A Semiparametric Instrumented Difference-in-Differences Approach to Policy Learning

Pan Zhao, Yifan Cui

A Semiparametric Instrumented Difference-in-Differences Approach to Policy Learning

Pan Zhao[1]*, Yifan Cui[2]†

[1] PreMeDICaL, Inria

Desbrest Institute of Epidemiology and Public Health, University of Montpellier, France

[2] Center for Data Science, Zhejiang University, China

October 14, 2023

Abstract

Recently, there has been a surge in methodological development for the difference-in-differences (DiD) approach to evaluate causal effects. Standard methods in the literature rely on the parallel trends assumption to identify the average treatment effect on the treated. However, the parallel trends assumption may be violated in the presence of unmeasured confounding, and the average treatment effect on the treated may not be useful in learning a treatment assignment policy for the entire population. In this article, we propose a general instrumented DiD approach for learning the optimal treatment policy. Specifically, we establish identification results using a binary instrumental variable (IV) when the parallel trends assumption fails to hold. Additionally, we construct a Wald estimator, novel inverse probability weighting (IPW) estimators, and a class of semiparametric efficient and multiply robust estimators, with theoretical guarantees on consistency and asymptotic normality, even when relying on flexible machine learning algorithms for nuisance parameters estimation. Furthermore, we extend the instrumented DiD to the panel data setting. We evaluate our methods in extensive simulations and a real data application.

Keywords: individualized treatment rule, instrumental variable, multiple robustness, semiparametric efficiency, unmeasured confounding

*Email: pan.zhao@inria.fr

†Email: cuiyf@zju.edu.cn

1 Introduction

Data-driven individualized decision making has received increasing interests in many fields, such as precision medicine [Luedtke and van der Laan, 2016b, Tsiatis et al., 2019], econometrics and quantitative social sciences [Imai and van Dyk, 2004, Athey and Wager, 2021], computer science and operations research [Shi et al., 2022, Kallus et al., 2022]. The common goal is to learn optimal treatment assignment policies (also known as regimes, rules or plans) which map individual characteristics to treatment assignments so as to optimize some functional of the counterfactual outcome distributions, leveraging observational data where causal effects can be identified under various strategies and assumptions.

Popular existing methods in the statistical and machine learning literature include model-based approaches such as Q-learning [Watkins and Dayan, 1992, Murphy, 2003, Linn et al., 2017], A-learning [Robins et al., 2000, Shi et al., 2018], and direct model-free policy search approaches [Zhang et al., 2012, Zhao et al., 2012]. Recent advances of policy learning have also considered a variety of data structures, optimization objectives, criteria or constraints, such as survival and longitudinal data [Goldberg and Kosorok, 2012, Ertefaie and Strawderman, 2018, Zhao et al., 2023], networks [Viviano, 2019, Sherman et al., 2020], distributional robustness [Mo et al., 2021, Sahoo et al., 2022], budget, fairness, or interpretability constraints [Luedtke and van der Laan, 2016a, Fang et al., 2022], among others [Luedtke and Chambaz, 2020, Hadad et al., 2021, Nie et al., 2021, Hu et al., 2022, Jin et al., 2023].

With few exceptions, most methods in prior work rely on the pivotal assumption that there is no unmeasured confounding. This is a key threat to credible causal inference in observational studies, and may lead to suboptimal policies, because this assumption is impossible to verify or test in practice. An ad hoc work-around commonly adopted by practitioners is to collect and appropriately adjust for a large number of covariates, which still lacks theoretical guarantee and seems likely to be error-prone. To address this limitation, there has been recent progress made in several directions. Kallus and Zhou [2018] propose to minimize the worst-case regret of a policy under a marginal sensitivity model for the unmeasured confounding. Zhang et al. [2021] utilize a randomization test to rank by a partial order and select treatment rules within a given finite collection. While partial identification results provide certain improvement, the performance of

such a learned policy may still be suboptimal. Qi et al. [2023] build on the semiparametric proximal causal inference framework introduced by Cui et al. [2023] to establish point identification results on different policy classes and accordingly propose several classification-based approaches; but this framework requires the analyst to correctly classify the measured covariates into three types of proxies, and it may be difficult to estimate the confounding bridge functions.

Instrumental variable methods are widely used to handle unmeasured confounding in observational studies or randomized trials with non-compliance. The core requirements for a pretreatment variable to be a valid IV are: (i) it is associated with the treatment; (ii) it is independent of all unmeasured confounders; (iii) it does not have a direct causal effect on the outcome other than through the treatment. Along with the seminal work of Imbens and Angrist [1994], Angrist et al. [1996], extensive development has been made in using the IV to estimate the local average treatment effect [Tan, 2006, Ogburn et al., 2015], defined as the average treatment effect for the complier subgroup who would always comply with their treatment assignments. Since the complier subgroup is unknown and may have systematically different characteristics from the population, the population (conditional) average treatment effect is arguably the causal parameter of primary interest in most studies [Hernán and Robins, 2006, Aronow and Carnegie, 2013], especially for policy learning. More recently, Pu and Zhang [2021] consider a partial identification approach to optimal treatment rule estimation; and Wang and Tchetgen Tchetgen [2018] formally establish point identification of the population average treatment effect under alternative no-interaction assumptions, upon which Cui and Tchetgen Tchetgen [2021] propose various IV methods for estimating optimal treatment regimes. It is notable that all of these IV methods in the literature only consider the setting with a single time point, with the only exception of Xu et al. [2023], where the authors propose an IV approach to off-policy evaluation in confounded Markov decision processes with infinite horizons.

There has always been interest in exploiting the longitudinal structure common in datasets such as electronic health records and medical claims in epidemiology and biomedicine [Robins et al., 2000], as well as cross-sectional or panel data in program evaluations, economic censuses, and surveys [Athey and Imbens, 2017]. DiD methods have been an important tool widely used by empirical researchers [Card and Krueger, 1994]. The key identification assumption of DiD is that the trend in outcome of the control group over time is informative about what the trend would have been for the treatment group in the absence of the treatment. Specifically, under the standard (conditional)

parallel trends assumption, which states that the (conditional) expected trends in the potential outcomes of the two groups in the absence of the treatment are identical, the average treatment effect on the treated can be identified [Abadie, 2005, Sant'Anna and Zhao, 2020]; we refer interested readers to Lechner et al. [2011] and Roth et al. [2023] for detailed reviews. However, concerns often arise that the parallel trends assumption may be violated due to unmeasured confounding. Athey and Imbens [2006] develop a new changes-in-changes model that relates outcomes to an individual's group, time, and unobservable characteristics; and various recent extensions for DiD include partial identification [Ye et al., 2020], sensitivity analysis [Keele et al., 2019] and negative control [Sofer et al., 2016], among others [Dukes et al., 2022, Park and Tchetgen, 2023]. Moreover, DiD methods focus on the identification and estimation of the average treatment effect on the treated, which limits its application in policy learning since the treated cannot represent the population. To the best of our knowledge, this is the first work to systematically study policy learning under the DiD setting.

In this article, we combine the two natural experiments and propose an instrumented DiD approach to policy learning when the parallel trends assumption fails to hold in the presence of unmeasured confounding. Specifically, we adapt and extend the recent progress in Ye et al. [2022] and Vo et al. [2022], relaxing some key assumptions of the conventional IV and DiD methods. We allow for the violation of the parallel trends assumption by leveraging an IV which has no direct effect on the the trend in outcome, and does not modify the average treatment effect. Notably, this exogenous variable is not necessarily a valid instrument for the conventional treatment-outcome association, since we allow it to have a direct effect on the outcome not just through the treatment at each time point.

The contributions of this article are summarized as follows. First, we propose the direct policy search approach to learn optimal treatment assignment policies, based on the conditional average treatment effect estimators using instrumented DiD. This approach essentially allows us to learn the optimal policy that maximizes the estimated value within a restricted policy class. Second, we establish novel identification results of optimal policies for the instrumented DiD design subject to unmeasured confounding. The new results give rise to new inverse probability weighting estimators of optimal policies without necessarily identifying the value function for a given policy. Another interesting progress is also made towards identifying optimal policies without necessarily using the

subjects' realized treatment values. In summary, we construct a Wald estimator and novel inverse probability weighting estimators. A class of semiparametric efficient and multiply robust estimators is also proposed, which is consistent provided that a subset of several posited models indexing the observed data distribution is correctly specified. Third, we prove theoretical guarantees for the proposed multiply robust policy learning approaches. Specifically, we consider both parametric models and flexible data-adaptive machine learning algorithms with the cross-fitting procedure to estimate the nuisance parameters, to draw valid inferences under mild regularity conditions and certain rate of convergence conditions. In particular, we consider a restricted policy class indexed by an Euclidean parameter η and establish the $n^{-1/3}$ convergence rate of $\hat{\eta}$, even though its resultant limiting distribution is not standard. Fourth, we extend our proposed methods to the panel data setup. We establish identification of the conditional average treatment effect under alternative assumptions and provide the direct policy search approaches for panel data. The theoretical results for panel data can be similarly derived.

The rest of this article is organized as follows. In Section 2, we introduce the statistical framework of instrumental variable, DiD and policy learning. Section 3 develops our main methodology of learning the optimal policy using the instrumented DiD. Semiparametric efficiency results and multiply robust estimators are presented in Section 4. Section 5 establishes the asymptotic properties of the proposed estimators. Extensive simulations are reported in Section 6 to demonstrate the proposed methods, followed by a real data application in Section 7. Next, we consider the extension of our methods to panel data in Section 8. The article concludes in Section 9 with a discussion of some remarks and future work. All proofs and additional results are provided in the Supplementary Material.

2 Statistical framework

We first introduce some notation. Let X denote the p-dimensional vector of covariates that belongs to a covariate space $\mathcal{X} \subset \mathbb{R}^p$, $A \in \mathcal{A} = \{0, 1\}$ denote the binary treatment, $Y \in \mathbb{R}$ denote the outcome of interest, and $T \in \mathcal{T} = \{0, 1\}$ denote the time period. Suppose that $U = (U_0, U_1)$ is an unmeasured confounder of the effect of A on Y, and $Z \in \{0, 1\}$ is a binary instrumental variable; the observed data are $O = (X, A, Y, T, Z)$. We assume that the random samples $(O_1, \ldots, O_n)$

collected at the two time periods are independent and identically distributed (i.i.d.) observations of $O \sim P_0$, and there is no overlap between individuals in these two time periods. This setup is commonly known as the repeated cross-sectional data. Extension to panel data setting is studied in Section 8.

We use the potential outcomes framework [Neyman, 1923, Rubin, 1974] to define causal effects. Let $A_t(z)$ denote the potential exposure at time t if the instrument were set to level z, $Y_t(a)$ denote the potential outcome at time t if the exposure were set to level a and the instrument would take the same value it actually had, and $Y_t(z, a)$ denote the potential outcome at time t had the instrument and exposure been set to z, a respectively.

Without loss of generality, we assume that larger values of Y are more desirable. Our aim is to identify and estimate an policy $d : \mathcal{X} \to \mathcal{A}$, that maximizes the expected potential outcome in a counterfactual world had this policy been implemented on the population. The optimal policy at time t is given by $d_{\text{opt},t}(x) = I\{\tau_t(x) > 0\}$, where $\tau_t(x) = E[Y_t(1) - Y_t(0) \mid X = x]$ is the conditional average treatment effect (CATE) at time t.

Let $Y_t(d) = d(X)Y_t(1) + (1 - d(X))Y_t(0)$ denote the potential outcome under a hypothetical intervention that assigns treatment according to policy d. The value function of a policy d at time t is defined as $V_t(d) = E[Y_t(d)]$. Let $\mathcal{D}$ be the class of candidate policies of primary interest. The optimal policy can be obtained by directly maximizing the value function:

$$d_{\text{opt},t} = \arg\max_{d \in \mathcal{D}} V_t(d) = \arg\max_{d \in \mathcal{D}} E[\tau_t(X)d(X)]. \tag{1}$$

Throughout this article, we assume that the stable treatment effect over time assumption holds, which says that the CATE does not vary over time, and thus ensures that the optimal policy remains the same between the two time periods. The subscript t is omitted when it is clear from the context.

Remark 1. Our proposed instrumented DiD methodology can also be readily formulated in the weighted classification perspective. Pioneered by Zhang et al. [2012], this perspective has been widely used in the biostatistics and precision medicine literature, and enjoys certain robustness empirically. Specifically, the above maximization problem (1) can be transformed into the following

equivalent weighted classification problem:

$$d_{\text{opt}}(x) = \arg\max_{d\in\mathcal{D}} E[WI\{A = d(X)\}], \tag{2}$$

where W is regarded as a weight that is motivated by standard outcome regression, inverse probability weighting and doubly robust methods. Many robust classification methods and off-the-shelf implementations can be utilized.

3 Instrumented difference-in-differences

In this section, we introduce a general instrumented DiD framework for policy learning under endogeneity, and provide novel identification results. Let $\pi(t, z, x) = Pr(T = t, Z = z \mid X = x)$, and for any random variable $C \in \{A, Y\}$, we define $\mu_C(t, z, x) = E[C \mid T = t, Z = z, X = x]$, $\delta_C(x) = \mu_C(1, 1, x) - \mu_C(0, 1, x) - \mu_C(1, 0, x) + \mu_C(0, 0, x)$. We make the following identification assumptions.

Assumption 1 (Consistency). *$A = A_T(Z)$ and $Y = Y_T(A)$.*

Assumption 2 (Positivity). *$c_1 < \pi(t, z, x) < 1 - c_1$ for some $0 < c_1 < 1/2$.*

Assumption 3 (Random sampling). *$T \perp \{A_t(z), Y_t(a) : t = 0, 1, z = 0, 1, a = 0, 1\} \mid X, Z$.*

Assumption 4 (Stable treatment effect over time). *$E[Y_0(1) - Y_0(0) \mid X] = E[Y_1(1) - Y_1(0) \mid X]$.*

Assumption 1 is also known as the stable unit treatment value assumption, which states that there is no interference between subjects and no multiple versions of the instrument and treatment. Assumption 2 ensures the same support of X for each (T, Z) level. Assumption 3 is commonly assumed for repeated cross-sectional data [Abadie, 2005]. Assumption 4 requires that the CATE $\tau(x)$ does not vary over time, and thus ensures that the optimal policy remains the same between the two time periods.

Assumption 5 (Trend relevance). *$E[A_1(1) - A_0(1) \mid Z = 1, X] \neq E[A_1(0) - A_0(0) \mid Z = 0, X]$.*

Assumption 6 (Independence & exclusion restriction). *$Z \perp \{A_t(1), A_t(0), Y_t(1) - Y_t(0), Y_1(0) - Y_0(0) : t = 0, 1\} \mid X$.*

Assumption 7 (No unmeasured common effect modifier)**.** $Cov\{A_t(1) - A_t(0), Y_t(1) - Y_t(0) \mid X\} = 0$ *for* $t = 0, 1$.

Assumption 5 and 6 are parallel to the core assumptions in the standard IV literature. Directed acyclic graphs illustrating the causal structure are provided in Section A of the Supplementary Material. Assumption 5 states that the IV affects the trend in treatment. Assumption 6 requires that the IV is unconfounded, has no direct effect on the trend in outcome, and does not modify the treatment effect. This exogenous variable is not necessarily a valid instrument for the conventional treatment-outcome association, since we allow it to have a direct effect on the outcome not just through the treatment at each time point. Assumption 7 essentially states that there is no common effect modifier by an unmeasured confounder, of the additive effect of treatment on the outcome, and the additive effect of the IV on treatment. It has been studied in Cui and Tchetgen Tchetgen [2021], and relax certain no additive interaction assumptions in Wang and Tchetgen Tchetgen [2018]. We refer interested readers to Ye et al. [2022] for detailed discussion and concrete examples of an IV for DiD. Now we present our first identification result under the above assumptions.

Theorem 1. *Under Assumptions 1-7, the optimal policy is nonparametrically identified by*

$$\arg\max_{d\in\mathcal{D}} E\left[\frac{\delta_Y(X)}{\delta_A(X)} d(X)\right]. \tag{3}$$

Theorem 1 combines the Wald estimator for CATE and the direct policy search approach in Equation (1). Similarly, the IPW estimator proposed by Ye et al. [2022] can also be used to learn the optimal policy. Semiparametric efficient and multiply robust estimators are presented in Section 4. Next we propose our novel identification results, which also serves as basis for the estimators proposed in Section 4.

Theorem 2. *Under Assumptions 1-7, the optimal policy is nonparametrically identified by*

$$\arg\max_{d\in\mathcal{D}} E\left[\frac{(2Z-1)(2T-1)(2A-1)YI\{A = d(X)\}}{\pi(T, Z, X)\delta_A(X)}\right]. \tag{4}$$

Theorem 2 extends prior identification of CATE, and proposes a novel IPW estimator of the optimal policy without necessarily identifying the value function. Semiparametric efficiency results based on (4) are given in Section F and G of the Supplementary Material.

Theorem 3. *Under Assumptions 1-7, the optimal policy is nonparametrically identified by*

$$\arg\max_{d\in\mathcal{D}} E\left[\frac{(2T-1)YI\{Z=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\right]. \tag{5}$$

Theorem 3 essentially proves that we can identify the optimal policy without necessarily using the subjects' realized treatment values, for instance when $\delta_A(X)$ is known a priori, or when a separate sample with data on (A,X,T,Z) is available to estimate $\delta_A(X)$. To conclude this section, we propose the following estimators for optimal policies:

$$\begin{aligned}
\hat{d}_{\text{Wald}} &= \arg\max_{d\in\mathcal{D}} \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\delta}_Y(X_i)}{\hat{\delta}_A(X_i)} d(X_i), \\
\hat{d}_{\text{IPW1}} &= \arg\max_{d\in\mathcal{D}} \frac{1}{n}\sum_{i=1}^{n} \frac{(2Z_i-1)(2T_i-1)(2A_i-1)Y_iI\{A_i=d(X_i)\}}{\hat{\pi}(T_i,Z_i,X_i)\hat{\delta}_A(X_i)}, \\
\hat{d}_{\text{IPW2}} &= \arg\max_{d\in\mathcal{D}} \frac{1}{n}\sum_{i=1}^{n} \frac{(2T_i-1)Y_iI\{Z_i=d(X_i)\}}{\hat{\pi}(T_i,Z_i,X_i)\hat{\delta}_A(X_i)},
\end{aligned}$$

where $\hat{\delta}_Y$, $\hat{\delta}_A$ and $\hat{\pi}$ are estimated by parametric models or machine learning algorithms. Our simulation studies in Section 6 empirically shows comparable performance of the IPW estimators (4) and (5).

Remark 2. Similarly, classification-based estimators based on Theorem 2 and 3 can be proposed:

$$\arg\max_{d\in\mathcal{D}} E[\tilde{W}_1 I\{A=d(X)\}], \quad \arg\max_{d\in\mathcal{D}} E[\tilde{W}_2 I\{Z=d(X)\}], \tag{6}$$

respectively, where the weights are given by

$$\tilde{W}_1 = \frac{(2Z-1)(2T-1)(2A-1)Y}{\pi(T,Z,X)\delta_A(X)}, \quad \tilde{W}_2 = \frac{(2T-1)Y}{\pi(T,Z,X)\delta_A(X)}.$$

The Fisher consistency, excess risk bound and universal consistency of the estimated policy can also be established [Zhao et al., 2012].

4 Semiparametric efficiency and multiply robust estimators

In this section, we use semiparametric theory and propose multiply robust estimators. The Wald and the IPW approaches require the corresponding models to be correctly specified. Hence, methods that are robust against model misspecification are highly desired, where consistency is guaranteed when a subset of several posited models indexing the observed data distribution is correctly specified.

We consider the (uncentered) efficient influence function:

$$\Delta(O) = \frac{\delta_Y(X)}{\delta_A(X)} + \frac{(2Z-1)(2T-1)}{\pi(T,Z,X)\delta_A(X)} \left\{ Y - \mu_Y(T,Z,X) - \frac{\delta_Y(X)}{\delta_A(X)}(A - \mu_A(T,Z,X)) \right\},$$

which has been proposed in Ye et al. [2022]. Therefore, the optimal policy is identified by $\arg\max_{\mathcal{D}} E\left[\Delta(X)d(X)\right]$. Moreover, in light of the optimization tasks formulated in (6), we propose the following two choices of statistic:

$$W_1 = \frac{(2A-1)\delta_Y(X)}{\delta_A(X)} + \frac{(2A-1)(2Z-1)(2T-1)}{\pi(T,Z,X)\delta_A(X)} \left\{ Y - \mu_Y(T,Z,X) - \frac{\delta_Y(X)}{\delta_A(X)}(A - \mu_A(T,Z,X)) \right\},$$

and

$$W_2 = \frac{(2Z-1)\delta_Y(X)}{\delta_A(X)} + \frac{2T-1}{\pi(T,Z,X)\delta_A(X)} \left\{ Y - \mu_Y(T,Z,X) - \frac{\delta_Y(X)}{\delta_A(X)}(A - \mu_A(T,Z,X)) \right\},$$

which also enjoy the multiply robustness property.

First, we consider positing parametric models. Let $\mu_A(t,z,x;\alpha)$, $\mu_Y(t,z,x;\beta)$ and $\pi(t,z,x;\theta)$ denote the posited models. $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\theta}$ can be estimated by maximum likelihood estimation. In Theorem 4, we show the multiple robustness in the sense of maximizing the objective function (or minimizing the weighted classification error) in the union model of the following models:

$\mathcal{M}_1$: models for $\pi(t,z,x)$ and $\delta_A(x)$ are correct;

$\mathcal{M}_2$: models for $\pi(t,z,x)$ and $\delta_Y(x)/\delta_A(x)$ are correct;

$\mathcal{M}_3$: models for $\delta_Y(x)/\delta_A(x)$ and $\mu_C(0,0,x)$, $\mu_C(1,0,x)$, $\mu_C(0,1,x)$ for $C \in \{A,Y\}$ are correct.

Theorem 4. *Under Assumptions 1-7, the optimal policy is identified by*

$$\arg\max_{\mathcal{D}} E\left[W_1 I\{A = d(X)\}\right] = \arg\max_{\mathcal{D}} E\left[W_2 I\{Z = d(X)\}\right] = \arg\max_{\mathcal{D}} E\left[\Delta(X) d(X)\right], \tag{7}$$

under the union model $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$.

We also consider using modern machine learning methods to estimate these nuisance parameters. In practice, we apply the cross-fitting technique [Schick, 1986, Zheng and van der Laan, 2010, Chernozhukov et al., 2018], which is easy to implement. The cross-fitting procedure goes as follows. We randomly split data into K folds; the cross-fitted estimator is given by

$$\hat{M}_{CF} = \frac{1}{K} \sum_{k=1}^{K} P_{n,k}\{\Delta(O; \hat{\mu}_{A,-k}, \hat{\mu}_{Y,-k}, \hat{\pi}_{-k}) d(X)\},$$

where $P_{n,k}$ denote empirical averages only over the k-th fold, and $\hat{\mu}_{A,-k}$, $\hat{\mu}_{Y,-k}$ and $\hat{\pi}_{-k}$ denote the nuisance estimators constructed excluding the k-th fold. Similar cross-fitted estimators for $E\left[W_1 I\{A = d(X)\}\right]$ and $E\left[W_2 I\{Z = d(X)\}\right]$ can also be constructed in the same way.

5 Asymptotic analysis of policy learning

In this section, we study theoretical guarantees for our proposed policy learning approaches. While researchers have suggested applying machine learning algorithms to estimate the optimal policies from large classes which cannot be described by a finite dimensional parameter [Luedtke and van der Laan, 2016b, Künzel et al., 2019], it is also important to consider certain classes of policies for better interpretability and transparency, especially in clinical medicine and policy research [Zhang et al., 2015, Athey and Wager, 2021]. Specifically, here we focus on a class of feasible policies $\mathcal{D} = \left\{I\{\eta^\top X > 0\} : \eta \in \mathbb{H}\right\}$, where η indexes different policies and $\mathbb{H}$ is a compact subset of $\mathbb{R}^p$. That is, we analyze the following estimator:

$$\hat{\eta} = \arg\max_{\eta \in \mathbb{H}} \hat{M}(\eta) = \arg\max_{\eta \in \mathbb{H}} \frac{1}{n} \sum_{i=1}^{n} \hat{\Delta}(O_i) d(X_i; \eta),$$

where $\hat{M}(\eta)$ is estimated by posited parametric models, or the cross-fitted estimator. Let $\eta^* = \arg\max_{\eta \in \mathbb{H}} E[\Delta(X) d(X; \eta)]$ denote the Euclidean parameter that indexes the optimal policy. We

detail the main large sample property of our proposed estimator, that $\hat{\eta}$ converges to η^* at $n^{1/3}$ rate, and that $\hat{M}(\hat{\eta})$ is $n^{1/2}$-consistent and asymptotically normal under weak conditions (mostly requiring standard regularity conditions [White, 1982], or only that the nuisance parameters are estimated at faster than $n^{1/4}$ rates).

Remark 3. In order to obtain certain rates of convergence or regret bounds, it is necessary to require some control over the complexity of the class $\mathcal{D}$; see Athey and Wager [2021, Section 2.2] for examples of the VC-dimension of classes of linear rules, decision trees and monotone rules. Here we apply the empirical process techniques to establish theoretical guarantees for linear rules, which also hold on any other $\mathcal{D}$ indexed by finite-dimensional parameters. Also note that all identification and semiparametric efficiency results hold for any class of policies, and other optimization methods can be readily utilized.

We assume the following regularity conditions.

Condition 1. (i) *The supports of* X *and* Y *are bounded.* (ii) *The functions* $\mu_Y(t,z,x)$, $\mu_A(t,z,x)$ *and* $\pi(t,z,x)$ *are smooth and bounded for all* (t,z,x). (iii) *The function* $M(\eta)$ *is twice continuously differentiable in a neighborhood of* η^*; (iv) *For all* $\delta > 0$, *we have that* $Pr(|X^T\eta^*| \leq \delta) \leq c_2\delta$, *for some constant* $c_2 > 0$ *such that* $c_2\delta \leq 1$.

Condition 2. (i) $\sqrt{n}(\hat{\alpha} - \alpha^*) = O_p(1)$; (ii) $\sqrt{n}(\hat{\beta} - \beta^*) = O_p(1)$; (iii) $\sqrt{n}(\hat{\theta} - \theta^*) = O_p(1)$.

Theorem 5. *Under Assumptions 1-7, if Conditions 1 and 2 hold, we have* (i) $\|\hat{\eta} - \eta^*\|_2 = O_p(n^{-1/3})$; (ii) $\sqrt{n}\{M(\hat{\eta}) - M(\eta^*)\} = o_p(1)$; (iii) $\sqrt{n}\{\hat{M}(\hat{\eta}) - M(\eta^*)\} \to \mathcal{N}(0, \sigma_1^2)$, *where* σ_1^2 *is given in the Supplementary Material.*

Condition 1 (i), (ii) and (iii) are standard regularity conditions to establish uniform convergence. Condition 1 (iv), also known as the margin condition, is often assumed in the literature of classification [Tsybakov, 2004], reinforcement learning [Hu et al., 2022] and treatment assignment policies [Luedtke and Chambaz, 2020], to guarantee fast convergence rates. Condition 2 requires $\sqrt{n}$ convergence rates of parameter estimates of the posited models, which holds under mild conditions.

We assume the following conditions for the machine learning algorithms used to construct cross-fitted estimators.

Condition 3. $\|\hat{\mu}_A(t,z,X) - \mu_A(t,z,X)\|_{L_2} = o_p(n^{-1/4})$, $\|\hat{\mu}_Y(t,z,X) - \mu_Y(t,z,X)\|_{L_2} = o_p(n^{-1/4})$ *and* $\|\hat{\pi}(t,z,X) - \pi(t,z,X)\|_{L_2} = o_p(n^{-1/4})$, *for* $t, z = 0, 1$.

Theorem 6. *Under Assumptions 1-7, if Conditions 1 and 3 hold, we have* (i) $\|\hat{\eta} - \eta^*\|_2 = O_p(n^{-1/3})$; (ii) $\sqrt{n}\{M(\hat{\eta}) - M(\eta^*)\} = o_p(1)$; (iii) $\sqrt{n}\{\hat{M}(\hat{\eta}) - M(\eta^*)\} \to \mathcal{N}(0, \sigma_2^2)$, *where* σ_2^2 *is given in the Supplementary Material.*

Condition 3 says the nuisance estimators must be consistent and converge at a fast enough rate (essentially $n^{1/4}$ in L_2 norm). This is quite general and can be achieved by many existing algorithms under nonparametric smoothness, sparsity, or other structural constraints. According to Theorems 5 and 6 (ii), the regret of our estimated regime vanishes as the sample size increases. Theorems 5 and 6 (iii) imply that $\hat{M}(\hat{\eta})$ is a regular and asymptotic normal estimator of $M(\eta^*)$.

6 Simulations

In this section, we conduct extensive simulations to evaluate the finite-sample performance of the proposed estimators. Specifically, we compare them to the instrumental variable approach proposed by Cui and Tchetgen Tchetgen [2021], which is in principle valid only for a single time point. Replication code is available at GitHub.

We first describe the complete data generation process as follows. Baseline covariates $X = (X_1, X_2)^\top$ are generated from independent standard normal distributions. The time period indicator T is generated from a Bernoulli distribution with probability 0.5. The unmeasured confounders $U = (U_0, U_1)^\top$ are generated from independent bridge distributions with parameter 0.5 [1]. The instrumental variable Z is generated from a Bernoulli distribution with probability 0.5. The potential treatments and outcomes at time points $t = 0, 1$ are generated from the models:

$$
\begin{aligned}
Pr(A_0 = 1 \mid Z, U, X) &= \text{expit}(2 - 7Z + 0.2U_0 + 2X_1), \\
Pr(A_1 = 1 \mid Z, U, X) &= \text{expit}(-1.5 + 5Z - 0.15U_1 + 1.5X_2), \\
(Y_0 \mid Z, U, X, A_0) \sim \mathcal{N}(\mu_0, 1), &\quad (Y_1 \mid Z, U, X, A_1) \sim \mathcal{N}(\mu_1, 1),
\end{aligned}
$$

[1] The bridge density function is $p(u) = 1/(2\pi \cosh(u/2))$. We use the bridge distribution because by Wang and Louis [2003], the data generation process ensures that upon marginalizing over U, the model for $Pr(A_t = 1 \mid Z, X)$ remains a logistic regression.

where $\mu_0 = 200 + 10(A_0(1.5X_1 + 2X_2 - 0.5) + 0.5U_0 + 2Z + 1.5X_1 + 2X_2)$, and $\mu_1 = 240 + 10(A_1(1.5X_1 + 2X_2 - 0.5) + 0.5U_1 + 2Z + 2X_1 + 1.5X_2)$. Therefore, the optimal policy is $d_{\text{opt}}(x) = I\{3x_1 + 4x_2 - 1 > 0\}$. Let $A = TA_1 + (1-T)A_0, Y = TY_1 + (1-T)Y_0$; thus the observed cross-sectional data are (X, A, Y, T, Z).

A large test dataset of size $N = 1 \times 10^6$ is generated independently to evaluate the performance of different estimators. The percentage of correct decisions (PCD) of an estimated policy $\hat{d}(x)$ is computed by $1 - N^{-1}\sum_{i=1}^{N} |\hat{d}(X_i) - d_{\text{opt}}(X_i)|$.

We compare 7 estimators in our study: the two IPW estimators, the Wald estimator, and the two multiply robust estimators, along with the below IV estimators proposed by Cui and Tchetgen Tchetgen [2021]:

$$d_{\text{IV.t0}} = \arg\max_{d\in\mathcal{D}} \frac{1}{n_{t0}} \sum_{i=1}^{n} \frac{Z_i A_i Y_i I\{A_i = d(X_i)\} I\{T_i = 0\}}{\hat{\delta}_{t0}(X_i)\hat{\pi}_{t0}(Z_i, X_i)},$$
$$d_{\text{IV.t1}} = \arg\max_{d\in\mathcal{D}} \frac{1}{n_{t1}} \sum_{i=1}^{n} \frac{Z_i A_i Y_i I\{A_i = d(X_i)\} I\{T_i = 1\}}{\hat{\delta}_{t1}(X_i)\hat{\pi}_{t1}(Z_i, X_i)},$$

where n_{t0} and n_{t1} are the sample sizes at time point $0, 1$, respectively; $\delta_{t0}(x) = \mu_A(0,1,x) - \mu_A(0,0,x)$, $\delta_{t1}(x) = \mu_A(1,1,x) - \mu_A(1,0,x)$, $\pi_{t0}(z,x) = Pr(Z = z \mid X = x, T = 0)$, $\pi_{t1}(z,x) = Pr(Z = z \mid X = x, T = 1)$ are the nuisance parameters, and $\hat{\delta}_{t0}, \hat{\delta}_{t1}, \hat{\pi}_{t0}, \hat{\pi}_{t1}$ can be estimated using parametric models or machine learning algorithms. We utilize the genetic algorithm implemented in the R package `rgenoud` [Mebane Jr and Sekhon, 2011] to solve the optimization tasks.

First, we posit parametric models for the nuisance parameters. The linear/logistic regression models for $\mu_A(t,z,x;\alpha)$, $\mu_Y(t,z,x;\beta)$ and $\pi(t,z,x;\theta)$ are correctly specified. The sample size is $n = 5000$.

We also consider flexible machine learning algorithms for nuisance parameter estimation. Specifically, we apply the generalized random forests [Athey et al., 2019] implemented in the R package `grf` with default tuning parameters. For the cross-fitting procedure, we use $K = 4$ folds. The sample size is $n = 10^4$.

Figure 1 reports the main simulation results from 500 Monte Carlo replications. In both scenarios, the two standard IV estimators fail to learn the optimal policy, due to the direct effects of the treatment A on the outcomes Y_0, Y_1. The two IPW estimators perform much better, but the

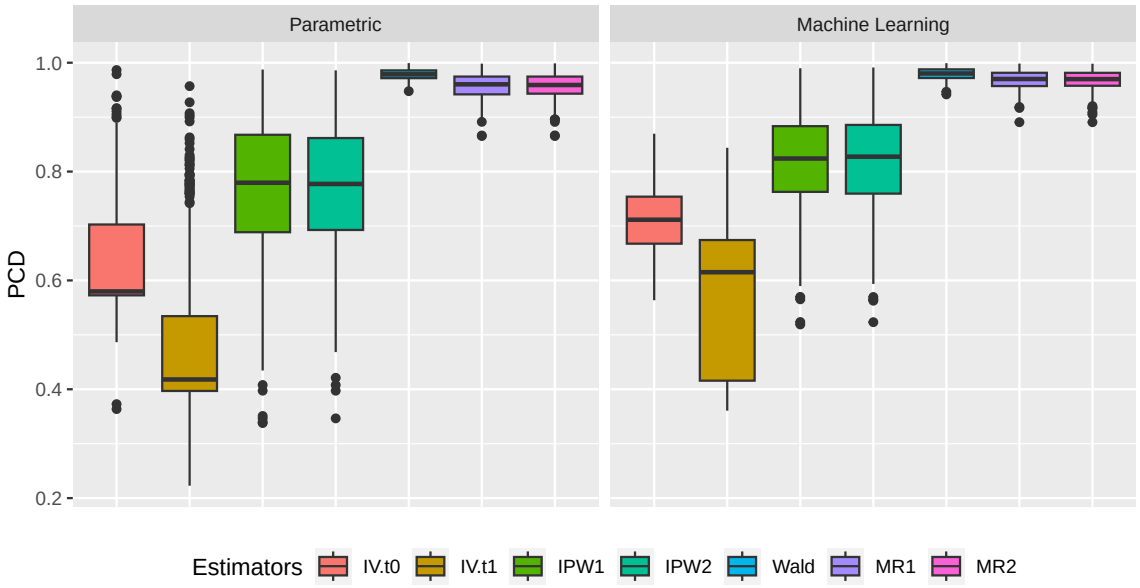

Figure 1: The percentage of correct decisions (PCD) results of the estimated optimal policies, using parametric models (left) or machine learning (right).

variability can be large due to possibly extreme weights. The Wald and multiply robust estimators generally lead to lower variability, and attain superior performance. Additional simulation results are reported in Section K of the Supplementary Material to illustrate how different sample sizes and the strength of the IV affect the performance of the estimated policies. We observe that a stronger strength of IV generally leads to lower variability and better accuracy, and also as sample size increases, our proposed methods have better performance.

7 Data application

In this section, we illustrate the use of the instrumented DiD approach for policy learning with a analysis of the Australian Longitudinal Survey (ALS) data. Researchers in labor economics have a longstanding interest in investigating the causal effect of education on earnings in the labor market. Card [2001] suggests that the endogeneity of education might partially explain the continuing interest "in this very difficult task of uncovering the causal effect of education in labor market outcomes", and argues that the effects of education are heterogeneous since the economic benefits are individual-specific. Besides the well acknowledged benefits of personal growth and social good from education, we aim to provide a personalized recommendation on whether an individual should

pursuit more education or not, in order to gain higher earnings.

The Australian Longitudinal Survey was conducted annually since 1984. Specifically, we include the 1984 and 1985 waves as cross sectional data in our analysis. The 1984 wave surveyed a sample of 3000 people aged $15-24$, and the 1985 wave consisted of 9000 interviews with people aged $16-25$. The surveys aim mainly at providing data on the dynamics of the youth labour market, and include basic demographic variables, labour market variables, background variables and topics related to the main labour market theme. We follow the guidelines from Su et al. [2013], Cai et al. [2006] and Vella [1994], who was among the first researchers extensively working with the ALS data. Finally, our data include 2401 subjects from the 1984 wave, and 8997 subjects from the 1985 wave. We consider the following baseline covariates: whether a person is born in Australia, marital status, union membership, government employment, age and work experience. The treatment is the education level, and the outcome is the hourly wage. We use an index of labor market attitudes as the instrumental variable [Su et al., 2013]. The details of our analysis are provided in Section L of the Supplementary Material.

The nuisance parameters are estimated by posited linear/logistic regression models, and we apply our proposed methods with the same configurations as Section 6. The policy coefficient estimates of all covariates are reported in Table 1.

Policies	`intercept`	`born_australia`	`married`	`uni_mem`	`gov_emp`	`age`	`year_expe`
IV.t0	0.4442	−0.4547	0.1311	−0.1179	−0.5181	0.0080	−0.5444
IV.t1	−0.2518	−0.3103	0.2445	−0.6157	−0.1406	0.2015	−0.5840
IPW1	−0.4203	−0.0847	0.5454	−0.3941	−0.5690	0.0299	0.1969
IPW2	−0.2503	−0.0529	0.6051	−0.4384	−0.5801	0.0207	0.1980
Wald	0.5032	0.3891	0.4738	0.5755	−0.1656	−0.0772	0.0793
MR1	−0.0513	0.1341	−0.6039	0.4127	0.5861	−0.0226	−0.3168
MR2	0.5480	−0.3937	−0.4072	0.4393	0.4167	−0.0302	−0.1064

Table 1: Coefficients of estimated optimal policy (normalized with L_2 norm 1). `born_australia`: whether a person is born in Australia; `married`: marital status; `uni_mem`: union membership; `gov_emp`: government employment; `age`: age; `year_expe`: work experience.

The coefficients should be interpreted cautiously. We also find that there exists some discrepancies among the treatment recommendations by our proposed estimators. The Wald and multiply robust estimators usually agree, but the variability of the IPW estimators are a bit large. Due to the potentially different recommendations by different estimated policies, one may conservatively

suggest a recommendation by the majority rule, and accordingly obtain an ensemble policy. It is also interesting to construct a decision tree to further explore which covariates indicate which treatment level [Qi et al., 2023].

8 Extension to panel data

In this section, we consider extending the instrumented DiD approach to the panel data setup where a random sample from the population is followed up over two time points [Abadie, 2005]. The observed data are $O = (X, Z, A_0, Y_0, A_1, Y_1)$. Let $\delta_{Y,z}(x) = E[Y_1 - Y_0 \mid X = x, Z = z]$, $\delta_{A,z}(x) = E[A_1 - A_0 \mid X = x, Z = z]$, and $\pi_Z(x) = Pr(Z = 1 \mid X = x)$. We make the following identification assumptions.

Assumption 8. *Suppose the following assumptions hold:* (consistency) $A_t = A_t(Z)$ *and* $Y_t = Y_t(A_t)$ *for* $t = 0, 1$*;* (positivity) $c_3 < \pi_Z(x) < 1 - c_3$ *for some* $0 < c_3 < 1/2$*;* (trend relevance) $E[A_1(1) - A_0(1) \mid Z = 1, X] \neq E[A_1(0) - A_0(0) \mid Z = 0, X]$*;* (stable treatment effect over time) $E[Y_0(1) - Y_0(0) \mid X] = E[Y_1(1) - Y_1(0) \mid X]$*;* (independence & exclusion restriction) $Z \perp \{A_t(1), A_t(0), Y_t(1) - Y_t(0), Y_1(0) - Y_0(0) : t = 0, 1\} \mid X$*;* (no unmeasured common effect modifier) $Cov\{A_t(1) - A_t(0), Y_t(1) - Y_t(0) \mid X\} = 0$ *for* $t = 0, 1$.

Assumption 8 is the counterpart of Assumptions 1-7 for the panel/longitudinal structure. Vo et al. [2022] use a structural mean model and consider alternative assumptions to the no unmeasured common effect modifier assumption above. In Section J of the Supplementary Material, we also prove the identification results under the following assumptions that replaces the no unmeasured common effect modifier assumption: (sequential ignorability) $Y_t(a) \perp A_t \mid U, X, Z$ for $t, a = 0, 1$, and there is no additive interaction of either (i) $E[A_1 - A_0 \mid X, U, Z = 1] - E[A_1 - A_0 \mid X, U, Z = 0] = E[A_1 - A_0 \mid X, Z = 1] - E[A_1 - A_0 \mid X, Z = 0]$ or (ii) $E[Y_t(1) - Y_t(0) \mid U, X] = E[Y_t(1) - Y_t(0) \mid X]$ for $t = 0, 1$. The sequential ignorability is intuitive, and commonly assumed in panel/longitudinal data analysis. We note that the no additive interaction assumption implies the no unmeasured common effect modifier assumption.

Theorem 7. *Under Assumption 8, the CATE is nonparametrically identified by*

$$\tau(x) = \frac{E[Y_1 - Y_0 \mid X = x, Z = 1] - E[Y_1 - Y_0 \mid X = x, Z = 0]}{E[A_1 - A_0 \mid X = x, Z = 1] - E[A_1 - A_0 \mid X = x, Z = 0]}, \tag{8}$$

and the efficient influence function is

$$\begin{aligned}\phi_{\text{panel}} =& \frac{\delta_{Y,1}(x) - \delta_{Y,0}(x)}{\delta_{A,1}(x) - \delta_{A,0}(x)} - \frac{z - \pi_Z(x)}{\pi_Z(x)(1 - \pi_Z(x))(\delta_{A,1}(x) - \delta_{A,0}(x))^2} \{(y_1 - y_0)(\delta_{A,1}(x) - \delta_{A,0}(x)) \\ & -(a_1 - a_0)(\delta_{Y,1}(x) - \delta_{Y,0}(x)) + \delta_{Y,1}(x)\delta_{A,0}(x) - \delta_{Y,0}(x)\delta_{A,1}(x)\} - \tau(x).\end{aligned}$$

Theorem 8. *Under Assumption 8, the optimal policy is nonparametrically identified by*

$$\arg\max_{\mathcal{D}} E\left[\frac{\delta_{Y,1}(X) - \delta_{Y,0}(X)}{\delta_{A,1}(X) - \delta_{A,0}(X)} d(X)\right] = \arg\max_{\mathcal{D}} E\left[\Delta_{\text{panel}}(X) d(X)\right], \tag{9}$$

where the uncentered efficient influence function Δ_{panel} *is*

$$\begin{aligned}\Delta_{\text{panel}} =& \frac{\delta_{Y,1}(X) - \delta_{Y,0}(X)}{\delta_{A,1}(X) - \delta_{A,0}(X)} - \frac{Z - \pi_Z(X)}{\pi_Z(X)(1 - \pi_Z(X))(\delta_{A,1}(X) - \delta_{A,0}(X))^2} \{(Y_1 - Y_0)(\delta_{A,1}(X) - \delta_{A,0}(X)) \\ & -(A_1 - A_0)(\delta_{Y,1}(X) - \delta_{Y,0}(X)) + \delta_{Y,1}(X)\delta_{A,0}(X) - \delta_{Y,0}(X)\delta_{A,1}(X)\}.\end{aligned}$$

Estimators of optimal policies can be constructed by the empirical versions of equations in Theorem 8, and the cross-fitting procedure can also be applied when using the efficient influence function. Similarly, asymptotic analysis of policy learning as Theorems 5 and 6 can be established for panel data.

9 Discussion

Similar approaches as the instrumented difference-in-differences design has long been employed by econometricians [Duflo, 2001] and has also been formally considered as fuzzy differences-in-differences by De Chaisemartin and d'Haultfoeuille [2018], where the individuals can switch treatment in only one direction within each treatment group. We refer interested readers to Ye et al. [2022] and its rejoinder for discussions on the differences, and applications in biomedicine and epidemiology.

There are several interesting directions for future research and application. Our approach is the first work to systematically study policy learning under the DiD setting. It may be possible to consider alternative assumptions or structures in DiD design to learn the optimal policy. Our instrumented DiD may also be generalized to multiple time points, continuous time, or continuous IV.

Note that Assumption 7 can be replaced by the *monotonicity* assumption, i.e. $A_t(1) \geq A_t(0)$ for $t = 0, 1$ with probability 1, which identifies the complier treatment effects. Then we can also target complier optimal policies that would optimize the potential outcome among compliers.

Acknowledgements

The authors are grateful for helpful comments and feedback from Julie Josse and Antoine Chambaz.

Pan Zhao is supported in part by the French National Research Agency ANR-16-IDEX-0006. Yifan Cui is supported in part by the National Natural Science Foundation of China and the Open Research Fund Key Laboratory of Advanced Theory and Application in Statistics and Data Science (East China Normal University), Ministry of Education of the People's Republic of China. The authors are grateful to the OPAL infrastructure from Université Côte d'Azur for providing resources and support.

References

Alberto Abadie. Semiparametric difference-in-differences estimators. *The review of economic studies*, 72(1):1–19, 2005.

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

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

Susan Athey and Guido W Imbens. Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497, 2006.

Susan Athey and Guido W Imbens. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic perspectives*, 31(2):3–32, 2017.

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

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *Annals of Statistics*, 47(2), 2019.

Zongwu Cai, Mitali Das, Huaiyu Xiong, and Xizhi Wu. Functional coefficient instrumental variables models. *Journal of Econometrics*, 133(1):207–241, 2006.

David Card. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160, 2001.

David Card and Alan B Krueger. Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *American Economic Review*, 84:772–793, 1994.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 01 2018. doi: 10.1111/ectj.12097.

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

Yifan Cui, Hongming Pu, Xu Shi, Wang Miao, and Eric Tchetgen Tchetgen. Semiparametric proximal causal inference. *Journal of the American Statistical Association*, pages 1–12, 2023.

Clément De Chaisemartin and Xavier d'Haultfoeuille. Fuzzy differences-in-differences. *The Review of Economic Studies*, 85(2):999–1028, 2018.

Esther Duflo. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American economic review*, 91(4):795–813, 2001.

Oliver Dukes, David Richardson, Zach Shahn, and Eric Tchetgen Tchetgen. Semiparametric bespoke instrumental variables. *arXiv preprint arXiv:2204.04119*, 2022.

Ashkan Ertefaie and Robert L Strawderman. Constructing dynamic treatment regimes over indefinite time horizons. *Biometrika*, 105(4):963–977, 2018.

Ethan X Fang, Zhaoran Wang, and Lan Wang. Fairness-oriented learning for optimal individualized treatment rules. *Journal of the American Statistical Association*, pages 1–14, 2022.

Yair Goldberg and Michael R Kosorok. Q-learning with censored data. *Annals of statistics*, 40(1): 529, 2012.

Vitor Hadad, David A Hirshberg, Ruohan Zhan, Stefan Wager, and Susan Athey. Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the national academy of sciences*, 118(15):e2014602118, 2021.

Miguel A Hernán and James M Robins. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*, pages 360–372, 2006.

Yichun Hu, Nathan Kallus, and Xiaojie Mao. Fast rates for contextual linear optimization. *Management Science*, 68(6):4236–4245, 2022.

Kosuke Imai and David A van Dyk. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99(467):854–866, 2004.

Guido W Imbens and Joshua D Angrist. Identification and estimation of local average treatment effects. *Econometrica: journal of the Econometric Society*, pages 467–475, 1994.

Ying Jin, Zhimei Ren, and Emmanuel J Candès. Sensitivity analysis of individual treatment effects: A robust conformal inference approach. *Proceedings of the National Academy of Sciences*, 120 (6):e2214889120, 2023.

Nathan Kallus and Angela Zhou. Confounding-robust policy improvement. *Advances in neural information processing systems*, 31, 2018.

Nathan Kallus, Xiaojie Mao, Kaiwen Wang, and Zhengyuan Zhou. Doubly robust distributionally robust off-policy evaluation and learning. In *International Conference on Machine Learning*, pages 10598–10632. PMLR, 2022.

Luke J Keele, Dylan S Small, Jesse Y Hsu, and Colin B Fogarty. Patterns of effects and sensitivity analysis for differences-in-differences. *arXiv preprint arXiv:1901.01869*, 2019.

Edward H. Kennedy, Sivaraman Balakrishnan, and Max G'Sell. Sharp instruments for classifying compliers and generalizing causal effects. *The Annals of Statistics*, 48(4):2008 – 2030, 2020. doi: 10.1214/19-AOS1874. URL https://doi.org/10.1214/19-AOS1874.

Michael R Kosorok. *Introduction to empirical processes and semiparametric inference*, volume 61. Springer, 2008.

Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10):4156–4165, 2019.

Michael Lechner et al. The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics*, 4(3):165–224, 2011.

Kristin A Linn, Eric B Laber, and Leonard A Stefanski. Interactive q-learning for quantiles. *Journal of the American Statistical Association*, 112(518):638–649, 2017.

Alexander Luedtke and Antoine Chambaz. Performance guarantees for policy learning. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(3):2162–2188, 2020.

Alexander R Luedtke and Mark J van der Laan. Optimal individualized treatments in resource-limited settings. *The international journal of biostatistics*, 12(1):283–303, 2016a.

Alexander R Luedtke and Mark J van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of statistics*, 44(2):713, 2016b.

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

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

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

Jersey Neyman. Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych*, 10(1):1–51, 1923.

Xinkun Nie, Emma Brunskill, and Stefan Wager. Learning when-to-treat policies. *Journal of the American Statistical Association*, 116(533):392–409, 2021.

Elizabeth L Ogburn, Andrea Rotnitzky, and James M Robins. Doubly robust estimation of the local average treatment effect curve. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 77(2):373, 2015.

Chan Park and Eric Tchetgen Tchetgen. A universal difference-in-differences approach for causal inference, 2023.

Hongming Pu and Bo 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.

Zhengling Qi, Rui Miao, and Xiaoke Zhang. Proximal learning for individualized treatment regimes under unmeasured confounding. *Journal of the American Statistical Association*, pages 1–14, 2023.

James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, pages 550–560, 2000.

Jonathan Roth, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 2023.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Roshni Sahoo, Lihua Lei, and Stefan Wager. Learning from a biased sample. *arXiv preprint arXiv:2209.01754*, 2022.

Pedro HC Sant'Anna and Jun Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020.

Anton Schick. On asymptotically efficient estimation in semiparametric models. *The Annals of Statistics*, pages 1139–1151, 1986.

Eli Sherman, David Arbour, and Ilya Shpitser. General identification of dynamic treatment regimes under interference. In *International Conference on Artificial Intelligence and Statistics*, pages 3917–3927. PMLR, 2020.

Chengchun Shi, Alin Fan, Rui Song, and Wenbin Lu. High-dimensional a-learning for optimal dynamic treatment regimes. *Annals of statistics*, 46(3):925, 2018.

Chengchun Shi, Jin Zhu, Shen Ye, Shikai Luo, Hongtu Zhu, and Rui Song. Off-policy confidence interval estimation with confounded markov decision process. *Journal of the American Statistical Association*, pages 1–12, 2022.

Tamar Sofer, David B Richardson, Elena Colicino, Joel Schwartz, and Eric J Tchetgen Tchetgen. On negative outcome control of unobserved confounding as a generalization of difference-in-differences. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 31 (3):348, 2016.

Liangjun Su, Irina Murtazashvili, and Aman Ullah. Local linear gmm estimation of functional coefficient iv models with an application to estimating the rate of return to schooling. *Journal of Business & Economic Statistics*, 31(2):184–207, 2013.

Zhiqiang Tan. Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association*, 101(476):1607–1618, 2006.

Anastasios A Tsiatis, Marie Davidian, Shannon T Holloway, and Eric B Laber. *Dynamic treatment regimes: Statistical methods for precision medicine*. CRC press, 2019.

Alexander B Tsybakov. Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, 32(1):135–166, 2004.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer New York, 1996.

Francis Vella. Gender roles and human capital investment: The relationship between traditional attitudes and female labour market performance. *Economica*, pages 191–211, 1994.

Davide Viviano. Policy targeting under network interference. *arXiv preprint arXiv:1906.10258*, 2019.

Tat-Thang Vo, Ting Ye, Ashkan Ertefaie, Samrat Roy, James Flory, Sean Hennessy, Stijn Vansteelandt, and Dylan S Small. Structural mean models for instrumented difference-in-differences. *arXiv preprint arXiv:2209.10339*, 2022.

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

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

Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8:279–292, 1992.

Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the econometric society*, pages 1–25, 1982.

Yang Xu, Jin Zhu, Chengchun Shi, Shikai Luo, and Rui Song. An instrumental variable approach to confounded off-policy evaluation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 38848–38880, 2023.

Ting Ye, Luke Keele, Raiden Hasegawa, and Dylan S Small. A negative correlation strategy for bracketing in difference-in-differences. *arXiv preprint arXiv:2006.02423*, 2020.

Ting Ye, Ashkan Ertefaie, James Flory, Sean Hennessy, and Dylan S Small. Instrumented difference-in-differences. *Biometrics*, 2022.

Baqun Zhang, Anastasios A Tsiatis, Marie Davidian, Min Zhang, and Eric Laber. Estimating optimal treatment regimes from a classification perspective. *Stat*, 1(1):103–114, 2012.

Bo Zhang, Jordan Weiss, Dylan S Small, and Qingyuan Zhao. Selecting and ranking individualized treatment rules with unmeasured confounding. *Journal of the American Statistical Association*, 116(533):295–308, 2021.

Yichi Zhang, Eric B Laber, Anastasios Tsiatis, and Marie Davidian. Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, 71(4):895–904, 2015.

Pan Zhao, Julie Josse, and Shu Yang. Efficient and robust transfer learning of optimal individualized treatment regimes with right-censored survival data. *arXiv preprint arXiv:2301.05491*, 2023.

Yingqi Zhao, Donglin Zeng, A John Rush, and Michael R Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118, 2012.

Wenjing Zheng and Mark J. van der Laan. Asymptotic theory for cross-validated targeted maximum likelihood estimation. Technical Report Working Paper 273, U.C. Berkeley Division of Biostatistics Working Paper Series, November 2010. URL https://biostats.bepress.com/ucbbiostat/paper273.

SUPPLEMENTARY MATERIAL

A Directed acyclic graphs

In this section, we present the directed acyclic graphs (DAGs) in Figures 2 and 3 illustrating the causal structure of the proposed instrumented DiD. The IV Z is associated with the trend in treatment $A_1 - A_0$, is independent of the unmeasured confounders U_0, U_1, cannot have direct effect on the trend in outcome $Y_1 - Y_0$, and does not modofy the treatment effect. But in comparison to a standard IV, here Z is allowed to have a direct effect on the outcomes Y_0, Y_1, as illustrated by the edges $Z \to Y_0$ and $Z \to Y_1$ in Figure 3.

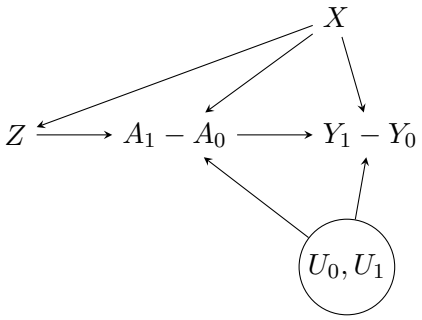

Figure 2: DAG for instrumented DiD on the trend scale.

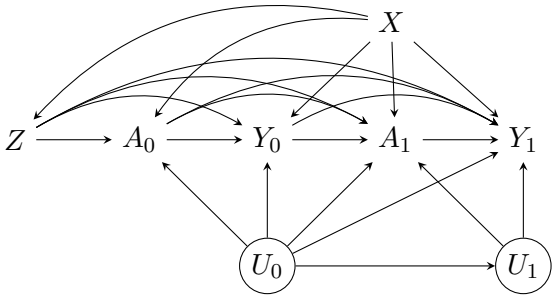

Figure 3: DAG for instrumented DiD over two time points.

B Proof of Theorem 1

In this section, we provide a proof of Theorem 1 for completeness. Similar proof can be found at Ye et al. [2022].

We first note that

$$
\begin{aligned}
&\delta_Y(X) = \mu_Y(1,1,X) - \mu_Y(0,1,X) - \mu_Y(1,0,X) + \mu_Y(0,0,X) \\
&= \sum_{z=0,1} (2z-1)(E[Y \mid T=1, Z=z, X] - E[Y \mid T=0, Z=z, X]) \\
&= \sum_{z=0,1} (2z-1)(E[Y_1(A_1(z)) \mid T=1, Z=z, X] - E[Y_0(A_0(z)) \mid T=0, Z=z, X]) \\
&= \sum_{z=0,1} (2z-1)(E[Y_1(A_1(z)) \mid Z=z, X] - E[Y_0(A_0(z)) \mid Z=z, X]) \\
&= \sum_{z=0,1} (2z-1)E[Y_1(A_1(z)) - Y_0(A_0(z)) \mid Z=z, X] \\
&= \sum_{z=0,1} (2z-1)E[A_1(z)Y_1(1) + (1-A_1(z))Y_1(0) - A_0(z)Y_0(1) - (1-A_0(z))Y_0(0) \mid Z=z, X] \\
&= \sum_{z=0,1} (2z-1)E[A_1(z)(Y_1(1) - Y_1(0)) - A_0(z)(Y_0(1) - Y_0(0)) + Y_1(0) - Y_0(0) \mid Z=z, X] \\
&= \sum_{z=0,1} (2z-1)(E[A_1(z)(Y_1(1) - Y_1(0)) \mid X] - E[A_0(z)(Y_0(1) - Y_0(0)) \mid X] + E[Y_1(0) - Y_0(0) \mid X]) \\
&= E[(A_1(1) - A_1(0))(Y_1(1) - Y_1(0)) \mid X] - E[(A_0(1) - A_0(0))(Y_0(1) - Y_0(0)) \mid X] \\
&= E[A_1(1) - A_1(0) \mid X]E[Y_1(1) - Y_1(0) \mid X] - E[A_0(1) - A_0(0) \mid X]E[Y_0(1) - Y_0(0) \mid X] \\
&= E[A_1(1) - A_1(0) - A_0(1) + A_0(0) \mid X]\tau(X).
\end{aligned}
$$

Then note that

$$
\begin{aligned}
&\delta_A(X) = \mu_A(1,1,X) - \mu_A(0,1,X) - \mu_A(1,0,X) + \mu_A(0,0,X) \\
&= \sum_{z=0,1} (2z-1)(E[A \mid T=1, Z=z, X] - E[A \mid T=0, Z=z, X]) \\
&= \sum_{z=0,1} (2z-1)(E[A_1(z) \mid T=1, Z=z, X] - E[A_0(z) \mid T=0, Z=z, X]) \\
&= E[A_1(1) - A_1(0) - A_0(1) + A_0(0) \mid X].
\end{aligned}
$$

Hence we have that $\delta_Y(X) = \delta_A(X)\tau(X)$. That is, the CATE $\tau(X)$ can be identified by

$\delta_Y(X)/\delta_A(X)$. It follows that the optimal policy is nonparametrically identified by

$$\arg\max_{d\in\mathcal{D}} E[\tau(X)d(X)] = \arg\max_{d\in\mathcal{D}} E\left[\frac{\delta_Y(X)}{\delta_A(X)}d(X)\right],$$

which completes the proof.

C Proof of Theorem 2

In this section, we prove our first novel identification results of the optimal policy.

First we note that

$$\begin{aligned}
&E\left[\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\right]\\
&=E\left[\sum_{a=0,1}\frac{(2Z-1)(2T-1)(2a-1)Y_T(a)I\{A=a\}I\{d(X)=a\}}{\pi(T,Z,X)\delta_A(X)}\right]\\
&=E\left[\sum_{a=0,1}\frac{(2Z-1)(2T-1)(2a-1)E[Y_T(a)\mid X,U]I\{A=a\}I\{d(X)=a\}}{\pi(T,Z,X)\delta_A(X)}\right]\\
&=E\left[\sum_{a=0,1}\frac{(2Z-1)(2T-1)(2a-1)E[Y_T(a)\mid X,U]Pr(A=a\mid X,U,T,Z)I\{d(X)=a\}}{\pi(T,Z,X)\delta_A(X)}\right]\\
&=E\left[\frac{Pr(A=1\mid X,U,T=1,Z=1)I\{d(X)=1\}E[Y_1(1)\mid X,U]}{\delta_A(X)}\right]\\
&\quad-E\left[\frac{Pr(A=1\mid X,U,T=0,Z=1)I\{d(X)=1\}E[Y_0(1)\mid X,U]}{\delta_A(X)}\right]\\
&\quad-E\left[\frac{Pr(A=1\mid X,U,T=1,Z=0)I\{d(X)=1\}E[Y_1(1)\mid X,U]}{\delta_A(X)}\right]\\
&\quad+E\left[\frac{Pr(A=1\mid X,U,T=0,Z=0)I\{d(X)=1\}E[Y_0(1)\mid X,U]}{\delta_A(X)}\right]\\
&\quad-E\left[\frac{Pr(A=0\mid X,U,T=1,Z=1)I\{d(X)=0\}E[Y_1(0)\mid X,U]}{\delta_A(X)}\right]\\
&\quad+E\left[\frac{Pr(A=0\mid X,U,T=0,Z=1)I\{d(X)=0\}E[Y_0(0)\mid X,U]}{\delta_A(X)}\right]\\
&\quad+E\left[\frac{Pr(A=0\mid X,U,T=1,Z=0)I\{d(X)=0\}E[Y_1(0)\mid X,U]}{\delta_A(X)}\right]\\
&\quad-E\left[\frac{Pr(A=0\mid X,U,T=0,Z=0)I\{d(X)=0\}E[Y_0(0)\mid X,U]}{\delta_A(X)}\right]\\
&=E\left[\frac{[Pr(A=1\mid X,U,T=1,Z=1)-Pr(A=1\mid X,U,T=1,Z=0)]I\{d(X)=1\}E[Y_1(1)\mid X,U]}{\delta_A(X)}\right]
\end{aligned}$$

$$+ E\left[\frac{[Pr(A=1 \mid X,U,T=1,Z=1) - Pr(A=1 \mid X,U,T=1,Z=0)]I\{d(X)=0\}E[Y_1(0) \mid X,U]}{\delta_A(X)}\right]$$
$$- E\left[\frac{[Pr(A=1 \mid X,U,T=0,Z=1) - Pr(A=1 \mid X,U,T=0,Z=0)]I\{d(X)=1\}E[Y_0(1) \mid X,U]}{\delta_A(X)}\right]$$
$$- E\left[\frac{[Pr(A=1 \mid X,U,T=0,Z=1) - Pr(A=1 \mid X,U,T=0,Z=0)]I\{d(X)=0\}E[Y_0(0) \mid X,U]}{\delta_A(X)}\right].$$

Since we have that for $t = 0, 1$,

$$\begin{aligned}
&I\{d(X)=1\}E[Y_t(1) \mid X,U] + I\{d(X)=0\}E[Y_t(0) \mid X,U] \\
&= d(X)(E[Y_t(1) \mid X,U] - E[Y_t(0) \mid X,U]) + E[Y_t(0) \mid X,U] \\
&= d(X)\tau(X) + E[Y_t(0) \mid X,U],
\end{aligned}$$

we continue by Assumption 7 that

$$\begin{aligned}
&E\left[\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\right] \\
&= E[d(X)\tau(X)] + E[\nu(X,U)],
\end{aligned}$$

where the second term $E[\nu(X,U)]$ does not depend on the policy d. That is,

$$\arg\max_{d\in\mathcal{D}} E\left[\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\right] = \arg\max_{d\in\mathcal{D}} E[\tau(X)d(X)],$$

which completes the proof.

D Proof of Theorem 3

In this section, we prove our second novel identification results of the optimal policy.

First we note that

$$\begin{aligned}
&E\left[\frac{(2T-1)YI\{Z=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\right] \\
&= E\left[\sum_{a=0,1}\frac{(2T-1)I\{Z=d(X)\}Y_T(a)I\{A=a\}}{\pi(T,Z,X)\delta_A(X)}\right]
\end{aligned}$$

$$
\begin{aligned}
&= E\left[\sum_{a=0,1} \frac{(2T-1)I\{Z=d(X)\}E[Y_T(a)\mid X,U]Pr(A=a\mid X,U,T,Z)}{\pi(T,Z,X)\delta_A(X)}\right] \\
&= E\left[\frac{Pr(A=1\mid X,U,T=1,Z=1)I\{d(X)=1\}E[Y_1(1)\mid X,U]}{\delta_A(X)}\right] \\
&\quad - E\left[\frac{Pr(A=1\mid X,U,T=0,Z=1)I\{d(X)=1\}E[Y_0(1)\mid X,U]}{\delta_A(X)}\right] \\
&\quad + E\left[\frac{Pr(A=1\mid X,U,T=1,Z=0)I\{d(X)=0\}E[Y_1(1)\mid X,U]}{\delta_A(X)}\right] \\
&\quad - E\left[\frac{Pr(A=1\mid X,U,T=0,Z=0)I\{d(X)=0\}E[Y_0(1)\mid X,U]}{\delta_A(X)}\right] \\
&\quad + E\left[\frac{Pr(A=0\mid X,U,T=1,Z=1)I\{d(X)=1\}E[Y_1(0)\mid X,U]}{\delta_A(X)}\right] \\
&\quad - E\left[\frac{Pr(A=0\mid X,U,T=0,Z=1)I\{d(X)=1\}E[Y_0(0)\mid X,U]}{\delta_A(X)}\right] \\
&\quad + E\left[\frac{Pr(A=0\mid X,U,T=1,Z=0)I\{d(X)=0\}E[Y_1(0)\mid X,U]}{\delta_A(X)}\right] \\
&\quad - E\left[\frac{Pr(A=0\mid X,U,T=0,Z=0)I\{d(X)=0\}E[Y_0(0)\mid X,U]}{\delta_A(X)}\right] \\
&= E\left[\frac{[Pr(A=1\mid X,U,T=1,Z=1)-Pr(A=1\mid X,U,T=1,Z=0)]I\{d(X)=1\}E[Y_1(1)\mid X,U]}{\delta_A(X)}\right] \\
&\quad + E\left[\frac{[Pr(A=1\mid X,U,T=1,Z=1)-Pr(A=1\mid X,U,T=1,Z=0)]I\{d(X)=0\}E[Y_1(0)\mid X,U]}{\delta_A(X)}\right] \\
&\quad - E\left[\frac{[Pr(A=1\mid X,U,T=0,Z=1)-Pr(A=1\mid X,U,T=0,Z=0)]I\{d(X)=1\}E[Y_0(1)\mid X,U]}{\delta_A(X)}\right] \\
&\quad - E\left[\frac{[Pr(A=1\mid X,U,T=0,Z=1)-Pr(A=1\mid X,U,T=0,Z=0)]I\{d(X)=0\}E[Y_0(0)\mid X,U]}{\delta_A(X)}\right] \\
&\quad + E\left[\frac{Pr(A=1\mid X,U,T=1,Z=0)E[Y_1(1)\mid X,U]+Pr(A=0\mid X,U,T=1,Z=1)E[Y_1(0)\mid X,U]}{\delta_A(X)}\right] \\
&\quad - E\left[\frac{Pr(A=1\mid X,U,T=0,Z=0)E[Y_0(1)\mid X,U]+Pr(A=0\mid X,U,T=0,Z=1)E[Y_0(0)\mid X,U]}{\delta_A(X)}\right].
\end{aligned}
$$

Then by the same arguments as in Section C, we have that

$$
E\left[\frac{(2T-1)YI\{Z=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\right] = E[d(X)\tau(X)] + E[\tilde{\nu}(X,U)],
$$

where the second term does not depend on the policy d. That is,

$$
\arg\max_{d\in\mathcal{D}} E\left[\frac{(2T-1)YI\{Z=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\right] = \arg\max_{d\in\mathcal{D}} E[\tau(X)d(X)],
$$

which completes the proof.

E Proof of Theorem 4

In this section, we prove our identification results of the optimal policy using the efficient influence functions.

First we note that

$$
\begin{aligned}
&E[W_1 I\{A = d(X)\}] \\
&= \frac{1}{2}E[W_1(2I\{A = d(X)\} - 1)] + \frac{1}{2}E[W_1] \\
&= \frac{1}{2}E[W_1(2A - 1)(2d(X) - 1)] + \frac{1}{2}E[W_1] \\
&= \frac{1}{2}E[\Delta(O)(2d(X) - 1)] + \frac{1}{2}E[W_1] \\
&= E[\Delta(O)d(X)] + \frac{1}{2}E[W_1 - \Delta(O)] \\
&= E[\tau(X)d(X)] + \frac{1}{2}E[W_1 - \Delta(O)],
\end{aligned}
$$

where the last equality holds under the union model $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$. The proof of the multiple robustness is omitted since it simply follows the same arguments of Theorem 1 in Ye et al. [2022].

We also note that

$$
\begin{aligned}
&E[W_2 I\{Z = d(X)\}] \\
&= \frac{1}{2}E[W_2(2I\{Z = d(X)\} - 1)] + \frac{1}{2}E[W_2] \\
&= \frac{1}{2}E[W_2(2Z - 1)(2d(X) - 1)] + \frac{1}{2}E[W_2] \\
&= \frac{1}{2}E[\Delta(O)(2d(X) - 1)] + \frac{1}{2}E[W_2] \\
&= E[\Delta(O)d(X)] + \frac{1}{2}E[W_2 - \Delta(O)] \\
&= E[\tau(X)d(X)] + \frac{1}{2}E[W_2 - \Delta(O)],
\end{aligned}
$$

where the last equality holds under the union model $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$.

F A locally efficient and multiply robust estimator

In this section, we present the semiparametric efficiency results for our proposed IPW formula:

$$\Psi(P) = E\left[\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\right].$$

We first characterize the efficient influence function, and then propose the multiply robust estimator.

Theorem 9. *The efficient influence function of* $\Psi(P)$ *is*

$$\begin{aligned}\phi_P = &\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\\ &-\frac{(2Z-1)(2T-1)E[(2A-1)YI\{A=d(X)\}\mid T,Z,X]}{\pi(T,Z,X)\delta_A(X)}+\gamma(X)\\ &-\frac{(2Z-1)(2T-1)(A-\mu_A(T,Z,X))\gamma(X)}{\pi(T,Z,X)\delta_A(X)}-\Psi(P),\end{aligned}$$

where $\gamma(x)=\sum_{t,z}(2z-1)(2t-1)E[(2A-1)YI\{A=d(X)\}\mid T=t,Z=z,X=x]/\delta_A(x)$.

By Theorem 9, we conclude that the optimal policy is nonparametrically identified by $\arg\max_{\mathcal{D}}\psi_P$, where

$$\begin{aligned}\psi_P = E\Bigg[&\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\\ &-\frac{(2Z-1)(2T-1)E[(2A-1)YI\{A=d(X)\}\mid T,Z,X]}{\pi(T,Z,X)\delta_A(X)}+\gamma(X)\\ &-\frac{(2Z-1)(2T-1)(A-\mu_A(T,Z,X))\gamma(X)}{\pi(T,Z,X)\delta_A(X)}\Bigg].\end{aligned}$$

In Theorem 10, we show the multiple robustness of the above formula under models:

$\tilde{\mathcal{M}}_1$: models for $\pi(t,z,x)$ and $\delta_A(x)$ are correct;

$\tilde{\mathcal{M}}_2$: models for $\pi(t,z,x)$ and $\gamma(x)$ are correct;

$\tilde{\mathcal{M}}_3$: models for $\mu_A(t,z,x)$, $\gamma(x)$ and $\nu(t,z,x)$ are correct, where $\nu(t,z,x)=E[(2A-1)YI\{A=d(X)\}\mid T=t,Z=z,X=x]$.

Theorem 10. *Under standard regularity conditions, we have that*

$$P_n\psi(\hat{P}) = P_n\Bigg[\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\hat{\pi}(T,Z,X)\hat{\delta}_A(X)} - \frac{(2Z-1)(2T-1)\hat{E}[(2A-1)YI\{A=d(X)\}\mid T,Z,X]}{\hat{\pi}(T,Z,X)\hat{\delta}_A(X)} + \hat{\gamma}(X) - \frac{(2Z-1)(2T-1)(A-\hat{\mu}_A(T,Z,X))\hat{\gamma}(X)}{\hat{\pi}(T,Z,X)\hat{\delta}_A(X)}\Bigg]$$

is a consistent and asymptotically normal estimator of $\Psi(P)$ *under the union model* $\tilde{\mathcal{M}}_1\cup\tilde{\mathcal{M}}_2\cup\tilde{\mathcal{M}}_3$. *Furthermore, it is locally efficient under the intersection model* $\tilde{\mathcal{M}}_1\cap\tilde{\mathcal{M}}_2\cap\tilde{\mathcal{M}}_3$.

Despite the fact that we characterize the efficient influence function and propose a multiply robust estimator, note that it is not straightforward to posit models for $\gamma(x)$ and $\nu(t,z,x)$.

G Proof of Theorem 9 and Theorem 10

We first prove Theorem 9 by deriving the efficient influence function.

For a given distribution P in the nonparametric statistical model $\mathcal{M}$, we let p denote the density of P with respect to some dominating measure ν. For all bounded $h\in L_2(P)$, define the parametric submodel $p_\epsilon=(1+\epsilon h)p$, which is valid for small enough ϵ and has score h at $\epsilon=0$.

We study the following statistical functional

$$\Psi(P) = E_P\left[\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\right],$$

and would establish that $\Psi(P)$ is pathwise differentiable with respect to $\mathcal{M}$ at P with efficient influence function ϕ_P if we have that for any $P\in\mathcal{M}$

$$\frac{\partial}{\partial\epsilon}\Psi(P_\epsilon)\Big|_{\epsilon=0} = \int\phi_P(o)h(o)dP(o).$$

We denote $\pi_\epsilon(t,z,x)=E_{P_\epsilon}[I\{T=t,Z=z\}\mid X=x]$, $\delta_{A,\epsilon}(x)=\mu_{A,\epsilon}(1,1,x)-\mu_{A,\epsilon}(0,1,x)-\mu_{A,\epsilon}(1,0,x)+\mu_{A,\epsilon}(0,0,x)$, $\mu_{A,\epsilon}(t,z,x)=E_{P_\epsilon}[A\mid T=t,Z=z,X=x]$, $S=\partial\log p_\epsilon/\partial\epsilon$, and

compute

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\Psi(P_\epsilon)\Big|_{\epsilon=0} &= \frac{\partial}{\partial\epsilon}E_{P_\epsilon}\left[\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi_\epsilon(T,Z,X)\delta_{A,\epsilon}(X)}\right]\Big|_{\epsilon=0} \\
&= \frac{\partial}{\partial\epsilon}E_P\left[(1+\epsilon S)\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi_\epsilon(T,Z,X)\delta_{A,\epsilon}(X)}\right]\Big|_{\epsilon=0} \\
&= E_P\left[S\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)}\right] \\
&\quad - E_P\left[\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi^2(T,Z,X)\delta_A^2(X)}\left(\delta_A(X)\frac{\partial}{\partial\epsilon}\pi_\epsilon(T,Z,X)\Big|_{\epsilon=0}\right.\right. \\
&\qquad \left.\left.+\pi(T,Z,X)\frac{\partial}{\partial\epsilon}\delta_{A,\epsilon}(X)\Big|_{\epsilon=0}\right)\right].
\end{aligned}$$

Then we need to compute

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\pi_\epsilon(t,z,X)\Big|_{\epsilon=0} &= \frac{\partial}{\partial\epsilon}E_{P_\epsilon}[I\{T=t,Z=z\}\mid X]\Big|_{\epsilon=0} \\
&= \frac{\partial}{\partial\epsilon}\frac{\pi(t,z,X)+\epsilon E_P[SI\{T=t,Z=z\}\mid X]}{1+\epsilon E_P[S\mid X]}\Big|_{\epsilon=0} \\
&= E_P[SI\{T=t,Z=z\}\mid X]-\pi(t,z,X)E_P[S\mid X],
\end{aligned}$$

$$\frac{\partial}{\partial\epsilon}\delta_{A,\epsilon}(X)\Big|_{\epsilon=0} = \frac{\partial}{\partial\epsilon}\{\mu_{A,\epsilon}(1,1,X)-\mu_{A,\epsilon}(0,1,X)-\mu_{A,\epsilon}(1,0,X)+\mu_{A,\epsilon}(0,0,X)\}\Big|_{\epsilon=0},$$

and

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\mu_{A,\epsilon}(t,z,X)\Big|_{\epsilon=0} &= \frac{\partial}{\partial\epsilon}E_{P_\epsilon}[A\mid T=t,Z=z,X]\Big|_{\epsilon=0} \\
&= \frac{\partial}{\partial\epsilon}\frac{\mu_A(t,z,X)+\epsilon E_P[SA\mid T=t,Z=z,X]}{1+\epsilon E_P[S\mid T=t,Z=z,X]}\Big|_{\epsilon=0} \\
&= E_P[SA\mid T=t,Z=z,X]-\mu_A(t,z,X)E_P[S\mid T=t,Z=z,X] \\
&= E_P\left[S\frac{(A-\mu_A(t,z,X))I\{T=t,Z=z\}}{\pi(t,z,X)}\mid X\right].
\end{aligned}$$

In summary, we obtain the efficient influence function

$$\begin{aligned}\phi_P = &\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)} \\ &- \frac{(2Z-1)(2T-1)E[(2A-1)YI\{A=d(X)\} \mid T,Z,X]}{\pi(T,Z,X)\delta_A(X)} + \gamma(X) \\ &- \frac{(2Z-1)(2T-1)(A-\mu_A(T,Z,X))\gamma(X)}{\pi(T,Z,X)\delta_A(X)} - \Psi(P),\end{aligned}$$

which completes the proof of Theorem 9.

Next, we prove Theorem 10 by verifying the multiple robustness property.

We first note the facts that $\mu_A(T,Z,X) = \mu_A(0,0,x)+Z(\mu_A(0,1,x)-\mu_A(0,0,x))+T(\mu_A(1,0,x)-\mu_A(0,0,x))+TZ\delta_A(X)$, $\nu(T,Z,X) = \nu(0,0,x)+Z(\nu(0,1,x)-\nu(0,0,x))+T(\nu(1,0,x)-\nu(0,0,x))+TZ\delta_A(X)$, $E[(2Z-1)(2T-1)/\pi(T,Z,X) \mid T,X] = E[(2Z-1)(2T-1)/\pi(T,Z,X) \mid Z,X] = 0$, and $E[\gamma(X)] = \Psi(P)$.

If $\tilde{\mathcal{M}}_1$ is correctly specified, we have that

$$\begin{aligned}E[\phi_P(O)] &= E\left[\frac{(2Z-1)(2T-1)(A-\mu_A(T,Z,X))\gamma(X)}{\pi(T,Z,X)\delta_A(X)}\right] \\ &= E\left[\frac{(2Z-1)(2T-1)\gamma(X)}{\pi(T,Z,X)\delta_A(X)}(A-\mu_A(T,Z,X))\right] = 0.\end{aligned}$$

If $\tilde{\mathcal{M}}_2$ is correctly specified, we have that

$$\begin{aligned}E[\phi_P(O)] = E\Bigg[&\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)} \\ &- \frac{(2Z-1)(2T-1)E[(2A-1)YI\{A=d(X)\} \mid T,Z,X]}{\pi(T,Z,X)\delta_A(X)} \\ &- \frac{(2Z-1)(2T-1)(A-\mu_A(T,Z,X))\gamma(X)}{\pi(T,Z,X)\delta_A(X)}\Bigg] = 0.\end{aligned}$$

If $\tilde{\mathcal{M}}_3$ is correctly specified, we have that

$$\begin{aligned}E[\phi_P(O)] = E\Bigg[&\frac{(2Z-1)(2T-1)(2A-1)YI\{A=d(X)\}}{\pi(T,Z,X)\delta_A(X)} \\ &- \frac{(2Z-1)(2T-1)E[(2A-1)YI\{A=d(X)\} \mid T,Z,X]}{\pi(T,Z,X)\delta_A(X)}\Bigg] = 0,\end{aligned}$$

which completes the proof.

H Proof of Theorem 5

We study the following maximization problem:

$$\hat{\eta} = \arg\max_{\eta \in \mathbb{H}} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\delta_Y(X_i;\hat{\beta})}{\delta_A(X_i;\hat{\alpha})} + \frac{(2Z_i-1)(2T_i-1)}{\pi(T_i,Z_i,X_i;\hat{\theta})\delta_A(X_i;\hat{\alpha})} \Big\{ Y_i - \mu_Y(T_i,Z_i,X_i;\hat{\beta}) - \frac{\delta_Y(X_i;\hat{\beta})}{\delta_A(X_i;\hat{\alpha})}(A_i - \mu_A(T_i,Z_i,X_i;\hat{\alpha})) \Big\} \right) d(X_i;\eta),$$

where $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\theta}$ are estimated by posited parametric models. We let $\hat{M}(\eta)$ denote the estimated objective function above, i.e. $\hat{\eta} = \arg\max_{\eta \in \mathbb{H}} \hat{M}(\eta)$.

Under standard regularity conditions, we have that

$$\begin{aligned}
\sqrt{n}(\hat{\alpha} - \alpha^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\alpha,i} + o_p(1), \\
\sqrt{n}(\hat{\beta} - \beta^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\beta,i} + o_p(1), \\
\sqrt{n}(\hat{\theta} - \theta^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\theta,i} + o_p(1),
\end{aligned}$$

where α^*, β^* and θ^* are the probability limits, ϕ_α, ϕ_β and ϕ_θ are the influence functions.

Now we start our proof which has three main parts as follows.

PART 1. First we note that, by the multiple robustness property, the strong law of large numbers and uniform consistency, $\hat{M}(\eta) = M(\eta) + o_p(1)$.

We denote

$$M_n^*(\eta) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\delta_Y(X_i;\beta^*)}{\delta_A(X_i;\alpha^*)} + \frac{(2Z_i-1)(2T_i-1)}{\pi(T_i,Z_i,X_i;\theta^*)\delta_A(X_i;\alpha^*)} \{ Y_i - \mu_Y(T_i,Z_i,X_i;\beta^*) - \frac{\delta_Y(X_i;\beta^*)}{\delta_A(X_i;\alpha^*)}(A_i - \mu_A(T_i,Z_i,X_i;\alpha^*)) \} \right) d(X_i;\eta),$$

and apply the Taylor expansion on $\hat{M}(\eta)$ at $(\alpha^*, \beta^*, \theta^*)$,

$$\hat{M}(\eta) = M_n^*(\eta) + H_{\alpha^*}^{\mathrm{T}}(\hat{\alpha} - \alpha^*) + H_{\beta^*}^{\mathrm{T}}(\hat{\beta} - \beta^*) + H_{\theta^*}^{\mathrm{T}}(\hat{\theta} - \theta^*) + o_p(n^{-1/2}),$$

where $H_{\alpha^*} = \lim_{n\to\infty} \partial\hat{M}(\eta)/\partial\alpha|_{\alpha=\alpha^*}$, $H_{\beta^*} = \lim_{n\to\infty} \partial\hat{M}(\eta)/\partial\beta|_{\beta=\beta^*}$, and $H_{\theta^*} = \lim_{n\to\infty} \partial\hat{M}(\eta)/\partial\theta|_{\theta=\theta^*}$.

Hence, we obtain that

$$\sqrt{n}\left\{\hat{M}(\eta)-M(\eta)\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(M_n^*(\eta)-M(\eta)+H_{\alpha^*}^{\mathrm{T}}\phi_{\alpha,i}+H_{\beta^*}^{\mathrm{T}}\phi_{\beta,i}+H_{\theta^*}^{\mathrm{T}}\phi_{\theta,i}\right)+o_p(1). \tag{10}$$

PART 2. We prove that $n^{1/3}\|\hat{\eta}-\eta^*\|_2=O_p(1)$.

First we note that, by Condition 1 (iii), $M(\eta)$ is twice continuously differentiable at a neighborhood of η^*. In PART 1, we show that $\hat{M}(\eta)=M(\eta)+o_p(1),\forall\eta$. Since $\hat{\eta}$ maximizes $\hat{M}(\eta)$, we have that $\hat{M}(\hat{\eta})\geq\sup_\eta\hat{M}(\eta)$; thus by the Argmax theorem, we obtain that $\hat{\eta}\xrightarrow{p}\eta^*$ as $n\to\infty$.

Then we apply Theorem 14.4 (Rate of convergence) of Kosorok [2008] to establish the $n^{-1/3}$ rate of convergence of $\hat{\eta}$, and need to find the suitable rate that satisfies three conditions below.

Condition 1 For every η in a neighborhood of η^* such that $\|\eta-\eta^*\|_2<\delta$, by Condition 1 (iii), we apply the second-order Taylor expansion,

$$\begin{aligned}M(\eta)-M(\eta^*)&=M'(\eta^*)\|\eta-\eta^*\|_2+\frac{1}{2}M''(\eta^*)\|\eta-\eta^*\|_2^2+o(\|\eta-\eta^*\|_2^2)\\&=\frac{1}{2}S''(\eta^*)\|\eta-\eta^*\|_2^2+o(\|\eta-\eta^*\|_2^2),\end{aligned}$$

and as $S''(\eta^*)<0$, there exists $c_0=-\frac{1}{2}S''(\eta^*)>0$ such that $S(t;\eta)-S(t;\eta^*)\leq c_0\|\eta-\eta^*\|_2^2$.

Condition 2 For all n large enough and sufficiently small δ, we consider the centered process $\hat{M}-M$, and have that

$$\begin{aligned}&E^*\left[\sqrt{n}\sup_{\|\eta-\eta^*\|_2<\delta}\left|\hat{M}(\eta)-M(\eta)-\left\{\hat{M}(\eta^*)-M(\eta^*)\right\}\right|\right]\\&=E^*\left[\sqrt{n}\sup_{\|\eta-\eta^*\|_2<\delta}\left|\hat{M}(\eta)-M_n^*(\eta)+M_n^*(\eta)-M(\eta)-\left\{\hat{M}(\eta^*)-M_n^*(\eta^*)+M_n^*(\eta^*)-M(\eta^*)\right\}\right|\right]\\&\leq E^*\left[\sqrt{n}\sup_{\|\eta-\eta^*\|_2<\delta}\left|\hat{M}(\eta)-M_n^*(\eta)-\left\{\hat{M}(\eta^*)-M_n^*(\eta^*)\right\}\right|\right]\\&\quad+E^*\left[\sqrt{n}\sup_{\|\eta-\eta^*\|_2<\delta}\left|M_n^*(\eta)-M(\eta)-\{M_n^*(\eta^*)-M(\eta^*)\}\right|\right]\\&=(I)+(II),\end{aligned}$$

where $E^*(\cdot)$ denote the outer expectation, and we bound (I) and (II) respectively as follows.

Condition 2.1 To bound (II), we note that

$$M_n^*(\eta) - M_n^*(\eta^*) = \frac{1}{n}\sum_{i=1}^{n} \Delta^*(O_i)(d(X_i;\eta) - d(X_i;\eta*)) \\ = \frac{1}{n}\sum_{i=1}^{n} \Delta^*(O_i)(I\{X_i^{\mathrm{T}}\eta > 0\} - I\{X_i^{\mathrm{T}}\eta^* > 0\}),$$

where

$$\Delta^*(o) = \frac{\delta_Y(x;\beta^*)}{\delta_A(x;\alpha^*)} + \frac{(2z-1)(2t-1)}{\pi(t,z,x;\theta^*)\delta_A(x;\alpha^*)}\left\{y - \mu_Y(t,z,x;\beta^*) - \frac{\delta_Y(x;\beta^*)}{\delta_A(x;\alpha^*)}(a - \mu_A(t,z,x;\alpha^*))\right\}.$$

We define a class of functions

$$\mathcal{F}_\eta^1(o) = \left\{\Delta^*(o)(I\{x^{\mathrm{T}}\eta > 0\} - I\{x^{\mathrm{T}}\eta^* > 0\}) : \|\eta - \eta^*\|_2 < \delta\right\},$$

and let $B_1 = \sup|\Delta^*(o)|$. By Assumption 2 and Condition 1, we have that $B_1 < \infty$.

When $\|\eta - \eta^*\|_2 < \delta$, by Condition 1 (i), there exists a constant $0 < k_0 < \infty$ such that $|x^{\mathrm{T}}(\eta - \eta^*)| < k_0\delta$. Furthermore, we show that $|d(x;\eta) - d(x;\eta^*)| = |I\{x^{\mathrm{T}}\eta > 0\} - I\{x^{\mathrm{T}}\eta^* > 0\}| \leq I\{-k_0\delta \leq x^{\mathrm{T}}\eta^* \leq k_0\delta\}$, by considering the three cases:

- when $-k_0\delta \leq x^{\mathrm{T}}\eta^* \leq k_0\delta$, we have $|d(x;\eta) - d(x;\eta^*)| \leq 1 = I\{-k_0\delta \leq x^{\mathrm{T}}\eta^* \leq k_0\delta\}$;
- when $x^{\mathrm{T}}\eta^* > k_0\delta > 0$, we have $x^{\mathrm{T}}\eta = x^{\mathrm{T}}(\eta - \eta^*) + x^{\mathrm{T}}\eta^* > 0$, so $|d(x;\eta) - d(x;\eta^*)| = 0 = I\{-k_0\delta \leq x^{\mathrm{T}}\eta^* \leq k_0\delta\}$;
- when $x^{\mathrm{T}}\eta^* < -k_0\delta < 0$, we have $x^{\mathrm{T}}\eta = x^{\mathrm{T}}(\eta - \eta^*) + x^{\mathrm{T}}\eta^* < 0$, so $|d(x;\eta) - d(x;\eta^*)| = 0 = I\{-k_0\delta \leq x^{\mathrm{T}}\eta^* \leq k_0\delta\}$.

Thus we define the envelope of $\mathcal{F}_\eta^1$ as $F_1 = B_1 I\{-k_0\delta \leq x^{\mathrm{T}}\eta^* \leq k_0\delta\}$. By Condition 1 (iv), there exists a constant $0 < k_1 < \infty$ such that

$$\|F_1\|_{P,2} \leq B_1\sqrt{Pr(-k_0\delta \leq x^{\mathrm{T}}\eta^* \leq k_0\delta)} \leq B_1\sqrt{2k_0k_1}\delta^{1/2} < \infty.$$

By Lemma 9.6 and Lemma 9.9 of Kosorok [2008], we have that $\mathcal{F}_\eta^1$, a class of indicator functions, is a Vapnik-Cervonenkis (VC) class with bounded bracketing entropy $J_{[]}^*(1, \mathcal{F}_\eta^1) < \infty$.

Next, we note that

$$\begin{aligned}\mathbb{G}_n\mathcal{F}_\eta^1 &= n^{-1/2}\sum_{i=1}^{n}\left\{\mathcal{F}_\eta^1(O_i) - E[\mathcal{F}_\eta^1(O)]\right\}\\ &= \sqrt{n}\left(M_n^*(\eta) - M_n^*(\eta^*) - \{M(\eta) - M(\eta^*)\}\right),\end{aligned}$$

and by Theorem 11.2 of Kosorok [2008], we obtain that there exists a constant $0 < c_1 < \infty$,

$$(II) = E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}|\mathbb{G}_n\mathcal{F}_\eta^1|\right] \le c_1 J_{[]}^*(1,\mathcal{F}_\eta^1)\|F_1\|_{P,2} \le c_1 J_{[]}^*(1,\mathcal{F}_\eta^1)B_1\sqrt{2k_0k_1}\delta^{1/2} = \tilde{c}_1\delta^{1/2},$$

hence we conclude that $(II) \le \tilde{c}_1\delta^{1/2}$, where $\tilde{c}_1 > 0$ is a finite constant.

Condition 2.2 To bound (I), first we note that

$$\begin{aligned}&\hat{M}(\eta) - M_n^*(\eta) - \{\hat{M}(\eta^*) - M_n^*(\eta^*)\} = \hat{M}(\eta) - \hat{M}(\eta^*) - \{M_n^*(\eta) - M_n^*(\eta^*)\}\\ &= \frac{1}{n}\sum_{i=1}^{n}(d(X_i;\eta) - d(X_i;\eta*))(\hat{\Delta}(O_i) - \Delta^*(O_i)),\end{aligned}$$

and then apply the Taylor expansion at $(\alpha^*, \beta^*, \theta^*)$

$$\begin{aligned}&\hat{M}(\eta) - M_n^*(\eta) - \{\hat{M}(\eta^*) - M_n^*(\eta^*)\}\\ &= \frac{1}{n}\sum_{i=1}^{n}(d(X_i;\eta) - d(X_i;\eta*))\left\{\left[g_1^*(O_i)\left(\frac{\partial\delta_A(X_i;\alpha^*)}{\partial\alpha}\right)^{\mathrm{T}} + g_2^*(O_i)\left(\frac{\partial\mu_A(T_i,Z_i,X_i;\alpha^*)}{\partial\alpha}\right)^{\mathrm{T}}\right](\hat{\alpha}-\alpha^*)\right.\\ &\quad + \left[g_3^*(O_i)\left(\frac{\partial\delta_Y(X_i;\beta^*)}{\partial\beta}\right)^{\mathrm{T}} + g_4^*(O_i)\left(\frac{\partial\mu_Y(T_i,Z_i,X_i;\beta^*)}{\partial\beta}\right)^{\mathrm{T}}\right](\hat{\beta}-\beta^*)\\ &\quad \left.+g_5^*(O_i)\left(\frac{\partial\pi(T_i,Z_i,X_i;\theta^*)}{\partial\theta}\right)^{\mathrm{T}}(\hat{\theta}-\theta^*)\right\} + o_p(n^{-1/2}),\end{aligned} \tag{11}$$

where

$$g_1^*(o) = -\frac{\delta_Y(x;\beta^*)}{\delta_A^2(x;\alpha^*)} - \frac{(2z-1)(2t-1)(y-\mu_Y(t,z,x;\beta^*))}{\pi(t,z,x;\theta^*)\delta_A^2(x;\alpha^*)} + \frac{2(2z-1)(2t-1)\delta_Y(x;\beta^*)}{\pi(t,z,x;\theta^*)\delta_A^3(x;\alpha^*)}(a-\mu_A(t,z,x;\alpha^*)),$$

$$g_2^*(o) = \frac{(2z-1)(2t-1)\delta_Y(x;\beta^*)}{\pi(t,z,x;\theta^*)\delta_A^2(x;\alpha^*)},$$

$$g_3^*(o) = \frac{1}{\delta_A^2(x;\alpha^*)} - \frac{2(2z-1)(2t-1)}{\pi(t,z,x;\theta^*)\delta_A^2(x;\alpha^*)}(a - \mu_A(t,z,x;\alpha^*)),$$

$$g_4^*(o) = -\frac{(2z-1)(2t-1)}{\pi(t,z,x;\theta^*)\delta_A(x;\alpha^*)},$$

$$g_5^*(o) = -\frac{(2z-1)(2t-1)}{\pi^2(t,z,x;\theta^*)\delta_A(x;\alpha^*)}\left\{y - \mu_Y(t,z,x;\beta^*) - \frac{\delta_Y(x;\beta^*)}{\delta_A(x;\alpha^*)}(a - \mu_A(t,z,x;\alpha^*))\right\}.$$

Similarly, we define the following classes of functions

$$\mathcal{F}_\eta^2(o) = \left\{\left[g_1^*(o)\left(\frac{\partial\delta_A(x;\alpha^*)}{\partial\alpha}\right)^{\mathrm{T}} + g_2^*(o)\left(\frac{\partial\mu_A(t,z,x;\alpha^*)}{\partial\alpha}\right)^{\mathrm{T}}\right](I\{x^{\mathrm{T}}\eta > 0\} - I\{x^{\mathrm{T}}\eta^* > 0\}) : \|\eta - \eta^*\|_2 < \delta\right\},$$

$$\mathcal{F}_\eta^3(o) = \left\{\left[g_3^*(o)\left(\frac{\partial\delta_Y(x;\beta^*)}{\partial\beta}\right)^{\mathrm{T}} + g_4^*(o)\left(\frac{\partial\mu_Y(t,z,x;\beta^*)}{\partial\beta}\right)^{\mathrm{T}}\right](I\{x^{\mathrm{T}}\eta > 0\} - I\{x^{\mathrm{T}}\eta^* > 0\}) : \|\eta - \eta^*\|_2 < \delta\right\},$$

$$\mathcal{F}_\eta^4(o) = \left\{g_5^*(o)\left(\frac{\partial\pi(t,z,x;\theta^*)}{\partial\theta}\right)^{\mathrm{T}}(I\{x^{\mathrm{T}}\eta > 0\} - I\{x^{\mathrm{T}}\eta^* > 0\}) : \|\eta - \eta^*\|_2 < \delta\right\},$$

and let $B_2 = \sup|g_1^*(o)\partial\delta_A(x;\alpha^*)/\partial\alpha + g_2^*(o)\partial\mu_A(t,z,x;\alpha^*)/\partial\alpha|$, $B_3 = \sup|g_3^*(o)\partial\delta_Y(x;\beta^*)/\partial\beta + g_4^*(o)\partial\mu_Y(t,z,x;\beta^*)/\partial\beta|$, and $B_4 = \sup|g_5^*(o)\partial\pi(t,z,x;\theta^*)/\partial\theta|$, where $B_2, B_3, B_4 > 0$ and the supremum is taken over all the coordinates. By Assumption 2 and Condition 1, we have that $B_2, B_3, B_4 < \infty$.

Using the same technique as in **Condition 2.1**, we define the envelop of $\mathcal{F}_\eta^j$ as $F_j = B_j I\{-k_0\delta \le x^{\mathrm{T}}\eta^* \le k_0\delta\}$ for $j = 2, 3, 4$, and obtain that

$$\|F_j\|_{P,2} \le \tilde{B}_j\delta^{1/2} < \infty, \quad j = 2, 3, 4,$$

where $\tilde{B}_2, \tilde{B}_3, \tilde{B}_4$ are some finite constants, and that $\mathcal{F}_\eta^j$ is a VC class with bounded bracketing entropy $J_{[]}^*(1, \mathcal{F}_\eta^j) < \infty$, for $j = 2, 3, 4$. By Theorem 11.2 of Kosorok [2008], we obtain that

$$E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}\left\|\mathbb{G}_N\mathcal{F}_\eta^j\right\|_1\right] \le c_j J_{[]}^*(1, \mathcal{F}_\eta^j)\|F_j\|_{P,2}, \quad j = 2, 3, 4,$$

where $c_2, c_3, c_4 > 0$ are some finite constants.

Furthermore, by Theorem 2.14.5 of van der Vaart and Wellner [1996], we obtain that

$$\begin{aligned}\left\{E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}\|\mathbb{G}_n\mathcal{F}_\eta^j\|_2^2\right]\right\}^{1/2} &\le l_j\left\{E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}\|\mathbb{G}_n\mathcal{F}_\eta^j\|_1\right]+\|F_j\|_{P,2}\right\}\\ &\le l_j\{c_j J_{[]}^*(1,\mathcal{F}_\eta^j)+1\}\|F_j\|_{P,2}\\ &\le \tilde{c}_j\delta^{1/2}, \qquad j=2,3,4,\end{aligned}$$

where l_2, l_3, l_4 and $\tilde{c}_2, \tilde{c}_3, \tilde{c}_4$ are some finite constants.

By Equation (11), we have that

$$\begin{aligned}(I) &= E^*\left[n^{1/2}\sup_{\|\eta-\eta^*\|_2<\delta}\left|\hat{M}(\eta)-M_n^*(\eta)-\{\hat{M}(\eta^*)-M_n^*(\eta^*)\}\right|\right]\\ &\le E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}\left\{|\mathbb{G}_n\mathcal{F}_\eta^2(\hat{\alpha}-\alpha^*)|+|\mathbb{G}_n\mathcal{F}_\eta^3(\hat{\beta}-\beta^*)|+|\mathbb{G}_n\mathcal{F}_\eta^4(\hat{\theta}-\theta^*)|+o_p(1)\right\}\right]\\ &\le n^{-1/2}\left\{E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}|\mathbb{G}_n\mathcal{F}_\eta^2\cdot n^{1/2}(\hat{\alpha}-\alpha^*)|\right]+E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}|\mathbb{G}_n\mathcal{F}_\eta^3\cdot n^{1/2}(\hat{\beta}-\beta^*)|\right]\right.\\ &\qquad\left.+E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}|\mathbb{G}_n\mathcal{F}_\eta^4\cdot n^{1/2}(\hat{\theta}-\theta^*)|\right]\right\},\end{aligned}$$

and then by the Cauchy-Schwarz inequality, we obtain that

$$\begin{aligned}(I) \le\, & n^{-1/2}\left\{E[n\|\hat{\alpha}-\alpha^*\|_2^2]\right\}^{1/2}\left\{E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}\|\mathbb{G}_n\mathcal{F}_\eta^2\|_2^2\right]\right\}^{1/2}\\ &+n^{-1/2}\left\{E[n\|\hat{\beta}-\beta^*\|_2^2]\right\}^{1/2}\left\{E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}\|\mathbb{G}_n\mathcal{F}_\eta^3\|_2^2\right]\right\}^{1/2}\\ &+n^{-1/2}\left\{E[n\|\hat{\theta}-\theta^*\|_2^2]\right\}^{1/2}\left\{E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta}\|\mathbb{G}_n\mathcal{F}_\eta^4\|_2^2\right]\right\}^{1/2}.\end{aligned}$$

By Condition 2, we have that $B_\alpha = \left\{E[n\|\hat{\alpha}-\alpha^*\|_2^2]\right\}^{1/2} < \infty$, $B_\beta = \left\{E[n\|\hat{\beta}-\beta^*\|_2^2]\right\}^{1/2} < \infty$, $B_\theta = \left\{E[n\|\hat{\theta}-\theta^*\|_2^2]\right\}^{1/2} < \infty$, hence

$$(I) \le n^{-1/2}(B_\alpha\tilde{c}_2 + B_\beta\tilde{c}_3 + B_\theta\tilde{c}_4)\delta^{1/2}.$$

In summary, we conclude that as $n \to \infty$, the centered process satisfies

$$E^* \left[\sqrt{n} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| \hat{M}(\eta) - M(\eta) - \{\hat{M}(\eta^*) - M(\eta^*)\} \right| \right] \leq (I) + (II) \leq \tilde{c}_1 \delta^{1/2}. \tag{12}$$

Let $\phi_n(\delta) = \delta^{1/2}$ and $b = \frac{3}{2} < 2$, thus we have $\frac{\phi_n(\delta)}{\delta^b} = \delta^{-1}$ is decreasing, and b does not depend on n.

Condition 3 By the facts that $\hat{\eta} \xrightarrow{p} \eta^*$ as $n \to \infty$, and that $\hat{M}(\hat{\eta}) \geq \sup_\eta \hat{M}(\eta)$, we choose $r_n = n^{1/3}$ such that $r_n^2 \phi_n(r_n^{-1}) = n^{2/3} \phi_n(n^{-1/3}) = n^{1/2}$.

In the end, the three conditions are satisfied with $r_n = n^{1/3}$; thus we conclude that $n^{1/3}\|\hat{\eta} - \eta^*\|_2 = O_p(1)$, which completes the proof of (i) of Theorem 5.

PART 3. We characterize the asymptotic distribution of $\hat{M}(\hat{\eta})$. First we note that

$$\sqrt{n}\{\hat{M}(\hat{\eta}) - M(\eta^*)\} = \sqrt{n}\{\hat{M}(\hat{\eta}) - \hat{M}(\eta^*)\} + \sqrt{n}\{\hat{M}(\eta^*) - M(\eta^*)\},$$

and then study the two terms in two steps.

Step 3.1 To establish $\sqrt{n}\{\hat{M}(\hat{\eta}) - \hat{M}(\eta^*)\} = o_p(1)$, it suffices to show that $\sqrt{n}\{M(\hat{\eta}) - M(\eta^*)\} = o_p(1)$ and $\sqrt{n}(\hat{M}(\hat{\eta}) - \hat{M}(\eta^*) - \{M(\hat{\eta}) - M(\eta^*)\}) = o_p(1)$.

First, as $n^{1/3}\|\hat{\eta} - \eta^*\|_2 = O_p(1)$, we apply the second-order Taylor expansion

$$\begin{aligned}
\sqrt{n}\{M(\hat{\eta}) - M(\eta^*)\} &= \sqrt{n} \left\{ M'(\eta^*)\|\hat{\eta} - \eta^*\|_2 + \frac{1}{2} M''(\eta^*)\|\hat{\eta} - \eta^*\|_2^2 + o_p(\|\hat{\eta} - \eta^*\|_2^2) \right\} \\
&= \sqrt{n} \left\{ \frac{1}{2} M''(\eta^*)\|\hat{\eta} - \eta^*\|_2^2 + o_p(\|\hat{\eta} - \eta^*\|_2^2) \right\} \\
&= \sqrt{n} \left\{ \frac{1}{2} M''(\eta^*) O_p(n^{-2/3}) + o_p(n^{-2/3}) \right\} = o_p(1),
\end{aligned}$$

which proves (ii) of Theorem 5.

Next, we follow the result (12) obtained in **PART 2**. As $n^{1/3}\|\hat{\eta} - \eta^*\|_2 = O_p(1)$, there exists

$\tilde{\delta} = c_5 n^{-1/3}$, where $c_5 < \infty$ is a finite constant, such that $\|\hat{\eta} - \eta^*\|_2 \leq \tilde{\delta}$. Therefore we have

$$\begin{aligned}
&\sqrt{n}(\hat{M}(\hat{\eta}) - \hat{M}(\eta^*) - \{M(\hat{\eta}) - M(\eta^*)\}) \\
&\leq E^* \left[\sqrt{n} \sup_{\|\hat{\eta}-\eta^*\|_2 < \tilde{\delta}} \left| \hat{M}(\hat{\eta}) - M(\hat{\eta}) - \{\hat{M}(\eta^*) - M(\eta^*)\} \right| \right] \\
&\leq \tilde{c}_1 \tilde{\delta}^{1/2} = \tilde{c}_1 \sqrt{c_5} n^{-1/6} = o_p(1),
\end{aligned}$$

which yields the result.

Step 3.2 To derive the asymptotic distribution of $\sqrt{n}\{\hat{M}(\eta^*) - M(\eta^*)\}$, we follow the result (10) obtained in **PART 1** and have that

$$\sqrt{n} \left\{ \hat{M}(\eta^*) - M(\eta^*) \right\} \xrightarrow{D} \mathcal{N}(0, \sigma_1^2),$$

where $\sigma_1^2 = E[(M^* - M + H_{\alpha^*}^{\mathrm{T}} \phi_{\alpha,i} + H_{\beta^*}^{\mathrm{T}} \phi_{\beta,i} + H_{\theta^*}^{\mathrm{T}} \phi_{\theta,i})^2]$.

Therefore we obtain in the end

$$\begin{aligned}
\sqrt{n}\{\hat{M}(\hat{\eta}) - M(\eta^*)\} &= \sqrt{n}\{\hat{M}(\hat{\eta}) - \hat{M}(\eta^*)\} + \sqrt{n}\{\hat{M}(\eta^*) - M(\eta^*)\} \\
&= o_p(1) + \sqrt{n}\{\hat{M}(\eta^*) - M(\eta^*)\} \\
&\xrightarrow{D} \mathcal{N}(0, \sigma_1^2),
\end{aligned}$$

which completes the proof.

I Proof of Theorem 6

We first review a useful lemma from Kennedy et al. [2020], which illustrates the basic technique of cross-fitting.

Lemma 1. *Consider two independent samples $\mathcal{O}_1 = (O_1, \ldots, O_n)$ and $\mathcal{O}_2 = (O_{n+1}, \ldots, O_{\tilde{n}})$, let $\hat{f}(o)$ be a function estimated from $\mathcal{O}_2$ and $\mathbb{P}_n$ the empirical measure over $\mathcal{O}_1$, then we have*

$$(\mathbb{P}_n - \mathbb{P})(\hat{f} - f) = O_{\mathbb{P}} \left(\frac{\|\hat{f} - f\|}{\sqrt{n}} \right)$$

Proof. First note that by conditioning on $\mathcal{O}_2$ we obtain

$$\mathbb{E}\left\{\mathbb{P}_n(\hat{f}-f)\,\middle|\,\mathcal{O}_2\right\} = \mathbb{E}(\hat{f}-f\,|\,\mathcal{O}_2) = \mathbb{P}(\hat{f}-f)$$

and the conditional variance is

$$var\{(\mathbb{P}_n-\mathbb{P})(\hat{f}-f)\,|\,\mathcal{O}_2\} = var\{\mathbb{P}_n(\hat{f}-f)\,|\,\mathcal{O}_2\} = \frac{1}{n}var(\hat{f}-f\,|\,\mathcal{O}_2) \leq \|\hat{f}-f\|^2/n$$

therefore by Chebyshev's inequality we have

$$\mathbb{P}\left\{\frac{|(\mathbb{P}_n-\mathbb{P})(\hat{f}-f)|}{\|\hat{f}-f\|^2/n} \geq t\right\} = \mathbb{E}\left[\mathbb{P}\left\{\frac{|(\mathbb{P}_n-\mathbb{P})(\hat{f}-f)|}{\|\hat{f}-f\|^2/n} \geq t\,\middle|\,\mathcal{O}_2\right\}\right] \leq \frac{1}{t^2}$$

thus for any $\epsilon > 0$ we can pick $t = 1/\sqrt{\epsilon}$ so that the probability above is no more than ϵ, which yields the result. □

We randomly split data into K folds. For $k = 1, \ldots, K$,

$$\hat{M}(\eta) = \frac{1}{K}\sum_{k=1}^{K}\hat{M}_k(\eta) = \frac{1}{K}\sum_{k=1}^{K}P_{n,k}\{\Delta(O;\hat{\mu}_{A,-k},\hat{\mu}_{Y,-k},\hat{\pi}_{-k})d(X)\},$$

where $P_{n,k}$ denote empirical averages only over the k-th fold, and $\hat{\mu}_{A,-k}$, $\hat{\mu}_{Y,-k}$ and $\hat{\pi}_{-k}$ denote the nuisance estimators constructed excluding the k-th fold.

Now we start our proof which has three main parts as follows.

PART 1. We prove that $\hat{M}(\eta) - M_n(\eta) = o_p(n^{-1/2})$, where $M_n(\eta) = P_n\{\Delta(O)d(X,\eta)\}$. Essentially it suffices to prove that $\hat{M}_k(\eta) - M_{n,k}(\eta) = o_p(n^{-1/2})$, where $M_{n,k}(\eta) = P_{n,k}\{\Delta(O)d(X,\eta)\}$.

First we note the following decomposition

$$
\begin{aligned}
&\hat{M}_k(\eta) - M_{n,k}(\eta) \\
&= P_{n,k}\, d(\eta) \left\{ \frac{\hat{\delta}_{Y,-k}}{\hat{\delta}_{A,-k}} - \frac{\delta_Y}{\delta_A} + (2Z-1)(2T-1) \left[\left(\frac{1}{\hat{\pi}_{-k}} - \frac{1}{\pi} \right) \left(\frac{1}{\hat{\delta}_{A,-k}} - \frac{1}{\delta_A} \right) \left(Y - \hat{\mu}_{Y,-k} - \frac{\hat{\delta}_{Y,-k}}{\hat{\delta}_{A,-k}} (A - \hat{\mu}_{A,-k}) \right) \right. \right. \\
&\quad + \frac{1}{\delta_A} \left(\frac{1}{\hat{\pi}_{-k}} - \frac{1}{\pi} \right) G_1 + \frac{1}{\pi} \left(\frac{1}{\hat{\delta}_{A,-k}} - \frac{1}{\delta_A} \right) G_1 + \frac{1}{\pi \delta_A} G_2 \\
&\quad + \frac{1}{\delta_A} \left(\frac{1}{\hat{\pi}_{-k}} - \frac{1}{\pi} \right) \left(Y - \mu_Y - \frac{\delta_Y}{\delta_A} (A - \mu_A) \right) + \frac{1}{\pi} \left(\frac{1}{\hat{\delta}_{A,-k}} - \frac{1}{\delta_A} \right) \left(Y - \mu_Y - \frac{\delta_Y}{\delta_A} (A - \mu_A) \right) \\
&\quad \left. \left. + \frac{1}{\pi \delta_A} \left(\mu_Y - \hat{\mu}_{Y,-k} - \frac{1}{\delta_A} (\hat{\delta}_{Y,-k} - \delta_Y)(A - \mu_A) - \delta_Y \left(\frac{1}{\hat{\delta}_{A,-k}} - \frac{1}{\delta_A} \right) (A - \mu_A) + \frac{\delta_Y}{\delta_A} (\hat{\mu}_{A,-k} - \mu_A) \right) \right] \right\},
\end{aligned}
$$

where we omit the arguments of the nuisance functions to simplify the notation, and denote

$$
\begin{aligned}
G_1 = \mu_Y - \hat{\mu}_{Y,-k} - (\hat{\delta}_{Y,-k} - \delta_Y) \left(\frac{1}{\hat{\delta}_{A,-k}} - \frac{1}{\delta_A} \right) (A - \mu_A) - \frac{1}{\delta_A} (\hat{\delta}_{Y,-k} - \delta_Y)(A - \mu_A) \\
+ \frac{1}{\delta_A} (\hat{\delta}_{Y,-k} - \delta_Y)(\hat{\mu}_{A,-k} - \mu_A) - \delta_Y \left(\frac{1}{\hat{\delta}_{A,-k}} - \frac{1}{\delta_A} \right) (A - \mu_A) \\
+ \delta_Y \left(\frac{1}{\hat{\delta}_{A,-k}} - \frac{1}{\delta_A} \right) (\hat{\mu}_{A,-k} - \mu_A) + \frac{\delta_Y}{\delta_A} (\hat{\mu}_{A,-k} - \mu_A),
\end{aligned}
$$

$$
G_2 = \frac{\hat{\delta}_{Y,-k} - \delta_Y}{\delta_A} (\hat{\mu}_{A,-k} - \mu_A) + \delta_Y \left(\frac{1}{\hat{\delta}_{A,-k}} - \frac{1}{\delta_A} \right) (\hat{\mu}_{A,-k} - \mu_A) - (\hat{\delta}_{Y,-k} - \delta_Y) \left(\frac{1}{\hat{\delta}_{A,-k}} - \frac{1}{\delta_A} \right) (A - \mu_A).
$$

In summary, we have two types of terms from this decomposition: product terms and mean zero terms (by multiple robustness). The product terms are $o_p(n^{-1/2})$ by Cauchy-Schwarz inequality and Condition 3 (rate of convergence). The mean zero terms are $o_p(n^{-1/2})$ by Lemma 1.

PART 2. We prove that $n^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$.

First we note that, by Condition 1 (iii), $M(\eta)$ is twice continuously differentiable at a neighborhood of η^*. In PART 1, we show that $\hat{M}(\eta) = M(\eta) + o_p(1), \forall \eta$. Since $\hat{\eta}$ maximizes $\hat{M}(\eta)$, we have that $\hat{M}(\hat{\eta}) \geq \sup_\eta \hat{M}(\eta)$; thus by the Argmax theorem, we obtain that $\hat{\eta} \xrightarrow{p} \eta^*$ as $n \to \infty$.

Then we apply Theorem 14.4 (Rate of convergence) of Kosorok [2008] to establish the $n^{-1/3}$ rate of convergence of $\hat{\eta}$, and need to find the suitable rate that satisfies three conditions below.

Condition 1 For every η in a neighborhood of η^* such that $\|\eta - \eta^*\|_2 < \delta$, by Condition 1 (iii),

we apply the second-order Taylor expansion,

$$\begin{aligned} M(\eta) - M(\eta^*) &= M'(\eta^*)\|\eta - \eta^*\|_2 + \frac{1}{2}M''(\eta^*)\|\eta - \eta^*\|_2^2 + o(\|\eta - \eta^*\|_2^2) \\ &= \frac{1}{2}S''(\eta^*)\|\eta - \eta^*\|_2^2 + o(\|\eta - \eta^*\|_2^2), \end{aligned}$$

and as $S''(\eta^*) < 0$, there exists $c_0 = -\frac{1}{2}S''(\eta^*) > 0$ such that $S(t;\eta) - S(t;\eta^*) \leq c_0\|\eta - \eta^*\|_2^2$.

Condition 2 For all n large enough and sufficiently small δ, we consider the centered process $\hat{M} - M$, and have that

$$\begin{aligned} & E^*\left[\sqrt{n}\sup_{\|\eta-\eta^*\|_2<\delta}\left|\hat{M}(\eta) - M(\eta) - \left\{\hat{M}(\eta^*) - M(\eta^*)\right\}\right|\right] \\ &= E^*\left[\sqrt{n}\sup_{\|\eta-\eta^*\|_2<\delta}\left|\hat{M}(\eta) - M_n^*(\eta) + M_n^*(\eta) - M(\eta) - \left\{\hat{M}(\eta^*) - M_n^*(\eta^*) + M_n^*(\eta^*) - M(\eta^*)\right\}\right|\right] \\ &\leq E^*\left[\sqrt{n}\sup_{\|\eta-\eta^*\|_2<\delta}\left|\hat{M}(\eta) - M_n^*(\eta) - \left\{\hat{M}(\eta^*) - M_n^*(\eta^*)\right\}\right|\right] \\ &\quad + E^*\left[\sqrt{n}\sup_{\|\eta-\eta^*\|_2<\delta}|M_n^*(\eta) - M(\eta) - \{M_n^*(\eta^*) - M(\eta^*)\}|\right] \\ &= (I) + (II), \end{aligned}$$

where $E^*(\cdot)$ denote the outer expectation, and we bound (I) and (II) respectively as follows.

It follows from the result in **PART 1** that $(I) = o_p(1)$.

To bound (II), we note that

$$\begin{aligned} M_n^*(\eta) - M_n^*(\eta^*) &= \frac{1}{n}\sum_{i=1}^n \Delta^*(O_i)(d(X_i;\eta) - d(X_i;\eta*)) \\ &= \frac{1}{n}\sum_{i=1}^n \Delta^*(O_i)(I\{X_i^{\mathrm{T}}\eta > 0\} - I\{X_i^{\mathrm{T}}\eta^* > 0\}), \end{aligned}$$

where

$$\Delta^*(o) = \frac{\delta_Y(x)}{\delta_A(x)} + \frac{(2z-1)(2t-1)}{\pi(t,z,x)\delta_A(x)}\left\{y - \mu_Y(t,z,x) - \frac{\delta_Y(x)}{\delta_A(x)}(a - \mu_A(t,z,x))\right\}.$$

We define a class of functions

$$\mathcal{F}_\eta^5(o) = \left\{\Delta^*(o)(I\{x^{\mathrm{T}}\eta > 0\} - I\{x^{\mathrm{T}}\eta^* > 0\}) : \|\eta - \eta^*\|_2 < \delta\right\},$$

and let $B_5 = \sup|\Delta^*(o)|$. By Assumption 2 and Condition 1, we have that $B_5 < \infty$.

Using the same technique as in Section **Condition 2.1**, we define the envelop of $\mathcal{F}_\eta^5$ as $F_5 = B_5 I\{-k_0\delta \le x^T\eta^* \le k_0\delta\}$, and obtain that $\|F_5\|_{P,2} \le \tilde{B}_9\delta^{1/2} < \infty$, where $\tilde{B}_9$ is a finite constant, and that $\mathcal{F}_\eta^5$ is a VC class with bounded entropy $J_{[]}^*(1, \mathcal{F}_\eta^5) < \infty$. By Theorem 11.2 of Kosorok [2008], we obtain that there exists a constant $0 < c_6 < \infty$,

$$(II) = E^*\left[\sup_{\|\eta-\eta^*\|_2<\delta} |\mathbb{G}_n\mathcal{F}_\eta^5|\right] \le c_6 J_{[]}^*(1, \mathcal{F}_\eta^5)\|F_5\|_{P,2} \le c_6 J_{[]}^*(1, \mathcal{F}_\eta^5) B_5\sqrt{2k_0k_1}\delta^{1/2} = \tilde{c}_5\delta^{1/2}.$$

In summary, we conclude that as $n \to \infty$, the centered process satisfies

$$E^*\left[\sqrt{n}\sup_{\|\eta-\eta^*\|_2<\delta}\left|\hat{M}(\eta) - M(\eta) - \{\hat{M}(\eta^*) - M(\eta^*)\}\right|\right] \le (I) + (II) \le \tilde{c}_5\delta^{1/2}. \tag{13}$$

Let $\phi_n(\delta) = \delta^{1/2}$ and $b = \frac{3}{2} < 2$, thus we have $\frac{\phi_n(\delta)}{\delta^b} = \delta^{-1}$ is decreasing, and b does not depend on n.

Condition 3 By the facts that $\hat{\eta} \xrightarrow{p} \eta^*$ as $n \to \infty$, and that $\hat{M}(\hat{\eta}) \ge \sup_\eta \hat{M}(\eta)$, we choose $r_n = n^{1/3}$ such that $r_n^2\phi_n(r_n^{-1}) = n^{2/3}\phi_n(n^{-1/3}) = n^{1/2}$.

In the end, the three conditions are satisfied with $r_n = n^{1/3}$; thus we conclude that $n^{1/3}\|\hat{\eta} - \eta^*\|_2 = O_p(1)$, which completes the proof of (i) of Theorem 6.

PART 3. We characterize the asymptotic distribution of $\hat{M}(\hat{\eta})$. First we note that

$$\sqrt{n}\{\hat{M}(\hat{\eta}) - M(\eta^*)\} = \sqrt{n}\{\hat{M}(\hat{\eta}) - \hat{M}(\eta^*)\} + \sqrt{n}\{\hat{M}(\eta^*) - M(\eta^*)\},$$

and then study the two terms in two steps.

Step 3.1 To establish $\sqrt{n}\{\hat{M}(\hat{\eta}) - \hat{M}(\eta^*)\} = o_p(1)$, it suffices to show that $\sqrt{n}\{M(\hat{\eta}) - M(\eta^*)\} = o_p(1)$ and $\sqrt{n}(\hat{M}(\hat{\eta}) - \hat{M}(\eta^*) - \{M(\hat{\eta}) - M(\eta^*)\}) = o_p(1)$.

First, as $n^{1/3}\|\hat{\eta} - \eta^*\|_2 = O_p(1)$, we apply the second-order Taylor expansion

$$\begin{aligned}
\sqrt{n}\{M(\hat{\eta}) - M(\eta^*)\} &= \sqrt{n}\left\{M'(\eta^*)\|\hat{\eta} - \eta^*\|_2 + \frac{1}{2}M''(\eta^*)\|\hat{\eta} - \eta^*\|_2^2 + o_p(\|\hat{\eta} - \eta^*\|_2^2)\right\} \\
&= \sqrt{n}\left\{\frac{1}{2}M''(\eta^*)\|\hat{\eta} - \eta^*\|_2^2 + o_p(\|\hat{\eta} - \eta^*\|_2^2)\right\} \\
&= \sqrt{n}\left\{\frac{1}{2}M''(\eta^*)O_p(n^{-2/3}) + o_p(n^{-2/3})\right\} = o_p(1),
\end{aligned}$$

which proves (ii) of Theorem 6.

Next, we follow the result (13) obtained in **PART 2**. As $n^{1/3}\|\hat{\eta} - \eta^*\|_2 = O_p(1)$, there exists $\tilde{\delta} = c_7 n^{-1/3}$, where $c_7 < \infty$ is a finite constant, such that $\|\hat{\eta} - \eta^*\|_2 \leq \tilde{\delta}$. Therefore we have

$$\begin{aligned}
&\sqrt{n}(\hat{M}(\hat{\eta}) - \hat{M}(\eta^*) - \{M(\hat{\eta}) - M(\eta^*)\}) \\
&\leq E^*\left[\sqrt{n}\sup_{\|\hat{\eta}-\eta^*\|_2<\tilde{\delta}}\left|\hat{M}(\hat{\eta}) - M(\hat{\eta}) - \{\hat{M}(\eta^*) - M(\eta^*)\}\right|\right] \\
&\leq \tilde{c}_5\tilde{\delta}^{1/2} = \tilde{c}_5\sqrt{c_7}n^{-1/6} = o_p(1),
\end{aligned}$$

which yields the result.

Step 3.2 To derive the asymptotic distribution of $\sqrt{n}\{\hat{M}(\eta^*) - M(\eta^*)\}$, we follow the result obtained in **PART 1** and have that

$$\sqrt{n}\left\{\hat{M}(\eta^*) - M(\eta^*)\right\} \xrightarrow{D} \mathcal{N}(0, \sigma_2^2),$$

where $\sigma_2^2 = E[(\Delta(O_i)d(X_i;\eta^*) - M(\eta^*))^2]$.

Therefore we obtain in the end

$$\begin{aligned}
\sqrt{n}\{\hat{M}(\hat{\eta}) - M(\eta^*)\} &= \sqrt{n}\{\hat{M}(\hat{\eta}) - \hat{M}(\eta^*)\} + \sqrt{n}\{\hat{M}(\eta^*) - M(\eta^*)\} \\
&= o_p(1) + \sqrt{n}\{\hat{M}(\eta^*) - M(\eta^*)\} \\
&\xrightarrow{D} \mathcal{N}(0, \sigma_2^2),
\end{aligned}$$

which completes the proof.

J Proof of Theorem 7 and 8

We first prove the identification result.

First we note that

$$\begin{aligned}
&\delta_{Y,1}(X) - \delta_{Y,0}(X) = E[Y_1 - Y_0 \mid X, Z = 1] - E[Y_1 - Y_0 \mid X, Z = 0] \\
&= \sum_{z=0,1} (2z-1) E[Y_1 - Y_0 \mid X, Z = z] \\
&= \sum_{z=0,1} (2z-1) E[Y_1(A_1(z)) - Y_0(A_0(z)) \mid X, Z = z] \\
&= \sum_{z=0,1} (2z-1) E[A_1(z)Y_1(1) + (1 - A_1(z))Y_1(0) - A_0(z)Y_0(1) - (1 - A_0(z))Y_0(0) \mid Z = z, X] \\
&= \sum_{z=0,1} (2z-1) E[A_1(z)(Y_1(1) - Y_1(0)) - A_0(z)(Y_0(1) - Y_0(0)) + Y_1(0) - Y_0(0) \mid Z = z, X] \\
&= \sum_{z=0,1} (2z-1) (E[A_1(z)(Y_1(1) - Y_1(0)) \mid X, Z = z] - E[A_0(z)(Y_0(1) - Y_0(0)) \mid X, Z = z] \\
&\qquad\qquad + E[Y_1(0) - Y_0(0) \mid X, Z = z]) \\
&= \sum_{z=0,1} (2z-1) (E[A_1(z)(Y_1(1) - Y_1(0)) \mid X, Z = z] - E[A_0(z)(Y_0(1) - Y_0(0)) \mid X, Z = z] \\
&= \sum_{z=0,1} (2z-1) (E[A_1(z)(Y_1(1) - Y_1(0)) \mid X] - E[A_0(z)(Y_0(1) - Y_0(0)) \mid X] \\
&= E[(A_1(1) - A_1(0))(Y_1(1) - Y_1(0)) \mid X] - E[(A_0(1) - A_0(0))(Y_0(1) - Y_0(0)) \mid X] \\
&= E[A_1(1) - A_1(0) \mid X]\tau(X) - E[A_0(1) - A_0(0) \mid X]\tau(X) \\
&= E[A_1(1) - A_1(0) - A_0(1) + A_0(0) \mid X]\tau(X).
\end{aligned}$$

We also note that

$$\begin{aligned}
&\delta_{A,1}(X) - \delta_{A,0}(X) = E[A_1 - A_0 \mid X, Z = 1] - E[A_1 - A_0 \mid X, Z = 0] \\
&= \sum_{z=0,1} (2z-1) E[A_1 - A_0 \mid X, Z = z] \\
&= \sum_{z=0,1} (2z-1) E[A_1(z) - A_0(z) \mid X, Z = z] \\
&= E[A_1(1) - A_1(0) - A_0(1) + A_0(0) \mid X].
\end{aligned}$$

Combining the above derivations, we obtain that $\delta_{Y,1}(X)-\delta_{Y,0}(X) = (\delta_{A,1}(X)-\delta_{A,0}(X))\tau(X)$. That is, the CATE is identified by

$$\tau(X) = \frac{\delta_{Y,1}(X) - \delta_{Y,0}(X)}{\delta_{A,1}(X) - \delta_{A,0}(X)}.$$

Alternatively, we consider the following assumptions: (sequential ignorability) $Y_t(a) \perp A_t \mid U, X, Z$ for $t, a = 0, 1$, and there is no additive interaction of either (i) $E[A_1 - A_0 \mid X, U, Z = 1] - E[A_1 - A_0 \mid X, U, Z = 0] = E[A_1 - A_0 \mid X, Z = 1] - E[A_1 - A_0 \mid X, Z = 0]$ or (ii) $E[Y_t(1) - Y_t(0) \mid U, X] = E[Y_t(1) - Y_t(0) \mid X]$ for $t = 0, 1$.

We can continue that

$$\begin{aligned}
&\delta_{Y,1}(X) - \delta_{Y,0}(X) \\
&= \sum_{z=0,1} (2z-1)(E[A_1(z)(Y_1(1) - Y_1(0)) \mid X, Z = z] - E[A_0(z)(Y_0(1) - Y_0(0)) \mid X, Z = z]) \\
&= E_U \sum_{z=0,1} (2z-1)(E[A_1(z)(Y_1(1) - Y_1(0)) \mid X, U, Z = z] - E[A_0(z)(Y_0(1) - Y_0(0)) \mid X, U, Z = z] \\
&= E_U \sum_{z=0,1} (2z-1)(E[A_1(z) \mid X, U, Z = z] E[Y_1(1) - Y_1(0) \mid X, U, Z = z] \\
&\qquad\qquad - E[A_0(z) \mid X, U, Z = z] E[Y_0(1) - Y_0(0) \mid X, U, Z = z]) \\
&= E_U[E[Y_t(1) - Y_t(0) \mid U, X](E[A_1 - A_0 \mid X, U, Z = 1] - E[A_1 - A_0 \mid X, U, Z = 0])].
\end{aligned}$$

Under Assumption (i), we have that

$$\begin{aligned}
&E[A_1 - A_0 \mid X, U, Z = 1] - E[A_1 - A_0 \mid X, U, Z = 0] \\
&= E[A_1 - A_0 \mid X, Z = 1] - E[A_1 - A_0 \mid X, Z = 0] \\
&= \delta_{A,1}(X) - \delta_{A,0}(X);
\end{aligned}$$

or under Assumption (ii), we have that

$$E[Y_t(1) - Y_t(0) \mid U, X] = E[Y_t(1) - Y_t(0) \mid X], t = 0, 1,$$

and also

$$E_U[E[A_1 - A_0 \mid X, U, Z = 1] - E[A_1 - A_0 \mid X, U, Z = 0]] = \delta_{A,1}(X) - \delta_{A,0}(X).$$

Hence combining the above derivations, we obtain the same identification results.

Next, we derive the efficient influence function.

For a given distribution P in the nonparametric statistical model $\mathcal{M}$, we let p denote the density of P with respect to some dominating measure ν. For all bounded $h \in L_2(P)$, define the parametric submodel $p_\epsilon = (1 + \epsilon h)p$, which is valid for small enough ϵ and has score h at $\epsilon = 0$.

We study the following statistical functional

$$\Psi(P) = E_P\left[\frac{E_P[Y_1 - Y_0 \mid X = x, Z = 1] - E_P[Y_1 - Y_0 \mid X = x, Z = 0]}{E_P[A_1 - A_0 \mid X = x, Z = 1] - E_P[A_1 - A_0 \mid X = x, Z = 0]}\right],$$

and would establish that $\Psi(P)$ is pathwise differentiable with respect to $\mathcal{M}$ at P with efficient influence function ϕ_P if we have that for any $P \in \mathcal{M}$

$$\frac{\partial}{\partial \epsilon}\Psi(P_\epsilon)\bigg|_{\epsilon=0} = \int \phi_P(o)h(o)dP(o).$$

We denote $\delta_{Y,z,\epsilon}(x) = E_{P_\epsilon}[Y_1 - Y_0 \mid X = x, Z = z]$, $\delta_{A,z,\epsilon}(x) = E_{P_\epsilon}[A_1 - A_0 \mid X = x, Z = z]$, $S = \partial \log p_\epsilon / \partial \epsilon$, and compute

$$\begin{aligned}
\frac{\partial}{\partial \epsilon}\Psi(P_\epsilon)\bigg|_{\epsilon=0} &= \frac{\partial}{\partial \epsilon}E_{P_\epsilon}\left[\frac{\delta_{Y,1,\epsilon}(X) - \delta_{Y,0,\epsilon}(X)}{\delta_{A,1,\epsilon}(X) - \delta_{A,0,\epsilon}(X)}\right]\bigg|_{\epsilon=0} \\
&= \frac{\partial}{\partial \epsilon}E_P\left[(1 + \epsilon S)\frac{\delta_{Y,1,\epsilon}(X) - \delta_{Y,0,\epsilon}(X)}{\delta_{A,1,\epsilon}(X) - \delta_{A,0,\epsilon}(X)}\right]\bigg|_{\epsilon=0} \\
&= E_P\left[S\frac{\delta_{Y,1}(X) - \delta_{Y,0}(X)}{\delta_{A,1}(X) - \delta_{A,0}(X)}\right] \\
&\quad + E_P\left[\frac{1}{\delta_{A,1}(X) - \delta_{A,0}(X)}\left(\frac{\partial}{\partial \epsilon}\delta_{Y,1,\epsilon}(X)\bigg|_{\epsilon=0} - \frac{\partial}{\partial \epsilon}\delta_{Y,0,\epsilon}(X)\bigg|_{\epsilon=0}\right)\right] \\
&\quad - E_P\left[\frac{\delta_{Y,1}(X) - \delta_{Y,0}(X)}{\{\delta_{A,1}(X) - \delta_{A,0}(X)\}^2}\left(\frac{\partial}{\partial \epsilon}\delta_{A,1,\epsilon}(X)\bigg|_{\epsilon=0} - \frac{\partial}{\partial \epsilon}\delta_{A,0,\epsilon}(X)\bigg|_{\epsilon=0}\right)\right].
\end{aligned}$$

Then we need to compute

$$\begin{aligned}
\frac{\partial}{\partial \epsilon}\delta_{Y,z,\epsilon}(X)\Big|_{\epsilon=0} &= \frac{\partial}{\partial \epsilon}E_{P_\epsilon}[Y_1 - Y_0 \mid X, Z=z]\Big|_{\epsilon=0} \\
&= \frac{\partial}{\partial \epsilon}\frac{\delta_{Y,z}(X) + \epsilon E_P[S(Y_1 - Y_0) \mid X, Z=z]}{1 + \epsilon E_P[S \mid X, Z=z]}\Big|_{\epsilon=0} \\
&= E_P[S(Y_1 - Y_0) \mid X, Z=z] - \delta_{Y,z}(X)E_P[S \mid X, Z=z] \\
&= E_P\left[S\frac{(Y_1 - Y_0 - \delta_{Y,z}(X))I\{Z=z\}}{z\pi_Z(X) + (1-z)(1-\pi_Z(X))} \mid X\right],
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial \epsilon}\delta_{A,z,\epsilon}(X)\Big|_{\epsilon=0} &= \frac{\partial}{\partial \epsilon}E_{P_\epsilon}[A_1 - A_0 \mid X, Z=z]\Big|_{\epsilon=0} \\
&= \frac{\partial}{\partial \epsilon}\frac{\delta_{A,z}(X) + \epsilon E_P[S(A_1 - A_0) \mid X, Z=z]}{1 + \epsilon E_P[S \mid X, Z=z]}\Big|_{\epsilon=0} \\
&= E_P[S(A_1 - A_0) \mid X, Z=z] - \delta_{A,z}(X)E_P[S \mid X, Z=z] \\
&= E_P\left[S\frac{(A_1 - A_0 - \delta_{A,z}(X))I\{Z=z\}}{z\pi_Z(X) + (1-z)(1-\pi_Z(X))} \mid X\right].
\end{aligned}$$

In summary, we obtain the efficient influence function

$$\begin{aligned}
\phi_P(O) = &\frac{E[Y_1 - Y_0 \mid X, Z=1] - E[Y_1 - Y_0 \mid X, Z=0]}{E[A_1 - A_0 \mid X, Z=1] - E[A_1 - A_0 \mid X, Z=0]} \\
&+ \frac{Z - \pi_Z(X)}{\pi_Z(X)(1-\pi_Z(X))(\delta_{A,1}(X) - \delta_{A,0}(X))^2}\{(Y_1 - Y_0)(\delta_{A,1}(X) - \delta_{A,0}(X)) \\
&-(A_1 - A_0)(\delta_{Y,1}(X) - \delta_{Y,0}(X)) + \delta_{Y,1}(X)\delta_{A,0}(X) - \delta_{Y,0}(X)\delta_{A,1}(X)\} - \Psi(P).
\end{aligned}$$

Finally, it follows to prove Theorem 8 by Equation (1).

K Additional simulations

In this section, we report additional simulation results to illustrate how different sample sizes and the strength of the IV affect the performance of the estimated policies.

K.1 Sensitivity analysis

In this section, we study how the strength of the IV affects the performance of the estimated policies. The data generation process is the same as Section 6, except that the treatment assignment mechanism is given by

$$Pr(A_0 = 1 \mid Z, U, X) = \text{expit}(1.5 - 3Z + 0.2U_0 + 2X_1),$$
$$Pr(A_1 = 1 \mid Z, U, X) = \text{expit}(-1.5 + 2Z - 0.15U_1 + 1.5X_2),$$

for weak IV strength, and

$$Pr(A_0 = 1 \mid Z, U, X) = \text{expit}(3 - 7Z + 0.2U_0 + 2X_1),$$
$$Pr(A_1 = 1 \mid Z, U, X) = \text{expit}(-3 + 5Z - 0.15U_1 + 1.5X_2),$$

for strong IV strength, respectively. Simulation results are reported in Figures 4 and 5.

K.2 Sample size

In this section, we study how different sample sizes affect the performance of the estimated policies. The data generation process is the same as Section 6. The sample sizes are $n = 2500, 10000$ when using parametric models, and $n = 5000, 20000$ when using machine learning. Simulation results are reported in Figures 6 and 7.

L Australian Longitudinal Survey

In this section, we provide supplementary information on our data analysis of the Australian Longitudinal Survey. The data can be accessed by making a request to the Australian Data Archive (Australian National University).

We follow Su et al. [2013], Cai et al. [2006] and use an index of labor market attitudes as the instrumental variable in our analysis. The survey includes seven questions about work, social roles and school attitudes towards working women. Individuals respond to these questions with scores (1) strongly agree, (2) agree, (3) don't know, (4) disagree, and (5) strongly disagree. This survey

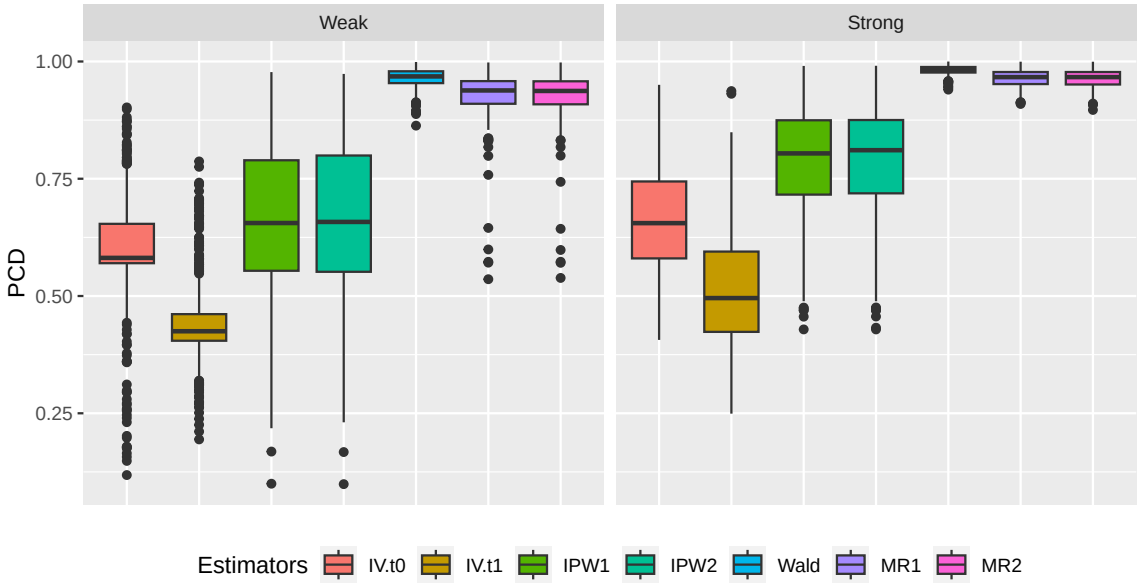

Figure 4: The percentage of correct decisions (PCD) results of the estimated optimal policies using parametric models, under weak (left) or strong (right) IV strength.

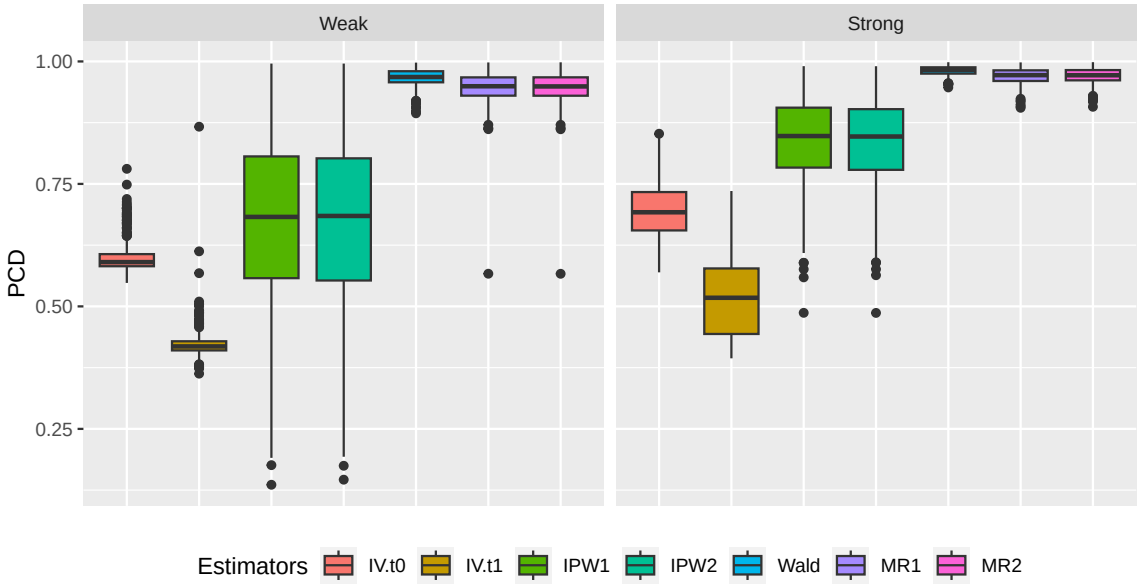

Figure 5: The percentage of correct decisions (PCD) results of the estimated optimal policies using machine learning, under weak (left) or strong (right) IV strength.

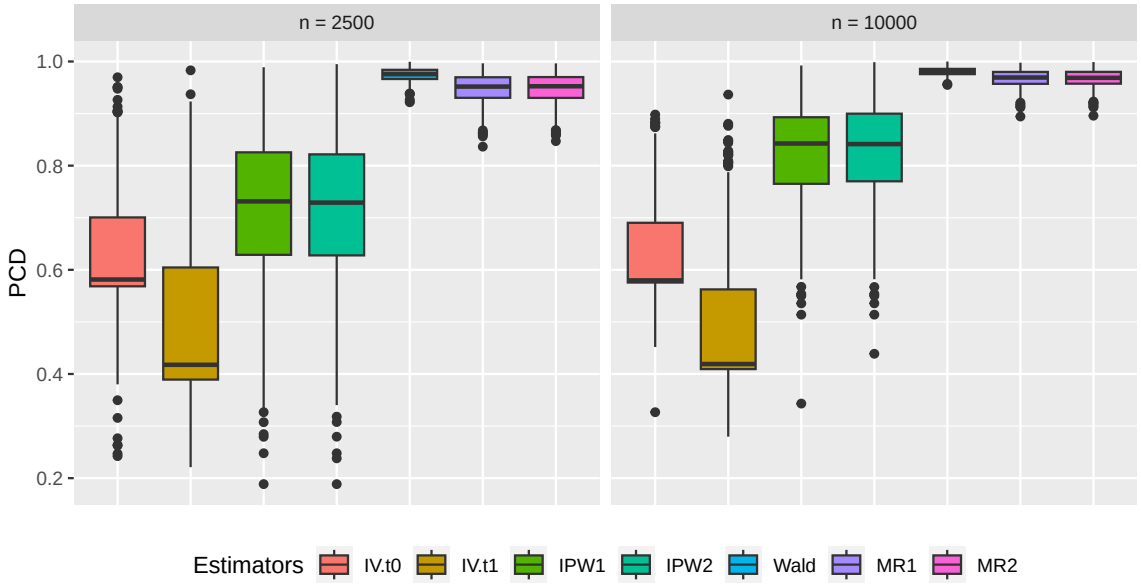

Figure 6: The percentage of correct decisions (PCD) results of the estimated optimal policies, using parametric models with sample size $n = 2500$ (left) or $n = 10000$ (right).

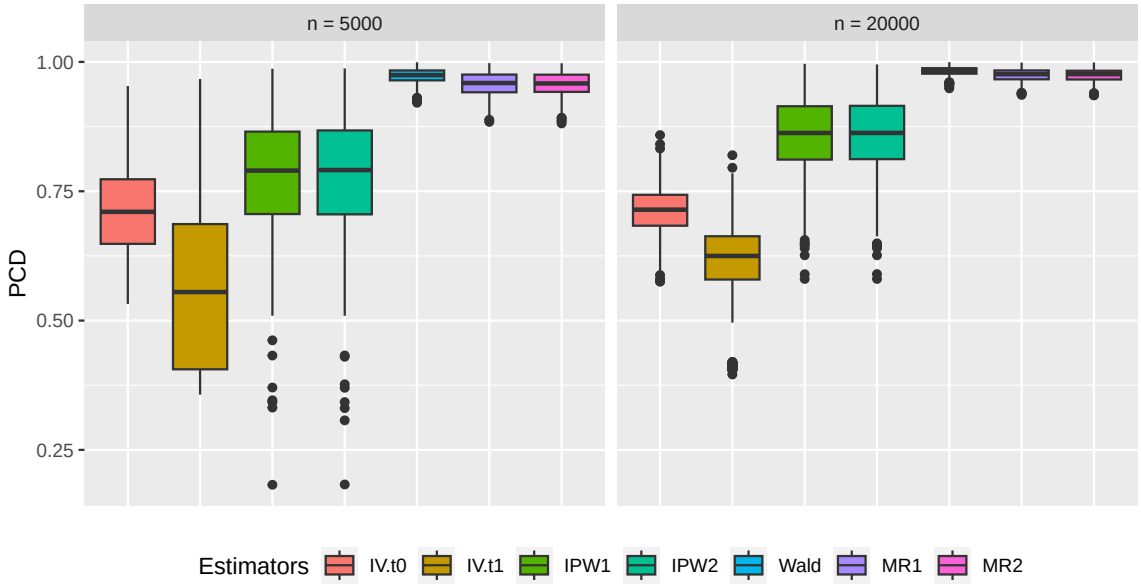

Figure 7: The percentage of correct decisions (PCD) results of the estimated optimal policies, using machine learning with sample size $n = 5000$ (left) or $n = 20000$ (right).

design implies that a response with a higher score indicates more positive attitude towards the education benefit of women and also their active role in the labor market. Following Su et al. [2013], we use only six out of the seven questions to construct our attitudes index, since questions 2 and 3 are actually very similar, thus might be repetitive. We choose question 2 over question 3. Summary statistics of our data from the 1984 and 1985 waves are reported in Table 2 and 3, respectively. Replication code is available at GitHub.

Variable	Source	Mean	SD	Min	Max
born_australia	A12	0.82	0.38	0	1
married	A9	0.07	0.25	0	1
uni_mem	G10	0.34	0.48	0	1
gov_emp	G9	0.21	0.41	0	1
age	A4	20.07	2.45	14	26
year_expe	F3-4, F7-10, F31-33, G21-23	0.94	1.40	0	11
attitude	O1-7	17.94	3.48	6	28
year_edu	E4, E7, E10, E14, E16, E23, E25	11.14	1.93	3	20
wage_hour	G3-5, G7-8	4.83	2.01	0.57	21.43

Table 2: The 1984 wave summary statistics of variables `born_australia`: whether a person is born in Australia; `married`: marital status; `uni_mem`: union membership; `gov_emp`: government employment; `age`: age; `year_expe`: work experience; `attitude`: index of labor market attitudes; `year_edu`: education levels; `wage_hour`: hourly wage. Source indicates which questions in the survey provide the information.

Variable	Source	Mean	SD	Min	Max
born_australia	B3	0.84	0.36	0	1
married	A7	0.15	0.36	0	1
uni_mem	G11	0.38	0.49	0	1
gov_emp	G10	0.22	0.42	0	1
age	A4	20.22	2.87	15	26
year_expe	F3-4, F7-10, F31-33, F23-25	1.82	2.13	0	16
attitude	O1-7	18.75	3.49	6	30
year_edu	E3, E5, E8, E12, E14, E21, E23	11.69	2.11	2	20
wage_hour	G3-5, G7-8	7.48	2.94	0.375	75.00

Table 3: The 1985 wave summary statistics of variables `born_australia`: whether a person is born in Australia; `married`: marital status; `uni_mem`: union membership; `gov_emp`: government employment; `age`: age; `year_expe`: work experience; `attitude`: index of labor market attitudes; `year_edu`: education levels; `wage_hour`: hourly wage. Source indicates which questions in the survey provide the information.