Proximal Learning for Individualized Treatment Regimes Under Unmeasured Confounding

Zhengling Qi\textsuperscript{a,}\textsuperscript{*}, Rui Miao\textsuperscript{b,}\textsuperscript{*}, and Xiaoke Zhang\textsuperscript{c}

\textsuperscript{a}Department of Decision Sciences, The George Washington University, Washington, DC; \textsuperscript{b}Department of Statistics, University of California, Irvine, Irvine, CA; \textsuperscript{c}Department of Statistics, The George Washington University, Washington, DC

\textbf{ABSTRACT}

Data-driven individualized decision making has recently received increasing research interest. However, most existing methods rely on the assumption of no unmeasured confounding, which cannot be ensured in practice especially in observational studies. Motivated by the recently proposed proximal causal inference, we develop several proximal learning methods to estimate optimal individualized treatment regimes (ITRs) in the presence of unmeasured confounding. Explicitly, in terms of two types of proxy variables, we are able to establish several identification results for different classes of ITRs respectively, exhibiting the tradeoff between the risk of making untestable assumptions and the potential improvement of the value function in decision making. Based on these identification results, we propose several classification-based approaches to finding a variety of restricted in-class optimal ITRs and establish their theoretical properties. The appealing numerical performance of our proposed methods is demonstrated via extensive simulation experiments and real data application. Supplementary materials for this article are available online.

\section{1. Introduction}

In recent years, there is a surge of interest in studying data-driven individualized decision making in various scientific fields. For example, in precision medicine, clinicians leverage biomedical data to discover the best personalized treatments for heterogeneous patients (e.g., Rashid et al. 2020). In mobile health, due to recent advances in smart devices and sensing technology, real time information can be collected and used to learn the most effective interventions for patients to promote healthy behaviors (e.g., Klasnja et al. 2015). In robotics, tremendous amounts of simulated data are generated to train robots for making optimal decisions to complete human tasks (e.g., Kober, Bagnell, and Peters 2013). In operations management, learning the optimal resource allocation based on current conditions, logistics and costs, etc, is necessary to improve the efficiency of operations (e.g., Seong, Mohseni, and Cioffi 2006). Apparently a common goal of the aforementioned applications is to find an optimal individualized treatment regime (ITR) that can optimize the utility of each instance.

Recently, many statistical learning methods have been developed for learning the optimal ITR. For example, Qian and Murphy (2011) proposed to learn the optimal ITR by first fitting a high-dimensional regression model for the so-called Q-function, which is the conditional expectation of the outcome given the treatment and covariates (Watkins and Dayan 1992), and then assigning the optimal treatment to each individual corresponding to the largest Q-function value. In the binary treatment setting, this method is equivalent to estimating the conditional average treatment effect. Methods of such type are usually referred to as model-based methods (e.g., Zhao, Kosorok, and Zeng 2009; Shi et al. 2018). Alternatively, one may obtain the optimal ITR by directly maximizing the value function, as defined in (1). In the literature, such method is called a direct method. For example, Dudik, Langford, and Li (2011) and Zhao et al. (2012) applied inverse probability weighting (IPW) to estimate the value function for each ITR, and then leveraged modern classification techniques to learn the optimal one. Along this line of research, various extensions have been proposed, such as Zhao et al. (2014) for censored outcomes, Chen, Zeng, and Kosorok (2016) for ordinal outcomes, and Wang et al. (2018) for quantile ITRs. To alleviate potential model misspecifications of the Q-function or the propensity score, augmented IPW has been used so that the value function estimator enjoys the doubly robust property (e.g., Zhang et al. 2012). Recently, borrowed from semiparametric statistics (Bickel 1982), cross-fitting techniques have been incorporated in ITR learning (Athey and Wager 2021; Zhao et al. 2019) so that flexible black-box machine learning methods can be used for estimating the Q-function and the propensity score without sacrificing the efficiency of the resulting estimated ITR. Finally, a review of various ITR learning methods can be found in Kosorok and Laber (2019) and references therein.

Most existing methods for learning the optimal ITR rely on the unconfoundedness assumption so that the value function
can be identified nonparametrically using the observed data. However, it is difficult, or even impossible, to verify this assumption in practice, especially in observational studies or randomized trials with noncompliance. Therefore, to remove confounding effects and thus identify optimal ITRs, practitioners often collect and adjust for as many variables as possible. While this might be the best approach in practice, it is often very costly and sometimes unethical. To address this problem, instrumental variables (IVs) have been used in the literature to find an optimal ITR in the presence of unmeasured confounding. For example, motivated by Wang and Tchetgen Tchetgen (2018), Cui and Tchetgen Tchetgen (2020), and Qiu et al. (2020) independently established similar identification results on the value function using an IV and proposed different optimal ITR learning methods, one for deterministic ITRs and the other for stochastic ITRs. While these two methods are particularly useful in randomized trials with noncompliance, their restrictions on the setting of binary treatments and IVs are very restrictive, which may limit their applicability. Recently, instead of aiming to exactly identify the value function under unmeasured confounding, Han (2019) and Pu and Zhang (2020) considered partial identification in terms of an IV to provide robustness in estimating the optimal ITR. In addition, Kallus and Zhou (2018) leveraged a sensitivity analysis in causal inference where the value function is partially identified and developed a confounding-robust policy improvement method. Although partial identification can still lead to valuable ITRs to policy makers, they are likely to be sub-optimal.

In this article, we propose an alternative remedy to estimate optimal ITRs under endogeneity. Our approach is built upon proximal causal inference recently developed by Miao and Tchetgen Tchetgen (2018) and Tchetgen Tchetgen et al. (2020). The salient idea behind proximal causal inference is to identify the causal effect under unmeasured confounding via either treatment-inducing or outcome-inducing confounding proxies, which connects existing identification results on the causal effect based on IVs and negative controls. The applicability of proximal causal inference is very promising since the existence of such proxies is common in many applications. See Miao and Tchetgen Tchetgen (2018) and Tchetgen Tchetgen et al. (2020) for examples. Moreover, in contrast with the aforementioned IV-based ITR learning approaches, there is no restriction on the data type of these proxy variables. Due to these merits, in this article, we adapt the idea of proximal causal inference to establish identification results for various classes of ITRs under unmeasured confounding and accordingly propose several classification-based methods to estimate corresponding in-class optimal ITRs.

The contribution of this article can be mainly summarized into three folds. First, we establish several new identification results for various classes of ITRs under unmeasured confounding in terms of treatment-inducing and/or outcome-inducing confounding proxies. All these results can show an interesting tradeoff between the risk of making untestable assumptions and the potential gain of value function in decision making. Note that since covariates are involved in ITRs, our identification results are focused on conditional treatment effect identification, which are different from those for average treatment effect estimation in the proximal causal inference literature. Second, based on these identification results, we propose several classification-based methods to estimate optimal ITRs. For nuisance functions involved in these methods, which are characterized by conditional moment restrictions, we apply the minimax learning approach in Dikkala et al. (2020) to estimate them nonparametrically. Similar approaches to estimating nuisance functions have also been used in Kallus, Mao, and Uehara (2021) and Ghassami et al. (2021), which were posted on arxiv.org very recently, and are independent works from this article. Under one specific setting where the value function can be identified via either a treatment-inducing confounding bridge function or an outcome-inducing confounding bridge function (see Section 3), motivated by Cui et al. (2020), we develop a doubly robust optimal ITR learning method with cross-fitting (see Section 4.3). Third, we establish a theoretical guarantee for the proposed doubly robust proximal learning method. Specifically, we provide a finite sample bound for the value function difference between the optimal ITR and our estimated in-class optimal ITR. The bound can be decomposed into four components: an irreducible error due to unmeasured confounding, an approximation error due to the restricted treatment regime class, and two estimation errors caused by the finite sample, illustrating different sources of errors in finding the optimal ITR under unmeasured confounding. Theoretical results for other proposed methods can be similarly derived.

The rest of our article is organized as follows. In Section 2, we briefly introduce the framework of learning optimal ITRs without unmeasured confounding. In Section 3, adapting the idea from the proximal causal inference, we establish nonparametric value function identification results for various classes of ITRs under unmeasured confounding. In Section 4, we develop several corresponding proximal learning methods based on the identification conditions established in Section 3. Theoretical guarantees of the proposed methods are presented in Section 5. In Sections 6 and 7, respectively, we demonstrate the superior performance of our methods via extensive simulation studies and one real data application. Discussions and future research directions in Section 8 conclude the article.

2. Optimal ITR without Unmeasured Confounding

In this section, we give a brief introduction to optimal ITR learning under no unmeasured confounding. Let \( A \) be a binary treatment which takes values in the space \( A = \{-1, 1\} \). Let \( Y(1) \) and \( Y(-1) \) be the potential outcomes when \( A = 1 \) and \( A = -1 \) respectively, but in practice \( Y(1) \) and \( Y(-1) \) are not both observable. Under the consistency assumption that \( Y = Y(A) \), we can write \( Y = Y(1)\mathbb{I}(A = 1) + Y(-1)\mathbb{I}(A = -1) \), where \( \mathbb{I}(\cdot) \) denotes the indicator function. Moreover, let \( X \) be the observed \( p \)-dimensional covariate that belongs to a covariate space \( X \subseteq \mathbb{R}^p \). Without loss of generality, we assume that a large outcome \( Y \) is always preferred.

For an ITR \( d \), which is a measurable function mapping from the covariate space \( X \) into the treatment space \( A \), the potential outcome under \( d \) is defined by \( Y(d) = Y(1)\mathbb{I}(d(X) = 1) + Y(-1)\mathbb{I}(d(X) = -1) \), and then the value function of \( d \) (Manski 2004; Qian and Murphy 2011) can be defined as

\[
V(d) = \mathbb{E} \{Y(d)\}.
\]
Under the following three standard assumptions in the potential outcome framework (Robins 1986): (i) Unconfoundedness: \( \{Y(1), Y(-1)\} \perp \perp A \mid X \) where \( \perp \perp \) represents independence, (ii) Positivity: \( \Pr(A = a(X)) > 0 \) for every \( a \in A \) almost surely, and (iii) Consistency: \( Y = Y(A) \), we can nonparametrically identify the value function \( V(d) \) using \( (X, A, Y) \) via

\[
V(d) = \mathbb{E}\left\{ \frac{Y \cdot I(A = d(X))}{\Pr(A = d(X))} \right\}.
\]

Then by maximizing \( V(d) \) in (2) over \( D \), the class of all ITRs, the global optimal ITR is

\[
d^*(X) = \text{sign}\left( \mathbb{E}(Y|X, A = 1) - \mathbb{E}(Y|X, A = -1) \right),
\]

almost surely. See Qian and Murphy (2011) and Zhao et al. (2012) for more details. Note that the optimal ITR remains same if we use other coding schemes, for example, 1/0, for the binary \( A \). Moreover, the optimal ITR learning method under no unmeasured confounding above is also applicable to multiple treatments and so are our proposed methods below.

3. Optimal ITR with Unmeasured Confounding

The optimal ITR learning without unmeasured confounding in Section 2 relies on the unconfoundedness assumption, that is, \( \{Y(1), Y(-1)\} \perp \perp A \mid X \). In practice, however, one typically cannot ensure the unconfoundedness assumption to hold. If there exist unmeasured confounders \( U \) that affect both the treatment \( A \) and the outcome \( Y \), unless we make proper assumptions, we are unable to identify the value function based on the observed data \( (X, A, Y) \) as in (2) or further to find optimal ITRs (Pearl 2009). Inspired by proximal casual inference that was recently proposed by Miao and Tchetgen Tchetgen (2018) and Tchetgen Tchetgen et al. (2020), we propose to adapt its idea to optimal ITR learning in the presence of unmeasured confounding.

Following Tchetgen Tchetgen et al. (2020), suppose that we can decompose \( X \) into three types \( X = (L, W, Z) \), where \( L \) are observable covariates that affect both \( A \) and \( Y \), and \( W \) are outcome-inducing confounding proxies that are only related to \( A \) through \( (L, U) \), and \( Z \) are treatment-inducing confounding proxies that are only related to \( Y \) through \( (L, U) \). The terminologies we adopt here for \( Z \) and \( W \) follow those in Tchetgen Tchetgen et al. (2020), but they may also be called negative control exposures and negative control outcomes respectively in other literature (e.g., Miao and Tchetgen Tchetgen 2018). Denote the spaces \( L, U, W, \) and \( Z \) belong to as \( L, U, W, \) and \( Z \), respectively. Figure 1 illustrates some of their relationships with \( A \) and \( Y \). In general, such decomposition does not guarantee the identifiability of the causal effect. Figure 2 shows an example where \( L, W, Z \) and

\[
U = (U_1, U_2, U_3) \text{ coexist, but the path } A \cdot U_2 \cdot U_3 \cdot U_1 \cdot Y \text{ prevents us from identifying the causal effect of } A \text{ on } Y. \text{ However, there exist scenarios where no unmeasured confounding still holds despite the presence of } U, W, \text{ and/or } Z \text{ to make the causal effect of } A \text{ on } Y \text{ identifiable. See Figure 1(b)-(c) and figure 1(d) of Tchetgen Tchetgen et al. (2020) for examples. As a result, given the decomposition } X = (L, W, Z), \text{ it is possible to relax the no unmeasured confounding assumption and still identify the causal effect of } A \text{ on } Y.

\]
links the potential outcome with the observed outcome while Assumption 2 states that each treatment has a positive probability of being assigned. Assumption 3 essentially states that $A$ and $Z$ do not have a causal effect on $Y$ except intervening on $A$. In many practical applications, Assumptions 3 and 4 can hold with the help of domain experts to identify valid $W$ and $Z$. See our real data application and also the air pollution example in Miao and Tchetgen Tchetgen (2018). Assumption 5 is also standard in the literature. It indicates that by adjusting for $(U, L)$, one is able to jointly identify the causal effect of $(Z, A)$ on $Y$ and $W$, which holds in principle as $U$ is not observed. A directed acyclic graph (DAG) of Assumptions 3–5 is depicted in Figure 3.

Note that if $D$ is now defined as the class of all ITRs mapping from $(L, U)$ to $A$, then the global optimal ITR within $D$ becomes $d^*(L, U) = \text{sign} \left( \mathbb{E}(Y|L, U, A = 1) - \mathbb{E}(Y|L, U, A = -1) \right)$, almost surely, but it is unattainable in practice since $U$ is unobservable.

Under Assumptions 1–5 and a few different sets of additional assumptions to be listed below, we are able to establish several value function identifications under unmeasured confounding.

### 3.2. Optimal ITR via Outcome Confounding Bridge

Our first value function identification requires the following technical assumptions, which were used by Miao, Geng, and Tchetgen Tchetgen (2018) to identify the population average treatment effect under unmeasured confounding

**Assumption 6 (Completeness).** (a) For any $a \in A, l \in \mathcal{L}$ and measurable function $g$ defined on $U$, if $\mathbb{E}(g(U)|Z, A = a, L = l) = 0$ almost surely, then $g(U) = 0$ almost surely.

(b) For any $a \in A, l \in \mathcal{L}$ and measurable function $g$ defined on $Z$, if $\mathbb{E}(g(Z)|W, A = a, L = l) = 0$ almost surely, then $g(Z) = 0$ almost surely.

**Assumption 7 (Outcome confounding bridge).** There exists an outcome confounding bridge function $h_0$ defined on $(W, A, \mathcal{L})$ such that

$$
\mathbb{E}(Y|Z, A, L) = \mathbb{E}(h_0(W, A, L)|Z, A, L),
$$

almost surely.

The completeness assumption is commonly seen in mathematical statistics and can be satisfied by many parametric or semiparametric models such as those for exponential families (Newey and Powell 2003). For more examples including non-parametric models, we refer readers to D’Haultfoeuille (2011) and Chen et al. (2014). Assumption 6 (a) essentially requires that the variability of $U$ can be accounted for by $Z$ and Assumption 6 (b) can be similarly interpreted. Unlike the requirement by Cui and Tchetgen Tchetgen (2020) and Qiu et al. (2020) that the IV must be binary, Assumption 6 does not require any specific type of $Z$, which is appealing. Assumption 7 basically states that there exists a solution to (3), which is called a linear integral equation of the first kind (Kress, Maz’ya, and Kozlov 1989). Assumption 6 (b), together with some regularity conditions given by Miao, Geng, and Tchetgen Tchetgen (2018), can ensure the existence of $h_0$ satisfying (3). For more details and practical examples where these assumptions are satisfied, we refer readers to Miao, Geng, and Tchetgen Tchetgen (2018), Shi et al. (2020), and Miao and Tchetgen Tchetgen (2018).

Based on the outcome confounding bridge $h_0$, we develop a nonparametric identification result for the value function of each ITR, which is similar to that by Miao and Tchetgen Tchetgen (2018) on the average treatment effect.

**Theorem 3.1.** Let $\mathcal{D}_1$ be the class of ITRs that map from $(L, Z)$ to $A$. Under Assumptions 1–7, for any $d_1 \in \mathcal{D}_1$, the value function $V(d_1)$ can be nonparametrically identified by

$$
V(d_1) = \mathbb{E}(h_0(W, d_1(L, Z), L)).
$$

Then the restricted in-class optimal ITR within $\mathcal{D}_1$, defined as $d^*_1 \in \text{arg max}_{d_1 \in \mathcal{D}_1} V(d_1)$, can be almost surely identified by

$$
d^*_1(L, Z) = \text{sign} \left( \mathbb{E}(h_0(W, 1, L)|L, Z) - \mathbb{E}(h_0(W, -1, L)|L, Z) \right).
$$

The proof of Theorem 3.1 is given in supplementary material S2. Theorem 3.1 indicates that the value function is identifiable over $\mathcal{D}_1$ in the presence of unmeasured confounders. Due to the use of the outcome confounding bridge $h_0$, we can only identify the value function over a restricted class of ITRs $\mathcal{D}_1$ instead of $\mathcal{D}$. That $W$ are not used as decision variables for the optimal ITR is somewhat reasonable since they have already been used as outcome-inducing confounding proxies. Moreover, in practice, $W$ may be collected after a decision is made, which prevents them from being used in decision making.

The form of $d^*_1$ in (5) shows that the optimal ITR in $\mathcal{D}_1$ incorporates the effect of treatment-inducing confounding proxies $Z$ on $Y$ unless $\mathbb{E}(h_0(W, 1, L)|L, Z) - \mathbb{E}(h_0(W, -1, L)|L, Z)$ is independent of $Z$. This is reasonable since $Z$ may contain some useful information of $U$, which can help improve the value function. This is studied by our simulation in Section 6. Obviously $V(d^*_1) \leq V(d^*)$, but due to unmeasured confounding, $V(d^*_1)$ is the best we can obtain within $\mathcal{D}_1$ under the assumptions in Theorem 3.1. To illustrate how to obtain $d^*_1$ via $h_0$, a concrete example is provided in supplementary material S1.

### 3.3. Optimal ITR via Treatment Confounding Bridge

In this section, we provide an alternative identification result for the value function, and thus that for a different restricted optimal ITR, without using the outcome confounding bridge as in Section 3.2. It requires the following assumptions, which are different from Assumptions 6 and 7 and were originally used by Cui et al. (2020) to study semiparametric proximal

![Figure 3. A causal DAG as a representation of Assumptions 3–5.](image-url)
causal inference to identify the average treatment effect under unmeasured confounding.

**Assumption 8 (Completeness).** (a) For any \( a \in \mathcal{A}, l \in \mathcal{L} \) and measurable function \( g \) defined on \( \mathcal{U} \), if \( \mathbb{E}[g(U)|W, A = a, L = l] = 0 \) almost surely, then \( g(U) = 0 \) almost surely.

(b) For any \( a \in \mathcal{A}, l \in \mathcal{L} \) and measurable function \( g \) defined on \( \mathcal{W} \), if \( \mathbb{E}[g(W)|Z, A = a, L = l] = 0 \) almost surely, then \( g(W) = 0 \) almost surely.

**Assumption 9 (Treatment confounding bridge).** For any \( a \in \mathcal{A} \), there exists a treatment confounding bridge function \( q_0 \) defined on \( (Z, \mathcal{A}, \mathcal{L}) \) such that

\[
\frac{1}{\Pr(A = a|W, L)} = \mathbb{E}\left\{ q_0(Z, A, L)|W, A = a, L \right\},
\]

almost surely.

Assumptions 8 and 9 play a similar role of Assumptions 6 and 7. In particular, Assumption 9 establishes a link between \( Z \) and \( A \) and (6) is also a linear integral equation of the first kind. The existence of such \( q_0 \) satisfying (6) can be guaranteed by Assumption 8 (b) combined with some regularity conditions given in Cui et al. (2020). Similar to Theorem 3.1, the value function can be nonparametrically identified, but via the treatment confounding bridge function \( q_0 \) and over a different class of ITRs.

**Theorem 3.2.** Let \( \mathcal{D}_2 \) be the class of ITRs that map from \( (\mathcal{L}, \mathcal{W}) \) to \( \mathcal{A} \). Under Assumptions 1–5, 8(a), and 9, for any \( d_2 \in \mathcal{D}_2 \), the value function \( V(d_2) \) can be nonparametrically identified by

\[
V(d_2) = \mathbb{E}\left\{ Y_q(Z, A, L)|d_2(L, W) = A \right\}.
\]

The restricted in-class optimal ITR within \( \mathcal{D}_2 \), defined as \( d^*_{\mathcal{D}_2}(L, W) = \arg\max_{d_2 \in \mathcal{D}_2} V(d_2) \), can be almost surely identified by

\[
d^*_{\mathcal{D}_2}(L, W) = \text{sign} \left[ \mathbb{E}\left\{ Y_q(A = 1)|q_0(Z, 1, L)|L, W \right\} \right] - \mathbb{E}\left\{ Y_q(A = -1)|q_0(Z, -1, L)|L, W \right\}.
\]

The proof of Theorem 3.2 is given in supplementary material S2. Similar to the discussion after Theorem 3.1, Theorem 3.2 shows that the value function is identifiable over \( \mathcal{D}_2 \) in terms of the treatment confounding bridge \( q_0 \) despite unmeasured confounding. As kindly pointed out by one reviewer, the result in Theorem 3.2 may be useful to identify a restricted optimal ITR when \( W \) should be included as a decision variable but \( Z \) should not. For instance, as in the time series example in Miao and Tchetgen Tchetgen (2018), if there is no feedback effect, the future exposure may serve as \( Z \) and the past outcomes can be \( W \), so it is reasonable to include \( W \) as a decision variable for the treatment regime but not \( Z \).

### 3.4. Optimal ITR via Both Confounding Bridges

In Sections 3.2 and 3.3, we have established the identification results for the value function and its corresponding optimal ITR in terms of the outcome confounding bridge and treatment confounding bridge respectively. A natural question is whether it is possible to obtain a broader identification result if both confounding bridges coexist. The answer is affirmative. Explicitly, if all Assumptions 1–5, 6(a), 7, 8(a), and 9 hold, by Theorems 3.1 and 3.2, clearly \( V(d_4) \) can be identified for any \( d_4 \in \mathcal{D}_1 \cup \mathcal{D}_2 \) by

\[
V(d_4) = \mathbb{E}\left\{ h_0(W, d_4(L, Z), L) \right\} + \mathbb{E}\left\{ h_0(Y_q(Z, A, L)|d_4(L, W) = A) \right\}.
\]

Then the restricted optimal ITR in \( \mathcal{D}_1 \cup \mathcal{D}_2 \) is defined as

\[
\bar{d}^*_4 \in \arg\max_{d_4 \in \mathcal{D}_1 \cup \mathcal{D}_2} V(d_4).
\]

Note that, despite the coexistence of both confounding bridges, we are still unable to identify the value function \( V(d) \) based on observed data over \( d \in \mathcal{D} \), which refers to the class of all ITRs mapping from \( (\mathcal{L}, \mathcal{W}, Z) \) to \( \mathcal{A} \), since \( \mathbb{E}\{Y(a)|L, W, Z\} \) for any \( a \in \mathcal{A} \) is not nonparametrically identifiable due to unmeasured confounding. It is unknown whether there exists a sufficient and necessary condition to identify the conditional average treatment effect given all observed covariates when there exists unmeasured confounding.

To conclude this section, we provide Table 1 that summarizes the identification results we have developed and their required assumptions. More discussions and practical suggestions on these results can be found at the end of Section 4.

### 4. Proximal Policy Learning

In this section, based on the identification results established in Section 3, we propose several methods to estimate restricted in-class optimal ITRs based on observed \( n \) independent and identically distributed samples \( \{(L_i, Z_i, W_i, A_i, Y_i): i = 1, \ldots, n\} \). In Section 4.1 we first propose the estimation methods for the confounding bridge functions \( h_0 \) and \( q_0 \) defined in Assumptions 7 and 9, respectively. In Section 4.2, based on the estimates of \( h_0 \) and \( q_0 \), we propose several classification-based methods to estimate the restricted optimal ITRs \( d^*_1, d^*_2 \), and \( d^*_3 \) defined in Section 3 under their corresponding assumptions respectively. In Section 4.3, under the condition that Assumptions 1–5, 6(a), 7, 8(a), and 9 hold, we propose an augmented inverse probability weighted (AIPW)-type classification-based method for estimating the optimal ITR in a new class of ITRs. For ease of presentation, let Assumptions 6(b) and 8(b) always hold hereafter so that \( h_0 \) and \( q_0 \) can be uniquely identified.

#### 4.1. Estimation of Confounding Bridge Functions

Here we introduce nonparametric estimations of outcome and treatment confounding bridge functions \( h_0 \) and \( q_0 \) defined in Assumptions 7 and 9, respectively.

**Estimating \( h_0 \):** Equation (3) in Assumption 7 is equivalent to

\[
\mathbb{E}\{Y - h_0(W, A, L)|Z, A, L\} = 0,
\]

which is known as the instrumental variable model or conditional moment restriction model and has been well studied in

| Table 1. A summary of optimal ITR identification results. |
|---------------------------------|-----------------|-----------------|
| Assumptions | ITR Class | Restricted Optimal ITR |
|----------------|-----------------|-----------------|
| 1–5, 6(a) and 7 | \( \mathcal{D}_1 \) : \( (\mathcal{L}, Z) \rightarrow \mathcal{A} \) | \( d^*_1 \) in (5) through \( h_0 \) |
| 1–5, 6(a) and 9 | \( \mathcal{D}_2 \) : \( (\mathcal{L}, W) \rightarrow \mathcal{A} \) | \( d^*_2 \) in (8) through \( q_0 \) |
| 1–5, 6(a), 7, 8(a), and 9 | \( \mathcal{D}_1 \cup \mathcal{D}_2 \) | \( d^*_3 \) in (9) through \( q_0 \) or \( h_0 \) |
econometrics and statistics (e.g., Chamberlain 1992; Newey and Powell 2003; Ai and Chen 2003; Blundell, Chen, and Kristensen 2007; Chen 2007; Chen and Pouzo 2012).

Here we adopt the min-max estimation method by Dikkala et al. (2020) to estimate $h_0$ nonparametrically as follows:

$$
\hat{h}_0 = \arg \min_{h \in \mathcal{H}} \sup_{f \in \mathcal{F}} \left[ \frac{1}{n} \sum_{i=1}^{n} \left( Y_i - h(W_i, A_i, L_i) \right) f(Z_i, A_i, L_i) \right. \\
- \left. \lambda_{1,n} \|f\|_{\mathcal{F}}^2 - \|f\|_{2,n}^2 \right] + \lambda_{2,n} \|h\|_{\mathcal{H}}^2, 
$$

where $\lambda_{1,n} > 0$ and $\lambda_{2,n} > 0$ are tuning parameters, $\|f\|_{\mathcal{F}}$ is the empirical $\ell^2$ norm, that is, $\|f\|_{\mathcal{F}}^2 = \sqrt{n^{-1} \sum_{i=1}^{n} f^2(Z_i, A_i, L_i)}$, and $\mathcal{H}$ and $\mathcal{F}$ are some functional classes, for example, reproducing kernel Hilbert spaces (RKHS), with their corresponding norms $\|\cdot\|_{\mathcal{F}}$ and $\|\cdot\|_{\mathcal{H}}$, respectively.

The rationale behind (11) is the following population version of the min-max optimization problem when $\lambda_{1,n}, \lambda_{2,n} \to 0$ as $n \to \infty$:

$$
\min_{h \in \mathcal{H}} \sup_{f \in \mathcal{F}} \left( \mathbb{E} \left[ (Y - h(W, A, L)) f(Z, A, L) \right] - \mathbb{E} \left[ f^2(Z, A, L) \right] \right). 
$$

(12)

If $2^{-1} \mathbb{E} \left[ h_0(W, A, L) - h(W, A, L) \right] \in \mathcal{F}$ for every $h \in \mathcal{H}$, then the optimization (12) above is equivalent to

$$
\min_{h \in \mathcal{H}} \mathbb{E} \left[ (Y - h(W, A, L)) [Z, A, L] \right]^2. 
$$

If we further assume $h_0 \in \mathcal{H}$, then $h_0$ is the unique global minimizer of (12). Hence, the min-max formulation (11) is valid.

Estimating $q_0$: Equation (6) in Assumption 9 indicates that for every $a \in \mathcal{A}$,

$$
\mathbb{E} \left[ (A = a) q_0(Z, A, L) - 1 \right] [W, L] = 0.
$$

Similar to (11), we estimate $q_0$ by the following min-max optimization: For each $a \in \mathcal{A}$,

$$
\hat{q}_0(a, \bullet, \bullet) = \arg \min_{q \in \mathcal{Q}} \sup_{g \in \mathcal{G}} \left[ \frac{1}{n} \sum_{i=1}^{n} \left[ (A_i = a) q(Z_i, a, L_i) - 1 \right] g(W_i, L_i) \right. \\
- \left. \mu_{1,n} \|q\|_{\mathcal{G}}^2 - \|g\|_{2,n}^2 + \mu_{2,n} \|q\|_{\mathcal{Q}}^2 \right], 
$$

(13)

where $\mu_{1,n} > 0$ and $\mu_{2,n} > 0$ are tuning parameters, and $\mathcal{Q}$ and $\mathcal{G}$ are functional classes with their corresponding norms $\|\cdot\|_{\mathcal{Q}}$ and $\|\cdot\|_{\mathcal{G}}$, respectively.

Generally one may use any functional class for $\mathcal{H}$ and $\mathcal{F}$ in (11) and for $\mathcal{Q}$ and $\mathcal{G}$ in (13). If they are all specified as RKHS, due to the represent theorem, both $\hat{h}_0$ and $\hat{q}_0$ will have finite-dimensional representations which lead to fast computations. See supplementary material S3 for details.

4.2. Outcome and Treatment Proximal Learning

With estimated $h_0$ and $q_0$, we propose to estimate the restricted optimal ITRs $d_1^*$, $d_2^*$, and $d_4^*$ defined in Section 3 using classification-based methods, which are similar to those of Zhao et al. (2012) and Zhang et al. (2012) under no unmeasured confounding.

**Outcome Proximal Learning of $d_1^*$**: According to (5), under the assumptions in Theorem 3.1, finding $d_1^*$ is equivalent to minimizing the following classification error

$$
\mathbb{E} \left[ [h_0(W, 1, L) - h_0(W, -1, L)] \right] (d_1(L, Z) \neq 1),
$$

over all $d_1 \in D_1$. Since each $d_1 \in D_1$ can be written as $d_1(L, Z) = \text{sign}(r_1(L, Z))$ for some measurable function $r_1$ defined on $(L, Z)$, we can rewrite $\|(d_1(L, Z) \neq 1) = \|r_1(L, Z) < 0$ where sign(0) = 1. Then the optimization problem above becomes

$$
\min_{r_1} \mathbb{E} \left[ [h_0(W, 1, L) - h_0(W, -1, L)] \right] (r_1(L, Z) < 0). 
$$

Given the observed data and $\hat{h}_0$, we solve its empirical version

$$
\min_{r_1} \frac{1}{n} \sum_{i=1}^{n} \hat{\Delta}(W_i, L_i) \|r_1(L_i, Z_i) < 0\),
$$

where $\hat{\Delta}(W, L) = \hat{h}_0(W, 1, L) - \hat{h}_0(W, -1, L)$, or equivalently

$$
\min_{r_1} \frac{1}{n} \sum_{i=1}^{n} \hat{\Delta}(W_i, L_i) \|\left(\text{sign} \left(\hat{\Delta}(W_i, L_i)\right) r_1(L_i, Z_i) < 0\right). 
$$

(14)

The equivalence above motivates us to use a convex surrogate function to replace the indicator function since all weights $\hat{\Delta}(W_i, L_i)$ are nonnegative. Similar to Zhao et al. (2012) and Zhao et al. (2019), we adopt the hinge loss $\phi(t) = \max(1-t, 0)$ and consider $r_1 \in \mathcal{R}_1$, a pre-specified class of functions defined on $(L, Z)$, for example, a RKHS, to obtain the estimated optimal ITR by $d_1^* = \text{sign}(\hat{r}_1)$ where

$$
\hat{r}_1 = \arg \min_{r_1 \in \mathcal{R}_1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\Delta}(W_i, L_i) \|\phi \left(\text{sign} \left(\hat{\Delta}(W_i, L_i)\right) r_1(L_i, Z_i)\right) + \rho_{1,n} \|r_1\|_{\mathcal{R}_1} \right\}. 
$$

Here $\|\cdot\|_{\mathcal{R}_1}$ is the norm of $\mathcal{R}_1$ and $\rho_{1,n} > 0$ is a tuning parameter. The optimization in (15) is convex and thus can be solved efficiently. See Algorithm 1 of supplementary material S3.

**Treatment Proximal Learning of $d_2^*$**: Similar to learning $d_1^*$, by (8) and with $\hat{q}_0$, we propose to first find $\hat{r}_2$, the solution to the following minimization

$$
\min_{r_2 \in \mathcal{R}_2} \left\{ \frac{1}{n} \sum_{i=1}^{n} Y_i\hat{q}_0(Z_i, A_i, L_i) \|\phi \left(A_i\text{sign} \left(Y_i\hat{q}_0(Z_i, A_i, L_i)\right) r_2(L_i, W_i) + \rho_{2,n} \|r_2\|_{\mathcal{R}_2} \right\}, 
$$

(16)
where \( \| \cdot \|_R \) is the norm of a pre-specified class of functions \( \mathcal{R}_2 \) defined on \((\mathcal{L}, \mathcal{W})\), for example, a RKHS. Then the estimated optimal ITR is obtained by \( \hat{d}^*_3 = \text{sign}(\hat{r}_3) \). See details in Algorithm 4 of supplementary material S3.

Maximum Proximal Learning of \( d^*_3 \): By (9) and Assumptions 1–9, obtaining \( d^*_3 \) is equivalent to solving

\[
\max \left\{ \max_{d_1 \in D_1} V(d_1), \max_{d_2 \in D_2} V(d_2) \right\}.
\]

(17)

Therefore, we combine the learning methods in (15) and (16) and use a cross-validation procedure to find the estimated optimal ITR \( \hat{d}^*_3 \), that is the better decision rule between \( \hat{d}^*_3 \) and \( \hat{d}^*_2 \). See details in Algorithm 5 of supplementary material S3.

4.3. Doubly Robust Proximal Learning

In some applications where \( W \) and \( Z \) are not observable in future decision making, one may be interested in ITRs only based on \( L \). Let \( D_3 \) be the class of all ITRs that map from \( L \) to \( A \) and denote \( d^*_3 \in \arg \max_{d \in D_3} V(d_3) \). In this section, we propose an estimation method for \( d^*_3 \).

Apparently, if all Assumptions 1–9 are satisfied, the value function \( V(d_3) \) can be identified for any \( d_3 \in D_3 \) via either \( h_0 \) or \( q_0 \). This motivates us to develop a doubly robust estimator for \( V(d_3) \) over \( D_3 \), in the sense that the value function estimator is consistent as long as one of \( h_0 \) and \( q_0 \) is modeled correctly. This can provide a protection against potential model misspecifications of \( h_0 \) and/or \( q_0 \). The foundation of our method is the efficient influence function of \( V(d_3) \) given below.

**Theorem 4.1.** Under Assumptions 1–9 and some regularity conditions given in Theorem 3.1 of Cui et al. (2020), the efficient influence function of \( V(d_3) \) is

\[
\Pi(A = d_3(L), q_0(Z, A, L), Y - h_0(W, A, L)) + h_0(W, d_3(L), L) - V(d_3),
\]

(18)

for any given \( d_3 \in D_3 \).

The proof is similar to that of Cui et al. (2020) and thus omitted. Define

\[
C_1(Y, L, W, Z; h_0, q_0) = \Pi(A = 1) q_0(Z, 1, L),
\]

\[
\{Y - h_0(W, 1, L) \} + h_0(W, 1, L), \quad \text{and}
\]

\[
C_{-1}(Y, L, W, Z; h_0, q_0) = \Pi(A = -1) q_0(Z, -1, L),
\]

\[
\{Y - h_0(W, -1, L) \} + h_0(W, -1, L).
\]

(19)

Based on (18), we can estimate \( V(d_3) \) for each \( d_3 \in D_3 \) by

\[
\hat{V}^{DR}(d_3) = \frac{1}{n} \sum_{i=1}^{n} \left\{ C_1(Y_i, L_i, W_i, Z_i; \hat{h}_0, \hat{q}_0) I(d_3(L_i) = 1) + C_{-1}(Y_i, L_i, W_i, Z_i; \hat{h}_0, \hat{q}_0) I(d_3(L_i) = -1) \right\}.
\]

(20)

It can be shown in Proposition 5.1 that \( \hat{V}^{DR}(d_3) \) enjoys the doubly robust property in the sense that \( \hat{V}^{DR}(d_3) \) is a consistent estimator for \( V(d_3) \) for each \( d_3 \in D_3 \) as long as one of \( q_0 \) and \( h_0 \) is modeled correctly, not necessarily both. Following similar arguments in Section 4.2, the estimated optimal ITR we propose is \( \hat{d}^{DR}_3 = \text{sign}(\hat{r}^{DR}_3) \) and \( \hat{r}^{DR}_3 \) is obtained by

\[
\hat{r}^{DR}_3 \in \arg \min_{r \in \mathcal{R}_3} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left\{ C_1(Y_i, L_i, W_i, Z_i; \hat{h}_0, \hat{q}_0) \right\} + \phi \left( \text{sign} \left( C_1(Y_i, L_i, W_i, Z_i; \hat{h}_0, \hat{q}_0) \right) r(L_i) \right) + \phi \left( -\text{sign} \left( C_{-1}(Y_i, L_i, W_i, Z_i; \hat{h}_0, \hat{q}_0) \right) r(L_i) \right) + \rho_3, n \| r \|^2_{\mathcal{R}_3}, \right\}.
\]

(21)

where \( \mathcal{R}_3 \) is a class of functions defined on \( L \) with norm \( \| \cdot \|_{\mathcal{R}_3} \) and \( \rho_3, n > 0 \) is a tuning parameter.

In practice, we can apply the cross-fitting technique (Bickel 1982) to remove the dependence between the nuisance function estimates \( \hat{h}_0 \) and \( \hat{q}_0 \), and the resulting estimated optimal ITR. Thanks to this technique, our proposed learning method does not require restrictive conditions on both nuisance function estimations (e.g., Donkser conditions) to avoid losing efficiency due to nuisance function estimations. See Chernozhukov et al. (2018) and Athey and Wager (2021) for more details.

To implement cross-fitting, we randomly split data into \( K \) folds and apply the following procedure: first use the \( k \)th fold to obtain \( \hat{h}^{(k)}_0 \) and \( \hat{q}^{(k)}_0 \), the estimates of \( h_0 \) and \( q_0 \), respectively, \( k = 1, \ldots, K \); then for each \( k = 1, \ldots, K \), compute the decision function by solving (21) based on \( \hat{h}^{(k)}_0 \) and \( \hat{q}^{(k)}_0 \) using all data except the \( k \)th fold; finally aggregate all \( K \) decision rules to obtain our final estimated optimal ITR. More details of this algorithm can be found in supplementary material S3.

With some abuse of notations, we denote the final decision function by \( \hat{r}^{DR}_3 \) and the corresponding estimated optimal ITR by \( \hat{d}^{DR}_3 = \text{sign}(\hat{r}^{DR}_3) \). To conclude this section, we present Table 2 which summarizes each optimal ITR learning and its corresponding assumptions. Note that compared with Table 1, we need additional Assumptions 6(b) and 8(b) so that \( q_0 \) and \( h_0 \) can be uniquely identified, respectively.

**Remark 1.** Table 2 reveals two important tradeoffs for proximal ITR learning. The first tradeoff is between improving the value function and imposing more untestable assumptions. For example, a comparison between \( \hat{d}^*_3 \) and \( \hat{d}^{DR}_3 \) shows that although we can identify a much larger class of ITRs and thus a potentially higher value function for \( \hat{d}^{DR}_3 \) than \( \hat{d}^*_3 \), learning \( \hat{d}^{DR}_3 \) requires an additional assumption on the existence of a treatment confounding bridge. Conversely, learning \( \hat{d}^*_3 \) needs fewer assumptions than learning \( \hat{d}^{DR}_3 \) and is thus more reliable, but its sub-optimality gap to the globally optimal ITR could be larger compared with \( \hat{d}^*_3 \). The second tradeoff is between the estimation robustness and the value function improvement. A comparison between \( \hat{d}^*_3 \) and \( \hat{d}^{DR}_3 \) shows that to estimate \( \hat{d}^{DR}_3 \) accurately which belongs to a larger class of ITRs and corresponds to a better value function than \( \hat{d}^*_3 \), we require both nuisance functions to be estimated consistently. However, if one is willing to consider \( D_3 \), a smaller
Table 2. A summary of the proposed proximal learning methods for optimal ITRs.

| Assumptions | ITR class | Proximal learning | Estimated optimal ITR |
|-------------|----------|------------------|----------------------|
| 1-5, 6 (a), 7 and 8 (b) | $D_1$ | Outcome proximal learning (15) | $\hat{d}_1$ |
| 1-5, 6 (b), 8 (a), and 9 | $D_2$ | Treatment proximal learning (16) | $\hat{d}_2$ |
| 1-9 | $D_3$ | Doubly robust proximal learning (21) | $\hat{d}_3^{DR}$ |
| 1-9 | $D_1 \cup D_2$ | Maximum proximal learning (17) | $\hat{d}_4$ |

class of ITRs, one can estimate $d_3^{DR}$ consistently as long as one of the two nuisance functions is estimated consistently, but its corresponding value function might be smaller than that of $d_4^*$.

Remark 2. Due to the aforementioned two tradeoffs, we suggest practitioners consider a conservative way of choosing the final optimal ITR estimate. For example, since learning $d_3^*$ or $d_4^*$ requires fewer assumptions than learning the other optimal ITRs, the optimal ITR obtained by either the outcome proximal learning or treatment proximal learning is likely to be more trustworthy than those by the other two methods. For example, to determine if a subgroup of patients can potentially benefit more from a new treatment than from the standard care, one may recommend the new treatment if both of these two optimal ITR estimates agree. In practice, to hopefully achieve conservativeness, robustness and value function improvement simultaneously, one may also recommend treatments for patients when most of the estimated optimal ITRs agree or use the ITR selected by cross-validation. The ensemble and cross-validation approaches have been applied in our real data application. More details are in Section 7 and supplementary material S6.

5. Theoretical Results

In this section, we develop the theoretical properties of our proposed methods, or specifically the finite sample excess risk bound for each estimated optimal ITR. For brevity, we only provide the results for the doubly robust optimal ITR estimator $\hat{d}_3^{DR}$, but similar results can be obtained for the other estimators. We first show the doubly robust property of $\hat{V}^{DR}(d_3)$ in (20) for any $d_3 \in D_3$.

Proposition 5.1. Under Assumptions 1–9, if either $\hat{h}_0$ can consistently estimate $h_0$ in the sup-norm or $\hat{q}_0$ can consistently estimate $q_0$ in the sup-norm, then $\hat{V}^{DR}(d_3)$ is a consistent estimator of $V(d_3)$ for any $d_3 \in D_3$.

The proof is given in supplementary material S2. Next we show Fisher consistency, that is, it is appropriate to replace the indicator function in $\hat{V}^{DR}$ with the hinge loss as in (21) to obtain $r_3^{DR}$ and $\hat{d}_3^{DR} = \text{sign}(r_3^{DR})$ accordingly. Define the hinge loss based $\phi$-risk by

$$ R_\phi(r) = \mathbb{E} \left\{ |C_1| \phi(\text{sign}(C_1) r(L)) + |C_{-1}| \phi(\text{sign}(C_{-1}) r(L)) \right\}, $$

(22)

where $C_1$ and $C_{-1}$ denote $C_1(Y, L, W, Z; h_0, q_0)$ and $C_{-1}(Y, L, W, Z; h_0, q_0)$ defined in (19), respectively for ease of presentation. Let $r^* \in \arg \min_r R_\phi(r)$. Then we have the following proposition.

Proposition 5.2. Under Assumptions 1–9, $d_3^*(L) = \text{sign}(r^*(L))$.

The proof of Proposition 5.2 is omitted since it can be derived by following similar arguments in the proof of Proposition 3.1 of Zhao et al. (2019). Proposition 5.2 essentially states that replacing the indicator function by the hinge loss function does not change the goal of finding the optimal ITR. We can further link the original value function with the $\phi$-risk as follows.

Proposition 5.3. Under Assumptions 1–9, $V(d_3^*) - V(d) \leq R_\phi(r) - R_\phi(r^*)$ for any $d = \text{sign}(r(L))$.

The proof of Proposition 5.3 is omitted due to the similar reason as that of Proposition 5.2. Proposition 5.3 implies that the value function difference between the in-class optimal ITR and any other ITR in $D_3$ can be bounded by their $\phi$-risk difference. Therefore, the convergence rate of the value function of $d_3^{DR}$ can be bounded by the convergence rate of the $\phi$-risk of our estimated decision function $\hat{r}_3^{DR}$. To establish the finite sample excess risk bound for $d_3^{DR}$, we make the following technical assumptions in addition to Assumptions 1–9.

Assumption 10. There exists a constant $C_1 > 0$ such that max $\{ |Y|, \|h_0\|_{\infty}, \|q_0\|_{\infty} \} \leq C_1$.

Assumption 11. There exist constants $A > 0$ and $\nu > 0$ such that sup$_Q N(R_3, Q, \epsilon \|F\|_{Q,2}) \leq (A/\epsilon^\nu)$ for all $0 < \epsilon \leq 1$, where $N(R_3, Q, \epsilon \|F\|_{Q,2})$ denotes the covering number of the space $R_3$, $F$ is the envelope function of $R_3$, $\| \bullet \|_{Q,2}$ denotes the $\ell_2$-norm under some finitely discrete probability measure $Q$ on $(L, C)$, and the supremum is taken over all such probability measures.

Assumption 12. The nuisance function estimators $\hat{h}_0^{(k)}$ and $\hat{q}_0^{(k)}$ obtained from the $k$th fold of the data in the cross-fitting procedure in Section 4.3 satisfy that there exist constants $\alpha > 0$ and $\beta > 0$ such that $\|h_0(W, a, L) - \hat{h}_0^{(k)}(W, a, L)\|_{P,2} = O(n^{-\alpha})$ and $\|q_0(Z, a, L) - \hat{q}_0^{(k)}(Z, a, L)\|_{P,2} = O(n^{-\beta})$ uniformly for all $a \in A$ and $1 \leq k \leq K$, where $P$ is the underlying probability measure associated with $(L, W, Z, A, Y)$. In addition, max $\{ \|\hat{q}_0\|_{\infty}, \|h_0\|_{\infty} \} \leq C_2$ for some constant $C_2 > 0$.

Assumption 10 requires that the outcome $Y$ is bounded, which is commonly assumed in the literature (e.g., Zhao et al. 2012; Zhou et al. 2017). Assumption 10 also requires $h_0$ and $q_0$ to be both uniformly bounded. Similar conditions on nuisance functions have also been used in ITR learning under no unmeasured confounding (e.g., Zhao et al. 2019). Assumption 11 essentially states that $R_3$ is of a certain Vapnik-Chervonenkis class (see Definition 2.1 of Chernozhukov, Chetverikov, and Kato 2014). Assumption 12 imposes high-level conditions on the convergence rates of the estimated nuisance functions when the cross-fitting technique is applied. Under proper conditions (see, Dikkala et al. 2020, for details), the
nuisance function estimators obtained via (11) and (13) can satisfy Assumption 12.

Define the approximation error of \( r^* \) in terms of the \( \phi \)-risk by
\[
\hat{A}(\rho_{3,n}) = \inf_{r \in \mathcal{R}_3} \{ R_\phi(r) + \rho_{3,n} \| r \|_{\mathcal{R}_3}^2 \} - R_\phi(r^*),
\]
and consider the irreducible value gap between \( d_{3}^* \) and the global optimal ITR \( d^* \) by
\[
\mathcal{G}(d_{3}^*) = V(d^*) - V(d_{3}^*).
\]
For two positive sequences \( \{a_n : n \geq 1\} \) and \( \{b_n : n \geq 1\} \), we denote \( a_n \lesssim b_n \) if \( \limsup_{n \to \infty} a_n/b_n < \infty \). The finite sample excess risk bound for \( d_{3}^{DR} \) is given as follows.

**Theorem 5.1.** Suppose that Assumptions 1–9 and 10–12 hold. If \( \rho_{3,n} \leq 1 \), then for any \( x > 0 \), with probability at least \( 1 - \exp(-x) \), we have
\[
V(d^*) - V(d_{3}^{DR}) \lesssim \mathcal{G}(d_{3}^*) + \hat{A}(\rho_{3,n}) + \sqrt{n} \rho_{3,n}^{-1/2} n^{-1/2} + \max\{1, x\} n^{-\alpha(-\beta)} \rho_{3,n}^{-\beta} + \max\{1, x\} \sqrt{\nu} n^{-1/2 - \min\{\alpha, \beta\}} \rho_{3,n}^{-1/2}. \tag{23}
\]

The proof of Theorem 5.1 is given in supplementary material S2. Theorem 5.1 decomposes the finite sample excess risk bound for \( d_{3}^{DR} \) into four components. The first component \( \mathcal{G}(d_{3}^*) \) is an irreducible error due to unmeasured confounding. The second component \( \hat{A}(\rho_{3,n}) \) is the approximation error determined by the size of \( \mathcal{R}_3 \) and tuning parameter \( \rho_{3,n} \). The third component \( \sqrt{n} \rho_{3,n}^{-1/2} n^{-1/2} \) is the estimation error with respect to \( \mathcal{R}_3 \) assuming that the nuisance functions are known. The last two components, \( \max\{1, x\} n^{-\alpha(-\beta)} \rho_{3,n}^{-\beta} \) and \( \max\{1, x\} \sqrt{\nu} n^{-1/2 - \min\{\alpha, \beta\}} \rho_{3,n}^{-1/2} \), are the errors resulted from estimating the nuisance functions \( h_0 \) and \( q_0 \). These two errors are negligible if \( \alpha + \beta > \frac{1}{2} \), which can be guaranteed by many nonparametric estimators, including those defined in (11) and (13) under certain assumptions.

The finite sample error bounds for other proposed proximal learning methods can be similarly derived and are thus omitted here for brevity. Similar to that for \( d_{3}^{DR} \), each of their value difference bounds can be decomposed into aforementioned four components. The magnitude of the irreducible error due to unmeasured confounding is determined by the size of the corresponding identifiable class of treatment regimes. Obviously, the irreducible error related to \( d_{3}^* \) is the smallest but at the expense of making additional untestable assumptions. The approximation and estimation errors with respect to different pre-specified classes of treatment regimes are similar for all our proposed methods. The errors incurred by estimating nuisance functions vary among these methods and they depend on the assumptions imposed on the nuisance function estimation. While the result in Theorem 5.1 is similar to those from Zhao et al. (2019), we consider the ITR estimation problem in the presence of unmeasured confounders while Zhao et al. (2019) considered the scenario of no unmeasured confounding. Therefore, the assumptions and proofs for obtaining our result are different from those in Zhao et al. (2019).

**6. Simulation**

In this section we perform a simulation study to evaluate the numerical performance of our proposed optimal ITR estimators under unmeasured confounding, where a state-of-the-art estimator under no unmeasured confounding (Zhao et al. 2019) and an oracle estimator that can access the unmeasured confounders are compared. Due to the length of presentation, we only report key simulation results in this section. Implementation details and additional numerical results such as statistical inference related to \( V(d_3) \) for \( d_3 \in \mathcal{D}_3 \) can be found in supplementary material S3–S5.

**6.1. Simulated Data Generation**

Here we briefly describe our simulated data generating mechanism, which is motivated by Cui et al. (2020). The details are given in supplementary material S4.

**Step 1.** We generate the data \((L, A, Z, W, U)\) by the following steps:

1.1 The covariates \( L \) are generated by \( L \sim N([0.25, 0.25]^T, \text{diag}(0.25^2, 0.25^2)) \). Given \( L \), the treatment \( A \) is generated by a conditional Bernoulli distribution with
\[
\Pr(A = 1|L) = \left[1 + \exp\{0.125, 0.125L\}\right]^{-1}.
\]

As shown in supplementary material S4, this is compatible with
\[
\Pr(A = 1|U, L) = \left[1 + \exp\{0.09375, 0.1875|L - 0.25U|\}\right]^{-1}, \tag{24}
\]
where \( U \) is generated below.

1.2 Then we generate \((Z, W, U)\) by the following conditional multivariate normal distribution given \((A, L)\) with parameters given in supplementary material S4:
\[
(Z, W, U)|(A, L) \sim N\left(\begin{bmatrix}
\alpha_0 + \omega_{0}\bar{I}(A = 1) + \omega_{1}^T L \\
\mu_0 + \omega_{1}\bar{I}(A = 1) + \mu_{1}^T L \\
\kappa_0 + \omega_{2}\bar{I}(A = 1) + \kappa_{1}^T L \\
\sigma_z^2 & \sigma_{zw} & \sigma_{zu} \\
\sigma_{zw} & \sigma_w^2 & \sigma_{wu} \\
\sigma_{zu} & \sigma_{wu} & \sigma_u^2
\end{bmatrix}\right).
\]

As shown in supplementary material S4, they lead to the following equivalent model of \( q_0 \):
\[
q_0(Z, A, L) = 1 + \exp\left\{A(t_0 + t_z Z + t_u \bar{I}(A = 1) + t_1^T L)\right\},
\]
where the values of \( t_0, t_z, t_u \), and \( t_1 \) are specified in supplementary material S4.

**Step 2.** Based on the generated \((L, A, Z, W, U)\), we generate \( Y \) by adding a random noise from the uniform distribution on \([-1, 1]\) to
\[
\mathbb{E}(Y|W, U, A, Z, L) = \mathbb{E}(Y|U, A, Z, L) + \omega[W - \mathbb{E}(W|U, A, Z, L)] = \mathbb{E}(Y|U, A, L) + \omega[W - \mathbb{E}(W|U, L)],
\]
Table 3. Simulation scenarios for linear h0.

| Scenario | c0 | c1 | c2 | c3 | c4 | c5 | ω |
|----------|----|----|----|----|----|----|----|
| L1       | 2  | 0.5| 8  | [0.25, 0.25]T | 0.25 | [3, -5]T | 2  |
| L2       | 2  | 0.5| 8  | [0.25, 0.25]T | 0   | [3, -5]T | 2  |

Table 4. Simulation scenarios for nonlinear h0.

| Scenario | c0 | c1(A) | c2(L) | c4 | c5(L) | c6(L) | ω |
|----------|----|-------|-------|----|-------|-------|----|
| N1       | 2  | 2.33(A = 1) | 4     | L1L | -2.5  | | 2  |
| N2       | 2  | 0.25(A = 1)  | 5     | L1L | 0     | -6L1L2 | 0  |

of which details will be specified in different scenarios below.

Next we consider the following outcome confounding bridge function $h_0$, which leads to a linear decision function of $d^*$, We let

$$h_0(W, A, L) = c_0 + c_1A + c_2W + c_3L + \mathbb{I}(A = 1)(c_4W + c_5^T L)$$

$$E(Y | W, U, A, Z, L)c_0 + c_1\mathbb{I}(A = 1) + c_3^T L$$

$$+ \mathbb{I}(A = 1)(c_4W + c_5^T L), \text{ and thus}$$

$$E(Y | W, U, A, Z, L)c_0 + c_1\mathbb{I}(A = 1) + c_3^T L$$

$$+ \mathbb{I}(A = 1)(c_4W + c_3^T L + \omega W + [c_2 + c_4\mathbb{I}(A = 1) - \omega]$$

$$\left\{ \mu_0 + \mu_1^T L + \frac{\sigma_{uu}}{\sigma_u^2} (U - \kappa_0 - \kappa_1^T L) \right\},$$

where the values of $c_0$−$c_5$ and $\omega$ are specified in Table 3 and the other parameters are given in supplementary material S4. Accordingly the global optimal ITR is

$$d^*(L, U) = \text{sign}[E(Y | L, U, A = 1) - E(Y | L, U, A = -1)]$$

$$= \text{sign} \left[ c_1 + c_3^T L + c_4 \left\{ \mu_0 + \mu_1^T L + \frac{\sigma_{uu}}{\sigma_u^2} (U - \kappa_0 - \kappa_1^T L) \right\} \right].$$

As shown in Table 3, we consider two scenarios. In Scenario L1, the global optimal ITR cannot be identified, but the conditional treatment effect $E(Y | L, U, A = 1) - E(Y | L, U, A = -1)$ depends on both U and L, so taking W or Z into ITRs can potentially improve their value if identifiable. In Scenario L2, the global optimal ITR only depends on L, which thus can be identified under our proximal learning framework.

We also consider a second case where $h_0$ is a nonlinear function of L and W. We let

$$h_0(W, A, L) = c_0 + c_1A + c_2W + c_3L + \mathbb{I}(A = 1)(c_4W + c_5^T L)$$

and thus

$$E(Y | W, U, A, Z, L)c_0 + c_1\mathbb{I}(A = 1) + c_3^T L$$

$$+ \mathbb{I}(A = 1)(c_4W + c_3^T L + \omega W + [c_2 + c_4\mathbb{I}(A = 1) - \omega]$$

$$\left\{ \mu_0 + \mu_1^T L + \frac{\sigma_{uu}}{\sigma_u^2} (U - \kappa_0 - \kappa_1^T L) \right\},$$

where functions $c_1, c_3, c_5, c_6$, and values of $c_0, c_2, c_4, \omega$ are specified in Table 4 and the other parameters are given in supplementary materials. Accordingly the global optimal ITR is

$$d^*(L, U) = \text{sign}[E(Y | L, U, A = 1) - E(Y | L, U, A = -1)]$$

$$= \text{sign} \left[ c_1(1) - c_1(-1) + c_5(L) + [c_4 + c_6(L)]$$

$$\left\{ \mu_0 + \mu_1^T L + \frac{\sigma_{uu}}{\sigma_u^2} (U - \kappa_0 - \kappa_1^T L) \right\} \right].$$

These two scenarios N1 and N2 shown in Table 4 are analogous to L1 and L2 in Table 3. In each simulation setting described above, we have 100 simulation runs with n = 2000 and 5000 subjects in each run, and also generate a noise-free test dataset of 500,000 used to obtain values of estimated ITRs by (2) with (24).

6.2. Estimators and Implementations

To increase the difficulty of this simulation study, we add eight independent variables from the uniform distribution on $[-1, 1]$ in L and used the resulting concatenated 10-dimensional L in all ITR learning methods.

We implement all four proposed learning methods in Table 2 denoted by d1 (L, Z), d2 (L, W), d3DR (L), and d4 here, respectively, together with an optimal ITR estimator denoted by d1 (L) here, which is assumed to only depend on L and obtained by (15), as well as the optimal ITR estimator d2 (L), which is assumed to only depend on L and obtained by (16). The latter two are used to compare with d3DR (L). We also compare our proposed optimal ITR estimators with that by sufficient augmentation and relaxation learning (EARL, Zhao et al. 2019) designed under no unmeasured confounding. We denote it by dEARL (L) here and implement it using the R package DynTxRegime. To obtain dEARL (L) when U is unobserved, we use all observed variables (L, W, Z) to construct tree-based nonparametric main effect models and propensity score models, and then fit the regime in terms of L. We also create an oracle optimal ITR estimator, denoted by NUC, which is obtained by EARL using (L, U). Due to the relatively unsatisfactory computing speed of the R package DynTxRegime, we only obtain dEARL (L) and NUC in Scenarios L1 and L2 but not in Scenarios N1 and N2. For Scenario L2, we only use L to construct an ITR for NUC since the optimal one only depends on L. Note that NUC is unattainable in practice since U is unobserved, but it can be regarded as a benchmark estimator for comparison.

We use the quadratic smoothed hinge loss as surrogate $\phi$ for all proximal learners. The nuisance functions involved in our proposed methods, that is, the confounding bridge functions $h_0$ and $g_0$, are estimated by (11) and (13), respectively, with RKHSes $\mathcal{F}, \mathcal{H}, \mathcal{G}, \mathcal{Q}$ and all equipped with Gaussian kernels. For Scenarios L1 and L2, all decision functions are fitted as linear functions of their corresponding covariates with the $\ell_2$ penalty on coefficients. For Scenarios N1 and N2, $\mathcal{R}_1, \mathcal{R}_2,$ and $\mathcal{R}_3$ are chosen as RKHSes equipped with Gaussian kernels. Tuning parameters of the penalties are selected by cross-validation (see supplementary material S3 for details). For computational acceleration, all kernels are approximated by the Nyström method with $2[\sqrt{n}]$ features (Yang et al. 2012).

6.3. Simulation Results

The values of all optimal ITR estimates for $n = 5000$ are illustrated in Figures 4 and 5. The figures for $n = 2000$ and detailed comparisons with EARL are given in supplementary material S5.

Figure 4 shows that the oracle estimator NUC is the best in both Scenarios L1 and L2. This is not surprising since in Scen-
In Scenarios L1 and N1, \( d_1(L, Z) \) and \( d_2(L, W) \) outperform \( d_1(L) \) and \( d_2(L) \), respectively, which demonstrates the improvement of decision making by including the proxy variables \( Z \) and \( W \) when the optimal decisions depend on unobserved \( U \). For Scenarios L2 and N2, the global optimal ITR can be identified by our proposed methods since the difference of outcome bridge only depends on \( L \). Figure 4(b) shows that all our proposed methods are comparable to \( \text{NUC} \) in Scenario L2 but the values of \( d_{\text{EARL}}(L) \) are less satisfactory. Figure 4 also shows that \( d_1(L) \), \( d_2(L) \) and \( d_{3\text{DR}}(L) \) have similar performances in Scenario L2 while \( d_1(L, Z) \), \( d_2(L, W) \), and \( d_4 \) lead to values with slightly or substantially higher variations. This is consistent with the fact that the global optimal ITR only depends on \( L \) in Scenario L2 so adding \( Z \) or \( W \) into an ITR may only increase its variability. For Scenario N2, similar patterns can be discovered in Figure 5(b). The values of the maximum proximal estimator \( d_4 \) mostly lie between those of \( d_1(L, Z) \) and of \( d_2(L, W) \). These observations are intuitively reasonable and consistent with the discussion in Section 4.

In all scenarios, the optimal ITR estimators which involve estimating \( q_0 \) typically have relatively longer lower shadows. Therefore, it is interesting to study how to improve the practical performance of the treatment proximal learning method in the future.

7. Real Data Application

In this section, we apply the five proposed proximal learning methods to a dataset from the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT, Connors et al. 1996). SUPPORT examined the effectiveness and safety of the direct measurement of cardiac function by Right Heart Catheterization (RHC) for certain critically ill patients in intensive care units (ICUs). This dataset has been previously analyzed for estimating the average treatment effect of using RHC (e.g., Lin, Psaty, and Kronmal 1998; Tan 2006; Tchetgen Tchetgen et al. 2020; Cui et al. 2020).

Our objective is to find an optimal ITR on the usage of RHC that maximizes 30-day survival rates of critically ill patients from the day admitted or transferred to ICU. The data include 5735 patients, of whom 2184 were measured by RHC in the first 24 hr (\( A = 1 \)) and 3551 were in the control group (\( A = -1 \)). Finally, 3817 patients survived or censored at day 30 (\( Y = 1 \)) and 1918 died within 30 days (\( Y = -1 \)). For each individual, we consider 71 covariates including demographics, diagnosis, estimated survival probability, comorbidity, vital signs, and...
physiological status among others. See the full list of covariates at https://hbiostat.org/data.repo/rhc.html. During the first 24 hours in the ICU, 10 variables were measured from a blood test for the assessment of the physiological status. Following Tchetgen Tchetgen et al. (2020), among those 10 physiological status, we let $Z = (pafi1, paco21)$ be treatment-inducing confounding proxies and $W = (ph1, hema1)$ be outcome-inducing confounding proxies respectively, where $pafi1$ is the ratio of arterial oxygen partial pressure to fractional inspired oxygen, $paco21$ is the partial pressure of carbon dioxide, $ph1$ is arterial blood pH, $hema1$ is hematocrit. We apply our proposed methods to obtain $d1(L, Z), d2(L, W), d3DR(L), d1(L),$ and $d2(L)$ using this dataset with the same configurations of Scenarios L1 and L2 in Section 6.

The coefficient estimates of all covariates are given in supplementary material S6. In Table 5, we provide a selected number of them with relatively large coefficient estimates in absolute value. First, the negative intercepts in Table 5 indicates that RHC may have a potential negative averaged treatment effect on the 30-day survival rate for critically ill patients, which is consistent with the existing literature (e.g., Tan 2006). Second, the negative coefficients of surv2md1 suggest not perform RHC to a patient with a higher survival prediction on day 1. In contrast, our estimated ITRs tend to suggest trauma patients (trauma), and patients diagnosed with coma (cat1_coma, cat2_coma), lung cancer (cat1_lung, cat2_lung) or congestive heart failure (cat1_chf) undergo RHC. Clinically, RHC is of value when the hemodynamic state of a patient is in question or changing rapidly. It is thus potentially helpful with patients in critical condition whose hemodynamic states are unstable (Kubiak et al. 2019). Those findings can be partially supported by Hernandez et al. (2019) and Tehrani et al. (2019), but require further investigations. Our estimated ITRs also imply that patients with upper gastrointestinal bleeding (gibledhx) or autoimmune polyglandular syndrome type 3 (aps1) might be harmed by RHC, which is also worthy of future studies.

The treatment recommendations by the five estimated ITRs for all 5735 patients, as illustrated in Figure 6, show some discrepancies among them. The estimated ITR by doubly robust proximal learning, $d3DR(L)$, only suggests 675 patients to receive RHC, the smallest number among the five. The estimated ITRs by outcome proximal learning, $d1(L, Z)$ and $d1(L)$, both suggest a similar group of around 1400 patients receive RHC, while those by treatment proximal learning, $d2(L, W)$ and $d2(L)$, suggest RHC to a similar group of around 2200 patients.

Due to the potentially different recommendations by the five estimated ITRs, we conservatively choose the ITR with the highest 40% quantile of the estimated values obtained by a 5-fold cross-validation. Table 6 lists the 40% quantiles of estimated values for the five ITRs, of which $d1(L, Z)$ is suggested to be applied. Based on the ITR $d1(L, Z)$ for the 5735 patients, we construct a decision tree in Figure 7 to illustrate which covariates indicate the usage of RHC. For example, the patients in coma, diagnosed with multiple organ system failure with malignancy or those who agree to "Do Not Resuscitate" on the first day of enrollment are recommended to undergo RHC in the first 24 hr. One may also obtain an ensemble ITR by majority voting from the five estimated ITRs. A similar decision tree can be found in supplementary material S6.

### Table 5. Important coefficients of estimated optimal ITRs.

| Covariate         | $d2(L, W)$ | $d2(L)$ | $d3DR(L)$ | $d1(L)$ | $d1(L, Z)$ |
|-------------------|------------|---------|-----------|---------|------------|
| intercept         |            | -0.204  | -0.258    | -0.722  | -1.215     | -1.222     |
| surv2md1          |            | -0.683  | -0.699    | -0.254  | -0.747     | -0.724     |
| gibledhx (Yes=1/No=0) |          | -0.511  | -0.490    | -0.065  | -0.343     | -0.328     |
| aps1 (Yes=1/No=0) |            | -0.361  | -0.405    | -0.184  | -0.451     | -0.461     |
| trauma (Yes=1/No=0)|            | 0.862   | 0.095     | 0.020   | 0.622      | 0.533      |
| cat1_coma (Yes=1/No=0) |          | -0.091  | -0.112    | 0.215   | 2.183      | 2.210      |
| cat2_coma (Yes=1/No=0) |          | 0.261   | 0.248     | 0.092   | 2.059      | 2.009      |
| cat1_lung (Yes=1/No=0) |         | -0.025  | -0.098    | 0.062   | 1.740      | 1.711      |
| cat2_lung (Yes=1/No=0) |         | 0.960   | 1.131     | 0.033   | 1.412      | 1.072      |
| cat1_chf (Yes=1/No=0) |          | 0.757   | 0.731     | -0.005  | 0.544      | 0.525      |

### Table 6. 40% quantiles of the 5-fold cross-validation values from estimated optimal ITRs.

| ITR    | $d2(L, W)$ | $d2(L)$ | $d3DR(L)$ | $d1(L)$ | $d1(L, Z)$ |
|--------|------------|---------|-----------|---------|------------|
| 40% quantile | 0.3655 | 0.3901 | 0.3735 | 0.4231 | 0.4287   |

### Figure 6. Optimal treatments suggested by the five estimated ITRs. Light portions represent that ITRs suggest the patients to undergo RHC while dark portions represent otherwise.

### Figure 7. Decision tree of RHC based on $d(L, Z)$. mosfm: multiple organism system failure with malignancy. dnrl: "Do Not Resuscitate" status on day 1.
8. Discussions

In this article, we propose several proximal learning methods to find optimal ITRs under unmeasured confounding. Our methods are built upon the recently developed proximal causal inference. When the observed covariates can be decomposed into variables that are common causes of the outcome and treatment, namely outcome-inducing treatment-inducing confounding proxies, we establish several identification results on a variety of classes of ITRs under different assumptions. Based on these results, we propose several classification-based methods to estimate restricted in-class optimal ITRs. The superior performance of our methods is demonstrated by simulation. The real data application above shows the potential of the proposed methods to shed light on the recommended use of RHC on subgroups of patients, although this requires additional studies for confirmation and validation.

There are several interesting directions for future research. First, for the existence of treatment and outcome confounding bridges, if Assumptions 6 and 8 on completeness hold, then Assumptions 7 and 9 can be satisfied under some mild regularity conditions. As suggested by Cobzaru et al. (2022), one can perform a sensitivity analysis on the violation of completeness assumptions, for example, Assumptions 6 and 8 under specified structural equation models. Studying this issue in a more general setting would be an interesting future work. Second, as shown in the simulation study, although a flexible nonparametric method is used to estimate \( q_0 \) to alleviate the issue of the model misspecification, the finite sample performance of most optimal ITR estimators that use the estimated \( q_0 \) is not as good as that of those that use the estimated \( h_0 \). This demonstrates the difficulty in estimating \( q_0 \), analogous to that in estimating the propensity score in average treatment effect estimation under no unmeasured confounding. One possible approach to addressing this limitation is to develop weighted estimators similar to Wong and Chan (2017) and Athey, Imbens, and Wager (2018). Moreover, our proposed methods are developed for a single decision time point. It will be interesting to extend them to the longitudinal data to estimate the optimal dynamic treatment regimes where individualized decisions are needed at multiple time points. This may be practically useful, for example, to study the treatment of some chronic disease. Lastly, it is also meaningful to study the estimation of optimal ITRs for survival outcomes in the presence of unmeasured confounding.

Supplementary Materials

All technical proofs, implementation details and additional numerical results are given in supplementary materials.

Acknowledgments

The authors thank the editor, an associate editor, and three referees for their constructive comments and suggestions.

Funding

Xiaoke Zhang’s research is partially supported by the George Washington University University Facilitating Fund.

ORCID

Zhengling Qi http://orcid.org/0000-0003-0270-7969
Rui Mao http://orcid.org/0000-0001-5046-6341
Xiaoke Zhang http://orcid.org/0000-0002-7272-1523

References

Ai, C., and Chen, X. (2003), “Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions,” Econometrica, 71, 1795–1843. [920]

Athey, S., and Wager, S. (2021), “Policy Learning with Observational Data,” Econometrica, 89, 133–161. [915,921]

Athey, S., Imbens, G. W., and Wager, S. (2018), “Approximate Residual Balancing: Debiased Inference of Average Treatment Effects in High Dimensions,” Journal of the Royal Statistical Society, Series B, 80, 597–623. [927]

Bickel, P. J. (1982), “On Adaptive Estimation,” Annals of Statistics, 10, 647–671. [915,921]

Blundell, R., Chen, X., and Kristensen, D. (2007), “Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves,” Econometrica, 75, 1613–1669. [920]

Chamberlain, G. (1993), “Efficiency Bounds for Semiparametric Regression,” Econometrica, 60, 567–596. [920]

Chen, G., Zeng, D., and Kosorok, M. R. (2016), “Personalized Dose Finding using Outcome Weighted Learning,” Journal of the American Statistical Association, 111, 1509–1521. [915]

Chen, X. (2007), “Large Sample Sieve Estimation of Nonparametric Models,” Handbook of Econometrics, 6, 5549–5632. [920]

Chen, X., and Pouzo, D. (2012), “Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals,” Econometrica, 80, 277–321. [920]

Chen, X., Chernozhukov, V., Lee, S., and Newey, W. K. (2014), “Local Identification of Nonparametric and Semiparametric Models,” Econometrica, 82, 785–809. [918]

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014), “Gaussian Approximation of Suprema of Empirical Processes,” Annals of Statistics, 42, 1564–1597. [922]

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robinson, J. (2018), “Double/Debiased Machine Learning for Treatment and Structural Parameters,” The Econometrics Journal, 21, 1–68. [921]

Cobzaru, R., Welsch, R., Finkelstein, S., Ng, K., and Shahn, Z. (2022), “Bias Formulas for Violations of Partial Identification Assumptions,” arXiv preprint arXiv:2208.00105. [927]

Connors, Jr., A. F. J., Speroff, T., Dawson, N. V., Thomas, C., Harrell, Jr., F. E., Wagner, D., Desbiens, N., Goldman, L., Wu, A. W., Califf, R. M., Fulkerson, Jr., W. J., Vidailett, H., Broste, S., Bellamy, P., Lynn, J., and Knaus, W. A. (1996), “The Effectiveness of Right Heart Catheterization in the Initial Care of Critically Ill Patients,” Journal of the American Medical Association, 276, 889–897. [925]

Cui, Y., and Tchetgen Tchetgen, E. J. (2020), “A Semiparametric Instrumental Variable Approach to Optimal Treatment Regimes under Endogeneity,” Journal of the American Statistical Association, 116, 162–173. [916,918]

Cui, Y., Pu, H., Shi, X., Mao, W., and Tchetgen Tchetgen, E. J. (2020), “Semiparametric Proximal Causal Inference,” arXiv preprint arXiv:2011.08411. [916,918,919,921,923,925]

D’Haultfoeuille, X. (2011), “On the Completeness Condition in Nonparametric Instrumental Problems,” Econometric Theory, 27, 460–471. [918]

Dikkala, N., Lewis, G., Mackey, L., and Syrgkanis, V. (2020), “Minimax Estimation of Conditional Moment Models,” arXiv preprint arXiv:2006.07201. [916,920,922]

Dudík, M., Langford, J., and Li, L. (2011), “ Doubly Robust Policy Evaluation and Learning,” arXiv preprint arXiv:1103.4601. [915]

Ghassami, A., Ying, A., Shpitser, I., and Tchetgen Tchetgen, E. J. (2021), “Minimax Kernel Machine Learning for a Class of Doubly Robust Functionals,” arXiv preprint arXiv:2104.02929. [916]

Han, S. (2019), “Optimal Dynamic Treatment Regimes and Partial Welfare Ordering,” arXiv preprint arXiv:1912.10014. [916]
Hernandez, G. A., Lemor, A., Blumer, V., Rueda, C. A., Zalawadiya, S., Stevenson, L. W., and Lindenfeld, J. (2019), “Trends in Utilization and Outcomes of Pulmonary Artery Catheterization in Heart Failure with and without Cardiogenic Shock,” *Journal of Cardiac Failure*, 25, 364–371. [926]

Kallus, N., and Zhou, A. (2018), “Confounding-Robust Policy Improvement,” in *Advances in Neural Information Processing Systems* (Vol. 31). [916]

Kallus, N., Mao, X., and Uehara, M. (2021), “Causal Inference Under Unmeasured Confounding With Negative Controls: A Minimax Learning Approach,” arXiv preprint arXiv:2103.14029. [916]

Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., and Murphy, S. A. (2015), “Microrandomized Trials: An Experimental Design for Developing Just-in-Time Adaptive Interventions,” *Health Psychology*, 34, 1220–1228. [915]

Kober, J., Bagnell, J. A., and Peters, J. (2013), “Reinforcement Learning in Robotics: A Survey,” *The International Journal of Robotics Research*, 32, 1238–1274. [915]

Kosorok, M. R., and Laber, E. B. (2019), “Precision Medicine,” *Annual Review of Statistics and Its Application*, 6, 263–286. [916]

Kress, R., Mazya, V., and Kozlov, M. A. (1989), *Linear Integral Equations* (Vol. 82), New York: Springer. [918]

Kubiac, G. M., Ciarka, A., Biniecka, M., and Ceranowicz, P. (2019), “Right Heart Catheterization—Background, Physiological Basics, and Clinical Implications,” *Journal of Clinical Medicine*, 8, 1331. [926]

Lin, D. Y., Psaty, B. M., and Kronmal, R. A. (1998), “Assessing the Sensitivity of Regression Results to Unmeasured Confounders in Observational Studies,” *Biometrics*, 54, 948–963. [925]

Manski, C. F. (2004), “Statistical Treatment Rules for Heterogeneous Populations,” *Econometrica*, 72, 1221–1246. [916]

Miao, W., and Tchetgen Tchetgen, E. J. (2018), “A Confounding Bridge Approach for Double Negative Control Inference on Causal Effects,” arXiv preprint arXiv:1808.04945. [916,917,918,919]

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018), “Identifying Causal Effects with Proxy Variables of an Unmeasured Confounder,” *Biometrika*, 105, 987–993. [918]

Newey, W. K., and Powell, J. L. (2003), “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 71, 1565–1578. [918,920]

Pearl, J. (2009), *Causality*, Cambridge: Cambridge University Press. [917]

Pu, H., and Zhang, B. (2020), “Estimating Optimal Treatment Rules with an Instrumental Variable: A Semi-Supervised Learning Approach,” arXiv preprint arXiv:2002.02579. [916]

Qian, M., and Murphy, S. A. (2011), “Performance Guarantees for Individualized Treatment Rules,” *Annals of Statistics*, 39, 1180–1210. [915,916,917]

Qiu, H., Carone, M., Sadikova, E., Petukhova, M., Kessler, R. C., and Luedtke, A. (2020), “Optimal Individualized Decision Rules using Instrumental Variable Methods,” *Journal of the American Statistical Association*, 116, 174–191. [916,918]

Rashid, N. U., Luckett, D. J., Chen, J., Lawson, M. T., Wang, L., Zhang, Y., Laber, E. B., Liu, Y., Yeh, J. J., Zeng, D., and Kosorok, M. R. (2020), “High-Dimensional Precision Medicine From Patient-Derived Xenografts,” *Journal of the American Statistical Association*, 116, 1140–1154. [915]

Robins, J. (1986), “A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period—Application to Control of the Healthy worker survivor effect,” *Mathematical Modelling*, 7, 1393–1512. [917]

Seong, K., Mohseni, M., and Cioffi, J. M. (2006), “Optimal Resource Allocation for OFDMA Downlink Systems,” in *2006 IEEE International Symposium on Information Theory*, pp. 1394–1398, IEEE. [915]

Shi, C., Fan, A., Song, R., and Lu, W. (2018), “High-Dimensional A-Learning for Optimal Dynamic Treatment Regimes,” *Annals of Statistics*, 46, 925–957. [915]

Shi, X., Miao, W., Nelson, J. C., and Tchetgen Tchetgen, E. J. (2020), “Multiply Robust Causal Inference with Double-Negative Control Adjustment for Categorical Unmeasured Confounding,” *Journal of the Royal Statistical Society, Series B*, 82, 521–540. [918]

Tan, Z. (2006), “A Distributional Approach for Causal Inference using Propensity Scores,” *Journal of the American Statistical Association*, 101, 1619–1637. [925,926]

Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X., and Miao, W. (2020), “An Introduction to Proximal Causal Learning,” arXiv preprint arXiv:2009.10982. [916,917,925,926]

Tehrani, B. N., Truesdell, A. G., Sherwood, M. W., Desai, S., Tran, H. A., Epps, K. C., Singh, R., Psotka, M., Shah, P., Cooper, L. B., Rosner, C., Raja, A., Barnett, S. D., Saulino, P., deFilippi, C. R., Gurbel, P. A., Murphy, C. E., and O’Connor, C. M. (2019), “Standardized Team-based Care for Cardiogenic Shock,” *Journal of the American College of Cardiology*, 73, 1659–1669. [926]

Wang, L., and Tchetgen Tchetgen, E. (2018), “Bounded, Efficient and Multiply Robust Estimation of Average Treatment Effects using Instrumental Variables,” *Journal of the Royal Statistical Society, Series B*, 80, 531–550. [916]

Wang, L., Zhou, Y., Song, R., and Sherwood, B. (2018), “Quantile-Optimal Treatment Regimes,” *Journal of the American Statistical Association*, 113, 1243–1254. [915]

Watkins, C. J., and Dayan, P. (1992), “Q-learning,” *Machine Learning*, 8, 279–292. [915]

Wong, R. K., and Chan, K. C. G. (2017), “Kernel-based Covariate Functional Balancing for Observational Studies,” *Biometrika*, 105, 199–213. [927]

Yang, T., Li, Y.-F., Mahdavi, M., Jin, R., and Zhou, Z.-H. (2012), “Nystrom Method vs Random Fourier Features: A Theoretical and Empirical Comparison,” *Advances in Neural Information Processing Systems*, 25, 476–484. [924]

Zhang, B., Tsatis, A. A., Laber, E. B., and Davidian, M. (2012), “A Robust Method for Estimating Optimal Treatment Regimes,” *Biometrics*, 68, 1010–1018. [915,920]

Zhao, Y., Kosorok, M. R., and Zeng, D. (2009), “Reinforcement Learning Design for Cancer Clinical Trials,” *Statistics in Medicine*, 28, 3294–3315. [915]

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012), “Estimating Individualized Treatment Rules using Outcome Weighted Learning,” *Journal of the American Statistical Association*, 107, 1106–1118. [915,917,920,922]

Zhao, Y.-Q., Zeng, D., Laber, E. B., Song, R., Yuan, M., and Kosorok, M. R. (2014), “Doubly Robust Learning for Estimating Individualized Treatment with Censored Data,” *Biometrika*, 102, 151–168. [915]

Zhao, Y.-Q., Laber, E. B., Ning, Y., Saha, S., and Sands, B. E. (2019), “Efficient Augmentation and Relaxation Learning for Individualized Treatment Rules using Observational Data,” *Journal of Machine Learning Research*, 20, 1–23. [915,917,922,923,924]

Zhou, X., Mayer-Hamblett, N., Khan, U., and Kosorok, M. R. (2017), “Residual Weighted Learning for Estimating Individualized Treatment Rules,” *Journal of the American Statistical Association*, 112, 169–187. [922]