Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach

Hongming Pu | Bo Zhang

University of Pennsylvania, Philadelphia, USA

Correspondence
Bo Zhang, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: bozhan@wharton.upenn.edu

Abstract
Individualized treatment rules (ITRs) are considered a promising recipe to deliver better policy interventions. One key ingredient in optimal ITR estimation problems is to estimate the average treatment effect conditional on a subject’s covariate information, which is often challenging in observational studies due to the universal concern of unmeasured confounding. Instrumental variables (IVs) are widely used tools to infer the treatment effect when there is unmeasured confounding between the treatment and outcome. In this work, we propose a general framework of approaching the optimal ITR estimation problem when a valid IV is allowed to only partially identify the treatment effect. We introduce a novel notion of optimality called ‘IV-optimality’. A treatment rule is said to be IV-optimal if it minimizes the maximum risk with respect to the putative IV and the set of IV identification assumptions. We derive a bound on the risk of an IV-optimal rule that illuminates when an IV-optimal rule has favourable generalization performance. We propose a classification-based statistical learning method that estimates such an IV-optimal rule, design computationally efficient algorithms, and prove theoretical guarantees. We contrast our proposed method to the popular outcome weighted learning (OWL) approach via extensive simulations, and apply our method to study which mothers would benefit from travelling to deliver their premature babies at hospitals with high-level
INTRODUCTION

1.1 Estimating individualized treatment rules with a valid instrumental variable

Individualized treatment rules (ITRs) are now recognized as a general recipe for leveraging vast amount of clinical, prognostic and socioeconomic status data to deliver the best possible healthcare or other policy interventions. Researchers across many disciplines have responded to this trend by developing novel data-driven strategies that estimate ITRs. Some seminal works include Murphy (2003), Robins (2004), Qian and Murphy (2011), Zhang et al. (2012a, b) and Zhao et al. (2012), among others. See Kosorok and Laber (2019) and Tsiatis (2019) for comprehensive and up-to-date surveys. One key ingredient in ITR estimation problems is to estimate the average treatment effect conditional on patients’ clinical and prognostic features. However, estimation of the conditional average treatment effect (CATE) can be challenging in randomized control trials (RCTs) with high-dimensional covariates, limited sample size and individual non-compliance, and observational studies due to the universal concern of unmeasured confounding (Cui & Tchetgen Tchetgen, 2020; Kallus & Zhou, 2018; Kallus et al., 2019; Qiu et al., 2020; Zhang et al., 2020).

In non-ITR settings, instrumental variables (IVs) are commonly used tools to infer treatment effects in observational studies where observed covariates cannot adequately adjust for confounding between the treatment and outcome. However, few works have explored using an IV to estimate optimal ITRs. Two exceptions are Cui and Tchetgen Tchetgen (2020) and Qiu et al. (2020), both of which studied ITR estimation problems when an IV can be used to point identify the conditional average treatment effect under assumptions introduced in Wang and Tchetgen Tchetgen (2018). However, one limitation of their approaches is that assumptions that allow a valid IV to point identify the CATE may be quite stringent in some scenarios and do not necessarily hold, especially in ITR estimation settings where treatment effect heterogeneity is expected.

This article takes a distinct perspective. Although a valid IV in general cannot point identify the average treatment effect (ATE) and similarly the CATE, it can partially identify them, in the sense that a lower and upper bound of ATE and CATE can be obtained with a valid IV. Depending on the quality of the putative IV and various identification assumptions, lengths of partial identification intervals may vary, and in the extreme case the intervals collapse to points, that is, the ATE or CATE (Angrist et al., 1996; Balke & Pearl, 1997; Manski, 2003; Robins & Greenland, 1996). See Swanson et al. (2018) for an up-to-date literature review on partial identification of ATE using an IV. It is worth pointing out that a partial identification interval is fundamentally different from a confidence interval, in that a confidence interval shrinks to a point as sample size goes to infinity, while a partial identification interval remains, in the limit, an interval, and represents the intrinsic uncertainty of an IV analysis.
A popular approach to ITR estimation problems in non-IV settings is to transform the problem into a weighted classification problem, where the sign of the CATE constitutes a subject’s label \{-1, +1\}, and the magnitude the weight. In an IV setting, the point identified CATE is replaced with an interval \(I\). When such an interval avoids 0, say \(I = [1, 3]\), it is clear that the subject benefits from receiving the treatment and should be labelled as such. However, the situation becomes complicated when the interval covers 0, say \(I = [-1, 3]\), in which case the true CATE can be anything between −1 and 3, and such a subject can no longer be labelled as benefiting or not benefiting from the treatment. In this way, partial identification of the CATE in an IV analysis poses a fundamental challenge to optimal ITR estimation.

We have three objectives in this article. First, we propose a general classification-based (Zhang et al., 2012a; Zhao et al., 2012) framework for optimal ITR estimation problems with an IV. The putative IV is allowed to only partially identify the conditional average treatment effect. Second, we introduce a novel notion of optimality called IV-optimality: a treatment rule is said to be IV-optimal if it minimizes the maximum risk that could incur given a putative IV and a set of IV identification assumptions. One remarkable feature of ‘IV-optimality’ is that it is amenable to different IVs and identification assumptions, and allows empirical researchers to weigh estimation precision against credibility. We also derive bounds on the risk of an IV-optimal rule, which illuminates when an IV-optimal ITR has favourable generalization performance. Finally, we derive an estimator of such an IV-optimal rule which we call the IV-PILE estimator, develop computationally efficient algorithms, and prove theoretical guarantees. The worst-case risk of the IV-PILE estimator can be estimated and gives practitioners important guidance on the applicability of their estimated ITRs. Via extensive simulations, we demonstrate that our proposed method has favourable performance compared to applying the outcome weighted learning (OWL) (Zhao et al., 2012) to observational data in the presence of unmeasured confounding. Finally, we apply our developed method to study the differential impact of delivery hospital on neonatal health outcomes. R package ivitr implements the proposed methodology.

1.2 A motivating example: differential impact of delivery hospital on premature babies’ health outcomes

We consider a concrete example to carry forward our discussion. Lorch et al. (2012) constructed a cohort-based retrospective study out of all hospital-delivered premature babies in Pennsylvania and California between 1995 and 2005 and Missouri between 1995 and 2003. Using the differential travel time as an instrumental variable, Lorch et al. (2012) find that there is benefit to neonatal outcomes when premature babies are delivered at hospitals with high-level neonatal intensive care units (NICUs) compared to hospitals without high-level NICUs.

An IV analysis is crucial in this study and similar studies based on observational data. Studies that directly use the treatment, for example delivery hospitals in the NICU study, to infer the treatment effect often cannot sufficiently adjust for unmeasured and unrecorded factors, such as the severity of comorbidities or laboratory results (Lorch et al., 2012). Although Lorch et al. (2012) have found evidence supporting a positive treatment effect of mothers delivering premature infants at high-level NICUs, some important questions remain. First, it is of great scientific interest to understand which mothers had better be sent to hospitals with high-level NICUs as compared to which mothers are just as well off at low level NICUs. Given the current limited capacity of high-level NICUs, an answer to this question would facilitate our understanding of which mothers and their premature babies are most in need of high-level NICUs, and provide insight into how to construct optimal perinatal regionalization systems, systems that designate hospitals by the scope of perinatal service provided and designate
where infants are born or transferred according to the level of care they need at birth (Kroeling et al., 2018; Lasswell et al., 2010). Such scientific inquiries elicit estimating individualized treatment rules using observational data consisting of mothers’ observed characteristics, treatment received, outcome of interest, and a valid instrumental variable. We will revisit this example and apply our developed methodology to it near the end of the article.

2 | ITR ESTIMATION WITH AN IV: FROM POINT TO PARTIAL IDENTIFICATION

We briefly review the problem of estimating ITRs from a classification perspective, and discuss the key impediment to generalizing the estimation strategy from RCTs to observational data in Section 2.1. Section 2.2 discusses how to leverage a valid IV to partially identify the conditional average treatment effect, and Section 2.3 proposes a general framework of approaching the ITR estimation problem with a valid IV.

2.1 | ITR estimation from a classification perspective: from randomized control trials to observational studies

Suppose that the data \{ (X_i, A_i, Y_i), \ i = 1, \ldots, n \} are collected from a two-arm randomized trial. The \(d\)-dimensional vector \(X_i \in \mathcal{X}\) encodes subject \(i\)'s prognostic characteristics, \(A_i \in \mathcal{A} = \{-1, +1\}\) a binary treatment, and \(Y_i \in \mathbb{R}\) an outcome of interest. Let \(f(\cdot): \mathcal{X} \mapsto \mathbb{R}\) be a discriminant function such that the sign of \(f(\cdot)\) yields the desired treatment rule. Let \(\mu(1, X) = \mathbb{E}[Y(1)|X] = \mathbb{E}[Y|A=1, X]\) and \(\mu(-1, X) = \mathbb{E}[Y(-1)|X] = \mathbb{E}[Y|A=-1, X]\) denote the average potential outcomes conditional on \(X\) in each arm, and \(C(X) = \mu(1, X) - \mu(-1, X)\) the conditional average treatment effect, or CATE, following the notation in Zhang et al. (2012a).

The value function of a particular rule \(f(\cdot)\) is defined to be \(V(f) = \mathbb{E}[\text{sgn}(f(X))]\). An optimal rule is the one that maximizes \(V(f)\) among a class of functions \(\mathcal{F}\), or equivalently minimizes the following risk:

\[
R(f) = \mathbb{E} \left[ |C(X)| \cdot 1 \{ \text{sgn}(C(X)) \neq \text{sgn}(f(X)) \} \right],
\]

where \(\text{sgn}(x) = 1, \forall x > 0\) and \(-1\) otherwise.

Let \(B_i = \text{sgn}\{ C(X_i) \}\) be a latent class label that assigns \(+1\) to subject \(i\) if she would benefit from the treatment and \(-1\) otherwise. If a correct treatment decision is made, that is, \(\text{sgn}(f(X_i)) = \text{sgn}\{ C(X_i) \}\), there is no loss incurred; otherwise, the decision is not optimal and the corresponding loss has magnitude \(W_i = |C(X_i)|\). In this way, the optimal ITR estimation problem is recast as a weighted classification problem whose expected weighted misclassification error is specified in Equation (1). In a typical classification problem, the training data contain class labels and weights. In the context of ITR estimation problems, the contrast function \(C(X_i)\) for subject \(i\) is first estimated from data, say as \(\hat{C}(X_i)\), and the associated label and weight are then constructed accordingly: \(\hat{B}_i = \text{sgn}\{ \hat{C}(X_i) \}\) and \(\hat{W}_i = |\hat{C}(X_i)|\). To summarize, the original data \{ (X_i, A_i, Y_i), \ i = 1, \ldots, n \} are transformed into \{ (X_i, \hat{B}_i, \hat{W}_i), \ i = 1, \ldots, n \}, and a standard classification routine is then applied to this derived data set in order to estimate a function \(f(\cdot) \in \mathcal{F}\) that minimizes \(R(f)\). This framework, as discussed in more detail in Zhang et al. (2012a), covers many popular ITR methodologies. Notably, the popular outcome weighted learning (OWL) approach (Zhao et al., 2012) is a particular instance of this general
framework, where $C(X)$ is estimated via an inverse probability weighted estimator (IPWE) and a support vector machine (SVM) is used to perform classification.

One critical task in estimating optimal ITRs from this classification perspective is to estimate well the contrast $C(X)$. With a known propensity score as in a randomized control trial, an IPWE can be used to unbiasedly estimate $C(X)$. However, this task becomes much more challenging when data come from observational studies. A key assumption in drawing causal inference from observational data is the so-called treatment ignorability assumption (Rosenbaum & Rubin, 1983), also known as the no unmeasured confounding assumption (henceforth NUCA) (Robins, 1992), or treatment exogeneity (Imbens, 2004). A version of the no unmeasured confounding assumption states that

$$F(Y(1), Y(-1)|A=a, X=x) = F(Y(1), Y(-1)|X=x), \forall (a, x),$$

where $F(\cdot)$ denotes the cumulative distribution function. In words, the treatment assignment is effectively randomized within strata formed by observed covariates. However, when the NUCA fails, the conditional average treatment effect may not be unbiasedly estimated from observed data as $\mathbb{E}[Y(1)|X]$ is not necessarily equal to $\mathbb{E}[Y|A=1, X]$.

### 2.2 Instrumental variables: assumptions and partial identification

An instrumental variable is a useful tool to estimate the treatment effect when the treatment and outcome are believed to be confounded by unmeasured confounders. We consider the potential outcome framework that formalizes an IV as in Angrist et al. (1996). Let $Z_i \in \{-1, +1\}$ be a binary IV associated with subject $i$, and $Z$ a length-$n$ vector containing all IV assignments. Let $A_i(Z)$ be the indicator of whether subject $i$ would receive the treatment or not under IV assignment $Z$, $A$ a length-$n$ vector of treatment assignment status with $A_i(Z)$ being the $i$th entry, and $Y_i(Z, A)$ the outcome of subject $i$ under IV assignment $Z$ and treatment assignment $A$. We assume that the following core IV assumptions hold (Angrist et al., 1996):

1. Stable Unit Treatment Value Assumption (SUTVA): $Z_i = Z_i'$ implies $A_i(Z) = A_i(Z')$; $Z_i = Z_i'$ and $A_i = A_i'$ together imply $Y_i(Z, A) = Y_i(Z', A')$.
2. Positive correlation between IV and treatment: $P(A=1|Z=1, X=x) > P(A=1|Z=-1, X=x)$ for all $x$.
3. Exclusion restriction (ER): $Y_i(Z', A) = Y_i(Z', A)$ for all $Z, Z', A$.
4. IV unconfoundedness conditional on $X$: $Z \perp A(z), Y(z, a) | X$ for $z \in \{-1, +1\}$ and $a \in \{-1, +1\}$.

These four core IV assumptions do not allow point identification of the average treatment effect (ATE); however, they lead to the well-known Balke–Pearl bound on the ATE (Balke & Pearl, 1997). See Swanson et al. (2018) for other (possibly weaker) versions of assumptions that lead to the same Balke–Pearl bound. For a continuous but bounded outcome, Manski and Pepper (2000) derived a non-parametric bound under the monotone instrumental variable assumption. Additional assumptions are typically needed to tighten these bounds or to identify the treatment effect in an identifiable subgroup or the entire population.

In a binary IV analysis, a subject belongs to one of the four compliance classes: (1) an always-taker if $\{A(1), A(-1)\}=(1, 1)$; (2) a complier if $\{A(1), A(-1)\}=(1, -1)$; (3) a never-taker if $\{A(1), A(-1)\}=(-1, -1)$; (4) a defier if $\{A(1), A(-1)\}=(-1, 1)$. When the outcome is binary and the proportion of defiers is known, Richardson and Robins (2010) discussed how to bound the ATE.
among all four compliance classes as a function of the defier proportion. An important instance is when there is no defiers and a valid IV can be used to identify the local average treatment effect, that is, the average treatment effect among compliers, as shown in the seminal work by Angrist et al. (1996).

The fundamental limitation of an IV analysis is that even a valid IV provides no information regarding the counterfactual outcomes of always-takers and never-takers: we simply do not have information about what would happen had always-takers been forced to forgo the treatment, and never-takers had they been forced to accept the treatment. This suggests another strategy to further bound the population average treatment effect by setting bounds to $Y(−1)$ and $Y(1)$ always-takers and never-takers. Swanson et al. (2018) contains a detailed account of various proposals on how to set these bounds. Some versions of IV identification assumptions, for instance those established in Wang and Tchetgen Tchetgen (2018), would allow a valid IV to point identify the population average treatment effect. Estimating optimal treatment rules when the CATE can be point identified using a valid IV has been carefully studied in Cui and Tchetgen Tchetgen (2020) and Qiu et al. (2020), and is not the focus of the current paper, although it is a special case of our general framework. The rest of this article focuses on how to estimate useful individualized treatment rules when an IV only partially identifies the CATE, possibly under various IV-specific identification assumptions.

2.3 Estimating optimal ITR with an IV from a partial identification perspective

We now describe a general framework of approaching the problem of estimating optimal treatment rules using a valid IV in observational studies. Suppose that we have i.i.d data \{(X_i, Z_i, A_i, Y_i) \; | \; i = 1, \ldots, n\} with a binary IV $Z_i$, binary treatment $A_i$, observed covariates $X_i \in \mathbb{R}^d$, and outcome of interest $Y_i \in \mathbb{R}$. Let $C$ denote a set of IV identification assumptions and $I_i = [L(X_i), U(X_i)] \ni C(X_i)$ a partial identification interval of the conditional average treatment effect $C(X_i)$ associated with subject $i$ under $C$. We view each subject as belonging to one of the following three latent classes (with class label $B_i$):

1. $B_i = +1$ if $I_i > \Delta$ in the sense that $x > \Delta, \forall x \in I_i$;
2. $B_i = −1$ if $I_i < \Delta$ in the sense that $x < \Delta, \forall x \in I_i$;
3. $B_i = NA$ if $\Delta \in I_i$.

In words, the class $B = +1$ consists of those who would benefit at least $\Delta$ from the treatment, $B = −1$ those who would not benefit more than $\Delta$, and $B = NA$ those for whom the putative IV and the set of identification assumptions $C$ together cannot assert that subject $i$ would benefit at least $\Delta$ from the treatment or not. We will refer to subjects with $B_i \in \{-1, +1\}$ as labelled subjects and $B_i = NA$ unlabelled.

Remark 1 We let $\Delta$ denote a margin of practical relevance. In many ITR settings, the treatment may only do harm (or good), and we would recommend taking/not taking the treatment only when the margin is large. In our application, a high-level NICU might never do harm to mothers and preemies compared to a low-level NICU; however, given the limited capacity of high-level NICUs and long travel times for some mothers to a high-level NICUs, it may be more reasonable for mothers and their newborns to be sent to the nearest NICU, unless a high-level NICU is significantly better in reducing the mortality. This trade-off is reflected by the margin $\Delta$ set according to expert knowledge. One can always let $\Delta = 0$ and the problem is reduced to the more familiar setting.
In practice, \( I_i \) is estimated from the observed data, say as \( \hat{I}_i \), and \( B_i \) is constructed accordingly, say as \( \hat{B}_i \). Consider the derived data set \( D = \{ (X_i, \hat{I}_i, \hat{B}_i), \ i = 1, \ldots, n \} \). Write
\[
D = D_l \cup D_{ul} = \{ (X_i, \hat{I}_i, \hat{B}_i), \ i = 1, \ldots, l \} \cup \{ (X_i, \hat{I}_i), \ i = l + 1, \ldots, n \},
\]
where \( l \) subjects in \( D_l \) have labels \( \hat{B}_i \in \{-1, \ +1\} \) and \( u = n−l \) subjects in \( D_{ul} \) are unlabelled. Our goal is to still learn an ‘optimal’ treatment rule \( f(\cdot) \) such that some properly defined misclassification error is minimized.

## 3 | EXAMPLES OF PARTIAL IDENTIFICATION BOUNDS FOR A BINARY OUTCOME

### 3.1 | Balke–Pearl bound

Assume that the four core IV assumptions (IV.A1–IV.A4) stated in Section 2.2 hold within strata formed by observed covariates, that is,
\[
C_{BP} = \{ \text{IV.A1 - IV.A4 hold within strata of } X \}.
\]
The Balke–Pearl bounds state that the conditional average treatment effect \( C(X) \) is lower bounded by (Balke & Pearl, 1997; Cui & Tchetgen Tchetgen, 2020):
\[
L(X) = \max \begin{cases}
\frac{p_{-1,-1|1}X + p_{1,1|1}X - 1}{p_{-1,-1|1}X + p_{1,1|1}X - 1} \\
p_{1,1|1}X + p_{-1,-1|1}X - 1 \\
p_{-1,-1|1}X + p_{1,1|1}X - 1 \\
2p_{-1,-1|1}X + p_{1,1|1}X + p_{-1,-1|1}X - 2 \\
p_{-1,-1|1}X + 2p_{1,1|1}X + p_{-1,-1|1}X - 2 \\
p_{1,1|1}X + 2p_{-1,-1|1}X + p_{1,1|1}X - 2 \\
p_{-1,-1|1}X + p_{1,1|1}X + p_{-1,-1|1}X + 2p_{1,1|1}X - 2
\end{cases},
\]
and upper bounded by
\[
U(X) = \min \begin{cases}
1 - p_{-1,-1|1}X - p_{1,1|1}X \\
1 - p_{1,-1|1}X - p_{-1,-1|1}X \\
1 - p_{1,1|1}X - p_{-1,-1|1}X \\
2 - 2p_{-1,-1|1}X - p_{1,1|1}X - p_{-1,-1|1}X - p_{1,1|1}X \\
2 - p_{-1,-1|1}X - 2p_{1,1|1}X - p_{-1,-1|1}X - p_{1,1|1}X \\
2 - p_{1,1|1}X - 2p_{-1,-1|1}X - p_{1,1|1}X - p_{1,1|1}X \\
2 - p_{-1,-1|1}X - p_{-1,-1|1}X - p_{1,1|1}X - 2p_{1,1|1}X
\end{cases},
\]
where \( p_{y,a|X} \) is a shorthand for \( P(Y = y, A = a; Z = z, X) \). Note that all conditional probabilities \( \{ P(Y = y, A = a; Z = z, X = x) \}, y = ±1, a = ±1, z = ±1 \) can in principle be non-parametrically identified. In practice, we may estimate them by re-coding \( 2 \times 2 = 4 \) combinations of \( Y \in \{-1, +1\} \) and \( A \in \{-1, +1\} \) as
four categories and fitting a flexible and expressive multi-class classification routine, for example random forests (Breiman, 2001).

3.2 Bounds as in Siddique (2013)

Siddique (2013) considers an assumption (in addition to four core IV assumptions) that limits treatment heterogeneity in the following way:

1. Correct Non-Compliant Decision:

\[
E[Y(1)|A = 1, Z = -1] - E[Y(-1)|A = -1, Z = -1] \geq 0,
\]

\[
E[Y(-1)|A = -1, Z = 1] - E[Y(1)|A = -1, Z = 1] \geq 0.
\]

In words, this assumption states that for those who take a treatment different from the encouragement (i.e. \(A \neq Z\)), their decisions are on average favourable. Under the four core IV assumptions and this extra assumption, the bound on ATE can be further tightened. Let

\[C_{Sid} = \{ IV.A1 - IV.A4 plus IV.A5 hold within strata of X \}.
\]

Under IV identification set \(C_{Sid}\), the conditional average treatment effect is lower bounded by Siddique (2013) and Swanson et al. (2018):

\[
L(X) = \max \left\{ \frac{p_{1,1|x} + p_{1,-1|x}}{p_{1,1|x}}, \frac{p_{1,-1|x} + p_{1,1|x}}{p_{1,1|x}} \right\} - \min \left\{ \frac{p_{1,-1|x} + p_{1,1|x}}{p_{1,-1|x}}, \frac{p_{1,1|x} + p_{1,-1|x}}{p_{1,1|x}} \right\}, \tag{4}
\]

and upper bounded by

\[
U(X) = \min \left\{ \frac{p_{1,1|x} + p_{1,-1|x}}{p_{1,1|x}}, \frac{p_{1,-1|x} + p_{1,1|x}}{p_{1,1|x}} \right\} - \max \left\{ \frac{p_{1,-1|x} + p_{1,1|x}}{p_{1,-1|x}}, \frac{p_{1,1|x} + p_{1,-1|x}}{p_{1,1|x}} \right\}, \tag{5}
\]

where \(p_{y,a|z,x}\) again stands for \(P(Y = y, A = a|Z = z, X)\), and \(p_{a|z,x}\) is a shorthand for \(P(A = a|Z = z, X)\). Observe that \(P(A = a|Z = z, X) = \sum_{y \in \{0,1\}} P(Y = y, A = a|Z = z, X)\), and we can again estimate the lower bound and upper bound by estimating \(\{P(Y = y, A = a|Z = z, X = x), y = \pm 1, a = \pm 1, z = \pm 1\}\).

Remark 2 (Assumption set \(C\)) There are many other IV identification assumptions that help reduce the length of partial identification intervals in one way or another. Again, we would refer readers to Baiocchi et al. (2014) and Swanson et al. (2018) for bounds other than those considered above. We would like to point out that IV identification assumptions are typically not verifiable (although they might lead to testable implications), and depend largely on expert knowledge. Moreover, certain assumptions may be inappropriate in the context of ITR estimation problems, for example assumptions that largely restrict treatment heterogeneity, and should be made with caution.

Remark 3 (Continuous but bounded \(Y\)) When \(Y\) is continuous, partial identification bounds on \(Y\) require additional assumptions that bound the support of \(Y\). In Supplementary Material A, we
further review assumptions and partial identification results that allow a valid IV to partially identify the counterfactual mean (and hence the ATE and CATE) of a continuous but bounded outcome.

4 AN IV-PARTIAL IDENTIFICATION LEARNING (IV-PILE) APPROACH TO ESTIMATING OPTIMAL ITRS: IV-OPTIMALITY, RISK AND OPTIMIZATION

4.1 IV-optimality

Without loss of generality, we assume $\Delta = 0$. Let $f(\cdot) : \mathcal{X} \mapsto \mathbb{R}$ be a discriminant function and $\text{sgn}(f(\cdot))$ a decision rule to be learned. Recall that the risk function to be minimized in an ITR estimation problem is

$$R(f) = \mathbb{E} \left[ |C(X)| \cdot 1 \{ \text{sgn}(f(X)) \neq \text{sgn}(C(X)) \} \right].$$

As has been argued extensively, this optimal rule is in general not identifiable when the collected observed covariates $X$ cannot adequately address the confounding between the treatment and outcome.

To proceed, we define a new notion of optimality and a new estimand to target.

**Definition 1** IV-optimality A treatment rule $f(\cdot) \in \mathcal{F}$ is said to be IV-optimal if it is optimal with respect to the putative IV and assumption set $\mathcal{C}$ in the following sense:

$$f = \arg\min_{f \in \mathcal{F}} R_{\text{upper}}(f; L(\cdot), U(\cdot))$$

$$= \arg\min_{f \in \mathcal{F}} \mathbb{E} \left[ \sup_{C'(X) : L(X) \leq C' \leq U(X)} |C'(X)| \cdot 1 \{ \text{sgn}(f(X)) \neq \text{sgn}(C'(X)) \} \right],$$

where $[L(X), U(X)]$ is the partial identification interval under the putative IV and identification assumption set $\mathcal{C}$.

Proposition 1 asserts that $\mathbb{E}[\cdot]$ and sup operators in Definition 1 are exchangeable.

**Proposition 1**

$$f = \arg\min_{f \in \mathcal{F}} \mathbb{E} \left[ \sup_{C'(X) : L(X) \leq C' \leq U(X)} |C'(X)| \cdot 1 \{ \text{sgn}(f(X)) \neq \text{sgn}(C'(X)) \} \right]$$

$$= \arg\min_{f \in \mathcal{F}} \sup_{C'(\cdot) : L(\cdot) \leq C' \leq U(\cdot)} R(f; C'(\cdot)),$$

where $f_1 \leq f_2$ denotes $f_1(x) \leq f_2(x)$ for all $x$.

**Proof** All proofs in the article are in Supplementary Materials C, D and E.

**Remark 4** The risk function $R_{\text{upper}}(f; L(\cdot), U(\cdot))$ considered in Definition 1 represents the expected worst-case weighted misclassification error among all $C'(X)$ compatible with $L(X)$ and $U(X)$ informed by the putative IV and IV identification assumptions. $R_{\text{upper}}(f; L(\cdot), U(\cdot))$ is a natural upper bound on the risk $R(f)$. Proposition 1 further shows that $f$ in Definition 1 can
be understood as a min-max estimate.

**Remark 5** When \( C(\cdot) = L(\cdot) = U(\cdot) \), \( R_{\text{upper}}(f) \) would reduce to \( R(f) \), and IV-optimality reduces to the usual notation of optimality considered in Zhang et al. (2012a), Zhao et al. (2012), Cui and Tchetgen Tchetgen (2020), and Qiu et al. (2020).

This new optimality criterion has at least three desirable features. First, it is always well-defined for any valid IV and under minimal IV identification assumptions. Second, it facilitates using IV identification assumptions as ‘leading cases, not truths’ (Tukey, 1986, p. 72). According to Tukey, a statistical procedure is ‘safe’ if it is valid in a wide range of scenarios. The statistical procedure targeting the ‘IV-optimal’ rule is therefore ‘safe’ in the sense that the estimand is well-defined and can be learned under a wide range of IV identification assumptions and mild modelling assumptions. In sharp contrast, Cui and Tchetgen Tchetgen (2020) and Qiu et al. (2020) aimed at learning the optimal ITR with an IV; though the optimal ITR is always well-defined, it cannot be learned even with a valid IV unless some often stringent IV identification assumptions are met. Third and perhaps most importantly, the notion of ‘IV-optimality’ leaves to IV identification assumptions ‘the task of stringency’ (Tukey, 1986, p. 72), and captures the intuition that the quality of the estimated ITR should depend on the quality of the instrumental variable. According to Definition 1 and Proposition 1, an ‘IV-optimal’ ITR is more stringent, in the sense that it is ‘closer’ to the true underlying optimal ITR and has smaller risk and better generalization performance if the putative IV together with IV identification assumptions help narrow down the partial identification intervals. This is the case, for instance, when the putative IV is a very strong one and the compliance rate is very high, or when assumptions in addition to the core IV assumptions, for example the correct non-compliant decision assumption (IV.A5), apply to the putative IV. We study more closely the risk of an IV-optimal ITR in Section 4.2.

**Remark 6** Multiple IVs and weak IVs In many empirical studies, researchers have multiple putative IVs, for example excess tuition and excess distance in a study of the effect of community college on educational attainment (Rouse, 1995), and it is often unclear which one of these IVs, or if any of them, satisfies the point identification assumptions in order to identify the optimal ITR. However, these multiple IVs can be used to estimate their respective ‘IV-optimal’ ITRs, possibly under different, IV-specific, identification assumptions and the quality of each resulting ‘IV-optimal’ ITR depends on how much each of these multiple IVs can narrow down the partial identification intervals and pinpoint the CATE. Multiple IVs can even be combined into a single stronger IV, and this stronger IV is likely to yield an ‘IV-optimal’ ITR that is more stringent and has better generalization performance compared to using any of the multiple IVs alone. On the other hand, if researchers only have a very weak IV, the corresponding partial identification intervals may be excessively long and non-informative, and as a result, the ‘IV-optimal’ ITR may be far from the optimal ITR in its generalization performance. Indeed, with a weak IV, researchers should expect little information to be learned about the treatment effect and perhaps the wisest thing to do is switching to a stronger IV.

**Remark 7** Although not the primary focus of this paper, one can directly minimize the expectation in Definition 1 in a pointwise manner by estimating \( L(X) \) and \( U(X) \), just like one can estimate \( C(X) \) and then take \( \text{sgn} \{ \hat{C}(X) \} \) to be the estimated optimal ITR in non-IV settings. These methods are called *indirect methods* in the literature as they indirectly specify the form of the optimal ITR through postulated models for various aspects of \( C(X) \) in non-IV settings (Zhao et al., 2019), and \( L(X) \) and \( U(X) \) in our setting. In Supplementary Material G, we construct simple plug-in estimators for an IV-optimal ITR based on this idea, and prove that this straightforward plug-in estimator is in fact minimax optimal. We pursue a classification perspective as in Zhang et al.
(2012a) and Zhao et al. (2012) here rather than the indirect methods because of the following consideration. In many practical scenarios, we would like to have control over the complexity of the estimated ITR. This in general cannot be fulfilled by indirect methods unless we specify some simple models to estimate $L(\mathbf{X})$ and $U(\mathbf{X})$ in the first place; however, $L(\mathbf{X})$ and $U(\mathbf{X})$ are unlikely to admit simple parametric forms in our settings as they are complicated combinations of maxima and/or minima of many conditional probabilities (see Section 3). On the other hand, if we use flexible machine learning tools to estimate conditional probabilities involved in $L(\mathbf{X})$ and $U(\mathbf{X})$, the corresponding ITR is often complicated and lacks interpretability. It may also suffer from the problem of overfitting (Wang et al., 2016; Zhao et al., 2012; Zhao et al., 2019). These considerations motivate us to adopt the classification perspective as in Zhang et al. (2012a) and Zhao et al. (2012). A great appeal of the classification perspective is that by decoupling the task of estimating $L(\mathbf{X})$ and $U(\mathbf{X})$ from that of estimating the IV-optimal rule, the estimated ITR no longer suffers from the aforementioned problems. In principle, one can leverage flexible machine learning tools to estimate relevant conditional probabilities and hence $L(\mathbf{X})$ and $U(\mathbf{X})$, while still learning a parsimonious IV-optimal ITR within a pre-specified function class, for example the class of linear functions.

### 4.2 Bayes decision rule and Bayes risk

Define the Bayes risk $R_{\text{upper}}^* = \inf f R_{\text{upper}}(f)$, where infimum is taken over all measurable functions. A decision rule $f$ is called a Bayes decision rule if it attains the Bayes risk, that is, $R_{\text{upper}}(f^*) = R_{\text{upper}}^*$. Proposition 2 gives a representation of the Bayes decision rule.

**Proposition 2** Consider the risk function $R_{\text{upper}}(f)$ defined in Definition 1. Let

$$
\eta(\mathbf{x}) = |U(\mathbf{x})| \cdot I\{L(\mathbf{x}) > 0\} - |L(\mathbf{x})| \cdot I\{U(\mathbf{x}) < 0\} + (|U(\mathbf{x})| - |L(\mathbf{x})|) \cdot I\{|L(\mathbf{x}), U(\mathbf{x})| \geq 0\}.
$$

Consider a decision rule $f^*(\mathbf{x})$ such that

$$
\text{sgn}\{f^*(\mathbf{x})\} = \text{sgn}\{\eta(\mathbf{x})\} = \begin{cases} +1, & \text{if } \eta(\mathbf{x}) \geq 0, \\ -1, & \text{if } \eta(\mathbf{x}) < 0. \end{cases}
$$

Let $R_{\text{upper}}^* = R_{\text{upper}}(f^*)$. $f^*$ is the Bayes decision rule and $R_{\text{upper}}^*$ is the Bayes risk such that $R_{\text{upper}}(f) \geq R_{\text{upper}}^*$, \forall f measurable.

Proposition 3 further derives the excess risk of a measurable decision rule $f$.

**Proposition 3** For any measurable decision rule $f$, its excess risk is

$$
R_{\text{upper}}(f) - R_{\text{upper}}^* = \mathbb{E}\left[ I\{\text{sgn}\{f(\mathbf{X})\} \neq \text{sgn}\{f^*(\mathbf{X})\}\} \cdot |\eta(\mathbf{X})| \right],
$$

where $\eta(\mathbf{x})$ is defined in Proposition 2.
4.3 Risk of IV-optimal rules

An IV-optimal rule targets $R_{\text{upper}}$. What can be said about the risk of an IV-optimal rule? Proposition 4 provides insight into this important question.

Proposition 4 Risk of IV-optimal rules

1. For any measurable $f$, we have

$$0 \leq R_{\text{upper}}(f) - R(f) \leq \mathbb{E} \left[ U(X) - L(X) \right].$$

2. Let $f^*$ be the Bayes decision rule targeting $R_{\text{upper}}$ in Proposition 2 such that $R_{\text{upper}}^* = R_{\text{upper}}(f^*)$. The risk of $f^*$ satisfies

$$R(f^*) = \mathbb{E} \left[ |C(X)| \cdot 1 \{ \text{sgn}(C(X)) \neq \text{sgn}(f^*(X)) \} \right]$$

$$\leq \mathbb{E} \left[ 1 \{ L(X) < 0 < U(X) \} \cdot \left\{ U(X) - L(X) \right\} \cdot \left\{ \frac{1 - \rho(X; U, L)}{2} \right\} \cdot 1 \{ \rho'(X; U, L, C) > \rho(X; U, L) \} \right],$$

where $\rho(x; U, L) = \frac{|U(x) + L(x)|}{2}$ and $\rho'(x; U, L, C) = \frac{|C(x) - (U(x) + L(x))/2|}{|U(x) - L(x)|/2}$.

The first part of Proposition 4 states that for any decision rule $f$, $R_{\text{upper}}(f)$ is no larger than the risk $R(f)$ by a margin of $\mathbb{E} \left[ U(X) - L(X) \right]$. From this result, it is transparent that as $U(x)$ converges uniformly to $L(x)$, the gap between $R_{\text{upper}}(f)$ and $R(f)$ goes to 0 uniformly in $f$.

The second part of the proposition further states that if $f^*$ is a Bayes decision rule for $R_{\text{upper}}$ that attains the Bayes risk $R_{\text{upper}}^*$, its generalization error $R(f^*)$ is upper bounded by the expectation of the product of four terms, each of which bears its own meaning. Fix $x \in X$ and the partial identification interval $[L(x), U(x)]$. The first term $I = 1 \{ L(x) < 0 < U(x) \}$ measures if the interval $[L(x), U(x)]$ covers 0. If $L(x), U(x)$ does not cover 0, such an $x$ would not contribute to the risk of $f^*$. The second term $II=U(x)−L(x)$ measures the length of the interval. Not surprisingly, if the interval $[L(x), U(x)]$ covers 0, the narrower it is, the less it would contribute to the risk of $f^*$. The third term $III=1−\rho(x; U, L)/2$ measures how symmetric $[L(x), U(x)]$ is about 0. Suppose that $[L(x), U(x)]$ is such that $L(x) < 0 < U(x)$, that is, $[L(x), U(x)]$ covers 0 (the first term I is 1) and has a non-trivial interval length (the second term II is not 0). If $[L(x), U(x)]$ is symmetric about 0, that is, $L(x)=-U(x)$, $\rho(x; U, L)$ would attain its minimum at 0; on the other hand, $\rho(x; U, L)$ could be arbitrarily close to 1 if either $L(x)$ is arbitrarily close to 0 from the left, or $U(x)$ is arbitrarily close to 0 from the right. In other words, $\rho(x; U, L)$ (and hence the third term III) measures the skewness of the interval $[L(x), U(x)]$ with respect to 0. The fourth term IV measures how symmetric $[L(x), U(x)]$ is about $C(x)$, relative to how symmetric the interval is around 0. Observe that $\rho^c(x; U, L, C)$ is analogous to $\rho(x; U, L)$, except that $\rho^c(x; U, L, C)$ measures the skewness of the interval $[L(x), U(x)]$ with respect to $C(x)$. $\rho^c(x; U, L, C)$ would attain its minimum at 0 if the interval is symmetric about $C(x)$, and its maximum at 1 if the interval barely covers $C(x)$, that is, $L(x)=C(x)$ or $U(x)=C(x)$. Therefore, if the interval is more symmetric about $C(x)$ than it is about 0, the fourth term IV is 0; otherwise, it is 1. To conclude, the risk of $f^*$ is small if with high probability, the partial identification interval does not cover 0, is short in length,
is asymmetric about 0, and is more symmetric about \( C(\mathbf{x}) \) than about 0. This upper bound on the risk of \( f^* \) can also be understood as the maximum gap between the generalization performance of \( f^* \) and an optimal ITR. Figure 1 summarizes the above discussion with a graphical illustration.

### 4.4 Risk decomposition, structural risk minimization and surrogate loss

The risk function \( R_{\text{upper}}(f) \) can be decomposed into two parts: \( R_{\text{label, upper}}(f) \), corresponding to the risk associated with the labelled part, and \( R_{\text{unlabel, upper}}(f) \), corresponding to that associated with the unlabelled part:

\[
R_{\text{upper}}(f) = E \left[ \max_{C'(\mathbf{X}) \in [L(\mathbf{X}), U(\mathbf{X})]} |C'(\mathbf{X})| \cdot 1\{\sgn\{f(\mathbf{X})\} \neq \sgn\{C'(\mathbf{X})\}\} \right] = E \left[ |U(\mathbf{X})| \cdot 1\{\sgn\{f(\mathbf{X})\} \neq 1\} \cdot 1\{L(\mathbf{X}) > 0\} \right] + E \left[ |L(\mathbf{X})| \cdot 1\{\sgn\{f(\mathbf{X})\} 
eq -1\} \cdot 1\{U(\mathbf{X}) < 0\} \right] + E \left[ \max \left\{ |U(\mathbf{X})| \cdot 1\{\sgn\{f(\mathbf{X})\} \neq 1\}, |L(\mathbf{X})| \cdot 1\{\sgn\{f(\mathbf{X})\} \neq -1\} \right\} \cdot 1\{[L(\mathbf{X}), U(\mathbf{X})] \ni 0\} \right]
\]

\[
R_{\text{label, upper}}(f) + R_{\text{unlabel, upper}}(f)
\]

= \( R_{\text{label, upper}}(f) + R_{\text{unlabel, upper}}(f) \).

**Remark 8** It may be tempting to replace \( \max_{C'(\mathbf{X}) \in [L(\mathbf{X}), U(\mathbf{X})]} \) with \( \min_{C'(\mathbf{X}) \in [L(\mathbf{X}), U(\mathbf{X})]} \) in Definition 1; however, the definition would then become vacuous as is easily seen from Equation (6). Fix \( \mathbf{x} \in \mathcal{X} \) such that \( [L(\mathbf{x}), U(\mathbf{x})] \ni 0 \). The risk conditional on \( \mathbf{x} \) is always minimized by letting \( \sgn\{f(\mathbf{x})\} = \sgn\{C'(\mathbf{x})\} \); however, since \( [L(\mathbf{x}), U(\mathbf{x})] \ni 0 \) and \( C'(\mathbf{x}) \in [L(\mathbf{x}), U(\mathbf{x})] \), \( \sgn\{C'(\mathbf{x})\} \) can be either +1 or −1, suggesting that the risk conditional on \( \mathbf{x} \) is always 0 no matter what value \( f(\mathbf{x}) \) takes on. In other words, \( R_{\text{unlabel, upper}}(f) = 0 \ \forall f \), with max replaced with min in Equation (6), and unlabelled data become superfluous.
Decomposition (6) motivates estimating \( f(\cdot) \in \mathcal{F} \) using the following structural risk minimization approach (Vapnik, 1992):

\[
\hat{f}(\cdot) = \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{l} \hat{U}(X_i) \cdot \mathbb{1}\{\sgn(f(X_i)) \neq 1\} \cdot \mathbb{1}\{\hat{L}(X_i) > 0\} \\
+ \mathbb{1}\{-\hat{L}(X_i)\} \cdot \mathbb{1}\{\sgn(f(X_i)) = -1\} \cdot \mathbb{1}\{\hat{U}(X_i) < 0\} \\
+ \sum_{i=l+1}^{n} \max \left[ \hat{U}(X_i) \cdot \mathbb{1}\{\sgn(f(X_i)) \neq 1\}, -\hat{L}(X_i) \cdot \mathbb{1}\{\sgn(f(X_i)) = -1\} \right] \\
+ \frac{n \lambda_n}{2} \|f\|^2,
\]

where \([\hat{L}(X_i), \hat{U}(X_i)]\) is an estimated partial identification interval of \([L(X_i), U(X_i)]\). \(\|\cdot\|\) denotes some norm of \(f(\cdot)\), and \(\lambda_n\) is a possibly data-dependent tuning parameter that penalizes the norm of \(f(\cdot)\) to reduce overfitting. For instance, if we assume that \(f(\cdot)\) resides in a reproducing kernel Hilbert space (RKHS) \(\mathcal{H}_K\), then \(\|\cdot\|\) corresponds to the norm associated with \(\mathcal{H}_K\). The complexity of \(f(\cdot)\) is restricted by penalizing its norm.

It is well known in machine learning and optimization literature that directly minimizing the empirical risk as above is difficult due to the non-continuity and non-convexity of the indicator function, and it is customary to rewrite the loss function by replacing the 0-1-based loss with a convex upper bound. Table 1 summarizes the original 0-1-based loss and our choice of surrogate loss corresponding to each \([L(x), U(x)]\) configuration. Supplementary Material B.1 further plots the original loss and the corresponding surrogate loss in each case. Observe that the surrogate loss is indeed a continuous convex upper bound of the original discontinuous loss function in all cases. Moreover, it can be shown that our designed surrogate loss function is continuous in both \(L\) and \(U\) values.

Remark 9 When \([L(x), U(x)] > 0\) or \([L(x), U(x)] < 0\), the surrogate loss is a scaled hinge loss. When \([L(x), U(x)] \ni 0\), the surrogate loss is a lifted and scaled hinge loss.

Let \(\phi(x) = (1-x)^+\). Under the surrogate loss, the objective function (7) becomes:

\[
\hat{f}(\cdot) = \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{l} \left[ \hat{U}(X_i) \cdot \phi(f(X_i)) \cdot \mathbb{1}\{\hat{L}(X_i) > 0\} \right. \\
+ \left. \mathbb{1}\{-\hat{L}(X_i)\} \cdot \phi(-f(X_i)) \cdot \mathbb{1}\{\hat{U}(X_i) < 0\} \right] \\
+ \sum_{i=l+1}^{n} \left[ \left\{ \hat{U}(X_i) \right\} + \left\{ (\hat{L}(X_i) - |\hat{L}(X_i)|) \cdot \phi(f(X_i)) \right\} \right] \\
\cdot \mathbb{1}\{|\hat{L}(X_i)| > |\hat{U}(X_i)|\} \\
+ \left[ \hat{U}(X_i) + (|\hat{L}(X_i)| - |\hat{U}(X_i)|) \cdot \phi(-f(X_i)) \right] \\
\cdot \mathbb{1}\{|\hat{L}(X_i)| > |\hat{U}(X_i)|\} \\
\left. + \frac{n \lambda_n}{2} \|f\|^2 \right].
\]

| \([L(x), U(x)]\) | Original loss | Surrogate loss |
|-----------------|-------------|---------------|
| \([L(x), U(x)] > 0\) | \(L(x) \cdot \mathbb{1}\{\sgn(f(x)) \neq 1\}\) | \(L(x) \cdot (1-f(x))^+\) |
| \([L(x), U(x)] < 0\) | \(L(x) \cdot \mathbb{1}\{\sgn(f(x)) = -1\}\) | \(L(x) \cdot (1+f(x))^+\) |
| \([L(x), U(x)] \ni 0\), \(|L(x)| \geq |U(x)|\) | \(\max(L(x) \cdot \mathbb{1}\{\sgn(f(x)) \neq 1\}, -L(x) \cdot \mathbb{1}\{\sgn(f(x)) = -1\})\) | \(L(x) + (|L(x)| - |U(x)|) \cdot (1-f(x))^+\) |
| \([L(x), U(x)] \ni 0\), \(|L(x)| < |U(x)|\) | \(\max(L(x) \cdot \mathbb{1}\{\sgn(f(x)) \neq 1\}, -L(x) \cdot \mathbb{1}\{\sgn(f(x)) = -1\})\) | \(L(x) + (|L(x)| - |U(x)|) \cdot (1+f(x))^+\) |
Let $R_{\text{upper}}(f)$ denote the risk associated with the surrogate loss, $f^*_h$ the Bayes decision rule that minimizes $R_{\text{upper}}^h(f)$, and $R_{\text{upper}}^{h,*}(f)$ the corresponding Bayes risk. Theorem 1 establishes a relationship between $R_{\text{upper}}(f) - R_{\text{upper}}^*$ and $R_{\text{upper}}^h(f) - R_{\text{upper}}^{h,*}$, so that we can transfer assessments of statistical error in terms of the excess risk $R_{\text{upper}}^h(f) - R_{\text{upper}}^{h,*}$ into assessments of error in terms of $R_{\text{upper}}(f) - R_{\text{upper}}^*$, the excess risk of genuine interest (Bartlett et al., 2006).

**Theorem 1** For any measurable function $f$, we have

$$R_{\text{upper}}(f) - R_{\text{upper}}^* \leq R_{\text{upper}}^h(f) - R_{\text{upper}}^{h,*}. \quad (9)$$

Theorem 1 reassures us that using the surrogate loss displayed in Table 1 does not hinder the search for a function that achieves the optimal Bayes risk $R_{\text{upper}}^*$, and it is appropriate to employ surrogate-loss-based computationally efficient algorithms. In Supplementary Materials B.2 and B.3, we derive linear/non-linear $\hat{f}$ when $f(\cdot)$ is in a reproducing kernel Hilbert space and show that the associated optimization problem can be transformed into a particular instance of weighted SVM (Vapnik, 2013) and readily solved using standard solvers.

4.5 | IV-PILE algorithm

Before delving into theoretical properties, we summarize the IV-PILE algorithm in Algorithm 1.

| Algorithm 1: Pseudo Algorithm for IV-PILE |
|-------------------------------------------------|
| **Input:** $\{(X_i, Z_i, A_i, Y_i), i = 1, \cdots, n\}$ and IV identification assumption set $C$; |
| o Obtain appropriate estimates of $L(X_i)$ and $U(X_i)$, denoted as $\hat{L}_i$ and $\hat{U}_i$, under IV identification assumption set $C$. Parametric models or more flexible and expressive estimators like random forests can be used. |
| o Compute the label $\hat{\varepsilon}_i \in \{-1, +1\}$ associated with each observation: |
| $\hat{\varepsilon}_i = 1\{\hat{U}_i < 0\} - 1\{\hat{L}_i > 0\} - \text{sgn}(|\hat{U}_i| - |\hat{L}_i|) \cdot 1\{[\hat{L}_i, \hat{U}_i] \ni 0\}$, |
| for $i = 1, \cdots, n$; |
| o Compute the weight $\hat{\omega}_i$ associated with each observation: |
| $\hat{\omega}_i = |\hat{U}_i| \cdot 1\{\hat{L}_i > 0\} + |\hat{L}_i| \cdot 1\{\hat{U}_i < 0\} + ||\hat{U}_i| - |\hat{L}_i|| \cdot 1\{[\hat{L}_i, \hat{U}_i] \ni 0\}$, |
| for $i = 1, \cdots, n$; |
| o Solve a weighted SVM problem with labels and weights computed in Step 2 and 3 using a Gaussian kernel. Let $\hat{f}$ be the solution; |
| o Return $\hat{f}$. |

5 | THEORETICAL RESULTS

5.1 | IV-PILE estimator via sample splitting

To facilitate theoretical analysis of the IV-PILE estimator, we study an alternative sample-splitting estimator that is very close to the IV-PILE estimator. Let $I_1$, $I_2$ denote an equal-size mutually exclusive
random partition of indices \( \{1, \ldots, n\} \) such that \( |I_1| > |I_2| > n/2 \). Samples with indices in \( I_1 \) are used to construct estimates \( \widehat{L}(x) \) and \( \widehat{U}(x) \) for functions \( L(x) \) and \( U(x) \). We then plug \( \widehat{L} \) and \( \widehat{U} \) into expressions for the weight \( w \) and label \( e \) (see Algorithm 1 and Supplementary Material B.2) to construct \( \widehat{w}(\cdot) \) and \( \widehat{\epsilon}(\cdot) \), and use the other half of samples to obtain the following IV-PILE estimator:

\[
\widehat{f}^h_n = \arg\min_{f \in \mathcal{F}} \frac{1}{|I_2|} \sum_{i \in I_2} \left\{ 1 + \widehat{\epsilon}(x_i) \cdot f(x_i) \right\}^+ + \frac{\lambda_n}{2} \|f\|^2.
\]

Sample splitting here helps remove the dependence between estimating \( \widehat{w} \) and \( \widehat{\epsilon} \) and constructing the IV-PILE estimator, which in turn helps weaken the assumptions needed to establish convergence rate results by getting rid of the entropy conditions on \( L(x) \) and \( U(x) \)’s function classes. Similar sample splitting technique can be also found in Bickel (1982), Zheng and van der Laan (2011), Chernozhukov et al. (2018), Robins et al. (2017), and Zhao et al. (2019), among many others.

### 5.2 Theoretical properties

Let \( \inf_{f \in \mathcal{F}} R_{\text{upper}}^h(f) \) denote the minimal risk among rules in \( \mathcal{F} \) and define the approximation error incurred by optimizing over \( \mathcal{F} \) as \( A(\mathcal{F}) = \inf_{f \in \mathcal{F}} R_{\text{upper}}^h(f) - R_{\text{upper}}^h(f) \). In this section we establish the convergence rate properties of \( R_{\text{upper}}(\hat{f}^h_n) - \inf_{f \in \mathcal{F}} R_{\text{upper}}^h(f) \). We consider the following assumptions.

**Assumption 1** Existence of a finite minimizer

\[
\exists f^*_h \in \mathcal{F} \text{ s.t. } R_{\text{upper}}^h(f^*_h) = \inf_{f \in \mathcal{F}} R_{\text{upper}}^h(f).
\]

**Assumption 2** Boundedness conditions I

\[
\exists M_1 > 0 \text{ s.t. } |L(X)|, |U(X)|, |Y| \leq M_1 \text{ with probability 1,}
\]

\[
\exists M_2 \text{ s.t. } \forall i, |X[i]| \leq M_2 \text{ with probability 1.}
\]

**Assumption 3** Boundedness conditions II

Assume that the estimates of \( L \) and \( U \), that is, \( \hat{L} \) and \( \hat{U} \), satisfy

\[
\exists M_3 \text{ s.t. } |\hat{L}(X)|, |\hat{U}(X)| \leq M_3 \text{ with probability 1.}
\]

For any \( \epsilon > 0 \), let \( N(\epsilon, \mathcal{F}, L_2(P)) \) denote the covering number of \( \mathcal{F} \), that is \( N(\epsilon, \mathcal{F}, L_2(P)) \) is the minimal number of closed \( L_2(P) \)-balls of radius \( \epsilon \) that is required to cover \( \mathcal{F} \). We consider the following assumption on entropy condition:

**Assumption 4** Entropy condition

There exists a subclass \( \mathcal{F}_0 \) of \( \mathcal{F} \) and constants \( C_0 > 0, 0 < \nu < 2 \) such that \( \mathcal{F} = \{ a \cdot f \in \mathcal{F}_0, a \in \mathbb{R} \} \), the minimizer \( \hat{f}^h_n \) is in \( (C_0 / \lambda_n) \cdot \mathcal{F}_0 = \{ a \cdot f \in \mathcal{F}_0, a \in \mathbb{R}, |a| \leq C_0 \cdot \lambda_n^{-1} \} \) with probability 1, and the \( \sup_{P} \log N(\epsilon, \mathcal{F}_0, L_2(P)) \leq O(\epsilon^{-\nu}) \) for all \( 0 < \epsilon < 1 \), where the supremum is taken over all probability measures \( P \).
Assumption 1 says that there exists an $f$ with finite norm that minimizes the risk over $P$. This is a standard assumption made in the statistical learning literature. Assumption 2 requires that $L$, $U$, and $Y$ are bounded, and coordinates of observed covariates $X$ are bounded with probability 1. When $Y$ is a binary outcome, for example mortality as in the NICU study, $L$ and $U$ obtained via the Balke–Pearl bound and the Siddique bound are trivially bounded. When $Y$ is continuous, the partial identification literature typically requires $Y$ to be bounded (Swanson et al., 2018), and the partial identification interval endpoints $L$ and $U$ are therefore also bounded. Boundedness of $Y$ is reasonable for many health outcomes, for example length of stay in hospital, the cholesterol level, etc. Assumption 3 says that estimates $\hat{L}$ and $\hat{U}$ are bounded when $L$ and $U$ are bounded. This holds for any reasonable estimates of $L$ and $U$. Finally, the assumption on entropy condition is satisfied for many popular RKHS, for example the RKHS induced by the Gaussian kernel $k(x, x') := \exp(-\|x - x'\|^2 / \sigma^2)$, where we take $P_0$ to be the convex hull of $\{f_0(x) = k(x, x'), x' \in \mathcal{X}\}$, the minimizer $f_n$ takes an explicit form that satisfies the specified condition (see Supplementary Material B), and the covering number condition follows from Kosorok (2007) Corollary 9.5).

**Assumption 5** Rate of convergence of $L$ and $U$

Assume that $\hat{L}$ and $\hat{U}$ converge to $L$ and $U$ at the following rates:

$$
\mathbb{E}\left[|\hat{L}(X) - L(X)|\right] = O(n^{-\alpha}),
$$

$$
\mathbb{E}\left[|\hat{U}(X) - U(X)|\right] = O(n^{-\beta}).
$$

Consider a binary outcome and the associated Balke–Pearl bound. To obtain estimates $\hat{L}$ and $\hat{U}$ that satisfy Assumption 5, we first estimate $\{p_{y,a|z,X}: y = \pm 1, a = \pm 1, z = \pm 1\}$ (let the estimates be $\{\hat{p}_{y,a|z,X}: y = \pm 1, a = \pm 1, z = \pm 1\}$) and then plug these estimates into Equations (2) and (3) to obtain $\hat{U}$ and $\hat{L}$. In Supplementary Material D.2, we show that if $K$ functions are all $n^{-\theta}$ estimable then their linear combinations, maximum, and minimum are also $n^{-\theta}$ estimable. $L(X)$ and $U(X)$ in the Balke–Pearl bound are both maximum/minimum of a series of linear combinations of $p_{y,a|z,X}$; hence, if we have the following condition hold

**Condition 1** Convergence of conditional probabilities

$$\exists \theta, \mathbb{E}\left[|\hat{p}_{y,a|z,X} - p_{y,a|z,X}|\right] = O(n^{-\theta}), \quad y = \pm 1, a = \pm 1, z = \pm 1,$$

then we can deduce that Assumption 5 holds for $\alpha = \beta = \theta$. Condition 1 holds in many scenarios. For instance, if we fit parametric models to estimate $p_{y,a|z,X}: y = \pm 1, a = \pm 1, z = \pm 1$, and models are correctly specified, then this condition holds for $\theta = \frac{1}{2}$. We can also use flexible and expressive non-parametric regression methods to estimate functions $p_{y,a|z,X}$. Assuming that functions $\{p_{y,a|z,X}: y = \pm 1, a = \pm 1, z = \pm 1\}$ are in a H"older ball with smoothness parameter $\alpha$ and a constant radius, then Condition 1 holds for $\theta = -\frac{\alpha}{d + 2\alpha}$, where $d$ is the dimension of $X$, when $\{p_{y,a|z,X}: y = \pm 1, a = \pm 1, z = \pm 1\}$ are estimated via wavelets (Donoho et al., 1998; Cai et al., 2012) or a variant of the random forests algorithm known as Mondrian forests (Mourtada et al., 2018). Similar results hold when we use Siddique bounds for a binary outcome and Manski–Pepper bounds for a continuous but bounded outcome; see Supplementary Material D.2 for details.

**Assumption 6** Norm condition

There exists a constant $M_4$ s.t. for any $f \in P$: 
\[ \|f\| \geq M_4 \|f\|_\infty. \]

Assumption 6 guarantees that the sup-norm of the minimizer \( \hat{f}_n \) can be controlled by penalizing \( f \).

Under Assumption 1–6, it can be shown that \( \|\hat{f}_n\|_\infty \leq \frac{2\sqrt{M_2 \sqrt{M_3}}}{M_2 \sqrt{M_3}} \). Define \( B_n \) to be the set of functions \( f \in \mathcal{F} \) and \( \|f\|_\infty \leq \frac{2\sqrt{M_2 \sqrt{M_3}}}{M_2 \sqrt{M_3}} \), and \( B^* \) to be the intersection of \( B_n \) and \( (C_0 / \lambda_n) \cdot \mathcal{F}_0 \). Lemmas 1 and 2 below develop properties of functions in \( B_n \) and \( B^*_n \), respectively. From now on, we consider \( \lambda_n = o(1) \). We use \( \mathbb{E}_Z \) to denote expectation with respect to the random variable \( Z \), and \( \mathbb{E} \) with respect to all random variables.

**Lemma 1** Let

\[
\begin{align*}
    w_i &= |U_i| \cdot \mathbb{1} \{ L_i > 0 \} + |L_i| \cdot \mathbb{1} \{ U_i < 0 \} + |L_i| \cdot \mathbb{1} \{ L_i \leq 0 \} \cdot \mathbb{1} \{ L_i, U_i \in B \}, \\
    e_i &= \mathbb{1} \{ U_i < 0 \} - \mathbb{1} \{ L_i > 0 \} - \text{sgn} \{ |U_i| - |L_i| \} \cdot \mathbb{1} \{ L_i, U_i \in B \}, \\
\end{align*}
\]

and

\[
    l(x; w, e, f) = w(x) \{ 1 + e(x)f(x) \}^+, \]

where \( L_i = L(X_i) \), and \( U_i = U(X_i) \), and \( \hat{w} \) and \( \hat{e} \) be defined in Algorithm 1. We have

\[
    \mathbb{E}_{X_{(i_1)}} \sup_{f \in B_n} \left| \mathbb{E}_{X_{(i_2)}} \left\{ \frac{1}{|I_2|} \sum_{i \in I_2} l(X_i; w, e, f) - \frac{1}{|I_2|} \sum_{i \in I_2} l(X_i; \hat{w}, \hat{e}, f) \right\} \right| \leq O \left( n^{-(a + \beta) / \sqrt{\lambda_n}} \right). \tag{10}
\]

where \( X_{(i_1)} \) and \( X_{(i_2)} \) denote \( \{ X_i, i \in I_2 \} \) and \( \{ X_i, i \in I_1 \} \), respectively.

Function \( l(\cdot; w, e, f) \) in Lemma 1 denotes the loss function where \( w \) and \( e \) are set at truth, and \( l(\cdot; \hat{w}, \hat{e}, f) \) the loss function where \( w \) and \( e \) are estimated. Lemma 1 effectively bounds the risk induced by estimating \( w(\cdot) \) and \( e(\cdot) \). Lemma 2 below further quantifies the risk induced by estimating the risk function using its empirical analogue.

**Lemma 2** Let \( w(\cdot) \) and \( e(\cdot) \) be defined as in Lemma 1. We have

\[
    \mathbb{E}_{X_{(i_1)}} \left| \mathbb{E}_{X_{(i_2)}} \left| \mathbb{E} l(X; w, e, f) - \sum_{i \in I_2} \frac{1}{|I_2|} l(X_i; w, e, f) \right| \right| \leq O \left( \frac{1}{\sqrt{n} \cdot \lambda_n} \right). \tag{11}
\]

Lemmas 1 and 2 facilitate the derivation of the convergence rate of \( R_{\text{upper}}^{\hat{f}_n} \), as is formally stated in Theorem 2.

**Theorem 2** Assume that Assumption 1 to 6 hold. We have

\[
    R_{\text{upper}}^{\hat{f}_n} - \inf_{f \in \mathcal{F}} R_{\text{upper}}^{f} \leq O(\lambda_n + n^{-\frac{1}{2}} \lambda_n^{-1} + \lambda_n^{-\frac{3}{2}} (n^{-a} + n^{-\beta})). \tag{12}
\]
Theorem 2 gives concrete upper bounds of $R_{upper}^{b} (\hat{f}_{n}^{\lambda}) - \inf_{f \in \mathcal{F}} R_{upper}^{b} (f)$, which implies that for a wide range of $\lambda_{n}$ satisfying $\frac{1}{n^{\frac{1}{2} + 2a + 2b}} \leq \lambda_{n} \leq \frac{1}{\ln(\ln(n))}$, as $n$ goes to infinity, $R_{upper}^{b} (\hat{f}_{n}^{\lambda})$ converges to $\inf_{f \in \mathcal{F}} R_{upper}^{b} (f)$, the minimal risk over $\mathcal{F}$, and $R_{upper}^{b} (\hat{f}_{n}^{\lambda}) - R_{upper}^{b} (f)$ converges to approximation error $\mathcal{A} (\mathcal{F})$. Combining Theorem 1 with Theorem 2, we have the following proposition that establishes the convergence rate of $R_{upper}^{b} (\hat{f}_{n}^{\lambda}) - R_{upper}^{*} - \mathcal{A} (\mathcal{F})$, that is, the excess risk minus approximation error under the true 0-1-based loss.

**Proposition 5** Under Assumption 1–6, we have

$$R_{upper}^{b} (\hat{f}_{n}^{\lambda}) \leq R_{upper}^{*} + \mathcal{A} (\mathcal{F}) + O (\lambda_{n} + n^{-\frac{1}{2}} \lambda_{n}^{-1} + \lambda_{n}^{-\frac{1}{2}} (n^{-a} + n^{-b})) .$$

(13)

Proposition 5 shows that $\hat{f}_{n}^{\lambda}$ is a consistent estimator in the sense that its risk converges to the Bayes risk plus the approximation error of the chosen function class $\mathcal{F}$, as long as $\lambda_{n}$ is selected from $\left[ \frac{1}{n^{\frac{1}{2} + 2a + 2b}} , \frac{1}{\ln(\ln(n))} \right]$. In other words, $\hat{f}_{n}^{\lambda}$ converges in its risk to the best IV-optimal ITR in the function class $\mathcal{F}$ for a wide range of $\lambda_{n}$ choices.

6 | SIMULATION STUDIES

In Section 6.1, we define the benchmark of our experiments. In Section 6.2, we demonstrate that outcome weighted learning (OWL) may have performance in the presence of unmeasured confounding. In Sections 6.3 and 6.4, we investigate the performance of our proposed IV-PILE estimator and contrast it to OWL. Section 6.5 summarizes results from additional simulations, details of which can be found in Supplementary Material F.2 through F.7.

6.1 | Experiment benchmark

There are at least two relevant benchmarks against which we can evaluate the performance of a candidate ITR in the presence of unmeasured confounding. Let $f_{opt}$ be an ITR that only has access to the observed covariates $X$, assigns $A=+1$ when $\text{CATE}(X) = \mathbb{E} [Y(1) - Y(0) | X] \geq 0$, and assigns $A=-1$ otherwise. Let $V(f_{opt})$ denote the value of $f_{opt}$. One may compare a candidate ITR $f$ to the benchmark $f_{opt}$ by calculating:

$$R (f; f_{opt}) : = V(f_{opt}) - V(f) = \mathbb{E} [||\text{CATE}(X)|| \cdot \mathbb{1} \{ \text{sgn} \{ \text{CATE}(X) \} \neq \text{sgn} \{ f(X) \} \}] .$$

Another relevant benchmark ITR is an ‘omniscient’ ITR $f_{omni}$ that has access to both observed covariate $X$ and the unmeasured confounder $U$, assigns $A=+1$ when $\text{CATE}(X, U) = \mathbb{E} [Y(1) - Y(0) | X, U] \geq 0$, and assigns $A=-1$ otherwise. One may compare a candidate ITR $f$ to the benchmark $f_{omni}$ by calculating:

$$R (f; f_{omni}) : = V(f_{omni}) - V(f) = \mathbb{E} [||\text{CATE}(X, U)|| \cdot \mathbb{1} \{ \text{sgn} \{ \text{CATE}(X, U) \} \neq \text{sgn} \{ f(X) \} \}] .$$
The value functions of the two benchmarks \( f_{\text{opt}} \) and \( f_{\text{omni}} \) differ by a constant that depends only on the data-generating process and encodes the loss of information due to not observing the unmeasured confounder, that is,

\[
V(f_{\text{omni}}) = V(f_{\text{opt}}) + C_{\text{DGP}}
\]

for some constant \( C_{\text{DGP}} \geq 0 \). We then immediately have

\[
\mathcal{R}(f, f_{\text{opt}}) + C_{\text{DGP}} = \mathcal{R}(f, f_{\text{omni}}).
\]

We note that neither \( f_{\text{opt}} \) nor \( f_{\text{omni}} \) is identified from observed data under our general partial identification framework, and an IV-optimal ITR neither converges to \( f_{\text{opt}} \) nor \( f_{\text{omni}} \). Moreover, for the purpose of comparing two or more candidate ITRs, it is equivalent to report \( \mathcal{R}(f, f_{\text{opt}}) \) or \( \mathcal{R}(f, f_{\text{omni}}) \) as they differ only by a constant \( C_{\text{DGP}} \), that is, \( \mathcal{R}(f_1, f_{\text{omni}}) \leq \mathcal{R}(f_2, f_{\text{omni}}) \) implies \( \mathcal{R}(f_1, f_{\text{opt}}) \leq \mathcal{R}(f_2, f_{\text{opt}}) \) and vice versa for two candidate ITRs \( f_1 \) and \( f_2 \) and a fixed data-generating process.

If we further assume identification assumptions that enable us to point identify \( \text{CATE}(X) \) as in Cui and Tchetgen Tchetgen (2020) and Qiu et al. (2020), then \( f_{\text{opt}} \) becomes identified from observed data, the estimated IV-optimal ITR would converge to \( f_{\text{opt}} \), and \( \mathcal{R}(f, f_{\text{opt}}) \) may be of greater importance; however, we would like to point out that even in this special case, \( \mathcal{R}(f, f_{\text{omni}}) \) is still of interest as \( C_{\text{DGP}} \) illuminates the gains from collecting more covariates.

In the simulation studies, we will estimate and report \( \mathcal{R}(f, f_{\text{omni}}) \) for each candidate ITR \( f \) and data-generating process. Weighted misclassification error rate refers to \( \mathcal{R}(f, f_{\text{omni}}) \). We also calculate \( C_{\text{DGP}} \) for each data-generating process. Readers can then immediately deduce \( \mathcal{R}(f, f_{\text{opt}}) \) from \( \mathcal{R}(f, f_{\text{omni}}) \) and \( C_{\text{DGP}} \).

### 6.2 Failure of the outcome weighted learning in the presence of unmeasured confounding

We illustrate how ITR-estimation methods could dramatically fail in the presence of unmeasured confounding. We consider the following simple data-generating process of covariates and treatment:

\[
X_1, X_2, U \sim \text{Unif} \{-1, 1\},
\]

\[
P(A = 1|X_1, X_2, U) = \expit(1 + X_1 - X_2 + \lambda U),
\]

where \((X_1, X_2)\) are observed covariates and \(U\) an unmeasured covariate. We consider two outcomes, one continuous \((Y_1)\) and the other binary \((Y_2)\):

\[
Y_1 = 1 + X_1 + X_2 + \xi U + 0.442(1 - X_1 - X_2 + \delta U) \cdot A + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 1),
\]

\[
P(Y_2 = 1|X_1, X_2, U, A) = \expit\{1 - X_1 + X_2 + \xi U + 0.442(1 - X_1 + X_2 + \delta U) \cdot A\}.
\]

Observed data consists only of \((X_1, X_2, A, Y_1)\) (or \((X_1, X_2, A, Y_2)\)) since \(U\) is not observed. Parameters \((\lambda, \xi, \delta)\) control the degree of unmeasured confounding, and \((\lambda, \xi, \delta) = (0, 0, 0)\) corresponds to no unmeasured confounding. We adapted the strategy proposed in Zhao et al. (2012) and Zhang et al. (2012a) to the setting of observational studies by fitting a propensity score model \(\hat{f}(a)\) based on \(X_1\) and \(X_2\) alone and estimating the conditional average treatment effect \(C(X)\) using an IPW estimator based on \(\hat{f}(a)\) in a training data set consisting of \(n_{\text{train}} = 300\) subjects. We then labelled each subject as +1 or −1 based on
\[ \text{sgn}\{ \hat{C}(\mathbf{X}) \} \] and attached to her a weight of magnitude \(| \hat{C}(\mathbf{X}) |\). A support vector machine with Gaussian RBF kernel was then applied to this derived data set. For various \((\lambda, \xi, \delta)\) combinations, we repeated the experiment 500 times and reported the average weighted misclassification error rate evaluated on a testing data set of size 100,000, for both outcomes. For the binary outcome \(Y_2\), the average weighted misclassification error is 0.02 when NUCA holds, suggesting a favourable performance of OWL when NUCA holds. However, the error rate jumps to 0.06 when \((\lambda, \xi, \delta)=(4, 0, 4)\). To get a sense of how poor the performance is, note that a classifier based on random coin flips yields an average error of 0.07. Similar qualitatively results hold for the continuous outcome \(Y_1\), and when we use \(R(f, f_{\text{opt}})\) in place of \(R(f, f_{\text{omni}})\). Supplementary Material F.1 summarizes the results.

### 6.3 | Comparing IV-PILE to OWL: experiment setup

#### 6.3.1 | Data generating process

We considered the following data-generating process with a binary IV \(Z\), a binary treatment \(A\), a binary outcome \(Y\), a 10-dimensional observed covariates \(\mathbf{X}\), and an unmeasured confounder \(U\):

\[
Z \sim \text{Bern}(0.5), \; X_1, \ldots, X_{10} \sim \text{Unif} \, [-1, 1], \; U \sim \text{Unif} \, [-1, 1],
\]

\[
P(A=1|\mathbf{X}, U, Z) = \expit\{8Z + X_1 - 7X_2 + \lambda(1 + X_1)U\},
\]

\[
P(Y=1|A, \mathbf{X}, U) = \expit\{g_1(\mathbf{X}, U) + g_2(\mathbf{X}, U, A)\},
\]

with the following choices of \(g_1(\mathbf{X}, U)\):

**Model (1):** \(g_1(\mathbf{X}, U) = 1 - X_1 + X_2 + \xi U\),

**Model (2):** \(g_1(\mathbf{X}, U) = 1 - X_1^2 + X_2^2 + \xi X_1 X_2 U\),

and \(g_2(\mathbf{X}, U)\):

**Model (1):** \(g_2(\mathbf{X}, U, A) = 0.442(1 - X_1 + X_2 + \delta U)A\),

**Model (2):** \(g_2(\mathbf{X}, U, A) = (X_2 - 0.25X_1^2 - 1 + \delta U)A\).

In the above specifications, \(\lambda\) controls the level of interaction between \(U\) and \(\mathbf{X}\) on \(P(A=1)\), and \(\delta\) controls the level of interaction between \(U\) and \(A\) on the outcome. Assumptions underpinning naive methods (OWL, EARL, etc) hold when \((\lambda, \xi, \delta) = (0, 0, 0)\). Moreover, the data-generating process considered here does not necessarily satisfy the IV identification assumptions in Cui and Tchetgen Tchetgen (2020) and Qiu et al. (2020). We direct interested readers to Supplementary Material F.1 for more details.

#### 6.3.2 | IV identification assumptions and estimators of \(L(\mathbf{x})\) and \(U(\mathbf{x})\)

We considered the IV identification set \(C_{BP}\) discussed in Section 3.1. Note that \(Z \sim \text{Bern}(0.5)\) trivially satisfies \(C_{BP}\). Under \(C_{BP}\), \(L(\mathbf{X})\) and \(U(\mathbf{X})\) are calculated as in Equations (2) and (3), and require estimating conditional probabilities \(P(Y=y, A=a|Z=z, \mathbf{X} = \mathbf{x})\) for \(2 \times 2 \times 2 = 8 \) \((y = \pm 1, \; a = \pm 1, \; z = \pm 1)\) different \((y, a, z)\) combinations. These conditional probabilities do not involve \(U\) and are identified from the observed data. In general, these conditional probabilities may not admit simple and familiar parametric
form, and researchers are advised to use some flexible estimation routines, for example random forest (Breiman, 2001), to fit the data. We fit all conditional probabilities using random forest models with default settings as implemented in the R package randomForest (Liaw & Wiener, 2002). We also considered estimating these conditional probabilities using simple (but misspecified) multinomial logistic regression models. Likewise, when implementing a naive OWL method, we estimated the propensity score model using a random forest and a logistic regression model.

### 6.3.3 Training and testing data set

Our training data set consisted of \( n_{\text{train}} = 300 \) or 500 independent samples of \((Z, X, A, Y)\). Although the true data-generating process involved the unmeasured confounder \( U \), we did not observe or make use of \( U \) throughout the training process. The testing data set consisted of \( n_{\text{test}} = 100,000 \) independently drawn copies of \((X, U)\). Their true conditional average treatment effects were calculated (since we had both \( X \) and \( U \) information). We reported estimated \( \mathcal{R}(f, f_{\text{omni}}) \) of various candidate ITRs on this testing data set. We also calculated and reported \( C_{\text{DGP}} \) for each data-generating process so that \( \mathcal{R}(f, f_{\text{opt}}) = \mathcal{R}(f, f_{\text{omni}}) - C_{\text{DGP}} \) can be immediately inferred.

### 6.4 Numerical results

We report simulation results in this section. We considered three classifiers:

1. IV-PILE-RF: IV-PILE with relevant conditional probabilities in \( L(X) \) and \( U(X) \) estimated via random forests and classification performed with a Gaussian kernel;
2. OWL-RF: OWL with the propensity score estimated via random forests and classification performed with a Gaussian kernel;
3. COIN-FLIP: Classifier based on random coin flips.

Table 2 reports the estimated weighted misclassification error rate \( \hat{\mathcal{R}}(f, f_{\text{omni}}) \) of IV-PILE-RF, OWL-RF and COIN-FLIP for different \((\lambda, \xi, \delta, g_1(X, U))\) and \( n_{\text{train}} \) combinations when \( g_2(X, U, A) \) is taken to be Model (1). Supplementary Material F.1 reports the same numerical results when \( g_2(X, U, A) \) is taken to be Model (2).

Table 2 suggests two consistent trends that align well with our theory and intuition. First, we would expect that neither \( \hat{\mathcal{R}}(\text{IV-PILE-RF}, f_{\text{omni}}) \) nor \( \hat{\mathcal{R}}(\text{IV-PILE-RF}, f_{\text{opt}}) \) would go to 0 when \( n_{\text{train}} \to \infty \). This is verified by noting that increasing \( n_{\text{train}} \) does not drive either error on the testing data set to 0 (See Supplementary Material F.3 for results when \( n_{\text{train}} \) is larger than 500). Moreover, we observe that \( \hat{\mathcal{R}}(\text{OWL-RF}, f_{\text{omni}}) \) and \( \hat{\mathcal{R}}(\text{OWL-RF}, f_{\text{opt}}) \) also remain large as \( n_{\text{train}} \) grows, which reflects that the problem of unmeasured confounding is fundamental, and does not go away as the training sample size grows.

Second, the IV-PILE estimator seems to be robust and outperforms the naive OWL estimator in all simulation settings considered here. However, we would like to point out that this is not suggesting that our approach always outperforms OWL or similar methods. When the assumptions underpinning these methods are not met, it is difficult to predict what would happen to these methods in practice.
We conducted abundant additional simulations and reported results in Supplementary Material F.2 through F.7. In particular, in Supplementary Material F.2, we compared the performance of IV-PILE and OWL when relevant conditional probabilities in $L(X)$ and $U(X)$ were fit via a misspecified multinomial logistic regression model, and with random forest models with node sizes equal to 5 or 10. Qualitative behaviours described in Section 6.4 still held, and the IV-PILE algorithm seemed to be robust against misspecification of models used to fit relevant probabilities in $L(X)$ and $U(X)$. We also observed that using a larger node size in a random forest model seemed to largely improve the performance in some scenarios. In Supplementary Material F.3, we reported simulation results with larger training sample size $n_{\text{train}}$. In Supplementary Material F.4, we repeated a subset of simulation studies with a sample-splitting version of the IV-PILE algorithm, whose theoretical properties were studied in Section 5. We found that the sample-splitting version of the IV-PILE had slightly inferior finite-sample performance compared to the non-splitting version when $n_{\text{train}}$ is small; however, the sample-splitting version still largely outperformed the non-sample-splitting OWL estimator in simulation settings considered here.

**Table 2** Estimated weighted misclassification error rate for different $(\lambda, \delta, \xi)$ and $g_1(X, U)$ combinations. We take $g_2(X, U, A)$ to be Model (1) throughout. Training data sample size $n_{\text{train}} = 300$ or 500. Each number in the cell is averaged over 500 simulations. Standard errors are in parentheses.

| $\xi = 0$, $g_1 = \text{Model (1)}$ | IV-PILE-RF | OWL-RF | COIN-FLIP | $C_{\text{DGP}}$ |
|---|---|---|---|---|
| $(\lambda, \delta) = (0.5, 0.5)$ | 0.005 (0.000) | 0.005 (0.000) | 0.014 (0.004) | 0.015 (0.003) | 0.031 (0.000) | 0.001 |
| $(\lambda, \delta) = (1.0, 1.0)$ | 0.008 (0.000) | 0.008 (0.000) | 0.018 (0.004) | 0.018 (0.003) | 0.033 (0.000) | 0.005 |
| $(\lambda, \delta) = (1.5, 1.5)$ | 0.014 (0.001) | 0.013 (0.000) | 0.022 (0.004) | 0.023 (0.003) | 0.037 (0.000) | 0.010 |
| $(\lambda, \delta) = (2.0, 2.0)$ | 0.020 (0.000) | 0.020 (0.000) | 0.029 (0.003) | 0.029 (0.003) | 0.043 (0.000) | 0.016 |

| $\xi = 1$, $g_1 = \text{Model (1)}$ | IV-PILE-RF | OWL-RF | COIN-FLIP | $C_{\text{DGP}}$ |
|---|---|---|---|---|
| $(\lambda, \delta) = (0.5, 0.5)$ | 0.005 (0.000) | 0.004 (0.000) | 0.013 (0.004) | 0.014 (0.003) | 0.029 (0.000) | 0.001 |
| $(\lambda, \delta) = (1.0, 1.0)$ | 0.007 (0.000) | 0.007 (0.000) | 0.016 (0.003) | 0.016 (0.003) | 0.029 (0.000) | 0.005 |
| $(\lambda, \delta) = (1.5, 1.5)$ | 0.012 (0.000) | 0.012 (0.000) | 0.019 (0.003) | 0.020 (0.002) | 0.031 (0.000) | 0.009 |
| $(\lambda, \delta) = (2.0, 2.0)$ | 0.018 (0.000) | 0.018 (0.000) | 0.025 (0.003) | 0.025 (0.002) | 0.035 (0.000) | 0.016 |

| $\xi = 0$, $g_1 = \text{Model (2)}$ | IV-PILE-RF | OWL-RF | COIN-FLIP | $C_{\text{DGP}}$ |
|---|---|---|---|---|
| $(\lambda, \delta) = (0.5, 0.5)$ | 0.005 (0.000) | 0.005 (0.000) | 0.017 (0.006) | 0.018 (0.005) | 0.040 (0.000) | 0.001 |
| $(\lambda, \delta) = (1.0, 1.0)$ | 0.007 (0.000) | 0.007 (0.000) | 0.020 (0.006) | 0.021 (0.005) | 0.042 (0.000) | 0.004 |
| $(\lambda, \delta) = (1.5, 1.5)$ | 0.012 (0.000) | 0.012 (0.000) | 0.024 (0.006) | 0.025 (0.005) | 0.045 (0.000) | 0.008 |
| $(\lambda, \delta) = (2.0, 2.0)$ | 0.019 (0.000) | 0.019 (0.000) | 0.031 (0.006) | 0.031 (0.005) | 0.050 (0.000) | 0.014 |

| $\xi = 1$, $g_1 = \text{Model (2)}$ | IV-PILE-RF | OWL-RF | COIN-FLIP | $C_{\text{DGP}}$ |
|---|---|---|---|---|
| $(\lambda, \delta) = (0.5, 0.5)$ | 0.004 (0.000) | 0.004 (0.000) | 0.017 (0.004) | 0.017 (0.005) | 0.040 (0.000) | 0.001 |
| $(\lambda, \delta) = (1.0, 1.0)$ | 0.007 (0.000) | 0.007 (0.000) | 0.019 (0.006) | 0.020 (0.005) | 0.042 (0.000) | 0.004 |
| $(\lambda, \delta) = (1.5, 1.5)$ | 0.012 (0.000) | 0.012 (0.000) | 0.024 (0.006) | 0.025 (0.005) | 0.045 (0.000) | 0.009 |
| $(\lambda, \delta) = (2.0, 2.0)$ | 0.019 (0.000) | 0.019 (0.000) | 0.030 (0.005) | 0.031 (0.005) | 0.049 (0.000) | 0.015 |
In Supplementary Material F.5, we varied the association between the IV Z and the treatment A. We found that the performance of IV-PILE became worse when the association between the IV and the treatment became smaller. This observation aligns very well with Proposition 4 and our intuition because a weaker IV corresponds to wider and less informative partial identification intervals. In Supplementary Material F.6, we allowed the IV to violate the exclusion restriction assumption and have a direct effect on the outcome. We found that the IV-PILE estimator seemed to be robust to slight violation of the exclusion restriction assumption. Finally, in Supplementary Material F.7, we considered settings where the outcome was continuous but bounded, and estimated the partial identification interval using the Manski–Pepper bounds. We found that the qualitative conclusions that held for a binary outcome still held for the continuous but bounded outcome.

7 | DIFFERENTIAL IMPACT OF DELIVERY HOSPITAL ON PREEMIES’ OUTCOMES REVISITED

We now revisit the NICU data and apply our developed method to it. We considered data describing all premature babies in the Commonwealth of Pennsylvania from the year 1995 to 2005, with the following observed covariates describing mothers and their preemies: birth weight, gestational age in weeks, age of mother, insurance type of mother (fee for service, HMO, federal/state, other, uninsured), mother’s race (white, African American, Hispanic, other), prenatal care, mother’s education level, mother’s parity, and the following covariates describing the zip code the mother lives in: median income, median home value, percentage of people who rent, percentage below poverty, percentage with high school degree, percentage with college degree. The treatment is 1 if the baby is delivered at a hospital with high-level NICUs and 0 otherwise. The treatment is believed to be still confounded because there are other important covariates not accounted for, for example the severity of mother’s comorbidities. Mothers with severe comorbidities are more likely to be sent to high-level NICUs and their babies are at higher risk.

To resolve this concern, we followed Lorch et al. (2012) and used the differential travel time, defined as mother’s travel time to the nearest high-level NICU minus time to the nearest low-level NICU, as an instrumental variable. We constructed a binary IV Z out of the excess travel time as follows: \( Z = 1 \) if excess travel time in its highest 10 percentile, and \( Z = -1 \) if in its lowest 10 percentile. The outcome of interest is premature infant mortality, including both fetal and non-fetal death. Our goal is to estimate an individualized treatment rule that recommends whether to send mothers to hospitals with high-level NICUs based on her observed covariates. In the case of having a premature baby, mothers should be directed to a hospital with high-level NICUs instead of the closest hospital only when a high-level NICU is significantly better for preemie mortality. As has been discussed in Section 2.3, we let \( \Delta \) control for what margin is significant. In practice, \( \Delta \) should be informed based on some

| Assumption set            | \( \Delta = 0 \) | \( \Delta = 0.02 \) | \( \Delta = 0.05 \) |
|---------------------------|-----------------|-----------------|-----------------|
| \( |D_l| \)  | \( |D_{ul}| \)  | \( |D_l| \)  | \( |D_{ul}| \)  | \( |D_l| \)  | \( |D_{ul}| \)  |
| \( C_{BP}: \) Balke–Pearl bound | 2132            | 23,617           | 4926            | 20,776           | 7463            | 18,239           |
| \( C_{Sid}: \) Siddique bound   | 23,253          | 2449            | 25,549          | 153              | 25,535          | 167              |
practical knowledge of the situation, for example how far the hospital with high-level NICUs is, and how urgent the situation is.

Before proceeding to estimation, we saved 5000 data points as testing data and used the rest 25,702 as training data. We considered two IV assumption sets: \( C_{BP} \) that underpins the Balke–Pearl bound and \( C_{Sid} \) that underpins the Siddique bound (see Section 3). Table 3 summarizes the sizes of the labelled and unlabelled parts of the training data, that is, \(|D_l|\) and \(|D_{ul}|\) under assumption sets \( C_{BP} \) and \( C_{Sid} \) and for assorted \( \Delta \) values. We observed that the additional ‘Correct Non-Compliant Decision’ assumption in \( C_{Sid} \) helped significantly reduce the length of the partial identification intervals for our data, and therefore largely reduced \(|D_{ul}|\). Nevertheless, \(|D_{ul}| > 0\) in all cases for both \( C_{BP} \) and \( C_{Sid} \).

We applied the developed IV-PILE approach to the training data, and used a 5-fold cross validation to select the tuning parameters \( \lambda_u \) (controlling the model complexity) and \( \sigma \) (Gaussian kernel parameter) on a logarithmic grid from \( 10^{-3} \) to \( 10^{3} \). Let \( \hat{f}_{BP} \) denote the estimated ITR with optimal tuning parameters under the assumption set \( C_{BP} \) and \( \hat{f}_{Sid} \) under \( C_{Sid} \). We then applied \( \hat{f}_{BP} \) and \( \hat{f}_{Sid} \) on the 5000 testing data and estimated \( R_{upper} \). When \( \Delta = 0 \), we had \( \hat{R}_{upper}(\hat{f}_{BP}) = 0.149 \) and \( \hat{R}_{upper}(\hat{f}_{Sid}) = 0.020 \). This suggests that \( \hat{f}_{BP} \) would incur a weighted misclassification error at most as large as 0.149 under the four core IV assumptions. If we further assumed the ‘Correct Non-Compliant Decision’ assumption as in \( C_{Sid} \), the estimated ITR \( \hat{f}_{Sid} \) would incur a weighted misclassification error at most as large as 0.020. It is not surprising that the expected worst-case loss is much smaller under \( C_{Sid} \) given that the additional assumption in \( C_{Sid} \) largely reduced the length of \([L(X),U(X)]\). Importantly, although the optimal rule is not identifiable under \( C_{upper} \) or \( C_{Sid} \), we can estimate the worst-case risk associated with the IV-PILE estimator, and this gives practitioners important guidance: the learned treatment rule is potentially useful and beneficial when the identification assumption set is agreed upon by experts, and the worst-case risk is deemed reasonable.

\section{Discussion}

We study in detail the problem of estimating individualized treatment rules with a valid instrumental variable in this article. We have two major contributions. First, we point out the connection and a fundamental distinction between ITR estimation problems with and without an IV: both problems can be viewed as a classification task; however, the partial identification nature of an IV analysis creates a third latent class, those for whom we cannot assert if the treatment is beneficial or harmful, in addition to those who would ‘benefit’ or ‘not benefit’ from the treatment. This perspective provides a unifying framework that facilitates thinking and framing the problem under distinct and problem-specific IV identification assumptions.

Second, we approach this unique classification problem by defining a new notion of ‘IV-optimality’: an IV-optimal rule minimizes the worst-case weighted misclassification error with respect to the putative IV and under the set of IV identification assumptions. IV-optimality is a sensible criterion that is always well-defined, and an IV-optimal rule can be estimated even under minimal IV identification assumptions and mild modelling assumptions. Our proposed IV-PILE estimator estimates such an IV-optimal rule, and may be advantageous compared to naively applying OWL or similar methods to observational data when NUCA fails, or when the putative IV does not allow point identifying the conditional average treatment effect. Although the focus of the article is estimating ITRs using observational data, the method developed here also applies to randomized control trials with individual non-compliance, as is commonly seen in clinical decision support systems.

Works most related to our proposed approach are Kallus and Zhou (2018) and Kallus et al. (2019), both of which consider the problem of improving a baseline policy when a \( \Gamma \)-sensitivity
analysis model is used to control the degree of unmeasured confounding. Both Kallus and Zhou (2018) and Kallus et al. (2019) consider minimizing the maximum risk relative to the baseline policy, and the CATE is also partially identified under the prescribed sensitivity analysis model. There are two main differences between their approach and ours. First, their approach necessarily requires a baseline policy/ITR and their derived policy/ITR is only guaranteed to do no worse than this baseline under their prescribed sensitivity analysis model. On the other hand, our approach does not require a baseline, and our method can be thought of as delivering a reasonably ‘good’ baseline policy/ITR. Second, their ‘improved policy’ always mimics the baseline when the CATE under the sensitivity analysis model covers 0. On the other hand, our IV-PILE estimator has a very different target, that is, ‘IV-optimality’, and would recommend a treatment based on the partial identification region alone. One promising research direction is to study how to improve a baseline policy/ITR using one or several valid instrumental variables, instead of relying on a sensitivity analysis model.

Finally, we outline several broad future directions. First, it is of great interest to develop alternative optimality criteria when an IV only partially identifies the conditional average treatment effect, and compare these alternative criteria with IV-optimality. Second, it is of great importance to restrict the function class under consideration to some parsimonious and scientifically meaningful classes, for example the class of decision trees as considered in Laber and Zhao (2015), and the decision lists as considered in Zhang et al. (2015) and Zhang et al. (2018), and develop more interpretable treatment rules under the IV setting. The ‘IV-optimality’ developed in this article is still a relevant criterion for such an interpretable decision rule. Third, the ‘min-max’ approach as developed in this article can be made less conservative in some settings. One possibility is to consider additional structural assumptions on C(x). For instance, it is conceivable that C(x), subject to L(x) ≤ C(x) ≤ U(x), is smooth in x. Lastly, instead of the single-decision setting considered in this article, it is interesting to consider multi-stage problems, that is, dynamic treatment rules (Moodie et al., 2007; Murphy, 2003; Murphy et al., 2001; Robins, 2004; Zhao et al., 2011), and investigate learning optimal dynamic treatment rules with a potentially time-varying instrumental variable.

ACKNOWLEDGEMENTS
The authors thank Dylan S. Small and reading group participants at the University of Pennsylvania for helpful thoughts and feedback. The authors acknowledge Scott A. Lorch for access to the NICU data.

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

Baiocchi, M., Cheng, J. & Small, D.S. (2014) Instrumental variable methods for causal inference. *Statistics in Medicine*, 33, 2297–2340.

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

Bartlett, P.L., Jordan, M.I. & McAuliffe, J.D. (2006) Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101, 138–156.

Bickel, P.J. (1982) On adaptive estimation. *The Annals of Statistics*, 10, 647–671.

Breiman, L. (2001) Random forests. *Machine Learning*, 45, 5–32.

Cai, T.T. et al. (2012) Minimax and adaptive inference in nonparametric function estimation. *Statistical Science*, 27, 31–50.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. et al. (2018) Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21, C1–C68. Available from: https://doi.org/10.1111/ectj.12097.
Cui, Y. & Tchetgen Tchetgen, E. (2020) A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, 1–34.

Donoho, D.L., Johnstone, I.M. et al. (1998) Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, 26, 879–921.

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

Kallus, N. & Zhou, A. (2018) Confounding-robust policy improvement. In *Advances in neural information processing systems*, pp. 9269–9279.

Kallus, N., Mao, X. & Zhou, A. (2019) Interval estimation of individual-level causal effects under unobserved confounding. In *The 22nd international conference on artificial intelligence and statistics*, pp. 2281–2290.

Kosorok, M.R. (2007) *Introduction to empirical processes and semiparametric inference*. Berlin: Springer Science & Business Media.

Kosorok, M.R. & Laber, E.B. (2019) Precision medicine. *Annual Review of Statistics and its Application*, 6, 263–286.

Kroelinger, C.D., Okoroh, E.M., Goodman, D.A., Lasswell, S.M. & Barfield, W.D. (2018) Comparison of state risk-appropriate neonatal care policies with the 2012 AAP policy statement. *Journal of Perinatology*, 38, 411–420.

Laber, E. & Zhao, Y. (2015) Tree-based methods for individualized treatment regimes. *Biometrika*, 102, 501–514.

Lasswell, S.M., Barfield, W.D., Rochat, R.W. & Blackmon, L. (2010) Perinatal regionalization for very low-birth-weight and very preterm infants: A meta-analysis. *JAMA*, 304, 992–1000.

Liaw, A. & Wiener, M. (2002) Classification and regression by randomforest. *R News*, 2, 18–22. Available from: https://CRAN.R-project.org/doc/Rnews/.

Manski, C.F. (2003) *Partial identification of probability distributions*. Berlin: Springer Science & Business Media.

Manski, C.F. & Pepper, J.V. (2000) Monotone instrumental variables: with an application to the returns to schooling. *Econometrica*, 68, 997–1010.

Moodie, E.E., Richardson, T.S. & Stephens, D.A. (2007) Demystifying optimal dynamic treatment regimes. *Biometrics*, 63, 447–455.

Mourtada, J., Gaïffas, S. & Scornet, E. (2018) Minimax optimal rates for mondrian trees and forests. *arXiv preprint arXiv:1803.05784*.

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

Murphy, S.A., van der Laan, M.J., Robins, J.M. & Group, C.P.P.R. (2001) Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96, 1410–1423.

Qian, M. & Murphy, S.A. (2011) Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39, 1180.

Qiu, H., Carone, M., Sadikova, E., Petukhova, M., Kessler, R.C. & Luedtke, A. (2020) Optimal individualized decision rules using instrumental variable methods. *Journal of the American Statistical Association*, 1–46.

Richardson, T.S. & Robins, J.M. (2010) Analysis of the binary instrumental variable model. R. Dechter, H. Geffner, & J.Y. Halpern (Eds.), *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, London: College Publications, 415–444.

Robins, J.M. (1992) Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, 79, 321–334.

Robins, J.M. (2004) Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*. Springer, pp. 189–326.

Robins, J.M. & Greenland, S. (1996) Identification of causal effects using instrumental variables: Comment. *Journal of the American Statistical Association*, 91, 456–458.

Robins, J.M., Li, L., Mukherjee, R., Tchetgen, E.T., van der Vaart, A. et al. (2017) Minimax estimation of a functional on a structured high-dimensional model. *The Annals of Statistics*, 45, 1951–1987.

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

Rouse, C.E. (1995) Democratization or diversion? The effect of community colleges on educational attainment. *Journal of Business & Economic Statistics*, 13, 217–224.
Siddique, Z. (2013) Partially identified treatment effects under imperfect compliance: The case of domestic violence. *Journal of the American Statistical Association*, 108, 504–513.

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

Tsatis, A.A. (2019) *Dynamic treatment regimes: Statistical methods for precision medicine*. Boca Raton, FL: CRC Press.

Tukey, J.W. (1986) Sunset salvo. *The American Statistician*, 40, 72–76.

Vapnik, V. (1992) Principles of risk minimization for learning theory. In D.S. Lippman, J.E. Moody, & D.S. Touretzky (Eds.), *Advances in neural information processing systems*. Vol. 4. Morgan Kaufmann, pp. 831–838.

Vapnik, V. (2013) *The nature of statistical learning theory*. Berlin: Springer Science & Business Media.

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

Wang, Y., Wu, P., Liu, Y., Weng, C. & Zeng, D. (2016) Learning optimal individualized treatment rules from electronic health record data. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, pp. 65–71.

Zhang, B., Tsatis, A.A., Davidian, M., Zhang, M. & Laber, E. (2012a) Estimating optimal treatment regimes from a classification perspective. *Stat*, 1, 103–114.

Zhang, B., Tsatis, A.A., Laber, E.B. & Davidian, M. (2012b) A robust method for estimating optimal treatment regimes. *Biometrics*, 68, 1010–1018.

Zhang, Y., Laber, E.B., Tsatis, A. & Davidian, M. (2015) Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, 71, 895–904.

Zhang, Y., Laber, E.B., Davidian, M. & Tsatis, A.A. (2018) Interpretable dynamic treatment regimes. *Journal of the American Statistical Association*, 113, 1541–1549.

Zhang, B., Weiss, J., Small, D.S. & Zhao, Q. (2020) Selecting and ranking individualized treatment rules with unmeasured confounding. *Journal of the American Statistical Association*, 1–14.

Zhao, Y., Zeng, D., Socinski, M.A. & Kosorok, M.R. (2011) Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics*, 67, 1422–1433.

Zhao, Y., Zeng, D., Rush, A.J. & Kosorok, M.R. (2012) Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107, 1106–1118.

Zhao, Y.-Q., Laber, E.B., Ning, Y., Saha, S. & Sands, B.E. (2019) Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *Journal of Machine Learning Research*, 20, 1–23.

Zheng, W. & van der Laan, M.J. (2011) Cross-validated targeted minimum-loss-based estimation. In M.J. van der Laan & S. Rose (Eds.), *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York: Springer, pp. 459–474.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

---

**How to cite this article:** Pu H, Zhang B. Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach. *J R Stat Soc Series B*. 2021;83:318–345. [https://doi.org/10.1111/rssb.12413](https://doi.org/10.1111/rssb.12413)