ABSTRACT

CHU, JIANING. Advances in Policy Evaluation and Learning: Targeting, Truncation by Death, and One-sided Feedback. (Under the direction of Wenbin Lu and Shu Yang).

The motivation behind my dissertation research is rooted in real-world applications of individualized decision-making. In precision medicine, treatment plans are tailored to each patient's unique characteristics, aiming for better outcomes. In advertising, personalized content recommendations are used to increase click-through rates. In finance, customized investment strategies can enhance an individual's financial well-being. In many applications, experimentation and exploration can be costly, risky, or even unethical. This motivates the study of off-policy evaluation (OPE) and off-policy learning (OPL). The goal of OPE is to estimate the expected outcome of a given decision policy (also known as a treatment regime) using historical data generated by a potentially different policy. OPL, on the other hand, aims to derive an optimal policy that maximizes the expected outcome based on the available historical data. Although there is extensive literature on OPE and OPL, most existing studies focus on simple settings, such as single datasets with complete data observations, making these methods less applicable to real-world scenarios where datasets often come from diverse sources, exhibit heterogeneity, and involve complex structures. In this thesis, we mainly focus on developing OPE and OPL methods for complex data from real-world applications such as healthcare, finance, and social science. This thesis is structured as follows.

In Chapter 2, we study how to learn the optimal treatment regime for a target population when the source and target populations are heterogeneous. We consider the scenario where individual-level data is available from the source population, but only summary information about the covariate distribution is available from the target population. This situation is common in healthcare, where privacy and confidentiality concerns often prevent sharing comprehensive individual-level data with researchers. However, summary statistics of patient characteristics from the target population are typically available and can be easily shared for research purposes. We propose a general framework to calibrate the source individual-level data to be closer to the target population. This framework provides a more accurate estimator for treatment regime evaluation and yields an optimal treatment regime outperforms standard methods. Additionally, when the source and target populations are identical, our proposed estimator is also more efficient than standard estimators.

In Chapter 3, we study the OPE and OPL problems under the "truncation by death" setting, where the outcome of interest is undefined after certain events, such as death. For example, the quality of life of patients cannot be assessed after death. We formulate OPE and OPL using prin-

cipal stratification under this setting. 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 semiparametric or nonparametric models for nuisance functions approximation. Moreover, under mild rate conditions of nuisance functions approximation, the estimators achieve the semiparametric efficiency bound.

In Chapter 4, we study a class of decision-making problems with one-sided feedback, where outcomes are only observable for specific actions. An example of this is bank loans, where the repayment status is known only if the loan is approved and remains undefined if rejected. Such one-sided feedback is prevalent in various domains, including school admissions, job employment, and criminal recidivism prediction. In such scenarios, conventional OPE and OPL are infeasible. We introduce a novel value function for OPE that addresses the issue of undefined counterfactual outcomes. Without assuming no unmeasured confounders, we establish the identification of the value function using shadow variables. Furthermore, we derive the semiparametric efficiency bound for the proposed value function and develop efficient methods for OPE and OPL.

Advances in Policy Evaluation and Learning: Targeting, Truncation by Death, and One-sided Feedback

by
Jianing Chu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina
2024

APPROVED BY:

Marie Davidian	Daowen Zhang
Wenbin Lu Co-chair of Advisory Committee	Shu Yang Co-chair of Advisory Committee

BIOGRAPHY

Jianing Chu was born in Yuyao, Zhejiang, China, in 1996. After graduating from Zhejiang Yuyao High School, he was admitted to Zhejiang University in 2015 and graduated in 2019 with a Bachelor's degree in Statistics. He then continued his education at North Carolina State University, pursuing a Doctoral degree in Statistics under the supervision of Dr. Wenbin Lu and Dr. Shu Yang.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my advisors, Dr. Wenbin Lu and Dr. Shu Yang, for their exceptional guidance, excellent expertise, and unwavering support throughout my research journey. Their mentorship has been instrumental in the successful completion of this work. Over the past four years, we have worked closely together, meeting almost every week to discuss and refine my dissertation research. Their commitment to my development has not only helped me overcome numerous challenges but has also fostered an environment of intellectual curiosity and critical thinking. I would also like to extend my heartfelt gratitude to Dr. Marie Davidian and Dr. Daowen Zhang for their invaluable feedback and dedicated service on my dissertation committee.

I would like to express my sincere appreciation to the faculty members of the Department of Statistics for offering a diverse and enriching array of courses that have greatly enhanced my academic journey. I am also thankful to the department staff for their outstanding service and timely support throughout my studies.

I would like to acknowledge the invaluable companionship from my friends. They have consistently provided help in my daily life and offered emotional support, making my journey fulfilling and memorable.

My sincere thanks also go to my internship managers and mentors at Amazon. They provided me with the opportunity to apply my research knowledge and skills to real-world industry challenges and offered valuable advice for my career development.

Finally, I want to express my deepest love and appreciation to my family. They have consistently encouraged me to pursue my passions and have stood by every decision I have made. Their unconditional love and unwavering support have been my greatest sources of strength.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

1.1 Background

Personalized decision-making, an artificial intelligence paradigm tailored to individual characteristics, has recently attracted significant attention across various fields, such as healthcare (Qian and Murphy 2011), education (Mandel et al. 2014), e-commerce (Swaminathan et al. 2017), and social science (Kitagawa and Tetenov 2018). A decision policy, also known as a treatment regime, formalizes treatment decisions as a function mapping from individual information to a recommended treatment. The expected outcome for the population under a given policy is referred to as the value function. In these decision-making processes, there are two primary objectives: (1) evaluating a policy based on the value function it achieves, and (2) deriving an optimal policy that maximizes the value function.

In many real-world applications, experimentation and exploration can be costly, risky, or even unethical. This motivates the study of off-policy evaluation (OPE) (Dudík et al. 2014; Chakraborty et al. 2014; Luedtke and Van Der Laan 2016; Wang et al. 2017; Shi et al. 2020) and off-policy learning (OPL) (Murphy 2003; Zhang et al. 2012, 2013; Schulte et al. 2014; Jiang and Li 2016; Shi et al. 2018; Athey and Wager 2021) methods in decision-making. The goal of OPE is to estimate the value function of a given policy using historical data generated by potentially different policies, while OPL aims to derive an optimal policy based on the

available historical data. Although there is extensive literature on OPE and OPL, most existing studies focus on simple settings, such as single datasets with complete data observations, making these methods less applicable to real-world scenarios where datasets often come from diverse sources, exhibit heterogeneity, and involve complex structures. In this thesis, we study various complex scenarios and develop corresponding OPE and OPL methods to address these challenges.

1.1.1 Heterogeneous Source and Target Populations

In many real-world applications, there is significant heterogeneity across different populations. Consequently, the optimal policy derived from one source population may not generalize effectively to another target population. In medical research, for example, it is well recognized that results of a randomized controlled trial cannot be directly transported because the covariate distribution in a target population may be different (Cole and Stuart 2010). Factors such as study design, inclusion, and exclusion criteria can render the source sample unrepresentative of the population of interest. Similarly, in the field of education, an intervention proven effective in one school district may not yield the same results when implemented in a different district with varying socioeconomic conditions and student demographics. Educational research often faces challenges when trying to generalize findings from a study conducted in a specific setting to different educational environments (Cheung and Slavin 2013). This challenge has garnered increasing attention in the fields of OPE and OPL, as researchers strive to develop methods that account for and address such population heterogeneity.

1.1.2 Truncation by Death

In many cases, the goals of decision-making and outcomes of interest may vary across subpopulations. For example, in critical care, the primary goal of treating is to save seriously ill patients' lives (i.e., decrease mortality). With the advanced development in medicine, the short-term mortality has been reduced in many clinical situations (Wunsch et al. 2010; Guérin et al. 2013). However, surviving critical illness frequently comes at a cost, with many survivors experiencing worsened pain, cognition, physical function, and mental health during or after treatment. To account for a patient perspective, the significance of improving "patient-important" outcomes, such as functional outcomes and quality of life in survivors has gained wide acceptance 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 change to the treatment

effect on symptom severity (Paltiel et al. 2021). In the above examples, the outcome of interest in the 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 measured in survivors. In the second case, the symptom severity can only be measured after infection. The events here can be flexibly defined based on different study purposes. However, to be aligned with the causal inference literature, we term such problems as "truncation by death" (Frangakis and Rubin 1999). Truncation by death leads outcomes to be undefined and thus cannot simply be regarded as missing data problems.

1.1.3 One-sided Feedback

Binary decision-making problems are pervasive in the real world, encompassing domains such as bank loan approval, school admission, advertising, and criminal recidivism prediction (Lakkaraju et al. 2017; Baker and Hawn 2022; Fuster et al. 2022; Wang et al. 2023). Often, feedback in these scenarios is one-sided. For example, in bank loan approval, a decision-maker evaluates an applicant based on their covariates and decides whether to approve or deny the loan. If the loan is approved, feedback on the applicant's repayment behavior is later received. However, if the loan is denied, no additional information is gathered. This one-sided feedback introduces significant challenges due to the inherent heterogeneity between the approved and rejected groups caused by selection bias. While outcome information is available for the approved group, it does not fully represent the entire population, as this group is a non-random and potentially biased sample. Consequently, employing the outcome model trained with approved samples for the rejected sample is not feasible.

1.2 Outline

The remainder of this thesis is organized as follows. In Chapter 2, we study how to learn the optimal treatment regime for a target population when the source and target populations are heterogeneous. We consider the scenario where individual-level data is available from the source population, but only summary information about the covariate distribution is available from the target population. This situation is common in healthcare, where privacy and confidentiality concerns often prevent sharing comprehensive individual-level data with researchers. However, summary statistics of patient characteristics from the target population are typically available and can be easily shared for research purposes. We propose a general framework to calibrate the source individual-level data to be closer to the target population. This framework provides a more accurate estimator for treatment regime evaluation and yields

an optimal treatment regime outperforms standard methods. Additionally, when the source and target populations are identical, our proposed estimator is also more efficient than standard estimators.

In Chapter 3, we study the OPE and OPL problems under the "truncation by death" setting, where the outcome of interest is undefined after certain events, such as death. For example, the quality of life of patients cannot be assessed after death. We formulate OPE and OPL using principal stratification under this setting. 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 semiparametric or nonparametric models for nuisance functions approximation. Moreover, under mild rate conditions of nuisance functions approximation, the estimators achieve the semiparametric efficiency bound.

In Chapter 4, we study a class of decision-making problems with one-sided feedback, where outcomes are only observable for specific actions. An example of this is bank loans, where the repayment status is known only if the loan is approved and remains undefined if rejected. Such one-sided feedback is prevalent in various domains, including school admissions, job employment, and criminal recidivism prediction. In such scenarios, conventional OPE and OPL are infeasible. We introduce a novel value function for OPE that addresses the issue of undefined counterfactual outcomes. Without assuming no unmeasured confounders, we establish the identification of the value function using shadow variables. Furthermore, we derive the semiparametric efficiency bound for the proposed value function and develop efficient methods for OPE and OPL.

CHAPTER

2

TARGETED OPTIMAL TREATMENT REGIME LEARNING USING SUMMARY STATISTICS

2.1 Introduction

Personalized decision-making, a pseudo intelligence paradigm tailored to individual characteristics, has recently attracted a great deal of attention in many fields, such as precision medicine, social services, economics, and recommendation system. An individualized treatment rule (ITR) formalizes treatment decisions as a function mapping from patient information to a recommended treatment. An optimal ITR is the one that leads to the greatest expected outcome in the population of interest, known as the value function.

A variety of approaches have been developed for estimating optimal ITRs. One class of approaches is model-based as they directly model the conditional mean outcome given covariates and treatment, known as the Q-function, and then use the estimated Q-function to infer the optimal ITR. Such methods include Q-learning (Qian and Murphy 2011) and its semiparametric extension, A-leaning (Murphy 2003; Shi et al. 2018), where only the contrast function is modeled while the baseline mean function is completely unspecified. Alternatively, direct value search methods have been developed and extensively studied recently (e.g. Zhang et al. 2012, 2013; Luckett et al. 2020; Athey and Wager 2021). These methods learn the optimal treatment

regime by regime evaluation. They first establish a flexible estimator of the value function, such as the augmented inverse probability weighted (AIPW) estimator, and the optimal ITR is then estimated by maximizing the estimated value function within a class of pre-specified ITRs, such as linear decision rules and tree-based decision rules. The AIPW value estimator possesses the double robustness property, i.e., it is consistent for the value function if either the Q-function or the propensity score model is correctly specified.

Though the double robustness of the AIPW value estimator is appealing, it is only maintained when the source and target populations are identical. In other words, when there exists heterogeneity between the source and target populations, the AIPW value estimator obtained based on the source sample may no longer be consistent for the value function of the target population. Thus, the optimal ITR learned from the source data may not be optimal for the target population. In many real-world applications, the value function of an ITR over the distribution of the target population is of significant interest, which can be different from that of the source population. For example, in medical studies, it is known that the results of a randomized controlled trial cannot be directly transported because the covariate distribution in a target population may be different (Cole and Stuart 2010). Due to study design and inclusion/exclusion criteria, the source sample can be unrepresentative of the target population we are interested in. When there is heterogeneity between the source and target populations, an estimated optimal ITR from the source sample may not generalize well to the target population (Lee et al. 2023). Such problems gain increasing attention in the ITR learning fields recently. Zhao et al. (2019) and Mo et al. (2021) proposed different collections of possible target covariate distributions and estimated the optimal ITR by optimizing the worst-case quality assessment among the collection. Uehara et al. (2020) considered a nonparametric estimator for the density ratio of the covariate distributions of the source and target populations and constructed a weighted estimator for the target value function based on the estimated density ratio. However, all these methods require the availability of individual-level data from both the source and target populations, which may be unrealistic in many applications. For example, while large-population based databases, such as the Surveillance, Epidemiology and End Results database, can provide reliable summary statistics for covariates, such as means and medians, and overall survival statistics for the disease population, critical information about individual factors that influence the choice of treatment and clinical outcomes of interest may not be available (Huang and Qin 2020; Chen et al. 2021). Moreover, due to privacy and confidentiality concerns, comprehensive individual-level data is often prohibited to share with researchers. In contrast, summary statistics of patient characteristics of the target population are often available and can be easily shared for research purposes.

In this paper, we consider the targeted optimal treatment regime learning where we have

individual-level data from the source sample but only a few summary statistics of covariate distributions from the target population. As we alluded to previously, when there is heterogeneity in covariate distributions between the source and target populations, the estimated optimal ITR obtained by maximizing the value estimator constructed based on the source sample may not be optimal for the target population. One way to address this issue is to assign different subject weights to the source sample and calibrate the source covariate distributions to the target covariate distributions. Calibration weighting is widely used to integrate auxiliary information in survey sampling and causal inference, such as empirical likelihood based methods (Qin and Zhang 2007), entropy-based covariate balancing methods (Hainmueller 2012), and quadratic loss based covariate balancing methods (Zubizarreta 2015). Such weighting methods allow adjusting covariate distributions of the source sample using various summary statistics of covariates in the target population, such as means, variances, correlations, and quantiles. We propose a calibrated AIPW estimator of the value function using summary statistics from the target population and then search for the optimal ITR for the target population by maximizing the calibrated AIPW value estimator over a pre-specified class of ITRs. Here, the subject weights for the source sample are estimated by solving a general convex optimization problem with constraints. The objective function in the optimization problem can be chosen from the Cressie-Read family (Cressie and Read 1984), while the constraints force the weighted summary statistics of source covariates to be the same as that from the target population. We show that the calibrated AIPW estimator for the target value function is consistent, asymptotically normal, and has the double robustness property if the estimated weight function converges to the density ratio of covariate distributions between the two populations. The double robustness entails that the value estimator remains root-n consistent if any one of the two parametric models for the propensity score and outcome mean is correctly specified or if both models are estimated nonparametrically satisfying a certain rate condition for convergence. Interestingly, if the source and target populations have the same covariate distribution, the calibrated optimal value estimator gains efficiency over the uncalibrated one by utilizing additional summary information. However, in general, the weights learned from the calibration methods may not consistently estimate the density ratio. Under such general cases, the proposed calibrated AIPW estimator can still converge to the value function of a pseudo population that may be closer to the target distribution compared with the source population. As such, it can give a more accurate estimator for the value function of the target population than the uncalibrated value estimator, and the optimal ITR obtained by maximizing the calibrated AIPW value estimator can be more favorable for the target population.

2.2 Statistical Framework

2.2.1 Value Function and Optimal ITR

In a randomized trial or observational study, suppose there are two treatment options, labeled as control/treatment 0 and experimental treatment/treatment 1. Let A taking values 0 or 1 in accordance with the two options, denote the treatment received. Let $X \in \mathbb{R}^p$ be a vector of baseline covariates and Y be the observed outcome of interest. We assume larger values of Y are preferred by convention. The observed data are then $\{O_i = (Y_i, A_i, X_i), i = 1, \ldots, n\}$, which are independent and identically distributed. Define the potential outcomes $Y(0)$ and $Y(1)$ as the outcomes that would be observed if a subject received treatment 0 or 1, respectively. As is customary in causal inference (Rubin 1978), we make the following assumptions.

Assumption 1 *(A1)* $Y = Y(1)A + Y(0)(1-A)$, *(A2)* $\{Y(1), Y(0)\} \perp\!\!\!\perp A \mid X$, *and (A3)* $0 < \Pr(A = 1 \mid X = x) < 1$ *for all* x *such that* $\Pr(X = x) > 0$.

An ITR is a function $d(\cdot)$ that maps values of X to $\{0, 1\}$, so that a subject with covariates value $X = x$ would receive treatment 1 if $d(x) = 1$ and treatment 0 if $d(x) = 0$. For any arbitrary ITR $d(\cdot)$, we can define the potential outcome as $Y(d) = Y(1)d(X) + Y(0)\{1-d(X)\}$, which would be observed if a randomly chosen individual had been assigned a treatment according to $d(\cdot)$, where we suppress the dependence of $Y(d)$ on X. We then define the value function under $d(\cdot)$ as the expectation of the potential outcome as $V(d) = \mathbb{E}\{Y(d)\} = \mathbb{E}[Y(1)d(X) + Y(0)\{1-d(X)\}]$.

Suppose $\mathscr{D}$ is a class of ITRs of interest. Then we define the optimal ITR as $d^{\text{opt}}(X) = \text{argmax}_{d \in \mathscr{D}} V(d)$. In clinical practice, it may be desirable to consider a class of ITRs indexed by a vector of parameters β for feasibility and interpretability. We denote such a class of rules as $\mathscr{D}_\beta$ and its element as $d(X;\beta)$. For example, we can consider a class of linear ITRs $\{d(X;\beta) = \mathbb{I}(\beta^{\mathrm{T}}\tilde{X} > 0) : \beta \in \mathbb{R}^{p+1}, \|\beta\|_2 = 1\}$, where $\tilde{X} = (1, X^{\mathrm{T}})^{\mathrm{T}}$, $\|\beta\|_2 = (\beta^{\mathrm{T}}\beta)^{1/2}$. Given a linear ITR $d(X;\beta)$, we use a shorthand to write its value function $V(d)$ as $V(\beta)$. Let $\beta^* = \text{argmax}_\beta V(\beta)$. Then, the optimal linear ITR is $d_\beta^{\text{opt}} = d(X;\beta^*)$. The true optimal ITR d^{opt} may not be in $\mathscr{D}_\beta$. Thus, d_β^{opt} may not be the same as d^{opt}. However, when attention focuses on the feasible class $\mathscr{D}_\beta$, estimation of d_β^{opt} is still of considerable interest. In this paper, we focus on linear ITRs.

2.2.2 Source and Target Populations

The difference between covariate distributions in the source and target populations is called a covariate shift (Sugiyama and Kawanabe 2012). In this paper, we assume that there is a pooled population $\mathbb{P}$ consisting of both the source population $\mathbb{P}^{\text{s}}$ and the target population $\mathbb{P}^{\text{t}}$. Let S be a binary indicator for selection action: $S = 1$ if the individual comes from the source population

and $S=0$ if the individual comes from the target population. A covariate shift results from the situation where $\Pr(S=1 \mid X) \neq \Pr(S=0 \mid X)$.

For the source population $\mathbb{P}^{\mathrm{s}}$, we observe individual-level data $\{O_i=(Y_i,A_i,X_i), i=1,\ldots,n\}$. For the target population $\mathbb{P}^{\mathrm{t}}$, only summary statistics of covariate distributions, such as mean, variance or quantiles are available. For $\mathbb{P}^{\mathrm{s}}$, denote the density or probability mass function of covariates as $f^{\mathrm{s}}(X)$ and its associated expectation as $\mathbb{E}$; for $\mathbb{P}^{\mathrm{t}}$, we use the notation $f^{\mathrm{t}}(X)$ and $\mathbb{E}^{\mathrm{t}}$ correspondingly. The summary statistics from the target population $\mathbb{P}^{\mathrm{t}}$ are denoted as $\mu_{g0}=\mathbb{E}^{\mathrm{t}}\{g(X)\}$, where $g(X)=\left\{g_1(X),g_2(X),\ldots,g_q(X)\right\}^{\mathrm{T}}$ is a $q\times 1$ specified function. For example, a common choice is $g(X)=(X_1,X_2,\ldots,X_p)$, and μ_{g0} gives the mean of all covariates in the target population. We assume that summary statistics from the target population are derived from large databases so that their uncertainty are negligible. With only the summary statistics, targeted ITR learning is impossible without further assumptions in order to borrow information from the source sample.

Assumption 2 *(A4)* $\mathbb{E}\{Y(a) \mid X\}=\mathbb{E}^{\mathrm{t}}\{Y(a) \mid X\}$, *and (A5)* $\Pr(S=1 \mid X)>0$.

Assumption (A4) implies that true Q-functions are identical in both the source and target populations (Dahabreh et al. 2019). A stronger version of this assumption is the ignorability assumption that $\{Y(1),Y(0)\} \perp\!\!\!\perp S \mid X$ (Buchanan et al. 2018). Assumption (A5) implies that the support of the target X distribution must be covered by the support of the source X distribution.

2.3 Proposed Method

2.3.1 Calibrated AIPW Estimator

For the source population, define the propensity score as $\pi(X)=\Pr(A=1 \mid X,S=1)$ and conditional mean outcome model as $\mu(X,A)=\mathbb{E}(Y \mid X,A)$. In practice, $\pi(\cdot)$ and $\mu(\cdot)$ can be estimated from the observed source data based on some posited parametric models $\pi(X;\eta)$ and $\mu(X,A;\theta)$, respectively. Alternatively, they can also be estimated nonparametrically, e.g. using kernel regression or random forest. Given an ITR $d(X;\beta)$, Zhang et al. (2012) proposed an AIPW estimator of the value function $V(\beta)$ as

$$\widehat{V}^{\mathrm{o}}(\beta)=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathbb{I}\{A_i=d(X_i;\beta)\}}{\varrho(A_i \mid X_i;\widehat{\eta})}\{Y_i-\mu_d(X_i;\beta,\widehat{\theta})\}+\mu_d(X_i;\beta,\widehat{\theta})\right],$$

where the superscript o is a shorthand for original, $\varrho(A_i \mid X_i;\widehat{\eta})=\pi(X_i;\widehat{\eta})A_i+\{1-\pi(X_i;\widehat{\eta})\}(1-A_i)$, $\mu_d(X_i;\beta,\widehat{\theta})=\mu(X_i,1;\widehat{\theta})\mathbb{I}\{d(X_i;\beta)=1\}+\mu(X_i,0;\widehat{\theta})\mathbb{I}\{d(X_i;\beta)=0\}$, and $\widehat{\eta}$ and $\widehat{\theta}$ are the estimates of η and θ, respectively, based on some posited parametric models.

Denote the value function of the target population under the ITR $d(X;\beta)$ as $V^{\mathrm{t}}(\beta)$. If the source and target populations have the same covariate distributions, i.e. $f^{\mathrm{s}}(X)=f^{\mathrm{t}}(X)$, then $V^{\mathrm{t}}(\beta)$ can be consistently estimated by $\widehat{V}^{\mathrm{o}}(\beta)$ based on the source sample. However, since covariate distributions between the source and target populations often differ in practice, $\widehat{V}^{\mathrm{o}}(\beta)$ may be biased for $V^{\mathrm{t}}(\beta)$. To reduce the bias, a natural approach is to consider calibration weights, i.e., to assign different weights to individual data points in the source sample so that the weighted data is more representative of the target distribution. Specifically, we consider the following calibrated AIPW estimator

$$\widehat{V}^{\mathrm{c}}(\beta)=\sum_{i=1}^{n} w_i\left[\frac{\mathbb{I}\{A_i=d(X_i;\beta)\}}{\varrho(A_i\mid X_i;\widehat{\eta})}\{Y_i-\mu_d(X_i;\beta,\widehat{\theta})\}+\mu_d(X_i;\beta,\widehat{\theta})\right],$$

where the superscript c is a shorthand for calibrated, w_i's are calibration weights satisfying $\sum_{i=1}^{n} w_i=1$ and other constraints. Using the summary statistics μ_{g0} from the target population, we can utilize methods, such as empirical likelihood (Qin and Zhang 2007) and entropy balancing (Hainmueller 2012; Zhao and Percival 2017), to learn the weights. In the next section, we propose a general framework to estimate the weights.

2.3.2 Weights Estimation

Let $h(w)$ denote a generic convex distance function between a scalar w and n^{-1}. We consider the following optimization problem

$$\min_{w_1,\ldots,w_n}\sum_{i=1}^{n} h(w_i),$$

under the constraints $\sum_{i=1}^{n} w_i\{g(X_i)-\mu_{g0}\}=0$, $\sum_{i=1}^{n} w_i=1$, and $w_i\geq 0$.

In the considered optimization problem, the function $h(w)$ plays a role in quantifying the discrepancy of calibration weights and the uniform distribution n^{-1}. We choose the function $h(w)$ from the Cressie-Read family of discrepancies (Cressie and Read 1984). The Cressie-Read family is defined through a class of additive convex functions that encompasses a broad family of distance functions. Specifically,

$$CR(\gamma)=\sum_{i=1}^{n} h(w_i)=\sum_{i=1}^{n}\{\gamma(\gamma+1)\}^{-1}\{(nw_i)^{\gamma+1}-1\}.$$

Three special cases with $\gamma\in\{-1,0,1\}$ are popular. In particular, $CR(-1)=\sum_{i=1}^{n}-\log(nw_i)$ and $CR(0)=\sum_{i=1}^{n}(nw_i)\log(nw_i)$. Minimizing $CR(-1)$ is equivalent to maximizing $\sum_{i=1}^{n}\log(w_i)$,

leading to the maximum empirical log-likelihood objective function. Minimizing $CR(0)$ is equivalent to maximizing $-\sum_{i=1}^{n} w_i \log(w_i)$, leading to the maximum empirical exponential likelihood or entropy. Finally, minimizing $CR(1)$ is equivalent to minimizing the sum of squares $\sum_{i=1}^{n}(w_i - n^{-1})^2$. To be consistent with the existing literature, we call the weight estimation method as the empirical likelihood method for $\gamma = -1$ and the entropy balancing method for $\gamma = 0$. We summarize the correspondence between γ and the form of $h(w)$ in Table 2.1.

Table 2.1: The formulation of $\rho(x)$ for the empirical likelihood, entropy balancing, and Cressie-Read family.

Method	Empirical Likelihood	Entropy Balancing	Cressie-Read
γ	-1	0	γ
$h(w)$	$-\ln(nw)$	$nw\ln(nw)$	$\frac{(nw)^{\gamma+1}-1}{\gamma(\gamma+1)}$
$\rho(x)$	$(1-x)^{-1}$	$\exp(x)$	$(1+\gamma x)^{1/\gamma}$

The first constraint is referred to as the balancing constraint, which calibrates the covariate distribution of the source sample to the target population in terms of $g(X)$. As a common premise to solve the above optimization problem, μ_{g0} should fall within the convex hull of $\{g(X_i), i = 1, ..., n\}$. Then, the optimization problem can be solved using the method of Lagrangian multipliers with the loss function

$$L = \{\gamma(\gamma+1)\}^{-1} \sum_{i=1}^{n} \{(nw_i)^{\gamma+1} - 1\} - n\lambda^{\mathrm{T}} \sum_{i=1}^{n} w_i \{g(X_i) - \mu_{g0}\} + n\varphi\left(1 - \sum_{i=1}^{n} w_i\right). \tag{2.1}$$

As noted in Newey and Smith (2004), by minimizing (2.1), the estimator for w_i is

$$w(X_i; \widehat{\lambda}) = \rho\left[\widehat{\lambda}^{\mathrm{T}}\{g(X_i) - \mu_{g0}\}\right] \Big/ \sum_{j=1}^{n} \rho\left[\widehat{\lambda}^{\mathrm{T}}\{g(X_j) - \mu_{g0}\}\right], \tag{2.2}$$

where the function $\rho(x)$ for different γ values are summarized in Table 2.1, and $\widehat{\lambda}$ solves the equation $\sum_{i=1}^{n} \rho\left[\lambda^{\mathrm{T}}\{g(X_i) - \mu_{g0}\}\right]\{g(X_i) - \mu_{g0}\} = 0$.

Let $W(X; \lambda) = n\, w(X; \lambda)$. The proposed calibrated AIPW estimator is then

$$\widehat{V}^{\mathrm{c}}(\beta) = \frac{1}{n} \sum_{i=1}^{n} W(X_i; \widehat{\lambda}) \left[\frac{\mathbb{I}\{A_i = d(X_i; \beta)\}}{\varrho(A_i \mid X_i; \widehat{\eta})} \{Y_i - \mu_d(X_i; \beta, \widehat{\theta})\} + \mu_d(X_i; \beta, \widehat{\theta}) \right].$$

Thus, the regime learning procedure can be summarized as a 3-step algorithm:
Step 1: Estimate calibration weights, e.g., using the empirical likelihood method or entropy

balancing method.

Step 2: Estimate the propensity score $\pi(\cdot)$ and the conditional mean outcome $\mu(\cdot)$ using either parametric models or nonparametric models.

Step 3: Construct the calibrated AIPW estimator with the components estimated in Steps 1 and 2, and obtain the optimal ITR by maximizing the calibrated AIPW estimator within a class of pre-specified ITRs, such as linear decision rules.

Before delving into theoretical analysis, it is important to define the underlying population for which $\widehat{V}^{\mathrm{c}}(\beta)$ is targeting unambiguously. Toward this end, let λ^* be the limit of $\widehat{\lambda}$ and

$$W^*(X;\lambda)=\rho\left[\lambda^{\mathrm{T}}\{g(X)-\mu_{g0}\}\right]/\mathbb{E}\left(\rho\left[\lambda^{\mathrm{T}}\{g(X)-\mu_{g0}\}\right]\right).$$

In general, the calibration weights are not guaranteed to be non-negative. As pointed out in Schennach (2007), when $\gamma\leq 0$, the estimated weights are non-negative by construction. It can be shown that $f^+(X)\propto f^{\mathrm{s}}(X)W^*(X;\lambda^*)$ is a valid density or probability mass function when $\gamma=-1$ or 0. Therefore, it defines a pseudo population $\mathbb{P}^+$. While for $\gamma>0$, the calibration weights can take on negative values, and thus the corresponding $f^+(X)$ is not always a valid density or probability mass function. Therefore, we focus on $\gamma=-1,0$ for illustration. It is expected that $\widehat{V}^{\mathrm{c}}(\beta)$ will converge to the value function under the ITR $d(X;\beta)$ for the pseudo population $\mathbb{P}^+$, when either the propensity score or the conditional mean outcome model is correctly specified.

Moreover, when $W(X;\lambda^*)\propto f^{\mathrm{t}}(X)/f^{\mathrm{s}}(X)$, we have $f^+(X)=f^{\mathrm{t}}(X)$. Then, $\widehat{V}^{\mathrm{c}}(\beta)$ is also a consistent estimator of the value function for the target population. Denote the density or probability mass function of covariates in the pooled population $\mathbb{P}$ as $q(X)$. Then $f^{\mathrm{s}}(X)$ and $f^{\mathrm{t}}(X)$ can be described as $q(X\mid S=1)$ and $q(X\mid S=0)$, respectively. By Bayesian Theorem, we have,

$$\frac{f^{\mathrm{t}}(X)}{f^{\mathrm{s}}(X)}=\frac{q(X\mid S=0)}{q(X\mid S=1)}\propto\frac{\Pr(S=0\mid X)}{\Pr(S=1\mid X)}.$$

If $\Pr(S=0\mid X)$ follows a logistic regression with covariates $g(X)$, we have $\Pr(S=0\mid X)/\Pr(S=1\mid X)\propto\exp\{\alpha^{\mathrm{T}}g(X)\}$. Moreover, based on (2.2), we have $W(X;\widehat{\lambda})\propto\exp\{\widehat{\lambda}^{\mathrm{T}}g(X)\}$ when $\gamma=0$. Therefore, the weights obtained by the entropy balancing method satisfy $W(X;\lambda^*)\propto f^{\mathrm{t}}(X)/f^{\mathrm{s}}(X)$ under the logistic regression model for $\Pr(S=0\mid X)$. Similarly, if $\Pr(S=0\mid X)$ can be represented by the following model

$$\Pr(S=0\mid X)=\frac{\kappa_0}{1-\alpha^{\mathrm{T}}\{g(X)-\mu_{g0}\}}\Bigg/\left[1+\frac{\kappa_0}{1-\alpha^{\mathrm{T}}\{g(X)-\mu_{g0}\}}\right],$$

where κ_0 is a positive constant and α satisfies $1-\alpha^{\mathrm{T}}\{g(X)-\mu_{g0}\}>0$, we have $\Pr(S=0\mid X)/\Pr(S=1\mid X)\propto[1-\alpha^{\mathrm{T}}\{g(X)-\mu_{g0}\}]^{-1}$. Therefore, under the above model, the weights obtained by the empirical likelihood method satisfy $W(X;\lambda^*)\propto f^{\mathrm{t}}(X)/f^{\mathrm{s}}(X)$.

In general, the calibration weights can not lead to a pseudo population with exactly the same covariate distribution as for the target population. However, it is expected that with more constraints based on summary statistics from the target population, the covariate distribution of the pseudo population will get closer to that of the target population. Therefore, the optimal ITR obtained based on $\widehat{V}^{\mathrm{c}}(\beta)$ would be better than that obtained based on $\widehat{V}^{\mathrm{o}}(\beta)$. Let $V^+(\beta)$ denote the value function under the ITR $d(X;\beta)$ for the pseudo population $\mathbb{P}^+$ and define $\beta^*=\mathrm{argmax}_\beta\, V^+(\beta)$. Then, the true optimal linear ITR for $\mathbb{P}^+$ is $d(X;\beta^*)$ and the estimated optimal linear ITR is $d(X;\widehat{\beta}^{\mathrm{c}})$, where $\widehat{\beta}^{\mathrm{c}}=\mathrm{argmax}_\beta\,\widehat{V}^{\mathrm{c}}(\beta)$. Similarly, define $\widehat{\beta}^{\mathrm{o}}=\mathrm{argmax}_\beta\,\widehat{V}^{\mathrm{o}}(\beta)$. The estimated optimal linear ITR without calibration is $d(X;\widehat{\beta}^{\mathrm{o}})$.

2.4 Theoretical Properties

In this section, we establish the asymptotic properties of the calibrated AIPW estimator $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$. The proofs of all theorems are given in Appendix A.1.

We first consider the case when the propensity score model $\pi(x)$ and conditional mean outcome model $\mu(x,a)$ are estimated based on some posited parametric models $\pi(X;\eta)$ and $\mu(X,A;\theta)$, respectively. Denote the estimating equations for λ, θ, η and $V^+(\beta^*)$ as

$$\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\rho\left[\lambda^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\right]\{g(X_i)-\mu_{g0}\}\\ C(X_i,A_i,Y_i;\theta)\\ S(X_i,A_i;\eta)\\ W(X_i;\lambda)\psi(X_i,A_i,Y_i;\beta,\theta,\eta)-V^+(\beta^*)\end{pmatrix}=0,$$

where

$$\psi(X_i,A_i,Y_i;\beta,\theta,\eta)=\frac{\mathbb{I}\{A_i=d(X_i;\beta)\}}{\varrho(A_i\mid X_i;\eta)}\{Y_i-\mu_d(X_i;\beta,\theta)\}+\mu_d(X_i;\beta,\theta).$$

Let $\widehat{\lambda}$, $\widehat{\theta}$ and $\widehat{\eta}$ denote the estimators of λ, η and θ obtained from the above equations and let λ^*, θ^* and η^* denote the limits of $\widehat{\lambda}$, $\widehat{\theta}$ and $\widehat{\eta}$, respectively. To establish the asymptotic properties of $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$, we impose the following regularity conditions.

Assumption 3 *Assume the following regularity conditions hold: (A6) The supports of X and Y are bounded. (A7) The function $\mu(x,a)$ is smooth and bounded for all (x,a). (A8) The weight function $W(x;\lambda)$ is smooth and bounded away from ∞, and it has bounded first derivatives with respect to λ. (A9) The value function $V^+(\beta)$ is twice continuously differentiable in a neighborhood*

of β^*. (A10) There exist some constants $\delta_0 > 0$ such that $\Pr(|\tilde{X}^{\mathrm{T}}\beta^*| \le \delta) = O(\delta)$, where the big-O term is uniform in $0 < \delta \le \delta_0$. (A11) (i) $\sqrt{n}(\widehat{\lambda} - \lambda^*) = O_p(1)$, (ii) $\sqrt{n}(\widehat{\theta} - \theta^*) = O_p(1)$, and (iii) $\sqrt{n}(\widehat{\eta} - \eta^*) = O_p(1)$.

Conditions (A6) - (A9) are standard regularity conditions used to establish the uniform convergence results. Condition (A10) excludes the situation with $\Pr(\tilde{X}^{\mathrm{T}}\beta^* = 0) > 0$ and ensures the true targeted optimal ITR 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 ITR (e.g. Luedtke and Van Der Laan 2016). Condition (A11) assumes the $\sqrt{n}$-convergence rates of parameter estimates in the calibration weight function, propensity score model, and conditional mean outcome model, which usually hold under mild conditions for posited parametric models, for example, a logistic or probit regression model for the propensity score, a linear model for the conditional mean outcome, and weights obtained using the empirical likelihood method or entropy balancing method.

Define

$$\xi_{i1} = W(X_i;\lambda^*)\psi(Y_i, A_i, X_i; \beta^*, \theta^*, \eta^*) - V^+(\beta^*), \qquad \xi_{i3} = H_\theta^{\mathrm{T}} G_\theta^{-1} C(X_i, A_i, Y_i; \theta^*),$$
$$\xi_{i2} = H_\lambda^{\mathrm{T}} G_\lambda^{-1} \rho\left[(\lambda^*)^{\mathrm{T}}\{g(X_i) - \mu_{g0}\}\right]\{g(X_i) - \mu_{g0}\}, \qquad \xi_{i4} = H_\eta^{\mathrm{T}} G_\eta^{-1} S(X_i, A_i; \eta^*),$$

where

$$H_\lambda = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial W(X_i;\lambda^*)}{\partial \lambda}\right\}\psi(Y_i, A_i, X_i; \beta^*, \theta^*, \eta^*),$$
$$H_s = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} W(X_i;\lambda^*)\frac{\partial \psi(Y_i, A_i, X_i; \beta^*, \theta^*, \eta^*)}{\partial s} \ (s = \theta, \eta),$$
$$G_\lambda = -\mathbb{E}\left(\rho'\left[(\lambda^*)^{\mathrm{T}}\{g(X) - \mu_{g0}\}\right]\{g(X) - \mu_{g0}\}\{g(X) - \mu_{g0}\}^{\mathrm{T}}\right),$$
$$G_\theta = -\mathbb{E}\{\partial C(X, A, Y; \theta^*)/\partial \theta^{\mathrm{T}}\}, G_\eta = -\mathbb{E}\left\{\partial S(X, A; \eta^*)/\partial \eta^{\mathrm{T}}\right\}.$$

Note that ξ_{i2}, ξ_{i3} and ξ_{i4} are the terms in the inference function of $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$ due to estimators $\widehat{\lambda}$, $\widehat{\theta}$ and $\widehat{\eta}$, respectively.

Theorem 1 *Assume either $\pi(X;\eta)$ or $\mu(X, A;\theta)$ is correctly specified. Under (A1)-(A11), we have, as $n \to \infty$, $\sqrt{n}\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}}) - V^+(\beta^*)\} \longrightarrow N(0, \sigma_1^2)$ in distribution, where $\sigma_1^2 = \mathbb{E}\left\{(\xi_{i1} + \xi_{i2} + \xi_{i3} + + \xi_{i4})^2\right\}$. In addition, σ_1^2 can be estimated by replacing expectation with empirical sum and true values $V^+(\beta^*)$, λ^*, θ^*, and η^* with $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$, $\widehat{\lambda}$, $\widehat{\theta}$, and $\widehat{\eta}$, respectively.*

Next, we consider the case when both propensity score model $\pi(x)$ and conditional mean outcome model $\mu(x, a)$ are estimated by flexible semi/nonparametric models with certain

convergence rates. For example, $\pi(x)$ and/or $\mu(x,a)$ are estimated using kernel regression or random forest. Let $\widehat{\pi}(x)$ and $\widehat{\mu}(x,a)$ denote the corresponding estimators. The calibrated AIPW estimator $\widehat{V}^{\mathrm{c}}(\beta)$ can be similarly defined by replacing $\pi(x;\widehat{\eta})$ and $\mu(x,a;\widehat{\theta})$ with $\widehat{\pi}(x)$ and $\widehat{\mu}(x,a)$, respectively. To derive the asymptotic distribution of $\widehat{V}^{\mathrm{c}}(\beta)$, we need the following modified condition. For a function $f(Z)$, where Z is a generic random variable, we define its L_2-norm as $\|f(Z)\| = \left\{\int f(z)^2 dF_Z(z)\right\}^{1/2}$.

(A11') (i) $\sqrt{n}(\widehat{\lambda}-\lambda^*) = O_p(1)$; (ii) $\|\widehat{\pi}(X)-\pi(X)\|\sum_{a=0}^{1}\|\widehat{\mu}(X,a)-\mu(X,a)\| = o_p(n^{-1/2})$.

Condition (A11') (ii) is commonly imposed in the causal inference literature to derive the asymptotic distribution of the AIPW estimators when the nuisance functions are estimated with certain convergence rates (Kennedy 2016; Farrell et al. 2021). For example, if $\pi(x)$ is estimated based on a correctly specified parametric model, $\widehat{\pi}(x)$ is $\sqrt{n}$-consistent. Then it only requires $\widehat{\mu}(x,a)$ to be consistent for (A11') to hold. This can be easily achieved by most nonparametric regression methods. However, when both $\mu(x,a)$ and $\pi(x)$ are estimated nonparametrically, it usually requires both terms to be estimated with the rate $o_p(n^{-1/4})$. This can be achieved by some nonparametric methods, such as kernel regression or random forest under certain conditions. With Condition (A11') (ii), we can establish the $\sqrt{n}$-consistency of $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$. In addition, the asymptotic variance of $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$ will not depend on the variances of estimates $\widehat{\pi}(x)$ and $\widehat{\mu}(x,a)$. The results are summarized in the following theorem.

Theorem 2 *Under (A1)-(A10) and (A11'), we have, as $n\to\infty$, $\sqrt{n}\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-V^{+}(\beta^*)\}\longrightarrow N(0,\sigma_2^2)$ in distribution, where $\sigma_2^2 = \mathbb{E}\left\{(\xi_{i1}+\xi_{i2})^2\right\}$. Here, ξ_{i1} and ξ_{i2} are defined the same as in Theorem 1 but replacing $\pi(x;\eta)$ and $\mu(x,a;\theta)$ with $\pi(x)$ and $\mu(x,a)$, respectively. In addition, σ_2^2 can be estimated by replacing expectation with empirical sum and true values $V^{+}(\beta^*)$, λ^*, $\pi(x)$ and $\mu(x,a)$ with $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$, $\widehat{\lambda}$, $\widehat{\pi}(x)$ and $\widehat{\mu}(x,a)$, respectively.*

Remark 1 *The theorems established above focus on the inference for the optimal value function. In the proof of Theorems 1 and 2, we show that $\widehat{\beta}^{\mathrm{c}}$ has the cubic root convergence rate. In addition, the asymptotic distribution of $\widehat{\beta}^{\mathrm{c}}$ can be established and its associated inference can be done by bootstrap-based methods (e.g. Cattaneo et al. 2020).*

Finally, we compare the efficiency of $\widehat{V}^{\mathrm{o}}(\beta)$ and $\widehat{V}^{\mathrm{c}}(\beta)$ when the source and target populations have the same covariate distributions, i.e. $\mathbb{P}^{\mathrm{s}} = \mathbb{P}^{\mathrm{t}}$. Under such case, $V^{+}(\beta) = V^{\mathrm{t}}(\beta)$, the value function under the ITR $d(X;\beta)$ for the target population.

Theorem 3 *Assume (A1)-(A10) and (A11') hold. When $\mathbb{P}^{\mathrm{s}} = \mathbb{P}^{\mathrm{t}}$, we have that both $\sqrt{n}\{\widehat{V}^{\mathrm{o}}(\beta)-V^{\mathrm{t}}(\beta)\}$ and $\sqrt{n}\{\widehat{V}^{\mathrm{c}}(\beta)-V^{\mathrm{t}}(\beta)\}$ are asymptotically normal with mean zero, while the latter one has the same or smaller asymptotic variance.*

Theorem 3 implies that even when the source and target populations have the same covariate distributions, the calibrated AIPW value estimator can be more efficient than the original AIPW value estimator without calibration. The efficiency gain of the calibrated estimator comes from the constraints imposed based on available summary statistics of the covariate distribution for the target population.

2.5 Simulation Studies

We have carried out extensive simulation studies to evaluate the performance of the proposed methods. Here we focus on two methods for computing the weights: empirical likelihood ($\gamma = -1$) and entropy balancing ($\gamma = 0$). The results for $\gamma = 1$ are provided in Appendix A.2. For illustration, we only considered means of all covariates as the summary statistics from the target population. The corresponding pseudo populations are denoted as $\mathbb{P}^+_{EB}$ for $\gamma = 0$ and $\mathbb{P}^+_{EL}$ for $\gamma = -1$, respectively. Table 2.2 defines additional notation for the simulation. Since estimated value functions are non-smooth and non-convex in β, following Zhang et al. (2012), we used the genetic algorithm (Whitley 1994) to find $\widehat{\beta}^{\mathrm{o}}$, $\widehat{\beta}^{\mathrm{c}}_{EB}$, and $\widehat{\beta}^{\mathrm{c}}_{EL}$. The optimization was implemented using the function `genoud` in the R package `rgenoud` (Mebane Jr and Sekhon 2011).

Table 2.2: Additional notation used in the simulation studies.

Population	Value	Optimal ITR	Estimators
$\mathbb{P}^{\mathrm{t}}$	$V^{\mathrm{t}}(\beta)$	$d(X;\beta^{\mathrm{t}})$; $\beta^{\mathrm{t}} = \mathrm{argmax}_\beta\, V^{\mathrm{t}}(\beta)$	
$\mathbb{P}^+_{EB}$	$V^+_{EB}(\beta)$	$d(X;\beta^*_{EB})$; $\beta^*_{EB} = \mathrm{argmax}_\beta\, V^+_{EB}(\beta)$	$\widehat{V}^{\mathrm{c}}_{EB}(\beta)$; $\widehat{\beta}^{\mathrm{c}}_{EB} = \mathrm{argmax}_\beta\, \widehat{V}^{\mathrm{c}}_{EB}(\beta)$
$\mathbb{P}^+_{EL}$	$V^+_{EL}(\beta)$	$d(X;\beta^*_{EL})$; $\beta^*_{EL} = \mathrm{argmax}_\beta\, V^+_{EL}(\beta)$	$\widehat{V}^{\mathrm{c}}_{EL}(\beta)$; $\widehat{\beta}^{\mathrm{c}}_{EL} = \mathrm{argmax}_\beta\, \widehat{V}^{\mathrm{c}}_{EL}(\beta)$

For the source sample, outcomes are generated from the model $Y = \mu(X,A) + \epsilon$, where $X = (X_1, X_2, X_3)^{\mathrm{T}}$,

$$\mu(X,A) = \exp\left\{2 - 0.1X_1 - 0.2X_2 + 0.2X_3 + A\frac{2\mathrm{sign}(X_3 - X_2^2 + 1)}{2 + |X_3 - X_2^2 + 1|}\right\},$$

and ϵ is generated from a normal distribution with mean 0 and variance 0.25. In addition, we considered two different propensity score models for the treatment indicator A: $\pi(X) = 0.5$, which represents a randomization study; $\mathrm{logit}\{\pi(X)\} = 0.5X_1 - 0.5X_2 + 0.5X_3$, which represents an observational study.

We considered four different scenarios of the covariate distributions for $\mathbb{P}^{\mathrm{s}}$ and $\mathbb{P}^{\mathrm{t}}$, which are summarized in Table 2.3 and Table 2.4. In Scenario 1, the covariate distributions of the source and target populations are the same. We have $\mathbb{P}^{\mathrm{s}} = \mathbb{P}^{+}_{EB} = \mathbb{P}^{+}_{EL} = \mathbb{P}^{\mathrm{t}}$. In Scenario 2, the ratio $f^{\mathrm{t}}(X)/f^{\mathrm{s}}(X)$ can be written as $\exp\{\ln(0.4)+\ln(4)X_1\}$ or $1/\{1-1.875(X_1-0.8)\}$. It can be shown that $W(X;\lambda^*) \propto f^{\mathrm{t}}(X)/f^{\mathrm{s}}(X)$ for both calibration methods. Therefore, we have $\mathbb{P}^{+}_{EB} = \mathbb{P}^{+}_{EL} = \mathbb{P}^{\mathrm{t}}$ even if we only use means of covariates as the summary statistics from the target population. This implies $V^{+}_{EB}(\beta) = V^{+}_{EL}(\beta) = V^{\mathrm{t}}(\beta)$, and both $\widehat{V}^{\mathrm{c}}_{EB}(\widehat{\beta}^{\mathrm{c}}_{EB})$ and $\widehat{V}^{\mathrm{c}}_{EL}(\widehat{\beta}^{\mathrm{c}}_{EL})$ are consistent estimators of $V^{\mathrm{t}}(\beta^{\mathrm{t}})$ when either the propensity score or conditional mean outcome model is correctly specified. However, in Scenarios 3 and 4, $W(X;\lambda^*)$ is no longer proportional to $f^{\mathrm{t}}(X)/f^{\mathrm{s}}(X)$. Thus, $\widehat{V}^{\mathrm{c}}_{EB}(\widehat{\beta}^{\mathrm{c}}_{EB})$ and $\widehat{V}^{\mathrm{c}}_{EL}(\widehat{\beta}^{\mathrm{c}}_{EL})$ are doubly robust estimators only for the value functions of their corresponding pseudo populations, but not for that of the target population.

Table 2.3: Covariate distributions for $\mathbb{P}^{\mathrm{s}}$ and $\mathbb{P}^{\mathrm{t}}$ used in the simulation studies.

Scenario	$f^{\mathrm{s}}(X)$	$f^{\mathrm{t}}(X)$
1	$X_1 \sim \text{Bernoulli}(0.5)$ $(X_2, X_3)^{\mathrm{T}} \sim N((-1,0)^{\mathrm{T}}, \Sigma_1)$	$X_1 \sim \text{Bernoulli}(0.5)$ $(X_2, X_3)^{\mathrm{T}} \sim N((-1,0)^{\mathrm{T}}, \Sigma_1)$
2	$X_1 \sim \text{Bernoulli}(0.5)$ $(X_2, X_3)^{\mathrm{T}} \mid X_1 = 1 \sim N((1,-1)^{\mathrm{T}}, \Sigma_1)$ $(X_2, X_3)^{\mathrm{T}} \mid X_1 = 0 \sim N((-1,1)^{\mathrm{T}}, \Sigma_2)$	$X_1 \sim \text{Bernoulli}(0.8)$ $(X_2, X_3)^{\mathrm{T}} \mid X_1 = 1 \sim N((1,-1)^{\mathrm{T}}, \Sigma_1)$ $(X_2, X_3)^{\mathrm{T}} \mid X_1 = 0 \sim N((-1,1)^{\mathrm{T}}, \Sigma_2)$
3	$X_1 \sim \text{Bernoulli}(0.7)$ $(X_2, X_3)^{\mathrm{T}} \sim N((0.1,-0.2)^{\mathrm{T}}, \Sigma_1)$	$X_1 \sim \text{Bernoulli}(0.8)$ $(X_2, X_3)^{\mathrm{T}} \mid X_1 = 1 \sim N((1,-1)^{\mathrm{T}}, \Sigma_1)$ $(X_2, X_3)^{\mathrm{T}} \mid X_1 = 0 \sim N((-1,1)^{\mathrm{T}}, \Sigma_2)$
4	$X_1 \sim \text{Bernoulli}(0.6)$ $(X_2, X_3)^{\mathrm{T}} \sim N((0,0)^{\mathrm{T}}, \Sigma_1)$	$X_1 \sim \text{Bernoulli}(0.8)$ $(X_2, X_3)^{\mathrm{T}} \mid X_1 = 1 \sim N((1,-1)^{\mathrm{T}}, \Sigma_1)$ $(X_2, X_3)^{\mathrm{T}} \mid X_1 = 0 \sim N((-1,1)^{\mathrm{T}}, \Sigma_2)$

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

Table 2.4: Summary statistics of X_1, X_2, X_3 in different scenarios.

Population	Statistics	Scenario			
		1	2	3	4
$\mathbb{P}^s$	Mean	$0.5, -1, 0$	$0.5, 0, 0$	$0.7, 0.1, -0.2$	$0.6, 0, 0$
	Variance	$0.25, 1, 1$	$0.25, 2, 2$	$0.21, 1, 1$	$0.24, 1, 1$
$\mathbb{P}^t$	Mean	$0.5, -1, 0$	$0.8, 0.6, -0.6$		
	Variance	$0.25, 1, 1$	$0.16, 1.64, 1.64$		

We considered a source sample with size $n = 250, 1000$. For each setting, we conducted 500 replications. In our implementation, the propensity score and conditional mean outcome models are estimated using two methods:

(I) Both are estimated based on posited parametric models. In particular, the propensity score is estimated using a correctly specified logistic regression model, while the conditional mean outcome is estimated using a linear model with all the covariates and covariate-treatment interactions, which is a misspecified model.

(II) The propensity score is estimated nonparametrically using a generalized additive model, and the conditional mean outcome model $\mu(x, a)$ is estimated nonparametrically using the random forest for $a = 0$ and 1, separately.

We also implemented Q-learning as a benchmark for comparison. Specifically, we fitted linear models for Q-functions and inferred optimal linear ITRs from the estimated Q-functions. An ITR estimated by Q-learning is denoted as $d(x; \widehat{\beta}_Q)$. To evaluate and compare the performance of estimated optimal ITRs obtained from the original AIPW estimator, proposed calibrated AIPW estimators, and Q-learning, we compute the corresponding value functions and percentages of correct decisions for the target population. Specifically, we generate covariates X^t for a large sample with size $N = 10^5$ from the target population. The value function of an estimated ITR $d(x; \widehat{\beta})$, where $\widehat{\beta} = \widehat{\beta}^o$, $\widehat{\beta}^c_{EB}$, $\widehat{\beta}^c_{EL}$, or $\widehat{\beta}_Q$ is computed by $V^t(\widehat{\beta}) = N^{-1} \sum_{i=1}^N \mu\{X_i^t, d(X_i^t; \widehat{\beta})\}$, and its associated percentage of correct decisions is $1 - N^{-1} \sum_{i=1}^N |d(X_i^t; \widehat{\beta}) - d(X_i^t; \beta^t)|$. Here, the true optimal ITR $d(X; \beta^t)$ for the target population is obtained by maximizing $V^t(\beta)$ over β using the grid-search method. We report the values and percentages of correct decisions results of $d(x; \widehat{\beta}^o)$, $d(x; \widehat{\beta}^c_{EB})$, $d(x; \widehat{\beta}^c_{EL})$, and $d(x; \widehat{\beta}_Q)$ for the observational study in Figure 2.1 (method I) and Figure 2.2 (method II). Similar results for the randomization study are provided in Appendix A.2.

We have the following observations. In Scenario 1, the optimal ITR estimated by Q-learning

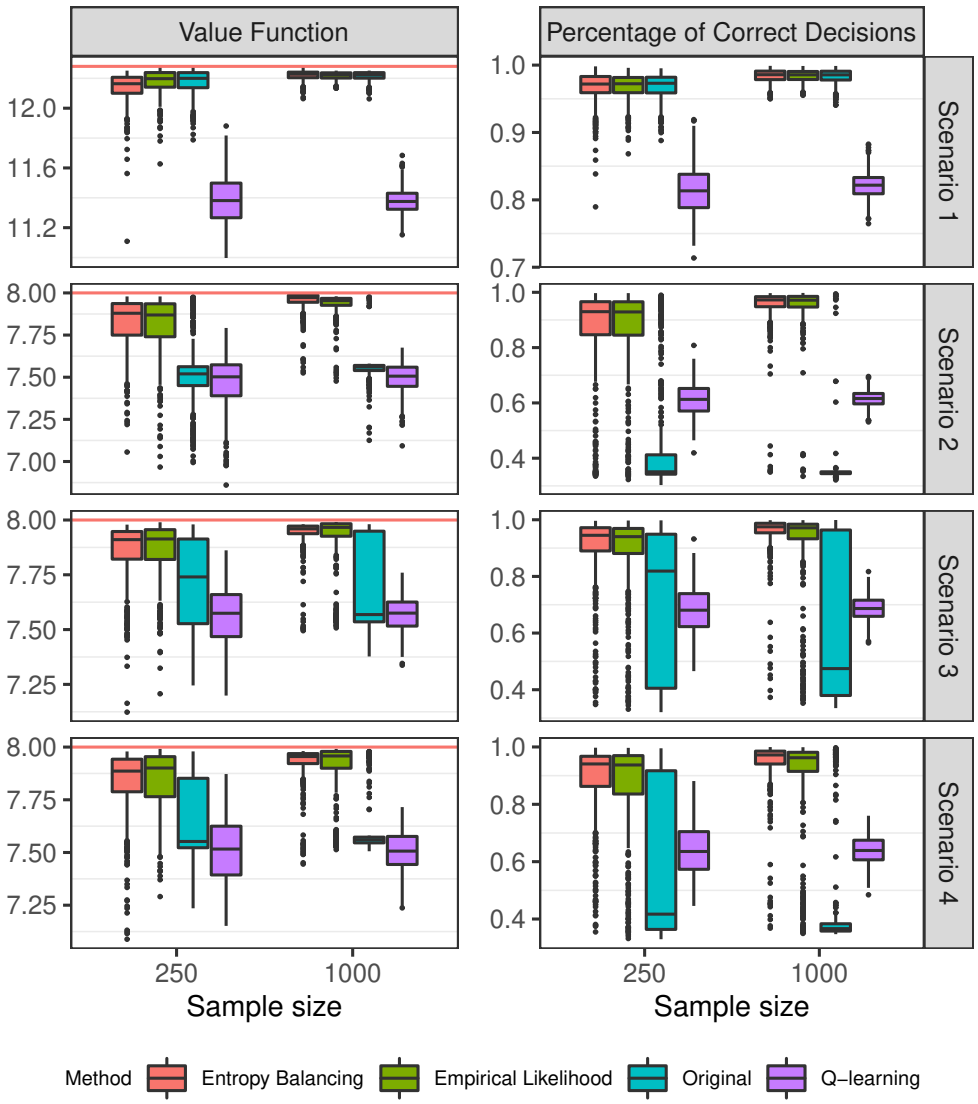

Figure 2.1: The value and percentage of correct decisions results of estimated optimal ITRs for the observational study with implementation method I. The red lines are the values of the true optimal ITRs for the target population.

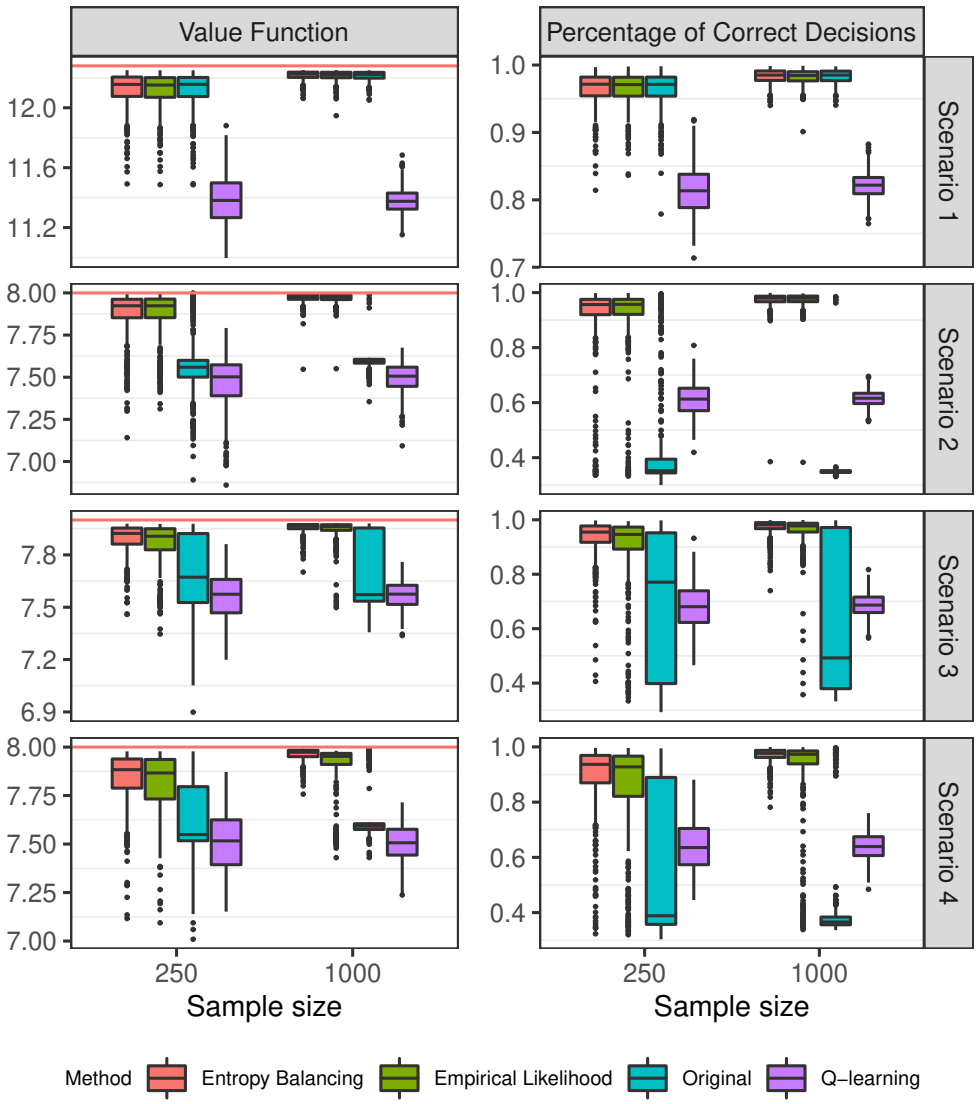

Figure 2.2: The value and percentage of correct decisions results of estimated optimal ITRs for the observational study with implementation method II. The red lines are the values of the true optimal ITRs for the target population.

has poor performance in terms of value and percentage of correct decisions, due to the misspecification of Q-function. All other three estimated optimal ITRs have good and comparable performance in terms of values and percentages of correct decisions, which is expected since $\mathbb{P}^{\mathrm{s}} = \mathbb{P}^{+}_{EB} = \mathbb{P}^{+}_{EL} = \mathbb{P}^{\mathrm{t}}$. In addition, as the sample size increases, the means of value functions become closer to the true optimal value for the target population, percentages of correct decisions get closer to 1, and the standard deviations of value functions and percentages of correct decisions become smaller. However, in Scenarios 2-4 where $\mathbb{P}^{\mathrm{s}} \neq \mathbb{P}^{\mathrm{t}}$, the estimated optimal ITR obtained using the original method has poor performance: the means of value functions are much smaller than the true optimal value for the target population and percentages of correct decisions are also much smaller than 1. This implies that the estimated optimal ITR obtained using the original method may not generalize well to the target population when $\mathbb{P}^{\mathrm{s}} \neq \mathbb{P}^{\mathrm{t}}$. The optimal ITR estimated by Q-learning still yields poor performance. However, the estimated optimal ITRs obtained using the proposed calibration methods still have competitive performance similar to those observed in Scenario 1. This supports that the proposed calibration using summary statistics can improve the treatment decision for the target population.

Next, we study the estimation and inference results of $\widehat{V}^{\mathrm{c}}_{EB}(\widehat{\beta}^{\mathrm{c}}_{EB})$ and $\widehat{V}^{\mathrm{c}}_{EL}(\widehat{\beta}^{\mathrm{c}}_{EL})$. For implementation method I, the asymptotic variances of $\widehat{V}^{\mathrm{c}}_{EB}(\widehat{\beta}^{\mathrm{c}}_{EB})$ and $\widehat{V}^{\mathrm{c}}_{EL}(\widehat{\beta}^{\mathrm{c}}_{EL})$ were estimated using the results established in Theorem 1, while for implementation method II, the corresponding asymptotic variances were estimated using the results established in Theorem 2 because Condition (A11') holds. In our simulations, we observed that the empirical likelihood method may produce a few extreme calibration weights in Scenarios 3 and 4. These extreme weights usually do not inflate the biases of $\widehat{V}^{\mathrm{c}}_{EL}(\widehat{\beta}^{\mathrm{c}}_{EL})$, but they do lead to overestimated standard errors due to the instability in variance estimation. To control the effects of these extreme weights, we stabilize the weights by reducing the large weights $\widehat{w}_i$ $(> a_n^{-1})$ to $\tilde{w}_i$ according to $(\tilde{w}_i)^{-1} = (\widehat{w}_i)^{-1} + a_n$ for $a_n = o_p(n)$. Such a stabilization leads all weights to be no larger than a_n^{-1}. The rationale for considering a_n to be $o_p(n)$ is that because $w_i \propto n^{-1}$, the stabilization does not affect the weights asymptotically. In the simulation study, we take $a_n = 12 \log n$. Based on our numerical studies, such a stabilization does not affect the biases of $\widehat{V}^{\mathrm{c}}_{EL}(\widehat{\beta}^{\mathrm{c}}_{EL})$ much but can give a reasonable standard error estimate. On the other hand, the calibration weights computed using the entropy balancing method do not have any extreme values in all four scenarios. Such an observation is consistent with the findings in the literature since the entropy balancing loss tends to penalize the deviation of the estimated weights $\widehat{w}_i$'s from n^{-1} more than the empirical likelihood method. We report the mean and standard deviation of $\widehat{V}^{\mathrm{c}}_{EB}(\widehat{\beta}^{\mathrm{c}}_{EB})$ and $\widehat{V}^{\mathrm{c}}_{EL}(\widehat{\beta}^{\mathrm{c}}_{EL})$, the mean of estimated standard errors and the empirical coverage probability (CP) of 95% Wald-type confidence intervals. The true optimal values $V^{+}_{EB}(\beta^{*}_{EB})$ and $V^{+}_{EL}(\beta^{*}_{EL})$ are computed using the grid-search method based on a large dataset generated from the corresponding pseudo

populations similar to the computation of $V^{\mathrm{t}}(\beta^{\mathrm{t}})$. In addition, we consider two types of CP: (1) CP^{+} for the optimal values $V^{+}_{EB}(\beta^{*}_{EB})$ or $V^{+}_{EL}(\beta^{*}_{EL})$ of the corresponding pseudo population; (2) CP^{t} for the optimal values $V^{\mathrm{t}}(\beta^{\mathrm{t}})$ of the target population. Simulation results for the observational study are summarized in Table 2.5 with implementation method I and Table 2.6 with implementation method II. Similar results for the randomization study are provided in Appendix A.2.

Table 2.5: Simulation results for the observational study with implementation method I. Mean, the average of estimates; SD, the empirical standard deviation of estimates; SE, the mean of estimated standard errors; CP^{+}(%), the empirical coverage probability of a 95% confidence interval for $V^{+}_{EB}(\beta^{*}_{EB})$ or $V^{+}_{EL}(\beta^{*}_{EL})$; CP^{t}(%), the empirical coverage probability of a 95% confidence interval for $V^{\mathrm{t}}(\beta^{\mathrm{t}})$.

	Scenario	1		2		3		4	
Method	n	250	1000	250	1000	250	1000	250	1000
	$V^{\mathrm{t}}(\beta^{\mathrm{t}})$	12.28		8.00		8.00		8.00	
Entropy Balancing	$V^{+}_{EB}(\beta^{*}_{EB})$	12.28		8.00		7.99		7.99	
	Mean	12.34	12.32	8.20	8.06	8.22	8.08	8.24	8.08
	SD	0.42	0.21	0.36	0.17	0.31	0.16	0.45	0.17
	SE	0.47	0.24	0.41	0.19	0.35	0.17	0.43	0.19
	CP^{+}	96.8	95.6	95.4	95.6	95.0	95.2	94.6	94.8
	CP^{t}	96.8	95.6	95.4	95.6	95.2	96.0	94.8	96.0
Empirical Likelihood	$V^{+}_{EL}(\beta^{*}_{EL})$	12.28		8.00		8.14		8.16	
	Mean	12.36	12.31	8.17	8.06	8.09	8.19	7.90	8.22
	SD	0.41	0.22	0.37	0.16	0.35	0.20	0.43	0.26
	SE	0.38	0.20	0.41	0.19	0.32	0.20	0.37	0.27
	CP^{+}	93.6	96.0	96.4	97.0	92.0	96.8	81.2	94.0
	CP^{t}	93.6	96.0	96.4	97.0	93.2	88.4	87.0	93.4

We have the following observations. In Scenarios 1 and 2, since $\mathbb{P}^{+}_{EB} = \mathbb{P}^{+}_{EL} = \mathbb{P}^{\mathrm{t}}$, we have $V^{+}_{EB}(\beta^{*}_{EB}) = V^{+}_{EL}(\beta^{*}_{EL}) = V^{\mathrm{t}}(\beta^{\mathrm{t}})$. Both calibrated value estimators are nearly unbiased. The mean of estimated standard errors is close to the standard deviation of the estimators, and the empirical coverage probability of 95% confidence intervals is close to the nominal level for all settings. In Scenarios 3 and 4, $V^{+}_{EB}(\beta^{*}_{EB})$ or $V^{+}_{EL}(\beta^{*}_{EL})$ is no longer equal to $V^{\mathrm{t}}(\beta^{\mathrm{t}})$. However, we can see that both $V^{+}_{EB}(\beta^{*}_{EB})$ and $V^{+}_{EL}(\beta^{*}_{EL})$ are close to $V^{\mathrm{t}}(\beta^{\mathrm{t}})$. In particular, the difference between $V^{+}_{EB}(\beta^{*}_{EB})$ and $V^{\mathrm{t}}(\beta^{\mathrm{t}})$ is nearly negligible. This implies that both calibration methods give good approximation of the target population, while the entropy balancing method is better than the empirical likelihood method for the considered Scenarios 3 and 4. A possible explanation is that the probability $\Pr(S = 0 \mid X)$ can be well approximated by a logistic regression under

Table 2.6: Simulation results for the observational study with implementation method II. Mean, the average of estimates; SD, the empirical standard deviation of estimates; SE, the mean of estimated standard errors; $CP^+(\%)$, the empirical coverage probability of a 95% confidence interval for $V^+_{EB}(\beta^*_{EB})$ or $V^+_{EL}(\beta^*_{EL})$; $CP^t(\%)$, the empirical coverage probability of a 95% confidence interval for $V^t(\beta^t)$.

Method	Scenario	1		2		3		4	
	n	250	1000	250	1000	250	1000	250	1000
	$V^t(\beta^t)$	12.28		8.00		8.00		8.00	
Entropy Balancing	$V^+_{EB}(\beta^*_{EB})$	12.28		8.00		7.99		7.99	
	Mean	12.42	12.32	8.12	8.03	8.15	8.06	8.19	8.06
	SD	0.50	0.22	0.33	0.15	0.29	0.13	0.34	0.15
	SE	0.54	0.25	0.38	0.18	0.30	0.14	0.37	0.17
	CP^+	97.4	97.0	96.8	96.2	93.8	95.4	94.6	95.4
	CP^t	97.4	97.0	96.8	96.2	94.6	95.6	95.0	96.0
Empirical Likelihood	$V^+_{EL}(\beta^*_{EL})$	12.28		8.00		8.14		8.16	
	Mean	12.42	12.32	8.11	8.03	8.06	8.18	7.86	8.16
	SD	0.50	0.23	0.33	0.15	0.31	0.17	0.43	0.23
	SE	0.46	0.21	0.37	0.17	0.29	0.17	0.35	0.24
	CP^+	95.4	95.4	96.6	96.0	91.6	95.4	74.4	93.2
	CP^t	95.4	95.4	96.6	96.0	94.0	84.0	83.4	96.2

Scenarios 3 and 4 so that the entropy balancing calibration method can approximate the target population very well. Moreover, the entropy balancing estimators are nearly unbiased, the mean of estimated standard errors is close to the standard deviation of the estimators, and the empirical coverage probabilities of 95% confidence intervals for both $V^+_{EB}(\beta^*_{EB})$ and $V^t(\beta^t)$ are close to the nominal level for all settings. For the empirical likelihood method, as n increases, the mean of estimators get closer to its true value $V^+_{EL}(\beta^*_{EL})$, the mean of estimated standard errors get closer to the standard deviation of estimators, and the empirical coverage probability of 95% confidence intervals for $V^+_{EL}(\beta^*_{EL})$ get closer to the nominal level as expected. However, because of the difference between $V^+_{EL}(\beta^*_{EL})$ and $V^t(\beta^t)$, the empirical coverage probability of 95% confidence intervals for $V^t(\beta^t)$ is lower than the nominal level for some settings even when n increases to 1000. Finally, standard deviations of the estimators reported in Table 2.6 for implementation method II are generally smaller than the corresponding values reported in Table 2.5 for implementation method I. Such efficiency gains are mainly due to the nonparametric fit of the conditional mean outcome model in implementation method II compared with the misspecified parametric conditional mean outcome model used in implementation method I.

We also compared the original AIPW estimator without calibration with the calibrated AIPW estimators under Scenario 1, where the source and target populations are identical. As expected,

all three estimators are consistent for the optimal value of the target population. The standard deviations of the original AIPW estimator for observational study with implementation method II are $0.55, 0.27$ for $n = 250, 1000$. These values are larger than the corresponding values of calibrated estimators in Table 2.6, which supports the results established in Theorem 3.

2.6 Real Data Application

We illustrate the proposed method using an application to data from the eICU collaborative research database (eICU-CRD) (Goldberger et al. 2000; Pollard et al. 2018, 2019) and the MIMIC-III clinical database (Goldberger et al. 2000; Johnson et al. 2016, 2019). Specifically, we use the eICU dataset as the source population, while treating the MIMIC-III dataset as the target population. Both MIMIC-III and eICU data consist of patients who suffered from sepsis. The eICU-CRD is a multi-center ICU database comprising de-identified health-related data associated with over 200,000 admissions to ICUs across the US between 2014-2015. The MIMIC-III database is a single-center ICU database comprising 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. It is likely that the populations in the two databases have some heterogeneity.

Both eICU and MIMIC-III data collected information from ICU patients with sepsis disease, and thus contain common baseline covariates and treatment. In our study, we consider $p = 7$ baseline covariates in both samples: 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). Here, treatment is coded as 1 if receiving the vasopressor, and 0 if receiving other medical supervisions such as IV fluid resuscitation. We consider the cumulative balance (mL) as the outcome of interest. A positive cumulative balance indicates that a patient's fluid input is higher than their output. The condition describing excess fluid is known as hypervolaemia or fluid overload. In critically ill patients, fluid overload is related to increased mortality and also leads to several complications like pulmonary edema, cardiac failure, tissue breakdown, and impaired bowel function (Claure-Del Granado and Mehta 2016). A negative cumulative balance indicates that a patient's fluid output is higher than their input. The condition describing inadequate fluid is known as hypovolaemia. Patients with severe hypovolemia can develop ischemic injury of vital organs, leading to multi-system organ failure (Taghavi and Askari 2021). We use $Y = -|\text{cumulative balance}|$ as the outcome, so a larger value of the outcome is better. After removing abnormal values, the MIMIC-III dataset consists of 10746 subjects, among which 2242 patients were treated with the vasopressor, while the rest were treated with other medical supervisions. The MIMIC-III data is treated as the

target population. We sample $n = 1000$ subjects from the eICU dataset as the source sample, among which 271 patients were treated with the vasopressor, while the rest were treated with other medical supervisions. Table 2.7 summarizes the mean and standard deviation of the outcome and covariates in the source and target samples. We can see some differences in the means of some covariates, such as glucose level, blood urea nitrogen amount, and WBC count.

Table 2.7: Mean and standard deviations (in parenthesis) of baseline characteristics in the source and target datasets.

	Source	Target
−|Cumulative Balance| (Y)	-1746.6 (1561.3)	-1785.0 (1246.6)
Age (X_1)	65.7 (15.1)	66.5 (16.5)
Admission Weights (X_2)	80.0 (22.9)	79.7 (20.7)
Admission Temperature (X_3)	36.5 (1.1)	36.8 (0.8)
Glucose (X_4)	158.6 (102.7)	145.6 (72.4)
BUN (X_5)	31.7 (20.3)	27.9 (18.4)
Creatinine (X_6)	1.8 (1.5)	1.5 (1.4)
WBC (X_7)	14.4 (8.4)	12.0 (6.5)

We used the means of all seven covariates of the target population as the summary statistics to estimate the calibration weights by the entropy balancing and empirical likelihood methods. We computed three optimal linear ITRs, $d(x;\widehat{\beta}^{\mathrm{o}})$, $d(x;\widehat{\beta}^{\mathrm{c}}_{EB})$, and $d(x;\widehat{\beta}^{\mathrm{c}}_{EL})$ by maximizing the original and calibrated AIPW value function estimators based on the source sample. In our implementation, the propensity score model was estimated using a logistic regression with all covariates and the conditional mean outcome model was estimated using the random forest for treatments 0 and 1 separately. To assess the performance of these three estimated optimal ITRs for treatment decisions in the target population, we apply them to random samples drawn from the target population. Specifically, we randomly sample $N = 1000$ subjects from the MIMIC-III data as the target sample and repeat this sampling procedure 100 times. We have individual-level data from the target population, which can be used as the benchmark for evaluation. For a given ITR $d(x;\beta)$, we computed the AIPW estimator of its value function based on the target sample by

$$\widehat{V}^{\mathrm{t}}(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\mathbb{I}\{A_i^{\mathrm{t}} = d(X_i^{\mathrm{t}};\beta)\}}{\varrho^{\mathrm{t}}(A_i^{\mathrm{t}} \mid X_i^{\mathrm{t}};\widehat{\eta})}\{Y_i^{\mathrm{t}} - \widehat{\mu}_d^{\mathrm{t}}(X_i^{\mathrm{t}};\beta)\} + \widehat{\mu}_d^{\mathrm{t}}(X_i^{\mathrm{t}};\beta)\right],$$

where $\varrho^{\mathrm{t}}(A_i^{\mathrm{t}} \mid X_i^{\mathrm{t}};\widehat{\eta}) = \pi^{\mathrm{t}}(X_i^{\mathrm{t}};\widehat{\eta})A_i^{\mathrm{t}} + \{1-\pi^{\mathrm{t}}(X_i^{\mathrm{t}};\widehat{\eta})\}(1-A_i^{\mathrm{t}})$, $\widehat{\mu}_d^{\mathrm{t}}(X_i^{\mathrm{t}};\beta) = \widehat{\mu}^{\mathrm{t}}(X_i^{\mathrm{t}},1)\mathbb{I}\{d(X_i^{\mathrm{t}};\beta)=1\} + \widehat{\mu}^{\mathrm{t}}(X_i^{\mathrm{t}},0)\mathbb{I}\{d(X_i^{\mathrm{t}};\beta)=0\}$, the propensity score $\pi^{\mathrm{t}}(X_i^{\mathrm{t}};\widehat{\eta})$ was estimated using a logistic regression

model, and the conditional mean outcome models $\widehat{\mu}^{\mathrm{t}}(X_i^{\mathrm{t}}, a)$, $a = 0, 1$, were estimated using random forest.

Let $\widehat{\beta}^{\mathrm{oracle}} = \mathrm{argmax}_{\beta}\, \widehat{V}^{\mathrm{t}}(\beta)$. Then, $d(x; \widehat{\beta}^{\mathrm{oracle}})$ is the optimal linear ITR for the target sample and $\widehat{V}^{\mathrm{t}}(\widehat{\beta}^{\mathrm{oracle}})$ is the associated optimal value, which can serve as the benchmark. We also computed the estimated values of three estimated ITRs $d(x; \widehat{\beta}^{\mathrm{o}})$, $d(x; \widehat{\beta}^{\mathrm{c}}_{EB})$, and $d(x; \widehat{\beta}^{\mathrm{c}}_{EL})$ by $\widehat{V}^{\mathrm{t}}(\widehat{\beta}^{\mathrm{o}})$, $\widehat{V}^{\mathrm{t}}(\widehat{\beta}^{\mathrm{c}}_{EB})$, and $\widehat{V}^{\mathrm{t}}(\widehat{\beta}^{\mathrm{c}}_{EL})$, respectively, and their associated percentages of correct decisions, defined as $1 - N^{-1}\sum_{i=1}^{N} |d(X_i^{\mathrm{t}}; \widehat{\beta}) - d(X_i^{\mathrm{t}}; \widehat{\beta}^{\mathrm{oracle}})|$ for an estimated ITR $d(x; \widehat{\beta})$. Table 2.8 summarize the means and standard deviations of the value estimators and percentages of correct decisions over 100 replications. We can see that the ITRs obtained using the proposed calibration methods have much better performance than the original AIPW estimator without calibration. Their estimated values are much closer to the optimal value computed using the target samples and the associated percentages of correct decisions are much closer to 1. Moreover, the ITR obtained using the entropy balancing method has slightly better performance than the one obtained using the empirical likelihood method in terms of both value and percentage of correct decisions.

Table 2.8: Mean and standard deviations (in parenthesis) of the value estimators and percentages of correct decisions (PCD).

	Oracle	Entropy Balancing	Empirical Likelihood	Original
Value	-1674.1 (47.2)	-1752.0 (49.5)	-1773.8 (49.3)	-1945.3 (75.6)
PCD	/	0.80 (0.1)	0.76 (0.1)	0.31 (0.1)

CHAPTER

3

MULTIPLY ROBUST OFF-POLICY EVALUATION AND LEARNING UNDER TRUNCATION BY DEATH

3.1 Introduction

In many real-world applications of personalized decision-making, experimentation and exploration can be costly, risky, or even unethical, such as healthcare (Qian and 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 and Li 2016; Fujimoto et al. 2019; Kallus and 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 and Van Der Laan 2016; Kitagawa and Tetenov 2018; Athey and 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. 2023). 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 and 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-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 and 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 and 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. 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 and Thompson 1952), and doubly robust (DR) (Zhang et al. 2012; Dudík et al. 2014) estimators, are not consistent for the survivor value function. Furthermore, it is 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.

3.2 Preliminaries

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

3.2.1 Setup

We consider a binary action $A \in \{0,1\}$. Let $X \in \mathscr{X} \subseteq \mathbb{R}^p$ 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 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 $\{Y_i S_i(1-C_i), S_i(1-C_i), C_i, A_i, X_i, i = 1,\ldots,n\}$ and we assume they are independent and identically distributed.

A policy $\pi : \mathscr{X} \to [0,1]$ 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

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

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

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

3.2.2 Problem Formulation

Following Frangakis and 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 3.1. The outcomes are only defined in the survivors, which are a mixture of different principal strata.

Table 3.1: 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

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,

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

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 $\{Y_i S_i(1-C_i), S_i(1-C_i), C_i, A_i, X_i, i = 1,\ldots,n\}$. Our second goal is OPL; i.e., estimating the survivor-optimal policy π^*.

3.2.3 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 $\varphi(x) = Pr(A = 1|X = x)$ 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 (DR) 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. 2020; Athey and 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.

3.3 Identification, EIF, and Efficiency Bound

We first make the following identification assumptions.

Assumption 4 *(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 4 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 5 *(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 4, Assumption 5 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 6 $S(1) \geq S(0)$ *almost surely.*

Assumption 6 is known as the monotonicity assumption (Sommer and 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 3.1, the observed survivors are a mixture of the always-survivor stratum ($U = 11$) and the protectable stratum ($U = 10$).

Assumption 7 $\mathbb{E}[Y(1) \mid U = 11, X] = \mathbb{E}[Y(1) \mid U = 10, X]$.

Assumption 7, known as the principal ignorability assumption, is widely used in principal causal analysis literature (Follmann 2000; Stuart and Jo 2015). Under Assumptions 4 and 6, Assumption 7 is equivalent to

$$\mathbb{E}[Y(1)|U = 11, A = 1, S = 1, X] = \mathbb{E}[Y(1)|U = 10, A = 1, S = 1, X]. \tag{3.1}$$

(3.1) 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 5, 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 B.3.

Define the propensity score $\varphi(x)=Pr(A=1 \mid X=x)$ and let $\varphi_a(x)=a\varphi(x)+(1-a)\{1-\varphi(x)\}$. Define the non-censoring probability $K_a(x)=Pr(C=0 \mid A=a, X=x)$, the observed survival probability $p_a(x)=Pr(S=1 \mid A=a, C=0, X=x)$, 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 8 $\{\varphi_a(x), K_a(x), p_a(x)\}>0$, $|\mu_a(x)|<L$, *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 mean 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 4 *Let* Π *be a policy class. Under Assumptions 4–8, for any given policy* $\pi \in \Pi$, *the survivor value function* $V(\pi)$ *is identified,*

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

Based on the identification formula provided in Theorem 4, we derive the EIF and the semiparametric efficiency bound for $V(\pi)$.

Define

$$\psi_{S(a)}=\frac{\mathbb{I}(A=a)\mathbb{I}(C=0)\{S-p_a(X)\}}{\varphi_a(X)K_a(X)}+p_a(X), \tag{3.2}$$

$$\psi_{Y(a)S(a)}=\frac{\mathbb{I}(A=a)\mathbb{I}(C=0)\{YS-\mu_a(X)p_a(X)\}}{\varphi_a(X)K_a(X)}+\mu_a(X)p_a(X). \tag{3.3}$$

Theorem 5 *Suppose* $V(\pi)$ *is identified in Theorem 4. 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 $\phi_\pi=\left[\mu_1(X)\left\{\psi_{S(0)}-\frac{p_0(X)}{p_1(X)}\psi_{S(1)}\right\}+\frac{p_0(X)}{p_1(X)}\psi_{Y(1)S(1)}\right]\pi(X)+\psi_{Y(0)S(0)}\{1-\pi(X)\}$.

3.4 Multiply Robust OPE

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

$$\widehat{V}(\pi)=\frac{\mathbb{P}_n(\widehat{\phi}_\pi)}{\mathbb{P}_n\{\widehat{\psi}_{S(0)}\}}, \tag{3.4}$$

where $\widehat{\phi}_\pi$ and $\widehat{\psi}_{S(0)}$ are the estimators for ϕ_π and $\psi_{S(0)}$, respectively. From (3.2) and (3.3), 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 $\mathscr{M}$ 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 $\mathscr{M}_{\text{ps}}$, 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, $\mathscr{M}_{\text{ps+cs}}$ 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, $\mathscr{M}_{\text{ps+cs}}\cup\mathscr{M}_{\text{sp}}$ 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 (3.4), we have a multiply robust (MR) estimator and we denote it as $\widehat{V}^{\text{MR}}(\pi)$.

Assumption 9 $\{\varphi_a(x;\widehat{\alpha}),K_a(x;\widehat{\eta}),p_a(x;\widehat{\gamma})\}>0$, $|\mu_a(x;\widehat{\zeta})|<L$, *for some* $L>0$, *and* $x\in\mathscr{X}$.

Theorem 6 *Suppose Assumptions 4–9 hold,* $\widehat{V}^{\text{MR}}(\pi)$ *is multiply robust in the sense that it is consistent for* $V(\pi)$ *under* $\mathscr{M}_{\text{ps+cs+sp}}\cup\mathscr{M}_{\text{ps+cs+om}}\cup\mathscr{M}_{\text{sp+om}}$. *Moreover, under* $\mathscr{M}_{\text{ps+cs+sp+om}}$, $\widehat{V}^{\text{MR}}(\pi)$ *has the influence function* ν_π *and achieves the semiparametric efficiency bound* $\Upsilon(\pi)$.

Remark 2 *Though in $\mathscr{M}_{\text{ps+cs+sp}}$ and $\mathscr{M}_{\text{ps+cs+om}}$, 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 $\varphi_a(x)K_a(x) = Pr(A=a, C=0 \mid X=x)$, to achieve the consistency. Therefore, the estimator is triply robust in the sense it is consistent for $V(\pi)$ if any two of $Pr(A=a, C=0 \mid X=x)$, $p_a(x)$, 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 10 *(i) $\{\widehat{\varphi}_a(x), \{\widehat{K}_a(x), \{\widehat{p}_a(x)\} > 0$, $|\widehat{\mu}_a(x)| < L$, for some $L > 0$, and $x \in \mathscr{X}$; (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 $x \in \mathscr{X}$; (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 7 *Suppose that Assumptions 4–8 and 10 hold. $\widehat{V}^{\text{MR}}(\pi)$ is asymptotically normal, has the influence function ν_π, and achieves the semiparametric efficiency bound $\Upsilon(\pi)$.*

Remark 3 *Assumption 10 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 and Laan 2011; Chernozhukov et al. 2018) and we summarize the procedure in Algorithm 1.*

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 $(\mathscr{D}_k)_{k=1}^{\xi}$ with $\mathscr{D}_k = \{Y_i S_i (1-C_i), S_i(1-C_i), C_i, A_i, X_i)\}_{i \in I_k^c}$.
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 $\mathscr{D}_k$.
    Construct an estimator $\widehat{V}_k(\pi)$ defined as (3.4).
end for
Construct an estimator $\widehat{V}^{\text{MR}}(\pi)$ by taking the average of $\widehat{V}_k(\pi)$, for $k = 1, \ldots, \xi$, i.e., $\widehat{V}^{\text{MR}}(\pi) = \frac{1}{\xi} \sum_{k=1}^{\xi} \widehat{V}_k(\pi)$.

3.5 From Robust OPE to OPL

In this section, we propose an OPL method based on the MR estimator $\widehat{V}^{\mathrm{MR}}(\pi)$ to estimate the survivor-optimal policy, which is defined as $\pi^* = \mathrm{argmax}_{\pi\in\Pi} V(\pi)$. A natural estimator for the survivor-optimal policy would be $\widehat{\pi} = \mathrm{argmax}_{\pi\in\Pi} \widehat{V}^{\mathrm{MR}}(\pi)$. 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 $\beta^* = \mathrm{argmax}_{\beta:\|\beta\|_2=1} V(\beta)$. 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 11 *(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 $P(|\tilde{X}^T\beta^*| \le \delta) = O(\delta)$, where the big-O term is uniform in $0 < \delta \le \delta_0$.*

Assumption 11(i) is a standard regularity condition used to establish the uniform convergence results; Assumption 11(ii) excludes the situation with $Pr(\tilde{X}^T\beta^* = 0) > 0$ 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 and Van Der Laan 2016).

We establish the following Lemma 1 and Theorem 8 when nuisance functions are estimated nonparametrically. Similar results for parametric estimation are provided in Appendix B.2.

Lemma 1 *Suppose that Assumptions 4–8, 10, and 11 hold, we have $n^{1/3}\|\widehat{\beta} - \beta^*\|_2 = O_p(1)$.*

Theorem 8 *Suppose that Assumptions 4–8, 10, and 11 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

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

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:

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

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 3.3, 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.

3.6 Experiments

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

$$\mu_A(X) = \exp\left\{0.2X_1 - 0.2X_2 + 0.1X_3 + 1.5A \cdot \text{sign}(X_1 - 2X_2^2 + X_3 > 0)\right\}, \tag{3.5}$$

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 4. 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

Table 3.2: OPE results with $n = 2000$. (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

(SD) results with $n = 2000$ are reported in Table 3.2 and the results with $n = 1000$ are reported in Appendix B.4. 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 $\widehat{V}^{\mathrm{MR}}(\beta)$ and the DR estimator $\widehat{V}^{\mathrm{DR}}(\beta)$ 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 Figures 3.1 and 3.2. We have the following observations. For censoring rate 30%, MR methods all have

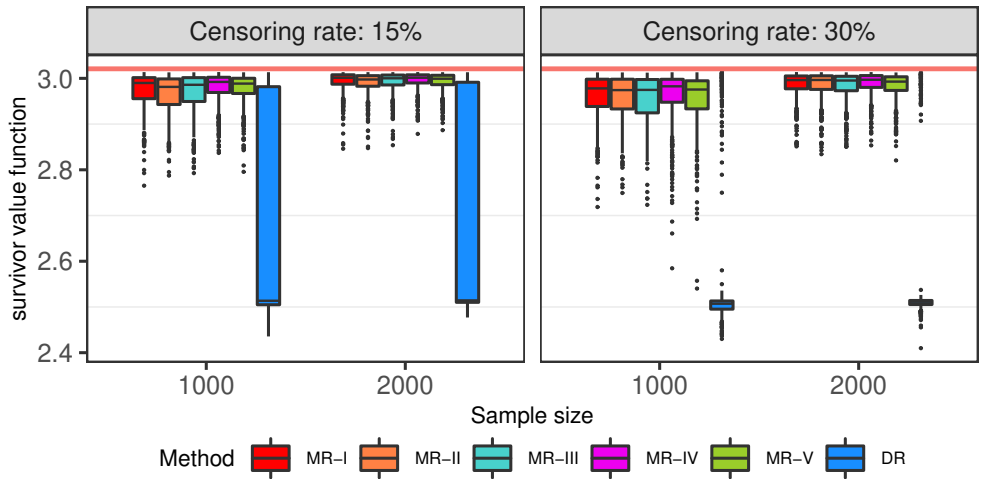

Figure 3.1: The value results of estimated survivor-optimal linear policies under different censoring rates. The red line is the true optimal survivor value function.

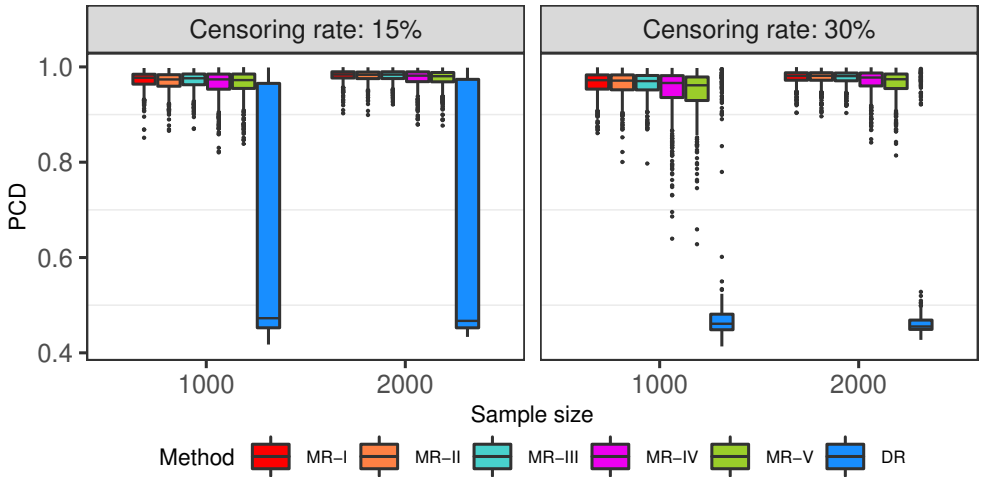

Figure 3.2: The PCD results of estimated survivor-optimal linear policies under different censoring rates.

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, 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.

Table 3.3: Inference results of $\widehat{V}(\widehat{\beta})$. Mean, the average of estimates; SD, the empirical standard deviation of estimates; SE, the mean of estimated standard errors; CP, the empirical coverage probability of a 95% confidence interval for $V(\beta^*)$.

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

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 3.3. 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.

3.6.2 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 3.4. We are interested to estimate the survivor-optimal linear policy in this case.

Table 3.4: 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 mean 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 $\widehat{\beta}^{\text{MR}}$ and $\widehat{\beta}^{\text{DR}}$.

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 $V^{\text{test}}(\widehat{\beta}^{\text{MR}})$, and the difference between $V^{\text{test}}(\beta^*)$ and $V^{\text{test}}(\widehat{\beta}^{\text{DR}})$. The training-testing procedure is repeated 50 times. We report the results of $V^{\text{test}}(\widehat{\beta}^{\text{MR}}) - V^{\text{test}}(\beta^*)$ and $V^{\text{test}}(\widehat{\beta}^{\text{DR}}) - V^{\text{test}}(\beta^*)$ in Figure 3.3. We can see that the average and variability of $V^{\text{test}}(\widehat{\beta}^{\text{MR}}) - V^{\text{test}}(\beta^*)$ is smaller than those of $V^{\text{test}}(\widehat{\beta}^{\text{DR}}) - V^{\text{test}}(\beta^*)$, which implies that OPL based on the MR method has better performance than that based on the DR method.

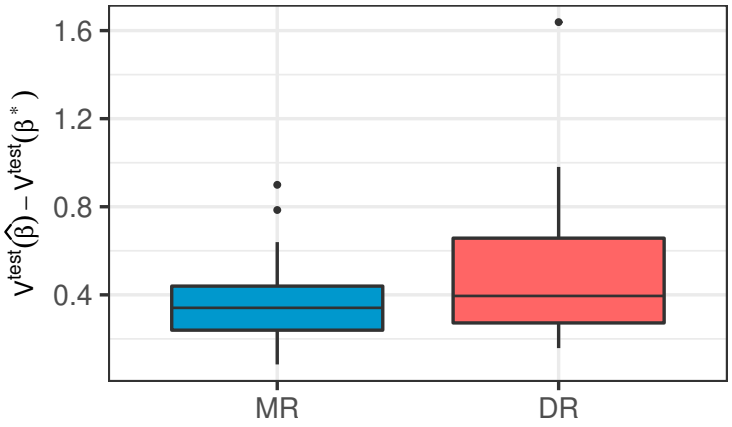

Figure 3.3: The boxplots of difference between testing values under true and estimated survivor-optimal linear policies by MR and DR methods.

CHAPTER

4

A SHADOW VARIABLE APPROACH TO POLICY EVALUATION AND LEARNING WITH ONE-SIDED FEEDBACK

4.1 Introduction

Binary decision-making problems are pervasive in the real world, encompassing domains such as bank loan approval (Pacchiano et al. 2021), job hiring (Raghavan et al. 2020), school admission (Baker and Hawn 2022), and criminal recidivism prediction (Lakkaraju et al. 2017). Often, feedback in these scenarios is one-sided. Take bank loan approval as an example: a decision-maker is presented with covariates describing a loan applicant and decides whether to grant or deny the loan. If the loan is approved, feedback regarding the applicant's repayment is subsequently received. However, if the loan is denied, no further information is obtained. There are two main objectives in these decision-making processes: (1) evaluating a policy that aims to approve loans for applicants likely to repay while denying loans to those unlikely to do so, based on the expected outcomes it achieves; and (2) deriving an optimal policy that maximizes the expected outcome.

Decision-making with one-sided feedback can be viewed as a special contextual bandit

problem with two actions, "approve" and "reject", where the outcome is observable exclusively when an individual is approved. Significant challenges arise due to the inherent heterogeneity between the approved and rejected groups—specifically, the conditional distribution of the outcome given the covariates may differ between these two groups. As a result, using an outcome model trained on approved samples to predict outcomes for the rejected group is generally unfeasible. To address model bias, one category of approaches uses exploration strategies to gather additional information from new samples, gradually reducing the bias over time (e.g. Jiang et al. 2021; Pacchiano et al. 2021). However, most existing works are restricted to binary outcomes and specific outcome models, lacking robustness to model misspecification and unable to generalize to numerical outcomes. Moreover, in real-world applications, exploration can be costly, risky, or even unethical, such as in healthcare, finance, and education. This motivates us to develop practical approaches to policy evaluation and learning for different types of outcomes from observational data (Dudík et al. 2014; Munos et al. 2016; Wang et al. 2017; Fujimoto et al. 2019; Kallus and Uehara 2020; Athey and Wager 2021).

As mentioned above, disparities between approved and rejected groups often lead to variations in outcome measures due to unobserved differences in action selection, which also serve as predictors for the outcomes. This phenomenon violates a critical assumption in the causal inference literature for identifying and estimating the value function, known as the no unmeasured confounders (NUC) assumption (Imbens 2004). This assumption, also referred to as strong ignorability (Rosenbaum and Rubin 1983) or exogeneity (Imbens and Rubin 2015), posits that actions are independent of potential outcomes given the covariates. Under this assumption, various approaches have been developed for estimating the value function, such as the inverse propensity weighting (IPW) method (Horvitz and Thompson 1952) and the doubly robust (DR) method (Dudík et al. 2011; Zhang et al. 2012; Jiang and Li 2016). The NUC assumption, however, can be often violated in many real-world scenarios. When the NUC assumption does not hold, the identifiability of the value function may be compromised, and existing estimators under this assumption may no longer be consistent for the value function.

To deal with such violations, the utilization of instrumental variables (IVs) emerges as a well-established strategy in the literature (Angrist et al. 1996; Hernán and Robins 2006; Aronow and Carnegie 2013; Wang and Tchetgen Tchetgen 2018). An IV is defined as a pretreatment variable that is independent of all unmeasured confounders, and does not have a direct causal effect on the outcome other than through the action. However, it is acknowledged that identifying suitable IVs poses a considerable challenge, given the potential existence of numerous unmeasured confounders and the difficulty in eliminating the possibility of an IV's dependence on all of them. In contrast to IVs, we consider an alternative approach using a distinct type of variables known as shadow variables (SVs) (Wang et al. 2014; Shao and Wang 2016; Miao et al.

2016; Li et al. 2024). SVs are independent of the action after conditioning on fully observed covariates and the outcome itself. Meanwhile, SVs are related to the outcome, potentially through unmeasured confounders. For example, in fairness-oriented employment, sensitive attributes such the age of candidates should be independent of the decision. However, these attributes may be related to the performance of candidates, thereby qualifying them as SVs. With the utilization of SVs, we show that the proposed value function is identifiable.

The contribution of this paper is multi-fold.

First, we propose a novel value function for decision-making with one-sided feedback. Without assuming the NUC condition, we consider a model that involves both outcomes and covariates for the action assignment mechanism. We provide identification for the proposed value function under this model by leveraging SVs.

Second, we derive the efficient influence function (EIF) and the semiparametric efficiency bound of the value function. Motivated by the EIF, we develop two different efficient estimators for the value function with binary and continuous outcomes, respectively. Our proposed estimation strategy does not require estimating the density when the outcome is continuous, thereby avoiding instability and distinguishing our methods from existing literature.

Third, we establish theoretical properties for the proposed estimators. We show the estimators are consistent and achieve semiparametric efficiency bound under mild conditions of nuisance functions approximation.

Fourth, we propose a classification-based framework for learning the optimal policy, which allows us to leverage a wide range of existing classification tools tailored to different classes of policies. Through numerical experiments, we demonstrate that the proposed method significantly outperforms conventional policy learning methods.

4.2 Preliminaries

We consider a binary action $A \in \{0,1\}$, where action 1 denotes "approve" and action 0 denotes "reject". Let $X \in \mathcal{X} \subseteq \mathbb{R}^p$ denote a vector of covariates, and $Y \in \mathbb{R}$ denote the observed outcome of interest. We assume larger values of Y are preferred by convention. The potential outcomes $Y(a)$, $a = 0,1$, which are the outcomes that would be observed if a subject received action $a = 0$ or $a = 1$, and both are well-defined in conventional policy evaluation and learning problems. Under the Stable Unit Treatment Value Assumption (SUTVA) (Rubin 2005), we have $Y = AY(1)+(1-A)Y(0)$. However, under the one-sided feedback setting, only $Y(1)$ is defined, and the outcome Y is only observed if an individual is approved ($A = 1$). In this case, the observed outcome is always $Y = Y(1)$. The observed data are then $\{O_i = (Y_i A_i, A_i, X_i), i = 1,\ldots,n\}$ and we assume they are independent and identically distributed.

A policy $\pi : \mathscr{X} \to [0,1]$ is a map from covariates to the probability of assigning action 1. In conventional policy evaluation and learning, where potential outcomes are defined for both actions, implementing a policy π in the population would yield the population mean outcome, commonly referred to as the value function, would be

$$V(\pi) = \mathbb{E}[Y(1)\pi(X) + Y(0)\{1-\pi(X)\}]. \tag{4.1}$$

Under the one-sided feedback setting, since $Y(0)$ is not defined, we can no longer use the definition of value function in (4.1). We define a new value function as

$$V_1(\pi) = \mathbb{E}\{Y(1)\pi(X)\}. \tag{4.2}$$

The interpretation of $V_1(\pi)$ is straightforward. Consider a practical example of bank loans and a deterministic policy π (where $\pi(X)$ can only take on values 0 or 1). Let $Y(1)$ denote the money earned by the bank if a loan is approved. For an applicant with covariates X, if $\pi(X) = 1$, indicating loan approval, then $Y(1)\pi(X) = Y(1)$ represents the potential financial outcome for the bank. On the other hand, if $\pi(X) = 0$, indicating loan rejection, the bank neither earns nor loses any money. Therefore, the newly defined value function $V_1(\pi)$ quantifies the expected monetary outcome for the bank when implementing policy π for loan approvals. We define the optimal policy as the one that maximizes the defined value function: $\pi^* = \underset{\pi\in\Pi}{\operatorname{argmax}}\, V_1(\pi)$. Our first goal is to evaluate a given policy π by estimating $V_1(\pi)$ using the historical data $\{O_i = (Y_i A_i, A_i, X_i), i = 1,\ldots,n\}$. Our second goal is to learn the optimal policy π^*.

4.3 Identification, EIF, and Efficiency Bound

In this section, we provide the identification of the value function $V_1(\pi)$, and establish the corresponding EIF and semiparametric efficiency bound.

4.3.1 Identification

Without assuming the NUC condition that $Y(1) \perp\!\!\!\perp A \mid X$, we consider a general action assignment mechanism that depends not only on covariates but also on the potential outcome:

$$\varphi(x,y) = \mathbb{P}\{A = 1 \mid X = x, Y(1) = y\},$$

and we assume $0 < \varphi(x,y) < 1$. Let $f(x)$ denote the marginal density of X, and let $f(y \mid x, 1)$ denote the conditional density of $Y(1)$ given $X = x$ and $A = 1$. Define $w(x) = \mathbb{P}(A = 1 \mid X = x)$.

We can show that the value function $V_1(\pi)$ has the following representation (details are given in Appendix C.1.1:

$$V_1(\pi) = \mathbb{E}\{Y(1)\pi(X)\} = \int f(x)w(x)\left\{\int y \frac{f(y \mid x,1)}{\varphi(x,y)} dy\right\}\pi(x)dx. \tag{4.3}$$

Therefore, we can identify $V_1(\pi)$ through identifying $f(x)$, $w(x)$, $f(y \mid x,1)$, and $\varphi(x,y)$. The likelihood function for a single observation is

$$f(x)w(x)^a\{1-w(x)\}^{1-a} f(y \mid x,1)^a. \tag{4.4}$$

Thus, $f(x)$, $w(x)$, and $f(y \mid x,1)$ can be identified from the observed data distribution. However, as noted in the literature (e.g. Wang et al. 2014; Miao et al. 2016), $\varphi(x,y)$ is not identifiable without further assumptions.

We assume that covariates X can be partitioned into two subsets of variables U and Z, i.e. $X = (U^T, Z^T)^T$. U and Z are variables satisfying the following assumptions.

Assumption 12 *$Z \perp\!\!\!\perp A \mid U, Y(1)$ and $Z \not\perp\!\!\!\perp Y(1) \mid U$.*

Assumption 13 *For any function $h(Y(1),U)$, $\mathbb{E}\{h(Y(1),U) \mid X, A=1\} = 0$ implies $h(Y(1),U) = 0$ almost surely.*

Assumption 12 indicates Z are SVs and $\varphi(x,y) = \mathbb{P}\{A=1 \mid X=x, Y(1)=y\} = \mathbb{P}\{A=1 \mid U=u, Y(1)=y\} = \varphi(u,y)$. For example, in fairness-oriented employment, sensitive attributes such as the age of candidates should be unrelated to the action assignment. If these attributes correlate with the performance of candidates, they can be considered SVs. SVs can be selected based on expert prior knowledge, or alternatively, representations that serve the role of shadow variables can be generated directly from observed covariates without the need for prior knowledge (Li et al. 2024). Assumption 13 is known as the conditional completeness assumption, which is widely used in identification problems (Newey and Powell 2003; Miao et al. 2015; Yang et al. 2019). This condition guarantees the uniqueness of $\varphi(u,y)$. When both $Y(1)$ and Z are categorical variables with l and m levels, respectively, Assumption 13 holds if $l < m$. When $Y(1)$ is continuous, Assumption 13 holds when $f(y \mid x,1)$ follows some common distributions, such as exponential families.

Theorem 9 *Under Assumptions 12 and 13, $f(x)$, $w(x)$, $f(y \mid x,1)$, and $\varphi(u,y)$ are identifiable, and thus $V_1(\pi)$ is identified by*

$$V_1(\pi) = \int f(x)w(x)\left\{\int y \frac{f(y \mid x,1)}{\varphi(u,y)} dy\right\}\pi(x)dx. \tag{4.5}$$

4.3.2 EIF and Efficiency Bound

The identification (4.5) motivates a rich class of estimators for the value function. However, to guide the construction of more principled estimators, we derive the EIF and the efficiency bound for the value function using semiparemetric theory (Bickel et al. 1993; Tsiatis 2006) in this section. Semiparametric models are sets of probability distributions that indexed by both finite-dimensional parametric and infinite-dimensional nonparametric components. The semiparametric efficiency bound is defined as the supremum of the Cramer-Rao lower bounds for all parametric submodels. The EIF is the influence function of a semiparametric regular and asymptotically linear estimator that achieves the semiparametric efficiency bound. We assume a general model for the action assignment mechanism, denoted as $\varphi(u, y; \eta)$, which is represented by a parameter η. Consider the Hilbert space $\mathscr{T}$ of all measurable functions of the observed data with mean zero and finite variance, equipped with covariance inner product $\langle h_1, h_2 \rangle = \mathbb{E}\{h_1(\cdot)^T h_2(\cdot)\}$, where $h_1, h_2 \in \mathscr{T}$. We first derive the nuisance tangent space and its orthogonal complement, where the nuisance tangent space is defined as the mean squared closure of all parametric submodel nuisance tangent spaces (Bickel et al. 1993; Tsiatis 2006). For the ease of exposition, we simplify $\varphi(U, Y(1); \eta)$ as $\varphi(\eta)$ and $\partial \varphi(U, Y(1); \eta)/\partial \eta$ as $\dot{\varphi}(\eta)$.

Theorem 10 *The Hilbert space $\mathscr{T}$ can be decomposed as*

$$\mathscr{T} = \Lambda_1 \oplus \Lambda_2 \oplus \Lambda_\perp,$$

where

$$\begin{aligned}
\Lambda_1 &= [h_1(X) : \mathbb{E}\{h_1(X) = 0\}], \\
\Lambda_2 &= \left[A h_2(X, Y(1)) + \frac{w(X) - A}{1 - w(X)} \mathbb{E}\{h_2(X, Y(1)) \mid X\} : \mathbb{E}\{h_2(X, Y(1)) \mid X, A = 1\} = 0 \right], \\
\Lambda_\perp &= \left\{ \frac{\varphi(\eta) - A}{\varphi(\eta)} g(X) \right\},
\end{aligned}$$

$g(X)$ is a function with the same dimension as η, and the notation $\oplus$ denotes the direct sum of two spaces that are orthogonal to each other.

Based on Theorem 10, the EIF for $V_1(\pi)$ has the following form

$$\phi_{\text{eff}} = \underbrace{h_1^*(X)}_{\in \Lambda_1} + \underbrace{A h_2^*(X) + \frac{w(X) - A}{1 - w(X)} \mathbb{E}\{h_2^*(X, Y(1)) \mid X\}}_{\in \Lambda_2} + \underbrace{D^T S_{\eta,\text{eff}}}_{\in \Lambda_\perp},$$

where $\mathbb{E}\{h_1^*(X) = 0\}, \mathbb{E}\{h_2^*(X, Y(1)) \mid X, A = 1\} = 0$, $S_{\eta,\text{eff}}$ is the efficient score for η, and D is a

vector with the same dimension as η. The efficient score $S_{\eta,\text{eff}}$ can be obtained by projecting the score function of η onto $\Lambda_\perp$, as stated in the following theorem.

Theorem 11 *Under Assumptions 12 and 13, the efficient score for η is*

$$S_{\eta,\text{eff}} = \frac{\varphi(\eta)-A}{\varphi(\eta)} \frac{\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2} \mid X, A=1\right\}}{\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2} \mid X, A=1\right\}}.$$

By projecting the value function identification (4.5) onto Λ_1,Λ_2, and $\Lambda_\perp$, we can derive $h_1^*(X)$, $h_2^*(X)$, and D. The EIF and semiparametric efficiency bound for the value function are given in the following theorem.

Theorem 12 *Under Assumptions 12 and 13, the EIF for $V_1(\pi)$ is*

$$\phi_{\text{eff}}(\pi) = \pi(X)\left[\frac{A}{\varphi(\eta)}Y + \left\{1-\frac{A}{\varphi(\eta)}\right\}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y \mid X, A=1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid X, A=1\right\}}\right] - V_1(\pi) + D S_{\eta,\text{eff}}, \tag{4.6}$$

where $D = \left(\mathbb{E}\left[\pi(X)\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y \mid X, A=1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid X, A=1\right\}}\frac{\dot{\varphi}(\eta)}{\varphi(\eta)}\right] - \mathbb{E}\left[\pi(X)\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}Y \mid X, A=1\right\}\right]\right)^T \{\text{Var}(S_{\eta,\text{eff}})\}^{-1}$. *The semiparametric efficiency bound for $V_1(\pi)$ is* $\Upsilon(\pi) = \mathbb{E}\{\phi_{\text{eff}}^2(\pi)\}$.

4.4 Efficient Policy Evaluation and Learning

4.4.1 Efficient Value Estimation

Based on the EIF (4.6), since D is a constant and $S_{\eta,\text{eff}}$ is a score function with mean zero, we propose the following estimator for $V_1(\pi)$:

$$\widehat{V}_1(\pi) = \mathbb{P}_n\left(\pi(x)\left[\frac{a}{\varphi(\widehat{\eta})}y + \left\{1-\frac{a}{\varphi(\widehat{\eta})}\right\}\frac{\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y \mid x, 1\right\}}{\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid x, 1\right\}}\right]\right), \tag{4.7}$$

where $\mathbb{P}_n[h(X)] = \frac{1}{n}\sum_{i=1}^n h(X_i)$ for any given function $h(X)$, and quantities marked with hats are estimates of their unmarked counterparts. To obtain the value estimator, we first need to estimate η and two conditional expectations $\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y \mid x, 1\right\}$ and $\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid x, 1\right\}$. A general semiparametric estimator for η can be obtained by solving the following equation:

$$\mathbb{P}_n\left[\frac{\varphi(u,y;\eta)-a}{\varphi(u,y;\eta)}g(x;\eta)\right] = 0, \tag{4.8}$$

where $g(x;\eta)$ is a calibration function with the same dimension as η. Although this estimator achieves consistency and asymptotic normality under certain regularity conditions, its efficiency is not guaranteed. To ensure minimum estimation variability introduced by $\widehat{\eta}$, we need to derive the efficient estimator of η, denoted as $\widehat{\eta}_{\text{eff}}$. This estimator can be obtained by solving the estimation equation based on the efficient score $S_{\eta,\text{eff}}$ given in Theorem 11,

$$\mathbb{P}_n\left[\frac{\varphi(\eta)-a}{\varphi(\eta)}\frac{\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}{\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2}\mid x,1\right\}}\right]=0. \tag{4.9}$$

However, the closed forms of the two conditional expectations in (4.9) are unknown and need to be approximated. We consider the following two scenarios.

Scenario I: When the outcome Y is binary, say $Y\in\{0,1\}$, we can specify a model for $\mathbb{P}(Y=1\mid X,A=1)$ and we denote its estimator as $\widehat{\mathbb{P}}(Y=1\mid X,A=1)$. The conditional expectations in (4.9) can be estimated by $\widehat{\mathbb{E}}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}\mid X,A=1\right\}=\frac{1}{\varphi(U,1;\eta)^2}\frac{\partial\varphi(U,1;\eta)}{\partial\eta}\widehat{\mathbb{P}}(Y=1\mid X,A=1)+\frac{1}{\varphi(U,0;\eta)^2}\frac{\partial\varphi(U,0;\eta)}{\partial\eta}\{1-\widehat{\mathbb{P}}(Y=1\mid X,A=1)\}$, and $\widehat{\mathbb{E}}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2}\mid X,A=1\right\}=\frac{\varphi(U,1;\eta)-1}{\varphi(U,1;\eta)^2}\widehat{\mathbb{P}}(Y=1\mid X,A=1)+\frac{\varphi(U,0;\eta)-1}{\varphi(U,0;\eta)^2}\{1-\widehat{\mathbb{P}}(Y=1\mid X,A=1)\}$. Thus we can get the efficient estimator $\widehat{\eta}_{\text{eff}}$ by solving (4.9). Next, the conditional expectations in (4.7) can be estimated by $\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y\mid X,A=1\right\}=\frac{1-\varphi(U,1;\widehat{\eta}_{\text{eff}})}{\varphi(U,1;\widehat{\eta}_{\text{eff}})^2}\widehat{\mathbb{P}}(Y=1\mid X,A=1)$, and $\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}\mid X,A=1\right\}=\frac{1-\varphi(U,1;\widehat{\eta}_{\text{eff}})}{\varphi(U,1;\widehat{\eta}_{\text{eff}})^2}\widehat{\mathbb{P}}(Y=1\mid X,A=1)+\frac{1-\varphi(U,0;\widehat{\eta}_{\text{eff}})}{\varphi(U,0;\widehat{\eta}_{\text{eff}})^2}\{1-\widehat{\mathbb{P}}(Y=1\mid X,A=1)\}$. By plugging the estimators $\widehat{\eta}_{\text{eff}}$, $\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y\mid X,A=1\right\}$, and $\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}\mid X,A=1\right\}$ into (4.7), we obtain the value estimator and denote it as $\widehat{V}_{\text{eff}}(\pi)$.

Scenario II: When the outcome Y is continuous, one can still first model the conditional density $f(Y\mid X,A=1)$. However, the density estimation often requires large sample sizes and complex algorithms to achieve accurate estimates. This can be computationally intensive and prone to high variance, particularly in high-dimensional spaces. Instead, we propose a two-step estimation strategy. In step 1, we find a root-n consistent estimator $\widehat{\eta}^{(1)}$. For example, we can choose a simple calibration function $g(x;\eta)$ and solve the equation (4.8). In step 2, we construct pseudo-outcomes $\frac{\dot{\varphi}(\widehat{\eta}^{(1)})}{\varphi^2(\widehat{\eta}^{(1)})}$ and $\frac{\varphi(\widehat{\eta}^{(1)})-1}{\varphi^2(\widehat{\eta}^{(1)})}$ and the estimators of the conditional expectations, $\widehat{\mathbb{E}}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}\mid X,A=1\right\}$ and $\widehat{\mathbb{E}}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2}\mid X,A=1\right\}$ can then be obtained using regression with these pseudo-outcomes. Thus we can get the efficient estimator $\widehat{\eta}_{\text{eff}}$ by solving (4.9). Similarly, to estimate the conditional expectations in (4.7), we can construct pseudo-outcomes $\frac{1-\varphi(\widehat{\eta}_{\text{eff}})}{\varphi(\widehat{\eta}_{\text{eff}})^2}Y$ and $\frac{1-\varphi(\widehat{\eta}_{\text{eff}})}{\varphi(\widehat{\eta}_{\text{eff}})^2}$. The estimators $\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y\mid X,A=1\right\}$, and $\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}\mid X,A=1\right\}$ can be obtained using regression with these pseudo-outcomes. By plugging the estimators $\widehat{\eta}_{\text{eff}}$, $\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y\mid X,A=1\right\}$, and $\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}\mid X,A=1\right\}$ into (4.7), we obtain the value estimator and denote it as $\widehat{V}_{\text{eff}}(\pi)$.

We now establish the theoretical results for the proposed value estimator. We first make the following assumptions for the nuisance functions and their approximations.

Assumption 14 *For all $x \in \mathcal{X}$, (i) $\{|k_1(x)|, |\widehat{k}_1(x)|\} > 0$, where $k_1(x) = \widehat{\mathbb{E}}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2} \mid x, 1\right\}$;*
(ii) for any $k_2(x) \in \left\{\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2} \mid x, 1\right\}, \mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} Y \mid x, 1\right\}\right\}$, $\{|k_2(x)|, |\widehat{k}_2(x)|\} < \infty$. (iii) for any $k_3(x) \in \left\{\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2} \mid x, 1\right\}, \mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} Y \mid x, 1\right\}, \mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2} \mid x, 1\right\}\right\}$, $\widehat{k}_3(x) \xrightarrow{p} k_3(x)$.

Assumption 14 (i) and (ii) require that the conditional expectations and their estimations are bounded. Assumption 14 (iii) requires that the conditional expectations are consistently estimated. In the case of a binary outcome, the estimation of $\mathbb{P}(Y = 1 \mid X, A = 1)$ is required to be consistent. For continuous outcomes, given the root-n consistency of $\widehat{\eta}^{(1)}$, we only require that the regression with constructed pseudo-outcomes is consistent. This can be achieved by various machine and deep learning models (e.g. Kennedy 2016; Farrell et al. 2021).

Theorem 13 *Under Assumptions 12, 13, and 14 (i) (ii), $\widehat{V}_{\text{eff}}(\pi)$ is a consistent estimator for $V_1(\pi)$. Additionally, if Assumption 14 (iii) holds, $\widehat{V}_{\text{eff}}(\pi)$ achieves the semiparametric efficiency bound $\Upsilon(\pi)$.*

4.4.2 From Efficient Policy Evaluation to Policy Leaning

In this section, we consider a deterministic policy class Π and propose a policy learning method based on the efficient estimator $\widehat{V}_{\text{eff}}(\pi)$ to estimate the optimal policy, $\pi^* = \text{argmax}_{\pi \in \Pi} V_1(\pi)$. A natural estimator for the optimal policy π^* would be $\widehat{\pi} = \text{argmax}_{\pi \in \Pi} \widehat{V}_{\text{eff}}(\pi)$. However, this direct search poses a significant challenge as it typically involves non-convex and non-smooth optimization problems and can be computationally expensive. We have the following proposition to transform it into a weighted classification problem.

Proposition 1 *Maximizing the value estimator $\widehat{V}_{\text{eff}}(\pi)$ is equivalent to a weighted classification problem of minimizing the following loss function over $\pi \in \Pi$,*

$$n^{-1} \sum_{i=1}^{n} \mathbb{I}\{\mathbb{I}\{\widehat{\psi}(x_i, y_i, a_i) > 0\} \neq \pi(x_i)\} |\widehat{\psi}(x_i, y_i, a_i)|, \tag{4.10}$$

where $\widehat{\psi}(x_i, y_i, a_i) = \frac{a_i}{\varphi_i(\widehat{\eta}_{\text{eff}})} y_i + \left\{1 - \frac{a_i}{\varphi_i(\widehat{\eta}_{\text{eff}})}\right\} \frac{\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} Y \mid x_i, 1\right\}}{\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid x_i, 1\right\}}$, for $1 \leq i \leq n$.

With Proposition 1, we have transformed the policy learning into a weighted classification problem (4.10) where for subject i with features x_i , the true label is $\mathbb{I}\{\widehat{\psi}(x_i, y_i, a_i) > 0\}$ and the sample weight is $|\widehat{\psi}(x_i, y_i, a_i)|$. The choice of classification approach dictates the restricted class Π. Compared to a direct search, a classification-based optimizer facilitates handling more complex functional classes and allows for the use of off-the-shelf machine learning and deep learning software packages.

4.5 Experiments

We have carried out extensive simulation studies and a real data application to evaluate the performance of the proposed methods.

4.5.1 Synthetic Scenarios

We compare the proposed method with three alternative methods. One consistent but not efficient estimator for η is the solution to the estimation equation (4.8) with a simple choice $g(x;\eta)$. We denote this estimator as $\widehat{\eta}_{\text{naive}}$. The first estimator for the value function is the IPW estimator with $\widehat{\eta}_{\text{naive}}$: $\widehat{V}_{\text{IPW-naive}}(\pi) = \mathbb{P}_n\Big[\frac{a}{\varphi(\widehat{\eta}_{\text{naive}})}\, y\, \pi(x)\Big]$. The second estimator is also an IPW estimator but with $\widehat{\eta}_{\text{eff}}$: $\widehat{V}_{\text{IPW-eff}}(\pi) = \mathbb{P}_n\Big[\frac{a}{\varphi(\widehat{\eta}_{\text{eff}})}\, y\, \pi(x)\Big]$. The third estimator is the DR estimator (Zhang et al. 2012): $\widehat{V}_{\text{DR}}(\pi) = \mathbb{P}_n\big(\pi(x)\big[\frac{a}{\widehat{w}(x)}\big\{y - \widehat{\mathbb{E}}(y \mid x)\big\} + \widehat{\mathbb{E}}(y \mid x)\big]\big)$.

Policy Evaluation: We first generate covariates $X = (X_1, X_2, X_3)^T \sim N((1,-1,0)^T, \Sigma)$, where

$$\Sigma = \begin{pmatrix} 1 & -0.25 & -0.25 \\ -0.25 & 1 & -0.25 \\ -0.25 & -0.25 & 1 \end{pmatrix}.$$

We consider two types of potential outcome, continuous and binary.

Case 1: The potential outcome $Y(1)$ is generated by $Y(1) = 8X_1 - 4X_1^2 - 4X_2 + 4X_3^2 + \epsilon$, where ϵ is generated from a normal distribution with mean 0 and standard deviation 0.5. The action A is generated from $A \sim \text{Bernoulli}\{\varphi(X, Y(1))\}$, and $\text{logit}\{\varphi(X, Y(1))\} = 1/[1 + \exp\{0.5 - X_1 - X_2 - 0.1\, Y(1)\}]$. Thus, X_3 is the shadow variable. We construct three different evaluation policies as mixtures of a deterministic policy $\pi_d(X) = \mathbb{I}(2X_1 - X_1^2 - X_2 + X_3^2 > 0)$ and the uniform random policy $\pi_u(X)$ by changing a mixture parameter α, i.e., $\pi(X) = \alpha\pi_d(X) + (1-\alpha)\pi_u(X)$. The candidates of the mixture parameter α are $\{0.6, 0.3, 0.0\}$.

Case 2: The potential outcome $Y(1)$ follows a Bernoulli distribution with probability of success $1/\{1 + \exp(X_1 + X_2 + X_3)\}$. The action A is generated from $A \sim \text{Bernoulli}\{\varphi(X, Y(1))\}$, and $\text{logit}\{\varphi(X, Y(1))\} = 1/[1 - \exp\{0.5 + X_1 + X_2 + 0.5\, Y(1)\}]$. Thus, X_3 is the shadow variable. We construct three different evaluation policies as mixtures of a deterministic policy $\pi_d(X) = \mathbb{I}(X_1 + X_2 + X_3 < 0)$ and the uniform random policy $\pi_u(X)$ by changing a mixture parameter α, i.e., $\pi(X) = \alpha\pi_d(X) + (1-\alpha)\pi_u(X)$. The candidates of the mixture parameter α are $\{0.7, 0.4, 0.0\}$.

For both cases, the true value function for each evaluation policy is obtained by generating a large sample $\{X_i, Y_i(1)\}_{i=1}^N$ with size $N = 10^5$ and applying the empirical version of $V(\pi) = \mathbb{E}[Y(1)\pi(X)]$. We consider a correctly specified logistic regression model for $\varphi(\eta)$. We obtain $\widehat{\eta}_{\text{naive}}$ using $g(x;\eta) = (1, x_1, x_2, x_3)^T$. We obtain the efficient estimators $\widehat{\eta}_{\text{eff}}$ and $\widehat{V}_{\text{eff}}(\pi)$ using the

approach introduced in Section 4.4. Specifically, in case 1, all the regressions with pseudo-outcomes are using random forest (RF) models. In case 2, we estimate $\mathbb{P}(Y=1 \mid X, A=1)$ using a generalized additive model (GAM). For the DR estimator, we estimate $w(x)$ using GAM in both cases. We estimate $\mathbb{E}(y \mid x)$ using RF in case 1 and using GAM in case 2.

We consider samples with size $n = 1000, 2000$. For each case, we conduct 500 replications. The root-mean-square error (RMSE), the standard deviation (SD), and the bias results for cases 1 and 2 are reported in Table 4.1 and Table 4.2.

Table 4.1: Policy evaluation results for case 1: (a) $0.0\pi_d + 1.0\pi_u$, (b) $0.3\pi_d + 0.7\pi_u$, (c) $0.6\pi_d + 0.4\pi_u$.

	(a)			(b)			(c)		
	RMSE	SD	Bias	RMSE	SD	Bias	RMSE	SD	Bias
	$n = 1000$								
$\widehat{V}_{\text{eff}}$	0.3512	0.3480	0.0468	0.5509	0.5483	0.0530	0.7999	0.7977	0.0591
$\widehat{V}_{\text{IPW-naive}}$	0.7893	0.7890	-0.0229	0.8279	0.8278	-0.0127	0.8740	0.8740	-0.0024
$\widehat{V}_{\text{IPW-eff}}$	0.6172	0.6119	0.0807	0.8426	0.8387	0.0809	1.0852	1.0822	0.0810
$\widehat{V}_{\text{DR}}$	0.4421	0.1559	0.4138	0.4371	0.1842	0.3964	0.4364	0.2162	0.3790
	$n = 2000$								
$\widehat{V}_{\text{eff}}$	0.2003	0.1985	0.0274	0.2016	0.2005	0.0209	0.2169	0.2165	0.0143
$\widehat{V}_{\text{IPW-naive}}$	0.7057	0.7026	-0.0662	0.7363	0.7341	-0.0575	0.7733	0.7718	-0.0489
$\widehat{V}_{\text{IPW-eff}}$	0.2563	0.2539	0.0353	0.2771	0.2761	0.0228	0.3121	0.3119	0.0103
$\widehat{V}_{\text{DR}}$	0.3647	0.1077	0.3485	0.3538	0.1245	0.3312	0.3455	0.1444	0.3139

We have the following observations. $\widehat{V}_{\text{eff}}$, $\widehat{V}_{\text{IPW-naive}}$, and $\widehat{V}_{\text{IPW-eff}}$ are nearly unbiased with sample size $n = 1000, 2000$. However, $\widehat{V}_{\text{DR}}$ has a significantly larger bias when compared to other estimators. This is because the NUC assumption is violated in this setting. Among three consistent estimators $\widehat{V}_{\text{eff}}, \widehat{V}_{\text{IPW-naive}}$, and $\widehat{V}_{\text{IPW-eff}}$, $\widehat{V}_{\text{eff}}$ has the smallest standard deviation and RMSE, which is expected. One interesting observation is that for case 1, when sample size $n = 1000$, the standard deviations of $\widehat{V}_{\text{IPW-naive}}$ with policies (b) and (c) are smaller than those of $\widehat{V}_{\text{IPW-eff}}$. One possible reason is that when the sample size is small, the performance of nonparametric regressions with pseudo-outcomes may have larger variation. As the sample size increases, the standard deviations and RMSEs of three consistent estimators $\widehat{V}_{\text{eff}}, \widehat{V}_{\text{IPW-naive}}$, and $\widehat{V}_{\text{IPW-eff}}$ become smaller.

Policy Learning: We consider the same covariates as policy evaluation. The potential outcome is generated by $Y(1) = 8X_1 - 6X_1^2 - 4X_2 + 2X_3^2 + \epsilon$, where ϵ is generated from a normal distribution with mean 0 and standard deviation 0.25. The action A is generated from

Table 4.2: Policy evaluation results for case 2. (a) $0.0\pi_d + 1.0\pi_u$, (b) $0.4\pi_d + 0.6\pi_u$, (c) $0.7\pi_d + 0.3\pi_u$.

	(a)			(b)			(c)		
	RMSE	SD	Bias	RMSE	SD	Bias	RMSE	SD	Bias
	$n = 1000$								
$\widehat{V}_{\text{eff}}$	0.0172	0.0172	-0.0005	0.0207	0.0207	-0.0008	0.0239	0.0239	-0.0011
$\widehat{V}_{\text{IPW-naive}}$	0.0204	0.0204	-0.0001	0.0246	0.0246	-0.0003	0.0282	0.0282	-0.0005
$\widehat{V}_{\text{IPW-eff}}$	0.0179	0.0179	-0.0006	0.0219	0.0219	-0.0009	0.0254	0.0253	-0.0012
$\widehat{V}_{\text{DR}}$	0.0196	0.0097	0.0170	0.0223	0.0124	0.0185	0.0248	0.0152	0.0196
	$n = 2000$								
$\widehat{V}_{\text{eff}}$	0.0119	0.0119	-0.0005	0.0142	0.0142	-0.0009	0.0163	0.0163	-0.0013
$\widehat{V}_{\text{IPW-naive}}$	0.0141	0.0141	-0.0003	0.0167	0.0167	-0.0006	0.0190	0.0190	-0.0009
$\widehat{V}_{\text{IPW-eff}}$	0.0122	0.0122	-0.0004	0.0148	0.0147	-0.0007	0.0171	0.0170	-0.0009
$\widehat{V}_{\text{DR}}$	0.0179	0.0069	0.0166	0.0198	0.0087	0.0178	0.0215	0.0106	0.0187

$A \sim \text{Bernoulli}\varphi(X, Y(1)) = 1/[1 + \exp\{0.5 - X_1 - X_2 - 0.15\,Y(1)\}]$. We construct four estimators following the same procedure as in policy evaluation. We use a tree-based classification algorithm introduced in Zhou et al. (2023) and focus on depth-2 decision trees for illustration. To evaluate and compare the performance of estimated optimal policies obtained by different methods, we compute the corresponding value functions and percentages of making correct decisions (PCD). Again, we generate a large sample $\{X_i, Y_i(1)\}_{i=1}^N$ with size $N = 10^5$. For a fixed policy π, its value function is computed using the empirical version of $V(\pi) = \mathbb{E}[Y(1)\pi(X)]$. We then maximize the value function and obtain the oracle optimal depth-2 decision tree, denoted as π^*. For each estimated optimal policy $\widehat{\pi}$, its associated value function is computed using the generated large sample and the PCD is computed by $N^{-1}\sum_{i=1}^N |\widehat{\pi}(X_i) - \pi^*(X_i)|$. We report the value and PCD results for the policies obtained by different methods in Figures 4.1 and 4.2. We observe that the policy obtained by our proposed method has best performance compared with other methods, in terms of values and PCDs. For our proposed method, as the sample size increases, the means of values become larger, PCDs get close to 1, and the standard deviations of values and PCDs become smaller.

4.5.2 Real Data Application

In this section, we apply our method to a loan application dataset from a fintech company. The fintech lender aims to provide short-term credit to young salaried professionals by using their mobile and social footprints to determine their credit-worthiness. To get a loan, a customer

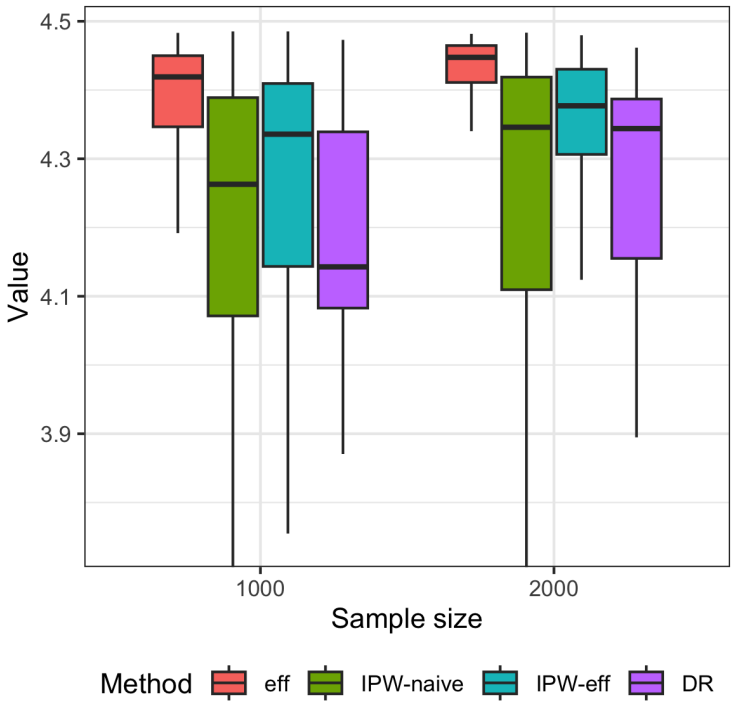

Figure 4.1: The value results of estimated policies.

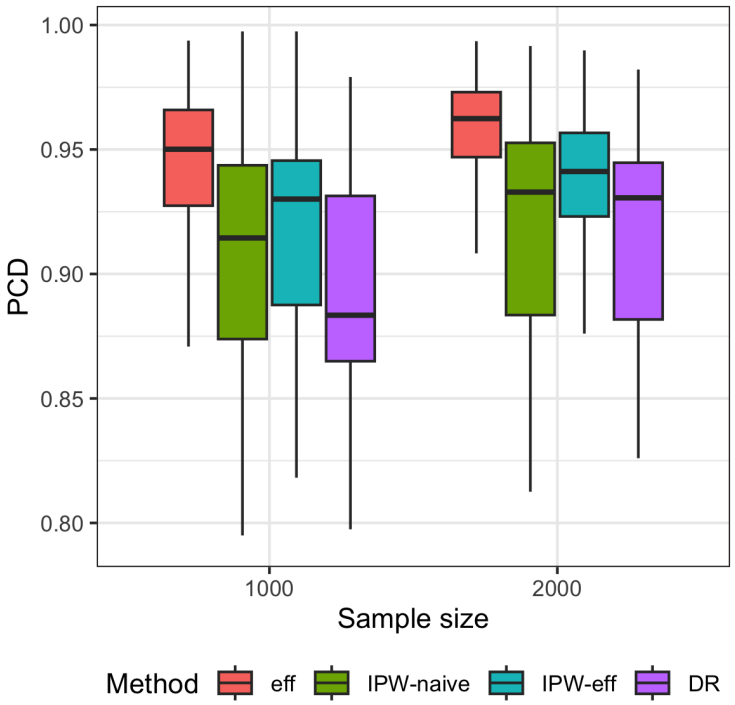

Figure 4.2: The PCD results of estimated policies.

needs to download the lending app, submit all the requisite details and documentation, and give permission to the lender to gather additional information from the smartphone, such as the number of apps and SMSs. We obtained data from the lending firm for all loans granted from February 2016 to November 2018. There are 42,777 customers in total. We select 8 covariates and they are applicants' age, salary, loan amount, CIBIL credit score, number of apps, number of SMSs, number of contacts, and number of social connections. The action A are whether or not the lender approves the loan applications. The outcome Y is defined as 1 if the loan is repaid, and -1 if the applicant defaults on the loan. We conduct hypothesis testing, and our analysis reveals no significant evidence suggesting that the number of social connections violates Assumption 12. Therefore, we consider it as a SV.

We randomly sample the training data with a size 3000 and 5000. We compare the four estimators introduced in Section 4.5.1, which are constructed using the same procedure for the binary outcome. Specifically, we estimate $\mathbb{E}(Y \mid X)$ for the DR method and $\mathbb{P}(Y \mid X, A=1)$ for the proposed method using GAM. For the DR method, we estimate $w(X)$ using GAM as well. We consider a logistic regression model for $\varphi(\eta)$ that uses all covariates (excluding the SV) and the potential outcome as predictors. We obtain $\widehat{\eta}_{\text{naive}}$ using $g(X,\eta)=(1,X^T)^T$, where X are covariates. We use the same classification algorithm as in the synthetic scenarios to estimate the optimal policy. The proposed efficient estimator over the entire dataset is used as the testing value. The training-testing procedure is repeated 100 times. We report the results of testing values in Figure 4.3. We observe that the average value of proposed method is much larger than those of other three methods, while the variability of proposed method is smaller. This implies the proposed method has better performance than other three methods.

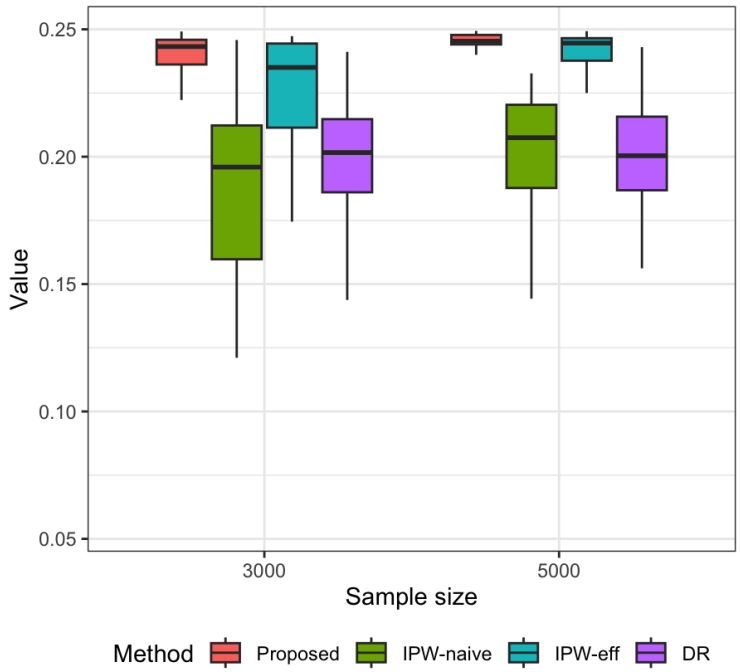

Figure 4.3: The boxplots of testing values under estimated optimal policies by different methods.

REFERENCES

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.

Aronow, P. M. and Carnegie, A. (2013). Beyond late: Estimation of the average treatment effect with an instrumental variable. *Political Analysis*, 21(4):492–506.

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

Baker, R. S. and Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4):1052–1092.

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

Buchanan, A. L., Hudgens, M. G., Cole, S. R., Mollan, K. R., Sax, P. E., Daar, E. S., Adimora, A. A., Eron, J. J., and Mugavero, M. J. (2018). Generalizing evidence from randomized trials using inverse probability of sampling weights. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(4):1193–1209.

Cattaneo, M. D., Jansson, M., and Nagasawa, K. (2020). Bootstrap-based inference for cube root asymptotics. *Econometrica*, 88(5):2203–2219.

Chakraborty, B., Laber, E. B., and Zhao, Y.-Q. (2014). Inference about the expected performance of a data-driven dynamic treatment regime. *Clinical Trials*, 11(4):408–417.

Chen, Z., Ning, J., Shen, Y., and Qin, J. (2021). Combining primary cohort data with external aggregate information without assuming comparability. *Biometrics*, 77(3):1024–1036.

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

Cheung, A. C. and Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in k-12 classrooms: A meta-analysis. *Educational Research Review*, 9:88–113.

Chu, J., Lu, W., and Yang, S. (2023). Targeted optimal treatment regime learning using summary statistics. *Biometrika*, 110(4):913–931.

Claure-Del Granado, R. and Mehta, R. L. (2016). Fluid overload in the icu: evaluation and management. *BMC Nephrology*, 17(1):109.

Cole, S. R. and Stuart, E. A. (2010). Generalizing evidence from randomized clinical trials to target populations: the actg 320 trial. *American Journal of Epidemiology*, 172(1):107–115.

Cressie, N. and Read, T. R. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):440–464.

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

Dahabreh, I. J., Robertson, S. E., Tchetgen, E. J., Stuart, E. A., and Hernán, M. A. (2019). Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics*, 75(2):685–694.

Dinglas, V. D., Faraone, L. N., and Needham, D. M. (2018). 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.

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

Dudík, M., Langford, J., and Li, L. (2011). Doubly robust policy evaluation and learning. *arXiv preprint arXiv:1103.4601*.

Fang, E. X., Wang, Z., and Wang, L. (2023). Fairness-oriented learning for optimal individualized treatment rules. *Journal of the American Statistical Association*, 118(543):1733–1746.

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

Follmann, D. (2006). Augmented designs to assess immune response in vaccine trials. *Biometrics*, 62(4):1161–1169.

Follmann, D. A. (2000). 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.

Frangakis, C. E. and Rubin, D. B. (1999). 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.

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

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

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

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., and Walther, A. (2022). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, 77(1):5–47.

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. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220.

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

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46.

Hernán, M. A. and Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream? *Epidemiology*, 17(4):360–372.

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

Huang, C.-Y. and Qin, J. (2020). A unified approach for synthesizing population-level covariate effect information in semiparametric estimation with survival data. *Statistics in Medicine*, 39(10):1573–1590.

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

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

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

Jiang, H., Jiang, Q., and Pacchiano, A. (2021). Learning the truth from only one side of the story. In *International Conference on Artificial Intelligence and Statistics*, pages 2413–2421. PMLR.

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

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

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

Johnson, A., Pollard, T., and Mark, R. (2019). Mimic-iii clinical database demo (version 1.4). *PhysioNet*. 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. (2016). Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3(1):1–9.

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

Kennedy, E. H. (2016). Semiparametric theory and empirical processes in causal inference. *Statistical Causal Inferences and Their Applications in Public Health Research*, pages 141–167.

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

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

Lakkaraju, H., Kleinberg, J., Leskovec, J., Ludwig, J., and Mullainathan, S. (2017). The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 275–284.

Lee, D., Yang, S., Dong, L., Wang, X., Zeng, D., and Cai, J. (2023). Improving trial generalizability using observational studies. *Biometrics*, 79(2):1213–1225.

Li, B., Li, H., Xiong, R., Wu, A., Wu, F., and Kuang, K. (2024). Learning shadow variable representation for treatment effect estimation under collider bias. In *Proceedings of the 41st International Conference on Machine Learning*, pages 28146–28163. PMLR.

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

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

Luckett, D. J., Laber, E. B., Kahkoska, A. R., Maahs, D. M., Mayer-Davis, E., and Kosorok, M. R. (2020). Estimating dynamic treatment regimes in mobile health using v-learning. *Journal of the American Statistical Association*, 115(530):692–706.

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

Mandel, T., Liu, Y.-E., Levine, S., Brunskill, E., and Popovic, Z. (2014). Offline policy evaluation across representations with applications to educational games. In *AAMAS*, volume 1077.

Mebane Jr, W. R. and Sekhon, J. S. (2011). Genetic optimization using derivatives: the rgenoud package for r. *Journal of Statistical Software*, 42(1):1–26.

Miao, W., Ding, P., and Geng, Z. (2016). Identifiability of normal and normal mixture models with nonignorable missing data. *Journal of the American Statistical Association*, 111(516):1673–1683.

Miao, W., Liu, L., Tchetgen, E. T., and Geng, Z. (2015). Identification, doubly robust estimation, and semiparametric efficiency theory of nonignorable missing data with a shadow variable. *arXiv preprint arXiv:1509.02556*.

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

Munos, R., Stepleton, T., Harutyunyan, A., and Bellemare, M. (2016). Safe and efficient off-policy reinforcement learning. *Advances in neural information processing systems*, 29.

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

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Newey, W. K. and Smith, R. J. (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255.

Pacchiano, A., Singh, S., Chou, E., Berg, A., and Foerster, J. (2021). Neural pseudo-label optimism for the bank loan problem. *Advances in Neural Information Processing Systems*, 34:6580–6593.

Paltiel, A. D., Schwartz, J. L., Zheng, A., and Walensky, R. P. (2021). 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.

Pollard, T., Johnson, A., Raffa, J., Celi, L., Badawi, O., and Mark, R. (2019). Icu collaborative research database (version 2.0). *PhysioNet*. Available at: `https://doi.org/10.13026/C2WM1R`.

Pollard, T. J., Johnson, A. E., Raffa, J. D., Celi, L. A., Mark, R. G., and Badawi, O. (2018). The eicu collaborative research database, a freely available multi-center database for critical care research. *Scientific Data*, 5(1):1–13.

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

Qin, J. and Zhang, B. (2007). Empirical-likelihood-based inference in missing response problems and its application in observational studies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):101–122.

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

Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 469–481.

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. (2017). Incidence and trends of sepsis in us hospitals using clinical vs claims data, 2009-2014. *Journal of the American Medical Association*, 318(13):1241–1249.

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

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6:34–58.

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

Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics*, 35(2):634–672.

Schulte, P. J., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2014). Q-and a-learning methods for estimating optimal dynamic treatment regimes. *Statistical Science*, 29(4):640.

Shao, J. and Wang, L. (2016). Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika*, 103(1):175–187.

Shi, C., Fan, A., Song, R., and Lu, W. (2018). High-dimensional a-learning for optimal dynamic treatment regimes. *The Annals of Statistics*, 46(3):925–957.

Shi, C., Lu, W., and Song, R. (2020). Breaking the curse of nonregularity with subagging—inference of the mean outcome under optimal treatment regimes. *Journal of Machine Learning Research*, 21(176):1–67.

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. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *Journal of the American Medical Association*, 315(8):801–810.

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

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

Sugiyama, M. and Kawanabe, M. (2012). *Machine learning in non-stationary environments: Introduction to covariate shift adaptation*. MIT Press.

Swaminathan, A., Krishnamurthy, A., Agarwal, A., Dudík, M., Langford, J., Jose, D., and Zitouni, I. (2017). Off-policy evaluation for slate recommendation. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 3635–3645.

Taghavi, S. and Askari, R. (2021). Hypovolemic shock. *StatPearls [Internet]*.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. Springer.

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

Van Der Vaart, A. W. and Wellner, J. (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media.

Van Der Vaart, A. W. and Wellner, J. A. (2007). Empirical processes indexed by estimated functions. *Lecture Notes-Monograph Series*, pages 234–252.

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

Wang, C., Han, B., Patel, B., and Rudin, C. (2023). In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction. *Journal of Quantitative Criminology*, 39(2):519–581.

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

Wang, S., Shao, J., and Kim, J. K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica*, pages 1097–1116.

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

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, 4(2):65–85.

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

Yang, S., Wang, L., and Ding, P. (2019). Causal inference with confounders missing not at random. *Biometrika*, 106(4):875–888.

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

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

Zhao, Q. and Percival, D. (2017). Entropy balancing is doubly robust. *Journal of Causal Inference*, 5(1).

Zhao, Y.-Q., Zeng, D., Tangen, C. M., and Leblanc, M. L. (2019). Robustifying trial-derived optimal treatment rules for a target population. *Electronic Journal of Statistics*, 13(1):1717–1743.

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

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

Zhou, Z., Athey, S., and Wager, S. (2023). Offline multi-action policy learning: Generalization and optimization. *Operations Research*, 71(1):148–183.

Zubizarreta, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, 110(511):910–922.

APPENDICES

APPENDIX

A

SUPPLEMENT TO CHAPTER 2

A.1 Technical Proofs

A.1.1 Proof of Theorem 1

Proof. The proof of Theorem 1 consists of three steps as follows.
Step 1: We show that $\widehat{V}^{\mathrm{c}}(\beta) = V^{+}(\beta) + o_p(1)$. Let

$$\psi(Y, A, X; \beta, \theta, \eta) = \frac{\mathbb{I}\{A = d(X;\beta)\}}{\pi(X;\eta)A + \{1-\pi(X;\eta)\}(1-A)}\{Y - \mu_d(X;\beta,\theta)\} + \mu_d(X;\beta,\theta).$$

By the definition of $V^{+}(\beta)$,

$$\begin{aligned} V^{+}(\beta) &= \mathbb{E}^{+}\left\{\psi(Y, A, X; \beta, \theta^*, \eta^*)\right\} \\ &= \mathbb{E}^{+}\left[\mathbb{E}\{\psi(Y, A, X; \beta, \theta^*, \eta^*) \mid X\}\right] \\ &= \int \mathbb{E}\{\psi(Y, A, X; \beta, \theta^*, \eta^*) \mid X\}\{W^*(X, \lambda^*) f^{\mathrm{s}}(X)\} dX \\ &= \mathbb{E}\left[W^*(X, \lambda^*)\mathbb{E}\{\psi(Y, A, X; \beta, \theta^*, \eta^*) \mid X\}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left\{W^*(X; \lambda^*)\psi(Y, A, X; \beta, \theta^*, \eta^*) \mid X\right\}\right] \\ &= \mathbb{E}\left\{W^*(X; \lambda^*)\psi(Y, A, X; \beta, \theta^*, \eta^*)\right\}, \end{aligned} \tag{A.1}$$

where $\mathbb{E}^+$ denotes the expectation of the pseudo population $\mathbb{P}^+$, and (A.1) follows from the double robustness property of the original AIPW estimator (Zhang et al. 2012). By the strong law of large numbers and uniform consistency, we have $\widehat{V}^{\mathrm{c}}(\beta) = V^+(\beta) + o_p(1)$.

Step 2: We show that $n^{1/3}\|\widehat{\beta}^{\mathrm{c}} - \beta^*\|_2 = O_p(1)$.

(A.) First, we show that $\widehat{\beta}^{\mathrm{c}}$ converges in probability to β^* as $n \to \infty$, by checking three conditions for the Argmax Theorem:

(a1.) By Condition (A9), the true value function $V^+(\beta)$ is twice continuously differentiable in a neighborhood of β^*.

(a2.) In Step 1, we have shown that for any β, $\widehat{V}^{\mathrm{c}}(\beta) = V^+(\beta) + o_p(1)$.

(a3.) Since $\widehat{\beta}^{\mathrm{c}} = \underset{\beta:\|\beta\|_2=1}{\operatorname{argmax}} \widehat{V}^{\mathrm{c}}(\beta)$, we have the estimated ITR as $d(X, \widehat{\beta}^{\mathrm{c}}) = \mathbb{I}(\tilde{X}^{\mathrm{T}}\widehat{\beta}^{\mathrm{c}} > 0)$ and the corresponding value function $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$ such that

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

Thus we have $\widehat{\beta}^{\mathrm{c}} \to \beta^*$, in probability, as $n \to \infty$.

(B.) Next, we show that the convergence rate of $\widehat{\beta}^{\mathrm{c}}$ is $n^{1/3}$, i.e., $n^{1/3}\|\widehat{\beta}^{\mathrm{c}} - \beta^*\|_2 = O_p(1)$. We check three conditions of the Theorem 14.4: Rate of convergence in Kosorok (2008):

(b1.) For every β in a neighborhood of β^*, i.e., $\|\beta - \beta^*\|_2 < \varepsilon$ for some $\varepsilon > 0$, by Condition (A9), we take the second order Taylor expansion on $V^+(\beta)$ at $\beta = \beta^*$,

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

Since $(V^+)''(\beta^*) < 0$, there exists $c_0 = -\frac{1}{2}(V^+)''(\beta^*) > 0$ such that $V^+(\beta) - V^+(\beta^*) < c_0\|\beta - \beta^*\|_2^2$ holds.

(b2.) Define

$$V_n^*(\beta) = \frac{1}{n}\sum_{i=1}^{n} W(\lambda^*; X_i)\left[\frac{\mathbb{I}\{A_i = d(X_i;\beta)\}}{\pi(X_i;\eta^*)A_i + \{1 - \pi(X_i;\eta^*)\}(1 - A_i)}\{Y_i - \mu_d(X_i;\beta,\theta^*)\} + \mu_d(X_i;\beta,\theta^*)\right].$$

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

$$
\begin{aligned}
&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-V^+(\beta)-\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^+(\beta^*)\}\right|\right]\\
=&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-V_n^*(\beta)+V_n^*(\beta)-V^+(\beta)-\{\widehat{V}^{\mathrm{c}}(\beta^*)-V_n^*(\beta^*)+V_n^*(\beta^*)-V^+(\beta^*)\}\right|\right]\\
\le&\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-V_n^*(\beta)-\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-V_n^*(\beta^*)\right\}\right|\right]}_{\tau_1}\\
&+\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|V_n^*(\beta)-V^+(\beta)-\left\{V_n^*(\beta^*)-V^+(\beta^*)\right\}\right|\right]}_{\tau_2},
\end{aligned}
\tag{A.2}
$$

where $\mathbb{E}^*$ is the outer expectation.

We first derive two results (b2.1) and (b2.2) to bound τ_1 and τ_2, respectively, and then we are able to show that the second condition of Theorem 14.4 in Kosorok (2008) is satisfied.

(b2.1)

$$
\begin{aligned}
&V_n^*(\beta)-V_n^*(\beta^*)\\
=&\frac{1}{n}\sum_{i=1}^n W(\lambda^*;X_i)\left[\frac{\mathbb{I}\{A_i=d(X_i;\beta)\}}{\pi(X_i;\eta^*)A_i+\{1-\pi(X_i;\eta^*)\}(1-A_i)}\{Y-\mu_d(X_i;\beta,\theta^*)\}+\mu_d(X_i;\beta,\theta^*)\right.\\
&\left.-\frac{\mathbb{I}\{A_i=d(X_i;\beta^*)\}}{\pi(X_i;\eta^*)A_i+\{1-\pi(X_i;\eta^*)\}(1-A_i)}\{Y-\mu_d(X_i;\beta^*,\theta^*)\}-\mu_d(X_i;\beta^*,\theta^*)\right]\\
=&\frac{1}{n}\sum_{i=1}^n W(\lambda^*;X_i)\left[\frac{(2A_i-1)Y_i-\mu(X_i,1;\theta^*)A_i+\mu(X_i,0;\theta^*)(1-A_i)}{\pi(X_i;\eta^*)A_i+\{1-\pi(X_i;\eta^*)\}(1-A_i)}+\mu(X_i,1;\theta^*)-\mu(X_i,0;\theta^*)\right]\\
&\left\{\mathbb{I}(\tilde{X}_i^{\mathrm{T}}\beta>0)-\mathbb{I}(\tilde{X}_i^{\mathrm{T}}\beta^*>0)\right\}.
\end{aligned}
$$

We define a class of functions

$$
\begin{aligned}
\mathscr{F}_\beta^1(y,a,x)=\Bigg\{&W(\lambda^*;x)\left[\frac{(2a-1)y-\mu(x,1;\theta^*)a+\mu(x,0;\theta^*)(1-a)}{\pi(x;\eta^*)a+\{1-\pi(x;\eta^*)\}(1-a)}+\mu(x,1;\theta^*)-\mu(x,0;\theta^*)\right]\\
&\left\{\mathbb{I}\left(\tilde{x}^{\mathrm{T}}\beta>0\right)-\mathbb{I}\left(\tilde{x}^{\mathrm{T}}\beta^*>0\right)\right\}:\|\beta-\beta^*\|_2<\varepsilon\Bigg\},
\end{aligned}
$$

where $\tilde{x}=(1,x^{\mathrm{T}})^{\mathrm{T}}$.

Let $M_1=\sup\left|W(\lambda^*;x)\left[\frac{(2a-1)y-\mu(x,1;\theta^*)a+\mu(x,0;\theta^*)(1-a)}{\pi(x;\eta^*)a+\{1-\pi(x;\eta^*)\}(1-a)}+\mu(x,1;\theta^*)-\mu(x,0;\theta^*)\right]\right|$. By Conditions (A6), (A8), and (A11), we have $M_1<\infty$. By Condition (A6), there exists a constant $0<k_0<\infty$ s.t. $|\tilde{x}^{\mathrm{T}}(\beta-\beta^*)|<k_0\varepsilon$ when $\|\beta-\beta^*\|_2<\varepsilon$. For the indicator function $\mathbb{I}\left(-k_0\varepsilon\le\tilde{x}^{\mathrm{T}}\beta^*\le k_0\varepsilon\right)$,

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

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

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

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

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

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

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

We then define the envelope of $\mathscr{F}^1_\beta(y,a,x)$ as $F_1 = M_1\mathbb{I}\big(-k_0\varepsilon \le \tilde{x}^{\mathrm{T}}\beta^* \le k_0\varepsilon\big)$. By Condition (A10), there exits a positive constant k_1 such that

$$\|F_1\|_{P,2} = M_1\sqrt{\Pr\big(-k_0\varepsilon \le \tilde{x}^{\mathrm{T}}\beta^* \le k_0\varepsilon\big)} \le M_1\sqrt{k_1\cdot 2k_0\varepsilon} = M_1\sqrt{2k_0k_1}\varepsilon^{1/2} < \infty.$$

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

$$\mathbb{G}_n\mathscr{F}^1_\beta = n^{-1/2}\sum_{i=1}^{n}\Big[\mathscr{F}^1_\beta(Y_i,A_i,X_i) - \mathbb{E}\Big\{\mathscr{F}^1_\beta(Y_i,A_i,X_i)\Big\}\Big].$$

Note that $\mathbb{G}_n\mathscr{F}^1_\beta = n^{1/2}\big[V_n^*(\beta) - V_n^*(\beta^*) - \big\{V^+(\beta) - V^+(\beta^*)\big\}\big]$. By applying Theorem 11.2 in Kosorok (2008), we have

$$\begin{aligned}\tau_2 &= \mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|V_n^*(\beta) - V^+(\beta) - \big\{V_n^*(\beta^*) - V^+(\beta^*)\big\}\right|\right]\\ &= \mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\mathbb{G}_n\mathscr{F}^1_\beta\right|\right) \le c_1 J^*_{[]}(1,\mathscr{F}^1)\|F_1\|_{P,2} \le c_1 J^*_{[]}(1,\mathscr{F}^1)M_1\sqrt{2k_0k_1}\varepsilon^{1/2},\end{aligned}$$

where c_1 is a finite constant.

Let $C_1^* \equiv c_1 J^*_{[]}(1,\mathscr{F}^1)M_1\sqrt{2k_0k_1}$, since c_1, $J^*_{[]}(1,\mathscr{F}^1)$, M_1, k_0 and k_1 are bounded, we have $C_1^* < \infty$, i.e.,

$$\tau_2 \le C_1^*\varepsilon^{1/2}. \tag{A.3}$$

(b2.2) We rewrite the form of $\widehat{V}^{\mathrm{c}}(\beta)-V_n^*(\beta)-\{\widehat{V}^{\mathrm{c}}(\beta^*)-V_n^*(\beta^*)\}$ as

$$
\begin{aligned}
&\widehat{V}^{\mathrm{c}}(\beta)-V_n^*(\beta)-\{\widehat{V}^{\mathrm{c}}(\beta^*)-V_n^*(\beta^*)\}=\widehat{V}^{\mathrm{c}}(\beta)-\widehat{V}^{\mathrm{c}}(\beta^*)-\{V_n^*(\beta)-V_n^*(\beta^*)\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\Bigg(W(\widehat{\lambda};X_i)\Bigg[\frac{(2A_i-1)Y_i-\mu(X_i,1;\widehat{\theta})A_i+\mu(X_i,0;\widehat{\theta})(1-A_i)}{\pi(X_i;\widehat{\eta})A_i+\{1-\pi(X_i;\widehat{\eta})\}(1-A_i)}+\mu(X_i,1;\widehat{\theta})-\mu(X_i,0;\widehat{\theta})\Bigg]\\
&\quad-W(\lambda^*;X_i)\Bigg[\frac{(2A_i-1)Y_i-\mu(X_i,1;\theta^*)A_i+\mu(X_i,0;\theta^*)(1-A_i)}{\pi(X_i;\eta^*)A_i+\{1-\pi(X_i;\eta^*)\}(1-A_i)}+\mu(X_i,1;\theta^*)-\mu(X_i,0;\theta^*)\Bigg]\Bigg)\\
&\quad\{\mathbb{I}(\tilde{X}_i^{\mathrm{T}}\beta>0)-\mathbb{I}(\tilde{X}_i^{\mathrm{T}}\beta^*>0)\}.
\end{aligned}
$$

We take the Taylor expansion on the above equation at $(\lambda^*,\theta^*,\eta^*)$,

$$
\begin{aligned}
&\widehat{V}^{\mathrm{c}}(\beta)-V_n^*(\beta)-\{\widehat{V}^{\mathrm{c}}(\beta^*)-V_n^*(\beta^*)\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\Bigg(\Bigg[\frac{(2A_i-1)Y_i-\mu(X_i,1;\theta^*)A_i+\mu(X_i,0;\theta^*)(1-A_i)}{\pi(X_i;\eta^*)A_i+\{1-\pi(X_i;\eta^*)\}(1-A_i)}+\mu(X_i,1;\theta^*)-\mu(X_i,0;\theta^*)\Bigg]\\
&\quad\left\{\frac{\partial W(X_i;\lambda^*)}{\partial\lambda}\right\}^{\mathrm{T}}(\widehat{\lambda}-\lambda^*)+\\
&\quad W(\lambda^*;X_i)\Bigg[\frac{-\{\partial\mu(X_i,1;\theta^*)/\partial\theta\}^{\mathrm{T}}A_i+\{\partial\mu(X_i,0;\theta^*)/\partial\theta\}^{\mathrm{T}}(1-A_i)}{\pi(X_i;\eta^*)A_i+\{1-\pi(X_i;\eta^*)\}(1-A_i)}+\\
&\quad\left\{\frac{\partial\mu(X_i,1;\theta^*)}{\partial\theta}\right\}^{\mathrm{T}}-\left\{\frac{\partial\mu(X_i,0;\theta^*)}{\partial\theta}\right\}^{\mathrm{T}}\Bigg](\widehat{\theta}-\theta^*)-\\
&\quad W(\lambda^*;X_i)(2A_i-1)\frac{(2A_i-1)Y_i-\mu(X_i,1;\theta^*)A_i+\mu(X_i,0;\theta^*)(1-A_i)}{\left[\pi(X_i;\eta^*)A_i+\{1-\pi(X_i;\eta^*)\}(1-A_i)\right]^2}\\
&\quad\left\{\frac{\partial\pi(X_i;\eta^*)}{\partial\eta}\right\}^{\mathrm{T}}(\widehat{\eta}-\eta^*)\Bigg)+o_p(n^{-1/2}).
\end{aligned}
$$

Next, we define three classes of functions,

$$
\begin{aligned}
\mathscr{F}_\beta^2(y,a,x)=&\Bigg\{\Bigg[\frac{(2a-1)y-\mu(x,1;\theta^*)a+\mu(x,0;\theta^*)(1-a)}{\pi(x;\eta^*)a+\{1-\pi(x;\eta^*)\}(1-a)}+\mu(x,1;\theta^*)-\mu(x,0;\theta^*)\Bigg]\\
&\left\{\frac{\partial W(x;\lambda^*)}{\partial\lambda}\right\}^{\mathrm{T}}\left\{\mathbb{I}\left(\tilde{x}^{\mathrm{T}}\beta>0\right)-\mathbb{I}\left(\tilde{x}^{\mathrm{T}}\beta^*>0\right)\right\}:\|\beta-\beta^*\|_2<\varepsilon\Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
\mathscr{F}_\beta^3(a,x)=&\Bigg\{W(\lambda^*;x)\Bigg[\frac{-\{\partial\mu(x,1;\theta^*)/\partial\theta\}^{\mathrm{T}}a+\{\partial\mu(x,0;\theta^*)/\partial\theta\}^{\mathrm{T}}(1-a)}{\pi(x;\eta^*)a+\{1-\pi(x;\eta^*)\}(1-a)}+\left\{\frac{\partial\mu(x,1;\theta^*)}{\partial\theta}\right\}^{\mathrm{T}}\\
&-\left\{\frac{\partial\mu(x,0;\theta^*)}{\partial\theta}\right\}^{\mathrm{T}}\Bigg]\left\{\mathbb{I}\left(\tilde{x}^{\mathrm{T}}\beta>0\right)-\mathbb{I}\left(\tilde{x}^{\mathrm{T}}\beta^*>0\right)\right\}:\|\beta-\beta^*\|_2<\varepsilon\Bigg\},
\end{aligned}
$$

$$\mathscr{F}_\beta^4(y,a,x)=\Bigg\{-W(\lambda^*;x)(2a-1)\frac{(2a-1)y-\mu(x,1;\theta^*)a+\mu(x,0;\theta^*)(1-a)}{[\pi(x;\eta^*)a+\{1-\pi(x;\eta^*)\}(1-a)]^2}$$
$$\left\{\frac{\partial\pi(x;\eta^*)}{\partial\eta}\right\}^{\mathrm{T}}\left\{\mathbb{I}\left(\tilde{x}^{\mathrm{T}}\beta>0\right)-\mathbb{I}\left(\tilde{x}^{\mathrm{T}}\beta^*>0\right)\right\}:\|\beta-\beta^*\|_2<\varepsilon\Bigg\}.$$

Let

$$M_2=\sup\left|\left[\frac{(2a-1)y-\mu(x,1;\theta^*)a+\mu(x,0;\theta^*)(1-a)}{\pi(x;\eta^*)a+\{1-\pi(x;\eta^*)\}(1-a)}+\mu(x,1;\theta^*)-\mu(x,0;\theta^*)\right]\left\{\frac{\partial W(x;\lambda^*)}{\partial\lambda}\right\}^{\mathrm{T}}\right|,$$

$$M_3=\sup\Bigg|W(\lambda^*;x)\Bigg[\frac{-\{\partial\mu(x,1;\theta^*)/\partial\theta\}^{\mathrm{T}}a+\{\partial\mu(x,0;\theta^*)/\partial\theta\}^{\mathrm{T}}(1-a)}{\pi(x;\eta^*)a+\{1-\pi(x;\eta^*)\}(1-a)}+\left\{\frac{\partial\mu(x,1;\theta^*)}{\partial\theta}\right\}^{\mathrm{T}}$$
$$-\left\{\frac{\partial\mu(x,0;\theta^*)}{\partial\theta}\right\}^{\mathrm{T}}\Bigg]\Bigg|,$$

$$M_4=\sup\left|-W(\lambda^*;x)(2a-1)\frac{(2a-1)y-\mu(x,1;\theta^*)a+\mu(x,0;\theta^*)(1-a)}{[\pi(x;\eta^*)a+\{1-\pi(x;\eta^*)\}(1-a)]^2}\left\{\frac{\partial\pi(x;\eta^*)}{\partial\eta}\right\}^{\mathrm{T}}\right|.$$

By Conditions (A6), (A8), and (A11), $M_2,M_3,M_4<\infty$. Define the envelop of $\mathscr{F}_\beta^j$ as $F_j=M_j\mathbb{I}\left(-k_0\varepsilon\le\tilde{x}^{\mathrm{T}}\beta^*\le k_0\varepsilon\right)$, for $j=2,3,4$. Similarly to (b2.1), we have

$$\|F_j\|_{P,2}=M_j\sqrt{2k_0k_1}\varepsilon^{1/2}<\infty,$$

and $\mathscr{F}_\beta^j$ is a VC class of functions. Thus, the entropy of $\mathscr{F}_\beta^j$, denoted as $J_{[]}^*(1,\mathscr{F}^j)$, is finite, i.e., $J_{[]}^*(1,\mathscr{F}^j)<\infty$, for $j=2,3,4$. We construct three empirical processes indexed by β,

$$\mathbb{G}_n\mathscr{F}_\beta^j=n^{-1/2}\sum_{i=1}^n\left[\mathscr{F}_\beta^j(Y_i,A_i,X_i)-\mathbb{E}\left\{\mathscr{F}_\beta^j(Y_i,A_i,X_i)\right\}\right].$$

By applying Theorem 11.2 in Kosorok (2008), we have

$$\mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left\|\mathbb{G}_n\mathscr{F}_\beta^j\right\|_1\right)\le c_jJ_{[]}^*(1,\mathscr{F}^j)\|F_j\|_{P,2},\quad j=2,3,4,$$

where c_2, c_3 and c_4 are finite constants.

By Theorem 2.14.5 in Van Der Vaart and Wellner (1996),

$$\left\{\mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left\|\mathbb{G}_n\mathscr{F}_\beta^j\right\|_2^2\right)\right\}^{1/2} \le l_j\left\{\mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left\|\mathbb{G}_n\mathscr{F}_\beta^j\right\|_1\right)+\|F_j\|_{P,2}\right\}$$
$$\le l_j\left\{c_j J_{[]}^*(1,\mathscr{F}^j)+1\right\}\|F_j\|_{P,2}$$
$$= l_j\left\{c_j J_{[]}^*(1,\mathscr{F}^j)+1\right\}M_j\sqrt{2k_0k_1}\varepsilon^{1/2},$$

where l_2, l_3 and l_4 are finite constants. Let

$$C_j^* \equiv l_j\left\{c_j J_{[]}^*(1,\mathscr{F}^j)+1\right\}M_j\sqrt{2k_0k_1} < \infty,$$

and we have

$$\left\{\mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left\|\mathbb{G}_n\mathscr{F}_\beta^j\right\|_2^2\right)\right\}^{1/2} \le C_j\varepsilon^{1/2}, \qquad j=2,3,4.$$

Note that

$$\tau_1 = \mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-V_n^*(\beta)-\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-V_n^*(\beta^*)\right\}\right|\right]$$
$$= \mathbb{E}^*\left\{\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|(\widehat{\lambda}-\lambda^*)^{\mathrm{T}}\mathbb{G}_n\mathscr{F}_\beta^2+(\widehat{\theta}-\theta^*)^{\mathrm{T}}\mathbb{G}_n\mathscr{F}_\beta^3+(\widehat{\eta}-\eta^*)^{\mathrm{T}}\mathbb{G}_n\mathscr{F}_\beta^4+o_p(1)\right|\right\}$$
$$\le n^{-1/2}\left[\mathbb{E}^*\left\{\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|n^{1/2}(\widehat{\lambda}-\lambda^*)^{\mathrm{T}}\mathbb{G}_n\mathscr{F}_\beta^2\right|\right\}+\mathbb{E}^*\left\{\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|n^{1/2}(\widehat{\theta}-\theta^*)^{\mathrm{T}}\mathbb{G}_n\mathscr{F}_\beta^3\right|\right\}\right.$$
$$\left.+\mathbb{E}^*\left\{\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|n^{1/2}(\widehat{\eta}-\eta^*)^{\mathrm{T}}\mathbb{G}_n\mathscr{F}_\beta^4\right|\right\}\right].$$

By the Cauchy-Schwarz Inequality, we have

$$\tau_1 \le n^{-1/2}\left\{\mathbb{E}\left(n\|\widehat{\lambda}-\lambda^*\|_2^2\right)\right\}^{1/2}\left\{\mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left\|\mathbb{G}_n\mathscr{F}_\beta^2\right\|_2^2\right)\right\}^{1/2}$$
$$+n^{-1/2}\left\{\mathbb{E}\left(n\|\widehat{\theta}-\theta^*\|_2^2\right)\right\}^{1/2}\left\{\mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left\|\mathbb{G}_n\mathscr{F}_\beta^3\right\|_2^2\right)\right\}^{1/2}$$
$$+n^{-1/2}\left\{\mathbb{E}\left(n\|\widehat{\eta}-\eta^*\|_2^2\right)\right\}^{1/2}\left\{\mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left\|\mathbb{G}_n\mathscr{F}_\beta^4\right\|_2^2\right)\right\}^{1/2}.$$

By Condition (A11), $\sqrt{n}(\widehat{\lambda}-\lambda^*)=O_p(1)$, $\sqrt{n}(\widehat{\theta}-\theta^*)=O_p(1)$ and $\sqrt{n}(\widehat{\eta}-\eta^*)=O_p(1)$, we have $M_\lambda \equiv \left\{\mathbb{E}\left(n\|\widehat{\lambda}-\lambda^*\|_2^2\right)\right\}^{1/2} < \infty$, $M_\theta \equiv \left\{\mathbb{E}\left(n\|\widehat{\theta}-\theta^*\|_2^2\right)\right\}^{1/2} < \infty$, and $M_\eta \equiv \left\{\mathbb{E}\left(n\|\widehat{\eta}-\eta^*\|_2^2\right)\right\}^{1/2} <$

∞, then

$$\tau_1 \leq n^{-1/2}\left(M_\lambda C_2^* + M_\theta C_3^* + M_\eta C_4^*\right)\varepsilon^{1/2}. \tag{A.4}$$

By (A.2), (A.3), and (A.4), we have the centered process $\widehat{V}^{\mathrm{c}} - V^+$ satisfies

$$\begin{aligned}&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-V^+(\beta)-\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^+(\beta^*)\}\right|\right]\\ \leq&\tau_1+\tau_2 \leq C_1^*\varepsilon^{1/2}+n^{-1/2}\left(M_\lambda C_2^*+M_\theta C_3^*+M_\eta C_4^*\right)\varepsilon^{1/2}.\end{aligned}$$

Let n goes infinite, we have

$$\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-V^+(\beta)-\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^+(\beta^*)\right\}\right|\right] \leq C_1^*\varepsilon^{1/2}. \tag{A.5}$$

Let $\phi_n(\varepsilon)=\varepsilon^{1/2}$, and $\alpha=\frac{3}{2}<2$, check $\frac{\phi_n(\varepsilon)}{\varepsilon^\alpha}=\varepsilon^{-1}$ is decreasing not depending on n. Therefore, the second condition holds.

(b3.) By $\widehat{\beta}^{\mathrm{c}} \to \beta^*$, in probability, as $n \to \infty$, and $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}}) \geq \sup_{\beta:\|\beta\|_2=1}\widehat{V}(\beta)$ shown previously, choose $r_n = n^{1/3}$, then r_n satisfies

$$\begin{aligned}&r_n^2\phi_n(r_n^{-1}) = n^{2/3}\phi_n(n^{-1/3})\\ =&n^{2/3}(n^{-1/3})^{1/2} = n^{1/2}.\end{aligned}$$

Thus, the third condition holds.

By the Theorem 14.4 in Kosorok (2008), we have $n^{1/3}\|\widehat{\beta}^{\mathrm{c}}-\beta^*\|_2 = O_p(1)$.

Step 3: We derive the asymptotic distribution of $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$ in this step. Note that

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

(C.) First, we show

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

which is sufficient to show $\sqrt{n}\left[\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-\widehat{V}^{\mathrm{c}}(\beta^*)\right\}-\left\{V^+(\widehat{\beta}^{\mathrm{c}})-V^+(\beta^*)\right\}\right] = o_p(1)$ and $\sqrt{n}\left\{V^+(\widehat{\beta}^{\mathrm{c}})-V^+(\beta^*)\right\} = o_p(1)$.

(c1.) First, by $n^{1/3}\|\widehat{\beta}-\beta^*\|_2 = O_p(1)$ and (A10), we take the Taylor expansion on $V^+(\widehat{\beta}^{\mathrm{c}})$ at

β^*,

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

(c2.) Next, recall the result (A.5) that

$$
\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-V^+(\beta)-\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^+(\beta^*)\right\}\right|\right]\le C_1^*\varepsilon^{1/2},
$$

where C_1^* is a finite constant. Since $\|\widehat{\beta}^{\mathrm{c}}-\beta^*\|_2=O_p(n^{-1/3})$, i.e., $\|\widehat{\beta}^{\mathrm{c}}-\beta^*\|_2=c_5n^{-1/3}$, where c_5 is a finite constant. We have

$$
\begin{aligned}
&\sqrt{n}\left[\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-\widehat{V}^{\mathrm{c}}(\beta^*)\right\}-\left\{V^+(\widehat{\beta}^{\mathrm{c}})-V^+(\beta^*)\right\}\right]\\
\le&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<c_5n^{-1/3}}\left|\widehat{V}^{\mathrm{c}}(\beta)-V^+(\beta)-\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^+(\beta^*)\right\}\right|\right]\\
\le&C_1^*\sqrt{c_5n^{-1/3}}=C_1^*\sqrt{c_5}n^{-1/6}=o_p(1).
\end{aligned}
\tag{A.7}
$$

(c3.) By (A.6) and (A.7), we have

$$
\begin{aligned}
&\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-\widehat{V}^{\mathrm{c}}(\beta^*)\right\}\\
=&\sqrt{n}\left[\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-\widehat{V}^{\mathrm{c}}(\beta^*)\right\}-\left\{V^+(\widehat{\beta}^{\mathrm{c}})-V^+(\beta^*)\right\}\right]+\sqrt{n}\left\{V^+(\widehat{\beta}^{\mathrm{c}})-V^+(\beta^*)\right\}\\
=&o_p(1)+o_p(1)=o_p(1).
\end{aligned}
\tag{A.8}
$$

(D.) Next, we only need to show the asymptotic distribution of $\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^+(\beta^*)\right\}$. By taking the Taylor expansion on $\widehat{V}^{\mathrm{c}}(\beta^*)$ at $(\lambda^*,\theta^*,\eta^*)$, we have

$$
\widehat{V}^{\mathrm{c}}(\beta^*)=V_n^*(\beta^*)+H_\lambda^{\mathrm{T}}(\widehat{\lambda}-\lambda^*)+H_\theta^{\mathrm{T}}(\widehat{\theta}-\theta^*)+H_\eta^{\mathrm{T}}(\widehat{\eta}-\eta^*)+o_p(n^{-1/2}),
$$

where

$$
H_\lambda=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial W(X_i;\lambda^*)}{\partial\lambda}\right\}\psi(Y_i,A_i,X_i;\beta^*,\theta^*,\eta^*),
$$

$$H_\theta = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} W(X_i;\lambda^*)\frac{\partial \psi(Y_i,A_i,X_i;\beta^*,\theta^*,\eta^*)}{\partial \theta},$$
$$H_\eta = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} W(X_i;\lambda^*)\frac{\partial \psi(Y_i,A_i,X_i;\beta^*,\theta^*,\eta^*)}{\partial \eta}.$$

Note that

$$\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \rho\left[\widehat{\lambda}^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\right]\{g(X_i)-\mu_{g0}\} \\ C(X_i,A_i,Y_i;\widehat{\theta}) \\ S(X_i,A_i;\widehat{\eta}) \end{pmatrix} = 0.$$

By taking the Taylor expansion on these equations at λ^*, θ^*, η^*, respectively, we have

$$\sqrt{n}(\widehat{\lambda}-\lambda^*) = G_\lambda^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\rho'\left[(\lambda^*)^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\right]\{g(X_i)-\mu_{g0}\} + o_p(1),$$
$$\sqrt{n}(\widehat{\theta}-\theta^*) = G_\theta^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} C(X_i,A_i,Y_i;\theta^*) + o_p(1),$$
$$\sqrt{n}(\widehat{\eta}-\eta^*) = G_\eta^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S(X_i,A_i;\eta^*) + o_p(1),$$

where

$$G_\lambda = -\mathbb{E}\left(\rho'\left[(\lambda^*)^{\mathrm{T}}\{g(X)-\mu_{g0}\}\right]\{g(X)-\mu_{g0}\}\{g(X)-\mu_{g0}\}^{\mathrm{T}}\right),$$
$$G_\theta = -\mathbb{E}\left\{\frac{\partial C(X,A,Y;\theta^*)}{\partial \theta^{\mathrm{T}}}\right\}, G_\eta = -\mathbb{E}\left\{\frac{\partial S(X,A;\eta^*)}{\partial \eta^{\mathrm{T}}}\right\}.$$

Thus,

$$\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\beta^*) - V^+(\beta^*)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_{i1}+\xi_{i2}+\xi_{i3}+\xi_{i4}) + o_p(1),$$

where

$$\xi_{i1} = W(X_i;\lambda^*)\psi(Y_i,A_i,X_i;\beta^*,\theta^*,\eta^*) - V^+(\beta^*),$$
$$\xi_{i2} = H_\lambda^{\mathrm{T}} G_\lambda^{-1}\rho\left[(\lambda^*)^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\right]\{g(X_i)-\mu_{g0}\},$$
$$\xi_{i3} = H_\theta^{\mathrm{T}} G_\theta^{-1} C(X_i,A_i,Y_i;\theta^*),$$
$$\xi_{i4} = H_\eta^{\mathrm{T}} G_\eta^{-1} S(X_i,A_i;\eta^*).$$

$\xi_{i1}, \xi_{i2}, \xi_{i3}$, and ξ_{i4} are i.i.d. mean zero variables. Then we have,

$$\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\beta^*) - V^+(\beta^*)\right\} \longrightarrow N(0,\sigma_1^2), \quad \text{in distribution,} \tag{A.9}$$

where $\sigma_1^2 = \mathbb{E}(\xi_{i1}+\xi_{i2}+\xi_{i3}+\xi_{i4})^2$.

By (A.8) and (A.9), we have

$$\begin{aligned}
&\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-V^{+}(\beta^*)\right\}=\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-\widehat{V}^{\mathrm{c}}(\beta^*)+\widehat{V}^{\mathrm{c}}(\beta^*)-V^{+}(\beta^*)\right\}\\
=&\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-\widehat{V}^{\mathrm{c}}(\beta^*)\right\}+\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^{+}(\beta^*)\right\}\\
=&o_p(1)+\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^{+}(\beta^*)\right\}\longrightarrow N(0,\sigma_1^2),\quad \text{in distribution.}
\end{aligned}$$

□

A.1.2 Proof of Theorem 2

The proof of Theorem 2 consists of three steps as follows.

Step 1: We show $\widehat{V}^{\mathrm{c}}(\beta)=\tilde{V}_n^*(\beta)+o_p(n^{-1/2})$, where

$$\tilde{V}_n^*(\beta)=\frac{1}{n}\sum_{i=1}^{n}W(X_i;\widehat{\lambda})\left[\frac{\mathbb{I}\{A_i=d(X_i;\beta)\}}{\pi(X_i)A_i+\{1-\pi(X_i)\}(1-A_i)}\{Y_i-\mu_d(X_i;\beta)\}+\mu_d(X_i;\beta)\right].$$

Define

$$\begin{aligned}
\psi(Y,A,X;\beta)&=\frac{\mathbb{I}\{A=d(X;\beta)\}}{\pi(X)A+\{1-\pi(X)\}(1-A)}\{Y-\mu_d(X;\beta)\}+\mu_d(X;\beta),\\
\widehat{\psi}(Y,A,X;\beta)&=\frac{\mathbb{I}\{A=d(X;\beta)\}}{\widehat{\pi}(X)A+\{1-\widehat{\pi}(X)\}(1-A)}\{Y-\widehat{\mu}_d(X;\beta)\}+\widehat{\mu}_d(X;\beta),\\
\tilde{\psi}(Y,A,X;\beta)&=\frac{\mathbb{I}\{A=d(X;\beta)\}}{\pi(X)A+\{1-\pi(X)\}(1-A)}\{Y-\widehat{\mu}_d(X;\beta)\}+\widehat{\mu}_d(X;\beta).
\end{aligned}$$

We assume $W(X;\lambda)\psi(Y,A,X;\beta)$ belongs to a Donsker class. By Theorem 2.1 in Van Der Vaart and Wellner (2007), we have

$$\widehat{V}^{\mathrm{c}}(\beta)-\tilde{V}_n^*(\beta)=P\{W(X;\widehat{\lambda})\widehat{\psi}(Y,A,X;\beta)-W(X;\widehat{\lambda})\psi(Y,A,X;\beta)\}+o_p(n^{-1/2}).$$

To show $\widehat{V}^{\mathrm{c}}(\beta)=\tilde{V}_n^*(\beta)+o_p(n^{-1/2})$, we only need to show

$$P\{W(X;\widehat{\lambda})\widehat{\psi}(Y,A,X;\beta)-W(X;\widehat{\lambda})\psi(Y,A,X;\beta)\}=o_p(n^{-1/2}).$$

Note that

$$\begin{aligned}
&P\{W(X;\widehat{\lambda})\widehat{\psi}(Y,A,X;\beta)-W(X;\widehat{\lambda})\psi(Y,A,X;\beta)\}\\
=&P\{W(X;\widehat{\lambda})\widehat{\psi}(Y,A,X;\beta)-W(X;\widehat{\lambda})\tilde{\psi}(Y,A,X;\beta)\}
\end{aligned}$$

$$+P\{W(X;\widehat{\lambda})\tilde{\psi}(Y,A,X;\beta)-W(X;\widehat{\lambda})\psi(Y,A,X;\beta)\}. \tag{A.10}$$

Define

$$\begin{aligned}\varrho(A\mid X)&=\pi(X)A+\{1-\pi(X)\}(1-A),\\ \widehat{\varrho}(A\mid X)&=\widehat{\pi}(X)A+\{1-\widehat{\pi}(X)\}(1-A).\end{aligned}$$

$$\begin{aligned}&P\{W(X;\widehat{\lambda})\widehat{\psi}(Y,A,X;\beta)-W(X;\widehat{\lambda})\tilde{\psi}(Y,A,X;\beta)\}\\ =&P\left(W(X;\widehat{\lambda})(2A-1)\{\pi(X)-\widehat{\pi}(X)\}\left[\frac{\mathbb{I}\{A=d(X;\beta)\}}{\widehat{\varrho}(A\mid X)\varrho(A\mid X)}\{Y-\mu_d(X,\beta)\}\right]\right)\\ &+P\left(W(X;\widehat{\lambda})(2A-1)\{\pi(X)-\widehat{\pi}(X)\}\left[\frac{\mathbb{I}\{A=d(X;\beta)\}}{\widehat{\varrho}(A\mid X)\varrho(A\mid X)}\{\mu_d(X;\beta)-\widehat{\mu}_d(X;\beta)\}\right]\right).\end{aligned} \tag{A.11}$$

The first term in (A.11) is 0 since we have

$$P[\mathbb{I}\{A=d(X;\beta)\}\{Y-\mu_d(X;\beta)\}]=0.$$

By Conditions (A3) and (A9), $W(X,\widehat{\lambda})$ and $\{\widehat{\varrho}(A\mid X)\varrho(A\mid X)\}^{-1}$ are bounded. By Cauchy-Schwarz inequality, the absolute value of the second term in (A.11) is bounded by

$$\left[P\{\widehat{\pi}(X)-\pi(X)\}^2\right]^{\frac{1}{2}}\sum_{a=0}^{1}\left[P\{\widehat{\mu}(X,a)-\mu(X,a)\}^2\right]^{\frac{1}{2}}=o_p(n^{-1/2}),$$

under (A11'). Therefore,

$$P\{W(X;\widehat{\lambda})\widehat{\psi}(Y,A,X;\beta)-W(X;\widehat{\lambda})\tilde{\psi}(Y,A,X;\beta)\}=o_p(n^{-1/2}). \tag{A.12}$$

Note that

$$P\left[\frac{\mathbb{I}\{A=d(X;\beta)\}}{\varrho(A\mid X)}\right]=1.$$

Thus, we have

$$\begin{aligned}&P\{W(X;\widehat{\lambda})\tilde{\psi}(Y,A,X;\beta)-W(X;\widehat{\lambda})\psi(Y,A,X;\beta)\}\\ =&P\left(W(X;\widehat{\lambda})\left[\frac{\mathbb{I}\{A=d(X;\beta)\}}{\varrho(A\mid X)}-1\right]\{\mu_d(X;\beta)-\widehat{\mu}_d(X;\beta)\}\right)\\ =&0.\end{aligned} \tag{A.13}$$

By (A.10), (A.12), and (A.13), we have

$$P\{W(X;\widehat{\lambda})\widehat{\psi}(Y,A,X;\beta)-W(X;\widehat{\lambda})\psi(Y,A,X;\beta)\}=o_p(n^{-1/2}),$$

and then

$$\widehat{V}^{\mathrm{c}}(\beta)=\tilde{V}_n^*(\beta)+o_p(n^{-1/2}). \tag{A.14}$$

Step 2: We show that $n^{1/3}\|\widehat{\beta}^{\mathrm{c}}-\beta^*\|_2=O_p(1)$.

(A.) First, we show that $\widehat{\beta}^{\mathrm{c}}$ converges in probability to β^* as $n\to\infty$, by checking three conditions for the Argmax Theorem:

(a1.) By Condition (A9), the true value function $V^+(\beta)$ is twice continuously differentiable in a neighborhood of β^*.

(a2.) By Condition (A11'), it can be easily shown

$$\widehat{V}^{\mathrm{c}}(\beta)=V^+(\beta)+o_p(1).$$

(a3.) Since $\widehat{\beta}^{\mathrm{c}}=\underset{\beta:\|\beta\|_2=1}{\operatorname{argmax}}\,\widehat{V}^{\mathrm{c}}(\beta)$, we have the estimated ITR as $d(X,\widehat{\beta}^{\mathrm{c}})=\mathbb{I}(\tilde{X}^{\mathrm{T}}\widehat{\beta}^{\mathrm{c}}>0)$ and the corresponding value function $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$ such that

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

Thus we have $\widehat{\beta}^{\mathrm{c}}\to\beta^*$, in probability, as $n\to\infty$.

(B). Next, we show that the convergence rate of $\widehat{\beta}^{\mathrm{c}}$ is $n^{1/3}$, i.e., $n^{1/3}\|\widehat{\beta}-\beta^*\|_2=O_p(1)$. We check three conditions of the Theorem 14.4: Rate of convergence in Kosorok (2008):

(b1.) For every β in a neighborhood of β^*, i.e., $\|\beta-\beta^*\|_2<\varepsilon$ for some $\varepsilon>0$, by Condition (A9), we take the second order Taylor expansion on $V^+(\beta)$ at $\beta=\beta^*$,

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

Since $(V^+)''(\beta^*)<0$, there exists $c_0=-\frac{1}{2}(V^+)''(\beta^*)>0$ such that $V^+(\beta)-V^+(\beta^*)<c_0\|\beta-\beta^*\|_2^2$ holds.

(b2.) For all n large enough and sufficiently small ε, the centered process $\widehat{V}^{\mathrm{c}}-V^+$ satisfies

$$
\begin{aligned}
&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-V^+(\beta)-\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^+(\beta^*)\}\right|\right]\\
=&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-\tilde{V}_n^*(\beta)+\tilde{V}_n^*(\beta)-V^+(\beta)-\{\widehat{V}^{\mathrm{c}}(\beta^*)-\tilde{V}_n^*(\beta^*)+\tilde{V}_n^*(\beta^*)-V^+(\beta^*)\}\right|\right]\\
\le&\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-\tilde{V}_n^*(\beta)-\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-\tilde{V}_n^*(\beta^*)\right\}\right|\right]}_{\zeta_1}\\
&+\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\tilde{V}_n^*(\beta)-V^+(\beta)-\left\{\tilde{V}_n^*(\beta^*)-V^+(\beta^*)\right\}\right|\right]}_{\zeta_2}. \qquad \text{(A.15)}
\end{aligned}
$$

By the result (A.14), $\widehat{V}^{\mathrm{c}}(\beta)-\tilde{V}_n^*(\beta)=o_p(n^{-1/2})$ and $\widehat{V}^{\mathrm{c}}(\beta^*)-\tilde{V}_n^*(\beta^*)=o_p(n^{-1/2})$. Thus,

$$
\zeta_1=\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-\tilde{V}_n^*(\beta)-\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-\tilde{V}_n^*(\beta^*)\right\}\right|\right]=o_p(1) \qquad \text{(A.16)}
$$

Define

$$
\bar{V}_n^*(\beta)=\frac{1}{n}\sum_{i=1}^n W(X_i;\lambda^*)\left\{\frac{\mathbb{I}\{A_i=d(X_i;\beta)\}}{\pi(X_i)A_i+\{1-\pi(X_i)\}(1-A_i)}\{Y_i-\mu_d(X_i;\beta)\}+\mu_d(X_i;\beta)\right\}.
$$

$$
\begin{aligned}
\zeta_2=&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\tilde{V}_n^*(\beta)-V^+(\beta)-\left\{\tilde{V}_n^*(\beta^*)-V^+(\beta^*)\right\}\right|\right]\\
=&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\tilde{V}_n^*(\beta)-\bar{V}_n^*(\beta)+\bar{V}_n^*(\beta)-V^+(\beta)-\left\{\tilde{V}_n^*(\beta^*)-\bar{V}_n^*(\beta^*)+\bar{V}_n^*(\beta^*)-V^+(\beta^*)\right\}\right|\right]\\
\le&\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\tilde{V}_n^*(\beta)-\bar{V}_n^*(\beta)-\left\{\tilde{V}_n^*(\beta^*)-\bar{V}_n^*(\beta^*)\right\}\right|\right]}_{\omega_1}\\
&+\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\bar{V}_n^*(\beta)-V^+(\beta)-\left\{\bar{V}_n^*(\beta^*)-V^+(\beta^*)\right\}\right|\right]}_{\omega_2}.
\end{aligned}
$$

We have

$$
\begin{aligned}
&\bar{V}_n^*(\beta)-\bar{V}_n^*(\beta^*)\\
=&\frac{1}{n}\sum_{i=1}^n W(\lambda^*;X)\left[\frac{\mathbb{I}\{A_i=d(X_i;\beta)\}\{Y_i-\mu_d(X_i;\beta)\}-\mathbb{I}\{A_i=d(X_i;\beta^*)\}\{Y_i-\mu_d(X_i;\beta^*)\}}{\pi(X_i)A_i+\{1-\pi(X_i)\}(1-A_i)}\right.
\end{aligned}
$$

$$
\begin{aligned}
&+\mu_d(X_i;\beta)-\mu_d(X_i;\beta^*)\Big] \\
=&\frac{1}{n}\sum_{i=1}^{n} W(\lambda^*;X_i)\Big[\frac{(2A_i-1)Y_i-\mu(X_i,1)A_i+\mu(X_i,0)(1-A_i)}{\pi(X_i)A_i+\{1-\pi(X_i)\}(1-A_i)}+\mu(X_i,1)-\mu(X_i,0)\Big] \\
&\big\{\mathbb{I}(\tilde{X}_i^{\mathrm{T}}\beta>0)-\mathbb{I}(\tilde{X}_i^{\mathrm{T}}\beta^*>0)\big\}.
\end{aligned}
$$

We define a class of functions

$$
\begin{aligned}
\mathscr{F}_\beta^1(y,a,x)=\Big\{ & W(\lambda^*;x)\Big[\frac{(2a-1)y-\mu(x,1)a+\mu(x,0)(1-a)}{\pi(x)a+\{1-\pi(x)\}(1-a)}+\mu(x,1)-\mu(x,0)\Big] \\
& \big\{\mathbb{I}\big(\tilde{x}^{\mathrm{T}}\beta>0\big)-\mathbb{I}\big(\tilde{x}^{\mathrm{T}}\beta^*>0\big)\big\}:\|\beta-\beta^*\|_2<\varepsilon\Big\}.
\end{aligned}
$$

Let $K_1=\sup\Big|W(\lambda^*;x)\Big[\frac{(2a-1)y-\mu(x,1)a+\mu(x,0)(1-a)}{\pi(x)a+\{1-\pi(x)\}(1-a)}+\mu(x,1)-\mu(x,0)\Big]\Big|$. By Conditions (A3), (A6), (A7), and (A8), $K_1<\infty$. Then we can define the envelope of $\mathscr{F}_\beta^1(y,a,x)$ as $F_1=K_1\mathbb{I}\big(-k_0\varepsilon\le\tilde{x}^{\mathrm{T}}\beta^*\le k_0\varepsilon\big)$. Similar to (b2.1) in the proof of Theorem 1, we have

$$
\|F_1\|_{P,2}=K_1\sqrt{\Pr\big(-k_0\varepsilon\le\tilde{x}^{\mathrm{T}}\beta^*\le k_0\varepsilon\big)}=K_1\sqrt{k_1\cdot 2k_0\varepsilon}=K_1\sqrt{2k_0k_1}\varepsilon^{1/2}<\infty.
$$

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

$$
\mathbb{G}_n\mathscr{F}_\beta^1=n^{-1/2}\sum_{i=1}^{n}\Big[\mathscr{F}_\beta^1(Y_i,A_i,X_i)-\mathbb{E}\big\{\mathscr{F}_\beta^1(Y_i,A_i,X_i)\big\}\Big].
$$

Note that $\mathbb{G}_n\mathscr{F}_\beta^1=n^{1/2}\big[\bar{V}_n^*(\beta)-\bar{V}_n^*(\beta^*)-\big\{V^+(\beta)-V^+(\beta^*)\big\}\big]$. By applying Theorem 11.2 in Kosorok (2008), we have

$$
\begin{aligned}
\omega_2&=\mathbb{E}^*\Big[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\Big|\bar{V}_n^*(\beta)-V^+(\beta)-\big\{\bar{V}_n^*(\beta^*)-V^+(\beta^*)\big\}\Big|\Big] \\
&=\mathbb{E}^*\Big(\sup_{\|\beta-\beta^*\|_2<\varepsilon}\Big|\mathbb{G}_n\mathscr{F}_\beta^1\Big|\Big)\le c_1J_{[]}^*(1,\mathscr{F}^1)\|F_1\|_{P,2}=c_1J_{[]}^*(1,\mathscr{F}^1)K_1\sqrt{2k_0k_1}\varepsilon^{1/2},
\end{aligned}
$$

where c_1 is a finite constant.

Let $C_1^*\equiv c_1J_{[]}^*(1,\mathscr{F}^1)K_1\sqrt{2k_0k_1}$, since c_1, $J_{[]}^*(1,\mathscr{F}^1)$, K_1, k_0 and k_1 are bounded, we have $C_1^*<\infty$, i.e.,

$$
\omega_2\le C_1^*\varepsilon^{1/2}. \tag{A.17}
$$

We rewrite the form of $\tilde{V}_n^*(\beta)-\bar{V}_n^*(\beta)-\{\tilde{V}_n^*(\beta^*)-\bar{V}_n^*(\beta^*)\}$ as

$$
\begin{aligned}
&\tilde{V}_n^*(\beta)-\bar{V}_n^*(\beta)-\{\tilde{V}_n^*(\beta^*)-\bar{V}_n^*(\beta^*)\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\{W(X_i;\widehat{\lambda})-W(X_i;\lambda^*)\}\\
&\left[\frac{\mathbb{I}\{A_i=d(X_i;\beta)\}\{Y_i-\mu_d(X_i;\beta)\}-\mathbb{I}\{A_i=d(X_i;\beta^*)\}\{Y_i-\mu_d(X_i;\beta^*)\}}{\pi(X_i)A_i+\{1-\pi(X_i)\}(1-A_i)}+\mu_d(X_i;\beta)-\mu_d(X_i;\beta^*)\right]\\
=&\frac{1}{n}\sum_{i=1}^{n}\{W(X_i;\widehat{\lambda})-W(X_i;\lambda^*)\}\left[\frac{(2A_i-1)Y_i-\mu(X_i,1)A_i+\mu(X_i,0)(1-A_i)}{\pi(X_i)A_i+\{1-\pi(X_i)\}(1-A_i)}+\mu(X_i,1)-\mu(X_i,0)\right]\\
&\{\mathbb{I}(\tilde{X}_i^{\mathrm{T}}\beta>0)-\mathbb{I}(\tilde{X}_i^{\mathrm{T}}\beta^*>0)\}.
\end{aligned}
$$

We take the Taylor expansion on the above equation at λ^*,

$$
\begin{aligned}
&\tilde{V}_n^*(\beta)-\bar{V}_n^*(\beta)-\{\tilde{V}_n^*(\beta^*)-\bar{V}_n^*(\beta^*)\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial W(X;\lambda^*)}{\partial\lambda}\right\}^{\mathrm{T}}\left[\frac{(2A_i-1)Y_i-\mu(X_i,1)A_i+\mu(X_i,0)(1-A_i)}{\pi(X_i)A_i+\{1-\pi(X_i)\}(1-A_i)}+\mu(X_i,1)-\mu(X_i,0)\right]\\
&\{\mathbb{I}(\tilde{X}_i^{\mathrm{T}}\beta>0)-\mathbb{I}(\tilde{X}_i^{\mathrm{T}}\beta^*>0)\}\left(\widehat{\lambda}-\lambda^*\right)+o_p(n^{-1/2}).
\end{aligned}
$$

Next, we define a class of functions,

$$
\begin{aligned}
\mathscr{F}_\beta^2(y,a,x)=\Bigg\{&\frac{\partial W(X;\lambda^*)}{\partial\lambda}\left[\frac{(2a-1)y-\mu(x,1)a+\mu(x,0)(1-a)}{\pi(x)a+\{1-\pi(x)\}(1-a)}+\mu(x,1)-\mu(x,0)\right]\\
&\{\mathbb{I}\left(\tilde{x}^{\mathrm{T}}\beta>0\right)-\mathbb{I}\left(\tilde{x}^{\mathrm{T}}\beta^*>0\right)\}:\|\beta-\beta^*\|_2<\varepsilon\Bigg\}.
\end{aligned}
$$

Let $K_2=\sup\left|\frac{\partial W(X;\lambda^*)}{\partial\lambda}\left[\frac{(2a-1)y-\mu(x,1)a+\mu(x,0)(1-a)}{\pi(x)a+\{1-\pi(x)\}(1-a)}+\mu(x,1)-\mu(x,0)\right]\right|$. By Conditions (A3), (A6), (A7) and (A8), $K_2<\infty$. Then we can define the envelope of $\mathscr{F}_\beta^2(y,a,x)$ as $F_2=K_2\mathbb{I}\left(-k_0\varepsilon\le\tilde{x}^{\mathrm{T}}\beta^*\le k_0\varepsilon\right)$. Similarly to above, we have

$$\|F_2\|_{P,2}\le K_2\sqrt{2k_0k_1}\varepsilon^{1/2}<\infty,$$

and $\mathscr{F}_\beta^2$ is a VC class of functions. Thus, the entropy of $\mathscr{F}_\beta^2$, denoted as $J_{[]}^*(1,\mathscr{F}^2)$, is finite, i.e., $J_{[]}^*(1,\mathscr{F}^2)<\infty$.

Next, we consider the following empirical process indexed by β,

$$\mathbb{G}_n\mathscr{F}_\beta^2=n^{-1/2}\sum_{i=1}^{n}\left[\mathscr{F}_\beta^2(Y_i,A_i,X_i)-\mathbb{E}\left\{\mathscr{F}_\beta^2(Y_i,A_i,X_i)\right\}\right].$$

By applying Theorem 11.2 in Kosorok (2008), we have

$$\mathbb{E}^*\left(n^{1/2} \sup_{\|\beta-\beta^*\|_2<\varepsilon} \left\| \mathbb{G}_n \mathscr{F}_\beta^2 \right\|_1 \right) \leq c_2 J_{[]}^*(1,\mathscr{F}^2) \|F_2\|_{P,2} \leq c_2 J_{[]}^*(1,\mathscr{F}^2) K_2 \sqrt{2k_0 k_1(\beta^*)}\varepsilon^{1/2},$$

where c_2 is a finite constant. By Theorem 2.14.5 in Van Der Vaart and Wellner (1996),

$$\begin{aligned}
\left\{ \mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon} \left\| \mathbb{G}_n \mathscr{F}_\beta^2 \right\|_2^2 \right) \right\}^{1/2} &\leq l_2 \left\{ \mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon} \left\| \mathbb{G}_n \mathscr{F}_\beta^2 \right\|_1 \right) + \|F_2\|_{P,2} \right\} \\
&\leq l_2 \left\{ c_2 J_{[]}^*(1,\mathscr{F}^2) + 1 \right\} \|F_2\|_{P,2} \\
&= l_2 \left\{ c_2 J_{[]}^*(1,\mathscr{F}^2) + 1 \right\} K_2 \sqrt{2k_0 k_1}\varepsilon^{1/2},
\end{aligned}$$

where l_2 is a finite constant. Let

$$C_2^* \equiv l_2 \left\{ c_2 J_{[]}^*(1,\mathscr{F}^2) + 1 \right\} K_2 \sqrt{2k_0 k_1} < \infty,$$

and we have

$$\left\{ \mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon} \left\| \mathbb{G}_n \mathscr{F}_\beta^2 \right\|_2^2 \right) \right\}^{1/2} \leq C_2^* \varepsilon^{1/2}.$$

Notice that

$$\begin{aligned}
\omega_1 &= \mathbb{E}^*\left[n^{1/2} \sup_{\|\beta-\beta^*\|_2<\varepsilon} \left| \tilde{V}_n^*(\beta) - \bar{V}_n^*(\beta) - \left\{ \tilde{V}_n^*(\beta^*) - \bar{V}_n^*(\beta^*) \right\} \right| \right] \\
&= \mathbb{E}^*\left\{ \sup_{\|\beta-\beta^*\|_2<\varepsilon} (\widehat{\lambda}-\lambda^*)^{\mathrm{T}} \mathbb{G}_n \mathscr{F}_\beta^2 + o_p(1) \right\} \\
&\leq n^{-1/2}\left[\mathbb{E}^*\left\{ \sup_{\|\beta-\beta^*\|_2<\varepsilon} \left| n^{1/2}(\widehat{\lambda}-\lambda^*)^{\mathrm{T}} \mathbb{G}_n \mathscr{F}_\beta^2 \right| \right\} \right].
\end{aligned}$$

By the Cauchy-Schwarz Inequality, we have

$$\omega_1 \leq n^{-1/2} \left\{ \mathbb{E}\left(n\|\widehat{\lambda}-\lambda^*\|_2^2 \right) \right\}^{1/2} \left\{ \mathbb{E}^*\left(\sup_{\|\beta-\beta^*\|_2<\varepsilon} \left\| \mathbb{G}_n \mathscr{F}_\beta^2 \right\|_2^2 \right) \right\}^{1/2}.$$

By Condition (A11'), $\sqrt{n}(\widehat{\lambda}-\lambda^*) = O_p(1)$, we have $M_\lambda \equiv \left\{ \mathbb{E}\left(n\|\widehat{\lambda}-\lambda^*\|_2^2 \right) \right\}^{1/2} < \infty$, then

$$\omega_1 \leq n^{-1/2} M_\lambda C_2^* \varepsilon^{1/2}. \tag{A.18}$$

By (A.17) and (A.18), we have

$$\zeta_2 \le \omega_1 + \omega_2 \le C_1^* \varepsilon^{1/2} + n^{-1/2} M_\lambda C_2^* \varepsilon^{1/2}. \tag{A.19}$$

By (A.15), (A.16), and (A.19), as $n \to \infty$, we have the centered process $\widehat{V}^{\mathrm{c}} - V^+$ satisfies

$$\mathbb{E}^* \left[n^{1/2} \sup_{\|\beta-\beta^*\|_2<\varepsilon} \left| \widehat{V}^{\mathrm{c}}(\beta) - V^+(\beta) - \left\{ \widehat{V}^{\mathrm{c}}(\beta^*) - V^+(\beta^*) \right\} \right| \right] \le \zeta_1 + \zeta_2 \le C_1^* \varepsilon^{1/2}. \tag{A.20}$$

Let $\phi_n(\varepsilon) = \varepsilon^{1/2}$, and $\alpha = \frac{3}{2} < 2$, check $\frac{\phi_n(\varepsilon)}{\varepsilon^\alpha} = \varepsilon^{-1}$ is decreasing not depending on n. Therefore, the second condition holds.

(b3.) By $\widehat{\beta}^{\mathrm{c}} \to \beta^*$, in probability, as $n \to \infty$, and $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}}) \ge \sup_{\beta:\|\beta\|_2=1} \widehat{V}^{\mathrm{c}}(\beta)$ shown previously, choose $r_n = n^{1/3}$, then r_n satisfies

$$\begin{aligned} r_n^2 \phi_n(r_n^{-1}) &= n^{2/3} \phi_n(n^{-1/3}) \\ &= n^{2/3} (n^{-1/3})^{1/2} = n^{1/2}. \end{aligned}$$

Thus, the third condition holds.

By the Theorem 14.4 in Kosorok (2008), we have $n^{1/3} \|\widehat{\beta}^{\mathrm{c}} - \beta^*\|_2 = O_p(1)$.

Step 3: We derive the asymptotic distribution of $\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})$ in this step. Note that

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

(C.) First, we show

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

which is sufficient to show $\sqrt{n} \left[\left\{ \widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}}) - \widehat{V}^{\mathrm{c}}(\beta^*) \right\} - \left\{ V^+(\widehat{\beta}^{\mathrm{c}}) - V^+(\beta^*) \right\} \right] = o_p(1)$ and $\sqrt{n} \left\{ V^+(\widehat{\beta}^{\mathrm{c}}) - V^+(\beta^*) \right\} = o_p(1)$.

(c1.) First, by $n^{1/3} \|\widehat{\beta}^{\mathrm{c}} - \beta^*\|_2 = O_p(1)$ and (A9), we take the Taylor expansion on $V^+(\widehat{\beta}^{\mathrm{c}})$ at β^*,

$$\begin{aligned} \sqrt{n} \left\{ V^+(\widehat{\beta}^{\mathrm{c}}) - V^+(\beta^*) \right\} &= \sqrt{n} \left\{ (V^+)'(\beta^*) \|\widehat{\beta}^{\mathrm{c}} - \beta^*\|_2 + \frac{1}{2} (V^+)''(\beta^*) \|\widehat{\beta}^{\mathrm{c}} - \beta^*\|_2^2 + o\left(\|\widehat{\beta}^{\mathrm{c}} - \beta^*\|_2^2 \right) \right\} \\ &= \sqrt{n} \left\{ \frac{1}{2} (V^+)''(\beta^*) \|\widehat{\beta}^{\mathrm{c}} - \beta^*\|_2^2 + o\left(\|\widehat{\beta}^{\mathrm{c}} - \beta^*\|_2^2 \right) \right\} \ \text{(by } (V^+)'(\beta^*) = 0) \\ &= \sqrt{n} \left\{ \frac{1}{2} (V^+)''(\beta^*) O_p(n^{-2/3}) + o_p(n^{-2/3}) \right\} \\ &= \frac{1}{2} (V^+)''(\beta^*) O_p(n^{-1/6}) = o_p(1). \end{aligned} \tag{A.21}$$

(c2.) Next, recall the result (A.20) that

$$\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{c}}(\beta)-V^+(\beta)-\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^+(\beta^*)\right\}\right|\right]\leq C_1^*\varepsilon^{1/2},$$

where C_1^* is a finite constant. Since $\|\widehat{\beta}^{\mathrm{c}}-\beta^*\|_2=O_p(n^{-1/3})$, i.e., $\|\widehat{\beta}^{\mathrm{c}}-\beta^*\|_2=c_3n^{-1/3}$, where c_3 is a finite constant. We have

$$\begin{aligned}
&\sqrt{n}\left[\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-\widehat{V}^{\mathrm{c}}(\beta^*)\right\}-\left\{V^+(\widehat{\beta}^{\mathrm{c}})-V^+(\beta^*)\right\}\right]\\
\leq&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<c_4n^{-1/3}}\left|\widehat{V}^{\mathrm{c}}(\beta)-V^+(\beta)-\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-V^+(\beta^*)\right\}\right|\right]\\
\leq&C_1^*\sqrt{c_3n^{-1/3}}=C_1^*\sqrt{c_3}n^{-1/6}=o_p(1).
\end{aligned}\tag{A.22}$$

(c3.) By (A.21) and (A.22), we have

$$\begin{aligned}
&\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-\widehat{V}^{\mathrm{c}}(\beta^*)\right\}\\
=&\sqrt{n}\left[\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-\widehat{V}^{\mathrm{c}}(\beta^*)\right\}-\left\{V^+(\widehat{\beta}^{\mathrm{c}})-V^+(\beta^*)\right\}\right]+\sqrt{n}\left\{V^+(\widehat{\beta}^{\mathrm{c}})-V^+(\beta^*)\right\}\\
=&o_p(1)+o_p(1)=o_p(1).
\end{aligned}\tag{A.23}$$

(D.) By the result (A.14) shown in Step 1, $\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-\tilde{V}_n^*(\beta^*)\right\}=o_p(1)$. With (A.23), we have

$$\begin{aligned}
&\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-V^+(\beta^*)\right\}\\
=&\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-\widehat{V}^{\mathrm{c}}(\beta^*)\right\}+\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\beta^*)-\tilde{V}_n^*(\beta^*)\right\}+\sqrt{n}\left\{\tilde{V}_n^*(\beta^*)-V^+(\beta^*)\right\}\\
=&o_p(1)+o_p(1)+\sqrt{n}\left\{\tilde{V}_n^*(\beta^*)-V^+(\beta^*)\right\}\\
=&\sqrt{n}\left\{\tilde{V}_n^*(\beta^*)-V^+(\beta^*)\right\}+o_p(1).
\end{aligned}\tag{A.24}$$

Thus, we only need to show the asymptotic distribution of $\sqrt{n}\left\{\tilde{V}_n^*(\beta^*)-V^+(\beta^*)\right\}$. By taking the Taylor expansion on $\tilde{V}_n^*(\beta^*)$ at λ^*, we have

$$\tilde{V}_n^*(\beta^*)=\tilde{V}_n^*(\beta^*)+H_\lambda^{\mathrm{T}}(\widehat{\lambda}-\lambda^*)+o_p(n^{-1/2}),$$

where

$$H_\lambda=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left\{\frac{\partial W(X_i;\lambda^*)}{\partial\lambda}\right\}\psi(Y_i,A_i,X_i;\beta^*).$$

Note that

$$\frac{1}{n}\sum_{i=1}^n\rho\left[\widehat{\lambda}^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\right]\{g(X_i)-\mu_{g0}\}=0.$$

By taking the Taylor expansion on the above equation at λ^*, we have

$$\sqrt{n}(\widehat{\lambda}-\lambda^*)=G_\lambda^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\rho\left[(\lambda^*)^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\right]\{g(X_i)-\mu_{g0}\}+o_p(1),$$

where

$$G_\lambda=-\mathbb{E}\left(\rho'\left[(\lambda^*)^{\mathrm{T}}\{g(X)-\mu_{g0}\}\right]\{g(X)-\mu_{g0}\}\{g(X)-\mu_{g0}\}^{\mathrm{T}}\right).$$

Thus,

$$\sqrt{n}\left\{\tilde{V}_n^*(\beta^*)-V^+(\beta^*)\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_{i1}+\xi_{i2})+o_p(1),$$

where

$$\begin{aligned}\xi_{i1}&=W(X_i;\lambda^*)\psi(Y_i,A_i,X_i;\beta^*)-V^+(\beta^*),\\ \xi_{i2}&=H_\lambda^{\mathrm{T}}G_\lambda^{-1}\rho'\left[(\lambda^*)^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\right]\{g(X_i)-\mu_{g0}\}.\end{aligned}$$

ξ_{i1} and ξ_{i2} are i.i.d. mean-zero random variables. Then,

$$\sqrt{n}\left\{\tilde{V}_n^*(\beta^*)-V^+(\beta^*)\right\}\longrightarrow N(0,\sigma_2^2),\quad\text{in distribution,}\tag{A.25}$$

where $\sigma_2^2=\mathbb{E}\left\{(\xi_{i1}+\xi_{i2})^2\right\}$. By (A.24) and (A.25), we have

$$\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\widehat{\beta}^{\mathrm{c}})-V^+(\beta^*)\right\}=\sqrt{n}\left\{\tilde{V}_n^*(\beta^*)-V^+(\beta^*)\right\}+o_p(1)\longrightarrow N(0,\sigma_2^2),\quad\text{in distribution.}$$

□

A.1.3 Proof of Theorem 3

When $\mathbb{P}^{\mathrm{s}}=\mathbb{P}^{\mathrm{t}}$, $f^{\mathrm{s}}(X)=f^{\mathrm{t}}(X)$, $\widehat{\lambda}\xrightarrow{p}\lambda^*=0$ and $W(X;\lambda^*)=W(X;0)=1=f^{\mathrm{t}}(X)/f^{\mathrm{s}}(X)$. Thus, $\mathbb{P}^+=\mathbb{P}^{\mathrm{s}}=\mathbb{P}^{\mathrm{t}}$. Therefore, for a fixed ITR $d(X;\beta)$, $V^+(\beta)=V^{\mathrm{t}}(\beta)$. Meanwhile, we have $\mathbb{E}\left\{g(X)\right\}=\mathbb{E}^{\mathrm{t}}\left\{g(X)\right\}=\mu_{g0}$. By Theorem 2, we have

$$\sqrt{n}\left\{\widehat{V}^{\mathrm{c}}(\beta)-V^{\mathrm{t}}(\beta)\right\}\longrightarrow N(0,(\sigma^{\mathrm{c}})^2),\quad\text{in distribution,}$$

where

$$\begin{aligned}(\sigma^{\mathrm{c}})^2&=\mathbb{E}\left\{(\xi_1+\xi_2)^2\right\},\\ \xi_1&=W(X;\lambda^*)\psi(Y,A,X;\beta)-V^+(\beta)=\psi(Y,A,X;\beta)-V^+(\beta),\\ \xi_2&=H_\lambda^{\mathrm{T}}G_\lambda^{-1}\rho\left[(\lambda^*)^{\mathrm{T}}\{g(X)-\mu_{g0}\}\right]\{g(X)-\mu_{g0}\}=H_\lambda^{\mathrm{T}}G_\lambda^{-1}\{g(X)-\mu_{g0}\}\quad(\rho(0)=1),\end{aligned}$$

$$H_\lambda = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial W(X_i;\lambda^*)}{\partial\lambda}\right\}\psi(Y_i,A_i,X_i;\beta^*),$$

$$\begin{aligned}
G_\lambda &= -\mathbb{E}\big(\rho'\big[(\lambda^*)^{\mathrm{T}}\{g(X)-\mu_{g0}\}\big]\{g(X)-\mu_{g0}\}\{g(X)-\mu_{g0}\}^{\mathrm{T}}\big)\\
&= \mathbb{E}\big[\{g(X)-\mu_{g0}\}\{g(X)-\mu_{g0}\}^{\mathrm{T}}\big] \quad (\rho'(0)=1)\\
&= \mathrm{Var}\{(g(X)\}.
\end{aligned}$$

Notice

$$\begin{aligned}
\frac{\partial W(X_i;\lambda^*)}{\partial\lambda} &= n\Bigg\{\frac{\rho'\big[(\lambda^*)^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\big]\{g(X_i)-\mu_{g0}\}\big(\sum_{i=1}^{n}\rho\big[(\lambda^*)^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\big]\big)}{\big(\sum_{i=1}^{n}\rho\big[(\lambda^*)^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\big]\big)^2}\\
&\qquad -\frac{\rho\big[(\lambda^*)^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\big]\big(\sum_{i=1}^{n}\rho'\big[(\lambda^*)^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\big]\{g(X_i)-\mu_{g0}\}\big)}{\big(\sum_{i=1}^{n}\rho\big[(\lambda^*)^{\mathrm{T}}\{g(X_i)-\mu_{g0}\}\big]\big)^2}\Bigg\}\\
&= \{g(X_i)-\mu_{g0}\}-\frac{\sum_{i=1}^{n}\{g(X_i)-\mu_{g0}\}}{n}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
H_\lambda &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial W(X_i;\lambda^*)}{\partial\lambda}\right\}\psi(Y_i,A_i,X_i;\beta^*)\\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[\{g(X_i)-\mu_{g0}\}-\frac{\sum_{i=1}^{n}\{g(X_i)-\mu_{g0}\}}{n}\right]\psi(Y_i,A_i,X_i;\beta^*)\\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\big\{g(X_i)-\mu_{g0}\big\}\psi(Y_i,A_i,X_i;\beta^*)\\
&\quad -\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{i=1}^{n}\{g(X_i)-\mu_{g0}\}}{n}\right]\psi(Y_i,A_i,X_i;\beta^*)\\
&= \mathbb{E}\big[\big\{g(X_i)-\mu_{g0}\big\}\psi(Y_i,A_i,X_i;\beta^*)\big].
\end{aligned}$$

The asymptotic variance of the original AIPW estimator is $\sigma^2=\mathbb{E}(\xi_1^2)$.

$$\begin{aligned}
(\sigma^{\mathrm{c}})^2-\sigma^2 &= \mathbb{E}\left\{(\xi_1+\xi_2)^2\right\}-\mathbb{E}(\xi_1^2)\\
&= \mathbb{E}\big(2\xi_1\xi_2+\xi_2^2\big)\\
&= -2H_\lambda^{\mathrm{T}}G_\lambda^{-1}\mathbb{E}\big[\big\{\psi(Y,A,X;\beta)\big\}\{g(X)-\mu_{g0}\}\big]+2H_\lambda^{\mathrm{T}}G_\lambda^{-1}V^{+}(\beta)\mathbb{E}\{g(X)-\mu_{g0}\}\\
&\quad +H_\lambda^{\mathrm{T}}G_\lambda^{-1}\mathrm{Var}\{g(X)\}G_\lambda^{-1}H_\lambda\\
&= -H_\lambda^{\mathrm{T}}G_\lambda^{-1}H_\lambda.
\end{aligned}$$

Since $G_\lambda = \text{Var}\{(g(X)\}$ is positive semidefinite, $(\sigma^{\text{c}})^2 - \sigma^2 = -H_\lambda^{\text{T}} G_\lambda^{-1} H_\lambda \leq 0$. Therefore, the calibrated AIPW estimator has the same or smaller asymptotic variance than the original AIPW estimator. □

A.2 Additional Simulation Results

We implemented the $CR(1)$ method under the simulation setting in Section 2.5. We call the $CR(1)$ method as the least squares method since minimizing $CR(1)$ is equivalent to minimizing the sum of squares $\sum_{i=1}^{n}(w_i - n^{-1})^2$. In Scenarios 3 and 4, the calibration weights obtained by the least squares method can be negative, and thus the pseudo populations are no longer well-defined. Therefore, for a given linear ITR $d(X;\beta)$, we define the value function for the least squares method as $V_{LS}^{+}(\beta) = \mathbb{E}\{W^*(X;\lambda^*)Y(d(X;\beta))\}$. Then, the true optimal linear ITR is $d(X;\beta_{LS}^*)$, where $\beta_{LS}^* = \text{argmax}_\beta\, V_{LS}^{+}(\beta)$. For a given linear ITR $d(X;\beta)$, we denote the calibrated value function estimator using the least squares method as $\widehat{V}_{LS}^{\text{c}}(\beta)$. The estimated optimal linear ITR is $d(X;\widehat{\beta}_{LS}^{\text{c}})$, where $\widehat{\beta}_{LS}^{\text{c}} = \text{argmax}_\beta\, \widehat{V}_{LS}^{\text{c}}(\beta)$. We report the simulation results of the least squares method together with the other methods for comparison.

A.2.1 Value and Percentage of Correct Decisions Results of Estimated Optimal ITRs

The value and percentage of correct decisions results of estimated optimal ITRs obtained by different methods for the randomization study are summarized in Figure A.1 (method I) and Figure A.2 (method II). The similar results for the observational study are summarized in Figure A.3 (method I) and Figure A.4 (method II).

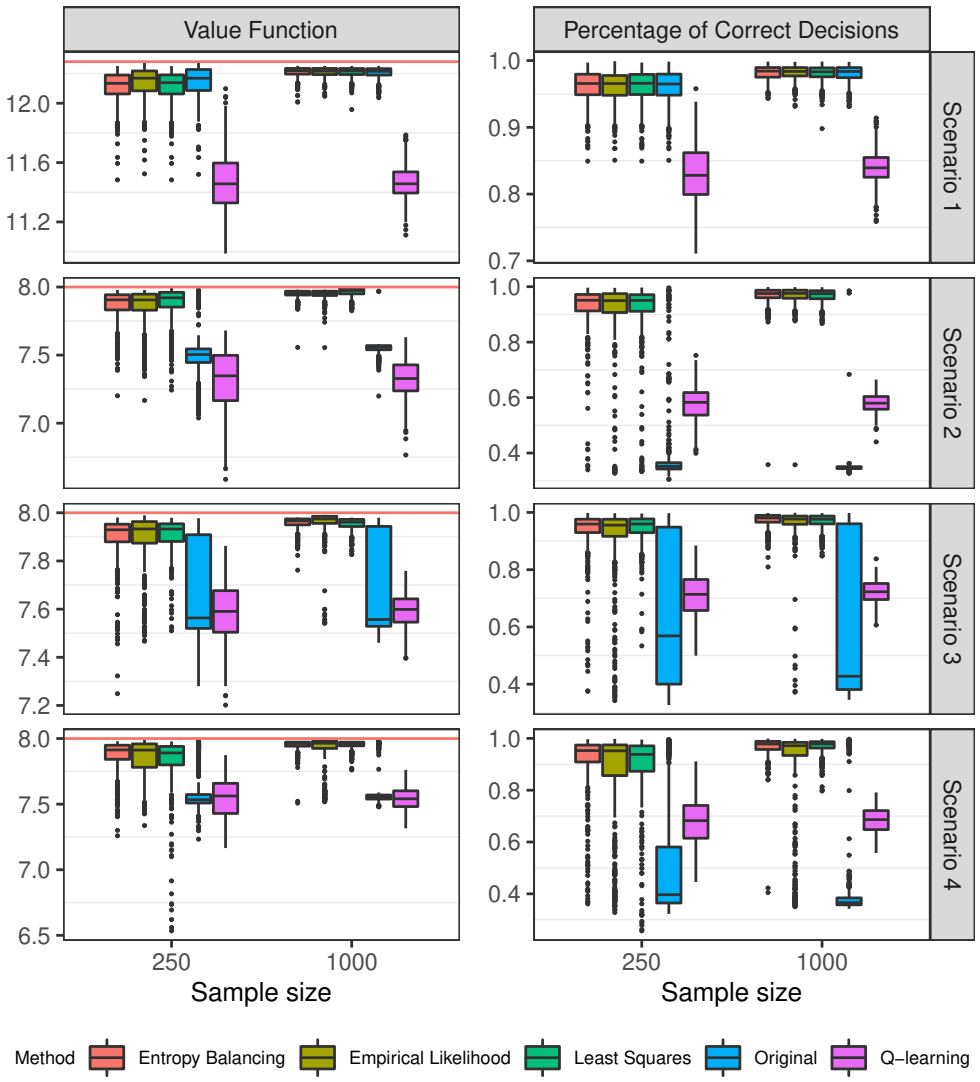

Figure A.1: The value and percentage of correct decisions results of estimated optimal ITRs for the randomization study with implementation method I. The red lines are the values of the true optimal ITRs for the target population.

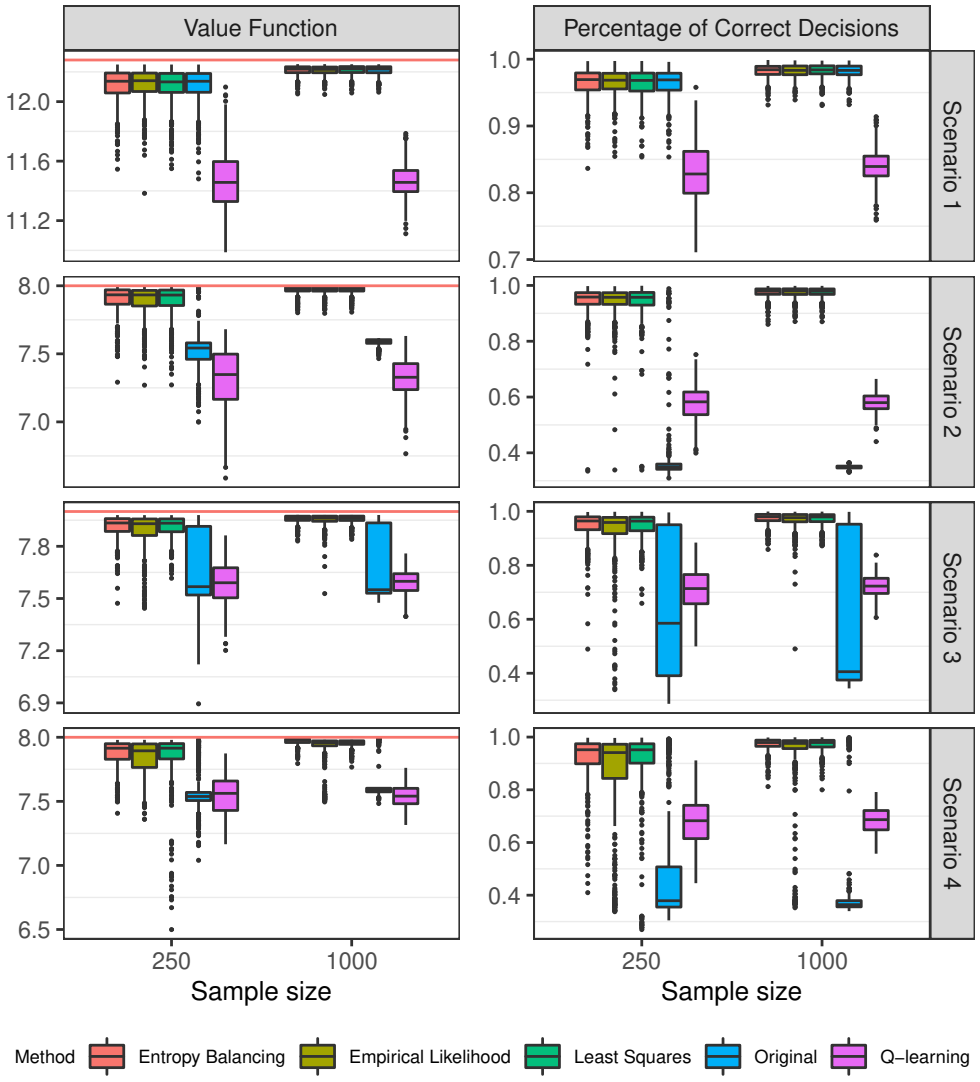

Figure A.2: The value and percentage of correct decisions results of estimated optimal ITRs for the randomization study with implementation method II. The red lines are the values of the true optimal ITRs for the target population.

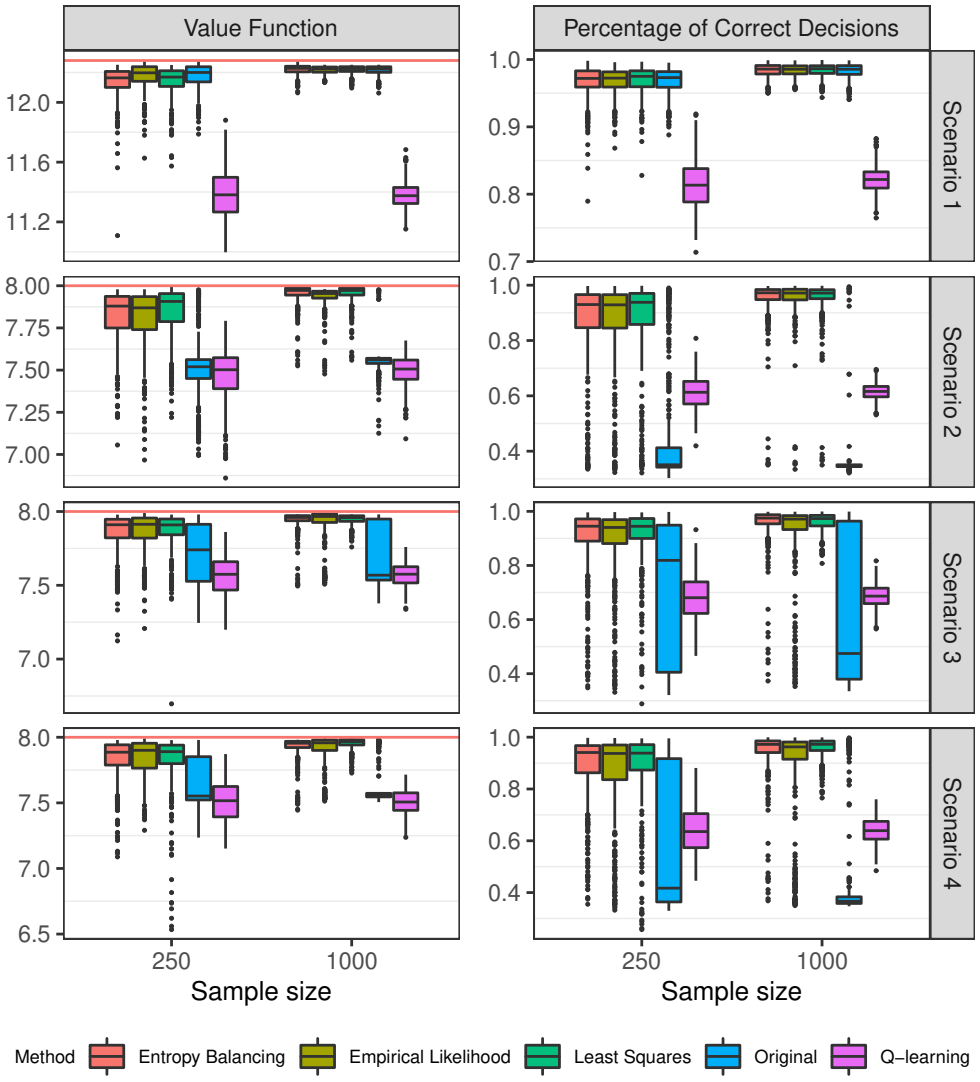

Figure A.3: The value and percentage of correct decisions results of estimated optimal ITRs for the observational study with implementation method I. The red lines are the values of the true optimal ITRs for the target population.

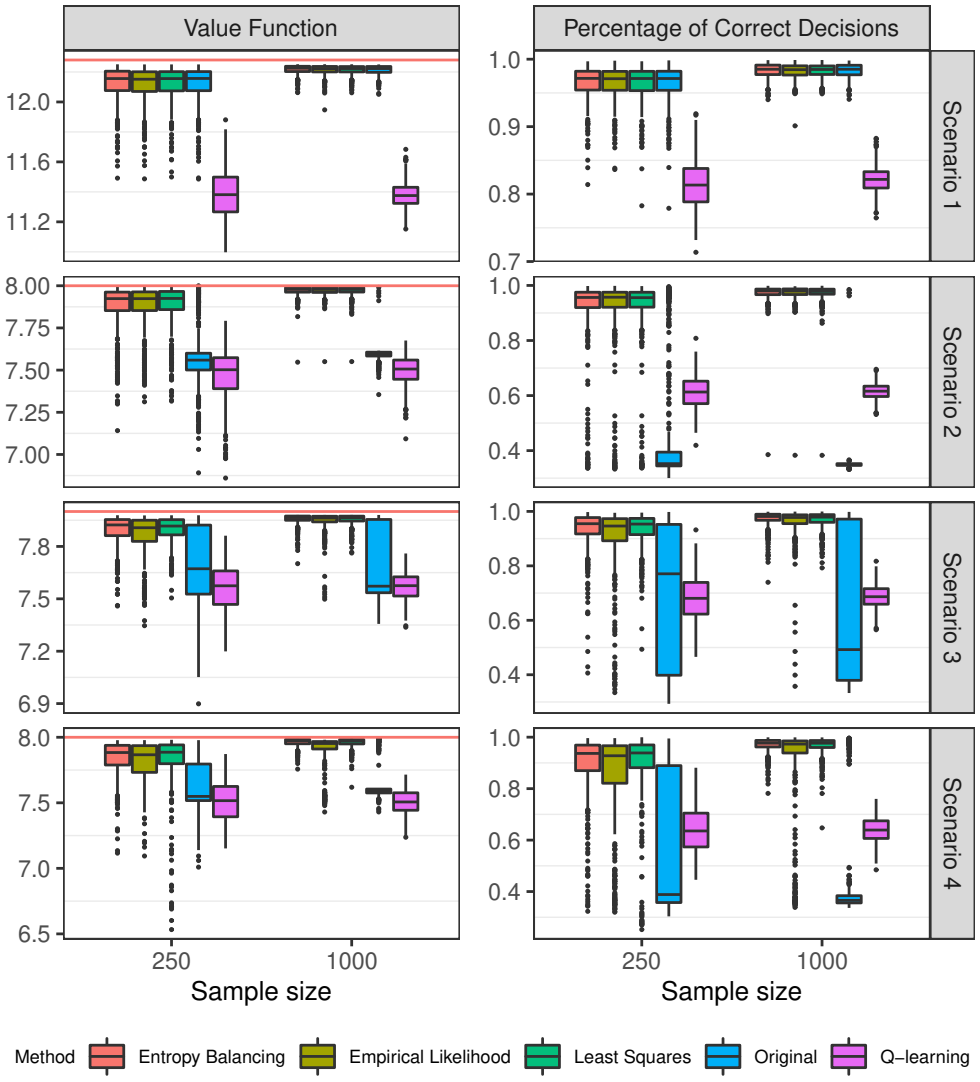

Figure A.4: The value and percentage of correct decisions results of estimated optimal ITRs for the observational study with implementation method II. The red lines are the values of the true optimal ITRs for the target population.

A.2.2 Inference Results of Optimal Value Estimators

The inference results of $\widehat{V}^{\mathrm{c}}_{EB}(\widehat{\beta}^{\mathrm{c}}_{EB})$, $\widehat{V}^{\mathrm{c}}_{EL}(\widehat{\beta}^{\mathrm{c}}_{EL})$, and $\widehat{V}^{\mathrm{c}}_{LS}(\widehat{\beta}^{\mathrm{c}}_{LS})$ for the randomization study are summarized in Table A.1 with implementation method I and Table A.2 with implementation method II. The similar results for the observational study are summarized in Table A.3 with implementation method I and Table A.4 with implementation method II.

Table A.1: Simulation results for the randomization study with implementation method I.

	Scenario	1		2		3		4	
Method	n	250	1000	250	1000	250	1000	250	1000
	$V^{\mathrm{t}}(\beta^{\mathrm{t}})$	12.28		8.00		8.00		8.00	
	$V^{+}_{EB}(\beta^{*}_{EB})$	12.28		8.00		7.99		7.99	
	Mean	12.38	12.32	8.13	8.05	8.13	8.06	8.14	8.06
Entropy	SD	0.41	0.20	0.37	0.18	0.32	0.15	0.37	0.18
Balancing	SE	0.47	0.23	0.41	0.20	0.34	0.17	0.40	0.19
	CP^{+}	96.6	96.4	94.8	95.0	94.8	95.0	95.4	95.2
	CP^{t}	96.6	96.4	94.8	95.0	94.8	95.2	95.8	95.2
	$V^{+}_{EL}(\beta^{*}_{EL})$	12.28		8.00		8.14		8.16	
	Mean	12.40	12.30	8.11	8.04	8.02	8.16	7.86	8.18
Empirical	SD	0.40	0.20	0.34	0.18	0.37	0.18	0.43	0.25
Likelihood	SE	0.38	0.19	0.40	0.20	0.33	0.19	0.38	0.25
	CP^{+}	93.0	93.8	96.4	96.0	88.6	95.4	78.0	94.0
	CP^{t}	93.0	93.8	96.4	96.0	92.2	89.2	85.8	92.0
	$V^{+}_{LS}(\beta^{*}_{LS})$	12.28		8.00		8.13		8.15	
	Mean	12.38	12.31	8.15	8.03	8.27	8.17	8.40	8.20
Least	SD	0.41	0.20	0.37	0.19	0.34	0.18	0.44	0.21
Squares	SE	0.38	0.19	0.49	0.24	0.45	0.22	0.60	0.30
	CP^{+}	92.2	93.4	98.6	98.6	98.2	98.2	98.4	99.6
	CP^{t}	92.2	93.4	98.6	98.6	96.6	93.2	97.2	96.8

Table A.2: Simulation results for the randomization study with implementation method II.

Method	Scenario	1		2		3		4	
	n	250	1000	250	1000	250	1000	250	1000
	$V^{\mathrm{t}}(\beta^{\mathrm{t}})$	12.28		8.00		8.00		8.00	
Entropy Balancing	$V^{+}_{EB}(\beta^{*}_{EB})$	12.28		8.00		7.99		7.99	
	Mean	12.39	12.31	8.12	8.02	8.14	8.05	8.14	8.06
	SD	0.40	0.19	0.33	0.16	0.28	0.14	0.35	0.17
	SE	0.47	0.23	0.39	0.19	0.32	0.15	0.38	0.18
	CP^{+}	96.8	97.4	96.2	96.6	95.4	94.6	95.2	95.4
	CP^{t}	96.8	97.4	96.2	96.6	95.4	95.2	95.2	95.2
Empirical Likelihood	$V^{+}_{EL}(\beta^{*}_{EL})$	12.28		8.00		8.14		8.16	
	Mean	12.38	12.31	8.12	8.02	8.04	8.16	7.82	8.17
	SD	0.40	0.19	0.33	0.16	0.32	0.18	0.45	0.24
	SE	0.38	0.18	0.39	0.19	0.31	0.18	0.36	0.25
	CP^{+}	93.2	94.6	95.6	96.6	92.2	95.2	73.0	93.4
	CP^{t}	93.2	94.6	95.6	96.6	94.4	88.8	81.0	94.0
Least Squares	$V^{+}_{LS}(\beta^{*}_{LS})$	12.28		8.00		8.13		8.15	
	Mean	12.39	12.31	8.13	8.02	8.26	8.16	8.31	8.19
	SD	0.40	0.19	0.34	0.17	0.32	0.16	0.41	0.20
	SE	0.39	0.19	0.48	0.23	0.43	0.21	0.58	0.28
	CP^{+}	93.8	94.4	98.8	99.0	98.8	98.8	99.6	99.6
	CP^{t}	93.8	94.4	98.8	99.0	96.8	94.8	98.0	96.6

Table A.3: Simulation results for the observational study with implementation method I.

Method	Scenario	1		2		3		4	
	n	250	1000	250	1000	250	1000	250	1000
	$V^{\mathrm{t}}(\beta^{\mathrm{t}})$	12.28		8.00		8.00		8.00	
Entropy Balancing	$V^{+}_{EB}(\beta^{*}_{EB})$	12.28		8.00		7.99		7.99	
	Mean	12.34	12.32	8.20	8.06	8.22	8.08	8.24	8.08
	SD	0.42	0.21	0.36	0.17	0.31	0.16	0.45	0.17
	SE	0.47	0.24	0.41	0.19	0.35	0.17	0.43	0.19
	CP^{+}	96.8	95.6	95.4	95.6	95.0	95.2	94.6	94.8
	CP^{t}	96.8	95.6	95.4	95.6	95.2	96.0	94.8	96.0
Empirical Likelihood	$V^{+}_{EL}(\beta^{*}_{EL})$	12.28		8.00		8.14		8.16	
	Mean	12.36	12.31	8.17	8.06	8.09	8.19	7.90	8.22
	SD	0.41	0.22	0.37	0.16	0.35	0.20	0.43	0.26
	SE	0.38	0.20	0.41	0.19	0.32	0.20	0.37	0.27
	CP^{+}	93.6	96.0	96.4	97.0	92.0	96.8	81.2	94.0
	CP^{t}	93.6	96.0	96.4	97.0	93.2	88.4	87.0	93.4
Least Squares	$V^{+}_{LS}(\beta^{*}_{LS})$	12.28		8.00		8.13		8.15	
	Mean	12.37	12.30	8.19	8.05	8.34	8.19	8.40	8.23
	SD	0.42	0.21	0.37	0.17	0.35	0.17	0.44	0.20
	SE	0.39	0.20	0.48	0.23	0.46	0.22	0.60	0.29
	CP^{+}	93.4	95.8	98.4	98.2	97.8	98.4	98.4	98.8
	CP^{t}	93.4	95.8	98.4	98.2	94.0	93.8	97.2	95.4

Table A.4: Simulation results for the observational study with implementation method II.

Method	Scenario	1		2		3		4	
	n	250	1000	250	1000	250	1000	250	1000
	$V^{\mathrm{t}}(\beta^{\mathrm{t}})$	12.28		8.00		8.00		8.00	
Entropy Balancing	$V^{+}_{EB}(\beta^{*}_{EB})$	12.28		8.00		7.99		7.99	
	Mean	12.42	12.32	8.12	8.03	8.15	8.06	8.19	8.06
	SD	0.50	0.22	0.33	0.15	0.29	0.13	0.34	0.15
	SE	0.54	0.25	0.38	0.18	0.30	0.14	0.37	0.17
	CP^{+}	97.4	97.0	96.8	96.2	93.8	95.4	94.6	95.4
	CP^{t}	97.4	97.0	96.8	96.2	94.6	95.6	95.0	96.0
Empirical Likelihood	$V^{+}_{EL}(\beta^{*}_{EL})$	12.28		8.00		8.14		8.16	
	Mean	12.42	12.32	8.11	8.03	8.06	8.18	7.86	8.16
	SD	0.50	0.23	0.33	0.15	0.31	0.17	0.43	0.23
	SE	0.46	0.21	0.37	0.17	0.29	0.17	0.35	0.24
	CP^{+}	95.4	95.4	96.6	96.0	91.6	95.4	74.4	93.2
	CP^{t}	95.4	95.4	96.6	96.0	94.0	84.0	83.4	96.2
Least Squares	$V^{+}_{LS}(\beta^{*}_{LS})$	12.28		8.00		8.13		8.15	
	Mean	12.42	12.32	8.13	8.03	8.31	8.17	8.38	8.21
	SD	0.50	0.23	0.34	0.16	0.32	0.15	0.40	0.18
	SE	0.46	0.22	0.46	0.22	0.42	0.21	0.56	0.27
	CP^{+}	95.8	95.6	98.8	98.4	98.0	99.2	98.6	99.6
	CP^{t}	95.8	95.6	98.8	99.0	96.0	95.0	97.6	96.2

A.2.3 Comparison of Calibration Using All Covariates Versus a Subset of Covariates

We conducted additional simulation studies under the setting in Section 2.5, but only use mean of X_1 as the summary statistics from the target population for calibration. The value and percentage of correct decisions results of the entropy balancing, empirical likelihood, and least squares methods using different covariates for calibration for the randomization study are summarized in Table A.5 (method I) and Table A.6 (method II). The similar results for the observational study are summarized in Table A.7 (method I) and Table A.8 (method II).

Table A.5: Mean and standard deviations (in parenthesis) of the values and percentages of correct decisions (PCD) using different covariates for calibration for the randomization study with implementation method I.

Scenario	n	Covariate	Entropy Balancing		Empirical Likelihood		Least Squares	
			Value	PCD	Value	PCD	Value	PCD
1	250	X_1	12.12(0.10)	0.96(0.02)	12.12(0.10)	0.96(0.02)	12.12(0.10)	0.96(0.02)
		X_1, X_2, X_3	12.11(0.10)	0.96(0.02)	12.14(0.11)	0.96(0.02)	12.11(0.11)	0.96(0.02)
	1000	X_1	12.21(0.03)	0.98(0.01)	12.21(0.03)	0.98(0.01)	12.21(0.03)	0.98(0.01)
		X_1, X_2, X_3	12.21(0.03)	0.98(0.01)	12.21(0.03)	0.98(0.01)	12.21(0.03)	0.98(0.01)
2	250	X_1	7.86(0.13)	0.91(0.11)	7.86(0.13)	0.91(0.11)	7.86(0.13)	0.91(0.11)
		X_1, X_2, X_3	7.87(0.11)	0.92(0.09)	7.86(0.13)	0.91(0.13)	7.88(0.13)	0.91(0.12)
	1000	X_1	7.95(0.04)	0.97(0.02)	7.95(0.03)	0.97(0.02)	7.95(0.04)	0.97(0.02)
		X_1, X_2, X_3	7.95(0.03)	0.97(0.04)	7.95(0.04)	0.97(0.04)	7.96(0.03)	0.97(0.02)
3	250	X_1	7.68(0.19)	0.65(0.25)	7.68(0.19)	0.65(0.25)	7.68(0.19)	0.65(0.25)
		X_1, X_2, X_3	7.90(0.10)	0.93(0.09)	7.89(0.12)	0.91(0.14)	7.91(0.07)	0.94(0.05)
	1000	X_1	7.70(0.20)	0.62(0.28)	7.69(0.20)	0.62(0.28)	7.70(0.20)	0.62(0.28)
		X_1, X_2, X_3	7.96(0.03)	0.97(0.03)	7.96(0.06)	0.96(0.07)	7.95(0.03)	0.97(0.03)
4	250	X_1	7.60(0.16)	0.52(0.23)	7.60(0.17)	0.53(0.23)	7.60(0.16)	0.52(0.23)
		X_1, X_2, X_3	7.87(0.12)	0.91(0.12)	7.84(0.16)	0.85(0.20)	7.83(0.20)	0.89(0.13)
	1000	X_1	7.58(0.12)	0.43(0.18)	7.57(0.11)	0.42(0.16)	7.58(0.12)	0.43(0.18)
		X_1, X_2, X_3	7.95(0.04)	0.97(0.04)	7.90(0.15)	0.88(0.20)	7.95(0.03)	0.97(0.03)

Table A.6: Mean and standard deviations (in parenthesis) of the values and percentages of correct decisions (PCD) using different covariates for calibration for the randomization study with implementation method II.

Scenario	n	Covariate	Entropy Balancing		Empirical Likelihood		Least Squares	
			Value	PCD	Value	PCD	Value	PCD
1	250	X_1	12.11(0.11)	0.96(0.02)	12.11(0.11)	0.96(0.02)	12.11(0.11)	0.96(0.02)
		X_1, X_2, X_3	12.11(0.11)	0.96(0.02)	12.11(0.11)	0.96(0.02)	12.10(0.12)	0.96(0.02)
	1000	X_1	12.21(0.04)	0.98(0.01)	12.21(0.03)	0.98(0.01)	12.21(0.04)	0.98(0.01)
		X_1, X_2, X_3	12.21(0.03)	0.98(0.01)	12.21(0.03)	0.98(0.01)	12.21(0.03)	0.98(0.01)
2	250	X_1	7.86(0.12)	0.93(0.07)	7.86(0.12)	0.93(0.08)	7.86(0.12)	0.93(0.07)
		X_1, X_2, X_3	7.90(0.10)	0.94(0.06)	7.90(0.10)	0.94(0.06)	7.89(0.11)	0.94(0.07)
	1000	X_1	7.96(0.03)	0.98(0.02)	7.95(0.03)	0.98(0.02)	7.96(0.03)	0.98(0.02)
		X_1, X_2, X_3	7.97(0.03)	0.97(0.02)	7.97(0.03)	0.97(0.02)	7.97(0.03)	0.97(0.02)
3	250	X_1	7.69(0.21)	0.66(0.26)	7.68(0.20)	0.66(0.26)	7.69(0.21)	0.66(0.26)
		X_1, X_2, X_3	7.91(0.07)	0.95(0.05)	7.89(0.10)	0.92(0.11)	7.91(0.06)	0.95(0.05)
	1000	X_1	7.70(0.20)	0.62(0.28)	7.70(0.20)	0.62(0.28)	7.70(0.20)	0.62(0.28)
		X_1, X_2, X_3	7.96(0.02)	0.97(0.02)	7.95(0.04)	0.97(0.04)	7.96(0.02)	0.97(0.02)
4	250	X_1	7.59(0.18)	0.51(0.24)	7.58(0.18)	0.51(0.24)	7.59(0.18)	0.51(0.24)
		X_1, X_2, X_3	7.88(0.10)	0.92(0.09)	7.83(0.15)	0.86(0.18)	7.85(0.21)	0.91(0.13)
	1000	X_1	7.58(0.11)	0.42(0.17)	7.58(0.11)	0.41(0.16)	7.58(0.11)	0.42(0.17)
		X_1, X_2, X_3	7.97(0.03)	0.97(0.02)	7.92(0.11)	0.93(0.14)	7.95(0.03)	0.97(0.02)

Table A.7: Mean and standard deviations (in parenthesis) of the values and percentages of correct decisions (PCD) using different covariates for calibration for the observational study with implementation method I.

Scenario	n	Covariate	Entropy Balancing		Empirical Likelihood		Least Squares	
			Value	PCD	Value	PCD	Value	PCD
1	250	X_1	12.14(0.10)	0.97(0.02)	12.14(0.09)	0.97(0.02)	12.14(0.10)	0.97(0.02)
		X_1, X_2, X_3	12.14(0.10)	0.97(0.02)	12.18(0.08)	0.97(0.02)	12.14(0.09)	0.97(0.02)
	1000	X_1	12.22(0.03)	0.98(0.01)	12.22(0.03)	0.98(0.01)	12.22(0.03)	0.98(0.01)
		X_1, X_2, X_3	12.22(0.03)	0.98(0.01)	12.22(0.03)	0.98(0.01)	12.22(0.03)	0.98(0.01)
2	250	X_1	7.82(0.15)	0.87(0.16)	7.82(0.15)	0.87(0.16)	7.82(0.15)	0.87(0.16)
		X_1, X_2, X_3	7.82(0.16)	0.87(0.17)	7.81(0.17)	0.86(0.16)	7.84(0.16)	0.87(0.17)
	1000	X_1	7.94(0.06)	0.95(0.07)	7.94(0.05)	0.96(0.07)	7.94(0.06)	0.95(0.07)
		X_1, X_2, X_3	7.95(0.06)	0.95(0.08)	7.93(0.06)	0.95(0.08)	7.95(0.06)	0.95(0.07)
3	250	X_1	7.72(0.19)	0.71(0.26)	7.73(0.19)	0.71(0.26)	7.72(0.19)	0.71(0.26)
		X_1, X_2, X_3	7.86(0.13)	0.90(0.12)	7.86(0.13)	0.89(0.14)	7.87(0.12)	0.91(0.10)
	1000	X_1	7.73(0.20)	0.66(0.28)	7.73(0.20)	0.66(0.28)	7.73(0.20)	0.66(0.28)
		X_1, X_2, X_3	7.94(0.07)	0.96(0.07)	7.92(0.12)	0.91(0.15)	7.95(0.03)	0.96(0.03)
4	250	X_1	7.64(0.18)	0.57(0.26)	7.64(0.18)	0.57(0.26)	7.64(0.18)	0.57(0.26)
		X_1, X_2, X_3	7.83(0.16)	0.89(0.13)	7.84(0.15)	0.86(0.17)	7.83(0.20)	0.89(0.13)
	1000	X_1	7.61(0.14)	0.45(0.21)	7.61(0.14)	0.46(0.21)	7.61(0.14)	0.45(0.21)
		X_1, X_2, X_3	7.93(0.09)	0.95(0.09)	7.89(0.15)	0.88(0.20)	7.95(0.04)	0.96(0.04)

Table A.8: Mean and standard deviations (in parenthesis) of the values and percentages of correct decisions (PCD) using different covariates for calibration for the observational study with implementation method II.

Scenario	n	Covariate	Entropy Balancing		Empirical Likelihood		Least Squares	
			Value	PCD	Value	PCD	Value	PCD
1	250	X_1	12.12(0.12)	0.97(0.03)	12.12(0.11)	0.97(0.02)	12.12(0.12)	0.97(0.03)
		X_1, X_2, X_3	12.12(0.11)	0.96(0.02)	12.12(0.12)	0.97(0.02)	12.12(0.12)	0.96(0.03)
	1000	X_1	12.21(0.04)	0.98(0.01)	12.21(0.03)	0.98(0.01)	12.21(0.04)	0.98(0.01)
		X_1, X_2, X_3	12.22(0.03)	0.98(0.01)	12.21(0.03)	0.98(0.01)	12.21(0.03)	0.98(0.01)
2	250	X_1	7.86(0.12)	0.91(0.13)	7.86(0.12)	0.92(0.13)	7.86(0.12)	0.91(0.13)
		X_1, X_2, X_3	7.87(0.14)	0.91(0.14)	7.88(0.13)	0.91(0.14)	7.88(0.13)	0.91(0.15)
	1000	X_1	7.95(0.04)	0.97(0.04)	7.95(0.04)	0.97(0.04)	7.95(0.04)	0.97(0.04)
		X_1, X_2, X_3	7.97(0.03)	0.97(0.03)	7.97(0.03)	0.97(0.03)	7.97(0.03)	0.97(0.03)
3	250	X_1	7.71(0.21)	0.68(0.27)	7.71(0.21)	0.69(0.27)	7.71(0.21)	0.68(0.27)
		X_1, X_2, X_3	7.89(0.09)	0.93(0.07)	7.87(0.12)	0.90(0.13)	7.90(0.07)	0.93(0.06)
	1000	X_1	7.71(0.20)	0.64(0.28)	7.72(0.20)	0.64(0.28)	7.71(0.20)	0.64(0.28)
		X_1, X_2, X_3	7.96(0.03)	0.97(0.03)	7.95(0.06)	0.96(0.06)	7.95(0.03)	0.97(0.03)
4	250	X_1	7.62(0.19)	0.56(0.26)	7.62(0.19)	0.57(0.26)	7.62(0.19)	0.56(0.26)
		X_1, X_2, X_3	7.84(0.14)	0.89(0.12)	7.81(0.16)	0.85(0.18)	7.82(0.22)	0.89(0.14)
	1000	X_1	7.60(0.14)	0.45(0.20)	7.59(0.12)	0.43(0.18)	7.60(0.14)	0.45(0.20)
		X_1, X_2, X_3	7.96(0.03)	0.97(0.03)	7.90(0.13)	0.90(0.17)	7.96(0.04)	0.97(0.03)

APPENDIX

B

SUPPLEMENT TO CHAPTER 3

B.1 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.

B.1.1 Proof of Theorem 4

Proof. To prove Theorem 4, we first state a lemma.

Lemma 2 *For any function $h(X)$ that has a finite moment $\mathbb{E}[h(X) \mid U = 11] < \infty$, the following balancing condition*

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

holds. Under Assumptions 4–6, the three expectations in (B.1) *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, (B.1) holds. Under Assumptions 4–6, we have

$$\begin{aligned}
&P(U=11\mid X)=P[S(0)=1,S(1)=1\mid X] \\
=&P[S(0)=1\mid X] \quad \text{(Assumption 6)} \\
=&P[S(0)=1\mid C(0)=0,X] \quad \text{(Assumption 5)} \\
=&P[S(0)=1\mid A=0,C(0)=0,X] \quad \text{(Assumption 4)} \\
=&P(S=1\mid A=0,C=0,X) \quad \text{(SUTVA)} \\
=&p_0(X),
\end{aligned}$$

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

$$\mathbb{E}[h(X)\mid U=11]=\mathbb{E}\left[\frac{f(X\mid U=11)}{f(X)}h(X)\right]=\mathbb{E}\left[\frac{P(U=11\mid X)}{P(U=11)}h(X)\right]=\mathbb{E}\left[\frac{p_0(X)}{p_0}h(X)\right]. \quad \text{(B.2)}$$

□

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

$$\begin{aligned}
V(\pi) &= \mathbb{E}[Y(1)\pi(X)+Y(0)\{1-\pi(X)\}\mid U=11] \\
&= \mathbb{E}[\mathbb{E}[Y(1)\pi(X)+Y(0)\{1-\pi(X)\}\mid U=11,X]\mid U=11] \\
&= \mathbb{E}[\mathbb{E}[Y(1)\mid U=11,X]\pi(X)\mid U=11]+\mathbb{E}[\mathbb{E}[Y(0)\mid U=11,X]\{1-\pi(X)\}\mid U=11] \\
&= \mathbb{E}[\mathbb{E}[Y(1)\mid U=11 \text{ or } 10,X]\pi(X)\mid U=11] \quad \text{(Assumption 7)} \\
&\quad +\mathbb{E}[\mathbb{E}[Y(0)\mid U=11,X]\{1-\pi(X)\}\mid U=11] \\
&= \mathbb{E}[\mathbb{E}[Y(1)\mid S(1)=1,X]\pi(X)\mid U=11]
\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}[\mathbb{E}[Y(0)\mid S(0)=1,X]\{1-\pi(X)\}\mid U=11] \quad \text{(Assumption 6)}\\
=\;& \mathbb{E}[\mathbb{E}[Y(1)\mid C(1)=0,S(1)=1,X]\pi(X)\mid U=11]\\
&+\mathbb{E}[\mathbb{E}[Y(0)\mid C(0)=0,S(0)=1,X]\{1-\pi(X)\}\mid U=11] \quad \text{(Assumption 5)}\\
=\;& \mathbb{E}[\mathbb{E}[Y\mid A=1,C(1)=0,S(1)=1,X]\pi(X)\mid U=11]\\
&+\mathbb{E}[\mathbb{E}[Y\mid A=0,C(0)=0,S(0)=1,X]\{1-\pi(X)\}\mid U=11] \quad \text{(Assumption 4)}\\
=\;& \mathbb{E}[\mathbb{E}[Y\mid A=1,C=0,S=1,X]\pi(X)\mid U=11]\\
&+\mathbb{E}[\mathbb{E}[Y\mid A=0,C=0,S=1,X]\{1-\pi(X)\}\mid U=11] \quad \text{(SUTVA)}\\
=\;& \mathbb{E}\big[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\mid U=11\big]. \qquad \text{(B.3)}
\end{aligned}
$$

By Lemma 2, 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 (B.3),

$$
V(\pi) \;= \mathbb{E}\Big[\tfrac{p_0(X)}{p_0}\big[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\big]\Big], \qquad \text{(B.4)}
$$

and

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

□

B.1.2 Proof of Theorem 5

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

$$f(O)=f(X)f(A\mid X)f(C\mid A,X)\{f(S\mid A,C=0,X)\}^{\mathbb{I}(C=0)}\{f(Y\mid A,C=0,S=1,X)\}^{\mathbb{I}(C=0,S=1)}. \quad \text{(B.5)}$$

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 (B.5), 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 3 *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*

$$\nu_R(O) = \frac{1}{D}\nu_N(O) - \frac{R}{D}\nu_D(O). \tag{B.6}$$

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

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

which yields (B.6). □

Lemma 4 *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 5 *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*

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

As a special case,

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

Proof. We first prove the general result:

$$\begin{aligned}
\dot{\mu}_{af,\theta}(X)|_{\theta=0} &= \frac{\partial}{\partial\theta}\mathbb{E}_\theta[f(Y,S,X) \mid A=a, C=0, X]|_{\theta=0} \\
&= \mathbb{E}[f(Y,S,X) \times s(Y,S \mid A=a, C=0, X) \mid A=a, C=0, X] \quad \text{(Lemma 4)} \\
&= \mathbb{E}\big[\{f(Y,S,X) - \mu_{af}(X)\}s(Y,S \mid A=a, C=0, X) \mid A=a, C=0, X\big] \\
&= \mathbb{E}\left[\frac{\mathbb{I}\{A=a\}\mathbb{I}(C=0)\{f(Y,S,X) - \mu_{af}(X)\}}{\varphi_a(X)K_a(X)}s(Y,S \mid A, C=0, X) \mid X\right] \\
&= \mathbb{E}\big[\{\psi_{f(Y(a),S(a),X)} - \mu_{af}(X)\}s(Y,S \mid A, C=0, X) \mid X\big],
\end{aligned} \tag{B.7}$$

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

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

□

Lemma 6 *Define $\mu_{af}=\mathbb{E}[\mu_{af}(X)]$ for any $f(Y,S,X)$, where $a=0,1$. We have*

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

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

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

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

Proof. We prove the general result:

$$\begin{aligned}
\dot{\mu}_{af,\theta}|_{\theta=0} &= \mathbb{E}[\mu_{af}(X)s(O)]+\mathbb{E}[\dot{\mu}_{af,\theta}(X)|_{\theta=0}] \quad \text{(Lemma 4)}\\
&= \mathbb{E}[\mu_{af}(X)s(O)]+\mathbb{E}[\{\psi_{f(Y(a),S(a),X)}-\mu_{af}(X)\}s(Y,S\mid A,C=0,X)] \quad \text{(Lemma 5)}\\
&= \mathbb{E}[\{\psi_{f(Y(a),S(a),X)}-\mu_{af}\}s(Y,S\mid A,C=0,X)]\\
&= \mathbb{E}[\{\psi_{f(Y(a),S(a),X)}-\mu_{af}\}s(O)],
\end{aligned}$$

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 7 *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 5, the numerator satisfies

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

The denominator satisfies

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

We then use Lemma 3 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 $\mathbb{E}\big[\{\psi_{Y(a)S(a)} - \mu_a(X)\psi_{S(a)}\} s(S \mid A, C=0, X) \mid X\big] = 0$. □

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

First, from (B.4), we can write $V(\pi) = N/D$, where

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

From Lemma 6,

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

From the chain rule, we have

$$\begin{aligned}
\dot{N}_\theta|_{\theta=0} &= \frac{\partial}{\partial\theta}\mathbb{E}_\theta\big[p_{0,\theta}(X)\big[\mu_{1,\theta}(X)\pi(X) + \mu_{0,\theta}(X)\{1-\pi(X)\}\big]\big]\Big|_{\theta=0} \\
&= \mathbb{E}\big[p_0(X)\big[\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\big] s(X)\big] \quad \text{(Lemma 4)} \\
&\quad + \mathbb{E}_\theta\big[\dot{p}_{0,\theta}(X)\big[\mu_{1,\theta}(X)\pi(X) + \mu_{0,\theta}(X)\{1-\pi(X)\}\big]\big]|_{\theta=0} \\
&\quad + \mathbb{E}_\theta\big[p_{0,\theta}(X)\big[\dot{\mu}_{1,\theta}(X)\pi(X) + \dot{\mu}_{0,\theta}(X)\{1-\pi(X)\}\big]\big]|_{\theta=0}. \qquad \text{(B.8)}
\end{aligned}$$

Because $E\{N s(X)\} = 0$, the first term in (B.8) equals

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

From Lemma 5, the second term in (B.8) reduces to

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

From Lemma 7, the third term in (B.8) reduces to

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

Plugging the above three formulas into (B.8) gives

$$\begin{aligned}
&\dot{N}_\theta\,|_{\theta=0}\\
=\;&\mathbb{E}\left[\left\{p_0(X)\left[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\right]-N\right\}s(X)\right]\\
&+\mathbb{E}\left[\{\psi_{S(0)}-p_0(X)\}\left\{\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\right\}s(S\mid A,C=0,X)\right]\\
&+\mathbb{E}\left[p_0(X)\left\{\frac{\psi_{Y(1)S(1)}-\mu_1(X)\psi_{S(1)}}{p_1(X)}\pi(X)+\frac{\psi_{Y(0)S(0)}-\mu_0(X)\psi_{S(0)}}{p_0(X)}\{1-\pi(X)\}\right\}s(Y\mid A,C=0,S=1,X)\right].
\end{aligned}\tag{B.9}$$

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 (B.9) as

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

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 3, the EIF for $V(\pi)$ is

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

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

B.1.3 Proof of Theorem 6

Proof. Recall the MR estimator

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

From (B.4), we can write $V(\pi) = N/D$, where

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

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 $\mathscr{M}_{\text{ps+cs+sp}} \cup \mathscr{M}_{\text{ps+cs+om}} \cup \mathscr{M}_{\text{sp+om}}$.

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

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

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

$$\begin{aligned}\mathbb{E}[\phi^*_\pi]=&\mathbb{E}\left[\left[\mu_1^*(X)\left\{\psi^*_{S(0)}-\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{S(1)}\right\}+\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{Y(1)S(1)}\right]\pi(X)+\psi^*_{Y(0)S(0)}\{1-\pi(X)\}\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\psi^*_{S(0)}\pi(X)\right]-\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{S(1)}\pi(X)\right]+\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{Y(1)S(1)}\pi(X)\right]\\ &+\mathbb{E}\left[\psi^*_{Y(0)S(0)}\{1-\pi(X)\}\right].\end{aligned}\tag{B.10}$$

(i) Under $\mathscr{M}_{\text{ps+cs+sp}}$, $\varphi_a(X)=\varphi_a^*(X)$, $K_a(X)=K_a^*(X)$, $p_a(X)=p_a^*(X)$. The first term in (B.10) equals

$$\begin{aligned}&\mathbb{E}\left[\mu_1^*(X)\psi^*_{S(0)}\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\mu_1^*(X)\psi^*_{S(0)}\pi(X)\mid A=0,C=0,X\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)\left[\frac{P(A=0,C=0\mid X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mathbb{E}[S=1\mid A=0,C=0,X]-p_0^*(X)\right\}+p_0^*(X)\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)\left[\frac{\varphi_0(X)K_0(X)}{\varphi_0(X)K_0(X)}\left\{p_0(X)-p_0(X)\right\}+p_0(X)\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)p_0(X)\right].\end{aligned}\tag{B.11}$$

The second term in (B.10) equals

$$\begin{aligned}&\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{S(1)}\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{S(1)}\pi(X)\mid A=1,C=0,X\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left[\frac{P(A=1,C=0\mid X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mathbb{E}[S=1\mid A=1,C=0,X]-p_1^*(X)\right\}+p_1^*(X)\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)\frac{p_0(X)}{p_1(X)}\left[\frac{\varphi_1(X)K_1(X)}{\varphi_1(X)K_1(X)}\left\{p_1(X)-p_1(X)\right\}+p_1(X)\right]\right]\\ =&\mathbb{E}\left[\mu_1^*(X)\pi(X)p_0(X)\right].\end{aligned}\tag{B.12}$$

The third term in (B.10) equals

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

$$=\mathbb{E}\big[\mu_1(X)\pi(X)p_0(X)\big]. \tag{B.13}$$

The fourth term in (B.10) equals

$$\begin{aligned}
&\mathbb{E}\Big[\psi^*_{Y(0)S(0)}\{1-\pi(X)\}\Big]=\mathbb{E}\Big[\mathbb{E}\Big[\psi^*_{Y(0)S(0)}\{1-\pi(X)\}\mid A=0,C=0,X\Big]\Big]\\
=&\mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{P(A=0,C=0\mid X)}{\varphi^*_0(X)K^*_0(X)}\big\{\mathbb{E}[YS\mid A=0,C=0,X]-\mu^*_0(X)p^*_0(X)\big\}+\mu^*_0(X)p^*_0(X)\right\}\right]\\
=&\mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{\varphi_0(X)K_0(X)}{\varphi_0(X)K_0(X)}\big\{\mu_0(X)p_0(X)-\mu^*_0(X)p_0(X)\big\}+\mu^*_0(X)p_0(X)\right\}\right]\\
=&\mathbb{E}\big[\mu_0(X)\{1-\pi(X)\}p_0(X)\big].
\end{aligned} \tag{B.14}$$

Combining (B.10)–(B.14), we have

$$\begin{aligned}
&\mathbb{E}\left[\left[\mu^*_1(X)\left\{\psi^*_{S(0)}-\frac{p^*_0(X)}{p^*_1(X)}\psi^*_{S(1)}\right\}+\frac{p^*_0(X)}{p^*_1(X)}\psi^*_{Y(1)S(1)}\right]\pi(X)+\psi^*_{Y(0)S(0)}\{1-\pi(X)\}\right]\\
=&\mathbb{E}\big[\mu^*_1(X)\pi(X)p_0(X)\big]-\mathbb{E}\big[\mu^*_1(X)\pi(X)p_0(X)\big]+\mathbb{E}\big[\mu_1(X)\pi(X)p_0(X)\big]\\
&+\mathbb{E}\big[\mu_0(X)\{1-\pi(X)\}p_0(X)\big]\\
=&\mathbb{E}\big[p_0(X)\big[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\big]\big]=N.
\end{aligned}$$

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

(ii) Under $\mathscr{M}_{\text{ps+cs+om}}$, $\varphi_a(X)=\varphi^*_a(X)$, $K_a(X)=K^*_a(X)$, $\mu_a(X)=\mu^*_a(X)$. The first term in (B.10) equals

$$\begin{aligned}
&\mathbb{E}\Big[\mu^*_1(X)\psi^*_{S(0)}\pi(X)\Big]=\mathbb{E}\Big[\mathbb{E}\Big[\mu^*_1(X)\psi^*_{S(0)}\pi(X)\mid A=0,C=0,X\Big]\Big]\\
=&\mathbb{E}\left[\mu^*_1(X)\pi(X)\left[\frac{P(A=0,C=0\mid X)}{\varphi^*_0(X)K^*_0(X)}\big\{\mathbb{E}[S=1\mid A=0,C=0,X]-p^*_0(X)\big\}+p^*_0(X)\right]\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)\left[\frac{\varphi_0(X)K_0(X)}{\varphi_0(X)K_0(X)}\big\{p_0(X)-p^*_0(X)\big\}+p^*_0(X)\right]\right]\\
=&\mathbb{E}\big[\mu_1(X)\pi(X)p_0(X)\big].
\end{aligned} \tag{B.15}$$

The second term in (B.10) equals

$$\begin{aligned}
&\mathbb{E}\left[\mu^*_1(X)\frac{p^*_0(X)}{p^*_1(X)}\psi^*_{S(1)}\pi(X)\right]=\mathbb{E}\left[\mathbb{E}\left[\mu^*_1(X)\frac{p^*_0(X)}{p^*_1(X)}\psi^*_{S(1)}\pi(X)\mid A=1,C=0,X\right]\right]\\
=&\mathbb{E}\left[\mu^*_1(X)\pi(X)\frac{p^*_0(X)}{p^*_1(X)}\left[\frac{P(A=1,C=0\mid X)}{\varphi^*_1(X)K^*_1(X)}\big\{\mathbb{E}[S=1\mid A=1,C=0,X]-p^*_1(X)\big\}+p^*_1(X)\right]\right]\\
=&\mathbb{E}\left[\mu_1(X)\pi(X)\frac{p^*_0(X)}{p^*_1(X)}\left[\frac{\varphi_1(X)K_1(X)}{\varphi_1(X)K_1(X)}\big\{p_1(X)-p^*_1(X)\big\}+p^*_1(X)\right]\right]
\end{aligned}$$

$$= \mathbb{E}\left[\mu_1(X)\pi(X)p_0^*(X)\frac{p_1(X)}{p_1^*(X)}\right]. \tag{B.16}$$

The third term in (B.10) equals

$$\begin{aligned}
&\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{Y(1)S(1)}\pi(X)\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi^*_{Y(1)S(1)}\pi(X) \mid A=1, C=0, X\right]\right] \\
=&\mathbb{E}\left[\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left\{\frac{P(A=1, C=0 \mid X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mathbb{E}[YS \mid A=1, C=0, X] - \mu_1^*(X)p_1^*(X)\right\} + \mu_1^*(X)p_1^*(X)\right\}\right] \\
=&\mathbb{E}\left[\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left\{\frac{\varphi_1(X)K_1(X)}{\varphi_1(X)K_1(X)}\left\{\mu_1(X)p_1(X) - \mu_1(X)p_1^*(X)\right\} + \mu_1(X)p_1^*(X)\right\}\right] \\
=&\mathbb{E}\left[\mu_1(X)\pi(X)p_0^*(X)\frac{p_1(X)}{p_1^*(X)}\right].
\end{aligned} \tag{B.17}$$

The fourth term in (B.10) equals

$$\begin{aligned}
&\mathbb{E}\left[\psi^*_{Y(0)S(0)}\{1-\pi(X)\}\right] = \mathbb{E}\left[\mathbb{E}\left[\psi^*_{Y(0)S(0)}\{1-\pi(X)\} \mid A=0, C=0, X\right]\right] \\
=&\mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{P(A=0, C=0 \mid X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mathbb{E}[YS \mid A=0, C=0, X] - \mu_0^*(X)p_0^*(X)\right\} + \mu_0^*(X)p_0^*(X)\right\}\right] \\
=&\mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{\varphi_0(X)K_0(X)}{\varphi_0(X)K_0(X)}\left\{\mu_0(X)p_0(X) - \mu_0(X)p_0^*(X)\right\} + \mu_0(X)p_0^*(X)\right\}\right] \\
=&\mathbb{E}\left[\mu_0(X)\{1-\pi(X)\}p_0(X)\right].
\end{aligned} \tag{B.18}$$

Combining (B.10) and (B.15)–(B.18) , 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 $\mathscr{M}_{\text{ps+cs+om}}$, $\widehat{V}^{\text{MR}}(\pi)$ is consistent for $V(\pi)$.

(iii) Under $\mathscr{M}_{\text{sp+om}}$, $p_a(X) = p_a^*(X)$, $\mu_a(X) = \mu_a^*(X)$. The first term in (B.10) equals

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

$$
\begin{aligned}
&= \mathbb{E}\left[\mu_1(X)\pi(X)\left[\frac{\varphi_0(X)K_0(X)}{\varphi_0^*(X)K_0^*(X)}\left\{p_0(X)-p_0(X)\right\}+p_0(X)\right]\right] \\
&= \mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right]. \qquad \text{(B.19)}
\end{aligned}
$$

The second term in (B.10) equals

$$
\begin{aligned}
&\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi_{S(1)}^*\pi(X)\right] = \mathbb{E}\left[\mathbb{E}\left[\mu_1^*(X)\frac{p_0^*(X)}{p_1^*(X)}\psi_{S(1)}^*\pi(X)\mid A=1,C=0,X\right]\right] \\
&= \mathbb{E}\left[\mu_1^*(X)\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left[\frac{P(A=1,C=0\mid X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mathbb{E}[S=1\mid A=1,C=0,X]-p_1^*(X)\right\}+p_1^*(X)\right]\right] \\
&= \mathbb{E}\left[\mu_1(X)\pi(X)\frac{p_0(X)}{p_1(X)}\left[\frac{\varphi_1(X)K_1(X)}{\varphi_1^*(X)K_1^*(X)}\left\{p_1(X)-p_1(X)\right\}+p_1(X)\right]\right] \\
&= \mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right]. \qquad \text{(B.20)}
\end{aligned}
$$

The third term in (B.10) equals

$$
\begin{aligned}
&\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi_{Y(1)S(1)}^*\pi(X)\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{p_0^*(X)}{p_1^*(X)}\psi_{Y(1)S(1)}^*\pi(X)\mid A=1,C=0,X\right]\right] \\
&= \mathbb{E}\left[\pi(X)\frac{p_0^*(X)}{p_1^*(X)}\left\{\frac{P(A=1,C=0\mid X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mathbb{E}[YS\mid A=1,C=0,X]-\mu_1^*(X)p_1^*(X)\right\}+\mu_1^*(X)p_1^*(X)\right\}\right] \\
&= \mathbb{E}\left[\pi(X)\frac{p_0(X)}{p_1(X)}\left\{\frac{\varphi_1(X)K_1(X)}{\varphi_1^*(X)K_1^*(X)}\left\{\mu_1(X)p_1(X)-\mu_1(X)p_1(X)\right\}+\mu_1(X)p_1(X)\right\}\right] \\
&= \mathbb{E}\left[\mu_1(X)\pi(X)p_0(X)\right]. \qquad \text{(B.21)}
\end{aligned}
$$

The fourth term in (B.10) equals

$$
\begin{aligned}
&\mathbb{E}\left[\psi_{Y(0)S(0)}^*\{1-\pi(X)\}\right] = \mathbb{E}\left[\mathbb{E}\left[\psi_{Y(0)S(0)}^*\{1-\pi(X)\}\mid A=0,C=0,X\right]\right] \\
&= \mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{P(A=0,C=0\mid X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mathbb{E}[YS\mid A=0,C=0,X]-\mu_0^*(X)p_0^*(X)\right\}+\mu_0^*(X)p_0^*(X)\right\}\right] \\
&= \mathbb{E}\left[\{1-\pi(X)\}\left\{\frac{\varphi_0(X)K_0(X)}{\varphi_0^*(X)K_0^*(X)}\left\{\mu_0(X)p_0(X)-\mu_0(X)p_0(X)\right\}+\mu_0(X)p_0(X)\right\}\right] \\
&= \mathbb{E}\left[\mu_0(X)\{1-\pi(X)\}p_0(X)\right]. \qquad \text{(B.22)}
\end{aligned}
$$

Combining (B.10) and (B.19)–(B.22) , we have

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

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

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

Under $\mathscr{M}_{\text{ps+cs+sp+om}}$, we show that the influence function of $\widehat{V}^{\text{MR}}(\pi)$ is the same as the EIF in Theorem 5 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

$$
\mathbb{P}_n\{g(O;\widehat{\theta})\}=\mathbb{P}_n\{g(O;\theta^*)\}+\mathbb{P}\{\dot{g}(O;\theta^*)\}(\widehat{\theta}-\theta)+o_p(n^{-1/2})=0. \tag{B.23}
$$

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

$$
\frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}_n\{D(O;\widehat{\theta})\}}=\frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}\{D(O;\theta*)\}}-\frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta*)\}]^2}\big[\mathbb{P}_n\{D(O;\widehat{\theta})\}-\mathbb{P}\{D(O;\theta^*)\}\big]+o_p(n^{-1/2}). \tag{B.24}
$$

$$
\mathbb{P}_n\{N(O;\widehat{\theta})\}=\mathbb{P}_n\{N(O;\theta^*)\}+\mathbb{P}\{\dot{N}(O;\theta^*)\}(\widehat{\theta}-\theta)+o_p(n^{-1/2}), \tag{B.25}
$$

$$
\mathbb{P}_n\{D(O;\widehat{\theta})\}=\mathbb{P}_n\{D(O;\theta^*)\}+\mathbb{P}\{\dot{D}(O;\theta^*)\}(\widehat{\theta}-\theta)+o_p(n^{-1/2}). \tag{B.26}
$$

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 (B.23)–(B.26), we have

$$
\begin{aligned}
&\widehat{V}^{\text{MR}}(\pi)-V(\pi)=\frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}_n\{D(O;\widehat{\theta})\}}-V(\pi)\\
=&\frac{\mathbb{P}_n\{N(O;\widehat{\theta})\}}{\mathbb{P}\{D(O;\theta*)\}}-\frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta*)\}]^2}\big[\mathbb{P}_n\{D(O;\widehat{\theta})\}-\mathbb{P}\{D(O;\theta^*)\}\big]-V(\pi)+o_p(n^{-1/2})\\
=&\frac{\mathbb{P}_n\{N(O;\theta^*)\}+\mathbb{P}\{\dot{N}(O;\theta^*)\}(\widehat{\theta}-\theta)}{\mathbb{P}\{D(O;\theta*)\}}-\frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta*)\}]^2}\big[\mathbb{P}_n\{D(O;\theta^*)\}+\mathbb{P}\{\dot{D}(O;\theta^*)\}(\widehat{\theta}-\theta)\big]+o_p(n^{-1/2})\\
=&\mathbb{P}_n\Bigg[\frac{N(O;\theta^*)}{\mathbb{P}\{D(O;\theta*)\}}-\frac{\mathbb{P}\{\dot{N}(O;\theta^*)\}}{\mathbb{P}\{D(O;\theta*)\}\mathbb{P}\{\dot{g}(O;\theta*)\}}g(O;\theta^*)-\frac{\mathbb{P}\{N(O;\theta^*)\}}{[\mathbb{P}\{D(O;\theta*)\}]^2\}}D(O;\theta^*)+\\
&\qquad\frac{\mathbb{P}\{\dot{D}(O;\theta^*)\}}{\mathbb{P}\{\dot{g}(O;\theta*)\}}g(O;\theta^*)\Bigg]+o_p(n^{-1/2}).
\end{aligned} \tag{B.27}
$$

By the central limit theorem (CLT), $\widehat{V}^{\text{MR}}(\pi)-V(\pi)$ is asymptotically normal. Under $\mathscr{M}_{\text{ps+cs+sp+om}}$, $\phi_\pi(\theta^*)=\phi_\pi$, $\psi_{S(0)}(\theta^*)=\psi_{S(0)}$, $\mathbb{P}\{N(O;\theta^*)\}=\mathbb{P}\big[p_0(X)\big[\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\big]\big]$, 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 (B.27),

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

This completes the proof. □

B.1.4 Proof of Theorem 7

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 5, 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 (B.24). Assumption 10(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

$$
\begin{aligned}
&\mathbb{P}_n\{N(O;\widehat{\theta})\}-\mathbb{P}\{N(O;\theta^*)\}=(\mathbb{P}_n-\mathbb{P})\{N(O;\widehat{\theta})\}+\mathbb{P}\{N(O;\widehat{\theta})-N(O;\theta^*)\}\\
=&(\mathbb{P}_n-\mathbb{P})\{N(O;\theta^*)\}+\mathbb{P}\{N(O;\widehat{\theta})-N(O;\theta^*)\}+o_p(n^{-1/2}),
\end{aligned}
\tag{B.28}
$$

where the second equality follows by Assumption 10(ii). It remains to analyze the second term $\mathbb{P}\{N(O;\widehat{\theta})-N(O;\theta^*)\}$.

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

$$
\begin{aligned}
&\left|\mathbb{P}\left[\pi(X)\widehat{\mu}_1(X)\frac{\widehat{p}_0(X)}{\widehat{p}_1(X)}\frac{\{\varphi_1(X)K_1(X)-\widehat{\varphi}_1(X)\widehat{K}_1(X)\}\{p_1(X)-\widehat{p}_1(X)\}}{\widehat{\varphi}_1(X)\widehat{K}_1(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\pi(X)\mu_1(X)\frac{\widehat{p}_0(X)}{\widehat{p}_1(X)}\frac{\{\varphi_1(X)K_1(X)-\widehat{\varphi}_1(X)\widehat{K}_1(X)\}\{p_1(X)-\widehat{p}_1(X)\}}{\widehat{\varphi}_1(X)\widehat{K}_1(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\pi(X)\widehat{p}_0(X)\frac{\{\varphi_1(X)K_1(X)-\widehat{\varphi}_1(X)\widehat{K}_1(X)\}\{\mu_1(X)-\widehat{\mu}_1(X)\}}{\widehat{\varphi}_1(X)\widehat{K}_1(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\{1-\pi(X)\}\mu_0(X)\frac{\{\varphi_0(X)K_0(X)-\widehat{\varphi}_0(X)\widehat{K}_0(X)\}\{p_0(X)-\widehat{p}_0(X)\}}{\widehat{\varphi}_0(X)\widehat{K}_0(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\{1-\pi(X)\}\widehat{p}_0(X)\frac{\{\varphi_0(X)K_0(X)-\widehat{\varphi}_0(X)\widehat{K}_0(X)\}\{\mu_0(X)-\widehat{\mu}_0(X)\}}{\widehat{\varphi}_0(X)\widehat{K}_0(X)}\right]\right|+\\
&\left|\mathbb{P}\left[\pi(X)\frac{p_0(X)}{\widehat{p}_1(X)}\{\mu_1(X)-\widehat{\mu}_1(X)\}\{p_1(X)-\widehat{p}_1(X)\}\right]\right|+\\
&\left|\mathbb{P}\left[\pi(X)\frac{p_1(X)}{\widehat{p}_1(X)}\{\mu_1(X)-\widehat{\mu}_1(X)\}\{p_0(X)-\widehat{p}_0(X)\}\right]\right|.
\end{aligned}
$$

By the Cauchy–Schwarz inequality and Assumption 10(iii), it follows that for some constant $l>0$, we have

$$
\begin{aligned}
&|\mathbb{P}\{N(O;\widehat{\theta})-N(O;\theta^*)\}|\\
\leq& l\times\|\varphi_0(X)K_0(X)-\widehat{\varphi}_0(X)\widehat{K}_0(X)\|_2\times\left[\|p_0(X)-\widehat{p}_0(X)\|_2+\|\mu_0(X)-\widehat{\mu}_0(X)\|_2\right]+\\
&l\times\|p_1(X)-\widehat{p}_1(X)\|_2\times\left[\|\varphi_1(X)K_1(X)-\widehat{\varphi}_1(X)\widehat{K}_1(X)\|_2+\|\mu_1(X)-\widehat{\mu}_1(X)\|_2\right]+\\
&l\times\|p_0(X)-\widehat{p}_0(X)\|_2\times\|\mu_0(X)-\widehat{\mu}_0(X)\|_2=o_p(-n^{1/2}).
\end{aligned}
\tag{B.29}
$$

Continuing with (B.28), 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 (B.24), we have

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

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

This completes the proof. □

B.1.5 Proof of Lemma 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 11(i), $V(\beta)$ is twice continuously differentiable at a neighborhood of β^*.

(2) In Section B.1.4, 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 $\pi(x;\widehat{\beta}) = \mathbb{I}\big(\tilde{x}^T\widehat{\beta} > 0\big)$ 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 $\beta = \beta^*$,

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

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

$$
V(\beta) - V(\beta^*) < c_0\|\beta - \beta^*\|_2^2. \tag{B.30}
$$

(2) From Section B.1.4, 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

$$\begin{aligned}
&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{MR}}(\beta) - V(\beta) - \{\widehat{V}^{\mathrm{MR}}(\beta^*) - V(\beta^*)\}\right|\right] \\
=&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\mathbb{P}_n\left\{\frac{\phi_{\pi(X;\beta)} - V(\beta)\psi_{S(0)}}{p_0}\right\} - \mathbb{P}_n\left\{\frac{\phi_{\pi(X;\beta^*)} - V(\beta^*)\psi_{S(0)}}{p_0}\right\}\right|\right] \\
=&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\mathbb{P}_n\left\{\frac{\phi_{\pi(X;\beta)} - V(\beta)p_0 + V(\beta)p_0 - V(\beta)\psi_{S(0)}}{p_0}\right\}\right.\right. \\
&\left.\left. - \mathbb{P}_n\left\{\frac{\phi_{\pi(X;\beta^*)} - V(\beta^*)p_0 + V(\beta^*)p_0 - V(\beta^*)\psi_{S(0)}}{p_0}\right\}\right|\right] \\
=&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\frac{\mathbb{P}_n\{\phi_{\pi(X;\beta)} - \phi_{\pi(X;\beta^*)}\}}{p_0} - \{V(\beta) - V(\beta^*)\}\right.\right. \\
&\left.\left. + \{V(\beta) - V(\beta^*)\}\left\{1 - \frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right] \\
\le&\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\frac{\mathbb{P}_n\{\phi_{\pi(X;\beta)} - \phi_{\pi(X;\beta^*)}\}}{p_0} - \{V(\beta) - V(\beta^*)\}\right|\right]}_{\tau_1} \\
&+\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\{V(\beta) - V(\beta^*)\}\left\{1 - \frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right]}_{\tau_2}.
\end{aligned} \tag{B.31}$$

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

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

where

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

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 8, 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 $\mathscr{F}_\beta(y,a,c,s,x)$ as $F=M\mathbb{I}(-k_0\varepsilon\le\tilde{x}^T\beta^*\le k_0\varepsilon)$. By Assumption 11(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 $\mathscr{F}_\beta$ is a class of indicator functions, by the conclusion of Lemma 9.6 and Lemma 9.9 in Kosorok (2008), $\mathscr{F}_\beta$ is a VC class of functions. Thus, the entropy of $\mathscr{F}_\beta$, denoted as $J^*_{[]}(1,\mathscr{F})$, is finite, i.e., $J^*_{[]}(1,\mathscr{F})<\infty$. Next, we consider the following empirical process indexed by β,

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

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

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

where l is a finite constant.

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

$$\tau_1 \le c_1\varepsilon^{1/2}. \tag{B.32}$$

$$\begin{aligned}\tau_2 =& \mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\{V(\beta)-V(\beta^*)\}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right]\\ =& \mathbb{E}^*\left[\left\{\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\{V(\beta)-V(\beta^*)\}\right|\right\}\left|n^{1/2}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right]\\ \le& \left[\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\{V(\beta)-V(\beta^*)\}\right|\right]\mathbb{E}^*\left[\left|n^{1/2}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right]\\ \le& O(\varepsilon),\end{aligned} \tag{B.33}$$

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 (B.31), (B.32), and (B.33), we have

$$\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{MR}}(\beta)-V(\beta)-\{\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\}\right|\right] \le \tau_1+\tau_2 \le c_1\varepsilon^{1/2}+O(\varepsilon) \le c_2\lambda(\varepsilon), \tag{B.34}$$

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 $n^{1/3}\|\widehat{\beta}-\beta^*\|_2 = O_p(1)$. □

B.1.6 Proof of Theorem 8

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^*,

$$\begin{aligned}\sqrt{n}\left\{V(\widehat{\beta})-V(\beta^*)\right\}&=\sqrt{n}\left\{V'(\beta^*)\|\widehat{\beta}-\beta^*\|_2+\frac{1}{2}V''(\beta^*)\|\widehat{\beta}-\beta^*\|_2^2+o_p\left(\|\widehat{\beta}-\beta^*\|_2^2\right)\right\}\\ &=\sqrt{n}\left\{\frac{1}{2}V''(\beta^*)\|\widehat{\beta}-\beta^*\|_2^2+o_p\left(\|\widehat{\beta}-\beta^*\|_2^2\right)\right\}\quad(V'(\beta^*)=0)\\ &=\sqrt{n}\left\{\frac{1}{2}V''(\beta^*)O_p(n^{-2/3})+o_p(n^{-2/3})\right\}\quad(\text{Lemma 1})\\ &=\frac{1}{2}V''(\beta^*)O_p(n^{-1/6})=o_p(1).\end{aligned}\tag{B.35}$$

By Lemma 1, we have $\|\widehat{\beta}-\beta^*\|_2=c_3n^{-1/3}$, for some constant $0<c_3<\infty$. From (B.34), we have

$$\begin{aligned}&\sqrt{n}\left[\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-\widehat{V}^{\mathrm{MR}}(\beta^*)\right\}-\left\{V(\widehat{\beta})-V(\beta^*)\right\}\right]\\ \le&\mathbb{E}^*\left[n^{1/2}\sup_{\|\widehat{\beta}-\beta^*\|_2<c_3n^{-1/3}}\left|\widehat{V}^{\mathrm{MR}}(\beta)-V(\beta)-\{\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\}\right|\right]\\ \le&c_2\left(c_3^{1/2}n^{-1/6}+c_3n^{-1/3}\right)=o_p(1).\end{aligned}\tag{B.36}$$

By (B.35) and (B.36), 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 7, $\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)$ is asymptotically normal with mean zero and variance $\Upsilon(\pi(x;\beta^*))$. This completes the proof. □

B.2 Additional Theoretical Properties for OPL

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

Lemma 8 *Suppose that Assumptions 4–8, 9, and 11 hold, under* $\mathscr{M}_{\mathrm{ps+cs+sp}}\cup\mathscr{M}_{\mathrm{ps+cs+om}}\cup\mathscr{M}_{\mathrm{sp+om}}$, *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 11(i), $V(\beta)$ is twice continuously differentiable at a neighborhood of β^*.

(2) In Section B.1.3, 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 $\beta=\beta^*$,

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

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 $\mathscr{M}_{\mathrm{ps+cs+sp}}\cup\mathscr{M}_{\mathrm{ps+cs+om}}\cup\mathscr{M}_{\mathrm{sp+om}}$, $\mathbb{P}\{N(O;\theta^*,\beta)/\mathbb{P}\{D(O;\theta^*)\}=V(\beta)$, $\mathbb{P}\{D(O;\theta^*)\}=p_0$. By (B.27), 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

$$
\begin{aligned}
&\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{MR}}(\beta)-V(\beta)-\{\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\}\right|\right]\\
=&\mathbb{E}^*\Bigg[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\Bigg|\frac{\mathbb{P}_n\{\phi_{\pi(X;\beta)}(\theta^*)-\phi_{\pi(X;\beta^*)}(\theta^*)\}}{p_0}-\{V(\beta)-V(\beta^*)\}\\
&+\{V(\beta)-V(\beta^*)\}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}(\theta^*)\}}{p_0}\right\}-\frac{\mathbb{P}\{\dot{N}(O;\theta^*,\beta)\}-\mathbb{P}\{\dot{N}(O;\theta^*,\beta^*)\}}{p_0\mathbb{P}\{\dot{g}(O;\theta^*)\}}\mathbb{P}_n\{g(O;\theta^*)\}\Bigg|\Bigg]\\
\le&\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\frac{\mathbb{P}_n\{\phi_{\pi(X;\beta)}-\phi_{\pi(X;\beta^*)}\}}{p_0}-\{V(\beta)-V(\beta^*)\}\right|\right]}_{\tau_1}\\
&+\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\{V(\beta)-V(\beta^*)\}\left\{1-\frac{\mathbb{P}_n\{\psi_{S(0)}\}}{p_0}\right\}\right|\right]}_{\tau_2}\\
&+\underbrace{\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\frac{\mathbb{P}\{\dot{N}(O;\theta^*,\beta)\}-\mathbb{P}\{\dot{N}(O;\theta^*,\beta^*)\}}{p_0\mathbb{P}\{\dot{g}(O;\theta^*)\}}\mathbb{P}_n\{g(O;\theta^*)\}\right|\right]}_{\tau_3}.
\end{aligned}
\tag{B.37}
$$

We can show similarly as in Section B.1.5, $\tau_1+\tau_2<c_1(\varepsilon^{1/2}+\varepsilon)$ 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 (B.37), we have

$$\mathbb{E}^*\left[n^{1/2}\sup_{\|\beta-\beta^*\|_2<\varepsilon}\left|\widehat{V}^{\mathrm{MR}}(\beta)-V(\beta)-\{\widehat{V}^{\mathrm{MR}}(\beta^*)-V(\beta^*)\}\right|\right]\leq\tau_1+\tau_2+\tau_3\leq c_1(\varepsilon^{1/2}+\varepsilon)+O(\varepsilon)\leq c_2\lambda(\varepsilon), \tag{B.38}$$

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 14 *Suppose that Assumptions 4–8, 9, and 11 hold, under* $\mathscr{M}_{\text{ps+cs+sp}}\cup\mathscr{M}_{\text{ps+cs+om}}\cup\mathscr{M}_{\text{sp+om}}$, $\widehat{V}^{\mathrm{MR}}(\widehat{\beta})-V(\beta)$ *is asymptotically normal with mean zero. Moreover, under* $\mathscr{M}_{\text{ps+cs+sp+om}}$, $\widehat{V}^{\mathrm{MR}}(\widehat{\beta})$ *achieves the semiparametric efficiency bound* $\Upsilon(\pi(x;\beta^*))$.

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^*,

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

$$= \frac{1}{2} V''(\beta^*) O_p(n^{-1/6}) = o_p(1). \tag{B.39}$$

By Lemma 1, we have $\|\widehat{\beta} - \beta^*\|_2 = c_3 n^{-1/3}$, for some constant $0 < c_3 < \infty$. From (B.38), we have

$$\begin{aligned}
&\sqrt{n}\left[\left\{\widehat{V}^{\mathrm{MR}}(\widehat{\beta}) - \widehat{V}^{\mathrm{MR}}(\beta^*)\right\} - \left\{V(\widehat{\beta}) - V(\beta^*)\right\}\right] \\
\leq &\mathbb{E}^*\left[n^{1/2} \sup_{\|\widehat{\beta}-\beta^*\|_2 < c_3 n^{-1/3}} \left|\widehat{V}^{\mathrm{MR}}(\beta) - V(\beta) - \{\widehat{V}^{\mathrm{MR}}(\beta^*) - V(\beta^*)\}\right|\right] \\
\leq &c_2\left(c_3^{1/2} n^{-1/6} + c_3 n^{-1/3}\right) = o_p(1).
\end{aligned} \tag{B.40}$$

By (B.39) and (B.40), 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 $\widehat{V}^{\mathrm{MR}}(\beta^*) - V(\beta^*)$. From Section B.1.3, $\widehat{V}^{\mathrm{MR}}(\beta^*) - V(\beta^*)$ is asymptotically normal with mean zero. Under $\mathscr{M}_{\mathrm{ps+cs+sp+om}}$, $\widehat{V}^{\mathrm{MR}}(\widehat{\beta})$ achieves the semiparametric efficiency bound $\Upsilon(\pi(x;\beta^*))$. This completes the proof. □

B.3 Sensitivity Analysis

Assumption 7 may be violated if there are latent confounders between the principal strata and outcome. For the sensitivity analysis, we assume the following tilting model:

$$\frac{\mathbb{E}[Y(1) \mid U = 10, X]}{\mathbb{E}[Y(1) \mid U = 11, X]} = \rho(X). \tag{B.41}$$

Then, we can use $\rho(X)$ as the sensitivity parameter. If $\rho(X) = 1$, then Assumptions 7 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 15 *Let Π be a policy class. Under Assumptions 4–6, 8, and (B.41) with known $\rho(X)$, for any given policy $\pi \in \Pi$, the survivor value function $V(\pi)$ is identified,*

$$V(\pi) = \mathbb{E}\left[\frac{p_0(X)}{p_0}\left[\omega(X)\mu_1(X)\pi(X) + \mu_0(X)\{1-\pi(X)\}\right]\right]. \tag{B.42}$$

Proof.

$$\begin{aligned}
\mu_1(X) =& \mathbb{E}[Y \mid A=1, C=0, S=1, X] \\
=& \mathbb{E}[Y \mid A=1, C=0, S=1, U=11, X] P(U=11 \mid A=1, C=0, S=1, X) \\
&+ \mathbb{E}[Y \mid A=1, C=0, S=1, U=10, X] P(U=10 \mid A=1, C=0, S=1, X) \\
=& \mathbb{E}[Y(1) \mid A=1, C(1)=0, S(1)=1, U=11, X] P(U=11 \mid A=1, C=0, S=1, X) \\
&+ \mathbb{E}[Y(1) \mid A=1, C(1)=0, S(1)=1, U=10, X] P(U=10 \mid A=1, C=0, S=1, X) \quad \text{(SUTVA)} \\
=& \mathbb{E}[Y(1) \mid C(1)=0, S(1)=1, U=11, X] P(U=11 \mid A=1, C=0, S=1, X) \\
&+ \mathbb{E}[Y(1) \mid C(1)=0, S(1)=1, U=10, X] P(U=10 \mid A=1, C=0, S=1, X) \quad \text{(Assumption 4)} \\
=& \mathbb{E}[Y(1) \mid S(1)=1, U=11, X] P(U=11 \mid A=1, C=0, S=1, X) \\
&+ \mathbb{E}[Y(1) \mid S(1)=1, U=10, X] P(U=10 \mid A=1, C=0, S=1, X) \quad \text{(Assumption 5)} \\
=& \mathbb{E}[Y(1) \mid U=11, X] P(U=11 \mid A=1, C=0, S=1, X) \\
&+ \mathbb{E}[Y(1) \mid U=10, X] P(U=10 \mid A=1, C=0, S=1, X) \\
=& \mathbb{E}[Y(1) \mid U=11, X] \frac{p_0(X)}{p_1(X)} + \mathbb{E}[Y(1) \mid U=10, X] \frac{p_1(X)-p_0(X)}{p_1(X)} \\
=& \mathbb{E}[Y(1) \mid U=11, X] \frac{p_0(X)}{p_1(X)} + \rho(X)\mathbb{E}[Y(1) \mid U=11, X] \frac{p_1(X)-p_0(X)}{p_1(X)} \quad \text{(by (B.41))} \\
=& \frac{\mathbb{E}[Y(1) \mid U=11, X]}{\omega(X)}.
\end{aligned} \tag{B.43}$$

$$\begin{aligned}
\mu_0(X) =& \mathbb{E}[Y \mid A=0, C=0, S=1, X] \\
=& \mathbb{E}[Y(0) \mid A=1, C(0)=0, S(0)=1, X] \quad \text{(SUTVA)} \\
=& \mathbb{E}[Y(0) \mid C(0)=0, S(0)=1, X] \quad \text{(Assumption 4)} \\
=& \mathbb{E}[Y(0) \mid S(0)=1, X] \quad \text{(Assumption 5)} \\
=& \mathbb{E}[Y(0) \mid U=11, X]. \quad \text{(Assumption 6)}
\end{aligned} \tag{B.44}$$

$$
\begin{aligned}
V(\pi) &= \mathbb{E}[Y(1)\pi(X)+Y(0)\{1-\pi(X)\} \mid U=11] \\
&= \mathbb{E}[\mathbb{E}[Y(1)\pi(X)+Y(0)\{1-\pi(X)\} \mid U=11, X] \mid U=11] \\
&= \mathbb{E}[\mathbb{E}[Y(1) \mid U=11, X]\pi(X) \mid U=11]+\mathbb{E}[\mathbb{E}[Y(0) \mid U=11, X]\{1-\pi(X)\} \mid U=11] \\
&= \mathbb{E}\big[\omega(X)\mu_1(X)\pi(X) \mid U=11\big] \quad \text{(by (B.43))} + \mathbb{E}\big[\mu_0(X)\{1-\pi(X)\} \mid U=11\big] \quad \text{(by (B.44))} \\
&= \mathbb{E}\bigg[\frac{p_0(X)}{p_0}\big[\omega(X)\mu_1(X)\pi(X)+\mu_0(X)\{1-\pi(X)\}\big].\bigg] \quad \text{(by (B.2))}
\end{aligned}
$$

Starting from the identification formula (B.42), we can derive the EIF for $V(\pi)$ using the same technique in Section B.1.2. Therefore, we directly give the following result and omit the proof.

Theorem 16 *Suppose $V(\pi)$ is identified in Theorem 15. 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

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

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

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

$$
\widehat{V}(\pi) = \frac{\mathbb{P}_n(\widehat{\phi'}_\pi)}{\mathbb{P}_n\{\widehat{\psi}_{S(0)}\}}, \tag{B.45}
$$

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

Theorem 17 *Suppose that Assumptions 4-6, 8, 10, and (B.41) with known $\rho(X)>0$ hold. $\widehat{V}(\pi)$ in (B.45) has the influence function ν'_π and therefore achieves the semiparametric efficiency bound $\Upsilon'(\pi)$.*

The proof of Theorem 17 is similar to that of Theorem 7 and thus omitted.

We re-visit the real data application in Section 3.6.2 to assess the violation of Assumption 7. 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 3.6.2. 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 B.1. We find that the result is not sensitive to the violation of Assumption 7.

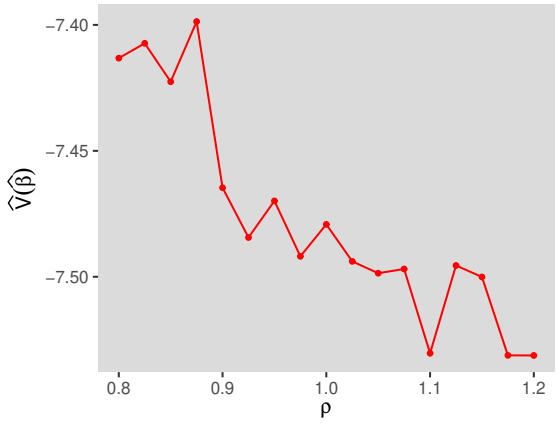

Figure B.1: Sensitivity analysis for real data application in Section 3.6.2.

B.4 Additional OPE Experiment Results

We report the OPE experiment results with sample size $n = 1000$ in Table B.1. The other setting is the same as Section 3.6.1.

Table B.1: 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

APPENDIX

C

SUPPLEMENT TO CHAPTER 4

C.1 Technical Proofs

C.1.1 Proof of Theorem 9

Proof.

$$
\begin{aligned}
&\mathbb{E}\{Y(1) \mid X = x\} \\
=&\mathbb{E}\{Y(1) \mid X = x, A = 1\} w(x) + \mathbb{E}\{Y(1) \mid X = x, A = 0\}\{1 - w(x)\} \\
=&w(x)\left\{\int y f(y \mid x, 1) d y\right\} + \{1 - w(x)\}\left\{\int y f(y \mid x, 0) d y\right\} \\
=&w(x)\left\{\int y f(y \mid x, 1) d y\right\} + \left\{\int y\{1 - w(x)\} f(y \mid x, 0) d y\right\} \\
=&w(x)\left\{\int y f(y \mid x, 1) d y\right\} + \left\{\int y f(y \mid x, 1)\left[\frac{f(y \mid x, 0)\{1 - w(x)\}}{f(y \mid x, 1)}\right] d y\right\} \\
=&w(x)\left\{\int y f(y \mid x, 1) d y\right\} + \left\{\int y f(y \mid x, 1)\left[w(x)\left\{\frac{1}{\varphi(x, y)} - 1\right\}\right] d y\right\} \\
=&w(x)\left\{\int y f(y \mid x, 1) d y\right\} + w(x)\left\{\int y f(y \mid x, 1)\left[\left\{\frac{1}{\varphi(x, y)} - 1\right\}\right] d y\right\}
\end{aligned}
$$

$$= w(x) \int y \frac{f(y \mid x,1)}{\varphi(x,y)} dy.$$

Therefore,

$$\begin{aligned} & V_1(\pi) = \mathbb{E}\{Y(1)\pi(X)\} \\ =& \mathbb{E}(\mathbb{E}[\{Y(1)\pi(X)\} \mid X]) \\ =& \int f(x)\pi(x)\mathbb{E}\{Y(1) \mid X = x\} dx \\ =& \int f(x) w(x) \left\{ \int y \frac{f(y \mid x,1)}{\varphi(x,y)} dy \right\} \pi(x) dx. \end{aligned}$$

To identify $V(\pi)$, we need to identify $f(x)$, $w(x)$, $f(y|x,1)$, and $\varphi(x,y)$. The likelihood function for a single observation is

$$f(x) w(x)^a \{1 - w(x)\}^{1-a} f(y \mid x,1)^a.$$

A key observation is that

$$w(x)^{-1} = \int \frac{f(y|x,1)}{\varphi(x,y)} dy.$$

Under Assumption 12, $\varphi(x,y) = \mathbb{P}\{A = 1 \mid X = x, Y(1) = y\} = \mathbb{P}\{A = 1 \mid U = u, Y(1) = y\} = \varphi(u,y)$, and the likelihood function becomes

$$f(x) \left\{ \int \frac{f(y|x,1)}{\varphi(u,y)} dy \right\}^{-a} \left[1 - \left\{ \int \frac{f(y|x,1)}{\varphi(u,y)} dy \right\}^{-1} \right]^{1-a} f(y|x,1)^a.$$

Assume we have two different sets of models $f(x)$, $f(y \mid x,1)$, $\varphi(u,y)$, and $\tilde{f}(x)$, $\tilde{f}(y \mid x,1)$, $\tilde{\varphi}(u,y)$, such that

$$\begin{aligned} & f(x) \left\{ \int \frac{f(y|x,1)}{\varphi(u,y)} dy \right\}^{-a} \left[1 - \left\{ \int \frac{f(y|x,1)}{\varphi(u,y)} dy \right\}^{-1} \right]^{1-a} f(y|x,1)^a \\ =& \tilde{f}(x) \left\{ \int \frac{\tilde{f}(y|x,1)}{\tilde{\varphi}(u,y)} dy \right\}^{-a} \left[1 - \left\{ \int \frac{\tilde{f}(y|x,1)}{\tilde{\varphi}(u,y)} dy \right\}^{-1} \right]^{1-a} \tilde{f}(y|x,1)^a. \end{aligned} \tag{C.1}$$

Taking $a = 0$ in (C.1), we have

$$f(x) \left[1 - \left\{ \int \frac{f(y|x,1)}{\varphi(u,y)} dy \right\}^{-1} \right] = \tilde{f}(x) \left[1 - \left\{ \int \frac{\tilde{f}(y|x,1)}{\tilde{\varphi}(u,y)} dy \right\}^{-1} \right]. \tag{C.2}$$

Taking $a=1$ and taking integration with respect to $Y(1)$ on both sides of the above equation, we have

$$f(x)\left\{\int \frac{f(y|x,1)}{\varphi(u,y)}dy\right\}^{-1} = \tilde{f}(x)\left\{\int \frac{\tilde{f}(y|x,1)}{\tilde{\varphi}(u,y)}dy\right\}^{-1}. \tag{C.3}$$

By Equations (C.2) and (C.3), we have

$$f(x)=\tilde{f}(x) \quad \text{and} \quad \int \frac{f(y|x,1)}{\varphi(u,y)}dy = \int \frac{\tilde{f}(y|x,1)}{\tilde{\varphi}(u,y)}dy.$$

Taking $a=1$ in (C.1), we have

$$f(x)\left\{\int \frac{f(y|x,1)}{\varphi(u,y)}dy\right\}^{-1} f(y|x,1) = \tilde{f}(x)\left\{\int \frac{\tilde{f}(y|x,1)}{\tilde{\varphi}(u,y)}dy\right\}^{-1} \tilde{f}(y|x,1).$$

Thus, we have

$$f(y|x,1)=\tilde{f}(y|x,1).$$

Finally, from

$$\int \frac{f(y|x,1)}{\varphi(u,y)}dy = \int \frac{f(y|x,1)}{\tilde{\varphi}(u,y)}dy,$$

and Assumption 13, we have

$$\varphi(u,y)=\tilde{\varphi}(u,y).$$

Thus, $f(x)$, $w(x)$, $f(y|x,1)$, and $\varphi(x,y)$ are all identified. The value function $V_1(\pi)$ is then identified. □

C.1.2 Proof of Theorem 10

Proof. Let $O=\{AY,A,X\}$ summarize the vector of observed variables with the likelihood factorized as

$$f(O)=f(X)w(X)^A\{1-w(X)\}^{1-A}f(Y\mid X,A=1)^A.$$

We consider a one-dimensional parametric submodel $f_{\theta_1}(X)$ for $f(X)$, and a one-dimensional parametric submodel $f_{\theta_2}(Y\mid X,A=1)$ for $f(Y\mid X,A=1)$, respectively. The submodel $f_{\theta_1}(X)$ contains the true model $f(X)$ at $\theta_1=0$, i.e., $f_{\theta_1}(X)\,|_{\theta_1=0}= f(X)$. Similarly, the submodel $f_{\theta_2}(Y\mid X,A=1)$ contains the true model $f(Y\mid X,A=1)$ at $\theta_2=0$, i.e., $f_{\theta_2}(Y\mid X,A=1)\,|_{\theta_2=0}= f(Y\mid$

$X, A=1$). The submodel for the likelihood can be represented as

$$f_{\theta_1,\theta_2}(O) = f_{\theta_1}(X) w_{\theta_2}(X)^A \{1-w_{\theta_2}(X)\}^{1-A} f_{\theta_2}(Y \mid X, A=1)^A.$$

$$\begin{aligned}
\frac{\partial \log f_{\theta_1,\theta_2}(O)}{\partial \theta_1} &= \frac{\partial \log f_{\theta_1}(X)}{\partial \theta_1}, \\
\frac{\partial \log f_{\theta_1,\theta_2}(O)}{\partial \theta_2} &= A\frac{\partial \log f_{\theta_2}(Y \mid X, A=1)}{\partial \theta_2} + \frac{w_{\theta_2}(X)-A}{1-w_{\theta_2}(X)}\mathbb{E}\left\{\frac{\partial \log f_{\theta_2}(Y \mid X, A=1)}{\partial \theta_2} \mid X\right\}.
\end{aligned}$$

By the semiparametric theory (Bickel et al. 1993; Tsiatis 2006), we have the nuisance tangent spaces

$$\begin{aligned}
\Lambda_1 &= [h_1(X) : \mathbb{E}\{h_1(X)=0\}], \\
\Lambda_2 &= \left[Ah_2(X, Y(1)) + \frac{w(X)-A}{1-w(X)}\mathbb{E}\{h_2(X, Y(1)) \mid X\} : \mathbb{E}\{h_2(X, Y(1)) \mid X, A=1\}=0\right].
\end{aligned}$$

It is easy to verify that $\Lambda_1 \perp \Lambda_2$. Consider a generic mean zero element in $\Lambda_\perp$, $Ag_1(X, Y(1)) + (1-A)g_2(X)$. Since $\Lambda_1 \perp \Lambda_\perp$, for any measurable mean zero function $h_1(X)$, we have

$$\begin{aligned}
&\mathbb{E}[\{Ag_1(X, Y(1)) + (1-A)g_2(X)\}h_1(X)] \\
=&\mathbb{E}(\mathbb{E}[\{Ag_1(X, Y(1)) + (1-A)g_2(X)\}h_1(X) \mid X]) \\
=&\mathbb{E}([w(X)\mathbb{E}\{g_1(X, Y(1)) \mid X, A=1\} + \{1-w(X)\}g_2(X)]h_1(X)) \\
=&0.
\end{aligned}$$

Therefore, $w(X)\mathbb{E}\{g_1(X, Y(1)) \mid X, A=1\} + \{1-w(X)\}g_2(X)$ is a constant and we denote it as c. Since $Ag_1(X, Y(1)) + (1-A)g_2(X)$ is mean zero, we have

$$\begin{aligned}
&\mathbb{E}\{Ag_1(X, Y(1)) + (1-A)g_2(X)\} \\
=&\mathbb{E}[w(X)\mathbb{E}\{g_1(X, Y(1)) \mid X, A=1\} + \{1-w(X)\}g_2(X)] \\
=&\mathbb{E}(c) = 0.
\end{aligned}$$

Therefore, we have

$$w(X)\mathbb{E}\{g_1(X, Y(1)) \mid X, A=1\} + \{1-w(X)\}g_2(X) = 0. \tag{C.4}$$

Since $\Lambda_2 \perp \Lambda_\perp$, we have

$$
\begin{aligned}
&\mathbb{E}\left(\{Ag_1(X,Y(1))+(1-A)g_2(X)\}\left[Ah_2(X,Y(1))+\frac{w(X)-A}{1-w(X)}\mathbb{E}\{h_2(X,Y(1))\mid X\}\right]\right)\\
=&\mathbb{E}\left[w(X)\mathbb{E}\{g_1(X,Y(1))h_2(X,Y(1))\mid X,A=1\}+g_2(X)\mathbb{E}\{h_2(X,Y(1))\mid X\}\right]\\
=&\mathbb{E}\left[w(X)\mathbb{E}\{g_1(X,Y(1))h_2(X,Y(1))\mid X,A=1\}+w(X)g_2(X)\mathbb{E}\left\{\frac{h_2(X,Y(1))}{\varphi(\eta)}\mid X,A=1\right\}\right]\\
=&\mathbb{E}\left(\mathbb{E}\left[w(X)\left\{g_1(X,Y(1))+\frac{g_2(X)}{\varphi(\eta)}\right\}h_2(X,Y(1))\mid X,A=1\right]\right)\\
=&0.
\end{aligned}
$$

Therefore, $g_1(X,Y(1))+\frac{g_2(X)}{\varphi(\eta)}$ is a function of X and we denote it as $k(X)$:

$$k(X)=g_1(X,Y(1))+\frac{g_2(X)}{\varphi(\eta)}.$$

Taking the conditional expectation on both sides, and by (C.4), we have

$$k(X)=\mathbb{E}\{g_1(X,Y(1))\mid X,A=1\}+\frac{g_2(X)}{w(X)}=g_2(X).$$

Therefore, we have

$$g_2(X)=g_1(X,Y(1))+\frac{g_2(X)}{\varphi(\eta)}.$$

Thus,

$$Ag_1(X,Y(1))+(1-A)g_2(X)=\frac{\varphi(\eta)-A}{\varphi(\eta)}g_1(X),$$

and $\Lambda_\perp=\left\{\frac{\varphi(\eta)-A}{\varphi(\eta)}g_1(X)\right\}$. This completes the proof. □

C.1.3 Proof of Theorem 11

Proof. The score function for η is

$$S_\eta=\frac{A-w(X)}{1-w(X)}\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)}\mid X\right\}.$$

The efficient score for η is the projection of the score function S_η onto the space $\Lambda_\perp$. Notice that $S_\eta \perp \Lambda_1$. Therefore, we can write

$$\frac{A-w(X)}{1-w(X)}\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)} \mid X\right\} = \underbrace{Ab(X,Y(1)) + \frac{w(X)-A}{1-w(X)}\mathbb{E}\{b(X,Y(1)) \mid X\}}_{\in\Lambda_2} + \underbrace{\frac{\varphi(\eta)-A}{\varphi(\eta)}c(X)}_{\Lambda_\perp}, \quad \text{(C.5)}$$

where $\mathbb{E}\{b(X,Y(1)) \mid X, A=1\} = 0$. Let $A=1$ in (C.5), we have

$$\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)} \mid X\right\} = b(X,Y(1)) - \mathbb{E}\{b(X,Y(1)) \mid X\} + \frac{\varphi(\eta)-1}{\varphi(\eta)}c(X).$$

By taking $\mathbb{E}(\cdot \mid X)$ on both sides, we have

$$c(X) = \frac{\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)} \mid X\right\}}{1-\mathbb{E}\left\{\frac{1}{\varphi(\eta)} \mid X\right\}} = \frac{\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2} \mid X, A=1\right\}}{\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2} \mid X, A=1\right\}}.$$

Therefore,

$$S_{\eta,\text{eff}} = \frac{\varphi(\eta)-A}{\varphi(\eta)} \frac{\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2} \mid X, A=1\right\}}{\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2} \mid X, A=1\right\}}.$$

Let $A=0$ in (C.5), we can further derive that

$$b(X,Y(1)) = \left\{\frac{1}{\varphi(\eta)} - \frac{1}{w(X)}\right\} c(X).$$

□

C.1.4 Proof of Theorem 12

Proof. We consider a one-dimensional parametric submodel $f_\alpha(X)$ for $f(X)$, and a one-dimensional parametric submodel $f_\beta(Y \mid X, A=1)$ for $f(Y \mid X, A=1)$, respectively. The submodel $f_\alpha(X)$ contains the true model $f(X)$ at $\alpha=\alpha_0$, i.e., $f_{\alpha_0}(X) = f(X)$. Similarly, the submodel $f_\beta(Y \mid X, A=1)$ contains the true model $f(Y \mid X, A=1)$ at $\beta=\beta_0$, i.e., $f_{\beta_0}(Y \mid X, A=1) = f(Y \mid X, A=1)$. Let $\theta=(\alpha,\beta)$. The submodel for the likelihood can be represented as

$$f_{\theta,\eta}(O) = f_\alpha(X)\{w_{\beta,\eta}(X)\}^A f_\beta(Y|X, A=1)\{1-w_{\beta,\eta}(X)\}^{1-A},$$

which contains the true model at $\theta_0=(\alpha_0,\beta_0)$. For the ease of exposition, we write $V_1(\pi)$ as $V(\pi)$. 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.

Let

$$S_{\alpha_0} = \left.\frac{\partial \log f_\theta(O)}{\partial \alpha}\right|_{\theta=\theta_0} = \left.\frac{\partial \log f_\alpha(X)}{\partial \alpha}\right|_{\alpha=\alpha_0},$$
$$S_{\beta_0} = \left.\frac{\partial \log f_\theta(O)}{\partial \beta}\right|_{\theta=\theta_0} = A\left.\frac{\partial \log f_\beta(Y|X,A=1)}{\partial \beta}\right|_{\beta=\beta_0} + \frac{w(X)-A}{1-w(X)}\mathbb{E}\left\{\left.\frac{\partial \log f_\beta(Y|X,A=1)}{\partial \beta}\right|_{\beta=\beta_0} \mid X\right\},$$
$$S_\eta = \left.\frac{\partial \log f_\theta(O)}{\partial \eta}\right|_{\theta=\theta_0} = \frac{A-w(X)}{1-w(X)}\mathbb{E}\left\{\frac{\partial \log \varphi(\eta)}{\partial \eta} \mid X\right\}.$$

Let $s_{\beta_0} = \left.\frac{\partial \log f_\beta(Y|X,A=1)}{\partial \beta}\right|_{\beta=\beta_0}$ and $s_\eta = \frac{\partial \log \varphi(\eta)}{\partial \eta}$.

By the semiparametric theory, the EIF for $V(\pi)$ must have the form

$$\phi_{\text{eff}} = \underbrace{h_1^*(X)}_{\in\Lambda_1} + \underbrace{Ah_2^*(X) + \frac{w(X)-A}{1-w(X)}\mathbb{E}\{h_2^*(X, Y(1)) \mid X\}}_{\in\Lambda_2} + \underbrace{D^T S_{\eta,\text{eff}}}_{\in\Lambda_\perp},$$

where $\mathbb{E}\{h_1^*(X)=0\}, \mathbb{E}\{h_2^*(X, Y(1)) \mid X, A=1\}=0$, and D is a vector with the same dimension as η. The EIF ϕ_{eff} for $V(\pi)$ must satisfy

$$\partial V_\theta(\pi)/\partial \alpha|_{\theta=\theta_0} = \mathbb{E}(\phi_{\text{eff}} S_{\alpha_0}),$$
$$\partial V_\theta(\pi)/\partial \beta|_{\theta=\theta_0} = \mathbb{E}(\phi_{\text{eff}} S_{\beta_0}),$$
$$\partial V_\theta(\pi)/\partial \eta|_{\theta=\theta_0} = \mathbb{E}(\phi_{\text{eff}} S_\eta).$$

(I)

$$\partial V_\theta(\pi)/\partial \alpha\,|_{\theta=\theta_0} = \mathbb{E}\left[\pi(X)w(X)\mathbb{E}\left\{\frac{Y}{\varphi(\eta)} \mid X, A=1\right\} S_{\alpha_0}\right],$$
$$\mathbb{E}(\phi_{\text{eff}} S_{\alpha_0}) = \mathbb{E}\{h_1^*(X) S_{\alpha_0}\}.$$

We have

$$h_1^*(X) = \pi(X)w(X)\mathbb{E}\left\{\frac{Y}{\varphi(\eta)} \mid X, A=1\right\} - V(\pi).$$

(II)

$$\partial V_\theta(\pi)/\partial \beta\,|_{\theta=\theta_0} = \mathbb{E}\left[\pi(X)\{Y(1) - \mathbb{E}(Y(1)|X)\} s_{\beta_0}\right],$$
$$\mathbb{E}(\phi_{\text{eff}} S_{\beta_0}) = \mathbb{E}\left(\left[\varphi(\eta)h_2^*(X, Y(1)) + \frac{w(X)}{1-w(X)}\mathbb{E}\{h_2^*(X, Y(1)) \mid X\}\right] s_{\beta_0}\right).$$

$$
\begin{aligned}
&\partial V_\theta(\pi)/\partial\beta\,|_{\theta=\theta_0} - \mathbb{E}(\phi_{\text{eff}} S_{\beta_0}) \\
=&\mathbb{E}\left(\left[\varphi(\eta)h_2^*(X,Y(1)) + \frac{w(X)}{1-w(X)}\mathbb{E}\{h_2^*(X,Y(1))\mid X\} - \pi(X)\{Y(1)-\mathbb{E}\{Y(1)|X\}\right]s_{\beta_0}\right) \\
=&\mathbb{E}\left\{\mathbb{E}\left(\left[h_2^*(X,Y(1)) + \frac{w(X)}{1-w(X)}\frac{\mathbb{E}\{h_2^*(X,Y(1))\}\mid X\}}{\varphi(\eta)} - \pi(X)\frac{Y(1)-\mathbb{E}\{Y(1)|X\}}{\varphi(\eta)}\right]\varphi(\eta)s_{\beta_0}\right)\mid X\right\}.
\end{aligned}
$$

Since $\mathbb{E}\{\varphi(\eta)s_{\beta_0}\mid X\}=0$, $h_2^*(X,Y(1)) + \frac{w(X)}{1-w(X)}\frac{\mathbb{E}\{h_2^*(X,Y(1))\}|X\}}{\varphi(\eta)} - \pi(X)\frac{Y(1)-\mathbb{E}\{Y(1)|X\}}{\varphi(\eta)}$ must be a function of X and we denote it as $m(X)$:

$$
m(X) = h_2^*(X,Y(1)) + \frac{w(X)}{1-w(X)}\frac{\mathbb{E}\{h_2^*(X,Y(1))\}\mid X\}}{\varphi(\eta)} - \pi(X)\frac{Y(1)-\mathbb{E}\{Y(1)|X\}}{\varphi(\eta)}. \tag{C.6}
$$

Taking the conditional expectation on both sides, we have

$$
m(X) = \frac{\mathbb{E}\{h_2^*(X,Y(1))\mid X\}}{1-w(X)}.
$$

Therefore, we have

$$
\frac{\mathbb{E}\{h_2^*(X,Y(1))\mid X\}}{1-w(X)} = h_2^*(X,Y(1)) + \frac{w(X)}{1-w(X)}\frac{\mathbb{E}\{h_2^*(X,Y(1))\}\mid X\}}{\varphi(\eta)} - \pi(X)\frac{Y(1)-\mathbb{E}\{Y(1)|X\}}{\varphi(\eta)}.
$$

Taking $\mathbb{E}(\cdot\mid X)$ on both sides,

$$
\begin{aligned}
&\frac{\mathbb{E}\{h_2^*(X,Y(1))\mid X\}}{1-w(X)} \\
=&\mathbb{E}\{h_2^*(X,Y(1))\mid X\} + \frac{w(X)}{1-w(X)}\mathbb{E}\{h_2^*(X,Y(1))\mid X\}\mathbb{E}\{1/\varphi(\eta)\mid X\} \\
&-\pi(X)\big[\mathbb{E}\{Y(1)/\varphi(\eta)\mid X\} - \mathbb{E}\{Y(1)\mid X\}\mathbb{E}\{1/\varphi(\eta)\mid X\}\big].
\end{aligned}
$$

We have

$$
\mathbb{E}\{h_2^*(X,Y(1))\mid X\} = \pi(X)\frac{1-w(X)}{w(X)}\frac{\mathbb{E}\left\{Y(1)/\varphi(\eta)\mid X\right\} - \mathbb{E}\{Y(1)\mid X\}\mathbb{E}\{1/\varphi(\eta)\mid X\}}{\mathbb{E}\{1/\varphi(\eta)\mid X\}-1}. \tag{C.7}
$$

By Equations (C.6) and (C.7),

$$
h_2^*(X,Y(1)) = \pi(X)\left[\left\{\frac{1}{w(X)} - \frac{1}{\varphi(\eta)}\right\}\frac{\mathbb{E}\left\{\frac{Y(1)}{\varphi(\eta)}\mid X\right\} - \mathbb{E}\{Y(1)\mid X\}\mathbb{E}\left\{\frac{1}{\varphi(\eta)}\mid X\right\}}{\mathbb{E}\{1/\varphi(\eta)\mid X\}-1} + \frac{Y(1)-\mathbb{E}\{Y(1)\mid X\}}{\varphi(\eta)}\right].
$$

(III)

$$\partial V_\theta(\pi)/\partial\eta|_{\theta=\theta_0} = \mathbb{E}\left[\pi(X)\frac{\mathbb{E}\left\{Y(1)\frac{1-\varphi(\eta)}{\varphi(\eta)} \mid X\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)} \mid X\right\}}\frac{\dot\varphi(\eta)}{\varphi(\eta)}\right] - \mathbb{E}\left\{\pi(X)Y(1)\frac{\dot\varphi(\eta)}{\varphi(\eta)}\right\}.$$

$$\mathbb{E}(\phi_{\text{eff}}S_\eta) = D^T\mathbb{E}\{S_{\text{eff}}(\eta)S_{\text{eff}}(\eta)^T\}.$$

By $\partial V_\theta(\pi)/\partial\eta|_{\theta=\theta_0} = \mathbb{E}(\phi_{\text{eff}}S_\eta)$,

$$D = \left(\mathbb{E}\left[\pi(X)\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y \mid X, A=1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid X, A=1\right\}}\frac{\dot\varphi(\eta)}{\varphi(\eta)}\right] - \mathbb{E}\left[\pi(X)\mathbb{E}\left\{\frac{\dot\varphi(\eta)}{\varphi(\eta)^2}Y \mid X, A=1\right\}\right]\right)^T \{\text{Var}(S_{\eta,\text{eff}})\}^{-1}.$$

By (I),(II), and (III), we complete the proof. □

C.1.5 Proof of Theorem 13

Proof.

$$\begin{aligned}
&\mathbb{E}\left(\pi(X)\left[\frac{A}{\varphi(\eta)}Y + \left\{1-\frac{A}{\varphi(\eta)}\right\}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y \mid x,1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid x,1\right\}}\right]\right)\\
=&\mathbb{E}\left\{\pi(X)\frac{A}{\varphi(\eta)}Y\right\}\\
=&\mathbb{E}\left\{\pi(X)\frac{A}{\varphi(\eta)}AY(1)\right\}\\
=&\mathbb{E}\left[\mathbb{E}\left\{\pi(X)\frac{A}{\varphi(\eta)}Y(1) \mid X, Y(1)\right\}\right]\\
=&\mathbb{E}\left[\pi(X)\frac{Y(1)}{\varphi(\eta)}\mathbb{E}\{A \mid X, Y(1)\}\right]\\
=&\mathbb{E}\{\pi(X)Y(1)\} = V_1(\pi).
\end{aligned}$$

Since a solution to Equation (4.8) is a root-n estimator of η, by the strong law of large numbers and uniform consistency, we have $\widehat{V}_{\text{eff}}(\pi) = V_1(\pi) + o_p(1)$.

By Assumption 14 and the empirical process theory, we have

$$\mathbb{P}_n\left[\frac{\varphi(\widehat\eta_{\text{eff}})-a}{\varphi(\widehat\eta_{\text{eff}})}\frac{\widehat{\mathbb{E}}\left\{\frac{\dot\varphi(\eta)}{\varphi(\eta)^2} \mid x,1\right\}}{\widehat{\mathbb{E}}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2} \mid x,1\right\}}\right] - \mathbb{P}_n\left[\frac{\varphi(\widehat\eta_{\text{eff}})-a}{\varphi(\widehat\eta_{\text{eff}})}\frac{\mathbb{E}\left\{\frac{\dot\varphi(\eta)}{\varphi(\eta)^2} \mid x,1\right\}}{\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2} \mid x,1\right\}}\right]$$

$$=\mathbb{P}\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\frac{\widehat{\mathbb{E}}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}{\widehat{\mathbb{E}}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2}\mid x,1\right\}}\right]-\mathbb{P}\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\frac{\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}{\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2}\mid x,1\right\}}\right]+o_p(n^{-1/2}). \tag{C.8}$$

For the ease of exposition, let $\mathbb{E}_1=\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}\mid x,1\right\}$ and $\mathbb{E}_2=\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2}\mid x,1\right\}$. By Assumptions 14, we have

$$\begin{aligned}
&\left|\mathbb{P}\left\{\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\frac{\widehat{\mathbb{E}}_1}{\widehat{\mathbb{E}}_2}\right\}-\mathbb{P}\left\{\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\frac{\mathbb{E}_1}{\mathbb{E}_2}\right\}\right|\\
=&\left|\mathbb{P}\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\left\{\frac{\widehat{\mathbb{E}}_1}{\widehat{\mathbb{E}}_2}-\frac{\mathbb{E}_1}{\mathbb{E}_2}\right\}\right]\right|\\
=&\left|\mathbb{P}\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\left\{\frac{\widehat{\mathbb{E}}_1}{\widehat{\mathbb{E}}_2}-\frac{\mathbb{E}_1}{\widehat{\mathbb{E}}_2}+\frac{\mathbb{E}_1}{\widehat{\mathbb{E}}_2}-\frac{\mathbb{E}_1}{\mathbb{E}_2}\right\}\right]\right|\\
=&\left|\mathbb{P}\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\left\{\frac{\widehat{\mathbb{E}}_1-\mathbb{E}_1}{\widehat{\mathbb{E}}_2}+\frac{\mathbb{E}_1(\mathbb{E}_2-\widehat{\mathbb{E}}_2)}{\mathbb{E}_2\widehat{\mathbb{E}}_2}\right\}\right]\right|\\
\leq&O_p(n^{-1/2})\times o_p(1)\\
=&o_p(n^{-1/2}).
\end{aligned} \tag{C.9}$$

By Equations (C.8) and (C.9), we have

$$\mathbb{P}_n\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\frac{\widehat{\mathbb{E}}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}{\widehat{\mathbb{E}}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2}\mid x,1\right\}}\right]=\mathbb{P}_n\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\frac{\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}{\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2}\mid x,1\right\}}\right]+o_p(n^{-1/2}).$$

By taking Taylor expansion, we have

$$\begin{aligned}
&\mathbb{P}_n\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\frac{\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}{\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2}\mid x,1\right\}}\right]\\
=&\mathbb{P}_n(S_{\eta,\text{eff}})+\mathbb{P}\left[\frac{a\dot{\varphi}(\eta)}{\varphi^2(\eta)}\frac{\mathbb{E}\left\{\frac{\dot{\varphi}(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}{\mathbb{E}\left\{\frac{\varphi(\eta)-1}{\varphi(\eta)^2}\mid x,1\right\}}\right]^T(\widehat{\eta}-\eta)+o_p(n^{-1/2})\\
=&\mathbb{P}_n(S_{\eta,\text{eff}})-\text{Var}(S_{\eta,\text{eff}})(\widehat{\eta}-\eta)+o_p(n^{-1/2}).
\end{aligned} \tag{C.10}$$

By Assumption 14 and the empirical process theory, we have

$$\begin{aligned}\widehat{V}_{\text{eff}}(\pi) =& \mathbb{P}_n\left(\pi(x)\left[\frac{a}{\varphi(\widehat{\eta}_{\text{eff}})}y+\left\{1-\frac{a}{\varphi(\widehat{\eta}_{\text{eff}})}\right\}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y\mid x,1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}\right]\right)\\ &+\mathbb{P}\left[\left\{1-\frac{a}{\varphi(\widehat{\eta}_{\text{eff}})}\right\}\frac{\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y\mid x,1\right\}}{\widehat{\mathbb{E}}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}\right]-\mathbb{P}\left[\left\{1-\frac{a}{\varphi(\widehat{\eta}_{\text{eff}})}\right\}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y\mid x,1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}\right]+o_p(n^{-1/2}).\end{aligned} \tag{C.11}$$

For the ease of exposition, let $\mathbb{E}_3=\mathbb{E}\left\{\frac{1-\varphi(\eta)Y}{\varphi(\eta)^2}\mid x,1\right\}$. By Assumptions 14, we have

$$\begin{aligned}&\left|\mathbb{P}\left\{-\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\frac{\widehat{\mathbb{E}}_3}{\widehat{\mathbb{E}}_2}\right\}+\mathbb{P}\left\{\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\frac{\mathbb{E}_3}{\mathbb{E}_2}\right\}\right|\\ =&\left|\mathbb{P}\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\left\{-\frac{\widehat{\mathbb{E}}_3}{\widehat{\mathbb{E}}_2}+\frac{\mathbb{E}_3}{\mathbb{E}_2}\right\}\right]\right|\\ =&\left|\mathbb{P}\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\left\{-\frac{\widehat{\mathbb{E}}_3}{\widehat{\mathbb{E}}_2}+\frac{\mathbb{E}_3}{\widehat{\mathbb{E}}_2}-\frac{\mathbb{E}_3}{\widehat{\mathbb{E}}_2}+\frac{\mathbb{E}_3}{\mathbb{E}_2}\right\}\right]\right|\\ =&\left|\mathbb{P}\left[\frac{\varphi(\widehat{\eta}_{\text{eff}})-a}{\varphi(\widehat{\eta}_{\text{eff}})}\left\{\frac{\mathbb{E}_3-\widehat{\mathbb{E}}_3}{\widehat{\mathbb{E}}_2}+\frac{\mathbb{E}_3(\widehat{\mathbb{E}}_2-\mathbb{E}_2)}{\mathbb{E}_2\widehat{\mathbb{E}}_2}\right\}\right]\right|\\ \leq& O_p(n^{-1/2})\times o_p(1)\\ =&o_p(n^{-1/2}).\end{aligned} \tag{C.12}$$

By Equations (C.11) and (C.12), we have

$$\widehat{V}_{\text{eff}}(\pi)=\mathbb{P}_n\left(\pi(x)\left[\frac{a}{\varphi(\widehat{\eta}_{\text{eff}})}y+\left\{1-\frac{a}{\varphi(\widehat{\eta}_{\text{eff}})}\right\}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y\mid x,1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}\right]\right)+o_p(n^{-1/2}).$$

By taking Taylor expansion, we have

$$\begin{aligned}\widehat{V}_{\text{eff}}(\pi) =& \mathbb{P}_n\left(\pi(x)\left[\frac{a}{\varphi(\eta)}y+\left\{1-\frac{a}{\varphi(\eta)}\right\}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y\mid x,1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}\right]\right)\\ &+\mathbb{P}\left(\pi(x)\left[-\frac{a\dot{\varphi}(\eta)}{\varphi^2(\eta)}y+\frac{a\dot{\varphi}(\eta)}{\varphi^2(\eta)}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y\mid x,1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}\mid x,1\right\}}\right]\right)^T(\widehat{\eta}-\eta)+o_p(n^{-1/2}).\end{aligned} \tag{C.13}$$

By Equations (C.10) and (C.13), we have

$$
\begin{aligned}
&\widehat{V}_{\text{eff}}(\pi) - V_1(\pi) \\
=&\mathbb{P}_n\left(\pi(x)\left[\frac{a}{\varphi(\eta)}y + \left\{1-\frac{a}{\varphi(\eta)}\right\}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y \mid x,1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid x,1\right\}}\right]\right) \\
&+\mathbb{P}\left(\pi(x)\left[-\frac{a\dot{\varphi}(\eta)}{\varphi^2(\eta)}y + \frac{a\dot{\varphi}(\eta)}{\varphi^2(\eta)}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y \mid x,1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid x,1\right\}}\right]\right)^T \{\text{Var}(S_{\eta,\text{eff}})\}^{-1}\mathbb{P}_n(S_{\eta,\text{eff}}) - V_1(\pi) + o_p(n^{-1/2}) \\
=&\mathbb{P}_n\left(\pi(x)\left[\frac{a}{\varphi(\eta)}y + \left\{1-\frac{a}{\varphi(\eta)}\right\}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y \mid x,1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid x,1\right\}}\right]\right) + D\mathbb{P}_n(S_{\eta,\text{eff}}) - V_1(\pi) + o_p(n^{-1/2}) \\
=&\mathbb{P}_n\left(\pi(x)\left[\frac{a}{\varphi(\eta)}y + \left\{1-\frac{a}{\varphi(\eta)}\right\}\frac{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2}Y \mid x,1\right\}}{\mathbb{E}\left\{\frac{1-\varphi(\eta)}{\varphi(\eta)^2} \mid x,1\right\}}\right] + DS_{\eta,\text{eff}} - V_1(\pi)\right) + o_p(n^{-1/2}) \\
=&\mathbb{P}_n\left\{\phi_{\text{eff}}(\pi)\right\} + o_p(n^{-1/2}).
\end{aligned}
$$

This completes the proof. □

C.1.6 Proof of Proposition 1

$$
\begin{aligned}
&\underset{\pi\in\Pi}{\text{argmax}}\,\widehat{V}_{\text{eff}}(\pi) \\
=&\underset{\pi\in\Pi}{\text{argmax}}\sum_{i=1}^{n}\pi(x_i)\widehat{\psi}(x_i,y_i,a_i) \\
=&\underset{\pi\in\Pi}{\text{argmax}}\sum_{i=1}^{n}\pi(x_i)|\widehat{\psi}(x_i,y_i,a_i)|[\mathbb{I}\{\widehat{\psi}(x_i,y_i,a_i)>0\} - \mathbb{I}\{\widehat{\psi}(x_i,y_i,a_i)\le 0\}] \\
=&\underset{\pi\in\Pi}{\text{argmax}}\sum_{i=1}^{n}|\widehat{\psi}(x_i,y_i,a_i)|\mathbb{I}\{\widehat{\psi}(x_i,y_i,a_i)>0\} \\
&\qquad -|\widehat{\psi}(x_i,y_i,a_i)|[\{1-\pi(x_i)\}\mathbb{I}\{\widehat{\psi}(x_i,y_i,a_i)>0\} + \pi(x_i)\mathbb{I}\{\widehat{\psi}(x_i,y_i,a_i)\le 0\}] \\
=&\underset{\pi\in\Pi}{\text{argmax}}\sum_{i=1}^{n}|\widehat{\psi}(x_i,y_i,a_i)|\mathbb{I}\{\widehat{\psi}(x_i,y_i,a_i)>0\} \\
&\qquad -|\widehat{\psi}(x_i,y_i,a_i)|[\pi(x_i) + \mathbb{I}\{\widehat{\psi}(x_i,y_i,a_i)>0\} - 2\pi(x_i)\mathbb{I}\{\widehat{\psi}(x_i,y_i,a_i)>0\}] \\
=&\underset{\pi\in\Pi}{\text{argmax}}\sum_{i=1}^{n}|\widehat{\psi}(x_i,y_i,a_i)|\mathbb{I}\{\widehat{\psi}(x_i,y_i,a_i)>0\} \\
&\qquad -|\widehat{\psi}(x_i,y_i,a_i)|[\pi^2(x) + \mathbb{I}^2\{\widehat{\psi}(x_i,y_i,a_i)>0\} - 2\pi(x_i)\mathbb{I}\{\widehat{\psi}(x_i,y_i,a_i)>0\}]
\end{aligned}
$$

$$
\begin{aligned}
&= \underset{\pi \in \Pi}{\operatorname{argmax}} \sum_{i=1}^{n} |\widehat{\psi}(x_i, y_i, a_i)| \mathbb{I}\{\widehat{\psi}(x_i, y_i, a_i) > 0\} - |\widehat{\psi}(x_i, y_i, a_i)| [\pi(x_i) - \mathbb{I}\{\widehat{\psi}(x_i, y_i, a_i) > 0\}]^2 \\
&= \underset{\pi \in \Pi}{\operatorname{argmax}} \sum_{i=1}^{n} -|\widehat{\psi}(x_i, y_i, a_i)| [\pi(x_i) - \mathbb{I}\{\widehat{\psi}(x_i, y_i, a_i) > 0\}]^2 \\
&= \underset{\pi \in \Pi}{\operatorname{argmin}} \sum_{i=1}^{n} |\widehat{\psi}(x_i, y_i, a_i)| [\pi(x_i) - \mathbb{I}\{\widehat{\psi}(x_i, y_i, a_i) > 0\}]^2 \\
&= \underset{\pi \in \Pi}{\operatorname{argmin}} \sum_{i=1}^{n} |\widehat{\psi}(x_i, y_i, a_i)| \mathbb{I}[\pi(x_i) \neq \mathbb{I}\{\widehat{\psi}(x_i, y_i, a_i) > 0\}].
\end{aligned}
$$

Therefore, the policy learning problem is equivalent to a weighted classification problem, where for subject i with features x_i, the true label is $\mathbb{I}\{\widehat{\psi}(x_i, y_i, a_i) > 0\}$ and the sample weight is $|\widehat{\psi}(x_i, y_i, a_i)|$. □