

# Optimal Treatment Regimes for Proximal Causal Learning

Tao Shen

National University of Singapore

Yifan Cui\*

Zhejiang University

## Abstract

A common concern when a policymaker draws causal inferences from and makes decisions based on observational data is that the measured covariates are insufficiently rich to account for all sources of confounding, i.e., the standard no confoundedness assumption fails to hold. The recently proposed proximal causal inference framework shows that proxy variables that abound in real-life scenarios can be leveraged to identify causal effects and therefore facilitate decision-making. Building upon this line of work, we propose a novel optimal individualized treatment regime based on so-called outcome and treatment confounding bridges. We then show that the value function of this new optimal treatment regime is superior to that of existing ones in the literature. Theoretical guarantees, including identification, superiority, excess value bound, and consistency of the estimated regime, are established. Furthermore, we demonstrate the proposed optimal regime via numerical experiments and a real data application.

---

\*Correspondence to Yifan Cui <[cuiyf@zju.edu.cn](mailto:cuiyf@zju.edu.cn)>

# 1 Introduction

Data-driven individualized decision-making has received tremendous attention nowadays due to its applications in healthcare, economics, marketing, etc. A large branch of work has focused on maximizing the expected utility of implementing the estimated optimal policy over a target population based on randomized controlled trials or observational studies, e.g., Athey and Wager (2021); Chakraborty and Moodie (2013); Jiang et al. (2019); Kitagawa and Tetenov (2018); Kosorok and Laber (2019); Murphy (2003); Qian and Murphy (2011); Robins (1986, 1994, 1997); Tsiatis et al. (2019); Wu et al. (2019); Zhao et al. (2012, 2019).

A critical assumption commonly made in these studies, known as unconfoundedness or exchangeability, precludes the existence of unmeasured confounding. Relying on an assumed ability of the decision-maker to accurately measure covariates relevant to a variety of confounding mechanisms present in a given observational study, causal effects, value functions, and other relevant quantities can be nonparametrically identified. However, such an assumption might not always be realistic in observational studies or randomized trials subject to non-compliance (Robins, 1994, 1997). Therefore, it is of great interest in recovering confounding mechanisms from measured covariates to infer causal effects and facilitate decision-making. A prevailing strand of work has been devoted to using instrumental variable (Angrist et al., 1996; Imbens and Angrist, 1994) as a proxy variable in dynamic treatment regimes and reinforcement learning settings (Cui, 2021; Cui and Tchetgen Tchetgen, 2021b,a; Han, 2023; Liao et al., 2021; Pu and Zhang, 2021; Qiu et al., 2021; Stensrud and Sarvet, 2022).

Recently, Tchetgen Tchetgen et al. proposed the so-called proximal causal inference framework, a formal potential outcome framework for proximal causal learning, which while explicitly acknowledging covariate measurements as imperfect proxies of confounding mechanisms, establishes causal identification in settings where exchangeability on the basis of measured covariates fails. Rather than as current practice dictates, assuming that adjusting for all measured covariates, unconfoundedness can be attained, proximal causal inference es-

sentially requires that the investigator can correctly classify a subset of measured covariates  $L \in \mathcal{L}$  into three types: i) variables  $X \in \mathcal{X}$  that may be common causes of the treatment and outcome variables; ii) treatment-inducing confounding proxies  $Z \in \mathcal{Z}$ ; and iii) outcome-inducing confounding proxies  $W \in \mathcal{W}$ .

There is a fast-growing literature on proximal causal inference since it has been proposed (Cui et al., 2023; Dukes et al., 2023; Ghassami et al., 2023; Kompa et al., 2022; Li et al., 2023; Mastouri et al., 2021; Miao et al., 2018b; Shi et al., 2020b, 2021; Shpitser et al., 2023; Singh, 2020; Tchetgen Tchetgen et al., 2020; Ying et al., 2023, 2022 and many others). In particular, Miao et al. (2018a); Tchetgen Tchetgen et al. (2020) propose identification of causal effects through an outcome confounding bridge and Cui et al. (2023) propose identification through a treatment confounding bridge. A doubly robust estimation strategy (Chernozhukov et al., 2018; Robins et al., 1994; Rotnitzky et al., 1998; Scharfstein et al., 1999) is further proposed in Cui et al. (2023). In addition, Ghassami et al. (2022) and Kallus et al. (2021) propose a nonparametric estimation of causal effects through a min-max approach. Moreover, by adopting the proximal causal inference framework, Qi et al. (2023) consider optimal individualized treatment regimes (ITRs) estimation, Sverdrup and Cui (2023) consider learning heterogeneous treatment effects, and Bennett and Kallus (2023) consider off-policy evaluation in partially observed Markov decision processes.

In this paper, we aim to estimate optimal ITRs under the framework of proximal causal inference. We start with reviewing two in-class ITRs that map from  $\mathcal{X} \times \mathcal{W}$  to  $\mathcal{A}$  and  $\mathcal{X} \times \mathcal{Z}$  to  $\mathcal{A}$ , respectively, where  $\mathcal{A}$  denotes the binary treatment space. The identification of value function and the learning strategy for these two optimal in-class ITRs are proposed in Qi et al. (2023). In addition, Qi et al. (2023) also consider a maximum proximal learning optimal ITR that maps from  $\mathcal{X} \times \mathcal{W} \times \mathcal{Z}$  to  $\mathcal{A}$  with the ITRs being restricted to either  $\mathcal{X} \times \mathcal{W} \rightarrow \mathcal{A}$  or  $\mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{A}$ . In contrast to their maximum proximal learning ITR, in this paper, we propose a brand new policy class whose ITRs map from measured covariates  $\mathcal{X} \times \mathcal{W} \times \mathcal{Z}$  to  $\mathcal{A}$ , which incorporates the predilection between these two in-class ITRs. Identification and

superiority of the proposed optimal ITRs compared to existing ones are further established.

The main contributions of our work are four-fold. Firstly, by leveraging treatment and outcome confounding bridges under the recently proposed proximal causal inference framework, identification results regarding the proposed class  $\mathcal{D}_{\mathcal{ZW}}^{\Pi}$  of ITRs that map  $\mathcal{X} \times \mathcal{W} \times \mathcal{Z}$  to  $\mathcal{A}$  are established. The proposed ITR class can be viewed as a generalization of existing ITR classes proposed in the literature. Secondly, an optimal subclass of  $\mathcal{D}_{\mathcal{ZW}}^{\Pi}$  is further introduced. Learning optimal treatment regimes within this subclass leads to a superior value function. Thirdly, we propose a learning approach to estimating the proposed optimal ITR. Our learning pipeline begins with the estimation of confounding bridges adopting the deep neural network method proposed by [Kompa et al. \(2022\)](#). Then we use optimal treatment regimes proposed in [Qi et al. \(2023\)](#) as preliminary regimes to estimate our optimal ITR. Lastly, we establish an excess value bound for the value difference between the estimated treatment regime and existing ones in the literature, and the consistency of the estimated regime is also demonstrated.

## 2 Methodology

### 2.1 Optimal individualized treatment regimes

We briefly introduce some conventional notation for learning optimal ITRs. Suppose  $A$  is a binary variable representing a treatment option that takes values in the treatment space  $\mathcal{A} = \{-1, 1\}$ . Let  $L \in \mathcal{L}$  be a vector of observed covariates, and  $Y$  be the outcome of interest. Let  $Y(1)$  and  $Y(-1)$  be the potential outcomes under an intervention that sets the treatment to values 1 and  $-1$ , respectively. Without loss of generality, we assume that larger values of  $Y$  are preferred.

Suppose the following standard causal assumptions hold: (1) Consistency:  $Y = Y(A)$ . That is, the observed outcome matches the potential outcome under the realized treatment. (2) Positivity:  $\mathbb{P}(A = a|L) > 0$  for  $a \in \mathcal{A}$  almost surely, i.e., both treatments are possible to

be assigned.

We consider an ITR class  $\mathcal{D}$  containing ITRs that are measurable functions mapping from the covariate space  $\mathcal{L}$  onto the treatment space  $\mathcal{A}$ . For any  $d \in \mathcal{D}$ , the potential outcome under a hypothetical intervention that assigns treatment according to  $d$  is defined as

$$Y(d(L)) \stackrel{\triangle}{=} Y(1)\mathbb{I}\{d(L) = 1\} + Y(-1)\mathbb{I}\{d(L) = -1\},$$

where  $\mathbb{I}\{\cdot\}$  denotes the indicator function. The value function of ITR  $d$  is defined as the expectation of the potential outcome, i.e.,

$$V(d) \stackrel{\triangle}{=} \mathbb{E}[Y(d(L))].$$

It can be easily seen that an optimal ITR can be expressed as

$$d^*(L) = \text{sign}\{\mathbb{E}(Y(1) - Y(-1)|L)\}$$

or

$$d^* = \arg \max_{d \in \mathcal{D}} \mathbb{E}[Y(d(L))].$$

There are many ways to identify optimal ITRs under different sets of assumptions. The most commonly seen assumption is the unconfoundedness:  $Y(a) \perp A|L$  for  $a = \pm 1$ , i.e., upon conditioning on  $L$ , there is no unmeasured confounder affecting both  $A$  and  $Y$ . Under this unconfoundedness assumption, the value function of a given regime  $d$  can be identified by (Qian and Murphy, 2011)

$$V(d) = \mathbb{E} \left[ \frac{Y\mathbb{I}\{A = d(L)\}}{f(A|L)} \right],$$

where  $f(A|L)$  denotes the propensity score (Rosenbaum and Rubin, 1983), and the optimal ITR is identified by

$$d^* = \arg \max_{d \in \mathcal{D}} V(d) = \arg \max_{d \in \mathcal{D}} \mathbb{E} \left[ \frac{Y\mathbb{I}\{A = d(L)\}}{f(A|L)} \right].$$

We refer to [Qian and Murphy \(2011\)](#); [Zhang et al. \(2012\)](#); [Zhao et al. \(2012\)](#) for more details of learning optimal ITRs in this unconfounded setting.

Because confounding by unmeasured factors cannot generally be ruled out with certainty in observational studies or randomized experiments subject to non-compliance, skepticism about the unconfoundedness assumption in observational studies is often warranted. To estimate optimal ITRs subject to potential unmeasured confounding, [Cui and Tchetgen Tchetgen \(2021b\)](#) propose instrumental variable approaches to learning optimal ITRs. Under certain instrumental variable assumptions, the optimal ITR can be identified by

$$\arg \max_{d \in \mathcal{D}} \mathbb{E} \left[ \frac{MAY\mathbb{I}\{A = d(L)\}}{\{\mathbb{P}(A = 1|M = 1, L) - \mathbb{P}(A = 1|M = -1, L)\}f(M|L)} \right],$$

where  $M$  denotes a valid binary instrumental variable. Other works including [Cui \(2021\)](#); [Cui and Tchetgen Tchetgen \(2021a\)](#); [Han \(2023\)](#); [Pu and Zhang \(2021\)](#) consider a sign or partial identification of causal effects to estimate suboptimal ITRs using instrumental variables.

## 2.2 Existing optimal ITRs for proximal causal inference

Another line of research in causal inference considers negative control variables as proxies to mitigate confounding bias ([Kuroki and Pearl, 2014](#); [Miao et al., 2018a](#); [Shi et al., 2020a](#); [Tchetgen Tchetgen, 2014](#)). Recently, a formal potential outcome framework, namely proximal causal inference, has been developed by [Tchetgen Tchetgen et al. \(2020\)](#), which has attracted tremendous attention since proposed.

Following the proximal causal inference framework proposed in [Tchetgen Tchetgen et al. \(2020\)](#), suppose that the measured covariate  $L$  can be decomposed into three buckets  $L = (X, W, Z)$ , where  $X \in \mathcal{X}$  affects both  $A$  and  $Y$ ,  $W \in \mathcal{W}$  denotes an outcome-inducing confounding proxy that is a potential cause of the outcome which is related with the treatment only through  $(U, X)$ , and  $Z \in \mathcal{Z}$  is a treatment-inducing confounding proxy that is a potential cause of the treatment which is related with the outcome  $Y$  through  $(U, X, A)$ . We now summarize several basic assumptions of the proximal causal inference framework.

**Assumption 1.** We make the following assumptions:

(1) *Consistency*:  $Y = Y(A, Z)$ ,  $W = W(A, Z)$ .

(2) *Positivity*:  $\mathbb{P}(A = a \mid U, X) > 0$ ,  $\forall a \in \mathcal{A}$ .

(3) *Latent unconfoundedness*:

$(Z, A) \perp (Y(a), W) \mid U, X$ ,  $\forall a \in \mathcal{A}$ .

The consistency and positivity assumptions are conventional in the causal inference literature. The latent unconfoundedness essentially states that  $Z$  cannot directly affect the outcome  $Y$ , and  $W$  is not directly affected by either  $A$  or  $Z$ . Figure 1 depicts a classical setting that satisfies Assumption 1. We refer to [Shi et al. \(2020b\)](#); [Tchetgen Tchetgen et al. \(2020\)](#) for other realistic settings for proximal causal inference.

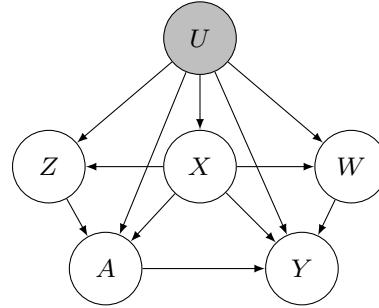


Figure 1: A causal DAG under the proximal causal inference framework.

We first consider two in-class optimal ITRs that map from  $\mathcal{X} \times \mathcal{Z}$  to  $\mathcal{A}$  and  $\mathcal{X} \times \mathcal{W}$  to  $\mathcal{A}$ , respectively. To identify optimal ITRs that map from  $\mathcal{X} \times \mathcal{Z}$  to  $\mathcal{A}$ , we make the following assumptions.

**Assumption 2.** *Completeness*: For any  $a \in \mathcal{A}, x \in \mathcal{X}$  and square-integrable function  $g$ ,  $\mathbb{E}[g(U) \mid Z, A = a, X = x] = 0$  almost surely if and only if  $g(U) = 0$  almost surely.

**Assumption 3.** *Existence of outcome confounding bridge*: There exists an outcome confounding bridge function  $h(w, a, x)$  that solves the following equation

$$\mathbb{E}[Y \mid Z, A, X] = \mathbb{E}[h(W, A, X) \mid Z, A, X],$$

almost surely.

The completeness Assumption 2 is a technical condition central to the study of sufficiency in foundational theory of statistical inference. It essentially assumes that  $Z$  has sufficient variability with respect to the variability of  $U$ . We refer to Tchetgen Tchetgen et al. (2020) and Miao et al. (2022) for further discussions regarding the completeness condition. Assumption 3 defines a so-called inverse problem known as a Fredholm integral equation of the first kind through an outcome confounding bridge. The technical conditions for the existence of a solution to a Fredholm integral equation can be found in Kress et al. (1989).

Let  $\mathcal{D}_Z$  be an ITR class that includes all measurable functions mapping from  $\mathcal{X} \times \mathcal{Z}$  to  $\mathcal{A}$ . As shown in Qi et al. (2023), under Assumptions 1, 2 and 3, for any  $d_z \in \mathcal{D}_Z$ , the value function  $V(d_z)$  can be nonparametrically identified by

$$V(d_z) = \mathbb{E}[h(W, d_z(X, Z), X)]. \quad (1)$$

Furthermore, the in-class optimal treatment regime  $d_z^* \in \arg \max_{d_z \in \mathcal{D}_Z} V(d_z)$  is given by

$$d_z^*(X, Z) = \text{sign}\{\mathbb{E}[h(W, 1, X) - h(W, -1, X)|X, Z]\}.$$

On the other hand, to identify optimal ITRs that map from  $\mathcal{X} \times \mathcal{W}$  to  $\mathcal{A}$ , we make the following assumptions.

**Assumption 4.** *Completeness:* For any  $a \in \mathcal{A}, x \in \mathcal{X}$  and square-integrable function  $g$ ,  $\mathbb{E}[g(U) | W, A = a, X = x] = 0$  almost surely if and only if  $g(U) = 0$  almost surely.

**Assumption 5.** *Existence of treatment confounding bridge:* There exists a treatment confounding bridge function  $q(z, a, x)$  that solves the following equation

$$\frac{1}{\mathbb{P}(A = a | W, X)} = \mathbb{E}[q(Z, a, X) | W, A = a, X],$$

almost surely.

Similar to Assumptions 2 and 3, Assumption 4 assumes that  $W$  has sufficient variability relative to the variability of  $U$ , and Assumption 5 defines another Fredholm integral equation of the first kind through a treatment confounding bridge  $q$ .

Let  $\mathcal{D}_W$  be another ITR class that includes all measurable functions mapping from  $\mathcal{X} \times \mathcal{W}$  to  $\mathcal{A}$ . As shown in Qi et al. (2023), under Assumptions 1, 4 and 5, for any  $d_w \in \mathcal{D}_W$ , the value function  $V(d_w)$  can be nonparametrically identified by

$$V(d_w) = \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}]. \quad (2)$$

The in-class optimal treatment regime  $d_w^* \in \arg \max_{d_w \in \mathcal{D}_W} V(d_w)$  is given by

$$d_w^*(X, W) = \text{sign}\{\mathbb{E}[Yq(Z, 1, X)\mathbb{I}\{A = 1\} - Yq(Z, -1, X)\mathbb{I}\{A = -1\}|X, W]\}.$$

Moreover, Qi et al. (2023) consider the ITR class  $\mathcal{D}_Z \cup \mathcal{D}_W$  and propose a maximum proximal learning optimal regime based on this ITR class. For any  $d_{z \cup w} \in \mathcal{D}_Z \cup \mathcal{D}_W$ , under Assumptions 1-5, the value function  $V(d_{z \cup w})$  for any  $d_{z \cup w} \in \mathcal{D}_Z \cup \mathcal{D}_W$  can be identified by

$$V(d_{z \cup w}) = \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_Z\}\mathbb{E}[h(W, d_{z \cup w}(X, Z), X)] + \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_W\}\mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_{z \cup w}(X, W) = A\}]. \quad (3)$$

The optimal ITR within this class is given by  $d_{z \cup w}^* \in \arg \max_{d_{z \cup w} \in \mathcal{D}_Z \cup \mathcal{D}_W} V(d_{z \cup w})$ , and they show that the corresponding optimal value function takes the maximum value between two optimal in-class ITRs, i.e.,

$$V(d_{z \cup w}^*) = \max\{V(d_z^*), V(d_w^*)\}.$$

### 2.3 Optimal decision-making based on two confounding bridges

As discussed in the previous section, given that neither  $\mathbb{E}[Y(a)|X, U]$  nor  $\mathbb{E}[Y(a)|X, W, Z]$  for any  $a \in \mathcal{A}$  may be identifiable under the proximal causal inference setting, one might nevertheless consider ITRs mapping from  $\mathcal{X} \times \mathcal{W}$  to  $\mathcal{A}$ , from  $\mathcal{X} \times \mathcal{Z}$  to  $\mathcal{A}$ , from  $\mathcal{X} \times \mathcal{W} \times \mathcal{Z}$  to  $\mathcal{A}$  as well as from  $\mathcal{X}$  to  $\mathcal{A}$ . Intuitively, policy-makers might want to use as much information as they can to facilitate their decision-making. Therefore, ITRs mapping from  $\mathcal{X} \times \mathcal{W} \times \mathcal{Z}$  to  $\mathcal{A}$  are of great interest if information regarding  $(X, W, Z)$  is available.

As a result, a natural question arises: is there an ITR mapping from  $\mathcal{X} \times \mathcal{W} \times \mathcal{Z}$  to  $\mathcal{A}$  which dominates existing ITRs proposed in the literature? In this section, we answer this

question by proposing a novel optimal ITR and showing its superiority in terms of global welfare.

We first consider the following class of ITRs that map from  $\mathcal{X} \times \mathcal{W} \times \mathcal{Z}$  to  $\mathcal{A}$ ,

$$\mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\Pi} \triangleq \{d_{zw}^{\pi} : d_{zw}^{\pi}(X, W, Z) = \pi(X)d_z(X, Z) + (1 - \pi(X))d_w(X, W), d_z \in \mathcal{D}_{\mathcal{Z}}, d_w \in \mathcal{D}_{\mathcal{W}}, \pi \in \Pi\},$$

where  $\Pi$  is the policy class containing all measurable functions  $\pi : \mathcal{X} \rightarrow \{0, 1\}$  that indicate the individualized predilection between  $d_z$  and  $d_w$ .

**Remark 1.** Note that  $\mathcal{D}_{\mathcal{Z}}$ ,  $\mathcal{D}_{\mathcal{W}}$  and  $\mathcal{D}_{\mathcal{Z}} \cup \mathcal{D}_{\mathcal{W}}$  are subsets of  $\mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\Pi}$  with a particular choice of  $\pi$ . For example,  $\mathcal{D}_{\mathcal{Z}}$  is  $\mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\Pi}$  with restriction on  $\pi(X) = 1$ ;  $\mathcal{D}_{\mathcal{W}}$  is  $\mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\Pi}$  with restriction on  $\pi(X) = 0$ ;  $\mathcal{D}_{\mathcal{Z}} \cup \mathcal{D}_{\mathcal{W}}$  is  $\mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\Pi}$  with restriction on  $\pi(X) = 1$  or  $\pi(X) = 0$ .

In the following theorem, we demonstrate that by leveraging the treatment and outcome confounding bridge functions, we can nonparametrically identify the value function over the policy class  $\mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\Pi}$ , i.e.,  $V(d_{zw}^{\pi})$  for  $d_{zw}^{\pi} \in \mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\Pi}$ .

**Theorem 1.** Under Assumptions 1-5, for any  $d_{zw}^{\pi} \in \mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\Pi}$ , the value function  $V(d_{zw}^{\pi})$  can be nonparametrically identified by

$$V(d_{zw}^{\pi}) = \mathbb{E}[\pi(X)h(W, d_z(X, Z), X) + (1 - \pi(X))Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}]. \quad (4)$$

One of the key ingredients of our constructed new policy class  $\mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\Pi}$  is the choice of  $\pi(\cdot)$ . It suggests an individualized strategy for treatment decisions between the two given treatment regimes. Because we are interested in policy learning, a suitable choice of  $\pi(\cdot)$  that leads to a larger value function is more desirable. Therefore, we construct the following  $\bar{\pi}(X; d_z, d_w)$ ,

$$\bar{\pi}(X; d_z, d_w) \triangleq \mathbb{I}\{\mathbb{E}[h(W, d_z(X, Z), X)|X] \geq \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]\}. \quad (5)$$

In addition, given any  $d_z \in \mathcal{D}_{\mathcal{Z}}$  and  $d_w \in \mathcal{D}_{\mathcal{W}}$ , we define

$$d_{zw}^{\bar{\pi}}(X, W, Z) \triangleq \bar{\pi}(X; d_z, d_w)d_z(X, Z) + (1 - \bar{\pi}(X; d_z, d_w))d_w(X, W).$$

We then obtain the following result which justifies the superiority of  $\bar{\pi}$ .

**Theorem 2.** Under Assumptions 1-5, for any  $d_z \in \mathcal{D}_{\mathcal{Z}}$  and  $d_w \in \mathcal{D}_{\mathcal{W}}$ ,

$$V(d_{zw}^{\bar{\pi}}) \geq \max\{V(d_z), V(d_w)\}.$$

Theorem 2 establishes that for the particular choice of  $\bar{\pi}$  given in (5), the value function of  $d_{zw}^{\bar{\pi}}$  is no smaller than that of  $d_z$  and  $d_w$  for any  $d_z \in \mathcal{D}_{\mathcal{Z}}$ , and  $d_w \in \mathcal{D}_{\mathcal{W}}$ . Consequently, Theorem 2 holds for  $d_z^*$  and  $d_w^*$ . Hence, we propose the following optimal ITR  $d_{zw}^{\bar{\pi}*}$ ,

$$d_{zw}^{\bar{\pi}*}(X, W, Z) \stackrel{\Delta}{=} \bar{\pi}(X; d_z^*, d_w^*)d_z^*(X, Z) + (1 - \bar{\pi}(X; d_z^*, d_w^*))d_w^*(X, W),$$

and we have the following corollary.

**Corollary 1.** Under Assumptions 1-5, we have that

$$V(d_{zw}^{\bar{\pi}*}) \geq \max\{V(d_z^*), V(d_w^*), V(d_{z \cup w}^*)\}.$$

Corollary 1 essentially states that the value of  $d_{zw}^{\bar{\pi}*}$  dominates that of  $d_z^*$ ,  $d_w^*$ , as well as  $d_{z \cup w}^*$ . Moreover, the proposition below demonstrates the optimality of  $d_{zw}^{\bar{\pi}*}$  within the proposed class.

**Proposition 1.** Under Assumptions 1-5, we have that

$$d_{zw}^{\bar{\pi}*} \in \arg \max_{d_{zw}^{\bar{\pi}} \in \mathcal{D}_{\mathcal{ZW}}^{\Pi}} V(d_{zw}^{\bar{\pi}}).$$

Therefore,  $d_{zw}^{\bar{\pi}*}$  is an optimal ITR of policymakers' interest.

### 3 Statistical Learning and Optimization

#### 3.1 Estimation of the optimal ITR $d_{zw}^{\bar{\pi}*}$

The estimation of  $d_{zw}^{\bar{\pi}*}$  consists of four steps: (i) estimation of confounding bridges  $h$  and  $q$ ; (ii) estimation of preliminary ITRs  $d_z^*$  and  $d_w^*$ ; (iii) estimation of  $\bar{\pi}(X; d_z^*, d_w^*)$ ; and (iv) learning  $d_{zw}^{\bar{\pi}*}$  based on (ii) and (iii). The estimation problem (i) has been developed by Cui et al. (2023); Miao et al. (2018b) using the generalized method of moments, Ghassami et al.

(2022); Kallus et al. (2021) by a min-max estimation (Dikkala et al., 2020) using kernels, and Kompa et al. (2022) using deep learning; and (ii) has been developed by Qi et al. (2023). We restate estimation of (i) and (ii) for completeness. With regard to (i), recall that Assumptions 3 and 5 imply the following conditional moment restrictions

$$\begin{aligned}\mathbb{E}[Y - h(W, A, X)|Z, A, X] &= 0, \\ \mathbb{E}[1 - \mathbb{I}\{A = a\}q(Z, a, X)|W, X] &= 0, \forall a \in \mathcal{A}.\end{aligned}$$

respectively. Kompa et al. (2022) propose a deep neural network approach to estimating bridge functions which avoids the reliance on kernel methods. We adopt this approach in our simulation and details can be found in the Appendix.

To estimate  $d_z^*$ , we consider classification-based approaches according to Zhang et al. (2012); Zhao et al. (2012). Under Assumptions 1, 2 and 3, maximizing the value function in (1) is equivalent to minimizing the following classification error

$$\mathbb{E}[\{h(W, 1, X) - h(W, -1, X)\}\mathbb{I}\{d_z(X, Z) \neq 1\}] \quad (6)$$

over  $d_z \in \mathcal{D}_z$ . By choosing some measurable decision function  $g_z \in \mathcal{G}_z : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}$ , we let  $d_z(X, Z) = \text{sign}(g_z(X, Z))$ . We consider the following empirical version of (6),

$$\min_{g_z \in \mathcal{G}_z} \mathbb{P}_n[\{\hat{h}(W, 1, X) - \hat{h}(W, -1, X)\}\mathbb{I}\{g_z(X, Z) < 0\}].$$

Due to the non-convexity and non-smoothness of the sign operator, we replace the sign operator with a smooth surrogate function and adopt the hinge loss  $\phi(x) = \max\{1 - x, 0\}$ . By adding a penalty term  $\rho_z \|g_z\|_{\mathcal{G}_z}^2$  to avoid overfitting, we solve

$$\hat{g}_z \in \arg \min_{g_z \in \mathcal{G}_z} \mathbb{P}_n[\{\hat{h}(W, 1, X) - \hat{h}(W, -1, X)\}\phi(g_z(X, Z))] + \rho_z \|g_z\|_{\mathcal{G}_z}^2, \quad (7)$$

where  $\rho_z > 0$  is a tuning parameter. The estimated ITR then follows  $\hat{d}_z(X, Z) = \text{sign}(\hat{g}_z(X, Z))$ . Similarly, under Assumptions 1, 4 and 5, maximizing the value function in (2) is equivalent to minimizing the following classification error

$$\mathbb{E}[\{Yq(Z, 1, X)\mathbb{I}\{A = 1\} - Yq(Z, -1, X)\mathbb{I}\{A = -1\}\}\mathbb{I}\{d_w(X, W) \neq 1\}]$$

over  $d_w \in \mathcal{D}_{\mathcal{W}}$ . By the same token, the problem is transformed into minimizing the following empirical error

$$\hat{g}_w \in \arg \min_{g_w \in \mathcal{G}_{\mathcal{W}}} \mathbb{P}_n[\{Y\hat{q}(Z, 1, X)\mathbb{I}\{A = 1\} - Y\hat{q}(Z, -1, X)\mathbb{I}\{A = -1\}\}\phi(g_w(X, W))] + \rho_w \|g_w\|_{\mathcal{G}_{\mathcal{W}}}^2. \quad (8)$$

The estimated ITR is obtained via  $\hat{d}_w(X, W) = \text{sign}(\hat{g}_w(X, W))$ .

For problem (iii), given two preliminary ITRs, we construct an estimator  $\hat{\pi}(X; \hat{d}_z, \hat{d}_w)$ , that is, for  $x \in \mathcal{X}$ ,

$$\hat{\pi}(x; \hat{d}_z, \hat{d}_w) = \mathbb{I}\{\hat{\delta}(x; \hat{d}_z, \hat{d}_w) \geq 0\},$$

where  $\hat{\delta}(x; \hat{d}_z, \hat{d}_w)$  denotes a generic estimator of

$$\delta(x; \hat{d}_z, \hat{d}_w) \triangleq \mathbb{E}[h(W, \hat{d}_z(X, Z), X) - Yq(Z, A, X)\mathbb{I}\{\hat{d}_w(X, W) = A\}|X = x],$$

where the expectation is taken with respect to everything except  $\hat{d}_z$  and  $\hat{d}_w$ . For example, the Nadaraya-Watson kernel regression estimator (Nadaraya, 1964) can be used, i.e.,  $\hat{\delta}(x; \hat{d}_z, \hat{d}_w)$  is expressed as

$$\frac{\sum_{i=1}^n \{\hat{h}(W_i, \hat{d}_z(x, Z_i), x) - Y_i\hat{q}(Z_i, A_i, x)\mathbb{I}\{\hat{d}_w(x, W_i) = A_i\}\}K(\frac{\|x - X_i\|_2}{\gamma})}{\sum_{i=1}^n K(\frac{\|x - X_i\|_2}{\gamma})},$$

where  $K : \mathbb{R} \rightarrow \mathbb{R}$  is a kernel function such as Gaussian kernel,  $\|\cdot\|_2$  denotes the  $L_2$ -norm, and  $\gamma$  denotes the bandwidth.

Finally, given  $\hat{d}_z, \hat{d}_w$  and  $\hat{\pi}(X; \hat{d}_z, \hat{d}_w)$ ,  $\hat{d}_{zw}^{\hat{\pi}}$  is estimated by the following plug-in regime,

$$\hat{d}_{zw}^{\hat{\pi}}(X, W, Z) = \hat{\pi}(X; \hat{d}_z, \hat{d}_w)\hat{d}_z(X, Z) + (1 - \hat{\pi}(X; \hat{d}_z, \hat{d}_w))\hat{d}_w(X, W). \quad (9)$$

### 3.2 Theoretical guarantees for $\hat{d}_{zw}^{\hat{\pi}}$

In this subsection, we first present an optimality guarantee for the estimated ITR  $\hat{d}_{zw}^{\hat{\pi}}$  in terms of its value function

$$V(\hat{d}_{zw}^{\hat{\pi}}) = \mathbb{E}[\hat{\pi}(X; \hat{d}_z, \hat{d}_w)h(W, \hat{d}_z(X, Z), X) + (1 - \hat{\pi}(X; \hat{d}_z, \hat{d}_w))Yq(Z, A, X)\mathbb{I}\{\hat{d}_w(X, W) = A\}],$$

where the expectation is taken with respect to everything except  $\hat{\pi}$ ,  $\hat{d}_z$  and  $\hat{d}_w$ .

We define an oracle optimal ITR which assumes  $\bar{\pi}(X; \hat{d}_z, \hat{d}_w)$  is known,

$$\hat{d}_{zw}^{\bar{\pi}}(X, W, Z) \triangleq \bar{\pi}(X; \hat{d}_z, \hat{d}_w) \hat{d}_z(X, Z) + (1 - \bar{\pi}(X; \hat{d}_z, \hat{d}_w)) \hat{d}_w(X, W).$$

The corresponding value function of this oracle optimal ITR is given by

$$V(\hat{d}_{zw}^{\bar{\pi}}) = \mathbb{E}[\bar{\pi}(X; \hat{d}_z, \hat{d}_w) h(W, \hat{d}_z(X, Z), X) + (1 - \bar{\pi}(X; \hat{d}_z, \hat{d}_w)) Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\}],$$

where the expectation is taken with respect to everything except  $\hat{d}_z$  and  $\hat{d}_w$ .

Then the approximation error incurred by estimating  $\hat{\pi}(X; \hat{d}_z, \hat{d}_w)$  is given by

$$\mathbb{K}(\hat{\pi}) \triangleq V(\hat{d}_{zw}^{\bar{\pi}}) - V(\hat{d}_{zw}^{\hat{\pi}}).$$

Moreover, we define the following gain

$$\mathbb{G}(\bar{\pi}) \triangleq \min\{V(\hat{d}_{zw}^{\bar{\pi}}) - V(\hat{d}_z), V(\hat{d}_{zw}^{\bar{\pi}}) - V(\hat{d}_w)\}.$$

It is clear that this gain  $\mathbb{G}(\bar{\pi})$  by introducing  $\bar{\pi}$  is always non-negative as indicated by Theorem 2. Then we have the following excess value bound for the value of  $\hat{d}_{zw}^{\hat{\pi}}$  compared to existing ones in the literature.

**Proposition 2.** *Under Assumptions 1-5,*

$$V(\hat{d}_{zw}^{\hat{\pi}}) = \max\{V(\hat{d}_z), V(\hat{d}_w)\} - \mathbb{K}(\hat{\pi}) + \mathbb{G}(\bar{\pi}) = V(\hat{d}_{z \cup w}) - \mathbb{K}(\hat{\pi}) + \mathbb{G}(\bar{\pi}).$$

Proposition 2 establishes a link between the value function of the estimated ITR  $\hat{d}_{zw}^{\hat{\pi}}$ , and that of  $\hat{d}_z$ ,  $\hat{d}_w$ , and  $\hat{d}_{z \cup w}$ . As shown in Appendix G,  $\mathbb{K}(\hat{\pi})$  diminishes as the sample size increases, therefore,  $\hat{d}_{zw}^{\hat{\pi}}$  has a significant improvement compared to other optimal ITRs depending on the magnitude of  $\mathbb{G}(\bar{\pi})$ .

Furthermore, we establish the consistency of the proposed regime based on the following assumption, which holds for example when  $\hat{d}_z$  and  $\hat{d}_w$  are estimated using indirect methods.

**Assumption 6.** *For  $\hat{d}_z, \hat{d}_w$ ,  $E[h(W, \hat{d}_z(X, Z), X)|X] - E[h(W, d_z^*(X, Z), X)|X] = o_p(n^{-\xi})$  almost surely and  $E[Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\}|X] - E[Y q(Z, A, X) \mathbb{I}\{d_w^*(X, W) = A\}|X] = o_p(n^{-\varphi})$  almost surely.*

**Proposition 3.** *Under Assumptions 1-6, we have  $V(\hat{d}_{zw}^{\hat{\pi}}) \xrightarrow{p} V(d_{zw}^{\bar{\pi}*})$ .*

## 4 Numerical Experiments

The data generating mechanism for  $(X, A, Z, W, U)$  follows the setup proposed in Cui et al. (2023) and is summarized in Appendix I. To evaluate the performance of the proposed framework, we vary  $b_1(X)$ ,  $b_2(X)$ ,  $b_3(X)$ ,  $b_a$  and  $b_w$  in  $\mathbb{E}[Y|X, A, Z, W, U]$  to incorporate heterogeneous treatment effects including the settings considered in Qi et al. (2023). The adopted data generating mechanism is compatible with the following  $h$  and  $q$ ,

$$h(W, A, X) = b_0 + \{b_1(X) + b_a W + b_3(X)W\} \frac{1+A}{2} + b_w W + b_2(X)X,$$

$$q(Z, A, X) = 1 + \exp \left\{ At_0 + At_z Z + t_a \frac{1+A}{2} + At_x X \right\},$$

where  $t_0 = 0.25$ ,  $t_z = -0.5$ ,  $t_a = -0.125$ , and  $t_x = (0.25, 0.25)^T$ . We derive preliminary optimal ITRs  $d_z^*$  and  $d_w^*$  in Appendix J, from which we can see that  $X, Z, W$  are relevant variables for individualized decision-making.

We consider six scenarios in total, and the setups of varying parameters are deferred to Appendix I. For each scenario, training datasets  $\{Y_i, A_i, X_i, Z_i, W_i\}_{i=1}^n$  are generated following the above mechanism with a sample size  $n = 1000$ . For each training dataset, we then apply the aforementioned methods to learn the optimal ITR. In particular, the preliminary ITRs  $\hat{d}_z$  and  $\hat{d}_w$  are estimated using a linear decision rule, and  $\hat{\pi}(x; \hat{d}_z, \hat{d}_w)$  is estimated using a Gaussian kernel. More details can be found in the Appendix K.

To evaluate the estimated treatment regimes, we consider the following generating mechanism for testing datasets:  $X \sim \mathcal{N}(\Gamma_x, \Sigma_x)$ ,

$$(Z, W, U)|X \sim \mathcal{N} \left\{ \begin{pmatrix} \alpha_0 + \alpha_a p_a + \alpha_x X \\ \mu_0 + \mu_a p_a + \mu_x X \\ \kappa_0 + \kappa_a p_a + \kappa_x X \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_z^2 & \sigma_{zw} & \sigma_{zu} \\ \sigma_{zw} & \sigma_w^2 & \sigma_{wu} \\ \sigma_{zu} & \sigma_{wu} & \sigma_u^2 \end{pmatrix} \right\},$$

where the parameter settings can be found in Appendix I. The testing dataset is generated with a size 10000, and the empirical value function for the estimated ITR is used as a performance measure. The simulations are replicated 200 times. To validate our approach

and demonstrate its superiority, we have also computed empirical values for other optimal policies, including existing optimal ITRs for proximal causal inference, as discussed in Section 2.2, along with optimal ITRs generated through causal forest (Athey and Wager, 2019) and outcome weighted learning (Zhao et al., 2012).

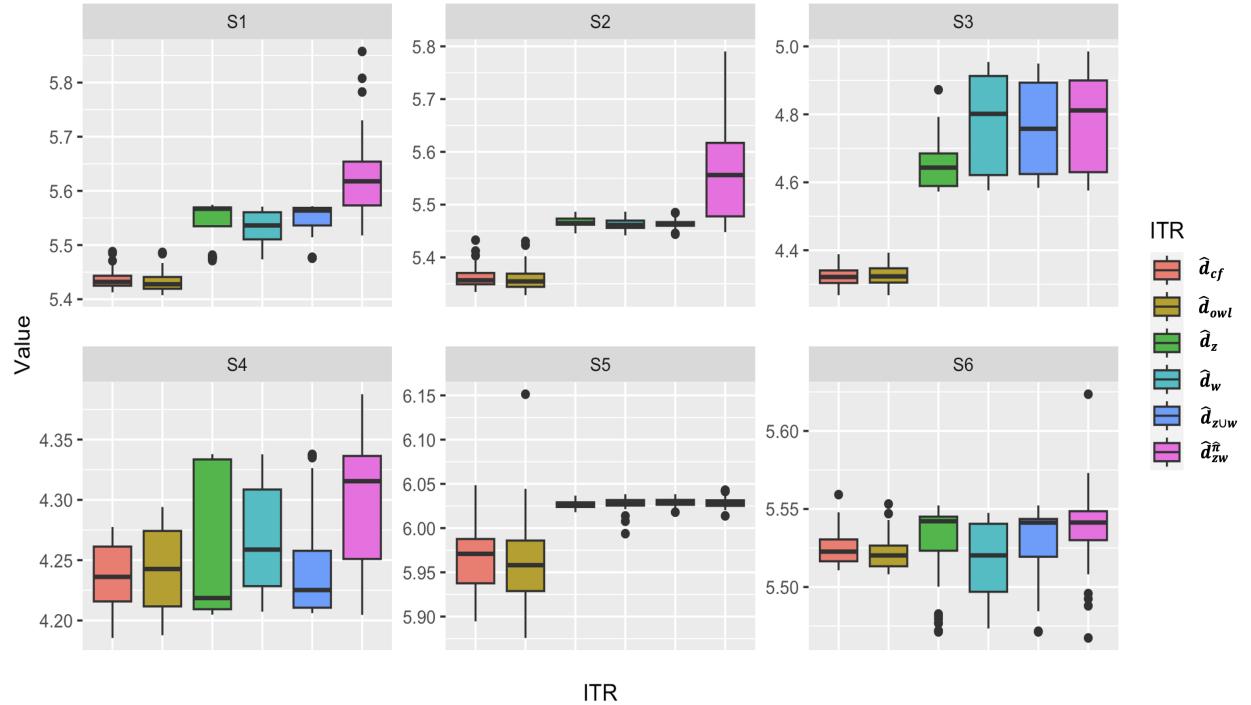


Figure 2: Boxplots of the empirical value functions ( $\hat{d}_{cf}$  and  $\hat{d}_{owl}$  denote estimated ITRs using causal forest and outcome weighted learning respectively).

Figure 2 presents the empirical value functions of different optimal ITRs for the six scenarios. As expected,  $\hat{d}_z$ ,  $\hat{d}_w$ ,  $\hat{d}_{z\cup w}$ , and  $\hat{d}_{z\hat{w}}$  consistently outperform  $\hat{d}_{cf}$  and  $\hat{d}_{owl}$ , which highlights their effectiveness in addressing unmeasured confounding. Meanwhile, across all scenarios,  $\hat{d}_{z\hat{w}}$  yields superior or comparable performance compared to the other estimated treatment regimes, which justifies the statements made in Sections 2 and 3. In addition, as can be seen in Scenario 5, all ITRs relying on the proximal causal inference framework perform similarly, which is not surprising as  $\hat{d}_z(X, Z)$  and  $\hat{d}_w(X, W)$  agree for most subjects. To further underscore the robust performance of our proposed approach, we include additional

results with a changed sample size and a modified behavior policy in Appendix L.

## 5 Real Data Application

In this section, we demonstrate the proposed optimal ITR via a real dataset originally designed to measure the effectiveness of right heart catheterization (RHC) for ill patients in intensive care units (ICU), under the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT, Connors et al. (1996)). These data have been re-analyzed in a number of papers in both causal inference and survival analysis literature with assuming unconfoundedness (Cui and Tchetgen Tchetgen, 2023; Tan, 2006, 2020, 2019; Vermeulen and Vansteelandt, 2015) or accounting for unmeasured confounding (Cui et al., 2023; Lin et al., 1998; Qi et al., 2023; Tchetgen Tchetgen et al., 2020; Ying et al., 2022).

There are 5735 subjects included in the dataset, in which 2184 were treated (with  $A = 1$ ) and 3551 were untreated (with  $A = -1$ ). The outcome  $Y$  is the duration from admission to death or censoring. Overall, 3817 patients survived and 1918 died within 30 days. Following Tchetgen Tchetgen et al. (2020), we collect 71 covariates including demographic factors, diagnostic information, estimated survival probability, comorbidity, vital signs, physiological status, and functional status (see Hirano and Imbens (2001) for additional discussion on covariates). Confounding in this study stems from the fact that ten physiological status measures obtained from blood tests conducted at the initial phase of admission may be susceptible to significant measurement errors. Furthermore, besides the lab measurement errors, whether other unmeasured confounding factors exist is unknown to the data analyst. Because variables measured from these tests offer only a single snapshot of the underlying physiological condition, they have the potential to act as confounding proxies. We consider a total of four settings, varying the number of selected proxies from 4 to 10. Within each setting, treatment-inducing proxies are first selected based on their strength of association with the treatment (determined through logistic regression of  $A$  on  $L$ ), and outcome-inducing

proxies are then chosen based on their association with the outcome (determined through linear regression of  $Y$  on  $A$  and  $L$ ). Excluding the selected proxy variables, other measured covariates are included in  $X$ . We then estimate  $\hat{d}_z$ ,  $\hat{d}_w$ ,  $\hat{d}_{z\cup w}$ , and  $\hat{d}_{zw}^{\hat{\pi}}$  using the SUPPORT dataset in a manner similar to that described in Section 4, with the goal optimizing the patients' 30-day survival after their entrance into the ICU.

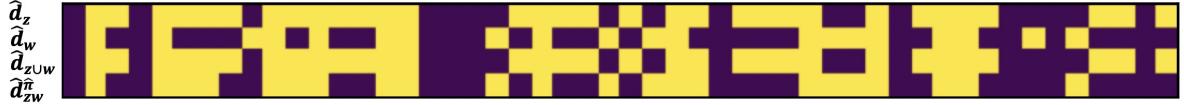


Figure 3: Graphical representation of concordance between estimated ITRs.

The estimated value functions of our proposed ITR, alongside existing ones, are summarized in Appendix M. As can be seen, our proposed regime has the largest value among all settings. For a visual representation of the concordance between the estimated optimal ITRs, we refer to Figure 3 (results from Setting 1). The horizontal ordinate represents the 50 selected subjects and the vertical axis denotes the decisions made from corresponding ITRs. The purple and yellow blocks stand for being recommended treatment values of -1 and 1 respectively. For the subjects with purple or yellow columns,  $\hat{d}_z(X, Z) = \hat{d}_w(X, W)$ , which leads to the same treatment decision for the other two ITRs. For columns with mixed colors,  $\hat{d}_z(X, Z)$  and  $\hat{d}_w(X, W)$  disagree. We see that in this case  $\hat{d}_{z\cup w}(X, W, Z)$  always agree with  $\hat{d}_z(X, Z)$ , while  $\hat{d}_{zw}^{\hat{\pi}}(X, W, Z)$  take values from  $\hat{d}_z(X, Z)$  or  $\hat{d}_w(X, W)$  depending on the individual criteria of the subjects as indicated by  $\hat{\pi}$ . In addition to the quantitative analysis, we have also conducted a qualitative assessment of the estimated regime to validate its performance. For further details, please refer to Appendix M.

## 6 Discussion

We acknowledge several limitations of our work. Firstly, the proximal causal inference framework relies on the validity of treatment- and outcome-inducing confounding proxies. When

the assumptions are violated, the proximal causal inference estimators can be biased even if unconfoundedness on the basis of measured covariates in fact holds. Therefore, one needs to carefully sort out proxies especially when domain knowledge is lacking. Secondly, while the proposed regime significantly improves upon existing methods both theoretically and numerically, it is not yet shown to be the sharpest under our considered model. It is still an open question to figure out if a more general policy class could be considered. Thirdly, our established theory provides consistency and superiority of our estimated regime. It is of great interest to derive convergence rates for Propositions 2 and 3 following [Jiang \(2017\)](#). In addition, it may be challenging to develop inference results for the value function of the estimated optimal treatment regimes, and further studies are warranted.

## Acknowledgement

Yifan Cui was supported by the National Natural Science Foundation of China.

## References

- J. D. Angrist, G. W. Imbens, and D. B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.
- S. Athey and S. Wager. Estimating treatment effects with causal forests: An application. *Observational Studies*, 5(2):37–51, 2019.
- S. Athey and S. Wager. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021.
- A. Bennett and N. Kallus. Proximal reinforcement learning: Efficient off-policy evaluation in partially observed markov decision processes. *Operations Research*, 09 2023.
- B. Chakraborty and E. Moodie. *Statistical methods for dynamic treatment regimes*. Springer, 2013.
- X. Chen and T. Christensen. Optimal uniform convergence rates for sieve nonparametric instrumental variables regression. *arXiv preprint arXiv:1311.0412*, 2013.
- Y. Chen, D. Zeng, T. Xu, and Y. Wang. Representation learning for integrating multi-domain outcomes to optimize individualized treatment. *Advances in Neural Information Processing Systems*, 33:17976–17986, 2020.
- Y.-C. Chen. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1(1):161–187, 2017.
- V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

- A. F. Connors, T. Speroff, N. V. Dawson, C. Thomas, F. E. Harrell, D. Wagner, N. Desbiens, L. Goldman, A. W. Wu, R. M. Califf, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients. *Jama*, 276(11):889–897, 1996.
- Y. Cui. Individualized decision-making under partial identification: Three perspectives, two optimality results, and one paradox. *Harvard Data Science Review*, 3(3), 2021.
- Y. Cui and E. Tchetgen Tchetgen. On a necessary and sufficient identification condition of optimal treatment regimes with an instrumental variable. *Statistics & Probability Letters*, 178:109180, 2021a. ISSN 0167-7152.
- Y. Cui and E. Tchetgen Tchetgen. A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, 116(533):162–173, 2021b.
- Y. Cui and E. Tchetgen Tchetgen. Selective machine learning of doubly robust functionals. *Biometrika*, page asad055, 2023. ISSN 1464-3510.
- Y. Cui, H. Pu, X. Shi, W. Miao, and E. Tchetgen Tchetgen. Semiparametric proximal causal inference. *Journal of the American Statistical Association*, pages 1–12, 2023.
- N. Dalmasso, T. Pospisil, A. B. Lee, R. Izbicki, P. E. Freeman, and A. I. Malz. Conditional density estimation tools in python and r with applications to photometric redshifts and likelihood-free cosmological inference. *Astronomy and Computing*, 30:100362, 2020.
- N. Dikkala, G. Lewis, L. Mackey, and V. Syrgkanis. Minimax estimation of conditional moment models. *Advances in Neural Information Processing Systems*, 33:12248–12262, 2020.
- L. Dinh, J. Sohl-Dickstein, and S. Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

- O. Dukes, I. Shpitser, and E. J. Tchetgen Tchetgen. Proximal mediation analysis. *Biometrika*, page asad015, 03 2023. ISSN 1464-3510.
- N. Galie, M. M. Hoeper, M. Humbert, A. Torbicki, J.-L. Vachiery, J. A. Barbera, M. Beghetti, P. Corris, S. Gaine, J. S. Gibbs, et al. Guidelines for the diagnosis and treatment of pulmonary hypertension: the task force for the diagnosis and treatment of pulmonary hypertension of the european society of cardiology (esc) and the european respiratory society (ers), endorsed by the international society of heart and lung transplantation (ishlt). *European heart journal*, 30(20):2493–2537, 2009.
- A. Ghassami, A. Ying, I. Shpitser, and E. Tchetgen Tchetgen. Minimax kernel machine learning for a class of doubly robust functionals with application to proximal causal inference. In *International Conference on Artificial Intelligence and Statistics*, pages 7210–7239. PMLR, 2022.
- A. Ghassami, I. Shpitser, and E. T. Tchetgen. Partial identification of causal effects using proxy variables. *arXiv preprint arXiv:2304.04374*, 2023.
- S. Han. Optimal dynamic treatment regimes and partial welfare ordering. *Journal of the American Statistical Association*, pages 1–11, 2023.
- K. Hirano and G. W. Imbens. Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes research methodology*, 2(3):259–278, 2001.
- G. W. Imbens and J. D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994. ISSN 00129682, 14680262.
- B. Jiang, R. Song, J. Li, and D. Zeng. Entropy learning for dynamic treatment regimes. *Statistica Sinica*, 29(4):1633, 2019.
- H. Jiang. Uniform convergence rates for kernel density estimation. In *International Conference on Machine Learning*, pages 1694–1703. PMLR, 2017.

- N. Kallus, X. Mao, and M. Uehara. Causal inference under unmeasured confounding with negative controls: A minimax learning approach. *arXiv preprint arXiv:2103.14029*, 2021.
- T. Kitagawa and A. Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.
- B. Kompa, D. Bellamy, T. Kolokotrones, A. Beam, et al. Deep learning methods for proximal inference via maximum moment restriction. *Advances in Neural Information Processing Systems*, 35:11189–11201, 2022.
- M. R. Kosorok and E. B. Laber. Precision medicine. *Annual Review of Statistics and Its Application*, 6:263–286, 2019.
- R. Kress, V. Maz'ya, and V. Kozlov. *Linear integral equations*, volume 82. Springer, 1989.
- M. Kuroki and J. Pearl. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2):423–437, 2014.
- K. Q. Li, X. Shi, W. Miao, and E. Tchetgen Tchetgen. Double negative control inference in test-negative design studies of vaccine effectiveness. *Journal of the American Statistical Association*, pages 1–12, 2023.
- L. Liao, Z. Fu, Z. Yang, Y. Wang, M. Kolar, and Z. Wang. Instrumental variable value iteration for causal offline reinforcement learning. *arXiv preprint arXiv:2102.09907*, 2021.
- D. Y. Lin, B. M. Psaty, and R. A. Kronmal. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*, pages 948–963, 1998.
- A. Mastouri, Y. Zhu, L. Gultchin, A. Korba, R. Silva, M. Kusner, A. Gretton, and K. Muandet. Proximal causal learning with kernels: Two-stage estimation and moment restriction. In *International Conference on Machine Learning*, pages 7512–7523. PMLR, 2021.
- W. Miao, Z. Geng, and E. J. Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018a.

- W. Miao, X. Shi, and E. Tchetgen Tchetgen. A confounding bridge approach for double negative control inference on causal effects. *arXiv preprint arXiv:1808.04945*, 2018b.
- W. Miao, W. Hu, E. L. Ogburn, and X.-H. Zhou. Identifying effects of multiple treatments in the presence of unmeasured confounding. *Journal of the American Statistical Association*, pages 1–15, 2022.
- S. A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.
- E. A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.
- H. Pu and B. Zhang. Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(2):318–345, 2021.
- Z. Qi, R. Miao, and X. Zhang. Proximal learning for individualized treatment regimes under unmeasured confounding. *Journal of the American Statistical Association*, pages 1–14, 2023.
- M. Qian and S. A. Murphy. Performance guarantees for individualized treatment rules. *Annals of Statistics*, 39(2):1180, 2011.
- H. Qiu, M. Carone, E. Sadikova, M. Petukhova, R. C. Kessler, and A. Luedtke. Optimal individualized decision rules using instrumental variable methods. *Journal of the American Statistical Association*, 116(533):174–191, 2021.
- A. Raghu, M. Komorowski, L. A. Celi, P. Szolovits, and M. Ghassemi. Continuous state-space models for optimal sepsis treatment: a deep reinforcement learning approach. In *Machine Learning for Healthcare Conference*, pages 147–163. PMLR, 2017.

- J. Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512, 1986.
- J. M. Robins. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics: Theory and Methods*, 23(8):2379–2412, 1994.
- J. M. Robins. Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality*, pages 69–117. Springer, 1997.
- J. M. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.
- P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.
- A. Rotnitzky, J. M. Robins, and D. O. Scharfstein. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*, 93(444):1321–1339, 1998.
- D. O. Scharfstein, A. Rotnitzky, and J. M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999.
- D. W. Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.
- X. Shi, W. Miao, J. C. Nelson, and E. J. Tchetgen Tchetgen. Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(2):521–540, 2020a.

- X. Shi, W. Miao, and E. Tchetgen Tchetgen. A selective review of negative control methods in epidemiology. *Current Epidemiology Reports*, 7(4):190–202, 2020b.
- X. Shi, W. Miao, M. Hu, and E. Tchetgen Tchetgen. Theory for identification and inference with synthetic controls: a proximal causal inference framework. *arXiv preprint arXiv:2108.13935*, 2021.
- I. Shpitser, Z. Wood-Doughty, and E. J. T. Tchetgen. The proximal id algorithm. *Journal of Machine Learning Research*, 23:1–46, 2023.
- R. Singh. Kernel methods for unobserved confounding: Negative controls, proxies, and instruments. *arXiv preprint arXiv:2012.10315*, 2020.
- K. Sohn, H. Lee, and X. Yan. Learning structured output representation using deep conditional generative models. *Advances in Neural Information Processing Systems*, 28, 2015.
- M. J. Stensrud and A. L. Sarvet. Optimal regimes for algorithm-assisted human decision-making. *arXiv preprint arXiv:2203.03020*, 2022.
- E. Sverdrup and Y. Cui. Proximal causal learning of heterogeneous treatment effects. In *International Conference on Machine Learning*, 2023.
- Z. Tan. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101(476):1619–1637, 2006.
- Z. Tan. Regularized calibrated estimation of propensity scores with model misspecification and high-dimensional data. *Biometrika*, 107(1):137–158, 12 2019. ISSN 0006-3444.
- Z. Tan. Model-assisted inference for treatment effects using regularized calibrated estimation with high-dimensional data. *The Annals of Statistics*, 48(2):811–837, 2020.
- E. Tchetgen Tchetgen. The control outcome calibration approach for causal inference with unobserved confounding. *American journal of epidemiology*, 179(5):633–640, 2014.

- E. J. Tchetgen Tchetgen, A. Ying, Y. Cui, X. Shi, and W. Miao. An introduction to proximal causal learning. *arXiv preprint arXiv:2009.10982*, 2020.
- A. A. Tsiatis, M. Davidian, S. T. Holloway, and E. B. Laber. *Dynamic treatment regimes: Statistical methods for precision medicine*. Chapman and Hall/CRC, 2019.
- K. Vermeulen and S. Vansteelandt. Bias-reduced doubly robust estimation. *Journal of the American Statistical Association*, 110(511):1024–1036, 2015.
- J. Wang, Z. Qi, and C. Shi. Blessing from experts: Super reinforcement learning in confounded environments. *arXiv preprint arXiv:2209.15448*, 2022.
- P. Wu, D. Zeng, and Y. Wang. Matched learning for optimizing individualized treatment strategies using electronic health records. *Journal of the American Statistical Association*, 2019.
- A. Ying, Y. Cui, and E. J. T. Tchetgen. Proximal causal inference for marginal counterfactual survival curves. *arXiv preprint arXiv:2204.13144*, 2022.
- A. Ying, W. Miao, X. Shi, and E. J. T. Tchetgen. Proximal causal inference for complex longitudinal studies. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2023. In press.
- J. Yoon, J. Jordon, and M. Van Der Schaar. Ganite: Estimation of individualized treatment effects using generative adversarial nets. *International Conference on Learning Representations*, 2018.
- B. Zhang, A. A. Tsiatis, E. B. Laber, and M. Davidian. A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018, 2012.
- Y. Zhao, D. Zeng, A. J. Rush, and M. R. Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118, 2012.

Y.-Q. Zhao, E. B. Laber, Y. Ning, S. Saha, and B. E. Sands. Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *Journal of Machine Learning Research*, 20(1):1821–1843, 2019.

## Supplementary Material

### A Proof of identification (3)

The proof is straightforward. We state it here for clarity and completeness. Note that

$$\mathbb{I}\{d_{z \cup w}(X, W, Z) = 1\} = \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{Z}}\} \mathbb{I}\{d_{z \cup w}(X, Z) = 1\} + \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{W}}\} \mathbb{I}\{d_{z \cup w}(X, W) = 1\},$$

$$\mathbb{I}\{d_{z \cup w}(X, W, Z) = -1\} = \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{Z}}\} \mathbb{I}\{d_{z \cup w}(X, Z) = -1\} + \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{W}}\} \mathbb{I}\{d_{z \cup w}(X, W) = -1\}.$$

Therefore, we have

$$\begin{aligned} \mathbb{E}[Y(1)\mathbb{I}\{d_{z \cup w}(X, W, Z) = 1\}] &= \mathbb{E}[Y(1)\mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{Z}}\} \mathbb{I}\{d_{z \cup w}(X, Z) = 1\}] \\ &\quad + Y(1)\mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{W}}\} \mathbb{I}\{d_{z \cup w}(X, W) = 1\}] \\ &= \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{Z}}\} \mathbb{E}[Y(1)\mathbb{I}\{d_{z \cup w}(X, Z) = 1\}] \\ &\quad + \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{W}}\} \mathbb{E}[Y(1)\mathbb{I}\{d_{z \cup w}(X, W) = 1\}]. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{E}[Y(-1)\mathbb{I}\{d_{z \cup w}(X, W, Z) = -1\}] &= \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{Z}}\} \mathbb{E}[Y(-1)\mathbb{I}\{d_{z \cup w}(X, Z) = -1\}] \\ &\quad + \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{W}}\} \mathbb{E}[Y(-1)\mathbb{I}\{d_{z \cup w}(X, W) = -1\}]. \end{aligned}$$

So

$$\begin{aligned} V(d_{z \cup w}) &= \mathbb{E}[Y(1)\mathbb{I}\{d_{z \cup w}(X, W, Z) = 1\}] + \mathbb{E}[Y(-1)\mathbb{I}\{d_{z \cup w}(X, W, Z) = -1\}] \\ &= \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{Z}}\} \mathbb{E}[Y(1)\mathbb{I}\{d_{z \cup w}(X, Z) = 1\} + Y(-1)\mathbb{I}\{d_{z \cup w}(X, Z) = -1\}] \\ &\quad + \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{W}}\} \mathbb{E}[Y(1)\mathbb{I}\{d_{z \cup w}(X, W) = 1\} + Y(-1)\mathbb{I}\{d_{z \cup w}(X, W) = -1\}] \\ &= \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{Z}}\} \mathbb{E}[h(W, d_{z \cup w}(X, Z), X)] + \mathbb{I}\{d_{z \cup w} \in \mathcal{D}_{\mathcal{W}}\} \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_{z \cup w}(X, W) = A\}], \end{aligned}$$

where the last equality holds due to identification results (1) and (2).

## B Proof of Theorem 1

Recall that  $V(d_{zw}^\pi) = \mathbb{E}[Y(1)\mathbb{I}\{d_{zw}^\pi(X, W, Z) = 1\} + Y(-1)\mathbb{I}\{d_{zw}^\pi(X, W, Z) = -1\}]$ , we essentially need to consider the first term  $\mathbb{E}[Y(1)\mathbb{I}\{d_{zw}^\pi(X, W, Z) = 1\}]$ . Note that

$$\mathbb{I}\{d_{zw}^\pi(X, W, Z) = 1\} = \mathbb{I}\{\pi(X) = 1\}\mathbb{I}\{d_z(X, Z) = 1\} + \mathbb{I}\{\pi(X) = 0\}\mathbb{I}\{d_w(X, W) = 1\},$$

we have

$$\begin{aligned} \mathbb{E}[Y(1)\mathbb{I}\{d_{zw}^\pi(X, W, Z) = 1\}] &= \mathbb{E}[Y(1)\mathbb{I}\{\pi(X) = 1\}\mathbb{I}\{d_z(X, Z) = 1\} + Y(1)\mathbb{I}\{\pi(X) = 0\}\mathbb{I}\{d_w(X, W) = 1\}] \\ &= \mathbb{E}[\mathbb{I}\{\pi(X) = 1\}\mathbb{E}[Y(1)\mathbb{I}\{d_z(X, Z) = 1\}|X]] \\ &\quad + \mathbb{I}\{\pi(X) = 0\}\mathbb{E}[Y(1)\mathbb{I}\{d_w(X, W) = 1\}|X]]. \end{aligned}$$

By leveraging the outcome confounding bridge, we have

$$\begin{aligned} \mathbb{E}[Y(1)\mathbb{I}\{d_z(X, Z) = 1\}|X] &= \mathbb{E}[\mathbb{E}[Y(1)|X, Z]\mathbb{I}\{d_z(X, Z) = 1\}|X] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y(1)|X, Z, U]|X, Z]\mathbb{I}\{d_z(X, Z) = 1\}|X] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y|X, U, A = 1]|X, Z]\mathbb{I}\{d_z(X, Z) = 1\}|X] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{E}[h(W, 1, X)|X, U]|X, Z]\mathbb{I}\{d_z(X, Z) = 1\}|X] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{E}[h(W, 1, X)|X, Z, U]|X, Z]\mathbb{I}\{d_z(X, Z) = 1\}|X] \\ &= \mathbb{E}[h(W, 1, X)\mathbb{I}\{d_z(X, Z) = 1\}|X], \end{aligned}$$

where the third equality is due to Assumption 1, the fourth equality can be verified by Theorem 1 in [Miao et al. \(2018a\)](#) under Assumptions 2 and 3, and the fifth equality is due

to Assumption 1. Moreover, by leveraging the treatment confounding bridge, we have

$$\begin{aligned}
\mathbb{E}[Y(1)\mathbb{I}\{d_w(X, W) = 1\}|X] &= \mathbb{E}[\mathbb{E}[Y(1)|X, W]\mathbb{I}\{d_w(X, W) = 1\}|X] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y(1)|X, W, U]|X, W]\mathbb{I}\{d_w(X, W) = 1\}|X] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y(1)|X, W, U, A = 1]|X, W]\mathbb{I}\{d_w(X, W) = 1\}|X] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y(1)|X, W, U, A = 1]\mathbb{E}[q(Z, 1, X)|X, U, A = 1] \\
&\quad \mathbb{P}(A = 1|X, U)|X, W]\mathbb{I}\{d_w(X, W) = 1\}|X] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[Yq(Z, 1, X)\mathbb{I}\{A = 1\}|X, U, W]|X, W]\mathbb{I}\{d_w(X, W) = 1\}|X] \\
&= \mathbb{E}[Yq(Z, 1, X)\mathbb{I}\{A = 1\}\mathbb{I}\{d_w(X, W) = 1\}|X],
\end{aligned}$$

where the third equality is due to Assumption 1, the fourth equality is implied by Theorem 2.2 of Cui et al. (2023) under Assumptions 4 and 5, and the fifth equality is due to Assumption 1. Therefore,

$$\begin{aligned}
\mathbb{E}[Y(1)\mathbb{I}\{d_{zw}^\pi(X, W, Z) = 1\}] &= \mathbb{E}[\mathbb{I}\{\pi(X) = 1\}\mathbb{E}[Y(1)\mathbb{I}\{d_z(X, Z) = 1\}|X] \\
&\quad + \mathbb{I}\{\pi(X) = 0\}\mathbb{E}[Y(1)\mathbb{I}\{d_w(X, W) = 1\}|X]] \\
&= \mathbb{E}[\mathbb{I}\{\pi(X) = 1\}\mathbb{E}[h(W, 1, X)\mathbb{I}\{d_z(X, Z) = 1\}|X] \\
&\quad + \mathbb{I}\{\pi(X) = 0\}\mathbb{E}[Yq(Z, 1, X)\mathbb{I}\{A = 1\}\mathbb{I}\{d_w(X, W) = 1\}|X]] \\
&= \mathbb{E}[\mathbb{I}\{\pi(X) = 1\}h(W, 1, X)\mathbb{I}\{d_z(X, Z) = 1\} \\
&\quad + \mathbb{I}\{\pi(X) = 0\}Yq(Z, 1, X)\mathbb{I}\{A = 1\}\mathbb{I}\{d_w(X, W) = 1\}]. \quad (10)
\end{aligned}$$

Similarly, as

$$\mathbb{I}\{d_{zw}^\pi(X, W, Z) = -1\} = \mathbb{I}\{\pi(X) = 1\}\mathbb{I}\{d_z(X, Z) = -1\} + \mathbb{I}\{\pi(X) = 0\}\mathbb{I}\{d_w(X, W) = -1\},$$

we have

$$\begin{aligned}
\mathbb{E}[Y(-1)\mathbb{I}\{d_{zw}^\pi(X, W, Z) = -1\}] &= \mathbb{E}[\mathbb{I}\{\pi(X) = 1\}h(W, -1, X)\mathbb{I}\{d_z(X, Z) = -1\} \\
&\quad + \mathbb{I}\{\pi(X) = 0\}Yq(Z, -1, X)\mathbb{I}\{A = -1\}\mathbb{I}\{d_w(X, W) = -1\}]. \quad (11)
\end{aligned}$$

Combining (10) and (11), we have

$$\begin{aligned}
V(d_{zw}^\pi) &= \mathbb{E}[Y(1)\mathbb{I}\{d_{zw}^\pi(X, W, Z) = 1\} + Y(-1)\mathbb{I}\{d_{zw}^\pi(X, W, Z) = -1\}] \\
&= \mathbb{E}[\mathbb{I}\{\pi(X) = 1\}h(W, 1, X)\mathbb{I}\{d_z(X, Z) = 1\} \\
&\quad + \mathbb{I}\{\pi(X) = 0\}Yq(Z, 1, X)\mathbb{I}\{A = 1\}\mathbb{I}\{d_w(X, W) = 1\}] \\
&\quad + \mathbb{I}\{\pi(X) = 1\}h(W, -1, X)\mathbb{I}\{d_z(X, Z) = -1\} \\
&\quad + \mathbb{I}\{\pi(X) = 0\}Yq(Z, -1, X)\mathbb{I}\{A = -1\}\mathbb{I}\{d_w(X, W) = -1\}] \\
&= \mathbb{E}[\mathbb{I}\{\pi(X) = 1\}h(W, d_z(X, Z), X) + \mathbb{I}\{\pi(X) = 0\}Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}] \\
&= \mathbb{E}[\pi(X)h(W, d_z(X, Z), X) + (1 - \pi(X))Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}],
\end{aligned}$$

which completes the proof.

## C Proof of Theorem 2

For any  $d_z \in \mathcal{D}_Z$  and  $d_w \in \mathcal{D}_W$ , we have

$$\begin{aligned}
V(d_{zw}^{\bar{\pi}}) &= \mathbb{E}[\bar{\pi}(X; d_z, d_w)h(W, d_z(X, Z), X) + (1 - \bar{\pi}(X; d_z, d_w))Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}] \\
&= \mathbb{E}[\bar{\pi}(X; d_z, d_w)\mathbb{E}[h(W, d_z(X, Z), X)|X] + (1 - \bar{\pi}(X; d_z, d_w))\mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]] \\
&= \mathbb{E}[\max\{\mathbb{E}[h(W, d_z(X, Z), X)|X], \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]\}],
\end{aligned}$$

where the last equality is due to the definition of  $\bar{\pi}(X; d_z, d_w)$ . As

$$\max\{\mathbb{E}[h(W, d_z(X, Z), X)|X], \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]\} \geq \mathbb{E}[h(W, d_z(X, Z), X)|X],$$

and

$$\max\{\mathbb{E}[h(W, d_z(X, Z), X)|X], \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]\} \geq \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X].$$

taking expectations on both sides, we have

$$\begin{aligned}
& \mathbb{E}[\max\{\mathbb{E}[h(W, d_z(X, Z), X)|X], \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]\}] \\
& \geq \mathbb{E}[\mathbb{E}[h(W, d_z(X, Z), X)|X]] \\
& = \mathbb{E}[h(W, d_z(X, Z), X)] \\
& = V(d_z), \\
& \mathbb{E}[\max\{\mathbb{E}[h(W, d_z(X, Z), X)|X], \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]\}] \\
& \geq \mathbb{E}[\mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]] \\
& = \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}] \\
& = V(d_w).
\end{aligned}$$

Therefore, we have  $V(d_{zw}^{\bar{\pi}}) \geq \max\{V(d_z), V(d_w)\}$ .

## D Proof of Corollary 1

Recall that

$$d_{zw}^{\bar{\pi}*}(X, W, Z) = \bar{\pi}(X; d_z^*, d_w^*)d_z^*(X, Z) + (1 - \bar{\pi}(X; d_z^*, d_w^*))d_w^*(X, W),$$

with

$$\bar{\pi}(X; d_z^*, d_w^*) = \mathbb{I}\{\mathbb{E}[h(W, d_z^*(X, Z), X|X] \geq \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w^*(X, W) = A\}|X]\},$$

we apply the result in Theorem 2, i.e.,

$$V(d_{zw}^{\bar{\pi}*}) \geq \max\{V(d_z^*), V(d_w^*)\}.$$

Due to  $V(d_{z \cup w}^*) = \max\{V(d_z^*), V(d_w^*)\}$  as shown in Qi et al. (2023), we then conclude that  $V(d_{zw}^{\bar{\pi}*}) \geq \max\{V(d_z^*), V(d_w^*), V(d_{z \cup w}^*)\}$ .

## E Proof of Proposition 1

In the following, we show

$$d_{zw}^{\bar{\pi}*} \in \arg \max_{d_{zw}^{\bar{\pi}} \in \mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\Pi}} V(d_{zw}^{\bar{\pi}}).$$

Recall that

$$\begin{aligned} V(d_{zw}^{\bar{\pi}}) &= \mathbb{E}[\bar{\pi}(X; d_z, d_w)h(W, d_z(X, Z), X) + (1 - \bar{\pi}(X; d_z, d_w))Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}] \\ &= \mathbb{E}[\bar{\pi}(X; d_z, d_w)\mathbb{E}[h(W, d_z(X, Z), X)|X] + (1 - \bar{\pi}(X; d_z, d_w))\mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]] \\ &= \mathbb{E}[\max\{\mathbb{E}[h(W, d_z(X, Z), X)|X], \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]\}] \\ &\geq \mathbb{E}[\pi(X)\mathbb{E}[h(W, d_z(X, Z), X)|X] + (1 - \pi(X))\mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]] \\ &= V(d_{zw}^{\pi}), \end{aligned}$$

for any  $d_z \in \mathcal{D}_{\mathcal{Z}}, d_w \in \mathcal{D}_{\mathcal{W}}$  and  $\pi(\cdot)$ . Therefore, we essentially need to show

$$d_{zw}^{\bar{\pi}*} \in \arg \max_{d_{zw}^{\bar{\pi}} \in \mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\bar{\pi}}} V(d_{zw}^{\bar{\pi}}),$$

where  $\mathcal{D}_{\mathcal{Z}\mathcal{W}}^{\bar{\pi}} \triangleq \{d_{zw}^{\bar{\pi}} : d_{zw}^{\bar{\pi}}(X, W, Z) = \bar{\pi}(X)d_z(X, Z) + (1 - \bar{\pi}(X))d_w(X, W), d_z \in \mathcal{D}_{\mathcal{Z}}, d_w \in \mathcal{D}_{\mathcal{W}}\}$ . Recall that

$$\begin{aligned} d_z^*(X, Z) &= \text{sign}\{\mathbb{E}[h(W, 1, X) - h(W, -1, X)|X, Z]\}, \\ d_w^*(X, W) &= \text{sign}\{\mathbb{E}[Yq(Z, A, X)\mathbb{I}\{A = 1\} - Yq(Z, A, X)\mathbb{I}\{A = -1\}|X, W]\}, \end{aligned}$$

we have

$$\begin{aligned} \mathbb{E}[h(W, d_z^*(X, Z), X)|X, Z] &\geq \mathbb{E}[h(W, d_z(X, Z), X)|X, Z], \\ \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{A = d_w^*(X, W)\}|X, W] &\geq \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{A = d_w(X, W)\}|X, W]. \end{aligned}$$

Taking expectation with respect to  $Z$  and  $W$  given  $X$  respectively, we have

$$\mathbb{E}[\mathbb{E}[h(W, d_z^*(X, Z), X)|X, Z]|X] \geq \mathbb{E}[\mathbb{E}[h(W, d_z(X, Z), X)|X, Z]|X], \quad (12)$$

$$\begin{aligned} \mathbb{E}[\mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w^*(X, W) = A\}|X, W]|X] &\geq \mathbb{E}[\mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X, W]|X]. \end{aligned} \quad (13)$$

By the proof given in Section C, we have that

$$V(d_{zw}^{\bar{\pi}*}) = \mathbb{E}[\max\{\mathbb{E}[h(W, d_z^*(X, Z), X)|X], \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w^*(X, W) = A\}|X]\}],$$

$$V(d_{zw}^{\bar{\pi}}) = \mathbb{E}[\max\{\mathbb{E}[h(W, d_z(X, Z), X)|X], \mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X]\}].$$

Therefore,

$$V(d_{zw}^{\bar{\pi}*}) = \mathbb{E}[\max\{\mathbb{E}[\mathbb{E}[h(W, d_z^*(X, Z), X)|X, Z]|X], \mathbb{E}[\mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w^*(X, W) = A\}|X, W]|X]\}],$$

$$V(d_{zw}^{\bar{\pi}}) = \mathbb{E}[\max\{\mathbb{E}[\mathbb{E}[h(W, d_z(X, Z), X)|X, Z]|X], \mathbb{E}[\mathbb{E}[Yq(Z, A, X)\mathbb{I}\{d_w(X, W) = A\}|X, W]|X]\}].$$

From (12) and (13), we have  $V(d_{zw}^{\bar{\pi}*}) \geq V(d_{zw}^{\bar{\pi}})$  for any  $d_{zw}^{\bar{\pi}} \in \mathcal{D}_{ZW}^{\bar{\pi}}$ , which implies that  $d_{zw}^{\bar{\pi}*}$  is the maximizer of  $V(d_{zw}^{\bar{\pi}})$ .

## F Proof of Proposition 2

By definition of  $\mathbb{K}(\hat{\pi})$  we have

$$V(\hat{d}_{zw}^{\hat{\pi}}) = V(\hat{d}_{zw}^{\bar{\pi}}) - \mathbb{K}(\hat{\pi}). \quad (14)$$

Then from

$$\mathbb{G}(\bar{\pi}) = \min\{V(\hat{d}_{zw}^{\bar{\pi}}) - V(\hat{d}_z), V(\hat{d}_{zw}^{\bar{\pi}}) - V(\hat{d}_w)\},$$

we can see

$$\max\{V(\hat{d}_z), V(\hat{d}_w)\} = V(\hat{d}_{zw}^{\bar{\pi}}) - \mathbb{G}(\bar{\pi}). \quad (15)$$

Finally, combining (14) and (15), we have

$$V(\hat{d}_{zw}^{\hat{\pi}}) = \max\{V(\hat{d}_z), V(\hat{d}_w)\} - \mathbb{K}(\hat{\pi}) + \mathbb{G}(\bar{\pi}).$$

As  $V(\hat{d}_{z \cup w}) = \max\{V(\hat{d}_z), V(\hat{d}_w)\}$ , we conclude that

$$V(\hat{d}_{zw}^{\hat{\pi}}) = \max\{V(\hat{d}_z), V(\hat{d}_w)\} - \mathbb{K}(\hat{\pi}) + \mathbb{G}(\bar{\pi}) = V(\hat{d}_{z \cup w}) - \mathbb{K}(\hat{\pi}) + \mathbb{G}(\bar{\pi}).$$

## G Asymptotics of $\mathbb{K}(\hat{\pi})$

Throughout this section, we assume that  $X \in [0, 1]^p$  has a bounded density  $f(x)$  and  $\max\{|Y|, \|h\|_\infty, \|q\|_\infty\} \leq M$  for some  $M > 0$ . In addition, we assume that  $\sup_{w,a,x} |\hat{h}(w, a, x) - h(w, a, x)| = o_p(n^{-\alpha})$ ,  $\sup_{z,a,x} |\hat{q}(z, a, x) - q(z, a, x)| = o_p(n^{-\beta})$  for some  $\alpha, \beta > 0$  (Chen and Christensen, 2013). Given the training dataset, we define an oracle estimator of  $\delta(x; \hat{d}_z, \hat{d}_w)$

$$\delta'(x; \hat{d}_z, \hat{d}_w) \triangleq \frac{\sum_{i=1}^n \{h(W_i, \hat{d}_z(x, Z_i), x) - Y_i q(Z_i, A_i, x)\mathbb{I}\{\hat{d}_w(x, W_i) = A_i\}\} K(\frac{\|x - X_i\|}{\gamma})}{\sum_{i=1}^n K(\frac{\|x - X_i\|}{\gamma})}.$$

We assume that with probability larger than  $1 - 1/n$ , for any  $d_z \in \mathcal{D}_{\mathcal{Z}}$  and  $d_w \in \mathcal{D}_{\mathcal{W}}$ ,  $\sup_x |\delta(x; d_z, d_w) - \delta'(x; d_z, d_w)| \leq C_1 n^{-\gamma}$  for some  $C_1 > 0$  and  $\gamma > 0$  under certain conditions (Jiang, 2017). If we further impose a restriction on the cardinality of preliminary policy classes and assume  $|\mathcal{D}_{\mathcal{Z}}| = o(n)$  and  $|\mathcal{D}_{\mathcal{W}}| = o(n)$ , by a straightforward calculation, we have  $\sup_{x, d_z \in \mathcal{D}_{\mathcal{Z}}, d_w \in \mathcal{D}_{\mathcal{W}}} |\hat{\delta}(x; d_z, d_w) - \delta(x; d_z, d_w)| \leq C_2 n^{-\zeta}$  on a set  $\mathcal{X}_0$  and  $\mathbb{P}(\mathcal{X}_0^c) \rightarrow 0$ , where  $C_2 > 0$ ,  $\zeta = \min\{\alpha, \beta, \gamma\}$ , and  $\mathcal{X}_0^c$  is the complement of  $\mathcal{X}_0$ .

To streamline the presentation, in the following, we abbreviate  $\delta(X; \hat{d}_z, \hat{d}_w)$ ,  $\delta'(X; \hat{d}_z, \hat{d}_w)$  and  $\hat{\delta}(X; \hat{d}_z, \hat{d}_w)$  as  $\delta(X)$ ,  $\delta'(X)$  and  $\hat{\delta}(X)$ , respectively. Two subsets of  $\mathcal{X}$ , namely  $\mathcal{X}_{f1}$  and  $\mathcal{X}_{f2}$ , are defined as

$$\begin{aligned}\mathcal{X}_{f1} &= \{x \in \mathcal{X} : \mathbb{I}\{\hat{\delta}(x) \geq 0\} = 1, \mathbb{I}\{\delta(x) \geq 0\} = 0\}, \\ \mathcal{X}_{f2} &= \{x \in \mathcal{X} : \mathbb{I}\{\hat{\delta}(x) \geq 0\} = 0, \mathbb{I}\{\delta(x) \geq 0\} = 1\},\end{aligned}$$

and we also define the complement set  $\mathcal{X}_c$  as

$$\mathcal{X}_c = \{x \in \mathcal{X} : \text{sign}(\hat{\delta}(x)) = \text{sign}(\delta(x))\},$$

with  $\text{sign}(0) = 1$ . We see that  $\mathcal{X}_{f1} \cap \mathcal{X}_{f2} = \emptyset$ ,  $\mathcal{X}_{f1} \cap \mathcal{X}_c = \emptyset$ ,  $\mathcal{X}_{f2} \cap \mathcal{X}_c = \emptyset$ , and  $\mathcal{X}_{f1} \cup \mathcal{X}_{f2} \cup \mathcal{X}_c =$

$\mathcal{X}$ . From the definition of  $\mathbb{K}(\hat{\pi})$ , we have

$$\begin{aligned}
\mathbb{K}(\hat{\pi}) &= \int_{x \in \mathcal{X}} \mathbb{E}[\bar{\pi}(X; \hat{d}_z, \hat{d}_w) h(W, \hat{d}_z(X, Z), X) + (1 - \bar{\pi}(X; \hat{d}_z, \hat{d}_w)) Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\}] f(x) dx \\
&- \hat{\pi}(X; \hat{d}_z, \hat{d}_w) h(W, \hat{d}_z(X, Z), X) + (1 - \hat{\pi}(X; \hat{d}_z, \hat{d}_w)) Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\} |X = x] f(x) dx \\
&= \int_{x \in \mathcal{X}_{f1}} -\delta(x) f(x) dx + \int_{x \in \mathcal{X}_{f2}} \delta(x) f(x) dx + \int_{x \in \mathcal{X}_c} 0 f(x) dx \\
&= - \int_{x \in \mathcal{X}_{f1}} \delta(x) f(x) dx + \int_{x \in \mathcal{X}_{f2}} \delta(x) f(x) dx.
\end{aligned}$$

The second equation holds because if  $x \in \mathcal{X}_{f1}$ ,  $\hat{\pi}(x; \hat{d}_z, \hat{d}_w) = 1$ ,  $\bar{\pi}(x; \hat{d}_z, \hat{d}_w) = 0$  and

$$\begin{aligned}
&\mathbb{E}[\bar{\pi}(X; \hat{d}_z, \hat{d}_w) h(W, \hat{d}_z(X, Z), X) + (1 - \bar{\pi}(X; \hat{d}_z, \hat{d}_w)) Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\}] \\
&- \hat{\pi}(X; \hat{d}_z, \hat{d}_w) h(W, \hat{d}_z(X, Z), X) + (1 - \hat{\pi}(X; \hat{d}_z, \hat{d}_w)) Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\} |X = x] = -\delta(x); \\
&\text{if } x \in \mathcal{X}_{f2}, \hat{\pi}(x; \hat{d}_z, \hat{d}_w) = 0, \bar{\pi}(x; \hat{d}_z, \hat{d}_w) = 1 \text{ and} \\
&\mathbb{E}[\bar{\pi}(X; \hat{d}_z, \hat{d}_w) h(W, \hat{d}_z(X, Z), X) + (1 - \bar{\pi}(X; \hat{d}_z, \hat{d}_w)) Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\}] \\
&- \hat{\pi}(X; \hat{d}_z, \hat{d}_w) h(W, \hat{d}_z(X, Z), X) + (1 - \hat{\pi}(X; \hat{d}_z, \hat{d}_w)) Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\} |X = x] = \delta(x); \\
&\text{if } x \in \mathcal{X}_c, \hat{\pi}(x; \hat{d}_z, \hat{d}_w) = \bar{\pi}(x; \hat{d}_z, \hat{d}_w) \text{ and}
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}[\bar{\pi}(X; \hat{d}_z, \hat{d}_w) h(W, \hat{d}_z(X, Z), X) + (1 - \bar{\pi}(X; \hat{d}_z, \hat{d}_w)) Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\}] \\
&- \hat{\pi}(X; \hat{d}_z, \hat{d}_w) h(W, \hat{d}_z(X, Z), X) + (1 - \hat{\pi}(X; \hat{d}_z, \hat{d}_w)) Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\} |X = x] = 0.
\end{aligned}$$

Therefore, we essentially need to bound  $-\int_{x \in \mathcal{X}_{f1}} \delta(x) f(x) dx$  and  $\int_{x \in \mathcal{X}_{f2}} \delta(x) f(x) dx$  follows a similar proof. In this regard, we further split  $\mathcal{X}_{f1}$  to  $\mathcal{X}_{f1,1} = \{x \in \mathcal{X}_{f1} : \delta(x) \in (-Cn^{-\zeta}, Cn^{-\zeta})\}$  and  $\mathcal{X}_{f1,2} = \{x \in \mathcal{X}_{f1} : \delta(x) \notin (-Cn^{-\zeta}, Cn^{-\zeta})\}$ . Then it is easy to see that  $-\int_{x \in \mathcal{X}_{f1,1}} \delta(x) f(x) dx$  is bounded by  $O(n^{-\zeta})$  and  $\mathbb{P}(\mathcal{X}_{f1,2})$  converges to 0 as  $\mathbb{P}(\mathcal{X}_0^c)$  converges to 0. We then conclude that  $\mathbb{K}(\hat{\pi}) = o(1)$  almost surely.

## H Proof of Proposition 3

We start with defining two subsets of  $\mathcal{X}$ ,

$$\begin{aligned}\mathcal{X}_{g1} &= \{x \in \mathcal{X} : \bar{\pi}(x, \hat{d}_z, \hat{d}_w) = 1, \bar{\pi}(x, d_z^*, d_w^*) = 0\}, \\ \mathcal{X}_{g2} &= \{x \in \mathcal{X} : \bar{\pi}(x, \hat{d}_z, \hat{d}_w) = 0, \bar{\pi}(x, d_z^*, d_w^*) = 1\},\end{aligned}$$

and we also define the complement set  $\mathcal{X}_{gc}$  as

$$\mathcal{X}_{gc} = \{x \in \mathcal{X} : \bar{\pi}(x, \hat{d}_z, \hat{d}_w) = \bar{\pi}(x, d_z^*, d_w^*)\},$$

which can also be split into

$$\begin{aligned}\mathcal{X}_{gc1} &= \{x \in \mathcal{X} : \bar{\pi}(x, \hat{d}_z, \hat{d}_w) = \bar{\pi}(x, d_z^*, d_w^*) = 0\}, \\ \mathcal{X}_{gc2} &= \{x \in \mathcal{X} : \bar{\pi}(x, \hat{d}_z, \hat{d}_w) = \bar{\pi}(x, d_z^*, d_w^*) = 1\}.\end{aligned}$$

We see that  $\mathcal{X}_{g1} \cap \mathcal{X}_{g2} = \emptyset$ ,  $\mathcal{X}_{gc1} \cap \mathcal{X}_{gc2} = \emptyset$ ,  $\mathcal{X}_{gc1} \cup \mathcal{X}_{gc2} = \mathcal{X}_{gc}$ ,  $\mathcal{X}_{g1} \cap \mathcal{X}_{gc} = \emptyset$ ,  $\mathcal{X}_{g2} \cap \mathcal{X}_{gc} = \emptyset$ , and  $\mathcal{X}_{g1} \cup \mathcal{X}_{g2} \cup \mathcal{X}_{gc} = \mathcal{X}$ .

From the definition of  $V(d_{zw}^{\bar{\pi}*})$  and  $V(\hat{d}_{zw}^{\bar{\pi}})$ , we have

$$\begin{aligned}& V(d_{zw}^{\bar{\pi}*}) - V(\hat{d}_{zw}^{\bar{\pi}}) \\ &= \int_{x \in \mathcal{X}} \mathbb{E}[\bar{\pi}(X; d_z^*, d_w^*)h(W, d_z^*(X, Z), X) + (1 - \bar{\pi}(X; d_z^*, d_w^*))Yq(Z, A, X)\mathbb{I}\{d_w^*(X, W) = A\} \\ &\quad - \bar{\pi}(X; \hat{d}_z, \hat{d}_w)h(W, \hat{d}_z(X, Z), X) - (1 - \bar{\pi}(X; \hat{d}_z, \hat{d}_w))Yq(Z, A, X)\mathbb{I}\{\hat{d}_w(X, W) = A\}|X = x]f(x)dx \\ &= \int_{x \in \mathcal{X}_{g1}} E[Yq(Z, A, X)\mathbb{I}\{d_w^*(X, W) = A\} - h(W, \hat{d}_z(X, Z), X)|X = x]f(x)dx \\ &\quad + \int_{x \in \mathcal{X}_{g2}} E[h(W, d_z^*(X, Z), X) - Yq(Z, A, X)\mathbb{I}\{\hat{d}_w(X, W) = A\}|X = x]f(x)dx \\ &\quad + \int_{x \in \mathcal{X}_{gc1}} E[Yq(Z, A, X)\mathbb{I}\{d_w^*(X, W) = A\} - Yq(Z, A, X)\mathbb{I}\{\hat{d}_w(X, W) = A\}|X = x]f(x)dx \\ &\quad + \int_{x \in \mathcal{X}_{gc2}} E[h(W, d_z^*(X, Z), X) - h(W, \hat{d}_z(X, Z), X)|X = x]f(x)dx.\end{aligned}$$

Then it is easy to see that

$$\int_{x \in \mathcal{X}_{g1}} E[Yq(Z, A, X)\mathbb{I}\{d_w^*(X, W) = A\} - Yq(Z, A, X)\mathbb{I}\{\hat{d}_w(X, W) = A\}|X = x]f(x)dx$$

and

$$\int_{x \in \mathcal{X}_{gc2}} E[h(W, d_z^*(X, Z), X) - h(W, \hat{d}_z(X, Z), X) | X = x] f(x) dx$$

converge to 0 in probability according to Assumption 6.

Therefore, we essentially need to bound

$$\int_{x \in \mathcal{X}_{g1}} E[Y q(Z, A, X) \mathbb{I}\{d_w^*(X, W) = A\} - h(W, \hat{d}_w(X, Z), X) | X = x] f(x) dx$$

and

$$\int_{x \in \mathcal{X}_{g2}} E[h(W, d_z^*(X, Z), X) - Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\} | X = x] f(x) dx.$$

We further split  $\mathcal{X}_{g1}$  to

$$\mathcal{X}_{g1,1} = \{x \in \mathcal{X}_{g1} : E[Y q(Z, A, X) \mathbb{I}\{d_w^*(X, W) = A\} - h(W, \hat{d}_z(X, Z), X) | X = x] \in (-Cn^{-\eta}, Cn^{-\eta})\}$$

and

$$\mathcal{X}_{g1,2} = \{x \in \mathcal{X}_{g1} : E[Y q(Z, A, X) \mathbb{I}\{d_w^*(X, W) = A\} - h(W, \hat{d}_z(X, Z), X) | X = x] \notin (-Cn^{-\eta}, Cn^{-\eta})\}$$

where  $\eta = \min\{\xi, \varphi\}$ . Then it is easy to see that

$$\int_{x \in \mathcal{X}_{g1,1}} E[Y q(Z, A, X) \mathbb{I}\{d_w^*(X, W) = A\} - h(W, \hat{d}_z(X, Z), X) | X = x] f(x) dx$$

is bounded by  $O(n^{-\eta})$  and  $\mathbb{P}(\mathcal{X}_{g1,2})$  converges to 0 in probability based on Assumption 6 and the definition of  $\mathcal{X}_{g1}$ . A similar proof can also be conducted to obtain  $\int_{x \in \mathcal{X}_{g2}} E[h(W, d_z^*(X, Z), X) - Y q(Z, A, X) \mathbb{I}\{\hat{d}_w(X, W) = A\} | X = x] f(x) dx$  is small enough. We then have that  $V(\hat{d}_{zw}^{\bar{\pi}}) \xrightarrow{p} V(d_{zw}^{\bar{\pi}})$ .

As we have proved that  $\mathbb{K}(\hat{\pi}) = V(\hat{d}_{zw}^{\bar{\pi}}) - V(d_{zw}^{\bar{\pi}}) = o(1)$  almost surely in Appendix G, we finally conclude that  $V(\hat{d}_{zw}^{\bar{\pi}}) \xrightarrow{p} V(d_{zw}^{\bar{\pi}})$ .

## I Data generating mechanism and parameter setup in Section 4

The data generating mechanism for  $(X, A, Z, W, U)$  is summarized in Table 1, and the setups of varying parameters in each scenario are summarized in Table 2.

Variables	Generating Mechanism	Fixed Parameter Setting
$X \in \mathbb{R}^2$	$X \sim \mathcal{N}(\Gamma_x, \Sigma_x)$	$\Gamma_x = (0.25, 0.25)^T, \Sigma_x = \begin{pmatrix} 0.25^2 & 0 \\ 0 & 0.25^2 \end{pmatrix}$
$A \in \{1, -1\}$	$(\frac{A+1}{2})   X \sim \text{Bern}(p_a)$	$p_a = \frac{1}{1+\exp\{(0.125, 0.125)^T X\}}$
$Z \in \mathbb{R}$	$Z = \alpha_0 + \alpha_a \frac{1+A}{2} + \alpha_x X$	$\alpha_0 = \alpha_a = \mu_0 = \kappa_0 = \kappa_a = \sigma_{zw} = 0.25,$
$W \in \mathbb{R}$	$(Z, W, U)   A, X \sim \mathcal{N}\left(\begin{pmatrix} \alpha_0 + \mu_a \frac{1+A}{2} + \mu_x X \\ \kappa_0 + \kappa_a \frac{1+A}{2} + \kappa_x X \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_z^2 & \sigma_{zw} & \sigma_{zu} \\ \sigma_{zw} & \sigma_w^2 & \sigma_{wu} \\ \sigma_{zu} & \sigma_{wu} & \sigma_u^2 \end{pmatrix}\right)$	$\mu_a = 0.125, \alpha_x = \mu_x = \kappa_x = (0.25, 0.25)^T,$
$U \in \mathbb{R}$	$U = \sigma_{zu} = \sigma_{wu} = 0.5, \sigma_z = \sigma_w = \sigma_u = 1$	$\sigma_{zu} = \sigma_{wu} = \sigma_z = \sigma_w = \sigma_u = 1$
$Y \in \mathbb{R}$	$Y \sim \mathcal{N}\{\mathbb{E}(Y W, U, A, Z, X), \sigma_y^2\}$	$\sigma_y = 0.25, b_0 = 2, \omega = 2$

\* As for generation of  $Y$ ,  $\mathbb{E}(Y|X, A, Z, W, U) = b_0 + b_1(X) \frac{1+A}{2} + b_2(X)X + (b_w + b_a \frac{1+A}{2} + b_3(X)A - \omega)\mathbb{E}(W|U, X) + \omega W$ , where  $\mathbb{E}(W|U, X) = \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X)$ .

Table 1: Data generating mechanism and setup for fixed parameters across scenarios.

Scenario	Parameter Setup				
Number	$b_1(X)$	$b_2(X)$	$b_3(X)$	$b_a$	$b_w$
1	$0.5 + 3X_{(1)} - 5X_{(2)}$	$(0.25, 0.25)^T$	0	0.25	8
2	$0.5 + 3X_{(1)} - 5X_{(2)}$	$(0.25, 0.25)^T$	0	0	8
3	$2.3 +  X_{(1)} - 1  -  X_{(2)} + 1 $	$X^T$	$\sin(X_{(1)}) - 2\cos(X_{(2)})$	-2.5	4
4	$0.25 - 6X_{(1)}X_{(2)}$	$X^T$	0	0	5
5	$0.1 - 2X_{(1)}^2$	$X^T$	$4X_{(2)}^2$	0.8	8
6	$-0.5 + \exp(X_{(1)}) - 3X_{(2)}$	$(0.25, 0.25)^T$	0	0	8

\*  $X_{(1)}, X_{(2)}$  denote the first and second dimensions of  $X$ .

\* The parameter settings in scenarios 1-4 are considered by [Qi et al. \(2023\)](#).

Table 2: The varying parameters for each scenario.

## J Derivation of optimal ITRs considered in Section 4

From

$$(Z, W, U)|A, X \sim \mathcal{N} \left\{ \begin{pmatrix} \alpha_0 + \alpha_a \frac{1+A}{2} + \alpha_x X \\ \mu_0 + \mu_a \frac{1+A}{2} + \mu_x X \\ \kappa_0 + \kappa_a \frac{1+A}{2} + \kappa_x X \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_z^2 & \sigma_{zw} & \sigma_{zu} \\ \sigma_{zw} & \sigma_w^2 & \sigma_{wu} \\ \sigma_{zu} & \sigma_{wu} & \sigma_u^2 \end{pmatrix} \right\},$$

and

$$(Z, W, U)|X \sim \mathcal{N} \left\{ \begin{pmatrix} \alpha_0 + \alpha_a \mathbb{P}(A = 1|X) + \alpha_x X \\ \mu_0 + \mu_a \mathbb{P}(A = 1|X) + \mu_x X \\ \kappa_0 + \kappa_a \mathbb{P}(A = 1|X) + \kappa_x X \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_z^2 & \sigma_{zw} & \sigma_{zu} \\ \sigma_{zw} & \sigma_w^2 & \sigma_{wu} \\ \sigma_{zu} & \sigma_{wu} & \sigma_u^2 \end{pmatrix} \right\},$$

the following results hold,

$$\mathbb{E}[W|X, A, U] = \mu_0 + \mu_a \frac{1+A}{2} + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2} (U - \kappa_0 - \kappa_a \frac{1+A}{2} - \kappa_x X), \quad (16)$$

$$\mathbb{E}[U|X, Z] = \kappa_0 + \kappa_a \mathbb{P}(A = 1|X) + \kappa_x X + \frac{\sigma_{zu}}{\sigma_z^2} (Z - \alpha_0 - \alpha_a \mathbb{P}(A = 1|X) - \alpha_x X), \quad (17)$$

$$\mathbb{E}[U|X, W] = \kappa_0 + \kappa_a \mathbb{P}(A = 1|X) + \kappa_x X + \frac{\sigma_{wu}}{\sigma_w^2} (W - \mu_0 - \mu_a \mathbb{P}(A = 1|X) - \mu_x X). \quad (18)$$

Recall that

$$\begin{aligned}\mathbb{E}(Y|X, A, Z, W, U) &= b_0 + b_1(X)\frac{1+A}{2} + b_2(X)X + \left( b_w + b_a\frac{1+A}{2} + b_3(X)A - \omega \right) \\ &\quad \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right) + \omega W,\end{aligned}$$

then we can find that

$$\begin{aligned}\mathbb{E}(Y|X, A, Z, U) &= b_0 + b_1(X)\frac{1+A}{2} + b_2(X)X + \left( b_w + b_a\frac{1+A}{2} + b_3(X)A - \omega \right) \\ &\quad \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right) + \omega \mathbb{E}[W|X, A, Z, U], \\ &= b_0 + b_1(X)\frac{1+A}{2} + b_2(X)X + \left( b_w + b_a\frac{1+A}{2} + b_3(X)A - \omega \right) \\ &\quad \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right) + \omega \mathbb{E}[W|X, A, U], \\ &= b_0 + b_1(X)\frac{1+A}{2} + b_2(X)X + \left( b_w + b_a\frac{1+A}{2} + b_3(X)A - \omega \right) \\ &\quad \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right) + \omega \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right), \\ &= b_0 + b_1(X)\frac{1+A}{2} + b_2(X)X + \left( b_w + b_a\frac{1+A}{2} + b_3(X)A \right) \\ &\quad \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right), \tag{19}\end{aligned}$$

where the first equality is due to Assumption 1, and the second equality is due to (16), and

$$\begin{aligned}\mathbb{E}(Y|X, A, W, U) &= \mathbb{E}(Y|X, A, Z, W, U) \\ &= b_0 + b_1(X)\frac{1+A}{2} + b_2(X)X + \left( b_w + b_a\frac{1+A}{2} + b_3(X)A - \omega \right) \\ &\quad \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right) + \omega W, \tag{20}\end{aligned}$$

where the first equality is due to Assumption 1. Furthermore, note that

$$\begin{aligned}\mathbb{E}[h(W, 1, X)|X, Z, U] &= \mathbb{E}[h(W, 1, X)|X, U] \\ &= \mathbb{E}[Y|X, A = 1, U] \\ &= \mathbb{E}[Y|X, A = 1, Z, U] \\ &= b_0 + b_1(X) + b_2(X)X + (b_w + b_a + b_3(X)) \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right),\end{aligned}$$

where the first and third equality is due to Assumption 1, the second equality follows from Theorem 1 of Miao et al. (2018a) under Assumptions 2 and 3, and the last equality is by (19). Similarly,

$$\begin{aligned}\mathbb{E}[h(W, -1, X)|X, Z, U] &= \mathbb{E}[Y|X, A = -1, Z, U] \\ &= b_0 + b_2(X)X + (b_w - b_3(X)) \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right).\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{E}[Yq(Z, 1, X)\mathbb{I}\{A = 1\}|X, W, U] &= \mathbb{P}(A = 1|X, W, U)\mathbb{E}[Yq(Z, 1, X)|X, A = 1, W, U] \\ &= \mathbb{P}(A = 1|X, U)\mathbb{E}[q(Z, 1, X)|X, A = 1, U]\mathbb{E}[Y|X, A = 1, W, U] \\ &= \mathbb{E}[Y|X, A = 1, W, U] \\ &= b_0 + b_1(X) + b_2(X)X + (b_w + b_a + b_3(X) - \omega) \\ &\quad \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right) + \omega W,\end{aligned}$$

where the second equality is due to Assumption 1, and the third equality is due to Theorem 2.2 of Cui et al. (2023) under Assumptions 4 and 5, and the last equality is due to (20). Similarly,

$$\begin{aligned}\mathbb{E}[Yq(Z, -1, X)\mathbb{I}\{A = -1\}|X, W, U] &= \mathbb{E}[Y|X, A = -1, W, U] \\ &= b_0 + b_2(X)X + (b_w - b_3(X) - \omega) \\ &\quad \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right) + \omega W.\end{aligned}$$

Then we can find that

$$\begin{aligned}\mathbb{E}[h(W, 1, X) - h(W, -1, X)|X, Z, U] &= b_1(X) + (b_a + 2b_3(X)) \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right), \\ \mathbb{E}[Yq(Z, 1, X)\mathbb{I}\{A = 1\} - Yq(Z, -1, X)\mathbb{I}\{A = -1\}|X, W, U] \\ &= b_1(X) + (b_a + 2b_3(X)) \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2}(U - \kappa_0 - \kappa_x X) \right).\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
\mathbb{E}[h(W, 1, X) - h(W, -1, X)|X, Z] &= \mathbb{E}[\mathbb{E}[h(W, 1, X) - h(W, -1, X)|X, Z, U]] \\
&= b_1(X) + (b_a + 2b_3(X)) \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2} (\mathbb{E}[U|X, Z] - \kappa_0 - \kappa_x X) \right),
\end{aligned} \tag{21}$$

$$\begin{aligned}
&\mathbb{E}[Yq(Z, 1, X)\mathbb{I}\{A = 1\} - Yq(Z, -1, X)\mathbb{I}\{A = -1\}|X, W] \\
&= \mathbb{E}[\mathbb{E}[Yq(Z, 1, X)\mathbb{I}\{A = 1\} - Yq(Z, -1, X)\mathbb{I}\{A = -1\}|X, W, U]] \\
&= b_1(X) + (b_a + 2b_3(X)) \left( \mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2} (\mathbb{E}[U|X, W] - \kappa_0 - \kappa_x X) \right).
\end{aligned} \tag{22}$$

Therefore, plug (17) and (18) into (21) and (22) respectively, we can find that

$$\begin{aligned}
\mathbb{E}[h(W, 1, X) - h(W, -1, X)|X, Z] &= b_1(X) + (b_a + 2b_3(X))(\mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2} (\kappa_0 + \kappa_a \mathbb{P}(A = 1|X) \\
&\quad + \kappa_x X + \frac{\sigma_{zu}}{\sigma_z^2} (Z - \alpha_0 - \alpha_a \mathbb{P}(A = 1|X) - \alpha_x X) - \kappa_0 - \kappa_x X)),
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}[Yq(Z, 1, X)\mathbb{I}\{A = 1\} - Yq(Z, -1, X)\mathbb{I}\{A = -1\}|X, W] \\
&= b_1(X) + (b_a + 2b_3(X))(\mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2} (\kappa_0 + \kappa_a \mathbb{P}(A = 1|X) \\
&\quad + \kappa_x X + \frac{\sigma_{wu}}{\sigma_w^2} (W - \mu_0 - \mu_a \mathbb{P}(A = 1|X) - \mu_x X) - \kappa_0 - \kappa_x X)).
\end{aligned}$$

Hence,

$$\begin{aligned}
d_z^*(X, Z) &= \text{sign}\{\mathbb{E}[h(W, 1, X) - h(W, -1, X)|X, Z]\} \\
&= \text{sign}\{b_1(X) + (b_a + 2b_3(X))(\mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2} (\kappa_0 + \kappa_a \mathbb{P}(A = 1|X) \\
&\quad + \kappa_x X + \frac{\sigma_{zu}}{\sigma_z^2} (Z - \alpha_0 - \alpha_a \mathbb{P}(A = 1|X) - \alpha_x X) - \kappa_0 - \kappa_x X))\},
\end{aligned}$$

$$\begin{aligned}
d_w^*(X, W) &= \text{sign}\{\mathbb{E}[Yq(Z, 1, X)\mathbb{I}\{A = 1\} - Yq(Z, -1, X)\mathbb{I}\{A = -1\}|X, W]\} \\
&= \text{sign}\{b_1(X) + (b_a + 2b_3(X))(\mu_0 + \mu_x X + \frac{\sigma_{wu}}{\sigma_u^2} (\kappa_0 + \kappa_a \mathbb{P}(A = 1|X) \\
&\quad + \kappa_x X + \frac{\sigma_{wu}}{\sigma_w^2} (W - \mu_0 - \mu_a \mathbb{P}(A = 1|X) - \mu_x X) - \kappa_0 - \kappa_x X))\}.
\end{aligned}$$

## K Implementation details of numerical experiments

**Step (i)** The method we adopt is neural maximum moment restriction (NMMR), which employs multilayer perceptron (MLP) to estimate the confounding bridges ([Kompa et al., 2022](#)). The target loss functions are set as

$$R(h) = \mathbb{E}[(Y - h(W, A, X))(Y' - h(W', A', X'))K_z((Z, A, X), (Z', A', X'))],$$

$$R(q, a) = \mathbb{E}[(1 - \mathbb{I}\{A = a\}q(Z, a, X))(1 - \mathbb{I}\{A' = a\}q(Z', a, X'))K_w((W, X), (W', X'))], \text{ for } a \in \mathcal{A},$$

where  $(Z', W', A', X', Y')$  are independent copies of  $(Z, W, A, X, Y)$ , and  $K_z : (\mathcal{Z} \times \mathcal{A} \times \mathcal{X})^2 \rightarrow \mathbb{R}$ ,  $K_w : (\mathcal{W} \times \mathcal{X})^2 \rightarrow \mathbb{R}$  denote continuous, bounded, and integrally strictly positive definite (ISPD) kernels. In practice, we use the empirical risk instead, i.e.,

$$\hat{R}(h) = \frac{1}{n(n-1)} \sum_{i,j=1, i \neq j}^n (y_i - h_i)(y_j - h_j)k_{z,ij}, \quad (23)$$

$$\hat{R}(q, a) = \frac{1}{n(n-1)} \sum_{i,j=1, i \neq j}^n (1 - \mathbb{I}\{a_i = a\}q_i)(1 - \mathbb{I}\{a_j = a\}q_j)k_{w,ij}, \text{ for } a \in \mathcal{A}, \quad (24)$$

where  $h_i = h(w_i, a_i, x_i)$ ,  $q_i = (z_i, a_i, x_i)$ ,  $k_{z,ij} = K_z((z_i, a_i, x_i)(z_j, a_j, x_j))$  and  $k_{w,ij} = K_w((w_i, x_i), (w_j, x_j))$ . In addition, we add a penalty term with respect to network weights to avoid overfitting.

As for the hyperparameters tuning procedure, we consider employing multilayer perceptrons with 2-8 fully connected layers with a variable number of hidden units. We then perform a grid search over the following parameters: learning rate, penalty coefficient, number of epochs, batch size, depth of the network, and width of the network. For every permutation of these parameters, we train a network based on the determined architecture and parameter values. Subsequently, we compute the empirical risk. Our aim is to pinpoint the parameter combination that yields the lowest empirical risk. These identified optimal parameters are then utilized to construct a refined neural network, which, in turn, serves as the foundation for conducting estimations. The parameter setup is summarized in Table 3. For detailed insights into the specific hyperparameter choices and architectural dimensions, we refer to supplementary Section B in [Kompa et al. \(2022\)](#).

Parameter	Value
Number of epoch	150
Batch size	250
Learning rate	0.003
Penalty coefficient	0.001, 0.01, 0.1
Depth of network	4 (for estimating $h$ ) 8 (for estimating $q$ )
Width of network	80

Table 3: Parameter setup for step (i)

**Step (ii)** For the estimation of preliminary ITRs, we follow the main text to solve the proposed optimization problems. For instance, to estimate  $d_z^*$ , we solve the following optimization problem:

$$\hat{g}_z \in \arg \min_{g_z \in \mathcal{G}_z} \mathbb{P}_n[\{\hat{h}(W, 1, X) - \hat{h}(W, -1, X)\}\phi(g_z(X, Z))] + \rho_z \|g_z\|_{\mathcal{G}_z}^2.$$

Here,  $g_z$  represents a measurable decision function in  $\mathcal{G}_z : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}$  used to indicate  $d_z$  (e.g.,  $d_z(X, Z) = \text{sign}(g_z(X, Z))$ ),  $\phi$  denotes the hinge loss function  $\phi(x) = \max\{1 - x, 0\}$ , and  $\rho_z > 0$  is a tuning parameter. As for the tuning procedure regarding  $\rho_z$ , when  $g_z$  is treated as a linear rule, for each predefined  $\rho_z$ , the data is divided into  $K$  folds. For each  $k \in [K]$ , we compute  $\hat{h}^{(-k)}$  and  $\hat{g}_z^{(-k)}$ , and then calculate the empirical value using the validation data. By averaging the empirical values across  $K$  folds for each value of  $\rho_z$ , we identify the parameter that maximizes the average empirical value. The finalized parameter is then employed to determine  $\hat{g}_z$ . Such a procedure can be extended. For example, when considering  $g_z$  as a RKHS, it is advisable to apply the cross-fitting procedure separately for each combination of pre-defined  $\rho_z$  and bandwidth, with details presented in [Qi et al. \(2023\)](#). And the estimation of  $\hat{d}_w$  can be approached in a similar manner.

For more estimators regarding  $d_z^*$  and  $d_w^*$ , we refer to [Bennett and Kallus \(2023\)](#); [Sverdrup and Cui \(2023\)](#); [Wang et al. \(2022\)](#). One could further expand the estimation pipeline

utilized in unconfounded scenarios and leverage state-of-the-art machine learning techniques (Chen et al., 2020; Raghu et al., 2017; Yoon et al., 2018) to tackle the weighted classification problems and construct estimates.

**Step (iii)** The estimation of  $\bar{\pi}$  follows the procedure given in the main text. As for the selection of bandwidth in the Nadaraya-Watson kernel regression estimator, we employ Scott's rule of thumb (Scott, 2015) and set  $\gamma = 1.06\hat{\sigma}n^{-1/5}$ , where  $\hat{\sigma}$  is the estimated standard deviation of  $X$ . For more methods regarding estimation of  $\delta(\cdot)$ , we refer to Chen (2017); Dalmasso et al. (2020); Dinh et al. (2016); Sohn et al. (2015).

For the convenience of readers to reproduce the results, the pseudo-code of the whole pipeline is presented in Algorithm 1. The code of implementation can also be accessed on GitHub<sup>1</sup>.

## L Additional results of numerical experiments

The experimental results with sample size  $n = 500$  are presented in Figure 4. The experimental results with sample size  $n = 500$  and an altered behavior policy (treatment is randomly assigned in this case) are presented in Figure 5.

## M Additional results of real data application

Regarding the quantitative analysis, Table 4 describes the estimated value functions of our proposed ITR, alongside existing approaches, under four settings with increasing numbers of proxies. For Setting 1,  $Z = (pafi1, paco21)$ ,  $W = (ph1, hema1)$ . For Setting 2,  $Z = (pafi1, paco21, pot1)$ ,  $W = (ph1, hema1, bili1)$ . For Setting 3,  $Z = (pafi1, paco21, pot1, wt0)$ ,  $W = (ph1, hema1, bili1, sod1)$ . For Setting 4,  $Z = (pafi1, paco21, pot1, wt0, crea1)$ ,  $W = (ph1, hema1, bili1, sod1, alb1)$ .

As for the qualitative analysis, we present an illustrative example below. Regarding the

---

<sup>1</sup><https://github.com/taoshen2022/Optimal-Treatment-Regimes-for-Proximal-Causal-Learning>

---

**Algorithm 1:** Estimation of optimal ITR  $d_{zw}^{\hat{\pi}*}$ 

---

1 **Input:** Training data

2 Construct MLP models to estimate  $h(w, a, x)$  and  $q(z, a, x)$ :

3 **Repeat** for different penalty coefficients:

4   **for** each epoch **do**

5     **for** each batch **do**

6       Compute loss function (23) and (24) based on the batch

7       Update the internal model parameter

8     **end**

9 **end**

10 **Finalize** the penalty coefficient which minimizes the empirical loss, and obtain  
 $\hat{h}(w, a, x)$  and  $\hat{q}(z, a, x)$

11 **Repeat** for different  $\rho_z$  and  $\rho_w$ :

12   **for** each batch **do**

13     Find  $\hat{g}_{z,b}, \hat{g}_{w,b}$  by (7) and (8) based on the  $b$ -th batch, estimated bridge  
functions, and specified  $\rho_z$  and  $\rho_w$ , and then obtain  $\hat{d}_{z,b}, \hat{d}_{w,b}$  based on  $\hat{g}_{z,b}, \hat{g}_{w,b}$

14     Compute empirical value of  $\hat{d}_{z,b}, \hat{d}_{w,b}$  respectively using the data not covered  
in the batch

15 **end**

16 **Finalize**  $\rho_z$  and  $\rho_w$  based on empirical values and then obtain  $\hat{d}_z, \hat{d}_w$

17 Select bandwidth by Scott's rule of thumb

18 Find  $\hat{\delta}(X; \hat{d}_z, \hat{d}_w)$  and then obtain  $\hat{\pi}(X; \hat{d}_z, \hat{d}_w)$

19 **Output:**  $\hat{d}_{zw}^{\hat{\pi}}$  constructed by (9)

---

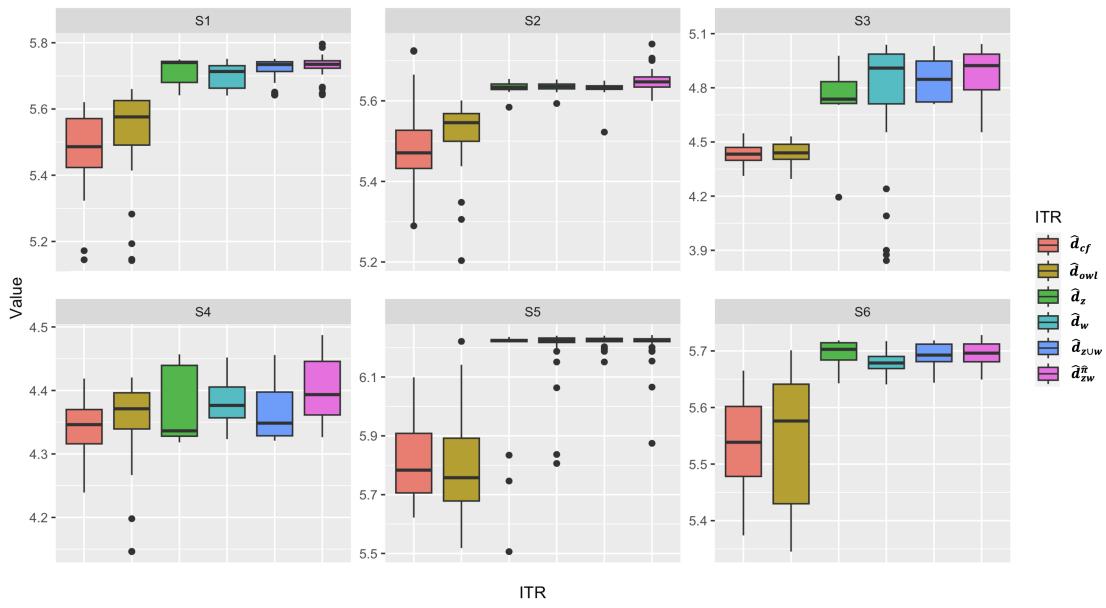


Figure 4: Boxplots of the empirical value functions with  $n = 500$ .

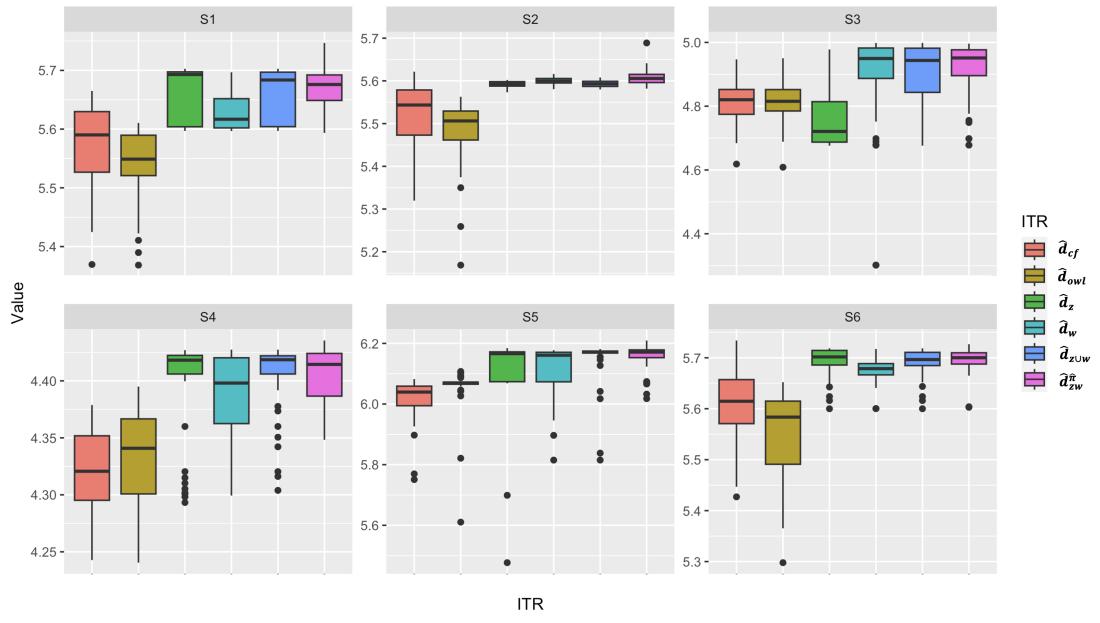


Figure 5: Boxplots of the empirical value functions with  $n = 500$  and an altered behavior policy.

	$\hat{V}(\hat{d}_{cf})$	$\hat{V}(\hat{d}_{owl})$	$\hat{V}(\hat{d}_z)$	$\hat{V}(\hat{d}_w)$	$\hat{V}(\hat{d}_{z \cup w})$	$\hat{V}(\hat{d}_{zw}^{\hat{\pi}})$
Setting 1	24.84 (3.06)	24.97 (2.93)	25.12 (4.69)	26.61 (3.34)	27.86 (2.28)	28.21 (3.28)
Setting 2	24.81 (3.11)	24.97 (2.94)	25.60 (3.73)	25.74 (3.57)	26.32 (2.29)	27.02 (2.95)
Setting 3	24.79 (3.02)	24.97 (2.93)	26.12 (3.61)	25.53 (3.29)	26.76 (2.76)	27.83 (3.03)
Setting 4	24.90 (3.18)	24.97 (2.93)	25.26 (4.76)	25.81 (3.03)	27.38 (2.74)	27.96 (3.07)

Table 4: Estimated values for different ITRs under different proxy variable settings.

estimated ITRs in Setting 1, the coefficient of *cat1\_lung* is negative with a minor magnitude for  $\hat{d}_z$ , contrasting with a positive and relatively large coefficient observed for  $\hat{d}_w$ , which mirror the outcomes outlined in [Qi et al. \(2023\)](#). This finding suggests that, within the primary disease category of patients with lung cancer,  $\hat{d}_z$  advocates for undergoing RHC, while  $\hat{d}_w$  displays a notably inconclusive trend. As evidenced by  $\hat{\pi}$ , the prevailing trajectory for patients with *cat1\_lung* = 1 involves a strong inclination toward undergoing RHC, i.e.,  $\hat{\pi}(X) = 1$ , aligning with the guidance offered by  $\hat{d}_z$ . Significantly, the domain knowledge underscores the potential for patients with advanced lung cancer to develop complications like pulmonary hypertension and coma, potentially warranting RHC for assessing pulmonary vascular changes and informing treatment strategies ([Galie et al., 2009](#)), which lends support to the recommendations offered by our proposed regime. Furthermore, it is important to note that the whole group of patients can be regarded as unions of multiple subgroups based on various distinct features, and the superiority of  $\hat{d}_w$  is evident in some subgroups (e.g., *amihx*). These results show that our proposed ITR offers superior efficacy compared to  $\hat{d}_z$ ,  $\hat{d}_w$  and  $\hat{d}_{z \cup w}$  as our methodology incorporates selection through  $\hat{\pi}$ .