p}_0(x)$ fall below a threshold δ_2, the patient will be assigned to the never-survivor stratum. In cases where neither threshold condition is met, the patient will be categorized as part of the protectable stratum.

For the always-survivor stratum, we implement the learned survivor-optimal policy. For the never-survivor stratum, we can derive a policy by maximizing $\widehat{B}(t; \pi)$ introduced above. Finally, for the protectable stratum, the optimal policy is always to assign treatment since subjects will survive if treated but die otherwise.

6. Experiments

6.1. Synthetic Scenarios

We generate the baseline covariates $X = (X_1, X_2, X_3)^T$ from the following distribution: $X_1 \sim$ Bernoulli(0.5),

$$(X_2, X_3)^T | X_1 = 1 \sim N((1, -1)^T, \Sigma_1),$$
$$(X_2, X_3)^T | X_1 = 0 \sim N((-1, 1)^T, \Sigma_2),$$

$$\Sigma_1 = \begin{pmatrix} 1 & -0.25 \\ -0.25 & 1 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 1 & -0.3 \\ -0.3 & 1 \end{pmatrix}.$$

The treatment is generated from $A \sim$ Bernoulli$\{\varphi(X)\}$, and logit$\{\varphi(X)\} = 0.5X_1 + 0.5X_2 + 0.2X_3^2$. The censoring indicator is generated from $C \sim$ Bernoulli$\{K_A(X)\}$, and we consider two models for $K_A(X)$: (i) logit$\{K_A(X)\} = -X_1 - X_2 - 0.5X_3^2 - A - 1$; (ii) logit$\{K_A(X)\} = -X_1 - X_2 - 0.5X_3^2 - A + 0.5$. Models (i) and (ii) result in censoring rates 15% and 30% on the population level, respectively. The survival indicator is generated from $S \sim$ Bernoulli$\{p_A(X)\}$, and logit$\{p_A(X)\} = -2X_1 - 0.5X_2^2 + X_3 + A(4X_1 + 1)$. The outcome is generated from the model $Y = \mu_A(X) + \epsilon$, where

$$\mu_A(X) = \exp\{0.2X_1 - 0.2X_2 + 0.1X_3 + 1.5A \cdot \text{sign}(X_1 - 2X_2^2 + X_3 > 0)\}, \quad (5)$$

and ϵ is generated from a normal distribution with mean 0 and variance 0.25.

We consider five different approaches of nuisance functions approximation for the proposed MR estimator:

(I) $\varphi(x), K_a(x)$, and $p_a(x)$ are estimated by correctly specified logistic regression models and $\mu_a(x)$ is estimated by a correctly specified log-linear regression model.

(II) $p_a(x)$ and $\mu_a(x)$ are estimated by correctly specified models the same as (I). $\varphi(x)$ and $K_a(x)$ are estimated by misspecified logistic regression models.

(III) $\varphi(x)$, $K_a(x)$, and $\mu_a(x)$ are estimated by correctly specified models the same as (I). $p_a(x)$ is estimated by a misspecified logistic regression model.

(IV) $\varphi(x)$, $K_a(x)$, and $p_a(x)$ are estimated by correctly specified models the same as (I). $\mu_a(x)$ is estimated by a misspecified linear regression model.

(V) $\varphi(x)$, $K_a(x)$, and $p_a(x)$ are estimated by generalized additive models (GAMs). $\mu_a(x)$ is estimated by a random forest (RF) model.

OPE: We construct three different evaluation policies as mixtures of a deterministic policy $\pi_d = \mathbb{I}(X_1 - 2X_2^2 + X_3 > 0)$ and the uniform random policy π_u by changing a mixture parameter w, i.e., $\pi = w\pi_d + (1-w)\pi_u$. The candidates of the mixture parameter w are $\{0.7, 0.4, 0.0\}$. The true survivor value function for each evaluation policy is obtained by generating a large sample using true nuisance functions and applying the empirical version of the identification formula provided in Theorem 3.6. We compare five MR methods, (I)-(V), with DM, IPW, and DR methods. For DM, IPW, and DR methods, we estimate the conditional mean outcome by RF and the propensity score by GAM. We consider samples with size $n = 1000, 2000$. For each setting, we conduct 500 replications. The root-mean-square error (RMSE) and the standard deviation (SD) results with $n = 2000$ are reported in Table 2 and the results with $n = 1000$ are reported in Appendix D. The MR estimators all outperform the standard estimators including DM, IPW, and DR estimators. Meanwhile, the RMSE and SD are very close to each other for all MR estimators, which shows the multiple robustness property that we established in the theorems.

OPL: We consider all policies in the policy class $\Pi_\beta = \{\pi(x;\beta) = \mathbb{I}(\beta^T\tilde{x} > 0) : \beta \in \mathbb{R}^4, \|\beta\|_2 = 1\}$ as candidates and learn the survivor-optimal linear policy by maximizing the MR estimator $\widehat{V}^{\text{MR}}(\beta)$ and the DR estimator $\widehat{V}^{\text{DR}}(\beta)$ over β. The MR and DR estimators are constructed in the same way as in OPE experiment. Since the value estimators are non-smooth and non-convex in β, we use the genetic algorithm to maximize the estimators with respect to β and obtain $\widehat{\beta}$. To evaluate and compare the performance of estimated survivor-optimal linear policies obtained by different methods, we compute the corresponding survivor value functions and percentages of making correct decisions (PCD) for the always-survivor stratum. Specifically, we first generate baseline covariates for a large sample with size $N = 10^5$ and then generate $S(0)$ for each subject in the sample. Always-survivors are those with $S(0) = 1$ under the monotonicity assumption. Let $N_s = \sum_{i=1}^N \mathbb{I}\{S(0)_i = 1\}$. The survivor value function for an estimated policy $\pi(x;\widehat{\beta})$ is computed by

$$V(\widehat{\beta}) = N_s^{-1}\sum_{i=1}^N \mathbb{I}\{S(0)_i = 1\}\Big[\mu_1(X_i)\pi(X_i;\widehat{\beta}) + \mu_0(X_i)\{1 - \pi(X_i;\widehat{\beta})\}\Big],$$

and its associated PCD is computed by

$$N_s^{-1}\sum_{i=1}^N \mathbb{I}\{S(0)_i = 1\}|\pi(X_i;\widehat{\beta}) - \pi(X_i;\beta^*)|,$$

where the true survivor-optimal linear policy $\pi(x;\beta^*)$ is obtained by maximizing $V(\beta)$ over β. We report the value and PCD results for the policies obtained by MR and DR methods in Figure 1. We have the following observations. For censoring rate 30%, MR methods all have good and comparable performance in terms of values and PCDs. In addition, as the sample size increases, the means of values become closer to the true optimal survivor value function,

Table 2. OPE results. (a) $0.7\pi_d + 0.3\pi_u$, (b) $0.4\pi_d + 0.6\pi_u$, (c) $0.0\pi_d + 1.0\pi_u$.

censoring rate:15%

	MR-I		MR-II		MR-III		MR-IV		MR-V		DM		IPW		DR	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
(a)	0.111	0.110	0.117	0.114	0.109	0.109	0.112	0.112	0.115	0.114	0.872	0.078	0.660	0.141	0.514	0.090
(b)	0.096	0.096	0.101	0.098	0.094	0.094	0.098	0.098	0.100	0.099	0.645	0.068	0.553	0.114	0.447	0.077
(c)	0.077	0.076	0.080	0.078	0.075	0.075	0.084	0.083	0.083	0.083	0.345	0.057	0.412	0.081	0.357	0.062

censoring rate: 30%

	MR-I		MR-II		MR-III		MR-IV		MR-V		DM		IPW		DR	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
(a)	0.138	0.138	0.148	0.141	0.139	0.139	0.142	0.142	0.149	0.148	0.887	0.093	0.713	0.125	0.543	0.112
(b)	0.120	0.120	0.128	0.122	0.121	0.121	0.125	0.125	0.131	0.131	0.666	0.082	0.594	0.105	0.472	0.096
(c)	0.096	0.096	0.102	0.097	0.097	0.097	0.109	0.108	0.115	0.114	0.372	0.068	0.436	0.082	0.377	0.077

PCDs get close to 1, and the standard deviations of values and PCDs become smaller. When the censoring rate decreases to 15%, the performance becomes even better. The values and PCDs become larger, and their standard deviations become smaller. However, the DR method has poor performance under all the settings: the means of values are much smaller than the true optimal survivor value function and PCDs are much smaller than 1. This implies that the estimated survivor-optimal linear policy obtained using the DR method may not be promising for the always-survivors.

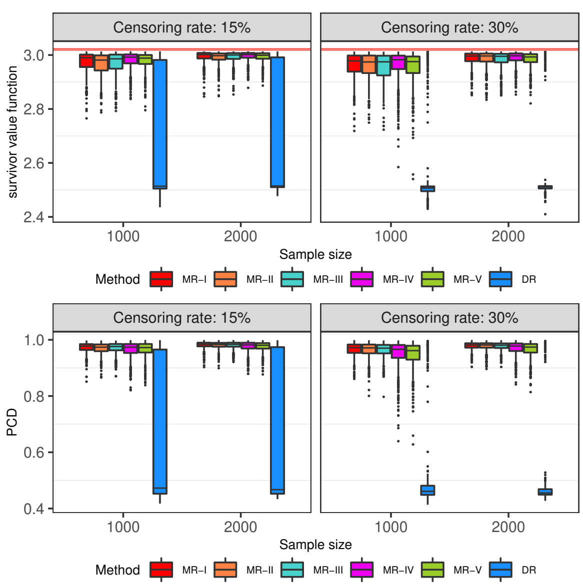

Figure 1. The value and PCD results of estimated survivor-optimal linear policies under different censoring rates. The red line is the true optimal survivor value function.

Next, we study the inference results of $\widehat{V}(\widehat{\beta})$ obtained by MR and DR methods. The standard errors (SE) are estimated using the bootstrap method. We report the mean and standard deviation of $\widehat{V}(\widehat{\beta})$, the mean of estimated standard errors, and the empirical coverage probability (CP) of 95% Wald-type confidence intervals for the true optimal survivor value function $V(\beta^*) = 3.02$. The results are summarized in Table 3. We have the following observations. For each MR method, the value estimator is nearly unbiased. The mean of estimated standard errors is close to the standard deviation of the estimators, and the empirical CP of 95% confidence intervals is close to the nominal level. The standard deviation under the censoring rate 15% is smaller than that under the censoring rate 30%. As the sample size increases, the mean gets closer to the true optimal survivor value function and the standard deviation becomes smaller. However, the DR estimator has a large bias and the empirical CP of 95% confidence intervals is 0.

Table 3. Inference results of $\widehat{V}(\widehat{\beta})$.

Method	MR-I		MR-II		MR-III		MR-IV		MR-V		DR	
	censoring rate: 15%											
$n(\times 10^3)$	1	2	1	2	1	2	1	2	1	2	1	2
Mean	3.05	3.04	3.08	3.06	3.03	3.02	3.08	3.05	3.03	3.02	2.27	2.24
SD	0.19	0.13	0.20	0.14	0.19	0.13	0.18	0.13	0.20	0.14	0.12	0.09
SE	0.18	0.13	0.19	0.13	0.19	0.13	0.19	0.13	0.20	0.14	0.12	0.09
CP	94.2	94.8	93.2	93.2	94.6	95.4	94.2	93.8	94.0	96.0	0	0
	censoring rate: 30%											
$n(\times 10^3)$	1	2	1	2	1	2	1	2	1	2	1	2
Mean	3.08	3.04	3.13	3.09	3.05	3.02	3.10	3.05	3.05	3.03	2.29	2.26
SD	0.22	0.16	0.22	0.17	0.22	0.16	0.22	0.16	0.23	0.17	0.15	0.11
SE	0.23	0.16	0.23	0.16	0.23	0.16	0.24	0.16	0.25	0.17	0.14	0.10
CP	96.2	94.6	94.4	93.4	95.8	94.6	96.8	94.4	96.4	96.0	0.02	0

6.2. Real Data Application

We illustrate the proposed methods using an application to data from the MIMIC-III clinical database (Goldberger et al., 2000; Johnson et al., 2016; 2019). The MIMIC-III database comprises de-identified health-related data associated with over 40,000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. The dataset contains time-stamped physiological measurements, lab values, and intake/output events. In this application, we include patients fulfilling the international consensus Sepsis-3 criteria (Singer et al., 2016). Sepsis is a life-threatening condition and causes organ failure (Rhee et al., 2017). The Sequential Organ Failure Assessment (SOFA) score is to numerically quantifies the severity of a person's organ dysfunction. We select those severely ill patients with SOFA scores larger than 12 at baseline as the study sample. There are 798 subjects in total, among which 496 patients were treated with mechanical ventilation $(A = 1)$, while the rest were not treated with mechanical ventilation $(A = 0)$. We consider 48 hours as the follow-up time point and the SOFA score at 48 hours is the outcome of interest (smaller values are preferred). The intermediate variable is the 48-hour mortality. If a patient survived the first 48 hours after admission to ICU, then $S = 1$; otherwise, $S = 0$. Furthermore, if a patient did not die within the first 48 hours but the outcome is missing at the follow-up time point, we consider it is censored $(C = 1)$. The censoring and survival information of the dataset is provided in Table 4. We are interested to estimate the survivor-optimal linear policy in this case.

Table 4. Censoring and survival information under two treatments.

$A=1$	$C=0$	$C=1$	$A=0$	$C=0$	$C=1$
$S=0$	199	53	$S=0$	41	132
$S=1$	244		$S=1$	129	

We consider $p = 7$ baseline covariates: age (years), admission weights (kg), admission temperature (Celsius), glucose level (mg/dL), blood urea nitrogen (BUN) amount (mg/dL), creatinine amount (mg/dL), white blood cell (WBC) count (K/uL). We randomly sample the training data with a size $798 \times 50\% = 399$ and the remaining sample is used for testing. We compare the performance of MR and DR methods. We first construct estimators using the training dataset. For the MR method, $\varphi(x)$, $K_a(x)$, and $p_a(x)$ are estimated using GAMs, and $\mu_a(x)$ is estimated using RF. For the DR method, propensity score and conditional outcome are estimated the same as in the MR method. Two survivor-optimal linear policies are obtained by maximizing the MR and DR estimators within the policy class $\Pi_\beta = \{\pi(x;\beta) = \mathbb{I}(\beta^T\tilde{x} > 0) : \beta \in \mathbb{R}^8, \|\beta\|_2 = 1\}$, respectively. We denote the estimated β's as $\widehat{\beta}^{\mathrm{MR}}$ and $\widehat{\beta}^{\mathrm{DR}}$.

We use the MR estimator over the testing sample as the testing value $V^{\text{test}}(\beta)$. We obtain the optimal survivor value function $V^{\text{test}}(\beta^*)$ by maximizing the testing value over β. We compute the difference between $V^{\text{test}}(\beta^*)$ and $V^{\text{test}}(\widehat{\beta}^{\mathrm{MR}})$, and the difference between $V^{\text{test}}(\beta^*)$ and $V^{\text{test}}(\widehat{\beta}^{\mathrm{DR}})$. The training-testing procedure is repeated 50 times. We report the results of $V^{\text{test}}(\widehat{\beta}^{\mathrm{MR}}) - V^{\text{test}}(\beta^*)$ and $V^{\text{test}}(\widehat{\beta}^{\mathrm{DR}}) - V^{\text{test}}(\beta^*)$ in Figure 2. We can see that the average and variability of $V^{\text{test}}(\widehat{\beta}^{\mathrm{MR}}) - V^{\text{test}}(\beta^*)$ is smaller than those of $V^{\text{test}}(\widehat{\beta}^{\mathrm{DR}}) - V^{\text{test}}(\beta^*)$, which implies that OPL based on the MR method has better performance than that based on the DR method.

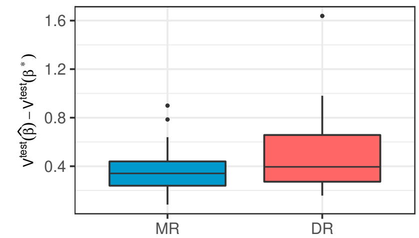

Figure 2. The boxplots of difference between testing values under true and estimated survivor-optimal linear policies by MR and DR methods.

7. Conclusion and Future Direction

To the best of our knowledge, this is the first paper formalizing OPE and OPL under the truncation by death setting. We established identification and the semiparametric efficiency bound for OPE and proposed OPE and OPL methods for this situation. In particular, our proposed estimators have multiple robustness property and achieve the semiparametric efficiency bound under mild rate conditions of nuisance functions approximation. The experiments showed that our proposed OPE and OPL methods outperform the existing methods.

Our identification strategy relies on some crucial assumptions such as principal ignorability and monotonicity, which may be violated in some real-world applications. In such cases, we may only have partial identification (Balke & Pearl, 1997; Imai, 2008; Swanson et al., 2018), instead of point identification of the survivor value function. The direction for OPE may change to construct bounds for the survivor value function and corresponding OPL methods need to be developed. The sensitivity analysis technique we provide opens a door to this interesting future direction.

References

Athey, S. and Wager, S. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021.

Balke, A. and Pearl, J. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997.

Bickel, P. J., Klaassen, C., Ritov, Y., and Wellner, J. *Efficient and Adaptive Inference in Semiparametric Models*. Johns Hopkins University Press, Baltimore, 1993.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21:C1–C68, 2018.

Chu, J., Lu, W., and Yang, S. Targeted optimal treatment regime learning using summary statistics. *Biometrika*, 2023. Available at: https://doi.org/10.1093/biomet/asad020.

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

Dinglas, V. D., Faraone, L. N., and Needham, D. M. Understanding patient-important outcomes after critical illness: a synthesis of recent qualitative, empirical, and consensus-related studies. *Current Opinion in Critical Care*, 24(5): 401, 2018.

Dudík, M., Erhan, D., Langford, J., and Li, L. Doubly robust policy evaluation and optimization. *Statistical Science*, 29(4):485–511, 2014.

Fang, E. X., Wang, Z., and Wang, L. Fairness-oriented learning for optimal individualized treatment rules. *Journal of the American Statistical Association*, pp. 1–14, 2022.

Farrell, M. H., Liang, T., and Misra, S. Deep neural networks for estimation and inference. *Econometrica*, 89(1): 181–213, 2021.

Follmann, D. Augmented designs to assess immune response in vaccine trials. *Biometrics*, 62:1161–1169, 2006.

Follmann, D. A. On the effect of treatment among would-be treatment compliers: An analysis of the multiple risk factor intervention trial. *Journal of the American Statistical Association*, 95:1101–1109, 2000.

Frangakis, C. E. and Rubin, D. B. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86:365–379, 1999.

Frangakis, C. E. and Rubin, D. B. Principal stratification in causal inference. *Biometrics*, 58:21–29, 2002.

Fried, T. R., Bradley, E. H., Towle, V. R., and Allore, H. Understanding the treatment preferences of seriously ill patients. *New England Journal of Medicine*, 346(14): 1061–1066, 2002.

Fujimoto, S., Meger, D., and Precup, D. Off-policy deep reinforcement learning without exploration. In *International Conference on Machine Learning*, pp. 2052–2062. PMLR, 2019.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23): e215–e220, 2000.

Guérin, C., Reignier, J., Richard, J.-C., Beuret, P., Gacouin, A., Boulain, T., Mercier, E., Badet, M., Mercat, A., Baudin, O., et al. Prone positioning in severe acute respiratory distress syndrome. *New England Journal of Medicine*, 368(23):2159–2168, 2013.

Horvitz, D. G. and Thompson, D. J. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952.

Imai, K. Sharp bounds on the causal effects in randomized experiments with "truncation-by-death". *Statistics & Probability Letters*, 78(2):144–149, 2008.

Iwashyna, T. J., Ely, E. W., Smith, D. M., and Langa, K. M. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *JAMA*, 304(16):1787–1794, 2010.

Jiang, N. and Li, L. Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*, pp. 652–661. PMLR, 2016.

Jiang, R., Lu, W., Song, R., and Davidian, M. 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, 2017.

Jiang, Z., Yang, S., and Ding, P. Multiply robust estimation of causal effects under principal ignorability. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1423–1445, 2022.

Johnson, A., Pollard, T., and Mark, R. Mimic-iii clinical database demo (version 1.4). *PhysioNet*, 2019. Available at: https://doi.org/10.13026/C2HM2Q.

Johnson, A. E., Pollard, T. J., Shen, L., Li-Wei, H. L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., and Mark, R. G. Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3(1):1–9, 2016.

Kallus, N. and Uehara, M. Double reinforcement learning for efficient off-policy evaluation in markov decision processes. *Journal of Machine Learning Research*, 21(167), 2020.

Kennedy, E. H. Semiparametric theory and empirical processes in causal inference. In *Statistical causal inferences and their applications in public health research*, pp. 141–167. Springer, 2016.

Kitagawa, T. and Tetenov, A. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

Kosorok, M. R. *Introduction to empirical processes and semiparametric inference*. Springer, 2008.

Lipkovich, I., Dmitrienko, A., and B D'Agostino Sr, R. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Statistics in Medicine*, 36 (1):136–196, 2017.

Lipkovich, I., Ratitch, B., Qu, Y., Zhang, X., Shan, M., and Mallinckrodt, C. Using principal stratification in analysis of clinical trials. *Statistics in Medicine*, 41(19): 3837–3877, 2022.

Luckett, D. J., Laber, E. B., Kahkoska, A. R., Maahs, D. M., Mayer-Davis, E., and Kosorok, M. R. Estimating dynamic treatment regimes in mobile health using v-learning. *Journal of the American Statistical Association*, 2019.

Luedtke, A. R. and Van Der Laan, M. J. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *The Annals of Statistics*, 44 (2):713–742, 2016.

Mandel, T., Liu, Y.-E., Levine, S., Brunskill, E., and Popovic, Z. Offline policy evaluation across representations with applications to educational games. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pp. 1077–1084, 2014.

Mo, W., Qi, Z., and Liu, Y. Learning optimal distributionally robust individualized treatment rules. *Journal of the American Statistical Association*, 116(534):659–674, 2021.

Munos, R., Stepleton, T., Harutyunyan, A., and Bellemare, M. Safe and efficient off-policy reinforcement learning. *Advances in Neural Information Processing Systems*, 29, 2016.

Paltiel, A. D., Schwartz, J. L., Zheng, A., and Walensky, R. P. Clinical outcomes of a covid-19 vaccine: Implementation over efficacy: Study examines how definitions and thresholds of vaccine efficacy, coupled with different levels of implementation effectiveness and background epidemic severity, translate into outcomes. *Health Affairs*, 40(1):42–52, 2021.

Qian, M. and Murphy, S. A. Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180–1210, 2011.

Qiu, H., Carone, M., Sadikova, E., Petukhova, M., Kessler, R. C., and Luedtke, A. Optimal individualized decision rules using instrumental variable methods. *Journal of the American Statistical Association*, 116(533):174–191, 2021.

Rhee, C., Dantes, R., Epstein, L., Murphy, D. J., Seymour, C. W., Iwashyna, T. J., Kadri, S. S., Angus, D. C., Danner, R. L., Fiore, A. E., et al. Incidence and trends of sepsis in us hospitals using clinical vs claims data, 2009-2014. *JAMA*, 318(13):1241–1249, 2017.

Rosenbaum, P. R. and Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Rubin, D. B. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J.-D., Coopersmith, C. M., et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*, 315(8):801–810, 2016.

Sommer, A. and Zeger, S. L. On estimating efficacy from clinical trials. *Statistics in Medicine*, 10:45–52, 1991.

Stuart, E. A. and Jo, B. Assessing the sensitivity of methods for estimating principal causal effects. *Statistical Methods in Medical Research*, 24:657–674, 2015.

Swaminathan, A., Krishnamurthy, A., Agarwal, A., Dudik, M., Langford, J., Jose, D., and Zitouni, I. Off-policy evaluation for slate recommendation. *Advances in Neural Information Processing Systems*, 30, 2017.

Swanson, S. A., Hernán, M. A., Miller, M., Robins, J. M., and Richardson, T. S. Partial identification of the average treatment effect using instrumental variables: review of methods for binary instruments, treatments, and outcomes. *Journal of the American Statistical Association*, 113(522):933–947, 2018.

Uehara, M., Kato, M., and Yasui, S. Off-policy evaluation and learning for external validity under a covariate shift. *Advances in Neural Information Processing Systems*, 33: 49–61, 2020.

VanderWeele, T. J., Luedtke, A. R., van der Laan, M. J., and Kessler, R. C. Selecting optimal subgroups for treatment using many covariates. *Epidemiology (Cambridge, Mass.)*, 30(3):334, 2019.

Wang, Y.-X., Agarwal, A., and Dudık, M. Optimal and adaptive off-policy evaluation in contextual bandits. In *International Conference on Machine Learning*, pp. 3589–3597. PMLR, 2017.

Wunsch, H., Linde-Zwirble, W. T., Angus, D. C., Hartman, M. E., Milbrandt, E. B., and Kahn, J. M. The epidemiology of mechanical ventilation use in the united states. *Critical Care Medicine*, 38(10):1947–1953, 2010.

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

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, 100(3): 681–694, 2013.

Zheng, W. and Laan, M. J. Cross-validated targeted minimum-loss-based estimation. In *Targeted Learning*, pp. 459–474. Springer, 2011.

Zhou, J., Zhang, J., Lu, W., and Li, X. On restricted optimal treatment regime estimation for competing risks data. *Biostatistics*, 22(2):217–232, 2021.

A. Technical Proofs

Throughout the proofs, we will use $f(\cdot)$ to denote the probability density functions for continuous random variables and the probability mass functions for discrete random variables.

A.1. Proof of Theorem 3.6

Proof. To prove Theorem 3.6, we first state a lemma.

Lemma A.1. *For any function $h(X)$ that has a finite moment $\mathbb{E}[h(X) \mid U = 11] < \infty$, the following balancing condition*

$$\mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{A}{\varphi_1(X)}\frac{1-C}{K_1(X)}\frac{S}{p_1(X)}h(X)\right] = \mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{1-A}{\varphi_0(X)}\frac{1-C}{K_0(X)}\frac{S}{p_0(X)}h(X)\right] = \mathbb{E}\left[\frac{p_0(X)}{p_0}h(X)\right]. \quad (6)$$

holds. Under Assumptions 3.1–3.3, the three expectations in (6) equal $\mathbb{E}[h(X) \mid U = 11]$.

Proof.

$$\begin{aligned}
\mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{A}{\varphi_1(X)}\frac{1-C}{K_1(X)}\frac{S}{p_1(X)}h(X)\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{A}{\varphi_1(X)}\frac{1-C}{K_1(X)}\frac{S}{p_1(X)}h(X) \mid X\right]\right] \\
&= \mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{P(A=1, C=0, S=1 \mid X)}{\varphi_1(X)K_1(X)p_1(X)}h(X)\right] \\
&= \mathbb{E}\left[\frac{p_0(X)}{p_0}h(X)\right]. \\
\mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{1-A}{\varphi_0(X)}\frac{1-C}{K_0(X)}\frac{S}{p_0(X)}h(X)\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{1-A}{\varphi_0(X)}\frac{1-C}{K_0(X)}\frac{S}{p_0(X)}h(X) \mid X\right]\right] \\
&= \mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{P(A=0, C=0, S=1 \mid X)}{\varphi_0(X)K_0(X)p_0(X)}h(X)\right] \\
&= \mathbb{E}\left[\frac{p_0(X)}{p_0}h(X)\right].
\end{aligned}$$

Therefore, (6) holds. Under Assumptions 3.1–3.3, we have

$$\begin{aligned}
&P(U=11 \mid X) = P\left[S(0)=1, S(1)=1 \mid X\right] \\
=&P\left[S(0)=1 \mid X\right] \quad \text{(Assumption 3.3)} \\
=&P\left[S(0)=1 \mid C(0)=0, X\right] \quad \text{(Assumption 3.2)} \\
=&P\left[S(0)=1 \mid A=0, C(0)=0, X\right] \quad \text{(Assumption 3.1)} \\
=&P\left(S=1 \mid A=0, C=0, X\right) \quad \text{(SUTVA)} \\
=&p_0(X),
\end{aligned}$$

and thus $P(U = 11) = p_0$. Then

$$\begin{aligned}
\mathbb{E}[h(X) \mid U=11] &= \mathbb{E}\left[\frac{f(X \mid U=11)}{f(X)}h(X)\right] \\
&= \mathbb{E}\left[\frac{P(U=11 \mid X)}{P(U=11)}h(X)\right] \\
&= \mathbb{E}\left[\frac{p_0(X)}{p_0}h(X)\right]. \quad (7)
\end{aligned}$$

□

For a given policy $\pi \in \Pi$, we have

$$\begin{aligned}
V(\pi) &= \mathbb{E}\left[Y(1)\pi(X) + Y(0)\{1-\pi(X)\} \mid U=11\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y(1)\pi(X) + Y(0)\{1-\pi(X)\} \mid U=11, X] \mid U=11\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathbb{E}[Y(1) \mid U=11, X]\pi(X) \mid U=11\right] \\
&\quad +\mathbb{E}\left[\mathbb{E}[Y(0) \mid U=11, X]\{1-\pi(X)\} \mid U=11\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y(1) \mid U=11 \text{ or } 10, X]\pi(X) \mid U=11\right] \quad \text{(Assumption 3.4)} \\
&\quad +\mathbb{E}\left[\mathbb{E}[Y(0) \mid U=11, X]\{1-\pi(X)\} \mid U=11\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y(1) \mid S(1)=1, X]\pi(X) \mid U=11\right] \\
&\quad +\mathbb{E}\left[\mathbb{E}[Y(0) \mid S(0)=1, X]\{1-\pi(X)\} \mid U=11\right] \quad \text{(Assumption 3.3)} \\
&= \mathbb{E}\left[\mathbb{E}[Y(1) \mid C(1)=0, S(1)=1, X]\pi(X) \mid U=11\right] \\
&\quad +\mathbb{E}\left[\mathbb{E}[Y(0) \mid C(0)=0, S(0)=1, X]\{1-\pi(X)\} \mid U=11\right] \quad \text{(Assumption 3.2)} \\
&= \mathbb{E}\left[\mathbb{E}[Y \mid A=1, C(1)=0, S(1)=1, X]\pi(X) \mid U=11\right] \\
&\quad +\mathbb{E}\left[\mathbb{E}[Y \mid A=0, C(0)=0, S(0)=1, X]\{1-\pi(X)\} \mid U=11\right] \quad \text{(Assumption 3.1)} \\
&= \mathbb{E}\left[\mathbb{E}[Y \mid A=1, C=0, S=1, X]\pi(X) \mid U=11\right] \\
&\quad +\mathbb{E}\left[\mathbb{E}[Y \mid A=0, C=0, S=1, X]\{1-\pi(X)\} \mid U=11\right] \quad \text{(SUTVA)} \\
&= \mathbb{E}\left[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\} \mid U=11\right]. \qquad (8)
\end{aligned}
$$

By Lemma A.1, for any $h(X)$,

$$
\begin{aligned}
\mathbb{E}\{h(X) \mid U=11\} &= \mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{A}{\varphi_1(X)}\frac{1-C}{K_1(X)}\frac{S}{p_1(X)}h(X)\right] \\
&= \mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{1-A}{\varphi_0(X)}\frac{1-C}{K_0(X)}\frac{S}{p_0(X)}h(X)\right] = \mathbb{E}\left[\frac{p_0(X)}{p_0}h(X)\right].
\end{aligned}
$$

Continuing from (8),

$$
V(\pi) = \mathbb{E}\left[\tfrac{p_0(X)}{p_0}\left[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\right]\right], \qquad (9)
$$

and

$$
\begin{aligned}
V(\pi) &= \mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{A}{\varphi_1(X)}\frac{1-C}{K_1(X)}\frac{S}{p_1(X)}\mu_1(X)\pi(X)\right. \\
&\quad \left.+\frac{p_0(X)}{p_0}\frac{1-A}{\varphi_0(X)}\frac{1-C}{K_0(X)}\frac{S}{p_0(X)}\mu_0(X)\{1-\pi(X)\}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{A}{\varphi_1(X)}\frac{1-C}{K_1(X)}\frac{S}{p_1(X)}\mu_1(X)\pi(X)\right.\right. \\
&\quad \left.\left.+\frac{p_0(X)}{p_0}\frac{1-A}{\varphi_0(X)}\frac{1-C}{K_0(X)}\frac{S}{p_0(X)}\mu_0(X)\{1-\pi(X)\} \mid X\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{p_0(X)}{p_0}\frac{A}{\varphi_1(X)}\frac{1-C}{K_1(X)}\frac{S}{p_1(X)}Y\pi(X)\right.\right. \\
&\quad \left.\left.+\frac{p_0(X)}{p_0}\frac{1-A}{\varphi_0(X)}\frac{1-C}{K_0(X)}\frac{S}{p_0(X)}Y\{1-\pi(X)\} \mid X\right]\right] \\
&= \mathbb{E}\left[\frac{p_0(X)}{p_0}\left\{\frac{A}{\varphi_1(X)}\frac{1-C}{K_1(X)}\frac{S}{p_1(X)}Y\pi(X)\right.\right. \\
&\quad \left.\left.+\frac{1-A}{\varphi_0(X)}\frac{1-C}{K_0(X)}\frac{S}{p_0(X)}Y\{1-\pi(X)\}\right\}\right].
\end{aligned}
$$

□

A.2. Proof of Theorem 3.7

Proof. We will use the semiparametric theory in Bickel et al. (1993) to derive the EIF. Let $O=\{X, A, C, (1-C)S, (1-C)SY\}$ summarize the vector of observed variables with the likelihood factorized as

$$
f(O) = f(X)f(A \mid X)f(C \mid A, X)\{f(S \mid A, C=0, X)\}^{\mathbb{I}(C=0)}\{f(Y \mid A, C=0, S=1, X)\}^{\mathbb{I}(C=0,S=1)}. \qquad (10)
$$

To derive the EIF for $V(\pi)$, we consider a one-dimensional parametric submodel, $f_\theta(O)$, which contains the true model $f(O)$ at $\theta = 0$, i.e., $f_\theta(O)|_{\theta=0} = f(O)$. We use θ in the subscript to denote the quantity with respect to the submodel, e.g., $V_\theta(\pi)$ is the value of $V(\pi)$ in the submodel. We use dot to denote the partial derivative with respect to θ, e.g., $\dot{V}_\theta(\pi) = \partial V_\theta(\pi)/\partial\theta$, and use $s_{\theta(\cdot)}$ to denote the score function of the submodel. From (10), the score function under the submodel can be decomposed as

$$s_\theta(O) = s_\theta(X) + s_\theta(A \mid X) + s_\theta(C \mid A, X) + s_\theta(S \mid A, C = 0, X) + s_\theta(Y \mid A, C = 0, S = 1, X),$$

where $s_\theta(X) = \partial \log f_\theta(X)/\partial\theta$, $s_\theta(A \mid X) = \partial \log f_\theta(A \mid X)/\partial\theta$, $s_\theta(C \mid A, X) = \partial \log f_\theta(C \mid A, X)/\partial\theta$, $s_\theta(S \mid A, C = 0, X) = \partial \log f_\theta(S \mid A, C = 0, X)/\partial\theta$, $s_\theta(Y \mid A, C = 0, S = 1, X) = \partial \log f_\theta(Y \mid A, C = 0, S = 1, X)/\partial\theta$ are the score functions corresponding to the five components of the likelihood. Analogous to $f_\theta(O)|_{\theta=0} = f(O)$, we write $s_\theta(\cdot)|_{\theta=0}$ as $s(\cdot)$, which is the score function evaluated at the true parameter under the one-dimensional submodel.

From the semiparametric theory, the tangent space,

$$\Lambda = H_1 \oplus H_2 \oplus H_3 \oplus H_4 \oplus H_5$$

is the direct sum of

$$\begin{aligned}
H_1 &= \{h(X) : \mathbb{E}[h(X)] = 0\}, \\
H_2 &= \{h(A, X) : \mathbb{E}[h(A, X) \mid X] = 0\}, \\
H_3 &= \{h(C, A, X) : \mathbb{E}[h(C, A, X) \mid A, X] = 0\}, \\
H_4 &= \{h(S, A, C = 0, X) : \mathbb{E}[h(S, A, C = 0, X) \mid A, C = 0, X] = 0\}, \\
H_5 &= \{h(Y, A, C = 0, S = 1, X) : \mathbb{E}[h(Y, A, C = 0, S = 1, X) \mid A, C = 0, S = 1, X] = 0\},
\end{aligned}$$

where H_1, H_2, H_3, H_4, H_5 are orthogonal to each other. The EIF for $V(\pi)$ denoted by $\nu_\pi(O) \in \Lambda$, must satisfy

$$\dot{V}_\theta(\pi)|_{\theta=0} = \mathbb{E}[\nu_\pi(O)s(O)].$$

We will derive the EIF by calculating $\dot{V}_\theta(\pi)|_{\theta=0}$. To simplify the proof, we introduce some lemmas.

Lemma A.2. *Consider a ratio-type parameter $R = N/D$. If $\dot{N}_\theta|_{\theta=0} = \mathbb{E}[\nu_N(O)s(O)]$ and $\dot{D}_\theta|_{\theta=0} = \mathbb{E}[\nu_D(O)s(O)]$, then $\dot{R}_\theta|_{\theta=0} = \mathbb{E}[\nu_R(O)s(O)]$ where*

$$\nu_R(O) = \frac{1}{D}\nu_N(O) - \frac{R}{D}\nu_D(O). \tag{11}$$

In particular, if $\nu_N(O)$ and $\nu_D(O)$ are the EIFs for N and D, then $\nu_R(O)$ is the EIF for R.

Proof. Let R_θ, N_θ, and D_θ denote the quantities R, N, and D evaluated with respect to the parametric submodel $f_\theta(O)$. By the chain rule, we have

$$\begin{aligned}
\dot{R}_\theta|_{\theta=0} &= \left.\frac{\dot{N}_\theta}{D}\right|_{\theta=0} - R_\theta \left.\frac{\dot{D}_\theta}{D}\right|_{\theta=0} \\
&= \frac{1}{D}\mathbb{E}[\nu_N(O)s(O)] - \frac{R}{D}\mathbb{E}[\nu_D(O)s(O)] \\
&= \mathbb{E}\left[\left\{\frac{1}{D}\nu_N(O) - \frac{R}{D}\nu_D(O)\right\} s(O)\right],
\end{aligned}$$

which yields (11). □

Lemma A.3. *For any $h(O)$ that does not depend on θ, $\partial\mathbb{E}_\theta[h(O)]/\partial\theta|_{\theta=0} = \mathbb{E}[h(O)s(O)]$.*

The proof is straightforward and thus omitted.

Lemma A.4. *Define $\mu_{af}(X) = \mathbb{E}[f(Y, S, X) \mid A = a, C = 0, X]$ for any $f(Y, S, X)$, where $a = 0, 1$. We have*

$$\dot{\mu}_{af,\theta}(X)|_{\theta=0} = \mathbb{E}\left[\{\psi_{f(Y(a),S(a),X)} - \mu_{af}(X)\}s(Y, S \mid A, C = 0, X) \mid X\right].$$

As a special case,

$$\dot{p}_{a,\theta}(X)|_{\theta=0} = \mathbb{E}\left[\{\psi_{S(a)} - p_a(X)\}s(S \mid A, C = 0, X) \mid X\right].$$

Proof. We first prove the general result:

$$
\begin{aligned}
\dot{\mu}_{af,\theta}(X)|_{\theta=0} &= \frac{\partial}{\partial\theta}\mathbb{E}_\theta[f(Y,S,X) \mid A=a, C=0, X]|_{\theta=0} \\
&= \mathbb{E}[f(Y,S,X) \times s(Y,S \mid A=a, C=0, X) \mid A=a, C=0, X] \quad \text{(Lemma A.3)} \\
&= \mathbb{E}\left[\{f(Y,S,X)-\mu_{af}(X)\}s(Y,S \mid A=a, C=0, X) \mid A=a, C=0, X\right] \\
&= \mathbb{E}\left[\frac{\mathbb{I}\{A=a\}\mathbb{I}(C=0)\{f(Y,S,X)-\mu_{af}(X)\}}{\varphi_a(X)K_a(X)} s(Y,S \mid A, C=0, X) \mid X\right] \\
&= \mathbb{E}\left[\{\psi_{f(Y(a),S(a),X)}-\mu_{af}(X)\}s(Y,S \mid A, C=0, X) \mid X\right], \qquad (12)
\end{aligned}
$$

where the third equality follows from $\mu_{af}(X)\mathbb{E}[s(Y,S \mid A=a, C=0, X) \mid A=a, C=0, X]=0$.

Choosing $f(Y,S,X)=S$, the result for $p_a(X)$ follows because

$$
\mathbb{E}\left[\{\psi_{S(a)}-p_a(X)\}s(Y \mid A, C=0, S=1, X) \mid X\right]=0.
$$

□

Lemma A.5. *Define* $\mu_{af} = \mathbb{E}[\mu_{af}(X)]$ *for any* $f(Y,S,X)$*, where* $a=0,1$*. We have*

$$
\dot{\mu}_{af,\theta}|_{\theta=0} = \mathbb{E}\left[\{\psi_{f(Y(a),S(a),X)}-\mu_{af}\}s(O)\right].
$$

Moreover, $\psi_{f(Y(a),S(a),X)}-\mu_{af}$ *is the EIF for* μ_{af}*,* $a=0,1$*. As a special case, for* $p_a=\mathbb{E}[p_a(X)]$*, we have*

$$
\dot{p}_{a,\theta}|_{\theta=0} = \mathbb{E}\left[\{\psi_{S(a)}-p_a\}s(O)\right]
$$

and $\psi_{S(a)}-p_a$ *is the EIF for* p_a*,* $a=0,1$*.*

Proof. We prove the general result:

$$
\begin{aligned}
\dot{\mu}_{af,\theta}|_{\theta=0} &= \mathbb{E}[\mu_{af}(X)s(O)] + \mathbb{E}[\dot{\mu}_{af,\theta}(X)|_{\theta=0}] \quad \text{(Lemma A.3)} \\
&= \mathbb{E}[\mu_{af}(X)s(O)] + \mathbb{E}[\{\psi_{f(Y(a),S(a),X)}-\mu_{af}(X)\}s(Y,S \mid A, C=0, X)] \quad \text{(Lemma A.4)} \\
&= \mathbb{E}[\{\psi_{f(Y(a),S(a),X)}-\mu_{af}\}s(Y,S \mid A, C=0, X)] \\
&= \mathbb{E}[\{\psi_{f(Y(a),S(a),X)}-\mu_{af}\}s(O)],
\end{aligned}
$$

where the last equality follows from $\mathbb{E}[\{\psi_{f(Y(a),S(a),X)}-\mu_{af}\}s(A \mid X)] = \mathbb{E}[\{\psi_{f(Y(a),S(a),X)}-\mu_{af}\}s(X)] = 0$. Because $\psi_{f(Y(a),S(a),X)}-\mu_{af}$ lies in the tangent space, it is the EIF for μ_{af}. The result for p_a follows by taking $f(Y,S,X)=S$. □

Lemma A.6. *For* $\mu_a(X)$*, we have*

$$
\dot{\mu}_{a,\theta}(X)|_{\theta=0} = \mathbb{E}\left[\frac{\psi_{Y(a)S(a)}-\mu_a(X)\psi_{S(a)}}{p_a(X)} s(Y \mid A, C=0, S=1, X) \mid X\right].
$$

Proof. A key observation is the ratio representation:

$$
\mu_a(X) = \mathbb{E}[Y \mid A=a, C=0, S=1, X] = \frac{\mathbb{E}[YS \mid A=a, C=0, X]}{\mathbb{E}[S \mid A=a, C=0, X]} = \frac{\mathbb{E}[YS \mid A=a, C=0, X]}{p_a(X)}
$$

From Lemma A.4, the numerator satisfies

$$
\frac{\partial}{\partial\theta}\mathbb{E}_\theta\{YS \mid A=a, C=0, X\}|_{\theta=0} = \mathbb{E}\left[\{\psi_{Y(a)S(a)}-p_a(X)\mu_a(X)\}s(Y,S \mid A, C=0, X) \mid X\right].
$$

and denominator satisfies

$$
\dot{p}_a(X)|_{\theta=0} = \mathbb{E}\left[\{\psi_{S(a)}-p_a(X)\}s(Y,S \mid A, C=0, X) \mid X\right]
$$

We then use Lemma A.2 to calculate the path derivative of $\mu_{a,\theta}(X)$ with all distributions conditional on X, yielding

$$
\dot{\mu}_{a,\theta}(X)|_{\theta=0} = \mathbb{E}\left[\frac{\psi_{Y(a)S(a)}-\mu_a(X)\psi_{S(a)}}{p_a(X)} s(Y,S \mid A, C=0, X) \mid X\right].
$$

The conclusion follows by using $\mathbb{E}\left[\{\psi_{Y(a)S(a)} - \mu_a(X)\psi_{S(a)}\}s(S \mid A, C=0, X) \mid X\right] = 0.$ □

Below, we derive the EIF for $V(\pi)$.

First, from (9), we can write $V(\pi) = N/D$, where

$$N = \mathbb{E}\left[p_0(X)\left[\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right]\right], \qquad D = p_0.$$

From Lemma A.5,

$$\nu_D(O) = \psi_{S_0} - p_0 = \psi_{S_0} - D.$$

From the chain rule, we have

$$\begin{aligned} \dot{N}_\theta \mid_{\theta=0} &= \frac{\partial}{\partial\theta}\mathbb{E}_\theta\left[p_{0,\theta}(X)\left[\mu_{1,\theta}(X)\pi(X) + \mu_{0,\theta}(X)\{1-\pi(X)\}\right]\right]\Big|_{\theta=0} \\ &= \mathbb{E}\left[p_0(X)\left[\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right]s(X)\right] \quad \text{(Lemma A.3)} \\ &\quad +\mathbb{E}_\theta\left[\dot{p}_{0,\theta}(X)\left[\mu_{1,\theta}(X)\pi(X) + \mu_{0,\theta}(X)\{1-\pi(X)\}\right]\right]\mid_{\theta=0} \\ &\quad +\mathbb{E}_\theta\left[p_{0,\theta}(X)\left[\dot{\mu}_{1,\theta}(X)\pi(X) + \dot{\mu}_{0,\theta}(X)\{1-\pi(X)\}\right]\right]\mid_{\theta=0}. \end{aligned} \tag{13}$$

Because $E\{Ns(X)\} = 0$, the first term in (13) equals

$$\begin{aligned} & \mathbb{E}\left[p_0(X)\left[\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right]s(X)\right] \\ = \ & \mathbb{E}\left[\{p_0(X)\left[\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right] - N\}s(X)\right]. \end{aligned}$$

From Lemma A.4, the second term in (13) reduces to

$$\begin{aligned} & \mathbb{E}_\theta\left[\dot{p}_{0,\theta}(X)\left[\mu_{1,\theta}(X)\pi(X) + \mu_{0,\theta}(X)\{1-\pi(X)\}\right]\right]\mid_{\theta=0} \\ = \ & \mathbb{E}\left[\mathbb{E}\left[\{\psi_{S(0)} - p_0(X)\}s(S \mid A, C=0, X)\{\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\} \mid X\right]\right] \\ = \ & \mathbb{E}\left[\{\psi_{S(0)} - p_0(X)\}\{\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\}s(S \mid A, C=0, X)\right]. \end{aligned}$$

From Lemma A.6, the third term in (13) reduces to

$$\begin{aligned} & \mathbb{E}_\theta\left[p_{0,\theta}(X)\left[\dot{\mu}_{1,\theta}(X)\pi(X) + \dot{\mu}_{0,\theta}(X)\{1-\pi(X)\}\right]\right]\mid_{\theta=0} \\ = \ & \mathbb{E}\Bigg[p_0(X)\left\{\frac{\psi_{Y(1)S(1)} - \mu_1(X)\psi_{S(1)}}{p_1(X)}\pi(X) + \frac{\psi_{Y(0)S(0)} - \mu_0(X)\psi_{S(0)}}{p_0(X)}\{1-\pi(X)\}\right\} \\ & \times s(Y \mid A, C=0, S=1, X)\Bigg]. \end{aligned}$$

Plugging the above three formulas into (13) gives

$$\begin{aligned} & \dot{N}_\theta \mid_{\theta=0} \\ = \ & \mathbb{E}\left[\{p_0(X)\left[\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right] - N\}s(X)\right] \\ & +\mathbb{E}\left[\{\psi_{S(0)} - p_0(X)\}\{\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\}s(S \mid A, C=0, X)\right] \\ & +\mathbb{E}\Bigg[p_0(X)\left\{\frac{\psi_{Y(1)S(1)} - \mu_1(X)\psi_{S(1)}}{p_1(X)}\pi(X) + \frac{\psi_{Y(0)S(0)} - \mu_0(X)\psi_{S(0)}}{p_0(X)}\{1-\pi(X)\}\right\} \qquad (14) \\ & \times s(Y \mid A, C=0, S=1, X)\Bigg]. \end{aligned} \tag{15}$$

We can verify that

$$\begin{aligned} p_0(X)\left[\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right] - N &\in H_1, \\ \{\psi_{S(0)} - p_0(X)\}\left[\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right] &\in H_4, \\ p_0(X)\left\{\frac{\psi_{Y(1)S(1)} - \mu_1(X)\psi_{S(1)}}{p_1(X)}\pi(X) + \frac{\psi_{Y(0)S(0)} - \mu_0(X)\psi_{S(0)}}{p_0(X)}\{1-\pi(X)\}\right\} &\in H_5. \end{aligned}$$

Because H_1, H_2, H_3, H_4, and H_5 are orthogonal to each other, we can write (15) as

$$\begin{aligned}\dot{N}_\theta \mid_{\theta=0} \;=\; & \mathbb{E}\left[\{p_0(X)\left[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\right]-N\}\, s(O)\right] \\ & +\mathbb{E}\left[\{\psi_{S(0)}-p_0(X)\}\left[\mu_{1,\theta}(X)\pi(X)+\mu_{0,\theta}(X)\{1-\pi(X)\}\right] s(O)\right] \\ & +\mathbb{E}\left[p_0(X)\left\{\frac{\psi_{Y(1)S(1)}-\mu_1(X)\psi_{S(1)}}{p_1(X)}\pi(X)+\frac{\psi_{Y(0)S(0)}-\mu_0(X)\psi_{S(0)}}{p(0)(X)}\{1-\pi(X)\}\right\} s(O)\right].\end{aligned}$$

As a result, we obtain the EIF for N:

$$\begin{aligned}\nu_N(O) \;=\; & p_0(X)\left[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\right]-N \\ & +\{\psi_{S(0)}-p_0(X)\}\left[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\right] \\ & +p_0(X)\left\{\frac{\psi_{Y(1)S(1)}-\mu_1(X)\psi_{S(1)}}{p_1(X)}\pi(X)+\frac{\psi_{Y(0)S(0)}-\mu_0(X)\psi_{S(0)}}{p_0(X)}\{1-\pi(X)\}\right\} \\ =\; & \left[\mu_1(X)\left\{\psi_{S(0)}-\frac{p_0(X)}{p_1(X)}\psi_{S(1)}\right\}+\frac{p_0(X)}{p_1(X)}\psi_{Y(1)S(1)}\right]\pi(X)+\psi_{Y(0)S(0)}\{1-\pi(X)\}-N\end{aligned}$$

From Lemma A.2, the EIF for $V(\pi)$ is

$$\nu_\pi=\frac{1}{p_0}\{\phi_\pi-V(\pi)\psi_{S(0)}\},$$

where $\phi_\pi=\left[\mu_1(X)\left\{\psi_{S(0)}-\frac{p_0(X)}{p_1(X)}\psi_{S(1)}\right\}+\frac{p_0(X)}{p_1(X)}\psi_{Y(1)S(1)}\right]\pi(X)+\psi_{Y(0)S(0)}\{1-\pi(X)\}$. □

A.3. Proof of Theorem 4.2

Proof. Recall the MR estimator

$$\widehat{V}^{\text{MR}}(\pi)=\frac{\mathbb{P}_n(\widehat{\phi}_\pi)}{\mathbb{P}_n\{\widehat{\psi}_{S(0)}\}}.$$

From (9), we can write $V(\pi)=N/D$, where

$$N=\mathbb{E}\left[p_0(X)\left[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\right]\right], \qquad D=p_0.$$

Therefore, we only need to prove the probability limit of $\mathbb{P}_n\{\widehat{\psi}_{S(0)}\}$ is equal to D and the probability limit of $\mathbb{P}_n(\widehat{\phi}_\pi)$ is equal to N under $\mathcal{M}_{\text{ps+cs+sp}}\cup\mathcal{M}_{\text{ps+cs+om}}\cup\mathcal{M}_{\text{sp+om}}$.

For the ease of exposition, we write $\varphi_a(X;\alpha^*)=\varphi_a^*(X)$, $K_a(X;\eta^*)=K_a^*(X)$, $p_a(X;\gamma^*)=p_a^*(X)$, $\mu_a(X;\zeta^*)=\mu_a^*(X)$ and let

$$\psi^*_{S(a)}=\frac{\mathbb{I}\{A=a\}\mathbb{I}(C=0)\{S-p_a^*(X)\}}{\varphi_a^*(X)K_a^*(X)}+p_a^*(X),$$

$$\psi^*_{Y(a)S(a)}=\frac{\mathbb{I}(A=a)\mathbb{I}(C=0)\{YS-\mu_a^*(X)p_a^*(X)\}}{\varphi_a^*(X)K_a^*(X)}+\mu_a^*(X)p_a^*(X),$$

$$\phi^*_\pi=\left[\mu_1^*(X)\left\{\psi^*_{S(0)}-\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{S(1)}\right\}+\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{Y(1)S(1)}\right]\pi(X)+\psi^*_{Y(0)S(0)}\{1-\pi(X)\}.$$

We first prove that under $\mathcal{M}_{\text{ps+cs}}\cup\mathcal{M}_{\text{sp}}\subset\mathcal{M}_{\text{ps+cs+sp}}\cup\mathcal{M}_{\text{ps+cs+om}}\cup\mathcal{M}_{\text{sp+om}}$, the probability limit of $\mathbb{P}_n\{\widehat{\psi}_{S(0)}\}$, $\mathbb{E}[\psi^*_{S(0)}]$, is equal to $D=p_0$.

$$\begin{aligned}\mathbb{E}[\psi^*_{S(0)}] =& \mathbb{E}\left[\frac{\mathbb{I}\{A=0\}\mathbb{I}(C=0)\{S-p_0^*(X)\}}{\varphi_a^*(X)K_0^*(X)}+p_0^*(X)\right] \\ =& \mathbb{E}\left\{\mathbb{E}\left[\frac{\mathbb{I}\{A=0\}\mathbb{I}(C=0)\{S-p_0^*(X)\}}{\varphi_a^*(X)K_0^*(X)}+p_0^*(X)\right]\mid A=0,C=0,X\right\} \\ =& \mathbb{E}\left[\frac{P(A=0,C=0\mid X)}{\varphi_0^*(X)K_0^*(X)}\{\mathbb{E}[S=1\mid A=0,C=0,X]-p_0^*(X)\}+p_0^*(X)\right]\end{aligned}$$

$$=\mathbb{E}\left[\frac{\varphi_0(X)K_0(X)}{\varphi_a^*(X)K_0^*(X)}\left\{p_0(X)-p_0^*(X)\right\}+p_0^*(X)\right].$$

Under $\mathcal{M}_{\mathrm{ps+cs}}$, $\varphi_0(X)=\varphi_0^*(X)$, $K_0(X)=K_0^*(X)$, and thus $\mathbb{E}[\psi_{S(0)}^*]=\mathbb{E}[p_0(X)]=p_0$. Under $\mathcal{M}_{\mathrm{sp}}$, $p_0(X)=p_0^*(X)$, and thus $\mathbb{E}[\psi_{S(0)}^*]=\mathbb{E}[p_0(X)]=p_0$.

Next, we prove the probability limit of $\mathbb{P}_n(\widehat{\phi}_\pi)$, $\mathbb{E}[\phi_\pi^*]$, is equal to N. We first decompose the probability limit as

$$\begin{aligned}\mathbb{E}[\phi_\pi^*]=&\mathbb{E}\left[\left[\mu_1^*(X)\left\{\psi_{S(0)}^*-\frac{p_0^*(X)}{p_1^*(X)}\psi_{S(1)}^*\right\}+\frac{p_0^*(X)}{p_1^*(X)}\psi_{Y(1)S(1)}^*\right]\pi(X)+\psi_{Y(0)S(0)}^*\{1-\pi(X)\}\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\psi_{S(0)}^*\pi(X)\right]-\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi_{S(1)}^*\pi(X)\right]+\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi_{Y(1)S(1)}^*\pi(X)\right]\\ &+\mathbb{E}\left[\psi_{Y(0)S(0)}^*\{1-\pi(X)\}\right].\end{aligned} \tag{16}$$

(i) Under $\mathcal{M}_{\mathrm{ps+cs+sp}}$, $\varphi_a(X)=\varphi_a^*(X)$, $K_a(X)=K_a^*(X)$, $p_a(X)=p_a^*(X)$. The first term in (16) equals

$$\begin{aligned}&\mathbb{E}\left[\mu_1^*(X)\psi_{S(0)}^*\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\mu_1^*(X)\psi_{S(0)}^*\pi(X)\mid A=0,C=0,X\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)\left[\frac{P(A=0,C=0\mid X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mathbb{E}\left[S=1\mid A=0,C=0,X\right]-p_0^*(X)\right\}+p_0^*(X)\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)\left[\frac{\varphi_0(X)K_0(X)}{\varphi_0(X)K_0(X)}\left\{p_0(X)-p_0(X)\right\}+p_0(X)\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)p_0(X)\right].\end{aligned} \tag{17}$$

The second term in (16) equals

$$\begin{aligned}&\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi_{S(1)}^*\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi_{S(1)}^*\pi(X)\mid A=1,C=0,X\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left[\frac{P(A=1,C=0\mid X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mathbb{E}\left[S=1\mid A=1,C=0,X\right]-p_1^*(X)\right\}+p_1^*(X)\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)\frac{p_0(X)}{p_1(X)}\left[\frac{\varphi_1(X)K_1(X)}{\varphi_1(X)K_1(X)}\left\{p_1(X)-p_1(X)\right\}+p_1(X)\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)p_0(X)\right].\end{aligned} \tag{18}$$

The third term in (16) equals

$$\begin{aligned}&\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi_{Y(1)S(1)}^*\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi_{Y(1)S(1)}^*\pi(X)\mid A=1,C=0,X\right]\right]\\ =&\mathbb{E}\left[\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left\{\frac{P(A=1,C=0\mid X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mathbb{E}[YS\mid A=1,C=0,X]-\mu_1^*(X)p_1^*(X)\right\}+\mu_1^*(X)p_1^*(X)\right\}\right]\\ =&\mathbb{E}\left[\pi(X)\frac{p_0(X)}{p_1(X)}\left\{\frac{\varphi_1(X)K_1(X)}{\varphi_1(X)K_1(X)}\left\{\mu_1(X)p_1(X)-\mu_1^*(X)p_1(X)\right\}+\mu_1^*(X)p_1(X)\right\}\right]\\ =&\mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right].\end{aligned} \tag{19}$$

The fourth term in (16) equals

$$\begin{aligned}&\mathbb{E}\left[\psi_{Y(0)S(0)}^*\{1-\pi(X)\}\right]=\mathbb{E}\left[\mathbb{E}\left[\psi_{Y(0)S(0)}^*\{1-\pi(X)\}\mid A=0,C=0,X\right]\right]\\ =&\mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{P(A=0,C=0\mid X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mathbb{E}[YS\mid A=0,C=0,X]-\mu_0^*(X)p_0^*(X)\right\}+\mu_0^*(X)p_0^*(X)\right\}\right]\\ =&\mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{\varphi_0(X)K_0(X)}{\varphi_0(X)K_0(X)}\left\{\mu_0(X)p_0(X)-\mu_0^*(X)p_0(X)\right\}+\mu_0^*(X)p_0(X)\right\}\right]\end{aligned}$$

$$
=\mathbb{E}\left[\mu_0(X)\{1-\pi(X)\}p_0(X)\right]. \tag{20}
$$

Combining (16)–(20), we have

$$
\begin{aligned}
&\mathbb{E}\left[\left[\mu_1^*(X)\left\{\psi^*_{S(0)}-\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{S(1)}\right\}+\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{Y(1)S(1)}\right]\pi(X)+\psi^*_{Y(0)S(0)}\{1-\pi(X)\}\right]\\
=&\mathbb{E}\left[\mu_1^*(X)\pi(X)p_0(X)\right]-\mathbb{E}\left[\mu_1^*(X)\pi(X)p_0(X)\right]+\mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right]\\
&+\mathbb{E}\left[\mu_0(X)\{1-\pi(X)\}p_0(X)\right]\\
=&\mathbb{E}\left[p_0(X)\left[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\right]\right]=N.
\end{aligned}
$$

Therefore, under $\mathcal{M}_{\text{ps+cs+sp}}$, $\widehat{V}^{\text{MR}}(\pi)$ is consistent for $V(\pi)$.

(ii) Under $\mathcal{M}_{\text{ps+cs+om}}$, $\varphi_a(X)=\varphi_a^*(X)$, $K_a(X)=K_a^*(X)$, $\mu_a(X)=\mu_a^*(X)$. The first term in (16) equals

$$
\begin{aligned}
&\mathbb{E}\left[\mu_1^*(X)\psi^*_{S(0)}\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\mu_1^*(X)\psi^*_{S(0)}\pi(X)\mid A=0,C=0,X\right]\right]\\
=&\mathbb{E}\left[\mu_1^*(X)\pi(X)\left[\frac{P(A=0,C=0\mid X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mathbb{E}\left[S=1\mid A=0,C=0,X\right]-p_0^*(X)\right\}+p_0^*(X)\right]\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)\left[\frac{\varphi_0(X)K_0(X)}{\varphi_0(X)K_0(X)}\left\{p_0(X)-p_0^*(X)\right\}+p_0^*(X)\right]\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right].
\end{aligned} \tag{21}
$$

The second term in (16) equals

$$
\begin{aligned}
&\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{S(1)}\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{S(1)}\pi(X)\mid A=1,C=0,X\right]\right]\\
=&\mathbb{E}\left[\mu_1^*(X)\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left[\frac{P(A=1,C=0\mid X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mathbb{E}\left[S=1\mid A=1,C=0,X\right]-p_1^*(X)\right\}+p_1^*(X)\right]\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left[\frac{\varphi_1(X)K_1(X)}{\varphi_1(X)K_1(X)}\left\{p_1(X)-p_1^*(X)\right\}+p_1^*(X)\right]\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)p_0^*(X)\frac{p_1(X)}{p_1^*(X)}\right].
\end{aligned} \tag{22}
$$

The third term in (16) equals

$$
\begin{aligned}
&\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{Y(1)S(1)}\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{Y(1)S(1)}\pi(X)\mid A=1,C=0,X\right]\right]\\
=&\mathbb{E}\left[\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left\{\frac{P(A=1,C=0\mid X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mathbb{E}[YS\mid A=1,C=0,X]-\mu_1^*(X)p_1^*(X)\right\}+\mu_1^*(X)p_1^*(X)\right\}\right]\\
=&\mathbb{E}\left[\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left\{\frac{\varphi_1(X)K_1(X)}{\varphi_1(X)K_1(X)}\left\{\mu_1(X)p_1(X)-\mu_1(X)p_1^*(X)\right\}+\mu_1(X)p_1^*(X)\right\}\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)p_0^*(X)\frac{p_1(X)}{p_1^*(X)}\right].
\end{aligned} \tag{23}
$$

The fourth term in (16) equals

$$
\begin{aligned}
&\mathbb{E}\left[\psi^*_{Y(0)S(0)}\{1-\pi(X)\}\right]=\mathbb{E}\left[\mathbb{E}\left[\psi^*_{Y(0)S(0)}\{1-\pi(X)\}\mid A=0,C=0,X\right]\right]\\
=&\mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{P(A=0,C=0\mid X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mathbb{E}[YS\mid A=0,C=0,X]-\mu_0^*(X)p_0^*(X)\right\}+\mu_0^*(X)p_0^*(X)\right\}\right]\\
=&\mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{\varphi_0(X)K_0(X)}{\varphi_0(X)K_0(X)}\left\{\mu_0(X)p_0(X)-\mu_0(X)p_0^*(X)\right\}+\mu_0(X)p_0^*(X)\right\}\right]
\end{aligned}
$$

$$=\mathbb{E}\left[\mu_0(X)\{1-\pi(X)\}p_0(X)\right]. \tag{24}$$

Combining (16) and (21)–(24) , we have

$$\begin{aligned}
&\mathbb{E}\left[\left[\mu_1^*(X)\left\{\psi_{S(0)}^*-\frac{p_0^*(X)}{p_1^*(X)}\psi_{S(1)}^*\right\}+\frac{p_0^*(X)}{p_1^*(X)}\psi_{Y(1)S(1)}^*\right]\pi(X)+\psi_{Y(0)S(0)}^*\{1-\pi(X)\}\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right]-\mathbb{E}\left[\mu_1(X)\pi(X)p_0^*(X)\frac{p_1(X)}{p_1^*(X)}\right]+\mathbb{E}\left[\mu_1(X)\pi(X)p_0^*(X)\frac{p_1(X)}{p_1^*(X)}\right]\\
&+\mathbb{E}\left[\mu_0(X)\{1-\pi(X)\}p_0(X)\right]\\
=&\mathbb{E}\left[p_0(X)\left[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\right]\right]=N.
\end{aligned}$$

Therefore, under $\mathcal{M}_{\text{ps+cs+om}}$, $\widehat{V}^{\text{MR}}(\pi)$ is consistent for $V(\pi)$.

(iii) Under $\mathcal{M}_{\text{sp+om}}$, $p_a(X)=p_a^*(X)$, $\mu_a(X)=\mu_a^*(X)$. The first term in (16) equals

$$\begin{aligned}
&\mathbb{E}\left[\mu_1^*(X)\psi_{S(0)}^*\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\mu_1^*(X)\psi_{S(0)}^*\pi(X)\mid A=0,C=0,X\right]\right]\\
=&\mathbb{E}\left[\mu_1^*(X)\pi(X)\left[\frac{P(A=0,C=0\mid X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mathbb{E}\left[S=1\mid A=0,C=0,X\right]-p_0^*(X)\right\}+p_0^*(X)\right]\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)\left[\frac{\varphi_0(X)K_0(X)}{\varphi_0^*(X)K_0^*(X)}\left\{p_0(X)-p_0(X)\right\}+p_0(X)\right]\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right].
\end{aligned} \tag{25}$$

The second term in (16) equals

$$\begin{aligned}
&\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi_{S(1)}^*\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi_{S(1)}^*\pi(X)\mid A=1,C=0,X\right]\right]\\
=&\mathbb{E}\left[\mu_1^*(X)\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left[\frac{P(A=1,C=0\mid X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mathbb{E}\left[S=1\mid A=1,C=0,X\right]-p_1^*(X)\right\}+p_1^*(X)\right]\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)\frac{p_0(X)}{p_1(X)}\left[\frac{\varphi_1(X)K_1(X)}{\varphi_1^*(X)K_1^*(X)}\left\{p_1(X)-p_1(X)\right\}+p_1(X)\right]\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right].
\end{aligned} \tag{26}$$

The third term in (16) equals

$$\begin{aligned}
&\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi_{Y(1)S(1)}^*\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi_{Y(1)S(1)}^*\pi(X)\mid A=1,C=0,X\right]\right]\\
=&\mathbb{E}\left[\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left\{\frac{P(A=1,C=0\mid X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mathbb{E}[YS\mid A=1,C=0,X]-\mu_1^*(X)p_1^*(X)\right\}+\mu_1^*(X)p_1^*(X)\right\}\right]\\
=&\mathbb{E}\left[\pi(X)\frac{p_0(X)}{p_1(X)}\left\{\frac{\varphi_1(X)K_1(X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mu_1(X)p_1(X)-\mu_1(X)p_1(X)\right\}+\mu_1(X)p_1(X)\right\}\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right].
\end{aligned} \tag{27}$$

The fourth term in (16) equals

$$\begin{aligned}
&\mathbb{E}\left[\psi_{Y(0)S(0)}^*\{1-\pi(X)\}\right]=\mathbb{E}\left[\mathbb{E}\left[\psi_{Y(0)S(0)}^*\{1-\pi(X)\}\mid A=0,C=0,X\right]\right]\\
=&\mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{P(A=0,C=0\mid X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mathbb{E}[YS\mid A=0,C=0,X]-\mu_0^*(X)p_0^*(X)\right\}+\mu_0^*(X)p_0^*(X)\right\}\right]\\
=&\mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{\varphi_0(X)K_0(X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mu_0(X)p_0(X)-\mu_0(X)p_0(X)\right\}+\mu_0(X)p_0(X)\right\}\right]
\end{aligned}$$

$$=\mathbb{E}\left[\mu_0(X)\{1-\pi(X)\}p_0(X)\right]. \tag{28}$$

Combining (16) and (25)–(28) , we have

$$\begin{aligned}
&\mathbb{E}\left[\left[\mu_1^*(X)\left\{\psi_{S(0)}^* - \frac{p_0^*(X)}{p_1^*(X)}\psi_{S(1)}^*\right\} + \frac{p_0^*(X)}{p_1^*(X)}\psi_{Y(1)S(1)}^*\right]\pi(X) + \psi_{Y(0)S(0)}^*\{1-\pi(X)\}\right] \\
=&\mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right] - \mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right] + \mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right] \\
&+ \mathbb{E}\left[\mu_0(X)\{1-\pi(X)\}p_0(X)\right] \\
=&\mathbb{E}\left[p_0(X)\left[\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right]\right] = N.
\end{aligned}$$

Therefore, under $\mathcal{M}_{\text{sp+om}}$, $\widehat{V}^{\text{MR}}(\pi)$ is consistent for $V(\pi)$.

Under $\mathcal{M}_{\text{ps+cs+sp+om}}$, we show that the influence function of $\widehat{V}^{\text{MR}}(\pi)$ is the same as the EIF in Theorem 3.7 and therefore $\widehat{V}^{\text{MR}}(\pi)$ achieves the local efficiency. Let θ denote the parameter vector containing α, η,γ, and ζ. Let θ^* be the probability limit of $\widehat{\theta}$. Let $\mathbb{P}_n\{g(O;\widehat{\theta})\} = 0$ be the estimation equations for θ. By Taylor expansion, we have

$$\mathbb{P}_n\{g(O;\widehat{\theta})\} = \mathbb{P}_n\{g(O;\theta^*)\} + \mathbb{P}\{\dot{g}(O;\theta^*)\}(\widehat{\theta}-\theta) + o_p(n^{-1/2}) = 0. \tag{29}$$

For a ratio estimator $\mathbb{P}_n\{N(O;\widehat{\theta})\}/\mathbb{P}_n\{D(O;\widehat{\theta})\}$, by Taylor expansion,

$$\frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}_n\{D(O;\widehat{\theta})\}} = \frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}\{D(O;\theta^*)\}} - \frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2}\left[\mathbb{P}_n\{D(O;\widehat{\theta})\} - \mathbb{P}\{D(O;\theta^*)\}\right] + o_p(n^{-1/2}). \tag{30}$$

$$\mathbb{P}_n\{N(O;\widehat{\theta})\} = \mathbb{P}_n\{N(O;\theta^*)\} + \mathbb{P}\{\dot{N}(O;\theta^*)\}(\widehat{\theta}-\theta) + o_p(n^{-1/2}), \tag{31}$$

$$\mathbb{P}_n\{D(O;\widehat{\theta})\} = \mathbb{P}_n\{D(O;\theta^*)\} + \mathbb{P}\{\dot{D}(O;\theta^*)\}(\widehat{\theta}-\theta) + o_p(n^{-1/2}). \tag{32}$$

For $\widehat{V}^{\text{MR}}(\pi) = \frac{\mathbb{P}_n(\widehat{\phi}_\pi)}{\mathbb{P}_n\{\widehat{\psi}_{S(0)}\}}$, we have $N(O;\theta) = \phi_\pi(\theta)$ and $D(O;\theta) = \psi_{S(0)}(\theta)$. Combining (29)–(32), we have

$$\begin{aligned}
&\widehat{V}^{\text{MR}}(\pi) - V(\pi) = \frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}_n\{D(O;\widehat{\theta})\}} - V(\pi) \\
=&\frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}\{D(O;\theta^*)\}} - \frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2}\left[\mathbb{P}_n\{D(O;\widehat{\theta})\} - \mathbb{P}\{D(O;\theta^*)\}\right] - V(\pi) + o_p(n^{-1/2}) \\
=&\frac{\mathbb{P}_n\{N(O;\theta^*)\} + \mathbb{P}\{\dot{N}(O;\theta^*)\}(\widehat{\theta}-\theta)}{\mathbb{P}\{D(O;\theta^*)\}} - \frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2}\left[\mathbb{P}_n\{D(O;\theta^*)\} + \mathbb{P}\{\dot{D}(O;\theta^*)\}(\widehat{\theta}-\theta)\right] + o_p(n^{-1/2}) \\
=&\mathbb{P}_n\left[\frac{N(O;\theta^*)}{\mathbb{P}\{D(O;\theta^*)\}} - \frac{\mathbb{P}\{\dot{N}(O;\theta^*)\}}{\mathbb{P}\{D(O;\theta^*)\}\mathbb{P}\{\dot{g}(O;\theta^*)\}}g(O;\theta^*) - \frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2\}}D(O;\theta^*) + \right. \\
&\left.\frac{\mathbb{P}\{\dot{D}(O;\theta^*)\}}{\mathbb{P}\{\dot{g}(O;\theta^*)\}}g(O;\theta^*)\right] + o_p(n^{-1/2}).
\end{aligned} \tag{33}$$

By the central limit theorem (CLT), $\widehat{V}^{\text{MR}}(\pi) - V(\pi)$ is asymptotically normal. Under $\mathcal{M}_{\text{ps+cs+sp+om}}$, $\phi_\pi(\theta^*) = \phi_\pi$, $\psi_{S(0)}(\theta^*) = \psi_{S(0)}$, $\mathbb{P}\{N(O;\theta^*)\} = \mathbb{P}\left[p_0(X)\left[\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right]\right]$, and $\mathbb{P}\{D(O;\theta^*)\} = p_0$. We can verify that $\mathbb{P}\{\dot{N}(O;\theta^*)\} = \mathbb{P}\{\dot{D}(O;\theta^*)\} = 0$. Continuing (33),

$$\begin{aligned}
&\widehat{V}^{\text{MR}}(\pi) - V(\pi) \\
=&\mathbb{P}_n\left[\frac{N(O;\theta^*)}{\mathbb{P}\{D(O;\theta^*)\}} - \frac{\mathbb{P}\{\dot{N}(O;\theta^*)\}}{\mathbb{P}\{D(O;\theta^*)\}\mathbb{P}\{\dot{g}(O;\theta^*)\}}g(O;\theta^*) - \frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2\}}D(O;\theta^*) + \right. \\
&\left.\frac{\mathbb{P}\{\dot{D}(O;\theta^*)\}}{\mathbb{P}\{\dot{g}(O;\theta^*)\}}g(O;\theta^*)\right] + o_p(n^{-1/2})
\end{aligned}$$

$$
\begin{aligned}
=&\mathbb{P}_n\left[\frac{N(O;\theta^*)}{\mathbb{P}\{D(O;\theta^*)\}}-\frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2}D(O;\theta^*)\right]+o_p(n^{-1/2})\\
=&\mathbb{P}_n\left[\frac{\phi_\pi-V(\pi)\psi_{S(0)}}{p_0}\right]+o_p(n^{-1/2}).
\end{aligned}
$$

This completes the proof. □

A.4. Proof of Theorem 4.5

Proof. We show that $\widehat{V}^{\mathrm{MR}}(\pi)$ constructed by $\{\widehat{\varphi}_a(x),\widehat{K}_a(x),\widehat{p}_a(x),\widehat{\mu}_a(x)\}$ is asymptotically normal and its influence function is the same as the EIF in Theorem 3.7, and therefore achieving the semiparametric efficiency. Let θ denote the nuisance functions $\{\varphi_a(x),K_a(x),p_a(x),\mu_a(x)\}$ and let θ^* be the probability limit of $\widehat{\theta}$.

Notice that $\widehat{V}^{\mathrm{MR}}(\pi)$ is a ratio estimator $\mathbb{P}_n\{N(O;\widehat{\theta})\}/\mathbb{P}_n\{D(O;\widehat{\theta})\}$, where $N(O;\theta)=\phi_\pi(\theta)$ and $D(O;\theta)=\psi_{S(0)}(\theta)$. We continue the Taylor expansion of a ratio estimator as in (30). Assumption 4.4(ii) implies that $\theta^*=\{\varphi_a(x),K_a(x),p_a(x),\mu_a(x)\}$, and thus $\phi_\pi(\theta^*)=\phi_\pi$ and $\psi_{S(0)}(\theta^*)=\psi_{S(0)}$. By the empirical process theory, we have

$$
\begin{aligned}
&\mathbb{P}_n\{N(O;\widehat{\theta})\}-\mathbb{P}\{N(O;\theta^*)\}=(\mathbb{P}_n-\mathbb{P})\{N(O;\widehat{\theta})\}+\mathbb{P}\{N(O;\widehat{\theta})-N(O;\theta^*)\}\\
=&(\mathbb{P}_n-\mathbb{P})\{N(O;\theta^*)\}+\mathbb{P}\{N(O;\widehat{\theta})-N(O;\theta^*)\}+o_p(n^{-1/2}),
\end{aligned}\tag{34}
$$

where the second equality follows by Assumption 4.4(ii). It remains to analyze the second term $\mathbb{P}\{N(O;\widehat{\theta})-N(O;\theta^*)\}$.

$$
\begin{aligned}
&|\mathbb{P}\{N(O;\widehat{\theta})-N(O;\theta^*)\}|\\
=&\Bigg|\mathbb{P}\left[\left[\widehat{\mu}_1(X)\left\{\widehat{\psi}_{S(0)}-\frac{\widehat{p}_0(X)}{\widehat{p}_1(X)}\widehat{\psi}_{S(1)}\right\}+\frac{\widehat{p}_0(X)}{\widehat{p}_1(X)}\widehat{\psi}_{Y(1)S(1)}\right]\pi(X)+\widehat{\psi}_{Y(0)S(0)}\{1-\pi(X)\}\right]\\
&-\mathbb{P}\left[p_0(X)\left[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\right]\right]\Bigg|\\
\leq&\left|\mathbb{P}\left[\pi(X)\widehat{\mu}_1(X)\frac{\{\varphi_0(X)K_0(X)-\widehat{\varphi}_0(X)\widehat{K}_0(X)\}\{p_0(X)-\widehat{p}_0(X)\}}{\widehat{\varphi}_0(X)\widehat{K}_0(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\pi(X)\widehat{\mu}_1(X)\frac{\widehat{p}_0(X)}{\widehat{p}_1(X)}\frac{\{\varphi_1(X)K_1(X)-\widehat{\varphi}_1(X)\widehat{K}_1(X)\}\{p_1(X)-\widehat{p}_1(X)\}}{\widehat{\varphi}_1(X)\widehat{K}_1(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\pi(X)\mu_1(X)\frac{\widehat{p}_0(X)}{\widehat{p}_1(X)}\frac{\{\varphi_1(X)K_1(X)-\widehat{\varphi}_1(X)\widehat{K}_1(X)\}\{p_1(X)-\widehat{p}_1(X)\}}{\widehat{\varphi}_1(X)\widehat{K}_1(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\pi(X)\widehat{p}_0(X)\frac{\{\varphi_1(X)K_1(X)-\widehat{\varphi}_1(X)\widehat{K}_1(X)\}\{\mu_1(X)-\widehat{\mu}_1(X)\}}{\widehat{\varphi}_1(X)\widehat{K}_1(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\{1-\pi(X)\}\mu_0(X)\frac{\{\varphi_0(X)K_0(X)-\widehat{\varphi}_0(X)\widehat{K}_0(X)\}\{p_0(X)-\widehat{p}_0(X)\}}{\widehat{\varphi}_0(X)\widehat{K}_0(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\{1-\pi(X)\}\widehat{p}_0(X)\frac{\{\varphi_0(X)K_0(X)-\widehat{\varphi}_0(X)\widehat{K}_0(X)\}\{\mu_0(X)-\widehat{\mu}_0(X)\}}{\widehat{\varphi}_0(X)\widehat{K}_0(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\pi(X)\frac{p_0(X)}{\widehat{p}_1(X)}\{\mu_1(X)-\widehat{\mu}_1(X)\}\{p_1(X)-\widehat{p}_1(X)\}\right]\right|+\\
&\left|\mathbb{P}\left[\pi(X)\frac{p_1(X)}{\widehat{p}_1(X)}\{\mu_1(X)-\widehat{\mu}_1(X)\}\{p_0(X)-\widehat{p}_0(X)\}\right]\right|.
\end{aligned}
$$

By the Cauchy–Schwarz inequality and Assumption 4.4(iii), it follows that for some constant $l>0$, we have

$$
\begin{aligned}
&|\mathbb{P}\{N(O;\widehat{\theta})-N(O;\theta^*)\}|\\
\leq& l\times\|\varphi_0(X)K_0(X)-\widehat{\varphi}_0(X)\widehat{K}_0(X)\|_2\times[\|p_0(X)-\widehat{p}_0(X)\|_2+\|\mu_0(X)-\widehat{\mu}_0(X)\|_2]+
\end{aligned}
$$

$$l \times \|p_1(X) - \widehat{p}_1(X)\|_2 \times \left[\|\varphi_1(X)K_1(X) - \widehat{\varphi}_1(X)\widehat{K}_1(X)\|_2 + \|\mu_1(X) - \widehat{\mu}_1(X)\|_2\right] +$$
$$l \times \|p_0(X) - \widehat{p}_0(X)\|_2 \times \|\mu_0(X) - \widehat{\mu}_0(X)\|_2 = o_p(-n^{1/2}). \tag{35}$$

Continuing with (34), we have $\mathbb{P}_n\{N(O;\widehat{\theta})\} - \mathbb{P}\{N(O;\theta^*)\} = (\mathbb{P}_n - \mathbb{P})\{N(O;\theta^*)\} + o_p(n^{-1/2})$. Similarly, we can show $\mathbb{P}_n\{D(O;\widehat{\theta})\} - \mathbb{P}\{D(O;\theta^*)\} = (\mathbb{P}_n - \mathbb{P})\{D(O;\theta^*)\} + o_p(n^{-1/2})$. Continuing with (30), we have

$$\begin{aligned}
&\widehat{V}^{\mathrm{MR}}(\pi) - V(\pi) = \frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}_n\{D(O;\widehat{\theta})\}} - V(\pi)\\
=&\frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}\{D(O;\theta^*)\}} - \frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2}\left[\mathbb{P}_n\{D(O;\widehat{\theta})\} - \mathbb{P}\{D(O;\theta^*)\}\right] - V(\pi) + o_p(n^{-1/2})\\
=&\frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}\{D(O;\theta^*)\}} - \frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2}(\mathbb{P}_n - \mathbb{P})\{D(O;\theta^*)\} - \frac{\mathbb{P}\{N(O;\theta^*)\}}{\mathbb{P}\{D(O;\theta^*)\}} + o_p(n^{-1/2})\\
=&\frac{\mathbb{P}_n\{N(O;\widehat{\theta})\} - \mathbb{P}\{N(O;\theta^*)\}}{\mathbb{P}\{D(O;\theta^*)\}} - \frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2}(\mathbb{P}_n - \mathbb{P})\{D(O;\theta^*)\} + o_p(n^{-1/2})\\
=&\frac{(\mathbb{P}_n - \mathbb{P})\{N(O;\theta^*)\}}{\mathbb{P}\{D(O;\theta^*)\}} - \frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2}(\mathbb{P}_n - \mathbb{P})\{D(O;\theta^*)\} + o_p(n^{-1/2})\\
=&\frac{\mathbb{P}_n\{N(O;\theta^*)\}}{\mathbb{P}\{D(O;\theta^*)\}} - \frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta^*)\}]^2}\mathbb{P}_n\{D(O;\theta^*)\} + o_p(n^{-1/2})\\
=&\frac{\mathbb{P}_n\{\phi_\pi\}}{p_0} - V(\pi)\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0} + o_p(n^{-1/2})\\
=&\mathbb{P}_n\left[\frac{\phi_\pi - V(\pi)\psi_{S(0)}}{p_0}\right] + o_p(n^{-1/2}).
\end{aligned}$$

This completes the proof. □

A.5. Proof of Lemma 5.2

Proof. Step 1: We first show that $\widehat{\beta}$ converges in probability to β^* as $n \to \infty$, by checking three conditions for the Argmax Theorem:

(1) By Assumption 5.1(i), $V(\beta)$ is twice continuously differentiable at a neighborhood of β^*.

(2) In Section A.4, we have shown that for any β, $\widehat{V}^{\mathrm{MR}}(\beta)$ is consistent for $V(\beta)$.

(3) Since $\widehat{\beta} = \mathrm{argmax}_{\beta:\|\beta\|_2=1}\widehat{V}^{\mathrm{MR}}(\beta)$, we have the estimated policy as $\pi(x;\widehat{\beta}) = \mathbb{I}\left(\tilde{x}^T\widehat{\beta} > 0\right)$ and the corresponding value estimator $\widehat{V}^{\mathrm{MR}}(\widehat{\beta})$ such that

$$\widehat{V}^{\mathrm{MR}}(\widehat{\beta}) \geq \sup_{\beta:\|\beta\|_2=1}\widehat{V}^{\mathrm{MR}}(\beta).$$

Thus we have $\widehat{\beta}$ converges in probability to β^* as $n \to \infty$.

Step 2: We show that $n^{1/3}\|\widehat{\beta} - \beta^*\|_2 = O_p(1)$. We check three conditions of the Theorem 14.4: Rate of convergence in Kosorok (2008):

(1) For every β in a neighborhood of β^*, i.e., $\|\beta - \beta^*\|_2 < \varepsilon$, for some constant $\varepsilon > 0$, we take the second order Taylor expansion on $V(\beta)$ at $\beta = \beta^*$,

$$\begin{aligned}
V(\beta) - V(\beta^*) &= V'(\beta^*)\|\beta - \beta^*\|_2 + \frac{1}{2}V''(\beta^*)\|\beta - \beta^*\|_2^2 + o\left(\|\beta - \beta^*\|_2^2\right)\\
&= \frac{1}{2}V''(\beta^*)\|\beta - \beta^*\|_2^2 + o\left(\|\beta - \beta^*\|_2^2\right) \quad (V'(\beta^*) = 0).
\end{aligned}$$

Since $V''(\beta^*) < 0$, there exists $c_0 = -\frac{1}{2}V''(\beta^*) > 0$ such that

$$V(\beta) - V(\beta^*) < c_0\|\beta - \beta^*\|_2^2. \tag{36}$$

(2) From Section A.4, we have

$$\widehat{V}^{\mathrm{MR}}(\beta)-V(\beta)=\mathbb{P}_n\left[\frac{\phi_{\pi(X;\beta)}-V(\beta)\psi_{S(0)}}{p_0}\right]+o_p(n^{-1/2}),$$
$$\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)=\mathbb{P}_n\left[\frac{\phi_{\pi(X;\beta^*)}-V(\beta^*)\psi_{S(0)}}{p_0}\right]+o_p(n^{-1/2}).$$

Let $\mathbb{E}^*(\cdot)$ denote the outer expectation. Then, for all n large enough and sufficiently small ε, the centered process $\widehat{V}^{\mathrm{MR}}-V$ satisfies

$$\begin{aligned}
&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{MR}}(\beta)-V(\beta)-\{\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\}\right|\right]\\
=&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\mathbb{P}_n\left\{\frac{\phi_{\pi(X;\beta)}-V(\beta)\psi_{S(0)}}{p_0}\right\}-\mathbb{P}_n\left\{\frac{\phi_{\pi(X;\beta^*)}-V(\beta^*)\psi_{S(0)}}{p_0}\right\}\right|\right]\\
=&\mathbb{E}^*\Bigg[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\Bigg|\mathbb{P}_n\left\{\frac{\phi_{\pi(X;\beta)}-V(\beta)p_0+V(\beta)p_0-V(\beta)\psi_{S(0)}}{p_0}\right\}\\
&-\mathbb{P}_n\left\{\frac{\phi_{\pi(X;\beta^*)}-V(\beta^*)p_0+V(\beta^*)p_0-V(\beta^*)\psi_{S(0)}}{p_0}\right\}\Bigg|\Bigg]\\
=&\mathbb{E}^*\Bigg[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\Bigg|\frac{\mathbb{P}_n\{\phi_{\pi(X;\beta)}-\phi_{\pi(X;\beta^*)}\}}{p_0}-\{V(\beta)-V(\beta^*)\}\\
&+\{V(\beta)-V(\beta^*)\}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\Bigg|\Bigg]\\
\leq&\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\frac{\mathbb{P}_n\{\phi_{\pi(X;\beta)}-\phi_{\pi(X;\beta^*)}\}}{p_0}-\{V(\beta)-V(\beta^*)\}\right|\right]}_{\tau_1}\\
&+\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\{V(\beta)-V(\beta^*)\}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right]}_{\tau_2}.
\end{aligned}\tag{37}$$

Note that

$$\begin{aligned}
&\phi_{\pi(X;\beta)}-\phi_{\pi(X;\beta^*)}\\
=&\left[\mu_1(X)\left\{\psi_{S(0)}-\frac{p_0(X)}{p_1(X)}\psi_{S(1)}\right\}+\frac{p_0(X)}{p_1(X)}\psi_{Y(1)S(1)}-\psi_{Y(0)S(0)}\right]\left\{\mathbb{I}(\tilde{X}^T\beta>0)-\mathbb{I}(\tilde{X}^T\beta^*>0)\right\}.
\end{aligned}$$

We define a class of functions

$$\begin{aligned}
\mathcal{F}_\beta(y,a,c,s,x)=\Bigg\{&\left[\mu_1(x)\left\{\psi_{S(0)}-\frac{p_0(x)}{p_1(x)}\psi_{S(1)}\right\}+\frac{p_0(x)}{p_1(x)}\psi_{Y(1)S(1)}-\psi_{Y(0)S(0)}\right]\times\\
&\left\{\mathbb{I}(\tilde{x}^T\beta>0)-\mathbb{I}(\tilde{x}^T\beta^*>0)\right\}:\|\beta-\beta^*\|_2<\varepsilon\Bigg\},
\end{aligned}$$

where

$$\begin{aligned}
\psi_{S(0)}&=\frac{\mathbb{I}\{a=0\}\mathbb{I}(c=0)\{s-p_0(x)\}}{\varphi_0(x)K_0(x)}+p_0(x),\\
\psi_{S(1)}&=\frac{\mathbb{I}\{a=1\}\mathbb{I}(c=0)\{s-p_1(x)\}}{\varphi_1(x)K_1(x)}+p_1(x),
\end{aligned}$$

$$\begin{aligned}\psi_{Y(0)S(0)} &= \frac{\mathbb{I}(a=0)\mathbb{I}(c=0)\{ys-\mu_0(x)p_0(x)\}}{\varphi_0(x)K_0(x)}\mu_0(x)p_0(x),\\ \psi_{Y(1)S(1)} &= \frac{\mathbb{I}(a=1)\mathbb{I}(c=0)\{ys-\mu_1(x)p_1(x)\}}{\varphi_1(x)K_1(x)}\mu_1(x)p_1(x).\end{aligned}$$

Let $M=\sup\left|\mu_1(x)\left\{\psi_{S(0)}-\frac{p_0(x)}{p_1(x)}\psi_{S(1)}\right\}+\frac{p_0(x)}{p_1(x)}\psi_{Y(1)S(1)}-\psi_{Y(0)S(0)}\right|$. By Assumption 3.5, we have $M<\infty$. When $\|\beta-\beta^*\|_2<\varepsilon$, there exists a constant $0<k_0<\infty$, such that $|\tilde{x}^T(\beta-\beta^*)|<k_0\varepsilon$. For the indicator function $\mathbb{I}(-k_0\varepsilon\le\tilde{x}^T\beta^*\le k_0\varepsilon)$,

(i) when $-k_0\varepsilon\le\tilde{x}^T\beta^*\le k_0\varepsilon$,

$$\mathbb{I}(-k_0\varepsilon\le\tilde{x}^T\beta^*\le k_0\varepsilon)=1\ge\left|I\left\{\tilde{x}^T\beta>0\right\}-I\left\{\tilde{x}^T\beta^*>0\right\}\right|;$$

(ii) when $\tilde{x}^T\beta^*>k_0\varepsilon$, $\tilde{x}^T\beta=\tilde{x}^T(\beta-\beta^*)+\tilde{x}^T\beta^*>-k_0\varepsilon+k_0\varepsilon>0$,

$$\mathbb{I}(-k_0\varepsilon\le\tilde{x}^T\beta^*\le k_0\varepsilon)=0=\left|I\left\{\tilde{x}^T\beta>0\right\}-I\left\{\tilde{x}^T\beta^*>0\right\}\right|;$$

(iii) when $\tilde{x}^T\beta^*<-k_0\varepsilon$, $\tilde{x}^T\beta=\tilde{x}^T(\beta-\beta^*)+\tilde{x}^T\beta^*<k_0\varepsilon+(-k_0\varepsilon)<0$,

$$\mathbb{I}(-k_0\varepsilon\le\tilde{x}^T\beta^*\le k_0\varepsilon)=0=\left|I\left\{\tilde{x}^T\beta>0\right\}-I\left\{\tilde{x}^T\beta^*>0\right\}\right|.$$

Therefore, we always have $\mathbb{I}(-k_0\varepsilon\le\tilde{x}^T\beta^*\le k_0\varepsilon)\ge\left|\mathbb{I}(\tilde{x}^T\beta>0)-\mathbb{I}(\tilde{x}^T\beta^*>0)\right|$ when $\|\beta-\beta^*\|_2<\varepsilon$.

We then define the envelope of $\mathcal{F}_\beta(y,a,c,s,x)$ as $F=M\mathbb{I}(-k_0\varepsilon\le\tilde{x}^T\beta^*\le k_0\varepsilon)$. By Assumption 5.1(ii), there exits a constant $0<k_1<\infty$, such that

$$\|F\|_{P,2}=M\sqrt{P(-k_0\varepsilon\le\tilde{x}^T\beta^*\le k_0\varepsilon)}\le M\sqrt{k_1\cdot 2k_0\varepsilon}=M\sqrt{2k_0k_1}\varepsilon^{1/2}<\infty.$$

Since $\mathcal{F}_\beta$ is a class of indicator functions, by the conclusion of Lemma 9.6 and Lemma 9.9 in Kosorok (2008), $\mathcal{F}_\beta$ is a VC class of functions. Thus, the entropy of $\mathcal{F}_\beta$, denoted as $J^*_{[]}(1,\mathcal{F})$, is finite, i.e., $J^*_{[]}(1,\mathcal{F})<\infty$. Next, we consider the following empirical process indexed by β,

$$\mathbb{G}_n\mathcal{F}_\beta=n^{-1/2}\sum_{i=1}^n\left[\mathcal{F}_\beta\left(O_i\right)-\mathbb{E}\left[\mathcal{F}_\beta\left(O\right)\right]\right].$$

Note that $\mathbb{G}_n\mathcal{F}_\beta=n^{1/2}\left[\frac{\mathbb{P}_n\{\phi_{\pi(X;\beta)}-\phi_{\pi(X;\beta^*)}\}}{p_0}-\{V(\beta)-V(\beta^*)\}\right]$. By applying Theorem 11.2 in Kosorok (2008), we have

$$\begin{aligned}\tau_1&=\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\frac{\mathbb{P}_n\{\phi_{\pi(X;\beta)}-\phi_{\pi(X;\beta^*)}\}}{p_0}-\{V(\beta)-V(\beta^*)\}\right|\right]\\&=\mathbb{E}^*\left[\sup_{\|\beta-\beta^*\|_2<\varepsilon}|\mathbb{G}_n\mathcal{F}_\beta|\right]\le lJ^*_{[]}(1,\mathcal{F})\|F\|_{P,2}\le lJ^*_{[]}(1,\mathcal{F})M\sqrt{2k_0k_1}\varepsilon^{1/2},\end{aligned}$$

where l is a finite constant.

Let $c_1\equiv lJ^*_{[]}(1,\mathcal{F})M\sqrt{2k_0k_1}$, since l, $J^*_{[]}(1,\mathcal{F})$, M, k_0 and k_1 are bounded, we have $c_1<\infty$, i.e.,

$$\tau_1\le c_1\varepsilon^{1/2}. \tag{38}$$

$$\tau_2=\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\{V(\beta)-V(\beta^*)\}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right]$$

$$
\begin{aligned}
=&\mathbb{E}^*\left[\left\{\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\{V(\beta)-V(\beta^*)\}\right|\right\}\left|n^{1/2}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right]\\
\leq&\left[\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\{V(\beta)-V(\beta^*)\}\right|\right]\mathbb{E}^*\left[\left|n^{1/2}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right]\\
\leq&O(\varepsilon),
\end{aligned}
\tag{39}
$$

where the last inequality follows from the mean value theory and $n^{1/2}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}=O_p(1)$. Combining (37), (38), and (39), we have

$$
\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{MR}}(\beta)-V(\beta)-\{\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\}\right|\right]\leq\tau_1+\tau_2\leq c_1\varepsilon^{1/2}+O(\varepsilon)\leq c_2\lambda(\varepsilon),\tag{40}
$$

where c_2 is a finite positive constant, and $\lambda(\varepsilon)=\varepsilon^{1/2}+\varepsilon$.

(3) Let $\alpha=\frac{3}{2}<2$. $\frac{\lambda(\varepsilon)}{\varepsilon^\alpha}=\varepsilon^{-1}+\varepsilon^{-1/2}$ is decreasing not depending on n. Let $r_n=n^{1/3}$, and then r_n satisfies

$$
r_n^2\lambda(r_n^{-1})=n^{2/3}\lambda(n^{-1/3})=n^{1/2}+n^{1/3}=O(n^{1/2}).
$$

Combining (1)-(3) in Step 2, by the Theorem 14.4 in (Kosorok, 2008), we have $n^{1/3}\|\widehat{\beta}-\beta^*\|_2=O_p(1)$. □

A.6. Proof of Theorem 5.3

Proof. Notice that

$$
\begin{aligned}
&\sqrt{n}\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-V(\beta^*)\right\}=\sqrt{n}\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-\widehat{V}^{\mathrm{MR}}(\beta^*)+\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\right\}\\
=&\sqrt{n}\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-\widehat{V}^{\mathrm{MR}}(\beta^*)\right\}+\sqrt{n}\left\{\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\right\}.
\end{aligned}
$$

First, we show

$$
\sqrt{n}\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-\widehat{V}^{\mathrm{MR}}(\beta^*)\right\}=o_p(1).
$$

which is sufficient to show

$$
\sqrt{n}\left[\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-\widehat{V}^{\mathrm{MR}}(\beta^*)\right\}-\left\{V(\widehat{\beta})-V(\beta^*)\right\}\right]=o_p(1),
$$

$$
\sqrt{n}\left\{V(\widehat{\beta})-V(\beta^*)\right\}=o_p(1).
$$

We take the Taylor expansion on $V(\widehat{\beta})$ at β^*,

$$
\begin{aligned}
\sqrt{n}\left\{V(\widehat{\beta})-V(\beta^*)\right\}&=\sqrt{n}\left\{V'(\beta^*)\|\widehat{\beta}-\beta^*\|_2+\frac{1}{2}V''(\beta^*)\|\widehat{\beta}-\beta^*\|_2^2+o_p\left(\|\widehat{\beta}-\beta^*\|_2^2\right)\right\}\\
&=\sqrt{n}\left\{\frac{1}{2}V''(\beta^*)\|\widehat{\beta}-\beta^*\|_2^2+o_p\left(\|\widehat{\beta}-\beta^*\|_2^2\right)\right\}\quad(V'(\beta^*)=0)\\
&=\sqrt{n}\left\{\frac{1}{2}V''(\beta^*)O_p(n^{-2/3})+o_p(n^{-2/3})\right\}\quad(\text{Lemma 5.2})\\
&=\frac{1}{2}V''(\beta^*)O_p(n^{-1/6})=o_p(1).
\end{aligned}
\tag{41}
$$

By Lemma 5.2, we have $\|\widehat{\beta}-\beta^*\|_2=c_3n^{-1/3}$, for some constant $0<c_3<\infty$. From (40), we have

$$
\begin{aligned}
&\sqrt{n}\left[\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-\widehat{V}^{\mathrm{MR}}(\beta^*)\right\}-\left\{V(\widehat{\beta})-V(\beta^*)\right\}\right]\\
\leq&\mathbb{E}^*\left[n^{1/2}\sup_{\|\widehat{\beta}-\beta^*\|_2<c_3n^{-1/3}}\left|\widehat{V}^{\mathrm{MR}}(\beta)-V(\beta)-\{\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\}\right|\right]
\end{aligned}
$$

$$\leq c_2\left(c_3^{1/2}n^{-1/6}+c_3n^{-1/3}\right)=o_p(1). \tag{42}$$

By (41) and (42), we have

$$\begin{aligned}&\sqrt{n}\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-\widehat{V}^{\mathrm{MR}}(\beta^*)\right\}\\ =&\sqrt{n}\left[\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-\widehat{V}^{\mathrm{MR}}(\beta^*)\right\}-\left\{V(\widehat{\beta})-V(\beta^*)\right\}\right]+\sqrt{n}\left\{V(\widehat{\beta})-V(\beta^*)\right\}\\ =&o_p(1)+o_p(1)=o_p(1).\end{aligned}$$

Then,

$$\begin{aligned}\sqrt{n}\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-V(\beta^*)\right\}=&\sqrt{n}\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-\widehat{V}^{\mathrm{MR}}(\beta^*)\right\}+\sqrt{n}\left\{\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\right\}\\ =&\sqrt{n}\left\{\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\right\}+o_p(1).\end{aligned}$$

Therefore, $\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-V(\beta^*)$ has the same asymptotic distribution as $\widehat{V}^{\mathrm{MR}}(\beta^*)$. From Thoerem 4.5, $\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)$ is asymptotically normal with mean zero and variance $\Upsilon(\pi(x;\beta^*))$. This completes the proof. □

B. Additional Theoretical Properties for OPL

When nuisance functions in the MR estimator are estimated using parametric models, we have the following results.

Lemma B.1. *Suppose that Assumptions 3.1–3.5, 4.1, and 5.1 hold, under* $\mathcal{M}_{\mathrm{ps+cs+sp}}\cup\mathcal{M}_{\mathrm{ps+cs+om}}\cup\mathcal{M}_{\mathrm{sp+om}}$, *we have* $n^{1/3}\|\widehat{\beta}-\beta^*\|_2=O_p(1)$.

Proof. Step 1: We first show that $\widehat{\beta}$ converges in probability to β^* as $n\to\infty$, by checking three conditions for the Argmax Theorem:

(1) By Assumption 5.1(i), $V(\beta)$ is twice continuously differentiable at a neighborhood of β^*.

(2) In Section A.3, we have shown that for any β, $\widehat{V}^{\mathrm{MR}}(\beta)$ is consistent for $V(\beta)$.

(3) Since $\widehat{\beta}=\mathrm{argmax}_{\beta:\|\beta\|_2=1}\widehat{V}^{\mathrm{MR}}(\beta)$, we have the estimated policy as $\pi(x;\widehat{\beta})=\mathbb{I}\left(\tilde{x}^T\widehat{\beta}>0\right)$ and the corresponding value estimator $\widehat{V}^{\mathrm{MR}}(\widehat{\beta})$ such that

$$\widehat{V}^{\mathrm{MR}}(\widehat{\beta})\geq\sup_{\beta:\|\beta\|_2=1}\widehat{V}^{\mathrm{MR}}(\beta).$$

Thus we have $\widehat{\beta}$ converges in probability to β^* as $n\to\infty$.

Step 2: We show that $n^{1/3}\|\widehat{\beta}-\beta^*\|_2=O_p(1)$. We check three conditions of the Theorem 14.4: Rate of convergence in Kosorok (2008):

(1) For every β in a neighborhood of β^*, i.e., $\|\beta-\beta^*\|_2<\varepsilon$, for some constant $\varepsilon>0$, we take the second order Taylor expansion on $V(\beta)$ at $\beta=\beta^*$,

$$\begin{aligned}V(\beta)-V(\beta^*)&=V'(\beta^*)\|\beta-\beta^*\|_2+\frac{1}{2}V''(\beta^*)\|\beta-\beta^*\|_2^2+o\left(\|\beta-\beta^*\|_2^2\right)\\ &=\frac{1}{2}V''(\beta^*)\|\beta-\beta^*\|_2^2+o\left(\|\beta-\beta^*\|_2^2\right)\quad(V'(\beta^*)=0).\end{aligned}$$

Since $V''(\beta^*)<0$, there exists $c_0=-\frac{1}{2}V''(\beta^*)>0$ such that

(2) Let $\widehat{V}^{\mathrm{MR}}(\beta)=\frac{\mathbb{P}_n\{N(O;\widehat{\theta},\beta)\}}{\mathbb{P}_n\{D(O;\widehat{\theta})\}}$, where $N(O;\theta,\beta)=\phi_{\pi(X;\beta)}(\theta)$ and $D(O;\theta)=\psi_{S(0)}(\theta)$. Under $\mathcal{M}_{\mathrm{ps+cs+sp}}\cup\mathcal{M}_{\mathrm{ps+cs+om}}\cup\mathcal{M}_{\mathrm{sp+om}}$, $\mathbb{P}\{N(O;\theta^*,\beta)/\mathbb{P}\{D(O;\theta^*)\}=V(\beta)$, $\mathbb{P}\{D(O;\theta^*)\}=p_0$. By (33), we have

$$
\begin{aligned}
&\widehat{V}^{\mathrm{MR}}(\beta) - V(\beta) \\
=&\mathbb{P}_n\left[\frac{\phi_{\pi(X;\beta)}(\theta^*)}{p_0} - \frac{V(\beta)\psi_{S(0)}(\theta^*)}{p_0} - \frac{\mathbb{P}\{\dot{N}(O;\theta^*,\beta)\}}{p_0\mathbb{P}\{\dot{g}(O;\theta^*)\}}g(O;\theta^*) + \frac{\mathbb{P}\{\dot{D}(O;\theta^*)\}}{\mathbb{P}\{\dot{g}(O;\theta^*)\}}g(O;\theta^*)\right] + o_p(n^{-1/2}).
\end{aligned}
$$

For all n large enough and sufficiently small ε, the centered process $\widehat{V}^{\mathrm{MR}} - V$ satisfies

$$
\begin{aligned}
&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{MR}}(\beta) - V(\beta) - \{\widehat{V}^{\mathrm{MR}}(\beta^*) - V(\beta^*)\}\right|\right] \\
=&\mathbb{E}^*\Bigg[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\Bigg|\frac{\mathbb{P}_n\{\phi_{\pi(X;\beta)}(\theta^*) - \phi_{\pi(X;\beta^*)}(\theta^*)\}}{p_0} - \{V(\beta) - V(\beta^*)\} \\
&+ \{V(\beta) - V(\beta^*)\}\left\{1 - \frac{\mathbb{P}_n\{\psi_{S(0)}(\theta^*)\}}{p_0}\right\} - \frac{\mathbb{P}\{\dot{N}(O;\theta^*,\beta)\} - \mathbb{P}\{\dot{N}(O;\theta^*,\beta^*)\}}{p_0\mathbb{P}\{\dot{g}(O;\theta^*)\}}\mathbb{P}_n\{g(O;\theta^*)\}\Bigg|\Bigg] \\
\le&\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\frac{\mathbb{P}_n\{\phi_{\pi(X;\beta)} - \phi_{\pi(X;\beta^*)}\}}{p_0} - \{V(\beta) - V(\beta^*)\}\right|\right]}_{\tau_1} \\
&+ \underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\{V(\beta) - V(\beta^*)\}\left\{1 - \frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right]}_{\tau_2} \\
&+ \underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\frac{\mathbb{P}\{\dot{N}(O;\theta^*,\beta)\} - \mathbb{P}\{\dot{N}(O;\theta^*,\beta^*)\}}{p_0\mathbb{P}\{\dot{g}(O;\theta^*)\}}\mathbb{P}_n\{g(O;\theta^*)\}\right|\right]}_{\tau_3}. \qquad (43)
\end{aligned}
$$

We can show similarly as in Section A.5, $\tau_1 + \tau_2 < c_1(\varepsilon^{1/2} + \varepsilon)$ for some constant $0 < c_1 < \infty$. It remains to analyze τ_3.

$$
\begin{aligned}
\tau_3 =& \mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\frac{\mathbb{P}\{\dot{N}(O;\theta^*,\beta)\} - \mathbb{P}\{\dot{N}(O;\theta^*,\beta^*)\}}{p_0\mathbb{P}\{\dot{g}(O;\theta^*)\}}\mathbb{P}_n\{g(O;\theta^*)\}\right|\right] \\
\le&\left[\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\frac{\mathbb{P}\{\dot{N}(O;\theta^*,\beta)\} - \mathbb{P}\{\dot{N}(O;\theta^*,\beta^*)\}}{p_0\mathbb{P}\{\dot{g}(O;\theta^*)\}}\right|\right]\mathbb{E}^*\left[\left|n^{1/2}\mathbb{P}_n\{g(O;\theta^*)\}\right|\right] \\
\le& O(\varepsilon).
\end{aligned}
$$

where the last inequality follows from the mean value theory and that $n^{1/2}\mathbb{P}_n\{g(O;\theta^*)\} = O_p(1)$.

Continuing (43), we have

$$
\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{MR}}(\beta) - V(\beta) - \{\widehat{V}^{\mathrm{MR}}(\beta^*) - V(\beta^*)\}\right|\right] \le \tau_1 + \tau_2 + \tau_3 \le c_1(\varepsilon^{1/2} + \varepsilon) + O(\varepsilon) \le c_2\lambda(\varepsilon), \qquad (44)
$$

where c_2 is a finite positive constant, and $\lambda(\varepsilon) = \varepsilon^{1/2} + \varepsilon$.

(3) Let $\alpha = \frac{3}{2} < 2$. $\frac{\lambda(\varepsilon)}{\varepsilon^\alpha} = \varepsilon^{-1} + \varepsilon^{-1/2}$ is decreasing not depending on n. Let $r_n = n^{1/3}$, and then r_n satisfies

$$
r_n^2\lambda(r_n^{-1}) = n^{2/3}\lambda(n^{-1/3}) = n^{1/2} + n^{1/3} = O(n^{1/2}).
$$

Combining (1)-(3) in Step 2, by the Theorem 14.4 in Kosorok (2008), we have $n^{1/3}\|\widehat{\beta} - \beta^*\|_2 = O_p(1)$. □

Theorem B.2. *Suppose that Assumptions 3.1–3.5, 4.1, and 5.1 hold, under* $\mathcal{M}_{\text{ps+cs+sp}} \cup \mathcal{M}_{\text{ps+cs+om}} \cup \mathcal{M}_{\text{sp+om}}$, $\widehat{V}^{\text{MR}}(\widehat{\beta}) - V(\beta)$ *is asymptotically normal with mean zero. Moreover, under* $\mathcal{M}_{\text{ps+cs+sp+om}}$, $\widehat{V}^{\text{MR}}(\widehat{\beta})$ *achieves the semiparametric efficiency bound* $\Upsilon(\pi(x;\beta^*))$.

Proof. Notice that

$$\sqrt{n}\left\{\widehat{V}^{\text{MR}}(\widehat{\beta}) - V(\beta^*)\right\} = \sqrt{n}\left\{\widehat{V}^{\text{MR}}(\widehat{\beta}) - \widehat{V}^{\text{MR}}(\beta^*) + \widehat{V}^{\text{MR}}(\beta^*) - V(\beta^*)\right\}$$
$$= \sqrt{n}\left\{\widehat{V}^{\text{MR}}(\widehat{\beta}) - \widehat{V}^{\text{MR}}(\beta^*)\right\} + \sqrt{n}\left\{\widehat{V}^{\text{MR}}(\beta^*) - V(\beta^*)\right\}.$$

First, we show

$$\sqrt{n}\left\{\widehat{V}^{\text{MR}}(\widehat{\beta}) - \widehat{V}^{\text{MR}}(\beta^*)\right\} = o_p(1).$$

which is sufficient to show

$$\sqrt{n}\left[\left\{\widehat{V}^{\text{MR}}(\widehat{\beta}) - \widehat{V}^{\text{MR}}(\beta^*)\right\} - \left\{V(\widehat{\beta}) - V(\beta^*)\right\}\right] = o_p(1),$$
$$\sqrt{n}\left\{V(\widehat{\beta}) - V(\beta^*)\right\} = o_p(1).$$

We take the Taylor expansion on $V(\widehat{\beta})$ at β^*,

$$\begin{aligned}
\sqrt{n}\left\{V(\widehat{\beta}) - V(\beta^*)\right\} &= \sqrt{n}\left\{V'(\beta^*)\|\widehat{\beta}-\beta^*\|_2 + \frac{1}{2}V''(\beta^*)\|\widehat{\beta}-\beta^*\|_2^2 + o_p\left(\|\widehat{\beta}-\beta^*\|_2^2\right)\right\} \\
&= \sqrt{n}\left\{\frac{1}{2}V''(\beta^*)\|\widehat{\beta}-\beta^*\|_2^2 + o_p\left(\|\widehat{\beta}-\beta^*\|_2^2\right)\right\} \quad (V'(\beta^*)=0) \\
&= \sqrt{n}\left\{\frac{1}{2}V''(\beta^*)O_p(n^{-2/3}) + o_p(n^{-2/3})\right\} \quad (\text{Lemma } 5.2) \\
&= \frac{1}{2}V''(\beta^*)O_p(n^{-1/6}) = o_p(1).
\end{aligned} \tag{45}$$

By Lemma 5.2, we have $\|\widehat{\beta}-\beta^*\|_2 = c_3 n^{-1/3}$, for some constant $0 < c_3 < \infty$. From (44), we have

$$\begin{aligned}
&\sqrt{n}\left[\left\{\widehat{V}^{\text{MR}}(\widehat{\beta}) - \widehat{V}^{\text{MR}}(\beta^*)\right\} - \left\{V(\widehat{\beta}) - V(\beta^*)\right\}\right] \\
\leq &\mathbb{E}^*\left[n^{1/2} \sup_{\|\widehat{\beta}-\beta^*\|_2 < c_3 n^{-1/3}} \left|\widehat{V}^{\text{MR}}(\beta) - V(\beta) - \{\widehat{V}^{\text{MR}}(\beta^*) - V(\beta^*)\}\right|\right] \\
\leq &c_2\left(c_3^{1/2} n^{-1/6} + c_3 n^{-1/3}\right) = o_p(1).
\end{aligned} \tag{46}$$

By (45) and (46), we have

$$\begin{aligned}
&\sqrt{n}\left\{\widehat{V}^{\text{MR}}(\widehat{\beta}) - \widehat{V}^{\text{MR}}(\beta^*)\right\} \\
=&\sqrt{n}\left[\left\{\widehat{V}^{\text{MR}}(\widehat{\beta}) - \widehat{V}^{\text{MR}}(\beta^*)\right\} - \left\{V(\widehat{\beta}) - V(\beta^*)\right\}\right] + \sqrt{n}\left\{V(\widehat{\beta}) - V(\beta^*)\right\} \\
=&o_p(1) + o_p(1) = o_p(1).
\end{aligned}$$

Then,

$$\sqrt{n}\left\{\widehat{V}^{\text{MR}}(\widehat{\beta}) - V(\beta^*)\right\} = \sqrt{n}\left\{\widehat{V}^{\text{MR}}(\widehat{\beta}) - \widehat{V}^{\text{MR}}(\beta^*)\right\} + \sqrt{n}\left\{\widehat{V}^{\text{MR}}(\beta^*) - V(\beta^*)\right\}$$
$$= \sqrt{n}\left\{\widehat{V}^{\text{MR}}(\beta^*) - V(\beta^*)\right\} + o_p(1).$$

Therefore, $\widehat{V}^{\text{MR}}(\widehat{\beta}) - V(\beta^*)$ has the same asymptotic distribution as $\widehat{V}^{\text{MR}}(\beta^*) - V(\beta^*)$. From Section A.3, $\widehat{V}^{\text{MR}}(\beta^*) - V(\beta^*)$ is asymptotically normal with mean zero. Under $\mathcal{M}_{\text{ps+cs+sp+om}}$, $\widehat{V}^{\text{MR}}(\widehat{\beta})$ achieves the semiparametric efficiency bound $\Upsilon(\pi(x;\beta^*))$. This completes the proof. □

C. Sensitivity Analysis

Assumption 3.4 may be violated if there are latent confounders between the principal strata and outcome. For the sensitivity analysis, we assume the following tilting model:

$$\frac{\mathbb{E}[Y(1) \mid U = 10, X]}{\mathbb{E}[Y(1) \mid U = 11, X]} = \rho(X). \tag{47}$$

Then, we can use $\rho(X)$ as the sensitivity parameter. If $\rho(X) = 1$, then Assumptions 3.4 holds. The following theorem establishes the nonparametric identification of $V(\pi)$ when the sensitivity parameter is known. Define

$$\omega(X) = \frac{p_1(X)}{\rho(X)p_1(X) + \{1 - \rho(X)\}p_0(X)}.$$

Theorem C.1. *Let* Π *be a policy class. Under Assumptions 3.1–3.3, 3.5, and (47) with known* $\rho(X)$*, for any given policy* $\pi \in \Pi$*, the survivor value function* $V(\pi)$ *is identified,*

$$V(\pi) = \mathbb{E}\left[\frac{p_0(X)}{p_0}\left[\omega(X)\mu_1(X)\pi(X) + \mu_0(X)\{1 - \pi(X)\}\right]\right]. \tag{48}$$

Proof.

$$\begin{aligned}
\mu_1(X) =& \mathbb{E}[Y \mid A = 1, C = 0, S = 1, X] \\
=& \mathbb{E}[Y \mid A = 1, C = 0, S = 1, U = 11, X]P(U = 11 \mid A = 1, C = 0, S = 1, X) \\
&+ \mathbb{E}[Y \mid A = 1, C = 0, S = 1, U = 10, X]P(U = 10 \mid A = 1, C = 0, S = 1, X) \\
=& \mathbb{E}[Y(1) \mid A = 1, C(1) = 0, S(1) = 1, U = 11, X]P(U = 11 \mid A = 1, C = 0, S = 1, X) \\
&+ \mathbb{E}[Y(1) \mid A = 1, C(1) = 0, S(1) = 1, U = 10, X]P(U = 10 \mid A = 1, C = 0, S = 1, X) \quad \text{(SUTVA)} \\
=& \mathbb{E}[Y(1) \mid C(1) = 0, S(1) = 1, U = 11, X]P(U = 11 \mid A = 1, C = 0, S = 1, X) \\
&+ \mathbb{E}[Y(1) \mid C(1) = 0, S(1) = 1, U = 10, X]P(U = 10 \mid A = 1, C = 0, S = 1, X) \quad \text{(Assumption 3.1)} \\
=& \mathbb{E}[Y(1) \mid S(1) = 1, U = 11, X]P(U = 11 \mid A = 1, C = 0, S = 1, X) \\
&+ \mathbb{E}[Y(1) \mid S(1) = 1, U = 10, X]P(U = 10 \mid A = 1, C = 0, S = 1, X) \quad \text{(Assumption 3.2)} \\
=& \mathbb{E}[Y(1) \mid U = 11, X]P(U = 11 \mid A = 1, C = 0, S = 1, X) \\
&+ \mathbb{E}[Y(1) \mid U = 10, X]P(U = 10 \mid A = 1, C = 0, S = 1, X) \\
=& \mathbb{E}[Y(1) \mid U = 11, X]\frac{p_0(X)}{p_1(X)} + \mathbb{E}[Y(1) \mid U = 10, X]\frac{p_1(X) - p_0(X)}{p_1(X)} \\
=& \mathbb{E}[Y(1) \mid U = 11, X]\frac{p_0(X)}{p_1(X)} + \rho(X)\mathbb{E}[Y(1) \mid U = 11, X]\frac{p_1(X) - p_0(X)}{p_1(X)} \quad \text{(by 47)} \\
=& \frac{\mathbb{E}[Y(1) \mid U = 11, X]}{\omega(X)}.
\end{aligned} \tag{49}$$

$$\begin{aligned}
\mu_0(X) =& \mathbb{E}[Y \mid A = 0, C = 0, S = 1, X] \\
=& \mathbb{E}[Y(0) \mid A = 1, C(0) = 0, S(0) = 1, X] \quad \text{(SUTVA)} \\
=& \mathbb{E}[Y(0) \mid C(0) = 0, S(0) = 1, X] \quad \text{(Assumption 3.1)} \\
=& \mathbb{E}[Y(0) \mid S(0) = 1, X] \quad \text{(Assumption 3.2)} \\
=& \mathbb{E}[Y(0) \mid U = 11, X]. \quad \text{(Assumption 3.3)}
\end{aligned} \tag{50}$$

$$V(\pi) \quad = \quad \mathbb{E}\left[Y(1)\pi(X) + Y(0)\{1 - \pi(X)\} \mid U = 11\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathbb{E}[Y(1)\pi(X) + Y(0)\{1-\pi(X)\} \mid U=11, X] \mid U=11\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y(1) \mid U=11, X]\pi(X) \mid U=11\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}[Y(0) \mid U=11, X]\{1-\pi(X)\} \mid U=11\right] \\
&= \mathbb{E}\left[\omega(X)\mu_1(X)\pi(X) \mid U=11\right] \quad \text{(by (49))} \\
&\quad + \mathbb{E}\left[\mu_0(X)\{1-\pi(X)\} \mid U=11\right] \quad \text{(by (50))} \\
&= \mathbb{E}\left[\frac{p_0(X)}{p_0}\left[\omega(X)\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right].\right] \quad \text{(by (7))}
\end{aligned}
$$

Starting from the identification formula (48), we can derive the EIF for $V(\pi)$ using the same technique in Section A.2. Therefore, we directly give the following result and omit the proof.

Theorem C.2. *Suppose $V(\pi)$ is identified in Theorem C.1. The EIF for $V(\pi)$ is*

$$
\nu'_\pi = \frac{1}{p_0}\{\phi'_\pi - V(\pi)\psi_{S(0)}\},
$$

and the semiparametric efficiency bound for $V(\pi)$ is

$$
\Upsilon'(\pi) = \mathbb{E}\left[\frac{1}{p_0}\{\phi'_\pi - V(\pi)\psi_{S(0)}\}\right]^2,
$$

where $\phi'_\pi = \left[\frac{\omega^2(X)\mu_1(X)}{\rho(X)}\left\{\psi_{S(0)} - \frac{p_0(X)}{p_1(X)}\psi_{S(1)}\right\} + \frac{\omega(X)p_0(X)}{p_1(X)}\psi_{Y(1)S(1)}\right]\pi(X) + \psi_{Y(0)S(0)}\{1-\pi(X)\}$.

The EIF ν'_π motivates the following estimator for $V(\pi)$:

$$
\widehat{V}(\pi) = \frac{\mathbb{P}_n(\widehat{\phi'}_\pi)}{\mathbb{P}_n\{\widehat{\psi}_{S(0)}\}}, \tag{51}
$$

where $\widehat{\phi'}_\pi \left[\frac{\omega^2(X)\widehat{\mu}_1(X)}{\rho(X)}\left\{\widehat{\psi}_{S(0)} - \frac{\widehat{p}_0(X)}{\widehat{p}_1(X)}\widehat{\psi}_{S(1)}\right\} + \frac{\omega(X)\widehat{p}_0(X)}{\widehat{p}_1(X)}\widehat{\psi}_{Y(1)S(1)}\right]\pi(X) + \widehat{\psi}_{Y(0)S(0)}\{1-\pi(X)\}$. This estimator achieves the semiparametric efficiency bound $\Upsilon'(\pi)$ under mild nonparametric rate conditions of nuisance functions estimation.

Theorem C.3. *Suppose that Assumptions 3.1-3.3, 3.5, 4.4, and (47) with known $\rho(X) > 0$ hold. $\widehat{V}(\pi)$ in (51) has the influence function ν'_π and therefore achieves the semiparametric efficiency bound $\Upsilon'(\pi)$.*

The proof of Theorem C.3 is similar to that of Theorem 4.5 and thus omitted.

We re-visit the real data application in Section 6.2 to assess the violation of Assumption 3.4. For the ease of presentation, we assume the sensitivity parameter is not dependent on X, i.e., $\rho(X) = \rho$, and vary them from 0.8 to 1.2. We estimate nuisance functions using the entire dataset and the same models as those in Section 6.2. We construct estimators $\widehat{V}(\beta) = \frac{\mathbb{P}_n\{\widehat{\phi'}_{\pi(X;\beta)}\}}{\mathbb{P}_n\{\widehat{\psi}_{S(0)}\}}$ with different values of ρ. We estimate the survivor-optimal linear policy by maximizing these estimators within the linear policy class $\Pi_\beta = \{\pi(x;\beta) = \mathbb{I}(\beta^T\tilde{x} > 0) : \beta \in \mathbb{R}^8, \|\beta\|_2 = 1\}$. We denote the estimated β as $\widehat{\beta}$. We report the values of $\widehat{V}(\widehat{\beta})$ in Figure 3. We find that the result is not sensitive to the violation of Assumption 3.4.

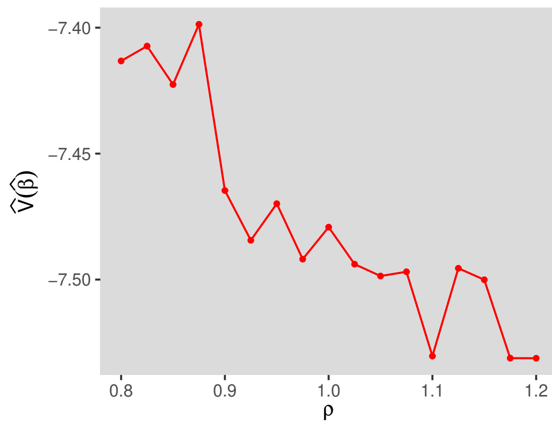

Figure 3. Sensitivity analysis for Section 6.2.

D. Additional OPE Experiment Results in Section 6

We report the OPE experiment results with sample size $n = 1000$ in Table 5. The other setting is the same as Section 6.1.

Table 5. OPE results with $n = 1000$. (a) $0.7\pi_d + 0.3\pi_u$, (b) $0.4\pi_d + 0.6\pi_u$, (c) $0.0\pi_d + 1.0\pi_u$.

censoring rate:15%

	MR-I		MR-II		MR-III		MR-IV		MR-V		DM		IPW		DR	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
(a)	0.159	0.158	0.167	0.164	0.159	0.159	0.161	0.160	0.172	0.172	0.882	0.103	0.691	0.118	0.517	0.129
(b)	0.137	0.137	0.144	0.142	0.138	0.138	0.144	0.143	0.148	0.148	0.652	0.092	0.585	0.087	0.449	0.111
(c)	0.110	0.109	0.115	0.113	0.110	0.110	0.126	0.125	0.123	0.122	0.347	0.078	0.445	0.051	0.359	0.090

censoring rate: 30%

	MR-I		MR-II		MR-III		MR-IV		MR-V		DM		IPW		DR	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
(a)	0.187	0.186	0.200	0.190	0.186	0.186	0.189	0.188	0.201	0.200	0.899	0.119	0.770	0.174	0.545	0.147
(b)	0.162	0.161	0.173	0.164	0.162	0.162	0.169	0.167	0.175	0.175	0.674	0.105	0.654	0.142	0.474	0.126
(c)	0.131	0.130	0.138	0.131	0.130	0.130	0.149	0.148	0.151	0.148	0.375	0.088	0.502	0.105	0.380	0.101