

# Robust Identification in Randomized Experiments with Noncompliance

Désiré Kédagni<sup>\*</sup>, Huan Wu<sup>†</sup>, and Yi Cui<sup>‡</sup>,

*University of North Carolina at Chapel Hill*

This version:<sup>§</sup> March 21, 2025.

## Abstract

Instrument variable (IV) methods are widely used in empirical research to identify causal effects of a policy. In the local average treatment effect (LATE) framework, the IV estimand identifies the LATE under three main assumptions: random assignment, exclusion restriction, and monotonicity. However, these assumptions are often questionable in many applications, leading some researchers to doubt the causal interpretation of the IV estimand. This paper considers a robust identification of causal parameters in a randomized experiment setting with noncompliance where the standard LATE assumptions could be violated. We discuss identification under two sets of weaker assumptions: random assignment and exclusion restriction (without monotonicity), and random assignment and monotonicity (without exclusion restriction). We derive sharp bounds on some causal parameters under these two sets of relaxed LATE assumptions. Finally, we apply our method to revisit the random information experiment conducted in [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#) and find that the standard LATE assumptions are jointly incompatible in this application. We then estimate the robust identified sets under the two sets of relaxed assumptions.

**Keywords:** Instrumental variable, monotonicity, exclusion restriction, local average controlled direct effect.

**JEL subject classification:** C14, C31, C35, C36.

---

<sup>\*</sup>Corresponding author, email address: [dkedagni@unc.edu](mailto:dkedagni@unc.edu)

<sup>†</sup>Email address: [huan.wu@unc.edu](mailto:huan.wu@unc.edu)

<sup>‡</sup>Email address: [yicui@unc.edu](mailto:yicui@unc.edu)

<sup>§</sup>We are grateful to Felipe Goncalves, Ismael Mourifié, Claudia Noack, Vira Semenova, and Kaspar Wuthrich for their helpful comments. We are also grateful to participants at the UNC-Chapel Hill Econometrics Workshop, and the 2024 Midwest Econometrics Group conference. First version: August 6, 2024.

# 1 Introduction

Researchers are often concerned about the credibility of their assumptions and the conclusions they reach (see *The Law of Decreasing Credibility* in [Manski, 2003, 2011](#)). When the stated assumptions are refuted by the data, analysts usually resort to weaker versions of the assumptions. There may exist multiple ways of relaxing a set of assumptions when they are rejected by the data. Some researchers may choose to relax the assumptions in a continuous way (as in [Masten and Poirier, 2021](#)), while others may consider discrete relaxations by dropping some of the assumptions (as in [Li, Kédagni, and Mourifié, 2024](#)). In this paper, we consider a robust identification of causal parameters in a randomized experiment setting with noncompliance where the standard local average treatment effect (LATE) assumptions could be violated. We follow [Li, Kédagni, and Mourifié’s \(2024\)](#) approach to propose a misspecification robust bound for a vector-valued parameter of various causal parameters.

Three main assumptions (random assignment, exclusion restriction, and monotonicity) are commonly used in the LATE framework. When these assumptions are jointly rejected, researchers could choose to relax some of the assumptions if they have some additional information on which assumptions may be causing the rejection. For example, [Kédagni \(2023\)](#) proposes an identification strategy that allows for violations of the random assignment assumption, [de Chaisemartin \(2017\)](#), [Huber, Laffers, and Mellace \(2017\)](#), [Small et al. \(2017\)](#), [Noack \(2021\)](#), [Dahl, Huber, and Mellace \(2023\)](#), among others have proposed different ways of relaxing the monotonicity assumption, while [Cai et al. \(2008\)](#) and [Flores and Flores-Lagunes \(2013\)](#) allow for violations of the exclusion restriction assumption. But these relaxations may appear arbitrary in some circumstances unless the analyst has an unambiguous reason to convince the scientific community that she knows exactly which assumptions are the source of the rejection in the data.

The current paper aims at developing a unifying approach that would provide the maximal combination of the three assumptions that are compatible with the data. We consider a randomized experiment setting where the random assignment assumption is maintained. We discuss identification under two sets of weaker assumptions: random assignment and exclusion restriction (without monotonicity), and random assignment and monotonicity (without exclusion restriction).

We first introduce two causal parameters: the local average treatment-controlled direct effect (LATCDE), and the local average instrument-controlled direct effect (LAICDE). We define a twenty-dimensional real-valued vector parameter of interest that contains the eight LATCDEs, the eight LAICDEs, and the four types’ probabilities. Second, we derive

sharp bounds for the twenty parameters under respectively (i) random assignment, exclusion restriction, and monotonicity; (ii) random assignment, and exclusion restriction; and (iii) random assignment, and monotonicity. Specifically, under the random assignment and monotonicity assumptions, we derive sharp bounds on the local average treatment-controlled direct effects for the always-takers and never-takers, respectively, and the total average controlled direct effect for the compliers. Additionally, we show that the intent-to-treat effect can be expressed as a convex weighted average of these three effects. Third, we propose a misspecification robust bound for the vector parameter of interest under the three possible sets of assumptions. Finally, we apply our method to analyze the randomized information experiment conducted by [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#). We find that the LATE assumptions are jointly rejected in this setting. However, when either the monotonicity or exclusion restriction is dropped, the remaining two assumptions cannot be rejected. We provide estimates of identified sets for our parameters of interest, which are robust to violations of monotonicity or exclusion assumptions.

The remainder of the paper is organized as follows. Section 2 presents the model, the assumptions, and the parameters of interest. In Sections 3, 4, and 5, we derive bounds on the parameters under respectively (i) random assignment, exclusion restriction, and monotonicity; (ii) random assignment, and exclusion restriction; and (iii) random assignment, and monotonicity. Section 6 presents the misspecification robust bound, Section 7 shows the empirical illustration, and finally Section 8 concludes. The proofs of the main results are relegated to the appendix from the supplementary material.

## 2 Analytical framework

We consider a potential outcome model with a binary treatment and a binary instrumental variable. Let  $D \in \{0, 1\}$  be the observed treatment, where  $D = 1$  indicates receiving the treatment and  $D = 0$  not receiving it. Let  $Y$  denote the observed outcome, and  $Z$  denote a binary instrumental variable, which takes values in  $\mathcal{Z} = \{0, 1\}$ . Following the convention, we use  $D_0$  and  $D_1$  to represent the potential treatment status, where  $D_z$  denotes the potential treatment if the instrument  $Z$  is externally set to  $z$ . Also,  $Y_{dz}$  represents the potential outcome if the treatment and instrument are externally set to  $d$  and  $z$ , respectively. Then, we have the following potential outcome model,

$$\begin{cases} Y = (Y_{11}Z + Y_{10}(1 - Z))D + (Y_{01}Z + Y_{00}(1 - Z))(1 - D), \\ D = D_1Z + D_0(1 - Z), \end{cases} \quad (1)$$

where  $(Y, D, Z)$  is the observed data,  $(Y_{00}, Y_{01}, Y_{10}, Y_{11}, D_0, D_1)$  is a vector of latent variables.

The vector  $(D_0, D_1)$  represents the four unobserved groups commonly known in the treatment effect literature as the *types* in the language of Angrist, Imbens, and Rubin (1996).

We have

$$\begin{aligned} (D_0 = 1, D_1 = 1) &: \text{always-takers } (a), & (D_0 = 0, D_1 = 0) &: \text{never-takers } (n), \\ (D_0 = 0, D_1 = 1) &: \text{compliers } (c), & (D_0 = 1, D_1 = 0) &: \text{defiers } (df). \end{aligned}$$

Let  $T$  denote the random type of an individual with support  $\{a, c, n, df\}$ . Let  $p_t$  denote the proportion of type  $t$  in the population. Thus, we have the corresponding probabilities for these four types (strata/groups) of the population:  $p_n, p_{df}, p_c$ , and  $p_a$ . We define our parameters of interest:

$$\begin{aligned} \theta_{zt} &\equiv \mathbb{E}[Y_{1z} - Y_{0z} \mid T = t] \text{ (local average instrument-controlled direct effect, LAICDE),} \\ \delta_{dt} &\equiv \mathbb{E}[Y_{d1} - Y_{d0} \mid T = t] \text{ (local average treatment-controlled direct effect, LATCDE),} \end{aligned}$$

for  $z \in \{0, 1\}$ ,  $d \in \{0, 1\}$ , and  $t \in \{a, c, df, n\}$ . We have eight  $\theta$ -parameters, and eight  $\delta$ -parameters. Combined with the four types' probabilities  $p_t$ , we define a  $20 \times 1$  column vector  $\Gamma$  of our parameters of interest,

$$\Gamma \equiv [\theta_{0a}, \theta_{1a}, \theta_{0c}, \theta_{1c}, \theta_{0df}, \theta_{1df}, \theta_{0n}, \theta_{1n}, \delta_{0a}, \delta_{1a}, \delta_{0c}, \delta_{1c}, \delta_{0df}, \delta_{1df}, \delta_{0n}, \delta_{1n}, p_a, p_c, p_{df}, p_n]'$$

Elements of the vector  $\Gamma$  can be viewed as building block parameters for many commonly used causal parameters. For instance, we can define some commonly used parameters as functionals of  $\Gamma$ :

$$\begin{aligned} \theta_t &\equiv \sum_{z \in \{0, 1\}} \mathbb{P}(Z = z) \theta_{zt} \text{ (local average treatment effect for type } t, LATE_t), \\ \theta &\equiv \sum_{t \in \{a, c, df, n\}} p_t \theta_t \text{ (average treatment effect, ATE),} \\ \delta_d &\equiv \sum_{t \in \{a, c, df, n\}} p_t \delta_{dt} \text{ (average controlled direct effect, ACDE).} \end{aligned}$$

In this paper, our goal is to provide a robust identified set for these parameters under the following set of assumptions considered in the original work by Imbens and Angrist (1994).

**Assumption 1** (Random assignment, RA). The instrument variable  $Z$  is independent of

potential outcomes and potential treatments:

$$Z \perp\!\!\!\perp (Y_{11}, Y_{10}, Y_{01}, Y_{00}, D_1, D_0).$$

Assumption 1 requires that the instrument  $Z$  be independent of all potential outcomes and potential treatments. This assumption will likely hold in randomized experiments. We may assume that it holds conditional on covariates in stratified randomized experiments. This assumption is maintained throughout the paper.

**Assumption 2** (Exclusion restriction, ER).  $Y_{dz} = Y_{dz'} = Y_d$  for all  $d, z$ , and  $z'$ .

Assumption 2 imposes that the instrument  $Z$  does not have a direct effect on the potential outcome. It is allowed to affect the outcome  $Y$  only through the treatment  $D$ . We allow for potential violations of this assumption in this paper.

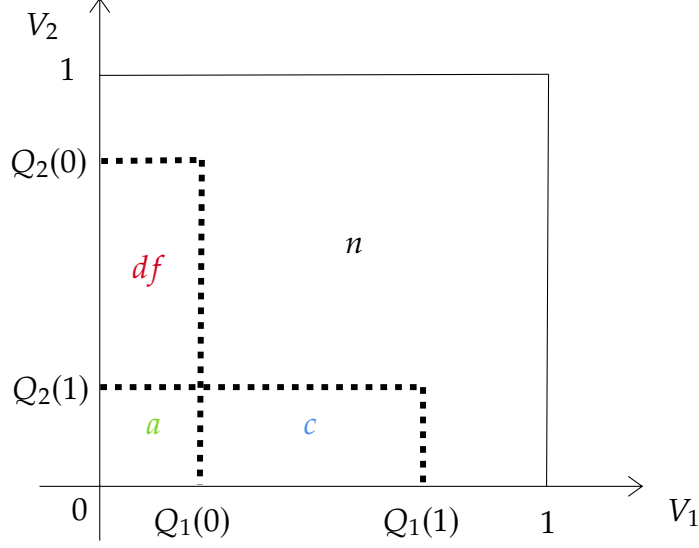
**Assumption 3** (Monotonicity, MON). Either  $D_1 \geq D_0$  or  $D_0 \geq D_1$ .

Assumption 3, also often referred to as *uniformity* (see Heckman and Pinto, 2018), rules out the existence of both compliers and defiers in the population. This assumption is questionable in many empirical settings, especially the judge IV design literature. To illustrate, suppose there are two judges who are randomly assigned to cases/defendants. A jury will make a decision on whether or not to convict a defendant. A researcher is interested in studying the effect of the jury's decision  $D$  on an outcome  $Y$  (recidivism/employment). The judges assignment variable  $Z$  can be considered as an IV. There is no *a priori* reason to assume monotonicity in the jury's decision with respect to the judges. Hence, all four types are likely present in this setting. Therefore, we allow for complete violations of this monotonicity assumption in this paper.

Fiorini and Stevens (2021) extensively discussed situations in which the monotonicity assumption might be questionable. For illustration purposes, we consider the measurement of the returns to schooling as our motivating example.

*Example 1* (Returns to schooling). Suppose that the college decision is based on whether the opportunity cost  $V_1$  is less than a threshold  $Q_1(Z)$  and the cost of attending college  $V_2$  (e.g., psychological cost of effort, tuition) is less than a different threshold  $Q_2(Z)$ , where  $Z$  denotes an IV, say college proximity. This two-dimensional decision model is called the double hurdle model in Lee and Salanié (2018). Let  $Y$  be the outcome of interest (wage),  $D$  be the treatment variable (college attendance), and  $Z$  be the instrument (college proximity) so that the college choice model is given as below:

$$D = \mathbb{1} \{V_1 < Q_1(Z), V_2 < Q_2(Z)\}, \quad (2)$$



**Figure 1:** Visualization of types (double hurdle model)

In this framework, the monotonicity assumption will likely fail if for example the instrument lowers  $Q_1$  while at the same time it increases  $Q_2$ , that is,  $Q_1(0) < Q_1(1)$ , and  $Q_2(1) < Q_2(0)$ . As a consequence, we will have all four types of individuals in the population: compliers ( $c$ ), defiers ( $df$ ), never-takers ( $n$ ), and always-takers ( $a$ ) as illustrated in Figure 1.<sup>1</sup>

In addition, growing up near a college could have a direct effect on someone's ability or skills necessary to complete some tasks that they would not have done otherwise. These skills would positively affect people's wages, which would make the college proximity instrument violate the exclusion restriction assumption.

We consider three different combinations (or menus) of assumptions.

$$A_1 \equiv \{RA, ER, MON\},$$

$$A_2 \equiv \{RA, ER\},$$

$$A_3 \equiv \{RA, MON\}.$$

We denote  $A$  the set of all three combinations of assumptions:  $A \equiv \{A_1, A_2, A_3\}$ . We assume that we are in a randomized experiment setting where Assumption 1 holds. For this reason, we do not consider the menu  $\{ER, MON\}$  as an option. In general, this

<sup>1</sup>Definition of the four types in the population:

$c = \{V_1 < Q_1(1), V_2 < Q_2(1)\} \cap \{V_1 \geq Q_1(0) \cup V_2 \geq Q_2(0)\},$   
 $df = \{V_1 < Q_1(0), V_2 < Q_2(0)\} \cap \{V_1 \geq Q_1(1) \cup V_2 \geq Q_2(1)\},$   
 $a = \{V_1 < Q_1(0), V_2 < Q_2(0)\} \cap \{V_1 < Q_1(1), V_2 < Q_2(1)\},$   
 $n = \{V_1 \geq Q_1(0) \cup V_2 \geq Q_2(0)\} \cap \{V_1 \geq Q_1(1) \cup V_2 \geq Q_2(1)\}.$

combination alone would not yield informative bounds without additional assumptions (see [Kédagni, 2023](#), for further details).

### 3 Identification under random assignment, monotonicity, and exclusion restriction

Let  $\mathcal{Y}$  denote the support of the outcome  $Y$ . We first note that under Assumption 2,  $Y_{dz} = Y_d$  for all  $d$  and  $z$ . Under Assumptions 1 and 2, we have for any Borel set  $A \subseteq \mathcal{Y}$ ,

$$\mathbb{P}(Y \in A, D = 1|Z = 1) = p_c \mathbb{P}(Y_1 \in A|T = c) + p_a \mathbb{P}(Y_1 \in A|T = a), \quad (3)$$

$$\mathbb{P}(Y \in A, D = 1|Z = 0) = p_{df} \mathbb{P}(Y_1 \in A|T = df) + p_a \mathbb{P}(Y_1 \in A|T = a), \quad (4)$$

$$\mathbb{P}(Y \in A, D = 0|Z = 1) = p_{df} \mathbb{P}(Y_0 \in A|T = df) + p_n \mathbb{P}(Y_0 \in A|T = n), \quad (5)$$

$$\mathbb{P}(Y \in A, D = 0|Z = 0) = p_c \mathbb{P}(Y_0 \in A|T = c) + p_n \mathbb{P}(Y_0 \in A|T = n). \quad (6)$$

Equation (3) shows that the distribution of  $Y$  in the treatment group for individuals assigned to this group is a mixture of the distribution of  $Y_1$  for the compliers and the always-takers. And Equation (4) shows that the distribution of  $Y$  in the treatment group for individuals assigned to the control group is a mixture of the distribution of  $Y_1$  for the defiers and the always-takers. A similar decomposition holds for the control group, as shown in Equations (5)-(6).

Let us take the difference between Equations (4) and (3). We have

$$\begin{aligned} \mathbb{P}(Y \in A, D = 1|Z = 0) - \mathbb{P}(Y \in A, D = 1|Z = 1) \\ = p_{df} \mathbb{P}(Y_1 \in A|T = df) - p_c \mathbb{P}(Y_1 \in A|T = c). \end{aligned} \quad (7)$$

Similarly, the following holds from Equations (5) and (6).

$$\begin{aligned} \mathbb{P}(Y \in A, D = 0|Z = 1) - \mathbb{P}(Y \in A, D = 0|Z = 0) \\ = p_{df} \mathbb{P}(Y_0 \in A|T = df) - p_c \mathbb{P}(Y_0 \in A|T = c). \end{aligned} \quad (8)$$

For  $A = \mathcal{Y}$ , Equation (7) implies

$$\mathbb{P}(D = 1|Z = 0) - \mathbb{P}(D = 1|Z = 1) = p_{df} - p_c. \quad (9)$$

Under Assumption 3, either  $p_{df} = 0$  or  $p_c = 0$ . Without loss of generality, assume that  $p_{df} = 0$ . Then, Equation (9) implies that  $p_c = \mathbb{P}(D = 1|Z = 1) - \mathbb{P}(D = 1|Z = 0)$ . As a result,

$p_a$  and  $p_n$  are also identified:  $p_a = \mathbb{P}(D = 1|Z = 0)$ , and  $p_n = \mathbb{P}(D = 0|Z = 1)$ . Under Assumption 2,  $\delta_{dt} = 0$  for all  $d$  and  $t$ , and  $\theta_{0t} = \theta_{1t}$  for all  $t$ . Under Assumptions 1-3,  $\theta_{zc}$  is identified if  $\mathbb{P}(D = 1|Z = 1) - \mathbb{P}(D = 1|Z = 0) > 0$  (Imbens and Angrist, 1994):

$$\theta_{1c} = \theta_{0c} = \frac{\mathbb{E}[Y|Z = 1] - \mathbb{E}[Y|Z = 0]}{\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]},$$

and from Equations (7)-(8), the following testable implications must hold:

$$\mathbb{P}(Y \in A, D = 1|Z = 0) - \mathbb{P}(Y \in A, D = 1|Z = 1) \leq 0, \quad (10)$$

$$\mathbb{P}(Y \in A, D = 0|Z = 1) - \mathbb{P}(Y \in A, D = 0|Z = 0) \leq 0, \quad (11)$$

for all Borel set  $A$  (Balke and Pearl, 1997; Heckman and Vytlacil, 2005; Kitagawa, 2015; Mourifié and Wan, 2017). The remaining  $\theta$ -parameters are not identified and lie in the real line  $\mathbb{R}$ . The identified set for  $\Gamma$  is

$$\Theta_I(A_1) = \begin{cases} \Theta_I^1(A_1), & \text{if } \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] > 0, \\ \Theta_I^2(A_1), & \text{if } \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] < 0, \\ \Theta_I^3(A_1), & \text{if } \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] = 0, \end{cases}$$

where

$$\begin{aligned} \Theta_I^1(A_1) = & \left\{ \Gamma \in \mathbb{R}^{20} : \theta_{0a} = \theta_{1a} \in [\mathbb{E}[Y | D = 1, Z = 0] - \sup \mathcal{Y}, \mathbb{E}[Y | D = 1, Z = 0] - \inf \mathcal{Y}] , \right. \\ & \theta_{0c} = \theta_{1c} = \frac{\mathbb{E}[Y|Z = 1] - \mathbb{E}[Y|Z = 0]}{\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]}, \\ & \theta_{0df} = \theta_{1df} \in [\inf \mathcal{Y} - \sup \mathcal{Y}, \sup \mathcal{Y} - \inf \mathcal{Y}] , \\ & \theta_{0n} = \theta_{1n} \in [\inf \mathcal{Y} - \mathbb{E}[Y | D = 0, Z = 1], \sup \mathcal{Y} - \mathbb{E}[Y | D = 0, Z = 1]] , \\ & \delta_{0a} = \delta_{1a} = \delta_{0c} = \delta_{1c} = \delta_{0df} = \delta_{1df} = \delta_{0n} = \delta_{1n} = 0, \\ & p_a = \mathbb{E}[D|Z = 0], p_c = \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0], p_{df} = 0, p_n = \mathbb{E}[1 - D|Z = 1], \\ & \left. \text{and inequalities (10)-(11) hold.} \right\}. \end{aligned}$$

Let  $\tilde{Z} \equiv 1 - Z$ , and let  $\tilde{T}$  and  $\tilde{\Gamma}$  be respectively the corresponding  $T$  and  $\Gamma$  defined based on the instrument  $\tilde{Z}$ . We have  $\tilde{a} = n$ ,  $\tilde{c} = df$ ,  $\tilde{df} = c$ ,  $\tilde{n} = a$ ,  $\tilde{\theta}_{\tilde{z}\tilde{a}} = \theta_{(1-z)n}$ ,  $\tilde{\theta}_{\tilde{z}\tilde{n}} = \theta_{(1-z)a}$ ,  $\tilde{\theta}_{\tilde{z}\tilde{c}} = \theta_{(1-z)df}$ ,  $\tilde{\theta}_{\tilde{z}\tilde{df}} = \theta_{(1-z)c}$ ,  $\tilde{\delta}_{d\tilde{a}} = \delta_{dn}$ ,  $\tilde{\delta}_{d\tilde{n}} = \delta_{da}$ ,  $\tilde{\delta}_{d\tilde{c}} = \delta_{ddf}$ , and  $\tilde{\delta}_{d\tilde{df}} = \delta_{dc}$ . Then  $\Theta_I^2(A_1) = \tilde{\Theta}_I^1(A_1)$ , where  $\tilde{\Theta}_I^1(A_1)$  is the corresponding  $\Theta_I^1(A_1)$  defined for  $\tilde{\Gamma}$  based on  $(Y, D, \tilde{Z})$  and  $\tilde{T}$ .



If  $\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0] = 0$ , then  $p_c - p_{df} = 0$ , and under Assumption 3,  $p_c = p_{df} = 0$ . This implies  $p_a = \mathbb{P}(D = 1 \mid Z = 1) = \mathbb{P}(D = 1 \mid Z = 0) = \mathbb{P}(D = 1)$ , and  $p_n = \mathbb{P}(D = 0)$ . Furthermore, inequalities (10)-(11) must hold with equality, i.e.,  $Z \perp\!\!\!\perp (Y, D)$ . Hence,

$$\Theta_I^3(A_1) = \left\{ \Gamma \in \mathbb{R}^{20} : \begin{aligned} &\theta_{0a} = \theta_{1a} \in [\mathbb{E}[Y \mid D = 1, Z = 0] - \sup \mathcal{Y}, \mathbb{E}[Y \mid D = 1, Z = 0] - \inf \mathcal{Y}], \\ &\theta_{0n} = \theta_{1n} \in \left[ \inf \mathcal{Y} - \mathbb{E}[Y \mid D = 0, Z = 0], \sup \mathcal{Y} - \mathbb{E}[Y \mid D = 0, Z = 0] \right], \\ &\theta_{0c} = \theta_{1c}, \theta_{0df} = \theta_{1df} \in [\inf \mathcal{Y} - \sup \mathcal{Y}, \sup \mathcal{Y} - \inf \mathcal{Y}], \\ &\delta_{0a} = \delta_{1a} = \delta_{0c} = \delta_{1c} = \delta_{0df} = \delta_{1df} = \delta_{0n} = \delta_{1n} = 0, \\ &p_a = \mathbb{E}[D], p_c = p_{df} = 0, p_n = \mathbb{E}[1 - D], \\ &\text{and inequalities (10)-(11) hold with equality.} \end{aligned} \right\}.$$

## 4 Identification under random assignment and exclusion restriction

Building on the above equations (7)-(9), we can derive bounds on the probability of being a defier as follows:

$$\begin{aligned} &\max \left\{ \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s \mid Z = 1 - s) - \mathbb{P}(Y \in A, D = s \mid Z = s) \} \right\}, 0 \right\} \\ &\leq p_{df} \leq \min \{ \mathbb{E}[D \mid Z = 0], \mathbb{E}[1 - D \mid Z = 1] \}. \end{aligned}$$

Kédagni and Mourifié (2020) and Kitagawa (2021) show that the following restriction is a sharp testable implication of Assumptions 1 and 2:

$$\max_{d \in \{0,1\}} \int_{\mathcal{Y}} \sup_z f_{Y,D|Z}(y, d|z) d\mu(y) \leq 1. \quad (12)$$

In the case when the instrument  $Z$  is binary, we can prove that the violation of testable implication in Equation (12) is equivalent to the identified bounds for the proportion of defiers being empty. Therefore, the following proposition holds.

**Proposition 1** (Sharp bounds for  $p_{df}$ ). *Under Assumptions 1 and 2, the identified set  $\Theta_I(p_{df})$*

for the probability of defiers is given by

$$\Theta_I(p_{df}) = \left[ \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s | Z = 1 - s) - \mathbb{P}(Y \in A, D = s | Z = s) \} \right\}, \right. \\ \left. \min\{\mathbb{E}[D | Z = 0], \mathbb{E}[1 - D | Z = 1]\} \right].$$

Moreover, the identified set  $\Theta_I(p_{df})$  is empty if and only if inequality (12) is violated.

*Proof.* The detailed proof is provided in Appendix A.1. □

The bounds in  $\Theta_I(p_{df})$  were first derived by [Richardson and Robins \(2010\)](#) for binary outcomes and recently generalized to any outcomes by [Noack \(2021\)](#). This is an intermediate result for our main results in this paper. If the lower bound of  $\Theta_I(p_{df})$  is greater than zero, the standard LATE assumptions are jointly rejected (or incompatible), aligning with [Balke and Pearl \(1997\)](#), [Imbens and Rubin \(1997\)](#), [Heckman and Vytlacil \(2005\)](#), [Kitagawa \(2015\)](#), and [Mourifié and Wan \(2017\)](#). If instead inequalities (10)-(11) hold, then the testable implications developed for the LATE assumptions hold, and inequality (12) becomes redundant, since the LATE testable implications are sharp ([Kitagawa, 2015; Mourifié and Wan, 2017](#)). In such a context, the lower bound of  $\Theta_I(p_{df})$  is 0, as the supremum in the lower bound is achieved when  $A = \emptyset$ . [Huber, Laffers, and Mellace \(2017\)](#) derive bounds on  $p_{df}$  under a weaker version of Assumption 1,  $\mathbb{E}[Y_{dz} | T, Z] = \mathbb{E}[Y_{dz} | T]$  and  $Z \perp\!\!\!\perp T$ , which implies the [Manski \(1990\)](#) mean independence assumption between the potential outcomes and the instrument,  $\mathbb{E}[Y_{dz} | Z] = \mathbb{E}[Y_{dz}]$ . This latter mean independence assumption together with the exclusion restriction assumption 2 imply similar conditions to (12):  $\sup_z \mathbb{E}[YD + (1 - D) \inf \mathcal{Y} | Z = z] \leq \inf_z \mathbb{E}[YD + (1 - D) \sup \mathcal{Y} | Z = z]$  and  $\sup_z \mathbb{E}[Y(1 - D) + D \inf \mathcal{Y} | Z = z] \leq \inf_z \mathbb{E}[Y(1 - D) + D \sup \mathcal{Y} | Z = z]$ . These conditions could be rejected if the support  $\mathcal{Y}$  is bounded. For example, when the outcome is binary (i.e.,  $\mathcal{Y} = \{0, 1\}$ ), the testable implication for mean independence and Assumption 2 is identical to inequality (12).

[Kitagawa \(2021\)](#) focuses on the identified region of marginal distributions of potential outcomes,  $Y_1$  and  $Y_0$ , under Assumptions 1 and 2, in the same context of a binary treatment and a binary instrument. This paper has a result that the identified region of marginal distributions of potential outcomes is non-empty if and only if inequality (12) holds. We focus on the identified set of type proportions and local average treatment effects here and also prove in Proposition 1 that the identified set of the defier proportion is non-empty if and only if inequality (12) is satisfied. Both of these results indicate that the non-emptiness of two different parameters (marginal potential outcome distributions and

defiers probability) under Assumptions 1 and 2 leads to the same testable inequality (12).

Kwon and Roth (2024) consider a more general framework with a binary instrument and a multivalued discrete treatment. They derive bounds on the proportion of always-takers. Their bounds on the fraction of always-takers can be used to derive bounds on the defiers probability in our framework, since we can express the defiers probability as a function of the always-takers probability.

While our ultimate goal is to derive the identified set for  $\Gamma$  under the bundle of assumptions  $A_2$ , we are first to going derive bounds on the LATE for compliers and defiers, respectively, as these are causal parameters that receive a particular attention in the causal inference literature. Before we move on, we introduce some additional notation. Denote  $\mu_{dt} \equiv \mathbb{E}[Y_d|T = t]$ ,  $d \in \{0, 1\}$  and  $t \in \{a, c, df, n\}$ . Under Assumptions 1 and 2, We have  $p_a = \mathbb{E}[D|Z = 0] - p_{df}$ , and  $p_n = \mathbb{E}[1 - D|Z = 1] - p_{df}$ . Lastly, we have  $p_c = 1 - p_{df} - p_n - p_a = \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] + p_{df}$ , where  $p_{df} \in \Theta_I(p_{df})$ .

To proceed, we are going to derive sharp bounds for  $\mathbb{P}(Y_1 \in A|T = a)$  using Equations (3)-(4). For simplicity, suppose  $p_{df}$  is an interior point of  $\Theta_I(p_{df})$  such that  $0 < p_{df} < \min\{\mathbb{E}[D|Z = 0], \mathbb{E}[1 - D|Z = 1]\}$ . Equation (3) implies

$$\mathbb{P}(Y_1 \in A|T = a) = \frac{\mathbb{P}(Y \in A, D = 1|Z = 1) - p_c \mathbb{P}(Y_1 \in A|T = c)}{p_a}.$$

Since  $\mathbb{P}(Y_1 \in A|T = c) \in [0, 1]$ , and  $\mathbb{P}(Y_1 \in A|T = a) \in [0, 1]$ , we have

$$\max \left\{ \frac{\mathbb{P}(Y \in A, D = 1|Z = 1) - p_c}{p_a}, 0 \right\} \leq \mathbb{P}(Y_1 \in A|T = a) \leq \min \left\{ \frac{\mathbb{P}(Y \in A, D = 1|Z = 1)}{p_a}, 1 \right\}.$$

Similarly, Equation (4) implies

$$\max \left\{ \frac{\mathbb{P}(Y \in A, D = 1|Z = 0) - p_{df}}{p_a}, 0 \right\} \leq \mathbb{P}(Y_1 \in A|T = a) \leq \min \left\{ \frac{\mathbb{P}(Y \in A, D = 1|Z = 0)}{p_a}, 1 \right\},$$

Hence, Equations (3)-(4) together imply

$$\begin{aligned} & \max \left\{ \frac{\mathbb{P}(Y \in A, D = 1|Z = 1) - p_c}{p_a}, \frac{\mathbb{P}(Y \in A, D = 1|Z = 0) - p_{df}}{p_a}, 0 \right\} \\ & \leq \mathbb{P}(Y_1 \in A|T = a) \leq \\ & \min \left\{ \frac{\mathbb{P}(Y \in A, D = 1|Z = 1)}{p_a}, \frac{\mathbb{P}(Y \in A, D = 1|Z = 0)}{p_a}, 1 \right\}. \end{aligned} \tag{13}$$

A similar reasoning holds for (5)-(6) and helps partially identify  $\mathbb{P}(Y_0 \in A|T = n)$ . The identified sets for  $F_{dt} \equiv F_{Y_d|T=t}$ ,  $d \in \{0, 1\}$ ,  $t \in \{c, df\}$  are given in Proposition 2.

**Proposition 2.** For a given  $p_{df}$  interior point of  $\Theta_I(p_{df})$ , pointwise sharp bounds for the distributions  $F_{dt}$  are given below:

$$\begin{aligned}
F_{1a}^{LB}(y) &\equiv \max \left\{ F_{1a}^{LB_1}(y), F_{1a}^{LB_0}(y) \right\} \leq F_{1a}(y) \leq \min \left\{ F_{1a}^{UB_1}(y), F_{1a}^{UB_0}(y) \right\} \equiv F_{1a}^{UB}(y), \\
F_{0n}^{LB}(y) &\equiv \max \left\{ F_{0n}^{LB_1}(y), F_{0n}^{LB_0}(y) \right\} \leq F_{0n}(y) \leq \min \left\{ F_{0n}^{UB_1}(y), F_{0n}^{UB_0}(y) \right\} \equiv F_{0n}^{UB}(y), \\
F_{1c}(y) &= \frac{\mathbb{P}(Y \leq y, D = 1 | Z = 1) - p_a F_{1a}(y)}{p_c}, \quad F_{0c}(y) = \frac{\mathbb{P}(Y \leq y, D = 0 | Z = 0) - p_n F_{0n}(y)}{p_c}, \\
F_{1df}(y) &= \frac{\mathbb{P}(Y \leq y, D = 1 | Z = 0) - p_a F_{1a}(y)}{p_{df}}, \quad F_{0df}(y) = \frac{\mathbb{P}(Y \leq y, D = 0 | Z = 1) - p_n F_{0n}(y)}{p_{df}},
\end{aligned}$$

where  $F_{1a}^\ell(y), F_{0n}^\ell(y), \ell \in \{LB_0, LB_1, UB_0, UB_1\}$  are defined in Appendix A.2, and  $p_a = \mathbb{E}[D | Z = 0] - p_{df}, p_n = \mathbb{E}[1 - D | Z = 1] - p_{df}, p_c = 1 - p_{df} - p_n - p_a = \mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0] + p_{df}$ .

While Proposition 2 shows *pointwise* sharp bounds for the distributions, we characterize their *functional* sharp bounds (in the terminology of Mourifié, Henry, and Méango (2020)) in Appendix A.2. In the following, we explain the intuition behind the reason why the bounds for  $F_{1a}(y)$  are not functionally sharp. Suppose that we are interested in bounding  $F_{1a}(y') - F_{1a}(y)$  where  $y < y'$ . One can take the difference of the bounds in Proposition 2 and obtain

$$F_{1a}^{LB}(y') - F_{1a}^{UB}(y) \leq F_{1a}(y') - F_{1a}(y) \leq F_{1a}^{UB}(y') - F_{1a}^{LB}(y).$$

But this approach will not lead to sharp bounds for  $F_{1a}(y') - F_{1a}(y)$ . Instead, we are going to replace the Borel set  $A$  by  $(y, y']$  in Equation (13). Doing this will yield tighter bounds than the pointwise difference of the bounds in Proposition 2. Pointwise sharp bounds for  $F_{1a}(y)$  yield sharp bounds on location parameters such as the mean and the quantiles, but they will not deliver sharp bounds on spread parameters like the interquantile range. Functional sharp bounds for  $F_{1a}(y)$  yield sharp bounds for interquantile range parameters.

From the results in Proposition 2, we derive bounds on the mean potential outcomes for types. Let  $\mu_F$  denote the expected value of a given cumulative distribution function (cdf)  $F$ , and  $\mu_{dt} \equiv \mathbb{E}[Y_d | T = t]$ . Corollary 1 provides sharp bounds on  $\mu_{dt}$ .

**Corollary 1.** For a given  $p_{df}$  interior point of  $\Theta_I(p_{df})$ , sharp bounds on  $\mu_{1a}$  and  $\mu_{0n}$  are given by:

$$\begin{aligned}
\mu_{F_{1a}^{UB}}(p_a) &\leq \mu_{1a}(p_a) \leq \mu_{F_{1a}^{LB}}(p_a), \\
\mu_{F_{0n}^{UB}}(p_n) &\leq \mu_{0n}(p_n) \leq \mu_{F_{0n}^{LB}}(p_n),
\end{aligned}$$

where  $p_a = \mathbb{E}[D | Z = 0] - p_{df}, p_n = \mathbb{E}[1 - D | Z = 1] - p_{df}$ .

The sharp bounds in Corollary 1 can be difficult to compute in practice. Tractable valid

outer sets are then proposed:

$$\begin{aligned}\max \left\{ \mu_{F_{1a}^{UB_1}}, \mu_{F_{1a}^{UB_0}} \right\} &\leq \mu_{1a} \leq \min \left\{ \mu_{F_{1a}^{LB_1}}, \mu_{F_{1a}^{LB_0}} \right\} \\ \max \left\{ \mu_{F_{0n}^{UB_1}}, \mu_{F_{0n}^{UB_0}} \right\} &\leq \mu_{0n} \leq \min \left\{ \mu_{F_{0n}^{LB_1}}, \mu_{F_{0n}^{LB_0}} \right\}.\end{aligned}$$

*Remark 1.*  $\max \left\{ \mu_{F_{1a}^{UB_1}}, \mu_{F_{1a}^{UB_0}} \right\} = \mu_{F_{1a}^{UB}}$  if  $F_{1a}^{UB_1}$  first-order stochastically dominates  $F_{1a}^{UB_0}$  or vice versa. Similarly,  $\min \left\{ \mu_{F_{1a}^{LB_1}}, \mu_{F_{1a}^{LB_0}} \right\} = \mu_{F_{1a}^{LB}}$  if  $F_{1a}^{LB_1}$  first-order stochastically dominates  $F_{1a}^{LB_0}$  or vice versa. However, in general,  $\max \left\{ \mu_{F_{1a}^{UB_1}}, \mu_{F_{1a}^{UB_0}} \right\} \leq \mu_{F_{1a}^{UB}}$ , and  $\min \left\{ \mu_{F_{1a}^{LB_1}}, \mu_{F_{1a}^{LB_0}} \right\} \geq \mu_{F_{1a}^{LB}}$ .

The bounds in Corollary 1 provides sharp bounds on  $\mu_{1a}(p_a)$  and  $\mu_{0n}(p_n)$  for discrete, continuous, or mixed outcomes. Closed-form expressions for the bounds can be obtained for continuous outcomes with strictly increasing cdfs. These are called Lee's (2009) bounds. Lemma 1 displays the bounds.

**Lemma 1.** *The following holds for continuous outcomes with strictly increasing cdfs.*

$$\begin{aligned}\mu_{F_{1a}^{LBz}} &= \mathbb{E} \left[ Y | D = 1, Z = z, Y > F_{Y|D=1, Z=z}^{-1} \left( 1 - \frac{p_a}{\mathbb{E}[D|Z=z]} \right) \right], \\ \mu_{F_{1a}^{UBz}} &= \mathbb{E} \left[ Y | D = 1, Z = z, Y < F_{Y|D=1, Z=z}^{-1} \left( \frac{p_a}{\mathbb{E}[D|Z=z]} \right) \right], \\ \mu_{F_{0n}^{LBz}} &= \mathbb{E} \left[ Y | D = 0, Z = z, Y > F_{Y|D=0, Z=z}^{-1} \left( 1 - \frac{p_n}{\mathbb{E}[1-D|Z=z]} \right) \right], \\ \mu_{F_{0n}^{UBz}} &= \mathbb{E} \left[ Y | D = 0, Z = z, Y < F_{Y|D=0, Z=z}^{-1} \left( \frac{p_n}{\mathbb{E}[1-D|Z=z]} \right) \right].\end{aligned}$$

Let  $\mathring{\Theta}_I(p_{df})$  denote the interior of  $\Theta_I(p_{df})$ . The following theorem proposes sharp bounds on the LATEs for compliers  $\theta_{0c} = \theta_{1c}$  and defiers  $\theta_{0df} = \theta_{1df}$ .

**Theorem 1.** *Suppose inequality (12) holds, and  $p_{df}$  is an interior point of  $\Theta_I(p_{df})$ . Under Assumptions 1 and 2, sharp bounds for  $\mu_{1c}$  and  $\mu_{1df}$  are as follows:*

$$\mu_{1c}(p_a) = \frac{\mathbb{E}[YD|Z=1] - p_a \mu_{1a}(p_a)}{\mathbb{E}[D|Z=1] - p_a}, \quad \mu_{1df}(p_a) = \frac{\mathbb{E}[YD|Z=0] - p_a \mu_{1a}(p_a)}{\mathbb{E}[D|Z=0] - p_a},$$

where  $p_a = \mathbb{E}[D|Z=0] - p_{df}$ ,  $p_{df} \in \mathring{\Theta}_I(p_{df})$ , and  $\mu_{1a}(p_a) \in \left[ \mu_{F_{1a}^{UB}}(p_a), \mu_{F_{1a}^{LB}}(p_a) \right]$ .

Similarly, sharp bounds for  $\mu_{0c}$  and  $\mu_{0df}$  are given by:

$$\mu_{0c}(p_n) = \frac{\mathbb{E}[Y(1-D)|Z=0] - p_n\mu_{0n}(p_n)}{\mathbb{E}[1-D|Z=0] - p_n}, \quad \mu_{0df}(p_n) = \frac{\mathbb{E}[Y(1-D)|Z=1] - p_n\mu_{0n}(p_n)}{\mathbb{E}[1-D|Z=1] - p_n},$$

where  $p_n = \mathbb{E}[1-D|Z=1] - p_{df}$ ,  $p_{df} \in \mathring{\Theta}_I(p_{df})$ ,  $\mu_{0n}(p_n) \in [\mu_{F_{0n}}^{UB}(p_n), \mu_{F_{0n}}^{LB}(p_n)]$ .

Sharp bounds for  $\theta_{0c}$ ,  $\theta_{1c}$ ,  $\theta_{0df}$ ,  $\theta_{1df}$  are given by:

$$\theta_{0c}(p_c) = \theta_{1c}(p_c) = \mu_{1c}(p_c) - \mu_{0c}(p_c), \quad \theta_{0df}(p_{df}) = \theta_{1df}(p_{df}) = \mu_{1df}(p_{df}) - \mu_{0df}(p_{df}),$$

where  $p_c = 1 - p_{df} - p_n - p_a = \mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0] + p_{df}$ , and  $p_{df} \in \mathring{\Theta}_I(p_{df})$ .

*Proof.* See detailed proofs in Appendix A.2. □

Theorem 1 provides sharp bounds on the LATEs for compliers and defiers under random assignment and exclusion restriction, without requiring any kind of monotonicity assumption. The bounds differ from those in [Huber, Laffers, and Mellace \(2017\)](#) in two ways. First, they are derived under a stronger assumption (random assignment instead of mean independence assumption). Second, as we previously discuss, these bounds take into account the testable implication (12), while the [Huber, Laffers, and Mellace \(2017\)](#) bounds do not take into account the mean independence version of this inequality. As explained in [Li, Kédagni, and Mourifié \(2024\)](#), these kinds of outer bounds could be misleading.

Proposition 2 and Theorem 1 differ from the results in [Kitagawa \(2021\)](#). While [Kitagawa \(2021\)](#) studies the identified sets of potential outcome distributions and derives sharp bounds for the average treatment effect, our focus is on the identified set of the vector parameter  $\Gamma$ , as it is a building block for many commonly used parameters, including the average and local average treatment effects. The proposed approach in Theorem 1 also differs from other existing papers ([de Chaisemartin, 2017](#); [Noack, 2021](#); [Dahl, Huber, and Mellace, 2023](#)) as it does not assume any weaker version of the monotonicity assumption 3.

The identified set for  $\Gamma$  is

$$\Theta_I(A_2) = \begin{cases} \Theta_I^1(A_2), & \text{if } p_{df} \in \Theta_I(p_{df}) \cap (0, \min\{\mathbb{E}[D|Z=0], \mathbb{E}[1-D|Z=1]\}), \\ \Theta_I^2(A_2) = \Theta_I(A_1), & \text{if } p_{df} = 0 \in \Theta_I(p_{df}), \\ \Theta_I^3(A_2), & \text{if } p_{df} = \min\{\mathbb{E}[D|Z=0], \mathbb{E}[1-D|Z=1]\}, \\ \emptyset, & \text{if } p_{df} \notin \Theta_I(p_{df}), \end{cases}$$

where

$$\begin{aligned} \Theta_I^1(A_2) = & \left\{ \Gamma \in \mathbb{R}^{20} : \theta_{0a}(p_a) = \theta_{1a}(p_a) = \mu_{1a}(p_a) - \mu_{0a}, \theta_{0n}(p_n) = \theta_{1n}(p_n) = \mu_{1n} - \mu_{0n}(p_n), \right. \\ & \theta_{0c}(p_c) = \theta_{1c}(p_c) = \frac{\mathbb{E}[YD|Z=1] - p_a\mu_{1a}(p_a)}{\mathbb{E}[D|Z=1] - p_a} - \frac{\mathbb{E}[Y(1-D)|Z=0] - p_n\mu_{0n}(p_n)}{\mathbb{E}[1-D|Z=0] - p_n}, \\ & \theta_{0df}(p_{df}) = \theta_{1df}(p_{df}) = \frac{\mathbb{E}[YD|Z=0] - p_a\mu_{1a}(p_a)}{\mathbb{E}[D|Z=0] - p_a} - \frac{\mathbb{E}[Y(1-D)|Z=1] - p_n\mu_{0n}(p_n)}{\mathbb{E}[1-D|Z=1] - p_n}, \\ & \delta_{0a} = \delta_{1a} = \delta_{0c} = \delta_{1c} = \delta_{0df} = \delta_{1df} = \delta_{0n} = \delta_{1n} = 0, \\ & p_a = \mathbb{E}[D|Z=0] - p_{df}, p_c = \mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0] + p_{df}, \\ & p_{df} \in \Theta_I(p_{df}) \cap (0, \min\{\mathbb{E}[D|Z=0], \mathbb{E}[1-D|Z=1]\}), p_n = \mathbb{E}[1-D|Z=1] - p_{df}, \\ & \mu_{1a}(p_a) \in [\mu_{F_{1a}}^{UB}(p_a), \mu_{F_{1a}}^{LB}(p_a)], \mu_{0n}(p_n) \in [\mu_{F_{0n}}^{UB}(p_n), \mu_{F_{0n}}^{LB}(p_n)], \\ & \left. \mu_{1n}, \mu_{0a} \in [\inf \mathcal{Y}, \sup \mathcal{Y}] \right\}, \end{aligned}$$

which arises when  $p_{df} > 0$  is in the interior of the identified set for the proportion of defiers,  $\Theta_I(p_{df})$ . When  $p_{df}$  is equal to 0, the identified set for  $\Gamma$  coincides with  $\Theta_I(A_1)$ , which corresponds to the setting where all assumptions in  $A_1$  (RA, ER, MON) hold. When  $p_{df}$  is equal to the upper bound of  $\Theta_I(p_{df})$ , the identified set for  $\Gamma$  under  $A_2$  is  $\Theta_I^3(A_2)$ , which is formally defined in Appendix A.2.

## 4.1 Numerical Illustration

In this section, we illustrate how informative our proposed bounds can be in the context of a double hurdle model for a given data generating process. This is an example where our bounds correctly identify the sign of the LATEs for compliers and defiers separately while the IV estimand yields the wrong sign. Consider the following data-generating process<sup>2</sup>

$$\begin{cases} Y &= \beta D + U \\ D &= \mathbb{1}\{V_1 \leq 2Z, V_2 > Z\} \\ Z &= \mathbb{1}\{\epsilon > 0\} \end{cases}$$

---

<sup>2</sup>Note that this is a version of the double hurdle model considered in the introduction, since we can equivalently write  $D$  as follows:

$D = \mathbb{1}\{V_1 \leq 2Z, -V_2 < -Z\} = \mathbb{1}\{\Phi(V_1) \leq \Phi(2Z), \Phi(-V_2) < \Phi(-Z)\} = \mathbb{1}\{\tilde{V}_1 \leq Q_1(Z), \tilde{V}_2 < Q_2(Z)\}$ , where  $\tilde{V}_1 \equiv \Phi(V_1)$ ,  $\tilde{V}_2 \equiv \Phi(-V_2)$ ,  $Q_1(Z) \equiv \Phi(2Z)$ ,  $Q_2(Z) \equiv \Phi(-Z)$ .

Parameters	Formulas	Value
$p_{df}$	$P(D_0 = 1, D_1 = 0)$	0.1708
$p_a$	$P(D_0 = 1, D_1 = 1)$	0.0793
$p_c$	$P(D_0 = 0, D_1 = 0)$	0.0756
$p_n$	$P(D_0 = 0, D_1 = 1)$	0.6743
$LATE_c$	$E[Y_1 T = c] - E[Y_0 T = c]$	4.9253
$LATE_{df}$	$E[Y_1 T = df] - E[Y_0 T = df]$	1.2317
	$E[D Z = 0] - E[D Z = 1]$	0.0955
IV estimand	$\frac{E[Y Z=1] - E[Y Z=0]}{E[D Z=1] - E[D Z=0]}$	-1.6874

**Table 1:** Numerical results (sample size  $n = 10,000,000$ )

where  $\beta = 5\Phi(2V_1 + V_2)$ ,  $U = \frac{1}{2}(V_1 + V_2)$ ,  $(V_1, V_2, \varepsilon)' \sim N(0, I)$ , and  $\Phi(\cdot)$  is the standard normal cdf.

The potential treatments are then defined as

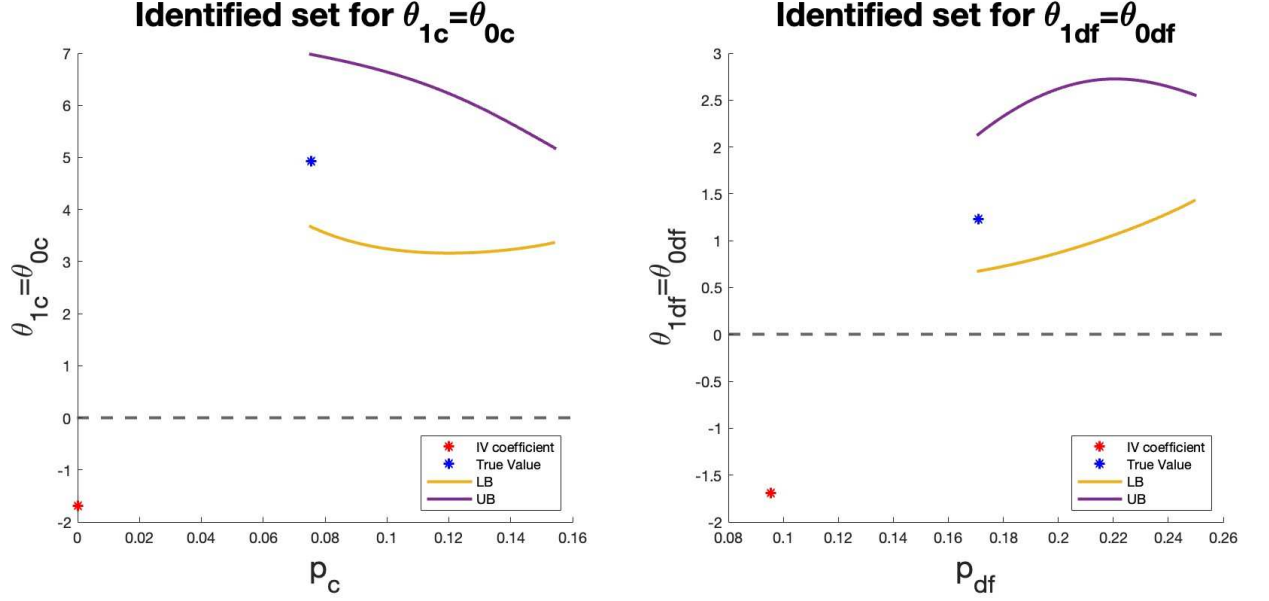
$$\begin{cases} D_0 &= \mathbb{1}\{V_1 \leq 0, V_2 > 0\}, \\ D_1 &= \mathbb{1}\{V_1 \leq 2, V_2 > 1\}. \end{cases}$$

In Table 1, we can compute the following some important parameters from this DGP. In this DGP, the true causal effects for compliers and defiers are  $LATE_c \equiv \mathbb{E}[Y_1 - Y_0|T = c] = 4.93$  and  $LATE_{df} \equiv \mathbb{E}[Y_1 - Y_0|T = df] = 1.23$ . Our lower bounds for compliers and defiers from Figure 2 are strictly larger than zero, showing our bound can clearly identify the positive sign and reasonable regions of LATE for two subgroups of people. However, when we add the IV estimand to Figure 2, we can see that the IV estimand is worse than our bounds under this DGP in two dimensions: magnitude and sign. The IV estimand is lower than the lower bounds. Also, compared to the positive true LATE for defiers ( $LATE_{df} = 1.23$ ), the sign of the IV estimand ( $LATE_{IV} = -1.69$ ) is negative, which is misleading to policymakers. Indeed, when there exist both compliers and defiers in the population, the IV estimand is a weighted average of the LATEs for compliers and defiers, with negative weights.<sup>3</sup> These negative weights lead to sign reversal of the IV estimand in this example.

---

<sup>3</sup>IV estimand =  $\frac{p_c LATE_c - p_{df} LATE_{df}}{p_c - p_{df}}$ .





**Figure 2:** Estimated bounds for LAICDEs, true values for LAICDEs, and IV estimand

## 5 Identification under random assignment and monotonicity

In this section, the potential outcome  $Y_{dz}$  is allowed to vary with  $z$  in such a way that we may have  $\theta_{0t} \neq \theta_{1t}$  and  $\delta_{dt} \neq 0$ . This section demonstrates how researchers can use a randomly assigned instrument that may violate the exclusion assumption 2. We maintain the monotonicity assumption 3. To proceed, as in the previous section, we write the identified probabilities in each of the observed four subgroups  $\{(D = d, Z = z) : z, d \in \{0, 1\}\}$  as a mixture of the potential outcome distributions for the unobserved types. More precisely, under Assumption 1, Equations (3)-(6) become

$$\mathbb{P}(Y \in A, D = 1|Z = 1) = p_c \mathbb{P}(Y_{11} \in A|T = c) + p_a \mathbb{P}(Y_{11} \in A|T = a), \quad (14)$$

$$\mathbb{P}(Y \in A, D = 1|Z = 0) = p_{df} \mathbb{P}(Y_{10} \in A|T = df) + p_a \mathbb{P}(Y_{10} \in A|T = a), \quad (15)$$

$$\mathbb{P}(Y \in A, D = 0|Z = 1) = p_{df} \mathbb{P}(Y_{01} \in A|T = df) + p_n \mathbb{P}(Y_{01} \in A|T = n), \quad (16)$$

$$\mathbb{P}(Y \in A, D = 0|Z = 0) = p_c \mathbb{P}(Y_{00} \in A|T = c) + p_n \mathbb{P}(Y_{00} \in A|T = n). \quad (17)$$

Suppose first that  $\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] > 0$ . Then, under Assumption 3, there are no defiers, i.e.,  $p_{df} = 0$ . As a result,  $p_c = \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]$ ,  $p_a = \mathbb{E}[D|Z = 0]$ , and

$p_n = \mathbb{E}[1 - D|Z = 1]$ . Equation (14) implies

$$\mathbb{P}(Y_{11} \in A|T = c) = \frac{\mathbb{P}(Y \in A, D = 1|Z = 1) - p_a \mathbb{P}(Y_{11} \in A|T = a)}{p_c}.$$

Since  $\mathbb{P}(Y_{11} \in A|T = c) \in [0, 1]$ , and  $\mathbb{P}(Y_{11} \in A|T = a) \in [0, 1]$ , we have

$$\max \left\{ \frac{\mathbb{P}(Y \in A, D = 1|Z = 1) - p_c}{p_a}, 0 \right\} \leq \mathbb{P}(Y_{11} \in A|T = a) \leq \min \left\{ \frac{\mathbb{P}(Y \in A, D = 1|Z = 1)}{p_a}, 1 \right\}.$$

Equation (15) implies  $\mathbb{P}(Y_{10} \in A|T = a) = \mathbb{P}(Y \in A|D = 1, Z = 0)$ , while Equation (16) implies  $\mathbb{P}(Y_{01} \in A|T = n) = \mathbb{P}(Y \in A|D = 0, Z = 1)$ . Finally, Equation (17) implies

$$\max \left\{ \frac{\mathbb{P}(Y \in A, D = 0|Z = 0) - p_n}{p_n}, 0 \right\} \leq \mathbb{P}(Y_{00} \in A|T = n) \leq \min \left\{ \frac{\mathbb{P}(Y \in A, D = 0|Z = 0)}{p_n}, 1 \right\},$$

and

$$\mathbb{P}(Y_{00} \in A|T = c) = \frac{\mathbb{P}(Y \in A, D = 0|Z = 0) - p_n \mathbb{P}(Y_{00} \in A|T = n)}{p_c}.$$

The above bounding approach builds on [Horowitz and Manski \(1995\)](#). We can then suitably take the expectations of the bounds and obtain the so-called [Lee \(2009\)](#) bounds. Let  $\mu_{dzt} \equiv \mathbb{E}[Y_{dz}|T = t]$ ,  $F_{dzt} \equiv F_{Y_{dz}|T=t}$ . The following proposition 3 derives sharp bounds on the distributions  $F_{dzt}$ .

**Proposition 3.** Suppose Assumptions 1 and 3 hold, and  $\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] > 0$ . Then,

$$\begin{aligned}
p_c &= \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0], \quad p_a = \mathbb{E}[D|Z = 0], \quad p_n = \mathbb{E}[1 - D|Z = 1], \quad p_{df} = 0, \\
F_{11a}^{LB}(y) &\equiv \max \left\{ \frac{\mathbb{P}(Y \leq y, D = 1|Z = 1) - p_c}{p_a}, 0 \right\} \leq F_{11a}(y) \\
&\leq \min \left\{ \frac{\mathbb{P}(Y \leq y, D = 1|Z = 1)}{p_a}, 1 \right\} \equiv F_{11a}^{UB}(y), \\
F_{00n}^{LB}(y) &\equiv \max \left\{ \frac{\mathbb{P}(Y \leq y, D = 0|Z = 0) - p_c}{p_n}, 0 \right\} \leq F_{00n}(y) \\
&\leq \min \left\{ \frac{\mathbb{P}(Y \leq y, D = 0|Z = 0)}{p_n}, 1 \right\} \equiv F_{00n}^{UB}(y), \\
F_{11c}(y) &= \frac{\mathbb{P}(Y \leq y, D = 1|Z = 1) - p_a F_{11a}(y)}{p_c}, \\
F_{00c}(y) &= \frac{\mathbb{P}(Y \leq y, D = 0|Z = 0) - p_n F_{00n}(y)}{p_c}, \\
F_{10a}(y) &= \mathbb{P}(Y \leq y|D = 1, Z = 0), \quad F_{01n}(y) = \mathbb{P}(Y \leq y|D = 0, Z = 1).
\end{aligned}$$

These above bounds are pointwise sharp.

Proposition 3 provides important results that help derive sharp bounds on the local average treatment-controlled direct effects for the always-takers and never-takers. A direct implication of this proposition is Corollary 3 in Appendix B, which derives sharp bounds on the parameters  $\mu_{dzt}$ . Building on the results in Corollary 3, Corollary 2 below provides sharp bounds on the parameters  $\delta_{1a}$ ,  $\delta_{0n}$ , and  $\theta_{1c} + \delta_{0c} = \theta_{0c} + \delta_{1c}$ . To the best of our knowledge, this is the first paper to formally derive sharp bounds on these parameters under the random assignment and monotonicity assumptions. Note however that  $\delta_{1a}$  and  $\delta_{0n}$  coincide with the local net average treatment effect (LNATE) for the always-takers and never-takers, respectively, defined in Flores and Flores-Lagunes (2013). In general,  $LNATE_t \equiv \mathbb{E}[Y_{D01} - Y_{D00}|T = t]$  is different from our LATCDE parameter  $\delta_{dt}$ .

**Corollary 2.** Suppose Assumptions 1 and 3 hold, and  $\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] > 0$ . Then,

$$\begin{aligned}
p_c &= \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0], \quad p_a = \mathbb{E}[D|Z = 0], \quad p_n = \mathbb{E}[1 - D|Z = 1], \quad p_{df} = 0, \\
\mu_{F_{11a}}^{UB} - \mathbb{E}[Y|D = 1, Z = 0] &\leq \delta_{1a} \leq \mu_{F_{11a}}^{LB} - \mathbb{E}[Y|D = 1, Z = 0], \\
\mathbb{E}[Y|D = 0, Z = 1] - \mu_{F_{00n}}^{LB} &\leq \delta_{0n} \leq \mathbb{E}[Y|D = 0, Z = 1] - \mu_{F_{00n}}^{UB}, \\
\delta_{0c} + \theta_{1c} = \delta_{1c} + \theta_{0c} &= \frac{\mathbb{E}[YD|Z = 1] - p_a \mu_{11a}}{p_c} - \frac{\mathbb{E}[Y(1 - D)|Z = 0] - p_n \mu_{00n}}{p_c}, \\
\mu_{F_{11a}}^{UB} \leq \mu_{11a} \leq \mu_{F_{11a}}^{LB}, \quad \mu_{F_{00n}}^{UB} \leq \mu_{00n} \leq \mu_{F_{00n}}^{LB}.
\end{aligned}$$

These bounds are sharp.

Proposition 3 and Corollaries 2-3 are intermediate results that are key to deriving the identified set for the vector of parameters  $\Gamma$ . The identified set for  $\Gamma$  is

$$\Theta_I(A_3) = \begin{cases} \Theta_I^1(A_3), & \text{if } \mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0] > 0, \\ \Theta_I^2(A_3), & \text{if } \mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0] < 0, \\ \Theta_I^3(A_3), & \text{if } \mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0] = 0, \end{cases}$$

where

$$\begin{aligned} \Theta_I^1(A_3) = & \left\{ \Gamma \in \mathbb{R}^{20} : \theta_{zt} = \mu_{1zt} - \mu_{0zt}, t \in \{a, c, df, n\}, z \in \{0, 1\}, \right. \\ & \delta_{dt} = \mu_{d1t} - \mu_{d0t}, t \in \{a, c, df, n\}, d \in \{0, 1\}, \\ & \mu_{10a} = \mathbb{E}[Y|D=1, Z=0], \mu_{01n} = \mathbb{E}[Y|D=0, Z=1], \\ & \mu_{F_{11a}}^{UB} \leq \mu_{11a} \leq \mu_{F_{11a}}^{LB}, \mu_{F_{00n}}^{UB} \leq \mu_{00n} \leq \mu_{F_{00n}}^{LB}, \\ & \mu_{11c} = \frac{\mathbb{E}[YD|Z=1] - p_a \mu_{11a}}{p_c}, \mu_{00c} = \frac{\mathbb{E}[Y(1-D)|Z=0] - p_n \mu_{00n}}{p_c}, \\ & \mu_{10n}, \mu_{11n}, \mu_{01a}, \mu_{00a}, \mu_{10c}, \mu_{01c}, \mu_{11df}, \mu_{10df}, \mu_{01df}, \mu_{00df} \in [\inf \mathcal{Y}, \sup \mathcal{Y}], \\ & \left. p_a = \mathbb{E}[D|Z=0], p_c = \mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0], p_{df} = 0, p_n = \mathbb{E}[1-D|Z=1] \right\}, \end{aligned}$$

$\Theta_I^2(A_3) = \tilde{\Theta}_I^1(A_3)$ , where  $\tilde{\Theta}_I^1(A_3)$  is the corresponding  $\Theta_I^1(A_3)$  defined for  $\tilde{\Gamma}$  based on  $(Y, D, \tilde{Z})$  and  $\tilde{T}$ , and

$$\begin{aligned} \Theta_I^3(A_3) = & \left\{ \Gamma \in \mathbb{R}^{20} : \theta_{zt} = \mu_{1zt} - \mu_{0zt}, t \in \{a, c, df, n\}, z \in \{0, 1\}, \right. \\ & \delta_{dt} = \mu_{d1t} - \mu_{d0t}, t \in \{a, c, df, n\}, d \in \{0, 1\}, \\ & \mu_{10a} = \mathbb{E}[Y|D=1, Z=0], \mu_{01n} = \mathbb{E}[Y|D=0, Z=1], \\ & \mu_{11a} = \mathbb{E}[Y | D=1, Z=1], \mu_{00n} = \mathbb{E}[Y | D=0, Z=0], \\ & \mu_{10n}, \mu_{11n}, \mu_{01a}, \mu_{00a}, \mu_{10c}, \mu_{01c}, \mu_{11c}, \mu_{00c}, \mu_{11df}, \mu_{10df}, \mu_{01df}, \mu_{00df} \in [\inf \mathcal{Y}, \sup \mathcal{Y}], \\ & \left. p_a = \mathbb{E}[D], p_c = p_{df} = 0, p_n = \mathbb{E}[1-D] \right\}. \end{aligned}$$

*Remark 2.* We can express  $\mathbb{E}[Y \mid Z = 1]$  as

$$\begin{aligned}\mathbb{E}[Y \mid Z = 1] &= \mathbb{E}[Y_{11}D_1 + Y_{01}(1 - D_1) \mid Z = 1] \\ &= \mathbb{E}[Y_{11}D_1 + Y_{01}(1 - D_1)] \\ &= \mathbb{E}[Y_{11} \mid T = a]p_a + \mathbb{E}[Y_{11} \mid T = c]p_c + \mathbb{E}[Y_{01} \mid T = n]p_n,\end{aligned}$$

where the second equality holds under Assumption 1, and we can apply the law of iterated expectation to obtain the third equality. Using the similar reasoning, we can also write  $\mathbb{E}[Y \mid Z = 0]$  as

$$\begin{aligned}\mathbb{E}[Y \mid Z = 0] &= \mathbb{E}[Y_{10}D_0 + Y_{00}(1 - D_0) \mid Z = 0] \\ &= \mathbb{E}[Y_{10}D_0 + Y_{00}(1 - D_0)] \\ &= \mathbb{E}[Y_{10} \mid T = a]p_a + \mathbb{E}[Y_{00} \mid T = c]p_c + \mathbb{E}[Y_{00} \mid T = n]p_n.\end{aligned}$$

As is customary, define the intent-to-treat (ITT) parameter as  $\text{ITT} \equiv \mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]$ . Then

$$\begin{aligned}\text{ITT} &= \mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0] \\ &= \mathbb{E}[Y_{11} - Y_{10} \mid T = a]p_a + \mathbb{E}[Y_{01} - Y_{00} \mid T = n]p_n + \mathbb{E}[Y_{11} - Y_{00} \mid T = c]p_c \\ &= \delta_{1a}p_a + \delta_{0n}p_n + [\delta_{1c} + \theta_{0c}]p_c.\end{aligned}$$

Since  $p_a, p_n, p_c \geq 0$ , and  $p_a + p_n + p_c = 1$  under Assumption 3, we have shown that ITT can be expressed as a convex combination of  $\delta_{1a}$ ,  $\delta_{0n}$ , and  $\delta_{1c} + \theta_{0c}$ .

*Remark 3.* (Testable implication) In Corollary 3, we point identify  $\mathbb{E}[Y_{10} \mid T = a]$  and  $\mathbb{E}[Y_{01} \mid T = n]$ , and partially identify  $\mathbb{E}[Y_{11} \mid T = a]$  and  $\mathbb{E}[Y_{00} \mid T = n]$ . Those identification results can be exploited as a testable implication for the exclusion restriction. Since the exclusion restriction states that  $Y_{11} = Y_{10}$  and  $Y_{01} = Y_{00}$ , it implies that  $\mathbb{E}[Y_{11} \mid T = a] = \mathbb{E}[Y_{10} \mid T = a]$  and  $\mathbb{E}[Y_{01} \mid T = n] = \mathbb{E}[Y_{00} \mid T = n]$ . Therefore, if the point identified  $\mathbb{E}[Y_{10} \mid T = a]$  is not in the identified set of  $\mathbb{E}[Y_{11} \mid T = a]$ , or the point identified  $\mathbb{E}[Y_{01} \mid T = n]$  is not contained in the identified set of  $\mathbb{E}[Y_{00} \mid T = n]$ , we have enough evidence to reject the exclusion restriction assumption 2. However, this testable implication is not sharp for testing the exclusion restriction assumption 2. Kitagawa (2021, Proposition 3.2 (ii)) showed that inequality (12) is the sharp testable implication for the random assignment and exclusion restriction assumptions 1 and 2 together. This result is confirmed in the sharpness proof of Proposition 1 in this paper. In a randomized experiment setting like the one considered in this paper, the testable inequality (12) can be interpreted as a testable implication for only the exclusion restriction assumption.

*Remark 4.* Notice that unlike  $\Theta_I(A_1)$  and  $\Theta_I(A_2)$  which could be empty,  $\Theta_I(A_3)$  is never empty. This implies that Assumptions 1 and 3 do not have any testable implication.

## 6 Misspecification robust bounds

In this section, we follow Li, Kédagni, and Mourifié (2024) to derive a misspecification robust bound for the model and the various assumptions considered in this paper. This is a discrete way of salvaging the model when it is refuted by the data. This relaxation is different from the continuous relaxation approach introduced in Masten and Poirier (2021). We find this discrete relaxation easier to interpret than the continuous one. Before we proceed to show the misspecification robust bound for the collection of assumptions  $A$ , we discuss when each of the assumptions  $A_1$ ,  $A_2$ , and  $A_3$  is data-consistent. According to Li, Kédagni, and Mourifié (2024), an assumption  $A_k$  is data-consistent if its identified set is nonempty. Therefore,  $A_1$  is data-consistent if  $\Theta_I^1(A_1)$  is nonempty when  $\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0] > 0$ ,  $\Theta_I^2(A_1)$  is nonempty when  $\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0] < 0$ , and  $\Theta_I^3(A_1)$  is nonempty when  $\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0] = 0$ . Assumption  $A_2$  is data-consistent if inequality (12) holds. Assumption  $A_3$  is always data-consistent as its identified set is never empty.

**Definition 1** (Li, Kédagni, and Mourifié, 2024). An assumption  $A_k \in A$  is a minimum data-consistent relaxation of  $A$  if  $\Theta_I(A_k) \neq \emptyset$  and  $\Theta_I(A_k \cup A_{k'}) = \emptyset$  for any  $A_{k'} \neq A_k : A_{k'} \in A$ .

The misspecification robust bound  $\Theta_I^*(A)$  is defined as the union of the identified sets of *all* minimum data-consistent relaxations of  $A$ . Note here that  $A_2 \cup A_3 = A_1$ . Then  $\Theta_I(A) = \Theta_I(A_1 \cup A_2 \cup A_3) = \Theta_I(A_1)$ . If  $\Theta(A) \neq \emptyset$ , there exists a unique minimum data-consistent relaxation by definition, it is  $A_1$ . But, if  $\Theta_I(A_1) = \emptyset$  and  $\Theta_I(A_2) \neq \emptyset$ , we have two minimum data-consistent relaxations,  $A_2$  and  $A_3$ . Finally, if  $\Theta(A_2) = \emptyset$  (i.e., inequality (12) is violated), then  $A_3$  will be the only minimum data-consistent relaxation. Hence, we have

$$\begin{aligned} \Theta_I^*(A) &= \begin{cases} \Theta_I(A_1) & \text{if } \Theta_I(A_1) \neq \emptyset, \\ \Theta_I(A_2) \cup \Theta_I(A_3) & \text{if } \Theta_I(A_1) = \emptyset \text{ and } \Theta_I(A_2) \neq \emptyset, \\ \Theta_I(A_3) & \text{if } \Theta_I(A_2) = \emptyset \end{cases} \\ &= \begin{cases} \Theta_I(A_1) & \text{if } \Theta_I(A_1) \neq \emptyset, \\ \Theta_I(A_2) \cup \Theta_I(A_3) & \text{if } \Theta_I(A_1) = \emptyset. \end{cases} \end{aligned}$$

In the next section, we illustrate the theoretical results with a real-world application.

## 7 Empirical illustration

In this section, we illustrate our methodology by revisiting [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#), which is also re-examined by [Kwon and Roth \(2024\)](#) in a mechanism framework. [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#) study how misperceived social norms affect women’s labor force participation in Saudi Arabia. They find that while many men privately support female employment, they incorrectly believe that their peers do not share this view. To study the effects of correcting these misperceptions, the authors design a randomized experiment in which men are randomly informed that a majority of their peers actually support female employment. The results show that women whose husbands receive the information are significantly more likely to register for a job-search service. In addition, the study finds persistent effects on women’s employment behavior, as women from treated households are more likely to be employed several months after the intervention.

To align this study with our framework, we define the women long-term labor supply outcome<sup>4</sup> as the outcome  $Y$ , the decision to sign up for a job-search service as the binary treatment variable  $D$ , and the information regarding social norms as a randomized experiment  $Z$ . Given that the information experiments in this study are randomly assigned, we maintain Assumption 1 throughout the analysis. The dataset applied here comes from the online appendix of [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#). Our sample is restricted to observations with non-missing data on women’s long-term labor supply outcomes, resulting in a final sample size of 381 observations. Some summary statistics are provided in Table 2. We observe that households assigned to the treatment group ( $Z = 1$ ) are more likely to enroll in job-search services compared to those in the control group ( $Z = 0$ ). Additionally, treated households, on average, have better long-term female labor force outcomes than the control households.

The results presented in this section are entirely based on estimation. We provide a brief discussion of the inference methods and present the corresponding inference results in Appendix D.

### 7.1 Identified set under $A_1$

We estimate the difference in probabilities,  $\mathbb{P}(D = 1 \mid Z = 1) - \mathbb{P}(D = 1 \mid Z = 0)$ , and obtain a result of  $\hat{\mathbb{P}}(D = 1 \mid Z = 1) - \hat{\mathbb{P}}(D = 1 \mid Z = 0) = 0.0929 > 0$ . Given this positive estimate, we want to test the inequalities (10)-(11), which are the sharp

---

<sup>4</sup>[Bursztyn, González, and Yanagizawa-Drott \(2020\)](#) construct an index by pooling six measures of women’s long-term labor supply outcomes, which we use as the outcome in our analysis.

**Table 2:** Summary Statistics

	Total	$Z = 1$	$Z = 0$
Observations	381	190	191
$D$ (job-search service)	0.294 (0.445)	0.340 (0.475)	0.247 (0.433)
$Z$ (exposure to information)	0.501 (0.500)	1.000 (0.000)	0.000 (0.000)
$Y$ (women labor force index)	0.006 (0.499)	0.130 (0.571)	-0.119 (0.374)
Average (standard deviation).			

testable implications of the joint assumptions in  $A_1 = \{RA, ER, MON\}$ . We apply the testing method proposed in [Mourifié and Wan \(2017\)](#) and reject the validity of inequalities (10)-(11) at the 1% level. This rejection implies that the identified set  $\Theta_I(A_1)$  is empty, suggesting that the assumptions in  $A_1$  are not consistent with the dataset used in [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#).

The application in [Kwon and Roth \(2024\)](#) differs from ours. First, we use a continuous outcome variable constructed in [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#), which is an index summarizing six measures of female labor force participation, including an indicator for whether the wife applies for jobs outside the home. In contrast, [Kwon and Roth \(2024\)](#) use only this binary indicator as their outcome variable. Second, in our notation, the primary objective of [Kwon and Roth \(2024\)](#) is to test the validity of the exclusion restriction assumption 2. For their implementation, they assume monotonicity (Assumption 3) and find that the set of assumptions  $A_1$  is rejected. So, the rejection of  $A_1$  implies that the random assignment  $Z$  violates the exclusion restriction if monotonicity holds. They also estimate the average direct effects of the random assignment on the outcome of always-takers and never-takers. Finally, they perform a sensitivity analysis by gradually relaxing the monotonicity assumption, assuming a certain proportion of defiers to exist. In contrast, our method constructs misspecification robust bounds, as discussed in Section 6, by applying the minimum data-consistent relaxation of the LATE assumptions. Unlike [Kwon and Roth \(2024\)](#), we do not impose any prior restrictions on the proportion of defiers. Instead, we discretely relax the LATE assumptions in a manner that ensures they remain consistent with the observed data.

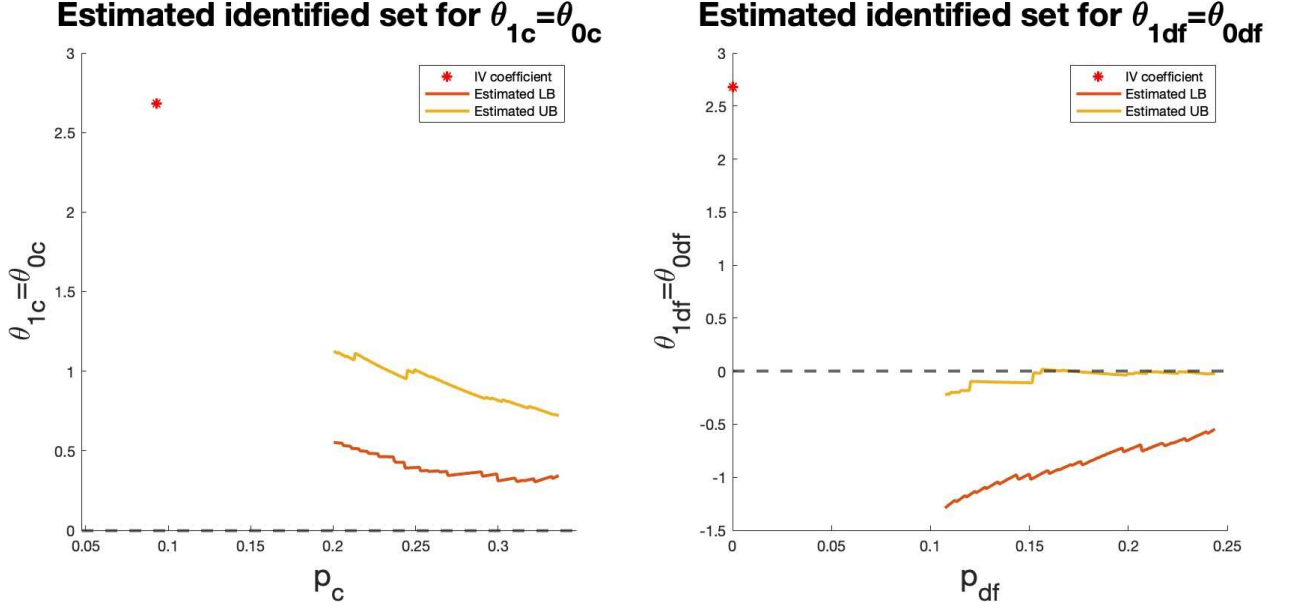
## 7.2 Identified set under $A_2$

Next, we relax Assumption 3 and consider identification under the assumption set  $A_2 = \{RA, ER\}$ . Based on Proposition 1, we estimate the identified set for the proportion of



**Table 3:** Estimated proportion of each type under  $A_2$

Parameter	Estimated identified set
$p_a$	$[0, 0.1436]$
$p_c$	$[0.1967, 0.3403]$
$p_n$	$[0.4123, 0.5559]$
$p_{df}$	$[0.1038, 0.2474]$
Observations	381



**Figure 3:** Estimated bounds for LAICDEs under  $A_2$  and IV estimate

defiers as

$$\widehat{\Theta}_I(p_{df}) = [0.1038, 0.2474].$$

Since this identified set is nonempty, Proposition 1 implies that we fail to reject the inequality (12). Therefore, the combination of assumptions in  $A_2$  cannot be rejected. Table 3 reports the estimated identified sets for the proportions of types under the assumption set  $A_2$ .

Building on the analysis in Section 4, we derive informative identified sets for the local average instrument-controlled direct effects for compliers,  $\theta_{1c} = \theta_{0c}$ , and for defiers,  $\theta_{1df} = \theta_{0df}$ . Figure 3 presents the estimated bounds for these identified sets.

Figure 3 shows the heterogeneous effects of job-search services across different household types. Specifically, we observe positive treatment effects for compliers across all values of  $p_c$  within the identified set. In complier households, individuals enroll in job-search services only when they are informed about the social norms. Upon learning that

**Table 4:** Estimated proportion of each type under  $A_3$ 

Parameter	Estimated proportion
$p_a$	0.2474
$p_c$	0.0929
$p_n$	0.7068
$p_{df}$	0
Observations	381

more women are employed than they initially thought, they may perceive an increased likelihood of getting a job, as a larger female workforce indicates greater opportunities in the labor market, which motivates them to comply. These characteristics of complier households align with the increase effects on long-term female labor force participation following their enrollment in job-search services.

In contrast, we observe negative treatment effects for defiers across most values of  $p_{df}$  within the identified set. Unlike compliers, women in defier households may regard increased female workforce as a signal of intensified competition. This may discourage them from seeking jobs, particularly if their relative advantage has diminished. Consequently, they choose not to enroll in job-search services once they are informed about the social norms. These characteristics of defier households are consistent with the decreased effects on long-term female labor force participation following enrollment in job-search services, as they may face greater competition in the female labor market.

Notably, the IV estimate (2.68) falls outside our estimated bounds. It overestimates the effect of job-search services for compliers relative to our identified set.

### 7.3 Identified set under $A_3$

Finally, we relax Assumption 2 while maintaining Assumption 3 and study identification under the assumption set  $A_3 = \{RA, MON\}$ . As discussed in Section 5, the identified set under  $A_3$  is always non-empty, implying that the joint assumptions in  $A_3$  can never be rejected.

With Assumption 3 imposed, defiers cannot exist and all proportions of types can be point identified in this framework. Table 4 presents the estimations of type proportions under  $A_3$ . The results show that never-takers constitute the largest proportion of the population, accounting for 70.68% of all observations. This finding suggests that the majority of households in the dataset would not sign up job-search services for women, regardless of whether or not they receive the information about correct social norms.

**Table 5:** Estimated informative identified sets under  $A_3$ 

Parameter	Estimated Bounds
$\delta_{1a}$	[0.0557, 0.7155]
$\delta_{0n}$	[0.0829, 0.2505]
$\delta_{1c} + \theta_{0c}$	[-0.7140, 1.9242]

According to the identified set  $\Theta_I(A_3)$ , we have informative identified sets for the local average treatment-controlled direct effects for always-takers and never-takers,  $\delta_{1a}$  and  $\delta_{0n}$ , and the total effects for compliers,  $\delta_{1c} + \theta_{0c} = \delta_{0c} + \theta_{1c}$ . Table 5 presents the estimated results for these identified sets.

The results in Table 5 indicate that the information experiment has positive direct effects on long-term female labor force outcomes for both always-takers and never-takers, who together account for nearly 90% of the population. This finding implies that, even though the information experiment does not affect job-search service enrollment decisions in these households, exposure to correct social norm information directly benefits women’s long-term labor force participation.

### Misspecification robust bound

Since  $\widehat{\Theta}_I(A_2)$  and  $\widehat{\Theta}_I(A_3)$  are nonempty, the estimated misspecification robust bound in this empirical example is  $\widehat{\Theta}_I^*(A) = \widehat{\Theta}_I(A_2) \cup \widehat{\Theta}_I(A_3)$ . We explained earlier in Section 6 how this result is achieved.

## 8 Conclusion

In this paper, we propose a robust identification approach in a randomized experiment setting with imperfect compliance where the LATE assumptions could potentially be rejected by the data. We propose bounds that are never empty and robust to detectable violations of the LATE assumptions. The bounds are also robust to violations of the monotonicity or exclusion restriction assumptions. We introduce two causal parameters: the local average treatment-controlled direct effect, and the local average instrument-controlled direct effect. We apply our methods to revisit [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#) and study the effect of randomly providing accurate social norm information on the decision to enroll women in job-search services, which may affect women’s labor force participation in Saudi Arabia. We show that the LATE assumptions are jointly rejected in this context. However, the assumptions cannot be rejected when either the monotonicity

or exclusion restriction is relaxed. Under the assumptions of random assignment and the exclusion restriction, we find that job-search service enrollment has a positive effect on long-term female labor force outcomes for compliers, whereas the effect is negative for defiers. Under the assumptions of random assignment and monotonicity, our results suggest that the information experiment itself has positive direct effects on long-term female labor force outcomes for both always-takers and never-takers, who together account for more than 90% of the population.

The approach developed in this paper works well for a binary instrument. An interesting area for future research would consist of extending the proposed method to a multivalued discrete or continuous instrument in the context of marginal treatment effects introduced in Heckman and Vytlačil (2001, 2005). One challenge is that the number of causal parameters would be infinite in this setting. Recently, Sun and Wüthrich (2022) proposed a pairwise validity approach for multivalued discrete instruments in the LATE framework. A combination of their method with the one developed in this paper could be an interesting area for future research. Another consideration for future research could be how to incorporate exogenous covariates in the analysis. While the approach still works conditional on covariates, it may suffer from the curse of dimensionality as the number of parameters grows with the cardinality of the support of the covariates. Semiparametric assumptions could also enter the set  $A$  of assumptions considered in the paper.

## Reference

- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434):444–455.
- Balke, Alexander and Judea Pearl. 1997. "Bounds on treatment effects from studies with imperfect compliance." *Journal of the American Statistical Association* 92 (439):1171–1176. URL <https://www.tandfonline.com/action/journalInformation?journalCode=uasa20>.
- Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott. 2020. "Misperceived social norms: Women working outside the home in Saudi Arabia." *American economic review* 110 (10):2997–3029.
- Cai, Zhihong, Manabu Kuroki, Judea Pearl, and Jin Tian. 2008. "Bounds on direct effects in the presence of confounded intermediate variables." *Biometrics* 64 (3):695–701.

- Dahl, Christian M, Martin Huber, and Giovanni Mellace. 2023. "It is never too LATE: a new look at local average treatment effects with or without defiers." *The Econometrics Journal* 26 (3):378–404.
- de Chaisemartin, Clément. 2017. "Tolerating defiance? Local average treatment effects without monotonicity." *Quantitative Economics* 8 (2):367–396.
- Fiorini, Mario and Katrien Stevens. 2021. "Scrutinizing the Monotonicity Assumption in IV and fuzzy RD designs." *Oxford Bulletin of Economics and Statistics* 83 (6):1475–1526.
- Flores, Carlos A and Alfonso Flores-Lagunes. 2013. "Partial identification of local average treatment effects with an invalid instrument." *Journal of Business & Economic Statistics* 31 (4):534–545.
- Heckman, James J and Rodrigo Pinto. 2018. "Unordered monotonicity." *Econometrica* 86 (1):1–35.
- Heckman, James J and Edward Vytlacil. 2001. "Policy-Relevant Treatment Effects." *American Economic Review* 91 (2):107–111.
- Heckman, James J. and Edward Vytlacil. 2005. "Structural equations, treatment effects, and econometric policy evaluation." *Econometrica* 73 (3):669–738.
- Horowitz, Joel L and Charles F Manski. 1995. "Identification and Robustness with Contaminated and Corrupted Data." *Econometrica* 63 (2):281–302.
- Huber, Martin, Lukas Laffers, and Giovanni Mellace. 2017. "Sharp IV Bounds on Average Treatment Effects on the Treated and Other Populations Under Endogeneity and Noncompliance." *Journal of Applied Econometrics* 32 (1):56–79.
- Imbens, Guido W and Joshua D Angrist. 1994. "Identification and estimation of local average treatment effects." *Econometrica* 62 (2):467–475.
- Imbens, Guido W and Charles F Manski. 2004. "Confidence intervals for partially identified parameters." *Econometrica* 72 (6):1845–1857.
- Imbens, Guido W and Donald B Rubin. 1997. "Estimating outcome distributions for compliers in instrumental variables models." *The Review of Economic Studies* 64 (4):555–574.

- Kédagni, Désiré. 2023. "Identifying treatment effects in the presence of confounded types." *Journal of Econometrics* 234 (2):479–511. URL <https://doi.org/10.1016/j.jeconom.2021.01.012>.
- Kédagni, Désiré and Ismael Mourifié. 2020. "Generalized instrumental inequalities: testing the instrumental variable independence assumption." *Biometrika* 107 (3):661–675. URL <https://academic.oup.com/biomet/article/107/3/661/5767137>.
- Kitagawa, Toru. 2015. "A Test for Instrument Validity." *Econometrica* 83 (5):2043–2063.
- . 2021. "The identification region of the potential outcome distributions under instrument independence." *Journal of Econometrics* 225 (2):231–253. URL <https://doi.org/10.1016/j.jeconom.2021.03.006>.
- Kwon, Soonwoo and Jonathan Roth. 2024. "Testing Mechanisms." *arXiv preprint arXiv:2404.11739*.
- Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76:1071–1102. URL <https://academic.oup.com/restud/article/76/3/1071/1590707>.
- Lee, Sokbae and Bernard Salanié. 2018. "Identifying Effects of Multivalued Treatments." *Econometrica* 86 (6):1939–1963.
- Li, Lixiong, Désiré Kédagni, and Ismaël Mourifié. 2024. "Discordant Relaxations of Misspecified Models." *Quantitative Economics* 15 (2):331–379.
- Manski, Charles F. 1990. "Nonparametric Bounds on Treatment Effects." *American Economic Review* 80 (2):319–323.
- . 2003. *Partial identification of probability distributions*. Springer Science & Business Media.
- . 2011. "Policy analysis with incredible certitude." *The Economic Journal* 121 (554):F261–F289.
- Masten, Matthew A. and Alexandre Poirier. 2021. "Salvaging Falsified Instrumental Variable Models." *Econometrica* 89 (3):1449–1469.
- Mourifié, Ismael, Marc Henry, and Romuald Méango. 2020. "Sharp bounds and testability of a Roy model of STEM major choices." *Journal of Political Economy* 128 (8):3220–3283.

- Mourifié, Ismael and Yuanyuan Wan. 2017. "Testing Local Average Treatment Effect Assumptions." *Review of Economics and Statistics* 99 (2):305–313. URL [http://direct.mit.edu/rest/article-pdf/99/2/305/1918591/rest\\_a\\_00622.pdf](http://direct.mit.edu/rest/article-pdf/99/2/305/1918591/rest_a_00622.pdf).
- Nevo, Aviv and Adam M. Rosen. 2012. "Identification with imperfect instruments." *Review of Economics and Statistics* 94 (3):659–671.
- Noack, Claudia. 2021. "Sensitivity of LATE Estimates to Violations of the Monotonicity Assumption." *Working paper, University of Mannheim* URL <http://arxiv.org/abs/2106.06421>.
- Richardson, Thomas S and James M Robins. 2010. "Analysis of the binary instrumental variable model." *Heuristics, Probability and Causality: A Tribute to Judea Pearl* 25:415–444.
- Small, Dylan S, Zhiqiang Tan, Roland R Ramsahai, Scott A Lorch, and M Alan Brookhart. 2017. "Instrumental variable estimation with a stochastic monotonicity assumption." *Statistical Science* 32 (4):561–579.
- Sun, Zhenting and Kaspar Wüthrich. 2022. "Pairwise Valid Instruments." *arXiv preprint arXiv:2203.08050*.

# A Proof of Identified Set Under Random Assignment and Exclusion Restriction

## A.1 Proof of Proposition 1

*Proof.* First, we demonstrate that Equation (12) is a testable implication of Assumptions 1 and 2. Given  $d \in \{0, 1\}$ , we have

$$\begin{aligned}
 \int_{\mathcal{Y}} \sup_z f_{Y,D|Z}(y, d|z) d\mu(y) &= \int_{\mathcal{Y}} \sup_z f_{Y_d, D_z|Z}(y, d|z) d\mu(y) \\
 &\leq \int_{\mathcal{Y}} \sup_z \{f_{Y_d, D_z|Z}(y, d|z) + f_{Y_d, D_z|Z}(y, 1-d|z)\} d\mu(y) \\
 &= \int_{\mathcal{Y}} \sup_z f_{Y_d|Z}(y|z) d\mu(y) \\
 &= \int_{\mathcal{Y}} f_{Y_d}(y) d\mu(y) = 1,
 \end{aligned}$$

where the first equality holds from Assumption 2, the first inequality holds because  $f_{Y_d, D_z|Z}(y, 1-d|z) \geq 0$  for any  $y$  and  $z$ , the second equality holds because  $f_{Y_d|Z}(y|z) = f_{Y_d, D_z|Z}(y, d|z) + f_{Y_d, D_z|Z}(y, 1-d|z)$  by the law of total probability, and the third equality follows from Assumption 1, and the last equality follows from the fact that a density function integrates to 1. Therefore, Assumptions 1 and 2 imply the inequality in Equation (12), and the identified set of  $p_{df}$  is empty if Equation (12) is violated.

For any Borel set  $A \subseteq \mathcal{Y}$ , under Assumptions 1 and 2, we have

$$\begin{aligned}
 &\mathbb{P}(Y \in A, D = 1 \mid Z = 0) - \mathbb{P}(Y \in A, D = 1 \mid Z = 1) \\
 &= \mathbb{P}(Y_1 \in A, D_0 = 1) - \mathbb{P}(Y_1 \in A, D_1 = 1) \\
 &= \mathbb{P}(Y_1 \in A, T = a) + \mathbb{P}(Y_1 \in A, T = df) - \mathbb{P}(Y_1 \in A, T = a) - \mathbb{P}(Y_1 \in A, T = c) \\
 &= \mathbb{P}(Y_1 \in A, T = df) - \mathbb{P}(Y_1 \in A, T = c) \\
 &\leq \mathbb{P}(Y_1 \in A, T = df) \leq \mathbb{P}(T = df),
 \end{aligned}$$



and

$$\begin{aligned}
& \mathbb{P}(Y \in A, D = 0 \mid Z = 1) - \mathbb{P}(Y \in A, D = 0 \mid Z = 0) \\
&= \mathbb{P}(Y_0 \in A, D_1 = 0) - \mathbb{P}(Y_0 \in A, D_0 = 0) \\
&= \mathbb{P}(Y_0 \in A, T = n) + \mathbb{P}(Y_0 \in A, T = df) - \mathbb{P}(Y_0 \in A, T = n) - \mathbb{P}(Y_0 \in A, T = c) \\
&= \mathbb{P}(Y_0 \in A, T = df) - \mathbb{P}(Y_0 \in A, T = c) \\
&\leq \mathbb{P}(Y_0 \in A, T = df) \leq \mathbb{P}(T = df).
\end{aligned}$$

Combined with the property probability that  $p_{df} \geq 0$ , we have

$$p_{df} \geq \max \left\{ \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s \mid Z = 1 - s) - \mathbb{P}(Y \in A, D = s \mid Z = s) \} \right\}, 0 \right\}.$$

In addition, according to the definitions and the property that  $p_a, p_n \geq 0$ ,

$$\mathbb{P}(D = 1 \mid Z = 0) = p_a + p_{df} \geq p_{df}, \quad \mathbb{P}(D = 0 \mid Z = 1) = p_n + p_{df} \geq p_{df}.$$

To summarize, an identified bound of  $p_{df}$  is given by

$$\Theta_I(p_{df}) = \begin{cases} \left[ \max \left\{ \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s \mid Z = 1 - s) - \mathbb{P}(Y \in A, D = s \mid Z = s) \} \right\}, 0 \right\}, \right. \\ \quad \left. \min \{ \mathbb{E}[D \mid Z = 0], \mathbb{E}[1 - D \mid Z = 1] \} \right], & \text{if inequality (12) holds} \\ \emptyset, & \text{otherwise.} \end{cases}$$

The next step is to prove the sharpness of the  $\Theta_I(p_{df})$ . We have to construct a joint distribution of  $(\tilde{Y}_{11}, \tilde{Y}_{10}, \tilde{Y}_{01}, \tilde{Y}_{00}, \tilde{T}, Z)$ ,  $\tilde{T} \in \{a, n, c, df\}$ ,  $Z \in \{0, 1\}$ , such that each point in  $\Theta_I(p_{df})$  can be achieved, and the constructed distribution is compatible with imposed assumptions and satisfies

$$\mathbb{P}(Y \in A, D = 1 \mid Z = 1) = \tilde{p}_c \mathbb{P}(\tilde{Y}_1 \in A \mid \tilde{T} = c) + \tilde{p}_a \mathbb{P}(\tilde{Y}_1 \in A \mid \tilde{T} = a), \quad (18)$$

$$\mathbb{P}(Y \in A, D = 1 \mid Z = 0) = \tilde{p}_{df} \mathbb{P}(\tilde{Y}_1 \in A \mid \tilde{T} = df) + \tilde{p}_a \mathbb{P}(\tilde{Y}_1 \in A \mid \tilde{T} = a), \quad (19)$$

$$\mathbb{P}(Y \in A, D = 0 \mid Z = 1) = \tilde{p}_{df} \mathbb{P}(\tilde{Y}_0 \in A \mid \tilde{T} = df) + \tilde{p}_n \mathbb{P}(\tilde{Y}_0 \in A \mid \tilde{T} = n), \quad (20)$$

$$\mathbb{P}(Y \in A, D = 0 \mid Z = 0) = \tilde{p}_c \mathbb{P}(\tilde{Y}_0 \in A \mid \tilde{T} = c) + \tilde{p}_n \mathbb{P}(\tilde{Y}_0 \in A \mid \tilde{T} = n). \quad (21)$$

Define  $\tilde{Y}_{dz} = \tilde{Y}_d$  for all  $d, z$ ,  $\tilde{p}_t \equiv \mathbb{P}(\tilde{T} = t) = \mathbb{P}(\tilde{T} = t \mid Z = z)$ ,  $\mathbb{P}(\tilde{Y}_d \leq y_d \mid \tilde{T} = t, Z = z) = \mathbb{P}(\tilde{Y}_d \leq y_d \mid \tilde{T} = t)$ , and  $\mathbb{P}(\tilde{Y}_1 \leq y_1, \tilde{Y}_0 \leq y_0, \tilde{T} = t \mid Z = z) = \tilde{p}_t \mathbb{P}(\tilde{Y}_1 \leq y_1 \mid \tilde{T} = t) \mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = t)$ .

For each  $\tilde{p}_{df} \in \Theta_I(p_{df})$ , we need to find  $\mathbb{P}(\tilde{Y}_d \leq y_d \mid \tilde{T} = t)$  such that Equations (18)-(21) hold. We will do this case by case.

**case 1:**  $\max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s | Z = 1 - s) - \mathbb{P}(Y \in A, D = s | Z = s) \} \right\} \leq 0$

In this case,  $\Theta_I(p_{df}) = [0, \min\{\mathbb{E}[D | Z = 0], \mathbb{E}[1 - D | Z = 1]\}]$ . Note that inequality (12) holds in this case, as  $\sup_z f_{Y,D|Z}(y, d | z) = f_{Y,D|Z}(y, d | d)$ , and  $\int_{\mathcal{Y}} f_{Y,D|Z}(y, d | d) d\mu(y) = \mathbb{P}(D = d | Z = d) \leq 1$  for all  $d$ .

1.  $\tilde{p}_{df} = 0$ .

We have  $\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0] = \mathbb{P}(Y \in \mathcal{Y}, D = 1 | Z = 1) - \mathbb{P}(Y \in \mathcal{Y}, D = 1 | Z = 0) \geq 0$ .

First, suppose  $\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0] > 0$ , and define

$$\begin{aligned} \tilde{p}_a &= \mathbb{E}[D | Z = 0], \quad \tilde{p}_n = \mathbb{E}[1 - D | Z = 1], \quad \tilde{p}_c = \mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0], \\ \mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = a) &= \mathbb{P}(Y \leq y_1 | D = 1, Z = 0), \\ \mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = n) &= \mathbb{P}(Y \leq y_0 | D = 0, Z = 1), \\ \mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = c) &= \frac{\mathbb{P}(Y \leq y_1, D = 1 | Z = 1) - \mathbb{P}(Y \leq y_1, D = 1 | Z = 0)}{\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0]}, \\ \mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = c) &= \frac{\mathbb{P}(Y \leq y_0, D = 0 | Z = 0) - \mathbb{P}(Y \leq y_0, D = 0 | Z = 1)}{\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0]}, \\ \mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = a) &= \mathbb{P}(Y \leq y_0 | D = 1, Z = 0), \\ \mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = n) &= \mathbb{P}(Y \leq y_1 | D = 0, Z = 1) \end{aligned}$$

Since  $\mathbb{P}(Y \in A, D = d | Z = d) - \mathbb{P}(Y \in A, D = d | Z = 1 - d) \geq 0$  for all Borel set  $A$  and all  $d$ , we can easily verify that the proposed distributions  $\mathbb{P}(\tilde{Y}_d \leq y_d | \tilde{T} = t)$  are well-defined cdf. We can also verify that Equations (18)-(21) hold.

Now, suppose  $\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0] = 0$ . Then,

$$\begin{aligned} \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s | Z = 1 - s) - \mathbb{P}(Y \in A, D = s | Z = s) \} \right\} \\ \leq 0 = \mathbb{P}(Y \in \mathcal{Y}, D = 1 | Z = 0) - \mathbb{P}(Y \in \mathcal{Y}, D = 1 | Z = 1). \end{aligned}$$

Therefore,

$$\begin{aligned} \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s | Z = 1 - s) - \mathbb{P}(Y \in A, D = s | Z = s) \} \right\} \\ = \mathbb{P}(Y \in \mathcal{Y}, D = 1 | Z = 0) - \mathbb{P}(Y \in \mathcal{Y}, D = 1 | Z = 1) = 0, \end{aligned}$$

which implies  $\mathbb{P}(Y \in A, D = s | Z = 1 - s) = \mathbb{P}(Y \in A, D = s | Z = s) = \mathbb{P}(Y \in A, D = s)$  for all

Borel set  $A$  and  $s$ , i.e.,  $Z \perp\!\!\!\perp (Y, D)$ . Define

$$\begin{aligned}\tilde{p}_a &= \mathbb{E}[D], \tilde{p}_n = \mathbb{E}[1 - D], \tilde{p}_c = 0, \\ \mathbb{P}(\tilde{Y}_1 \leq y_1 \mid \tilde{T} = a) &= \mathbb{P}(Y \leq y_1 \mid D = 1), \\ \mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = n) &= \mathbb{P}(Y \leq y_0 \mid D = 0), \\ \mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = a) &= \mathbb{P}(Y \leq y_0 \mid D = 1), \\ \mathbb{P}(\tilde{Y}_1 \leq y_1 \mid \tilde{T} = n) &= \mathbb{P}(Y \leq y_1 \mid D = 0).\end{aligned}$$

We only have two types  $a$  and  $n$ . It is straightforward that Equations (18)-(21) hold.

2.  $\tilde{p}_{df} = \min\{\mathbb{E}[D \mid Z = 0], \mathbb{E}[1 - D \mid Z = 1]\}$ .

First, suppose  $\tilde{p}_{df} = \mathbb{E}[D \mid Z = 0]$ . Then  $\mathbb{E}[D \mid Z = 0] \leq \mathbb{E}[1 - D \mid Z = 1]$ , and  $\tilde{p}_a = 0$ . Define

$$\begin{aligned}\tilde{p}_a &= 0, \tilde{p}_n = \mathbb{E}[1 - D \mid Z = 1] - \mathbb{E}[D \mid Z = 0], \tilde{p}_c = \mathbb{E}[D \mid Z = 1], \\ \mathbb{P}(\tilde{Y}_1 \leq y_1 \mid \tilde{T} = c) &= \mathbb{P}(Y \leq y_1 \mid D = 1, Z = 1), \\ \mathbb{P}(\tilde{Y}_1 \leq y_1 \mid \tilde{T} = df) &= \mathbb{P}(Y \leq y_1 \mid D = 1, Z = 0), \\ \mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = n) &= \mathbb{P}(Y \leq y_0 \mid D = 0, Z = 1), \\ \mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = df) &= \mathbb{P}(Y \leq y_0 \mid D = 0, Z = 1), \\ \mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = c) &= \frac{\mathbb{P}(Y \leq y_0, D = 0 \mid Z = 0) - (\mathbb{E}[1 - D \mid Z = 1] - \mathbb{E}[D \mid Z = 0])\mathbb{P}(Y \leq y_0 \mid D = 0, Z = 1)}{\mathbb{E}[D \mid Z = 1]}, \\ \mathbb{P}(\tilde{Y}_1 \leq y_1 \mid \tilde{T} = n) &= \mathbb{P}(Y \leq y_1 \mid D = 0, Z = 1)\end{aligned}$$

It is easy to verify that Equations (18)-(21) hold. It is obvious that  $\mathbb{P}(\tilde{Y}_1 \leq y_1 \mid \tilde{T} = c)$ ,  $\mathbb{P}(\tilde{Y}_1 \leq y_1 \mid \tilde{T} = df)$ ,  $\mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = n)$ ,  $\mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = df)$ , and  $\mathbb{P}(\tilde{Y}_1 \leq y_1 \mid \tilde{T} = n)$  are well-defined cdfs. We are going to check  $\mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = c)$  is also a well-defined cdf. It is clear that  $\lim_{y_0 \rightarrow -\infty} \mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = c) = 0$ , and  $\lim_{y_0 \rightarrow \infty} \mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = c) = 1$ . Right-continuity of  $\mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = c)$  holds by definition. It remains to check that  $\mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = c)$  is nondecreasing. We have

$$\mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = c) = \frac{\mathbb{P}(Y \leq y_0, D = 0 \mid Z = 0) - \mathbb{P}(Y \leq y_0, D = 0 \mid Z = 1) + \mathbb{E}[D \mid Z = 0]\mathbb{P}(Y \leq y_0 \mid D = 0, Z = 1)}{\mathbb{E}[D \mid Z = 1]}.$$

Since  $\mathbb{P}(Y \in A, D = d \mid Z = d) - \mathbb{P}(Y \in A, D = d \mid Z = 1 - d) \geq 0$  for all Borel set  $A$  and all  $d$ , the function  $\mathbb{P}(Y \leq y_0, D = 0 \mid Z = 0) - \mathbb{P}(Y \leq y_0, D = 0 \mid Z = 1)$  is nondecreasing. The function  $\mathbb{E}[D \mid Z = 0]\mathbb{P}(Y \leq y_0 \mid D = 0, Z = 1)$  is also nondecreasing, because the cdf  $\mathbb{P}(Y \leq y_0 \mid D = 0, Z = 1)$  is. Because the summation of two nondecreasing functions is nondecreasing, we conclude that  $\mathbb{P}(\tilde{Y}_0 \leq y_0 \mid \tilde{T} = c)$  as defined above is nondecreasing.

Now, suppose  $\tilde{p}_{df} = \mathbb{E}[1 - D \mid Z = 1]$ . Then  $\mathbb{E}[D \mid Z = 0] \geq \mathbb{E}[1 - D \mid Z = 1]$ , and  $\tilde{p}_n = 0$ . The argument is similar to the one above if we replace  $D$  and  $Z$  by  $\tilde{D} = 1 - D$  and  $\tilde{Z} = 1 - Z$ ,

respectively.

3.  $\tilde{p}_{df} = \lambda \times 0 + (1 - \lambda) \min\{\mathbb{E}[D|Z = 0], \mathbb{E}[1 - D|Z = 1]\}$ , where  $\lambda \in (0, 1)$ . As we can see,  $\tilde{p}_{df}$  is defined as a mixture of the previous sub-cases 1 and 2. Hence, It suffices to define the joint distribution based on the corresponding mixture distributions in the previous two points 1 and 2.

**case 2:**  $\max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s|Z = 1 - s) - \mathbb{P}(Y \in A, D = s|Z = s) \} \right\} > 0$

In this case,

$$\Theta_I(p_{df}) = \begin{cases} \left[ \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s|Z = 1 - s) - \mathbb{P}(Y \in A, D = s|Z = s) \} \right\} , \right. \\ \quad \left. \min\{\mathbb{E}[D|Z = 0], \mathbb{E}[1 - D|Z = 1]\} \right], \text{ if inequality (12) holds} \\ \emptyset, \text{ otherwise.} \end{cases}$$

1.  $\tilde{p}_{df} = \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s|Z = 1 - s) - \mathbb{P}(Y \in A, D = s|Z = s) \} \right\}.$

Define

$$\begin{aligned}
\tilde{p}_{df} &= \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s | Z = 1 - s) - \mathbb{P}(Y \in A, D = s | Z = s) \} \right\}, \\
\tilde{p}_a &= \mathbb{E}[D | Z = 0] - \tilde{p}_{df}, \\
\tilde{p}_c &= \mathbb{E}[D | Z = 1] - \tilde{p}_a = \mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0] + \tilde{p}_{df}, \\
\tilde{p}_n &= \mathbb{E}[1 - D | Z = 1] - \tilde{p}_{df}, \\
\mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = a) &= \frac{\int_{-\infty}^{y_1} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y)}{\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y)}, \\
\mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = n) &= \frac{\int_{-\infty}^{y_0} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)}{\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)}, \\
\mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = c) &= \frac{\mathbb{P}(Y \leq y_1, D = 1 | Z = 1) - \tilde{p}_a \mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = a)}{\tilde{p}_c}, \\
\mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = df) &= \frac{\mathbb{P}(Y \leq y_1, D = 1 | Z = 0) - \tilde{p}_a \mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = a)}{\tilde{p}_{df}}, \\
\mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = c) &= \frac{\mathbb{P}(Y \leq y_0, D = 0 | Z = 0) - \tilde{p}_n \mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = n)}{\tilde{p}_c}, \\
\mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = df) &= \frac{\mathbb{P}(Y \leq y_0, D = 0 | Z = 1) - \tilde{p}_n \mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = n)}{\tilde{p}_{df}}, \\
\mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = a) &= \frac{\int_{-\infty}^{y_0} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y)}{\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y)}, \\
\mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = n) &= \frac{\int_{-\infty}^{y_1} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)}{\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)}.
\end{aligned}$$

We need to show that  $\mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = c)$ ,  $\mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = df)$ ,  $\mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = c)$ , and  $\mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = df)$  as defined above are well-defined cdfs. To do so, we are going to show that their density versions are nonnegative and integrate to 1. Let  $f_{\tilde{Y}_d|\tilde{T}}(y_d | t)$  denote the density of  $\tilde{Y}_d$  conditional on  $\tilde{T} = t$ , that is absolutely continuous with respect to the dominating measure

$\mu$ . We have

$$\begin{aligned}
f_{\tilde{Y}_1|\tilde{T}}(y_1 | a) &= \frac{\min\{f_{Y,D|Z}(y_1, 1|1), f_{Y,D|Z}(y_1, 1|0)\}}{\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y)}, \\
f_{\tilde{Y}_0|\tilde{T}}(y_0 | n) &= \frac{\min\{f_{Y,D|Z}(y_0, 0|1), f_{Y,D|Z}(y_0, 0|0)\}}{\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)}, \\
f_{\tilde{Y}_1|\tilde{T}}(y_1 | c) &= \frac{f_{Y,D|Z}(y_1, 1|1) - \tilde{p}_a f_{\tilde{Y}_1|\tilde{T}}(y_1 | a)}{\tilde{p}_c}, \\
f_{\tilde{Y}_1|\tilde{T}}(y_1 | df) &= \frac{f_{Y,D|Z}(y_1, 1|0) - \tilde{p}_a f_{\tilde{Y}_1|\tilde{T}}(y_1 | a)}{\tilde{p}_{df}}, \\
f_{\tilde{Y}_0|\tilde{T}}(y_0 | c) &= \frac{f_{Y,D|Z}(y_0, 0|0) - \tilde{p}_n f_{\tilde{Y}_0|\tilde{T}}(y_0 | n)}{\tilde{p}_c}, \\
f_{\tilde{Y}_0|\tilde{T}}(y_0 | df) &= \frac{f_{Y,D|Z}(y_0, 0|1) - \tilde{p}_n f_{\tilde{Y}_0|\tilde{T}}(y_0 | n)}{\tilde{p}_{df}}, \\
f_{\tilde{Y}_0|\tilde{T}}(y_0 | a) &= \frac{\min\{f_{Y,D|Z}(y_0, 1|1), f_{Y,D|Z}(y_0, 1|0)\}}{\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y)}, \\
f_{\tilde{Y}_1|\tilde{T}}(y_1 | n) &= \frac{\min\{f_{Y,D|Z}(y_1, 0|1), f_{Y,D|Z}(y_1, 0|0)\}}{\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)}.
\end{aligned}$$

Clearly,  $f_{\tilde{Y}_1|\tilde{T}}(y_1 | a)$ ,  $f_{\tilde{Y}_0|\tilde{T}}(y_0 | a)$ ,  $f_{\tilde{Y}_0|\tilde{T}}(y_0 | n)$ , and  $f_{\tilde{Y}_1|\tilde{T}}(y_1 | n)$  are nonnegative and integrate to 1. To show that  $f_{\tilde{Y}_1|\tilde{T}}(y_1 | c) \geq 0$ , it suffices to show that  $f_{\tilde{Y}_1|\tilde{T}}(y_1 | a) \leq \frac{1}{\tilde{p}_a} f_{Y,D|Z}(y_1, 1|1)$ .

We have  $\min\{f_{Y,D|Z}(y_1, 1|1), f_{Y,D|Z}(y_1, 1|0)\} \leq f_{Y,D|Z}(y_1, 1|1)$ . We just need to show that  $\tilde{p}_a \leq \int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y)$ . Let  $B \equiv \{y \in \mathcal{Y} : f_{Y,D|Z}(y, 1|1) \geq f_{Y,D|Z}(y, 1|0)\}$ , and  $\bar{B} \equiv \{y \in \mathcal{Y} : f_{Y,D|Z}(y, 1|1) < f_{Y,D|Z}(y, 1|0)\}$ . On the one hand, we have

$$\begin{aligned}
\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y) &= \int_{\mathcal{Y}} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y) \\
&= \int_{\bar{B}} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y) + \int_B \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y) \\
&= \int_{\bar{B}} f_{Y,D|Z}(y, 1|1) d\mu(y) + \int_B f_{Y,D|Z}(y, 1|0) d\mu(y), \\
&= \mathbb{P}(Y \in \bar{B}, D = 1 | Z = 1) + \mathbb{P}(Y \in B, D = 1 | Z = 0)
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\tilde{p}_a &= \mathbb{E}[D|Z=0] - \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D=s|Z=1-s) - \mathbb{P}(Y \in A, D=s|Z=s) \} \right\}, \\
&= \min_{s \in \{0,1\}} \left\{ \inf_A \{ \mathbb{E}[D|Z=0] - \mathbb{P}(Y \in A, D=s|Z=1-s) + \mathbb{P}(Y \in A, D=s|Z=s) \} \right\}, \\
&\leq \inf_A \{ \mathbb{E}[D|Z=0] - \mathbb{P}(Y \in A, D=1|Z=0) + \mathbb{P}(Y \in A, D=1|Z=1) \} \quad (\text{for } s=1), \\
&= \inf_A \{ \mathbb{P}(Y \in \bar{A}, D=1|Z=0) + \mathbb{P}(Y \in A, D=1|Z=1) \}, \quad \text{where } \bar{A} \text{ denotes the complement of } A, \\
&\leq \mathbb{P}(Y \in \bar{B}, D=1|Z=1) + \mathbb{P}(Y \in B, D=1|Z=0), \\
&= \int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 1|1), f_{Y,D|Z}(y, 1|0)\} d\mu(y).
\end{aligned}$$

Similarly, we have  $f_{\tilde{Y}_1|\tilde{T}}(y_1 | a) \leq \frac{1}{\tilde{p}_a} f_{Y,D|Z}(y_1, 1|0)$ . Therefore,  $f_{\tilde{Y}_1|\tilde{T}}(y_1|df) \geq 0$ . To show that  $f_{\tilde{Y}_0|\tilde{T}}(y_0|c)$  and  $f_{\tilde{Y}_0|\tilde{T}}(y_0|df)$  are nonnegative, we are going to follow a similar reasoning as above.

Let  $C \equiv \{y \in \mathcal{Y} : f_{Y,D|Z}(y, 0|0) \geq f_{Y,D|Z}(y, 0|1)\}$ , and  $\bar{C} \equiv \{y \in \mathcal{Y} : f_{Y,D|Z}(y, 0|0) < f_{Y,D|Z}(y, 0|1)\}$ .

We have

$$\begin{aligned}
&\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y) = \int_{\mathcal{Y}} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y), \\
&= \int_{\bar{C}} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y) + \int_C \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y), \\
&= \int_{\bar{C}} f_{Y,D|Z}(y, 0|0) d\mu(y) + \int_C f_{Y,D|Z}(y, 0|1) d\mu(y), \\
&= \mathbb{P}(Y \in \bar{C}, D=0|Z=0) + \mathbb{P}(Y \in C, D=0|Z=1).
\end{aligned}$$

We also have

$$\begin{aligned}
\tilde{p}_n &= \mathbb{E}[1-D|Z=1] - \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D=s|Z=1-s) - \mathbb{P}(Y \in A, D=s|Z=s) \} \right\}, \\
&= \min_{s \in \{0,1\}} \left\{ \inf_A \{ \mathbb{E}[1-D|Z=1] - \mathbb{P}(Y \in A, D=s|Z=1-s) + \mathbb{P}(Y \in A, D=s|Z=s) \} \right\}, \\
&\leq \inf_A \{ \mathbb{E}[1-D|Z=1] - \mathbb{P}(Y \in A, D=0|Z=1) + \mathbb{P}(Y \in A, D=0|Z=0) \} \quad (\text{for } s=0), \\
&= \inf_A \{ \mathbb{P}(Y \in \bar{A}, D=0|Z=1) + \mathbb{P}(Y \in A, D=0|Z=0) \}, \\
&\leq \mathbb{P}(Y \in \bar{C}, D=0|Z=0) + \mathbb{P}(Y \in C, D=0|Z=1), \\
&= \int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y).
\end{aligned}$$

$$\min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} \leq f_{Y,D|Z}(y, 0|0) \text{ and } \tilde{p}_n \leq \int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)$$

imply  $f_{\tilde{Y}_0|\tilde{T}}(y_0|n) \leq \frac{1}{\tilde{p}_n} f_{Y,D|Z}(y, 0|0)$ , therefore  $f_{\tilde{Y}_0|\tilde{T}}(y_0|c) \geq 0$ . A similarly argument proves that  $f_{\tilde{Y}_0|\tilde{T}}(y_0|df) \geq 0$ .

Finally, given the definition of the types' probabilities  $\tilde{p}_t$ , it is easy to check that all the defined above densities integrate to 1. We can also check that the above proposed distributions satisfy Equations (18)-(21).

To complete this step of the proof, we need to check that the DGP we provide requires that inequality (12) holds. We have

$$\begin{aligned}
f_{\tilde{Y}_1}(y_1) &= \sum_{t \in \{a, c, df, n\}} \tilde{p}_t f_{\tilde{Y}_1|\tilde{T}}(y_1|t), \\
&= \tilde{p}_a f_{\tilde{Y}_1|\tilde{T}}(y_1|a) + f_{Y,D|Z}(y_1, 1|1) - \tilde{p}_a f_{\tilde{Y}_1|\tilde{T}}(y_1|a) + f_{Y,D|Z}(y_1, 1|0) - \tilde{p}_a f_{\tilde{Y}_1|\tilde{T}}(y_1|a) + \tilde{p}_n f_{\tilde{Y}_1|\tilde{T}}(y_1|n), \\
&= f_{Y,D|Z}(y_1, 1|1) + (f_{Y,D|Z}(y_1, 1|0) - \tilde{p}_a f_{\tilde{Y}_1|\tilde{T}}(y_1|a)) + \tilde{p}_n f_{\tilde{Y}_1|\tilde{T}}(y_1|n) \geq f_{Y,D|Z}(y_1, 1|1), \\
&= (f_{Y,D|Z}(y_1, 1|1) - \tilde{p}_a f_{\tilde{Y}_1|\tilde{T}}(y_1|a)) + f_{Y,D|Z}(y_1, 1|0) + \tilde{p}_n f_{\tilde{Y}_1|\tilde{T}}(y_1|n) \geq f_{Y,D|Z}(y_1, 1|0), \\
&\geq \max \{f_{Y,D|Z}(y_1, 1|1), f_{Y,D|Z}(y_1, 1|0)\} = \max_{z \in \{0,1\}} \{f_{Y,D|Z}(y_1, 1|z)\}.
\end{aligned}$$

Therefore,  $1 = \int_{\mathbf{y}} f_{\tilde{Y}_1}(y_1) d\mu(y_1) \geq \int_{\mathbf{y}} \max_{z \in \{0,1\}} \{f_{Y,D|Z}(y_1, 1|z)\} d\mu(y_1)$ .

Similarly,  $1 = \int_{\mathbf{y}} f_{\tilde{Y}_0}(y_0) d\mu(y_0) \geq \int_{\mathbf{y}} \max_{z \in \{0,1\}} \{f_{Y,D|Z}(y_0, 0|z)\} d\mu(y_0)$ . These last two inequalities combined are equivalent to inequality (12).

2.  $\tilde{p}_{df} = \min\{\mathbb{E}[D|Z = 0], \mathbb{E}[1 - D|Z = 1]\}$ .

First, suppose  $\tilde{p}_{df} = \mathbb{E}[D | Z = 0]$ . Then  $\mathbb{E}[D|Z = 0] \leq \mathbb{E}[1 - D|Z = 1]$ , and  $\tilde{p}_a = 0$ . Define

$$\begin{aligned}
\tilde{p}_a &= 0, \quad \tilde{p}_n = \mathbb{E}[1 - D | Z = 1] - \mathbb{E}[D | Z = 0], \quad \tilde{p}_c = \mathbb{E}[D | Z = 1], \\
\mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = c) &= \mathbb{P}(Y \leq y_1 | D = 1, Z = 1), \\
\mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = df) &= \mathbb{P}(Y \leq y_1 | D = 1, Z = 0), \\
\mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = n) &= \frac{\int_{-\infty}^{y_0} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)}{\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)}, \\
\mathbb{P}(\tilde{Y}_0 \leq y_0, |\tilde{T} = c) &= \frac{\mathbb{P}(Y \leq y_0, D = 0 | Z = 0) - \tilde{p}_n \mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = n)}{\tilde{p}_c}, \\
\mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = df) &= \frac{\mathbb{P}(Y \leq y_0, D = 0 | Z = 1) - \tilde{p}_n \mathbb{P}(\tilde{Y}_0 \leq y_0 | \tilde{T} = n)}{\tilde{p}_{df}}, \\
\mathbb{P}(\tilde{Y}_1 \leq y_1 | \tilde{T} = n) &= \frac{\int_{-\infty}^{y_1} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)}{\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)}.
\end{aligned}$$

It is clear that all the above quantities are well-defined cdfs, except  $\mathbb{P}(\tilde{Y}_0 \leq y_0, |\tilde{T} = c)$ , and  $\mathbb{P}(\tilde{Y}_0 \leq y_0, |\tilde{T} = df)$ . As in the previous bullet point, we show that the density analogs are



nonnegative and integrate to 1. We have

$$\begin{aligned} f_{\tilde{Y}_0|\tilde{T}}(y_0|c) &= \frac{f_{Y,D|Z}(y_0, 0|0) - \tilde{p}_n f_{\tilde{Y}_0|\tilde{T}}(y_0 | n)}{\tilde{p}_c}, \\ f_{\tilde{Y}_0|\tilde{T}}(y_0|df) &= \frac{f_{Y,D|Z}(y_0, 0|1) - \tilde{p}_n f_{\tilde{Y}_0|\tilde{T}}(y_0 | n)}{\tilde{p}_{df}}. \end{aligned}$$

As before, it suffices to show  $\tilde{p}_n \leq \int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y)$ . We have shown above that

$$\int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y) = \mathbb{P}(Y \in \overline{C}, D = 0|Z = 0) + \mathbb{P}(Y \in C, D = 0|Z = 1).$$

Therefore,

$$\begin{aligned} \tilde{p}_n - \int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y) \\ &= \mathbb{E}[1 - D | Z = 1] - \mathbb{E}[D | Z = 0] - \mathbb{P}(Y \in \overline{C}, D = 0|Z = 0) - \mathbb{P}(Y \in C, D = 0|Z = 1), \\ &= (\mathbb{E}[1 - D | Z = 1] - \mathbb{P}(Y \in C, D = 0|Z = 1)) + (\mathbb{E}[1 - D | Z = 0] - \mathbb{P}(Y \in \overline{C}, D = 0|Z = 0)) - 1, \\ &= \mathbb{P}(Y \in \overline{C}, D = 0|Z = 1) + \mathbb{P}(Y \in C, D = 0|Z = 0) - 1. \end{aligned}$$

Inequality (12) implies  $\int_{\mathcal{Y}} \max_{z \in \{0,1\}} f_{Y,D|Z}(y, 0|z) d\mu(y) \leq 1$ . Then,

$$\begin{aligned} 1 &\geq \int_{\mathcal{Y}} \max_{z \in \{0,1\}} f_{Y,D|Z}(y, 0|z) d\mu(y) = \int_{\overline{C}} \max_{z \in \{0,1\}} f_{Y,D|Z}(y, 0|z) d\mu(y) + \int_C \max_{z \in \{0,1\}} f_{Y,D|Z}(y, 0|z) d\mu(y), \\ &\geq \int_{\overline{C}} f_{Y,D|Z}(y, 0|1) d\mu(y) + \int_C f_{Y,D|Z}(y, 0|0) d\mu(y), \\ &= \mathbb{P}(Y \in \overline{C}, D = 0|Z = 1) + \mathbb{P}(Y \in C, D = 0|Z = 0). \end{aligned}$$

Hence,  $\tilde{p}_n - \int_{-\infty}^{\infty} \min\{f_{Y,D|Z}(y, 0|1), f_{Y,D|Z}(y, 0|0)\} d\mu(y) \leq 0$ .

Finally, we can check that the above proposed distributions satisfy Equations (18)-(21).

Second, if  $\tilde{p}_{df} = \mathbb{E}[1 - D | Z = 1]$ , then  $\mathbb{E}[D|Z = 0] \geq \mathbb{E}[1 - D|Z = 1]$ , and  $\tilde{p}_n = 0$ . The argument is similar to the one above if we replace  $D$  and  $Z$  by  $\tilde{D} = 1 - D$  and  $\tilde{Z} = 1 - Z$ , respectively.

3.  $\tilde{p}_{df} = \lambda \max_{s \in \{0,1\}} \{\sup_A \{\mathbb{P}(Y \in A, D = s|Z = 1 - s) - \mathbb{P}(Y \in A, D = s|Z = s)\}\} + (1 - \lambda) \min\{\mathbb{E}[D|Z = 0], \mathbb{E}[1 - D|Z = 1]\}$ , where  $\lambda \in (0, 1)$ .

As we can see,  $\tilde{p}_{df}$  is defined as a mixture of the previous sub-cases 1 and 2. Hence, it suffices to define the joint distribution based on the corresponding mixture distributions in the previous two points 1 and 2.

Finally, to finish the proof of Proposition 1, we show that the following identified set

$$\Theta_I(p_{df}) = \left[ \max \left\{ \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s | Z = 1 - s) - \mathbb{P}(Y \in A, D = s | Z = s) \} \right\}, 0 \right\}, \min \{ \mathbb{E}[D | Z = 0], \mathbb{E}[1 - D | Z = 1] \} \right] \quad (22)$$

is empty if and only if Equation (12) does not hold.

When the instrument and the treatment are binary,  $Z \in \{0, 1\}$  and  $D \in \{0, 1\}$ , we can write Equation (12) as

$$\max_{d \in \{0,1\}} \int_{\mathcal{Y}} \max_{z \in \{0,1\}} f_{Y,D|Z}(y, d | z) d\mu(y) \leq 1. \quad (23)$$

Let  $B_1$  denote the Borel set on  $\mathcal{Y}$  where  $f_{Y,D|Z}(y, 1 | 0)$  is no smaller than  $f_{Y,D|Z}(y, 1 | 1)$ ,  $B_1 = \{y : f_{Y,D|Z}(y, 1 | 0) \geq f_{Y,D|Z}(y, 1 | 1)\}$ , and  $B_0$  denote the Borel set on  $\mathcal{Y}$  where  $f_{Y,D|Z}(y, 0 | 1)$  is no smaller than  $f_{Y,D|Z}(y, 0 | 0)$ ,  $B_0 = \{y : f_{Y,D|Z}(y, 0 | 1) \geq f_{Y,D|Z}(y, 0 | 0)\}$ . The notation  $B^c$  represents the complement set of  $B$ . Given  $d = 1$ ,

$$\begin{aligned} & \int_{\mathcal{Y}} \max_{z \in \{0,1\}} f_{Y,D|Z}(y, 1 | z) d\mu(y) \\ &= \int_{B_1} f_{Y,D|Z}(y, 1 | 0) d\mu(y) + \int_{B_1^c} f_{Y,D|Z}(y, 1 | 1) d\mu(y) \\ &= \mathbb{P}(Y \in B_1, D = 1 | Z = 0) - \mathbb{P}(Y \in B_1, D = 1 | Z = 1) + \mathbb{P}(D = 1 | Z = 1), \end{aligned}$$

and given  $d = 0$ ,

$$\begin{aligned} & \int_{\mathcal{Y}} \max_{z \in \{0,1\}} f_{Y,D|Z}(y, 0 | z) d\mu(y) \\ &= \int_{B_0} f_{Y,D|Z}(y, 0 | 1) d\mu(y) + \int_{B_0^c} f_{Y,D|Z}(y, 0 | 0) d\mu(y) \\ &= \mathbb{P}(Y \in B_0, D = 0 | Z = 1) - \mathbb{P}(Y \in B_0, D = 0 | Z = 0) + \mathbb{P}(D = 0 | Z = 0). \end{aligned}$$

Therefore, the violation of Equation (23) can be represented as

$$\begin{aligned} & \max \{ \mathbb{P}(Y \in B_1, D = 1 | Z = 0) - \mathbb{P}(Y \in B_1, D = 1 | Z = 1) + \mathbb{P}(D = 1 | Z = 1), \\ & \quad \mathbb{P}(Y \in B_0, D = 0 | Z = 1) - \mathbb{P}(Y \in B_0, D = 0 | Z = 0) + \mathbb{P}(D = 0 | Z = 0) \} > 1 \\ \iff & \max_{s \in \{0,1\}} \left\{ \mathbb{P}(Y \in B_s, D = s | Z = 1 - s) - \mathbb{P}(Y \in B_s, D = s | Z = s) - \mathbb{P}(D = 1 - s | Z = s) \right\} > 0. \end{aligned} \quad (24)$$

In Equation (22), since the empty set  $\emptyset$  is in the Borel set generated by  $Y$ ,

$$\begin{aligned} & \sup_A \{ \mathbb{P}(Y \in A, D = s \mid Z = 1 - s) - \mathbb{P}(Y \in A, D = s \mid Z = s) \} \\ & \geq \mathbb{P}(Y \in \emptyset, D = s \mid Z = 1 - s) - \mathbb{P}(Y \in \emptyset, D = s \mid Z = s) = 0, \end{aligned}$$

for any  $s \in \{0, 1\}$ . Then, the identified set  $\Theta_I(p_{df})$  in Equation (22) would be empty when

$$\begin{aligned} & \min\{\mathbb{E}[D \mid Z = 0], \mathbb{E}[1 - D \mid Z = 1]\} \\ & < \max_{s \in \{0, 1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s \mid Z = 1 - s) - \mathbb{P}(Y \in A, D = s \mid Z = s) \} \right\}. \end{aligned}$$

The above inequality is equivalent to

$$\begin{aligned} & \max \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = 1 \mid Z = 0) - \mathbb{P}(Y \in A, D = 1 \mid Z = 1) - \mathbb{P}(D = 0 \mid Z = 1) \}, \right. \\ & \quad \sup_A \{ \mathbb{P}(Y \in A, D = 1 \mid Z = 0) - \mathbb{P}(Y \in A, D = 1 \mid Z = 1) - \mathbb{P}(D = 1 \mid Z = 0) \}, \\ & \quad \sup_A \{ \mathbb{P}(Y \in A, D = 0 \mid Z = 1) - \mathbb{P}(Y \in A, D = 0 \mid Z = 0) - \mathbb{P}(D = 1 \mid Z = 0) \}, \\ & \quad \left. \sup_A \{ \mathbb{P}(Y \in A, D = 0 \mid Z = 1) - \mathbb{P}(Y \in A, D = 0 \mid Z = 0) - \mathbb{P}(D = 0 \mid Z = 1) \} \right\} > 0. \end{aligned} \tag{25}$$

If the inequality in Equation (24) holds, it implies that either

$$\sup_A \{ \mathbb{P}(Y \in A, D = 1 \mid Z = 0) - \mathbb{P}(Y \in A, D = 1 \mid Z = 1) - \mathbb{P}(D = 0 \mid Z = 1) \} > 0$$

or

$$\sup_A \{ \mathbb{P}(Y \in A, D = 0 \mid Z = 1) - \mathbb{P}(Y \in A, D = 0 \mid Z = 0) - \mathbb{P}(D = 1 \mid Z = 0) \} > 0,$$

and the inequality in Equation (25) holds. Therefore, the violation of Equation (23) implies that the identified set in Equation (22) is empty.

Suppose now that the inequality in (25) holds so that the identified set in Equation (22) is empty. Since

$$\begin{aligned} & \sup_A \{ \mathbb{P}(Y \in A, D = 1 \mid Z = 0) - \mathbb{P}(Y \in A, D = 1 \mid Z = 1) - \mathbb{P}(D = 1 \mid Z = 0) \} \\ & \leq \sup_A \{ \mathbb{P}(Y \in A, D = 1 \mid Z = 0) - \mathbb{P}(D = 1 \mid Z = 0) \} = 0, \\ & \sup_A \{ \mathbb{P}(Y \in A, D = 0 \mid Z = 1) - \mathbb{P}(Y \in A, D = 0 \mid Z = 0) - \mathbb{P}(D = 0 \mid Z = 1) \} \\ & \leq \sup_A \{ \mathbb{P}(Y \in A, D = 0 \mid Z = 1) - \mathbb{P}(D = 0 \mid Z = 1) \} = 0, \end{aligned}$$

it should be that either the first or the third expression in Equation (25) is greater than zero. Assume

that the maximum in Equation (25) is achieved by the first expression. Based on the definition of  $B_1$ , the supremum is achieved when  $A = B_1$ . Then, we have

$$\mathbb{P}(Y \in B_1, D = 1 \mid Z = 0) - \mathbb{P}(Y \in B_1, D = 1 \mid Z = 1) - \mathbb{P}(D = 0 \mid Z = 1) > 0,$$

which implies that the inequality in (24) holds. Following similar reasoning, if the third term in Equation (25) attains its maximum, this would also imply that the inequality in (24) holds. Therefore, an empty identified set for the defier's proportion in Equation (22) would imply the failure of the testable implication in Equation (23).

To conclude, the violation of testable implication in Equation (23) is equivalent to the identified set  $\Theta_I(p_{df})$  in Equation (22) being empty. Consequently, we can write the identified set for defier's probability as

$$\Theta_I(p_{df}) = \left[ \max_{s \in \{0,1\}} \left\{ \sup_A \{ \mathbb{P}(Y \in A, D = s \mid Z = 1 - s) - \mathbb{P}(Y \in A, D = s \mid Z = s) \} \right\}, \right. \\ \left. \min\{\mathbb{E}[D \mid Z = 0], \mathbb{E}[1 - D \mid Z = 1]\} \right].$$

□

## A.2 Proof of Proposition 2 and Theorem 1

*Proof.* For each  $p_{df} \in \Theta_I(p_{df})$ , we derive in bounds on  $\mu_{1a}(p_a)$ ,  $\mu_{1c}(p_a)$ ,  $\mu_{1df}(p_a)$ ,  $\mu_{0n}(p_n)$ ,  $\mu_{0c}(p_n)$ ,  $\mu_{0df}(p_n)$  where  $p_a = \mathbb{E}[D \mid Z = 0] - p_{df}$ , and  $p_n = \mathbb{E}[1 - D \mid Z = 1] - p_{df}$ . We build on results from Horowitz and Manski (1995).

First, suppose  $0 < p_{df} < \min\{\mathbb{E}[D \mid Z = 0], \mathbb{E}[1 - D \mid Z = 1]\}$ . Equation (3) implies

$$\mathbb{P}(Y_1 \in A \mid T = a) = \frac{\mathbb{P}(Y \in A, D = 1 \mid Z = 1) - p_c \mathbb{P}(Y_1 \in A \mid T = c)}{p_a}.$$

Since  $\mathbb{P}(Y_1 \in A \mid T = c) \in [0, 1]$ , and  $\mathbb{P}(Y_1 \in A \mid T = a) \in [0, 1]$ , we have

$$\max \left\{ \frac{\mathbb{P}(Y \in A, D = 1 \mid Z = 1) - p_c}{p_a}, 0 \right\} \leq \mathbb{P}(Y_1 \in A \mid T = a) \leq \min \left\{ \frac{\mathbb{P}(Y \in A, D = 1 \mid Z = 1)}{p_a}, 1 \right\},$$

which we can equivalently rewrite as

$$\begin{aligned} & \frac{(\mathbb{P}(Y \in A, D = 1|Z = 1) - p_c) \mathbb{1}\{\mathbb{P}(Y \in A, D = 1|Z = 1) - p_c \geq 0\}}{p_a} \\ & \leq \mathbb{P}(Y_1 \in A|T = a) \leq \\ & \frac{\mathbb{P}(Y \in A, D = 1|Z = 1)}{p_a} \mathbb{1}\left\{\frac{\mathbb{P}(Y \in A, D = 1|Z = 1)}{p_a} \leq 1\right\} + \mathbb{1}\left\{\frac{\mathbb{P}(Y \in A, D = 1|Z = 1)}{p_a} > 1\right\}. \end{aligned}$$

Similarly, Equation (4) implies

$$\max\left\{\frac{\mathbb{P}(Y \in A, D = 1|Z = 0) - p_{df}}{p_a}, 0\right\} \leq \mathbb{P}(Y_1 \in A|T = a) \leq \min\left\{\frac{\mathbb{P}(Y \in A, D = 1|Z = 0)}{p_a}, 1\right\},$$

or equivalently

$$\begin{aligned} & \frac{(\mathbb{P}(Y \in A, D = 1|Z = 0) - p_{df}) \mathbb{1}\{\mathbb{P}(Y \in A, D = 1|Z = 0) - p_{df} \geq 0\}}{p_a} \\ & \leq \mathbb{P}(Y_1 \in A|T = a) \leq \\ & \frac{\mathbb{P}(Y \in A, D = 1|Z = 0)}{p_a} \mathbb{1}\left\{\frac{\mathbb{P}(Y \in A, D = 1|Z = 0)}{p_a} \leq 1\right\} + \mathbb{1}\left\{\frac{\mathbb{P}(Y \in A, D = 1|Z = 0)}{p_a} > 1\right\}. \end{aligned}$$

Hence, Equations (3)-(4) together imply

$$\begin{aligned} & \max\left\{\frac{\mathbb{P}(Y \in A, D = 1|Z = 1) - p_c}{p_a}, \frac{\mathbb{P}(Y \in A, D = 1|Z = 0) - p_{df}}{p_a}, 0\right\} \\ & \leq \mathbb{P}(Y_1 \in A|T = a) \leq \\ & \min\left\{\frac{\mathbb{P}(Y \in A, D = 1|Z = 1)}{p_a}, \frac{\mathbb{P}(Y \in A, D = 1|Z = 0)}{p_a}, 1\right\}. \end{aligned}$$

Therefore, the identified set for the distribution  $F_{1a} \equiv F_{Y_1|T=a}$  is

$$\begin{aligned}
\Theta_I(F_{1a}) = & \left\{ F_{1a} \in \mathcal{F} : \max \left\{ \frac{(\mathbb{P}(Y \in (y, y'], D = 1|Z = 1) - p_c)\mathbb{1}\{\mathbb{P}(Y \in (y, y'], D = 1|Z = 1) - p_c \geq 0\}}{p_a}, \right. \right. \\
& \left. \frac{(\mathbb{P}(Y \in (y, y'], D = 1|Z = 0) - p_{df})\mathbb{1}\{\mathbb{P}(Y \in (y, y'], D = 1|Z = 0) - p_{df} \geq 0\}}{p_a} \right\} \\
& \leq F_{1a}(y') - F_{1a}(y) \leq \\
& \min \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 1)}{p_a} \mathbb{1} \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 1)}{p_a} \leq 1 \right\} \right. \\
& \quad \left. + \mathbb{1} \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 1)}{p_a} > 1 \right\}, \right. \\
& \quad \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 0)}{p_a} \mathbb{1} \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 0)}{p_a} \leq 1 \right\} \\
& \quad \left. + \mathbb{1} \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 0)}{p_a} > 1 \right\} \right\} \text{ for all } y < y',
\end{aligned} \tag{26}$$

where  $\mathcal{F}$  is the set of all distributions.

Combining Equations (5)-(6), we obtain the following identified set for  $F_{0n} \equiv F_{Y_0|T=n}$ :

$$\begin{aligned}
\Theta_I(F_{0n}) = & \left\{ F_{0n} \in \mathcal{F} : \max \left\{ \frac{(\mathbb{P}(Y \in (y, y'], D = 0|Z = 0) - p_c)\mathbb{1}\{\mathbb{P}(Y \in (y, y'], D = 0|Z = 0) - p_c \geq 0\}}{p_n}, \right. \right. \\
& \left. \frac{(\mathbb{P}(Y \in (y, y'], D = 0|Z = 1) - p_{df})\mathbb{1}\{\mathbb{P}(Y \in (y, y'], D = 0|Z = 1) - p_{df} \geq 0\}}{p_n} \right\} \\
& \leq F_{0n}(y') - F_{0n}(y) \leq \\
& \min \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 0)}{p_n} \mathbb{1} \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 0)}{p_n} \leq 1 \right\} \right. \\
& \quad \left. + \mathbb{1} \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 0)}{p_n} > 1 \right\}, \right. \\
& \quad \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 1)}{p_n} \mathbb{1} \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 1)}{p_n} \leq 1 \right\} \\
& \quad \left. + \mathbb{1} \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 1)}{p_n} > 1 \right\} \right\} \text{ for all } y < y'.
\end{aligned} \tag{27}$$

Note that the above bounds are *functionally sharp*, as discussed in [Mourifié, Henry, and Méango \(2020\)](#). The pointwise bounds are given below:

$$F_{1a}^{LB}(y) \equiv \max \left\{ F_{1a}^{LB_1}(y), F_{1a}^{LB_0}(y) \right\} \leq F_{1a}(y) \leq \min \left\{ F_{1a}^{UB_1}(y), F_{1a}^{UB_0}(y) \right\} \equiv F_{1a}^{UB}(y), \tag{28}$$

where

$$\begin{aligned}
F_{1a}^{LB_1}(y) &\equiv \frac{(\mathbb{P}(Y \leq y, D = 1|Z = 1) - p_c)\mathbb{1}\{\mathbb{P}(Y \leq y, D = 1|Z = 1) - p_c \geq 0\}}{p_a}, \\
F_{1a}^{LB_0}(y) &\equiv \frac{(\mathbb{P}(Y \leq y, D = 1|Z = 0) - p_{df})\mathbb{1}\{\mathbb{P}(Y \leq y, D = 1|Z = 0) - p_{df} \geq 0\}}{p_a}, \\
F_{1a}^{UB_1}(y) &\equiv \frac{\mathbb{P}(Y \leq y, D = 1|Z = 1)}{p_a} \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 1|Z = 1)}{p_a} \leq 1\right\} + \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 1|Z = 1)}{p_a} > 1\right\}, \\
F_{1a}^{UB_0}(y) &\equiv \frac{\mathbb{P}(Y \leq y, D = 1|Z = 0)}{p_a} \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 1|Z = 0)}{p_a} \leq 1\right\} + \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 1|Z = 0)}{p_a} > 1\right\}.
\end{aligned}$$

$$F_{0n}^{LB}(y) \equiv \max \left\{ F_{0n}^{LB_1}(y), F_{0n}^{LB_0}(y) \right\} \leq F_{0n}(y) \leq \min \left\{ F_{0n}^{UB_1}(y), F_{0n}^{UB_0}(y) \right\} \equiv F_{0n}^{UB}(y), (29)$$

where

$$\begin{aligned}
F_{0n}^{LB_0}(y) &\equiv \frac{(\mathbb{P}(Y \leq y, D = 0|Z = 0) - p_c)\mathbb{1}\{\mathbb{P}(Y \leq y, D = 0|Z = 0) - p_c \geq 0\}}{p_n}, \\
F_{0n}^{LB_1}(y) &\equiv \frac{(\mathbb{P}(Y \leq y, D = 0|Z = 1) - p_{df})\mathbb{1}\{\mathbb{P}(Y \leq y, D = 0|Z = 1) - p_{df} \geq 0\}}{p_n}, \\
F_{0n}^{UB_0}(y) &\equiv \frac{\mathbb{P}(Y \leq y, D = 0|Z = 0)}{p_n} \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 0|Z = 0)}{p_n} \leq 1\right\} + \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 0|Z = 0)}{p_n} > 1\right\}, \\
F_{0n}^{UB_1}(y) &\equiv \frac{\mathbb{P}(Y \leq y, D = 0|Z = 1)}{p_n} \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 0|Z = 1)}{p_n} \leq 1\right\} + \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 0|Z = 1)}{p_n} > 1\right\}.
\end{aligned}$$

The identified sets for  $F_{dt} \equiv F_{Y_d|T=t}$ ,  $d \in \{0, 1\}$ ,  $t \in \{c, df\}$  are given by

$$\begin{aligned}
F_{1c}(y) &= \frac{\mathbb{P}(Y \leq y, D = 1|Z = 1) - p_a F_{1a}(y)}{p_c}, \\
F_{1df}(y) &= \frac{\mathbb{P}(Y \leq y, D = 1|Z = 0) - p_a F_{1a}(y)}{p_{df}}, \\
F_{0c}(y) &= \frac{\mathbb{P}(Y \leq y, D = 0|Z = 0) - p_n F_{0n}(y)}{p_c}, \\
F_{0df}(y) &= \frac{\mathbb{P}(Y \leq y, D = 0|Z = 1) - p_n F_{0n}(y)}{p_{df}},
\end{aligned}$$

where  $F_{1a} \in \Theta_I(F_{1a})$ , and  $F_{0n} \in \Theta_I(F_{0n})$ .

Let  $\mu_F$  denote the expected value of a given cdf  $F$ . Then sharp bounds for  $\mu_{1a}$  and  $\mu_{0n}$

are as follows:

$$\begin{aligned}\mu_{F_{1a}}^{UB} &\leq \mu_{1a} \leq \mu_{F_{1a}}^{LB} \\ \mu_{F_{0n}}^{UB} &\leq \mu_{0n} \leq \mu_{F_{0n}}^{LB}.\end{aligned}$$

Once obtaining the sharp bounds on  $\mu_{1a}, \mu_{0n}$ , we can characterize the sharp bounds for  $\mu_{dt}, d \in \{0, 1\}, t \in \{c, df\}$  as

$$\begin{aligned}\mu_{1c} &= \frac{\mathbb{E}[YD|Z=1] - p_a \mu_{1a}}{\mathbb{E}[D|Z=1] - p_a}, \quad \mu_{1df} = \frac{\mathbb{E}[YD|Z=0] - p_a \mu_{1a}}{\mathbb{E}[D|Z=0] - p_a}, \\ \mu_{0c} &= \frac{\mathbb{E}[Y(1-D)|Z=0] - p_n \mu_{0n}}{\mathbb{E}[1-D|Z=0] - p_n}, \quad \mu_{0df} = \frac{\mathbb{E}[Y(1-D)|Z=1] - p_n \mu_{0n}}{\mathbb{E}[1-D|Z=1] - p_n}.\end{aligned}$$

Finally, sharp bounds for parameters  $\theta_{0c}, \theta_{1c}, \theta_{0df}, \theta_{1df}$  are given by

$$\theta_{0c} = \theta_{1c} = \mu_{1c} - \mu_{0c}, \quad \theta_{0df} = \theta_{1df}(p_{df}) = \mu_{1df} - \mu_{0df}.$$

Therefore, when  $p_{df} \in \mathring{\Theta}_I(p_{df})$ , we can characterize the identified set for  $\Gamma$  with  $\Theta_I^1(A_2)$ .

These sharp bounds can be difficult to compute in practice. Valid outer sets can be computed:

$$\begin{aligned}\max \left\{ \mu_{F_{1a}}^{UB_1}, \mu_{F_{1a}}^{UB_0} \right\} &\leq \mu_{1a} \leq \min \left\{ \mu_{F_{1a}}^{LB_1}, \mu_{F_{1a}}^{LB_0} \right\} \\ \max \left\{ \mu_{F_{0n}}^{UB_1}, \mu_{F_{0n}}^{UB_0} \right\} &\leq \mu_{0n} \leq \min \left\{ \mu_{F_{0n}}^{LB_1}, \mu_{F_{0n}}^{LB_0} \right\}.\end{aligned}$$

*Remark 5.*  $\max \left\{ \mu_{F_{1a}}^{UB_1}, \mu_{F_{1a}}^{UB_0} \right\} = \mu_{F_{1a}}^{UB}$  if  $F_{1a}^{UB_1}$  first-order stochastically dominates  $F_{1a}^{UB_0}$  or vice versa. Similarly,  $\min \left\{ \mu_{F_{1a}}^{LB_1}, \mu_{F_{1a}}^{LB_0} \right\} = \mu_{F_{1a}}^{LB}$  if  $F_{1a}^{LB_1}$  first-order stochastically dominates  $F_{1a}^{LB_0}$  or vice versa. However, in general,  $\max \left\{ \mu_{F_{1a}}^{UB_1}, \mu_{F_{1a}}^{UB_0} \right\} \leq \mu_{F_{1a}}^{UB}$ , and  $\min \left\{ \mu_{F_{1a}}^{LB_1}, \mu_{F_{1a}}^{LB_0} \right\} \geq \mu_{F_{1a}}^{LB}$ .

### A.2.1 Proof of Lemma 1

*Proof.* To compute the expected value of  $F_{1a}^{LB_z}$ , we only need to identify the support and the density of a random variable with such cdf. Take  $\mu_{F_{1a}}^{LB_1}$  as the example. Since the outcome is continuous with a strictly increasing cdf, the support is characterized by  $\{y : 0 < F_{1a}^{LB_1}(y) < 1\}$ , which is this case equal to  $\{y : \mathbb{P}(Y \leq y, D=1|Z=1) - p_c > 0\} = \left\{y : \mathbb{P}(Y \leq y|D=1, Z=1) > \frac{p_c}{\mathbb{E}[D|Z=1]}\right\} = \left\{y : y > F_{Y|D=1, Z=1}^{-1}\left(\frac{p_c}{\mathbb{E}[D|Z=1]}\right)\right\}$ . Then for all  $y >$



$F_{Y|D=1,Z=1}^{-1}\left(\frac{p_c}{\mathbb{E}[D|Z=1]}\right)$ ,  $F_{1a}^{LB_1}(y) = \frac{\mathbb{P}(Y \leq y, D=1|Z=1) - p_c}{p_a}$  and its density  $f_{1a}^{LB_1}(y) = \partial F_{1a}^{LB_1}(y)/\partial y = \frac{f_{Y,D|Z}(y,1|1)}{p_a}$ . Therefore,

$$\begin{aligned}
\mu_{F_{1a}^{LB_1}} &= \int_{y > F_{Y|D=1,Z=1}^{-1}\left(\frac{p_c}{\mathbb{E}[D|Z=1]}\right)} y \frac{f_{Y,D|Z}(y,1|1)}{p_a} dy, \\
&= \int_{y > F_{Y|D=1,Z=1}^{-1}\left(1 - \frac{p_a}{\mathbb{E}[D|Z=1]}\right)} y \frac{f_{Y,D|Z}(y,1|1)}{p_a} dy \quad \text{since } p_c = \mathbb{E}[D|Z=1] - p_a, \\
&= \int_{y > F_{Y|D=1,Z=1}^{-1}\left(1 - \frac{p_a}{\mathbb{E}[D|Z=1]}\right)} y \frac{f_{Y|D,Z}(y|1,1)}{p_a/\mathbb{E}[D|Z=1]} dy \quad \text{from Bayes' rule} \\
&= \int_{y > F_{Y|D=1,Z=1}^{-1}\left(1 - \frac{p_a}{\mathbb{E}[D|Z=1]}\right)} y \frac{f_{Y|D,Z}(y|1,1)}{\mathbb{P}\left(Y > F_{Y|D=1,Z=1}^{-1}\left(1 - \frac{p_a}{\mathbb{E}[D|Z=1]}\right) | D=1, Z=1\right)} dy, \\
&\quad \text{because } \mathbb{P}\left(Y > F_{Y|D=1,Z=1}^{-1}\left(1 - \frac{p_a}{\mathbb{E}[D|Z=1]}\right) | D=1, Z=1\right) = p_a/\mathbb{E}[D|Z=1], \\
&= \mathbb{E}\left[Y | D=1, Z=1, Y > F_{Y|D=1,Z=1}^{-1}\left(1 - \frac{p_a}{\mathbb{E}[D|Z=1]}\right)\right].
\end{aligned}$$

A similar argument for  $\mu_{F_{1a}^{LB_0}}, \mu_{F_{0n}^{LB_z}}$  yields the formula in the lemma.

For  $\mu_{F_{1a}^{UB_z}}$ , the support is  $\{y : \frac{\mathbb{P}(Y \leq y, D=1|Z=z)}{p_a} < 1\} = \{y : \mathbb{P}(Y \leq y | D=1, Z=z) < \frac{p_a}{\mathbb{E}[D|Z=z]}\} = \left\{y : y < F_{Y|D=1,Z=z}^{-1}\left(\frac{p_a}{\mathbb{E}[D|Z=z]}\right)\right\}$ . Also, the density of  $F_{1a}^{UB_z}$  is  $f_{1a}^{UB_z}(y) = \partial F_{1a}^{UB_z}(y)/\partial y = \frac{f_{Y,D|Z}(y,1|z)}{p_a}$ . Therefore,

$$\begin{aligned}
\mu_{F_{1a}^{UB_z}} &= \int_{y < F_{Y|D=1,Z=z}^{-1}\left(\frac{p_a}{\mathbb{E}[D|Z=z]}\right)} y \frac{f_{Y,D|Z}(y,1|z)}{p_a} dy, \\
&= \int_{y < F_{Y|D=1,Z=z}^{-1}\left(\frac{p_a}{\mathbb{E}[D|Z=z]}\right)} y \frac{f_{Y,D|Z}(y,1|z)}{p_a} dy, \\
&= \int_{y < F_{Y|D=1,Z=z}^{-1}\left(\frac{p_a}{\mathbb{E}[D|Z=z]}\right)} y \frac{f_{Y|D,Z}(y|1,z)}{p_a/\mathbb{E}[D|Z=z]} dy \quad \text{from Bayes' rule} \\
&= \int_{y < F_{Y|D=1,Z=z}^{-1}\left(\frac{p_a}{\mathbb{E}[D|Z=z]}\right)} y \frac{f_{Y|D,Z}(y|1,z)}{\mathbb{P}\left(Y < F_{Y|D=1,Z=z}^{-1}\left(\frac{p_a}{\mathbb{E}[D|Z=z]}\right) | D=1, Z=z\right)} dy, \\
&\quad \text{because } \mathbb{P}\left(Y < F_{Y|D=1,Z=z}^{-1}\left(\frac{p_a}{\mathbb{E}[D|Z=z]}\right) | D=1, Z=z\right) = p_a/\mathbb{E}[D|Z=z], \\
&= \mathbb{E}\left[Y | D=1, Z=z, Y < F_{Y|D=1,Z=z}^{-1}\left(\frac{p_a}{\mathbb{E}[D|Z=z]}\right)\right].
\end{aligned}$$

A similar argument for  $\mu_{F_{0n}^{UB_Z}}$  yields the formula in the lemma.

□

Second, when  $p_{df} = 0$ , we have the bounds in  $\Theta_I(A_1)$ .

Third, suppose  $p_{df} = \min\{\mathbb{E}[D|Z = 0], \mathbb{E}[1 - D|Z = 1]\}$ . We distinguish three cases.

1.  $p_{df} = \mathbb{E}[D|Z = 0] < \mathbb{E}[1 - D|Z = 1]$

In this case,  $p_a = \mathbb{E}[D|Z = 0] - p_{df} = 0$ ,  $p_c = \mathbb{E}[D|Z = 1] - p_a = \mathbb{E}[D|Z = 1]$ , and  $p_n = \mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]$ . Equations (3)-(4) imply

$$\begin{aligned}\mathbb{P}(Y_1 \leq y|T = c) &= \mathbb{P}(Y \leq y|D = 1, Z = 1), \\ \mathbb{P}(Y_1 \leq y|T = df) &= \mathbb{P}(Y \leq y|D = 1, Z = 0),\end{aligned}$$

Following a similar reasoning as before and combining Equations (5)-(6), we obtain the following identified set for  $F_{0n}$ :

$$\begin{aligned}\Theta_I(F_{0n}) &= \left\{ F_{0n} \in \mathcal{F} : \right. \\ &\max \left\{ \frac{(\mathbb{P}(Y \in (y, y'], D = 0|Z = 0) - \mathbb{E}[D|Z = 1])}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} \cdot \right. \\ &\quad \mathbb{1}\{\mathbb{P}(Y \in (y, y'], D = 0|Z = 0) - \mathbb{E}[D|Z = 1] \geq 0\}, \\ &\quad \frac{(\mathbb{P}(Y \in (y, y'], D = 0|Z = 1) - \mathbb{E}[D|Z = 0])}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} \cdot \\ &\quad \left. \mathbb{1}\{\mathbb{P}(Y \in (y, y'], D = 0|Z = 1) - \mathbb{E}[D|Z = 0] \geq 0\} \right\} \\ &\leq F_{0n}(y') - F_{0n}(y) \leq \\ &\min \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 0)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} \mathbb{1}\left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 0)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} \leq 1 \right\} \right. \\ &\quad \left. + \mathbb{1}\left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 0)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} > 1 \right\}, \right. \\ &\quad \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 1)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} \mathbb{1}\left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 1)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} \leq 1 \right\} \\ &\quad \left. + \mathbb{1}\left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 0|Z = 1)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} > 1 \right\} \right\} \text{ for all } y < y'. \end{aligned} \tag{30}$$

$$F_{0n}^{LB}(y) \equiv \max \left\{ F_{0n}^{LB_1}(y), F_{0n}^{LB_0}(y) \right\} \leq F_{0n}(y) \leq \min \left\{ F_{0n}^{UB_1}(y), F_{0n}^{UB_0}(y) \right\} \equiv F_{0n}^{UB}(y),$$

where

$$\begin{aligned}
F_{0n}^{LB_0}(y) &\equiv \frac{(\mathbb{P}(Y \leq y, D = 0|Z = 0) - \mathbb{E}[D|Z = 1])\mathbb{1}\{\mathbb{P}(Y \leq y, D = 0|Z = 0) - \mathbb{E}[D|Z = 1] \geq 0\}}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]}, \\
F_{0n}^{LB_1}(y) &\equiv \frac{(\mathbb{P}(Y \leq y, D = 0|Z = 1) - \mathbb{E}[D|Z = 0])\mathbb{1}\{\mathbb{P}(Y \leq y, D = 0|Z = 1) - \mathbb{E}[D|Z = 0] \geq 0\}}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]}, \\
F_{0n}^{UB_0}(y) &\equiv \frac{\mathbb{P}(Y \leq y, D = 0|Z = 0)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 0|Z = 0)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} \leq 1\right\} \\
&\quad + \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 0|Z = 0)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} > 1\right\}, \\
F_{0n}^{UB_1}(y) &\equiv \frac{\mathbb{P}(Y \leq y, D = 0|Z = 1)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 0|Z = 1)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} \leq 1\right\} \\
&\quad + \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 0|Z = 1)}{\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0]} > 1\right\}.
\end{aligned}$$

The identified sets for  $F_{dt}$ ,  $d \in \{0, 1\}$ ,  $t \in \{c, df\}$  are given by

$$\begin{aligned}
F_{0c}(y) &= \frac{\mathbb{P}(Y \leq y, D = 0|Z = 0) - (\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0])F_{0n}(y)}{\mathbb{E}[D|Z = 1]}, \\
F_{0df}(y) &= \frac{\mathbb{P}(Y \leq y, D = 0|Z = 1) - (\mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0])F_{0n}(y)}{\mathbb{E}[D|Z = 0]},
\end{aligned}$$

where  $F_{0n} \in \Theta_I(F_{0n})$ .

$$\begin{aligned}
\Theta_I^{3,1}(A_2) &= \left\{ \Gamma \in \mathbb{R}^{20} : \theta_{0a}(p_a) = \theta_{1a}(p_a) = \mu_{1a} - \mu_{0a}, \theta_{0n}(p_n) = \theta_{1n}(p_n) = \mu_{1n} - \mu_{0n}(p_n), \right. \\
&\quad \theta_{0c}(p_c) = \theta_{1c}(p_c) = \mathbb{E}[Y|D = 1, Z = 1] - \frac{\mathbb{E}[Y(1 - D)|Z = 0] - p_n \mu_{0n}(p_n)}{\mathbb{E}[1 - D|Z = 0] - p_n}, \\
&\quad \theta_{0df}(p_{df}) = \theta_{1df}(p_{df}) = \mathbb{E}[Y|D = 1, Z = 0] - \frac{\mathbb{E}[Y(1 - D)|Z = 1] - p_n \mu_{0n}(p_n)}{\mathbb{E}[1 - D|Z = 1] - p_n}, \\
&\quad \delta_{0a} = \delta_{1a} = \delta_{0c} = \delta_{1c} = \delta_{0df} = \delta_{1df} = \delta_{0n} = \delta_{1n} = 0, \\
&\quad p_a = 0, p_c = \mathbb{E}[D|Z = 1], p_{df} = \mathbb{E}[D|Z = 0], p_n = \mathbb{E}[1 - D|Z = 1] - \mathbb{E}[D|Z = 0], \\
&\quad \left. \mu_{0n}(p_n) \in \left[ \mu_{F_{0n}^{UB}}(p_n), \mu_{F_{0n}^{LB}}(p_n) \right], \mu_{1n}, \mu_{0a}, \mu_{1a} \in [\inf \mathcal{Y}, \sup \mathcal{Y}] \right\},
\end{aligned}$$

2.  $p_{df} = \mathbb{E}[1 - D|Z = 1] < \mathbb{E}[D|Z = 0]$

This case is symmetric to the previous. By setting  $\tilde{D} = 1 - D$  and  $\tilde{Z} = 1 - Z$ , we

can adapt the previous identification results.

Hence,  $p_n = \mathbb{E}[1 - D|Z = 1] - p_{df} = 0$ ,  $p_c = \mathbb{E}[1 - D|Z = 0] - p_n = \mathbb{E}[1 - D|Z = 0]$ , and  $p_a = \mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]$ . Equations (3)-(4) imply

$$\begin{aligned}\mathbb{P}(Y_0 \leq y|T = c) &= \mathbb{P}(Y \leq y|D = 0, Z = 0), \\ \mathbb{P}(Y_0 \leq y|T = df) &= \mathbb{P}(Y \leq y|D = 0, Z = 1),\end{aligned}$$

Following a similar reasoning as before and combining Equations (5)-(6), we obtain the following identified set for  $F_{1a}$ :

$$\begin{aligned}\Theta_I(F_{1a}) &= \left\{ F_{1a} \in \mathcal{F} : \right. \\ &\max \left\{ \frac{(\mathbb{P}(Y \in (y, y'], D = 1|Z = 1) - \mathbb{E}[1 - D|Z = 0])}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \cdot \right. \\ &\quad \mathbb{1}\{\mathbb{P}(Y \in (y, y'], D = 1|Z = 1) - \mathbb{E}[1 - D|Z = 0] \geq 0\} \\ &\quad \frac{(\mathbb{P}(Y \in (y, y'], D = 1|Z = 0) - \mathbb{E}[1 - D|Z = 1])}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \cdot \\ &\quad \left. \mathbb{1}\{\mathbb{P}(Y \in (y, y'], D = 1|Z = 0) - \mathbb{E}[1 - D|Z = 1] \geq 0\} \right\} \\ &\leq F_{1a}(y') - F_{1a}(y) \leq \\ &\min \left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 1)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \mathbb{1}\left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 1)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \leq 1 \right\} \right. \\ &\quad \left. + \mathbb{1}\left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 1)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} > 1 \right\}, \right. \\ &\quad \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 0)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \mathbb{1}\left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 0)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \leq 1 \right\} \\ &\quad \left. + \mathbb{1}\left\{ \frac{\mathbb{P}(Y \in (y, y'], D = 1|Z = 0)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} > 1 \right\} \right\} \text{ for all } y < y'. \end{aligned} \tag{31}$$

$$F_{1a}^{LB}(y) \equiv \max \left\{ F_{1a}^{LB_1}(y), F_{1a}^{LB_0}(y) \right\} \leq F_{1a}(y) \leq \min \left\{ F_{1a}^{UB_1}(y), F_{1a}^{UB_0}(y) \right\} \equiv F_{1a}^{UB}(y),$$

where

$$\begin{aligned}
F_{1a}^{LB_1}(y) &\equiv \frac{(\mathbb{P}(Y \leq y, D = 1|Z = 1) - \mathbb{E}[1 - D|Z = 0])}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \cdot \\
&\quad \mathbb{1}\{\mathbb{P}(Y \leq y, D = 1|Z = 1) - \mathbb{E}[1 - D|Z = 0] \geq 0\}, \\
F_{1a}^{LB_0}(y) &\equiv \frac{(\mathbb{P}(Y \leq y, D = 1|Z = 0) - \mathbb{E}[1 - D|Z = 1])}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \cdot \\
&\quad \mathbb{1}\{\mathbb{P}(Y \leq y, D = 1|Z = 0) - \mathbb{E}[1 - D|Z = 1] \geq 0\}, \\
F_{1a}^{UB_1}(y) &\equiv \frac{\mathbb{P}(Y \leq y, D = 1|Z = 1)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 1|Z = 1)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \leq 1\right\} \\
&\quad + \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 1|Z = 1)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} > 1\right\}, \\
F_{1a}^{UB_0}(y) &\equiv \frac{\mathbb{P}(Y \leq y, D = 1|Z = 0)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 1|Z = 0)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} \leq 1\right\} \\
&\quad + \mathbb{1}\left\{\frac{\mathbb{P}(Y \leq y, D = 1|Z = 0)}{\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1]} > 1\right\}.
\end{aligned}$$

The identified sets for  $F_{dt}$ ,  $d \in \{0, 1\}$ ,  $t \in \{c, df\}$  are given by

$$\begin{aligned}
F_{1c}(y) &= \frac{\mathbb{P}(Y \leq y, D = 1|Z = 1) - (\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1])F_{1a}(y)}{\mathbb{E}[1 - D|Z = 0]}, \\
F_{1df}(y) &= \frac{\mathbb{P}(Y \leq y, D = 1|Z = 0) - (\mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1])F_{1a}(y)}{\mathbb{E}[1 - D|Z = 1]},
\end{aligned}$$

where  $F_{1a} \in \Theta_I(F_{1a})$ .

$$\begin{aligned}
\Theta_I^{3,2}(A_2) &= \left\{ \Gamma \in \mathbb{R}^{20} : \theta_{0a}(p_a) = \theta_{1a}(p_a) = \mu_{1a}(p_a) - \mu_{0a}, \theta_{0n}(p_n) = \theta_{1n}(p_n) = \mu_{1n} - \mu_{0n}, \right. \\
&\quad \theta_{0c}(p_c) = \theta_{1c}(p_c) = \frac{\mathbb{E}[YD|Z = 1] - p_a \mu_{1a}(p_a)}{\mathbb{E}[D|Z = 1] - p_a} - \mathbb{E}[Y|D = 0, Z = 0], \\
&\quad \theta_{0df}(p_{df}) = \theta_{1df}(p_{df}) = \frac{\mathbb{E}[YD|Z = 0] - p_a \mu_{1a}(p_a)}{\mathbb{E}[D|Z = 0] - p_a} - \mathbb{E}[Y|D = 0, Z = 1], \\
&\quad \delta_{0a} = \delta_{1a} = \delta_{0c} = \delta_{1c} = \delta_{0df} = \delta_{1df} = \delta_{0n} = \delta_{1n} = 0, \\
&\quad p_a = \mathbb{E}[D|Z = 0] - \mathbb{E}[1 - D|Z = 1], p_c = \mathbb{E}[1 - D|Z = 0], \\
&\quad p_{df} = \mathbb{E}[1 - D|Z = 1], p_n = 0, \\
&\quad \left. \mu_{1a}(p_a) \in \left[ \mu_{F_{1a}^{UB}}(p_a), \mu_{F_{1a}^{LB}}(p_a) \right], \mu_{1n}, \mu_{0n}, \mu_{0a} \in [\inf \mathcal{Y}, \sup \mathcal{Y}] \right\},
\end{aligned}$$

$$3. p_{df} = \mathbb{E}[D|Z = 0] = \mathbb{E}[1 - D|Z = 1]$$

Then  $p_a = p_n = 0$ ,  $p_c = \mathbb{E}[D|Z = 1] = \mathbb{E}[1 - D|Z = 0]$ .

$$\begin{aligned}\mathbb{P}(Y_1 \leq y|T = c) &= \mathbb{P}(Y \leq y|D = 1, Z = 1), \\ \mathbb{P}(Y_1 \leq y|T = df) &= \mathbb{P}(Y \leq y|D = 1, Z = 0), \\ \mathbb{P}(Y_0 \leq y|T = c) &= \mathbb{P}(Y \leq y|D = 0, Z = 0), \\ \mathbb{P}(Y_0 \leq y|T = df) &= \mathbb{P}(Y \leq y|D = 0, Z = 1),\end{aligned}$$

$$\begin{aligned}\Theta_I^{3,3}(A_2) &= \left\{ \Gamma \in \mathbb{R}^{20} : \theta_{0a}(p_a) = \theta_{1a}(p_a) = \mu_{1a} - \mu_{0a}, \theta_{0n}(p_n) = \theta_{1n}(p_n) = \mu_{1n} - \mu_{0n}, \right. \\ &\quad \theta_{0c}(p_c) = \theta_{1c}(p_c) = \mathbb{E}[Y|D = 1, Z = 1] - \mathbb{E}[Y|D = 0, Z = 0], \\ &\quad \theta_{0df}(p_{df}) = \theta_{1df}(p_{df}) = \mathbb{E}[Y|D = 1, Z = 0] - \mathbb{E}[Y|D = 0, Z = 1], \\ &\quad \delta_{0a} = \delta_{1a} = \delta_{0c} = \delta_{1c} = \delta_{0df} = \delta_{1df} = \delta_{0n} = \delta_{1n} = 0, \\ &\quad p_a = 0, p_c = \mathbb{E}[D|Z = 1] = \mathbb{E}[1 - D|Z = 0], \\ &\quad p_{df} = \mathbb{E}[D|Z = 0] = \mathbb{E}[1 - D|Z = 1], p_n = 0, \\ &\quad \left. \mu_{1a}, \mu_{0n}, \mu_{1n}, \mu_{0a} \in [\inf \mathcal{Y}, \sup \mathcal{Y}] \right\}.\end{aligned}$$

Finally,

$$\Theta_I^3(A_2) = \begin{cases} \Theta_I^{3,1}(A_2), & \text{if } p_{df} = \mathbb{E}[D|Z = 0] < \mathbb{E}[1 - D|Z = 1], \\ \Theta_I^{3,2}(A_2), & \text{if } p_{df} = \mathbb{E}[1 - D|Z = 1] < \mathbb{E}[D|Z = 0], \\ \Theta_I^{3,3}(A_2), & \text{if } p_{df} = \mathbb{E}[D|Z = 0] = \mathbb{E}[1 - D|Z = 1]. \end{cases}$$

□

## B Proof of Proposition 3

*Proof.* When  $\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] > 0$ , under Assumption 3, there are no defiers, i.e.,  $p_{df} = 0$ . This implies  $p_c = \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]$ ,  $p_a = \mathbb{E}[D|Z = 0]$ , and  $p_n = \mathbb{E}[1 - D|Z = 1]$ . From the main text, we have shown the following: for all Borel set

$A$ ,

$$\mathbb{P}(Y_{11} \in A|T = c) = \frac{\mathbb{P}(Y \in A, D = 1|Z = 1) - p_a \mathbb{P}(Y_{11} \in A|T = a)}{p_c},$$

$$\max \left\{ \frac{\mathbb{P}(Y \in A, D = 1|Z = 1) - p_c}{p_a}, 0 \right\} \leq \mathbb{P}(Y_{11} \in A|T = a) \leq \min \left\{ \frac{\mathbb{P}(Y \in A, D = 1|Z = 1)}{p_a}, 1 \right\},$$

$$\mathbb{P}(Y_{10} \in A|T = a) = \mathbb{P}(Y \in A|D = 1, Z = 0), \mathbb{P}(Y_{01} \in A|T = n) = \mathbb{P}(Y \in A|D = 0, Z = 1),$$

$$\max \left\{ \frac{\mathbb{P}(Y \in A, D = 0|Z = 0) - p_c}{p_n}, 0 \right\} \leq \mathbb{P}(Y_{00} \in A|T = n) \leq \min \left\{ \frac{\mathbb{P}(Y \in A, D = 0|Z = 0)}{p_n}, 1 \right\},$$

$$\mathbb{P}(Y_{00} \in A|T = c) = \frac{\mathbb{P}(Y \in A, D = 0|Z = 0) - p_n \mathbb{P}(Y_{00} \in A|T = n)}{p_c}.$$

The results of this proposition are straightforward if  $A$  is set to be  $(-\infty, y]$ , as one can easily check that the bounds are well-defined cdfs.  $\square$

**Corollary 3.** Suppose Assumptions 1 and 3 hold, and  $\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] > 0$ . Then,

$$p_c = \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0], \quad p_a = \mathbb{E}[D|Z = 0], \quad p_n = \mathbb{E}[1 - D|Z = 1], \quad p_{df} = 0,$$

$$\mu_{F_{11a}}^{UB} \leq \mu_{11a} \leq \mu_{F_{11a}}^{LB}, \quad \mu_{F_{00n}}^{UB} \leq \mu_{00n} \leq \mu_{F_{00n}}^{LB},$$

$$\mu_{11c} = \frac{\mathbb{E}[YD|Z = 1] - p_a \mu_{11a}}{p_c}, \quad \mu_{00c} = \frac{\mathbb{E}[Y(1 - D)|Z = 0] - p_n \mu_{00n}}{p_c},$$

$$\mu_{10a} = \mathbb{E}[Y|D = 1, Z = 0], \quad \mu_{01n} = \mathbb{E}[Y|D = 0, Z = 1].$$

These bounds are sharp. For continuous outcomes with strictly increasing cdfs, we have:

$$\begin{aligned} \mu_{F_{11a}}^{LB} &= \mathbb{E} \left[ Y|D = 1, Z = 1, Y > F_{Y|D=1, Z=1}^{-1} \left( 1 - \frac{p_a}{\mathbb{E}[D|Z = 1]} \right) \right], \\ \mu_{F_{11a}}^{UB} &= \mathbb{E} \left[ Y|D = 1, Z = 1, Y < F_{Y|D=1, Z=1}^{-1} \left( \frac{p_a}{\mathbb{E}[D|Z = 1]} \right) \right], \\ \mu_{F_{00n}}^{LB} &= \mathbb{E} \left[ Y|D = 0, Z = 0, Y > F_{Y|D=0, Z=0}^{-1} \left( 1 - \frac{p_n}{\mathbb{E}[1 - D|Z = 0]} \right) \right], \\ \mu_{F_{00n}}^{UB} &= \mathbb{E} \left[ Y|D = 0, Z = 0, Y < F_{Y|D=0, Z=0}^{-1} \left( \frac{p_n}{\mathbb{E}[1 - D|Z = 0]} \right) \right]. \end{aligned}$$

*Proof.* The results of this corollary hold from Proposition 3 and Lemma 1.  $\square$

## B.1 Proof of Corollary 2

*Proof.* By definition,  $\delta_{1a} = \mu_{11a} - \mu_{10a}$ ,  $\delta_{0n} = \mu_{01n} - \mu_{00n}$ ,  $\mu_{11c} - \mu_{00c} = \theta_{0c} + \delta_{1c} = \theta_{1c} + \delta_{0c}$ . Therefore, from Corollary 3, the results hold.  $\square$

## C Binary Outcomes

In this subsection, we provide bounds on the same causal parameters when the outcome of interest is binary. Equation (12) becomes

$$\max_{d \in \{0,1\}} \left\{ \max_{z \in \{0,1\}} \mathbb{P}(Y = 0, D = d | Z = z) + \max_{z \in \{0,1\}} \mathbb{P}(Y = 1, D = d | Z = z) \right\} \leq 1. \quad (32)$$

The bounds in Proposition 1 become

$$\Theta_I(p_{df}) = \begin{cases} \left[ \max \left\{ \max_{s \in \{0,1\}} \{ \mathbb{P}(Y = 0, D = s | Z = 1 - s) - \mathbb{P}(Y = 0, D = s | Z = s), \right. \right. \\ \mathbb{P}(Y = 1, D = s | Z = 1 - s) - \mathbb{P}(Y = 1, D = s | Z = s), \\ \left. \left. \mathbb{P}(D = s | Z = 1 - s) - \mathbb{P}(D = s | Z = s) \right\}, 0 \right], & \text{if inequality (32) holds} \\ \emptyset, & \text{otherwise.} \end{cases} \quad (33)$$

In the main text, we have shown that for any Borel set  $A$ ,

$$\begin{aligned} & \max \left\{ \frac{\mathbb{P}(Y \in A, D = 1 | Z = 1) - p_c}{p_a}, \frac{\mathbb{P}(Y \in A, D = 1 | Z = 0) - p_{df}}{p_a}, 0 \right\} \\ & \leq \mathbb{P}(Y_1 \in A | T = a) \leq \\ & \min \left\{ \frac{\mathbb{P}(Y \in A, D = 1 | Z = 1)}{p_a}, \frac{\mathbb{P}(Y \in A, D = 1 | Z = 0)}{p_a}, 1 \right\}. \end{aligned}$$

Therefore, for  $A = \{1\}$ , we have

$$\begin{aligned} & \max \left\{ \frac{\mathbb{P}(Y = 1, D = 1 | Z = 1) - p_c}{p_a}, \frac{\mathbb{P}(Y = 1, D = 1 | Z = 0) - p_{df}}{p_a}, 0 \right\} \\ & \leq \mu_{1a}(p_a) \leq \\ & \min \left\{ \frac{\mathbb{P}(Y = 1, D = 1 | Z = 1)}{p_a}, \frac{\mathbb{P}(Y = 1, D = 1 | Z = 0)}{p_a}, 1 \right\}. \end{aligned}$$



By a similar argument, we have

$$\begin{aligned} & \max \left\{ \frac{\mathbb{P}(Y = 1, D = 0 | Z = 0) - p_c}{p_n}, \frac{\mathbb{P}(Y = 1, D = 0 | Z = 1) - p_{df}}{p_n}, 0 \right\} \\ & \leq \mu_{0n}(p_n) \leq \\ & \min \left\{ \frac{\mathbb{P}(Y = 1, D = 0 | Z = 0)}{p_n}, \frac{\mathbb{P}(Y = 1, D = 0 | Z = 1)}{p_n}, 1 \right\}. \end{aligned}$$

## D Estimation and inference

### D.1 Estimation and inference under random assignment and exclusion restriction

Our identification results rely on the observable data  $(Y, D, Z)$  equipped with the probability measure  $\mathbb{P}$ . In this section, we provide a brief introduction of the estimation and inference methods based on a sample of size  $n$ , consisting of independently and identically distributed observations  $\{(Y_i, D_i, Z_i)\}_{i=1}^n$  drawn from  $\mathbb{P}$ .

First, we aim to estimate the identified set for the proportion of defiers, as characterized in Proposition 1. We can estimate the upper bound  $UB(p_{df}) \equiv \min\{\mathbb{E}[D | Z = 0], \mathbb{E}[1 - D | Z = 1]\}$  by

$$\begin{aligned} \widehat{UB}(p_{df}) &\equiv \min\{\widehat{\mathbb{E}}[D | Z = 0], \widehat{\mathbb{E}}[1 - D | Z = 1]\}, \\ \widehat{\mathbb{E}}[D | Z = 0] &= \frac{\sum_{i=1}^n D_i(1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)}, \widehat{\mathbb{E}}[1 - D | Z = 1] = \frac{\sum_{i=1}^n (1 - D_i)Z_i}{\sum_{i=1}^n Z_i}. \end{aligned}$$

It is straightforward to verify that  $\widehat{\mathbb{E}}[D | Z = 0]$  and  $\widehat{\mathbb{E}}[1 - D | Z = 1]$  are consistent estimators of  $\mathbb{E}[D | Z = 0]$  and  $\mathbb{E}[1 - D | Z = 1]$  by the law of large numbers. Consistency of  $\widehat{UB}(p_{df})$  for  $UB(p_{df})$  follows from the continuous mapping theorem, as the minimum function  $\min\{\cdot, \cdot\}$  is continuous everywhere on  $\mathbb{R}^2$ .

If the outcome  $Y$  has finite support,  $Y \in \{y_1, y_2, \dots, y_t\}$ ,  $t < \infty$ , we can estimate the lower bound of defier's proportion as

$$\widehat{LB}^{disc}(p_{df}) \equiv \max_{s \in \{0,1\}} \left\{ \max_{y \in \{y_1, y_2, \dots, y_t\}} \left\{ \frac{\sum_{i=1}^n \mathbb{1}\{Y_i = y\} \mathbb{1}\{D_i = s\} \mathbb{1}\{Z_i = 1 - s\}}{\mathbb{1}\{Z_i = 1 - s\}} - \frac{\sum_{i=1}^n \mathbb{1}\{Y_i = y\} \mathbb{1}\{D_i = s\} \mathbb{1}\{Z_i = s\}}{\mathbb{1}\{Z_i = s\}} \right\} \right\}.$$

According to the law of large number and the continuous mapping theorem,  $\widehat{LB}^{disc}(p_{df})$  is a consistent estimator of  $LB(p_{df}) \equiv \max_{s \in \{0,1\}} \left\{ \sup_A \{\mathbb{P}(Y \in A, D = s | Z = 1 - s) - \mathbb{P}(Y \in A, D = s | Z = s)\} \right\}$ .

If the outcome  $Y$  is continuous, taking the supremum over all Borel sets  $A$  in the support of  $Y$

becomes intractable. To address this, we partition the support of  $Y$  into  $d$  disjoint intervals, denoted as  $I_1, \dots, I_d$ , and define the lower bound for the proportion of defiers based on this partition as

$$LB^{part}(p_{df}) \equiv \max_{A \in \mathcal{A}^{part}} \{\mathbb{P}(Y \in A, D = s \mid Z = 1 - s) - \mathbb{P}(Y \in A, D = s \mid Z = s)\},$$

where  $\mathcal{A}^{part}$  is the finite  $\sigma$ -algebra generated by partitioning intervals  $I_1, \dots, I_d$ , and  $d$  is a fixed number specified by the researcher. Since  $\mathcal{A}^{part}$  is a subset of the Borel  $\sigma$ -algebra generated by  $Y$ , it follows that  $LB^{part}(p_{df}) \leq LB(p_{df})$ . As a result, when employing  $LB^{part}(p_{df})$ , the estimated lower bound on the proportion of defiers remains valid, though the sharpness of the bound may be compromised. Increasing the number of partition intervals can improve the approximation, making  $LB^{part}(p_{df})$  closer to the sharp lower bound  $LB(p_{df})$ . However, partitioning the support into a large number of intervals, especially when the number of intervals is large relative to the sample size or grows with the sample size, can introduce inference problems associated with many moment inequalities. For example, the confidence intervals might be too conservative when correcting for the large number of moment inequalities using critical values based on max-statistics. Therefore, we recommend choosing the number of intervals  $d$  sufficiently small relative to the sample size  $n$ . An estimator of  $LB^{part}(p_{df})$  is

$$\widehat{LB}^{part}(p_{df}) \equiv \max_{s \in \{0,1\}} \left\{ \max_{A \in \mathcal{A}^{part}} \left\{ \frac{\sum_{i=1}^n \mathbb{1}\{Y_i \in A\} \mathbb{1}\{D_i = s\} \mathbb{1}\{Z_i = 1 - s\}}{\mathbb{1}\{Z_i = 1 - s\}} - \frac{\sum_{i=1}^n \mathbb{1}\{Y_i \in A\} \mathbb{1}\{D_i = s\} \mathbb{1}\{Z_i = s\}}{\mathbb{1}\{Z_i = s\}} \right\} \right\},$$

which is consistent based on the law of large number and the continuous mapping theorem. To conclude, we estimate the identified bound of defier's proportion as

$$\widehat{\Theta}_I(p_{df}) = \begin{cases} [\widehat{LB}^{disc}(p_{df}), \widehat{UB}(p_{df})] & \text{if } Y \text{ is discrete,} \\ [\widehat{LB}^{part}(p_{df}), \widehat{UB}(p_{df})] & \text{if } Y \text{ is continuous.} \end{cases}$$

Next, we conduct inference on the identified bound for the proportion of defiers, which corresponds to an intersection of finitely many intervals when the number of partitions  $d$  is sufficiently small. We employ the inference procedure in [Nevo and Rosen \(2012\)](#), which is specifically designed for identified bounds in the form of intersections of finite intervals. The inference steps include selecting the set of moment inequalities close to binding at the estimated boundary points of the identified set. The corresponding estimates are then adjusted using their standard errors, along with critical values based on the distribution of the maximum or minimum over the moments. For a detailed description of the inference procedure and the underlying asymptotic arguments, we refer to Section IV of [Nevo and Rosen \(2012\)](#). Implementing their method yields a confidence band, denoted as  $CI(p_{df})$ , uniformly valid over points in the identified bound of defier's proportion.

The proportions of the remaining types can be identified as  $p_a = \mathbb{P}(D = 1 \mid Z = 0) - p_{df}$ ,

**Table 6:** Estimation and Inference of proportion of each type under  $A_2$ 

Parameter	Estimated identified bounds	Confidence band
$p_a$	[0, 0.1436]	[0, 0.5369]
$p_c$	[0.1967, 0.3403]	[0, 0.3800]
$p_n$	[0.4123, 0.5559]	[0.3375, 1]
$p_{df}$	[0.1038, 0.2474]	[0, 0.2813]
Observations		381

$p_n = \mathbb{P}(D = 0 \mid Z = 1)$ , and  $p_c = \mathbb{P}(D = 1 \mid Z = 1) - \mathbb{P}(D = 1 \mid Z = 0) + p_{df}$ . Given these expressions, we estimate their respective identified bounds as

$$\begin{aligned}\widehat{\Theta}_I(p_a) &= \left[ \frac{\sum_{i=1}^n D_i(1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)} - \widehat{UB}(p_{df}), \frac{\sum_{i=1}^n D_i(1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)} - \widehat{LB}(p_{df}) \right], \\ \widehat{\Theta}_I(p_n) &= \left[ \frac{\sum_{i=1}^n (1 - D_i)Z_i}{\sum_{i=1}^n Z_i} - \widehat{UB}(p_{df}), \frac{\sum_{i=1}^n (1 - D_i)Z_i}{\sum_{i=1}^n Z_i} - \widehat{LB}(p_{df}) \right], \\ \widehat{\Theta}_I(p_c) &= \left[ \frac{\sum_{i=1}^n D_i Z_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n D_i(1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)} + \widehat{LB}(p_{df}), \frac{\sum_{i=1}^n D_i Z_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n D_i(1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)} + \widehat{UB}(p_{df}) \right],\end{aligned}$$

where  $\widehat{LB}(p_{df}) = \widehat{LB}^{disc}(p_{df})$  if  $Y$  is discrete and  $\widehat{LB}(p_{df}) = \widehat{LB}^{part}(p_{df})$  if  $Y$  is continuous. It is straightforward that the proposed estimators of the bounds are consistent. We also apply the inference procedure in [Nevo and Rosen \(2012\)](#) to conduct uniform inference on the proportion of each type.

In the application of [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#), which investigates the effects of social norm perceptions on women's labor force participation in Saudi Arabia. The sample consists of 381 observations. The outcome is a continuous index measuring the long-term women labor force performance, so we equally partition the support of  $Y$  into 20 intervals. Table 6 presents the confidence bands for the proportions of all types.

After estimating the type proportions, we proceed to estimate the parameters  $\theta_{0c} = \theta_{1c}$  and  $\theta_{0df} = \theta_{1df}$ , which measure the local average direct effects of the treatment for compliers and defiers. Based on the identification results in Section 4, we estimate the tractable valid outer set of

$$\hat{\theta}_{0c} = \hat{\theta}_{1c} \text{ as}$$

$$\left[ \begin{array}{l} \max \left\{ \frac{\sum_{i=1}^n Y_i D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a \hat{\mu}_{F_{1a}^{LB_1}}(\hat{p}_a)}{\sum_{i=1}^n D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n \hat{\mu}_{F_{0n}^{UB_1}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n}, \right. \\ \frac{\sum_{i=1}^n Y_i D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a \hat{\mu}_{F_{1a}^{LB_1}}(\hat{p}_a)}{\sum_{i=1}^n D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n \hat{\mu}_{F_{0n}^{UB_0}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n}, \\ \frac{\sum_{i=1}^n Y_i D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a \hat{\mu}_{F_{1a}^{LB_0}}(\hat{p}_a)}{\sum_{i=1}^n D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n \hat{\mu}_{F_{0n}^{UB_1}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n}, \\ \left. \frac{\sum_{i=1}^n Y_i D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a \hat{\mu}_{F_{1a}^{LB_0}}(\hat{p}_a)}{\sum_{i=1}^n D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n \hat{\mu}_{F_{0n}^{UB_0}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n} \right\}, \\ \min \left\{ \frac{\sum_{i=1}^n Y_i D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a \hat{\mu}_{F_{1a}^{UB_1}}(\hat{p}_a)}{\sum_{i=1}^n D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n \hat{\mu}_{F_{0n}^{LB_1}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n}, \right. \\ \frac{\sum_{i=1}^n Y_i D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a \hat{\mu}_{F_{1a}^{UB_1}}(\hat{p}_a)}{\sum_{i=1}^n D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n \hat{\mu}_{F_{0n}^{LB_0}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n}, \\ \frac{\sum_{i=1}^n Y_i D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a \hat{\mu}_{F_{1a}^{UB_0}}(\hat{p}_a)}{\sum_{i=1}^n D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n \hat{\mu}_{F_{0n}^{LB_1}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n}, \\ \left. \frac{\sum_{i=1}^n Y_i D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a \hat{\mu}_{F_{1a}^{UB_0}}(\hat{p}_a)}{\sum_{i=1}^n D_i Z_i / \sum_{i=1}^n Z_i - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n \hat{\mu}_{F_{0n}^{LB_0}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i)(1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_n} \right\} \end{array} \right],$$

and the tractable valid outer set of  $\hat{\theta}_{0df} = \hat{\theta}_{1df}$  as

$$\left[ \begin{array}{l} \max \left\{ \frac{\sum_{i=1}^n Y_i D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a \hat{\mu}_{F_{1a}^{LB_1}}(\hat{p}_a)}{\sum_{i=1}^n D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n \hat{\mu}_{F_{0n}^{UB_1}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n}, \right. \\ \frac{\sum_{i=1}^n Y_i D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a \hat{\mu}_{F_{1a}^{LB_1}}(\hat{p}_a)}{\sum_{i=1}^n D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n \hat{\mu}_{F_{0n}^{UB_0}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n}, \\ \frac{\sum_{i=1}^n Y_i D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a \hat{\mu}_{F_{1a}^{LB_0}}(\hat{p}_a)}{\sum_{i=1}^n D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n \hat{\mu}_{F_{0n}^{UB_1}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n}, \\ \left. \frac{\sum_{i=1}^n Y_i D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a \hat{\mu}_{F_{1a}^{LB_0}}(\hat{p}_a)}{\sum_{i=1}^n D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n \hat{\mu}_{F_{0n}^{UB_0}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n} \right\}, \\ \min \left\{ \frac{\sum_{i=1}^n Y_i D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a \hat{\mu}_{F_{1a}^{UB_1}}(\hat{p}_a)}{\sum_{i=1}^n D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n \hat{\mu}_{F_{0n}^{LB_1}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n}, \right. \\ \frac{\sum_{i=1}^n Y_i D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a \hat{\mu}_{F_{1a}^{UB_1}}(\hat{p}_a)}{\sum_{i=1}^n D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n \hat{\mu}_{F_{0n}^{LB_0}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n}, \\ \frac{\sum_{i=1}^n Y_i D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a \hat{\mu}_{F_{1a}^{UB_0}}(\hat{p}_a)}{\sum_{i=1}^n D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n \hat{\mu}_{F_{0n}^{LB_1}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n}, \\ \left. \frac{\sum_{i=1}^n Y_i D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a \hat{\mu}_{F_{1a}^{UB_0}}(\hat{p}_a)}{\sum_{i=1}^n D_i (1 - Z_i) / \sum_{i=1}^n (1 - Z_i) - \hat{p}_a} - \frac{\sum_{i=1}^n Y_i (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n \hat{\mu}_{F_{0n}^{LB_0}}(\hat{p}_n)}{\sum_{i=1}^n (1 - D_i) Z_i / \sum_{i=1}^n Z_i - \hat{p}_n} \right\} \end{array} \right],$$

given  $(\hat{p}_a, \hat{p}_n)$  in the confidence band of identified set of types' proportions. If  $\hat{\mu}_{F_{1a}^{LB_z}}, \hat{\mu}_{F_{1a}^{UB_z}}, \hat{\mu}_{F_{0n}^{LB_z}}, \hat{\mu}_{F_{0n}^{UB_z}}$  are consistent estimators of  $\mu_{F_{1a}^{LB_z}}, \mu_{F_{1a}^{UB_z}}, \mu_{F_{0n}^{LB_z}}, \mu_{F_{0n}^{UB_z}}, z \in \{0, 1\}$ , in Lemma 1, then the proposed estimators consistently estimate the lower and upper bounds of the tractable valid outer sets by the law of large numbers and the continuous mapping theorem. We impose the estimators

as

$$\begin{aligned}
\hat{\mu}_{F_{1a}^{LBz}} &= \frac{\sum_{i=1}^n Y_i D_i \mathbb{1}\{Z_i = z\} \mathbb{1}\{Y_i > \hat{F}_{Y|D=1, Z=z}^{-1}(1 - \hat{\alpha})\}}{\sum_{i=1}^n D_i \mathbb{1}\{Z_i = z\} \mathbb{1}\{Y_i > \hat{F}_{Y|D=1, Z=z}^{-1}(1 - \hat{\alpha})\}}, \\
\hat{\mu}_{F_{1a}^{UBz}} &= \frac{\sum_{i=1}^n Y_i D_i \mathbb{1}\{Z_i = z\} \mathbb{1}\{Y_i < \hat{F}_{Y|D=1, Z=z}^{-1}(\hat{\alpha})\}}{\sum_{i=1}^n D_i \mathbb{1}\{Z_i = z\} \mathbb{1}\{Y_i < \hat{F}_{Y|D=1, Z=z}^{-1}(\hat{\alpha})\}}, \\
\hat{\mu}_{F_{0n}^{LBz}} &= \frac{\sum_{i=1}^n Y_i (1 - D_i) \mathbb{1}\{Z_i = z\} \mathbb{1}\{Y_i > \hat{F}_{Y|D=0, Z=z}^{-1}(1 - \hat{\gamma})\}}{\sum_{i=1}^n (1 - D_i) \mathbb{1}\{Z_i = z\} \mathbb{1}\{Y_i > \hat{F}_{Y|D=0, Z=z}^{-1}(1 - \hat{\gamma})\}}, \\
\hat{\mu}_{F_{0n}^{UBz}} &= \frac{\sum_{i=1}^n Y_i (1 - D_i) \mathbb{1}\{Z_i = z\} \mathbb{1}\{Y_i < \hat{F}_{Y|D=0, Z=z}^{-1}(\hat{\gamma})\}}{\sum_{i=1}^n (1 - D_i) \mathbb{1}\{Z_i = z\} \mathbb{1}\{Y_i < \hat{F}_{Y|D=0, Z=z}^{-1}(\hat{\gamma})\}}, \\
\text{where } \hat{F}_{Y|D=1, Z=z}^{-1}(q) &= \min \left\{ y : \frac{\sum_{i=1}^n D_i \mathbb{1}\{Z_i = z\} \mathbb{1}\{Y_i \leq y\}}{\sum_{i=1}^n D_i \mathbb{1}\{Z_i = z\}} \geq q \right\}, \\
\hat{F}_{Y|D=0, Z=z}^{-1}(q) &= \min \left\{ y : \frac{\sum_{i=1}^n (1 - D_i) \mathbb{1}\{Z_i = z\} \mathbb{1}\{Y_i \leq y\}}{\sum_{i=1}^n (1 - D_i) \mathbb{1}\{Z_i = z\}} \geq q \right\}, \\
\hat{\alpha} &= \frac{\hat{p}_a}{\sum_{i=1}^n D_i \mathbb{1}\{Z_i = z\} / \sum_{i=1}^n \mathbb{1}\{Z_i = z\}}, \\
\hat{\gamma} &= \frac{\hat{p}_n}{\sum_{i=1}^n (1 - D_i) \mathbb{1}\{Z_i = z\} / \sum_{i=1}^n \mathbb{1}\{Z_i = z\}},
\end{aligned}$$

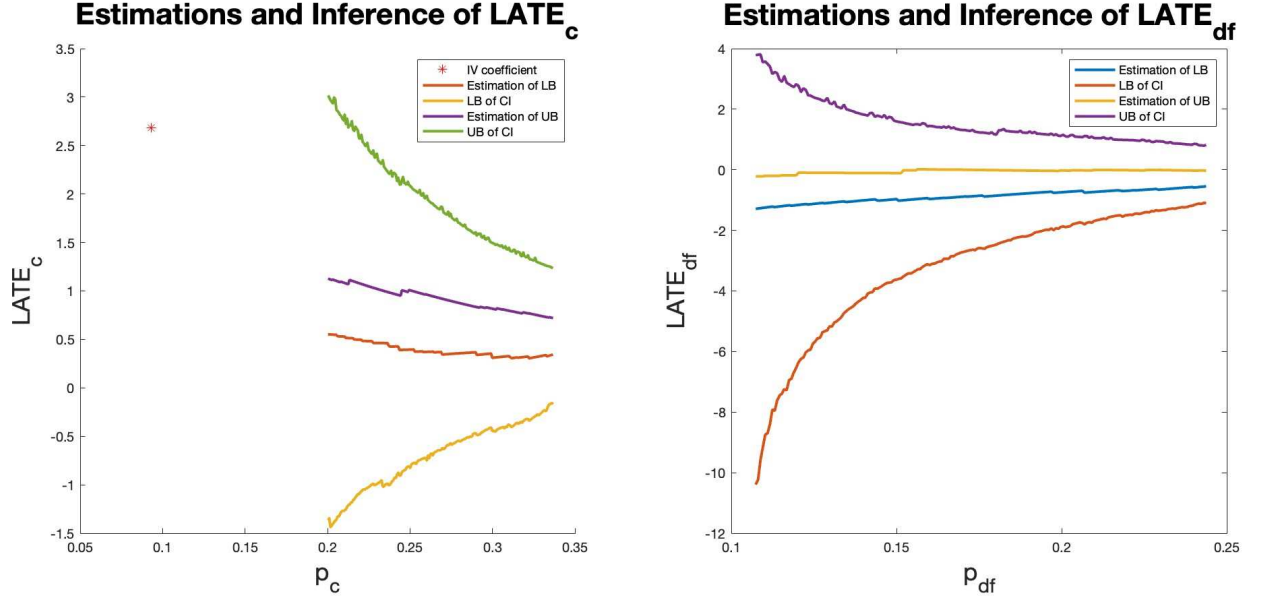
and Proposition 2 of [Lee \(2009\)](#) proves that the above estimators are consistent under the condition that the outcome variable  $Y$  has bounded support.

The estimated outer sets of  $\theta_{1c} = \theta_{0c}$  take the form of finite intersections of intervals. Therefore, we apply the inference procedure in [Nevo and Rosen \(2012\)](#) to conduct uniform inference on these parameters. We apply our estimation and inference methods to the study in [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#). Figure 4 presents the estimated identified bounds and the corresponding confidence bands for  $\theta_{1c} = \theta_{0c}$  and  $\theta_{1df} = \theta_{0df}$  given that  $p_c$  and  $p_{df}$  lie in their respective estimated identified sets.

## D.2 Estimation and inference under random assignment and monotonicity

We can use the sample analog to estimate the bounds derived in Corollary 2. First, we need to estimate the trimming probabilities  $\alpha \equiv \frac{p_a}{\mathbb{E}[D|Z=1]}$  and  $\gamma \equiv \frac{p_n}{\mathbb{E}[1-D|Z=0]}$  by

$$\begin{aligned}
\hat{\alpha} &= \frac{\hat{p}_a}{(\sum_{i=1}^n D_i Z_i) / (\sum_{i=1}^n Z_i)}, \\
\hat{\gamma} &= \frac{\hat{p}_n}{(\sum_{i=1}^n (1 - D_i)(1 - Z_i)) / (\sum_{i=1}^n (1 - Z_i))}.
\end{aligned}$$



**Figure 4:** Estimated bounds and inference bands for LAICDEs under  $A_2$  and IV estimand

Then, we can estimate the  $\hat{\alpha}$ th and  $(1-\hat{\alpha})$ th quantile of the empirical distribution of  $Y \mid D = 1, Z = 1$ , and calculate the estimators of bounds on  $\delta_{1a}$  as

$$\begin{aligned}\widehat{LB}_1 &= \frac{\sum_{i=1}^n Y_i D_i Z_i \mathbb{1}\{Y_i \leq \hat{F}_{Y|D=1,Z=1}^{-1}(\hat{\alpha})\}}{\sum_{i=1}^n D_i Z_i \mathbb{1}\{Y_i \leq \hat{F}_{Y|D=1,Z=1}^{-1}(\hat{\alpha})\}} - \frac{\sum_{i=1}^n Y_i D_i (1 - Z_i)}{\sum_{i=1}^n D_i (1 - Z_i)}, \\ \widehat{UB}_1 &= \frac{\sum_{i=1}^n Y_i D_i Z_i \mathbb{1}\{Y_i > \hat{F}_{Y|D=1,Z=1}^{-1}(1 - \hat{\alpha})\}}{\sum_{i=1}^n D_i Z_i \mathbb{1}\{Y_i > \hat{F}_{Y|D=1,Z=1}^{-1}(1 - \hat{\alpha})\}} - \frac{\sum_{i=1}^n Y_i D_i (1 - Z_i)}{\sum_{i=1}^n D_i (1 - Z_i)}, \\ \hat{F}_{Y|D=1,Z=1}^{-1}(q) &= \min \left\{ y : \frac{\sum_{i=1}^n D_i Z_i \mathbb{1}\{Y_i \leq y\}}{\sum_{i=1}^n D_i Z_i} \geq q \right\}.\end{aligned}$$

Repeating the same process, we can estimate the bounds on  $\delta_{0n}$  as

$$\begin{aligned}\widehat{LB}_2 &= \frac{\sum_{i=1}^n Y_i (1 - D_i) Z_i}{\sum_{i=1}^n (1 - D_i) Z_i} - \frac{\sum_{i=1}^n Y_i (1 - D_i) (1 - Z_i) \mathbb{1}\{Y_i > \hat{F}_{Y|D=0,Z=0}^{-1}(1 - \hat{\gamma})\}}{\sum_{i=1}^n (1 - D_i) (1 - Z_i) \mathbb{1}\{Y_i > \hat{F}_{Y|D=0,Z=0}^{-1}(1 - \hat{\gamma})\}}, \\ \widehat{UB}_2 &= \frac{\sum_{i=1}^n Y_i (1 - D_i) Z_i}{\sum_{i=1}^n (1 - D_i) Z_i} - \frac{\sum_{i=1}^n Y_i (1 - D_i) (1 - Z_i) \mathbb{1}\{Y_i \leq \hat{F}_{Y|D=0,Z=0}^{-1}(\hat{\gamma})\}}{\sum_{i=1}^n (1 - D_i) (1 - Z_i) \mathbb{1}\{Y_i \leq \hat{F}_{Y|D=0,Z=0}^{-1}(\hat{\gamma})\}}, \\ \hat{F}_{Y|D=0,Z=0}^{-1}(q) &= \min \left\{ y : \frac{\sum_{i=1}^n (1 - D_i) (1 - Z_i) \mathbb{1}\{Y_i \leq y\}}{\sum_{i=1}^n (1 - D_i) (1 - Z_i)} \geq q \right\},\end{aligned}$$

and estimate the bounds on  $\theta_{1c} + \delta_{0c} = \theta_{0c} + \delta_{1c}$  as

$$\begin{aligned}\widehat{\text{LB}}_3 &= \frac{\sum_{i=1}^n Y_i D_i Z_i \mathbb{1}\{Y_i \leq \hat{F}_{Y|D=1,Z=1}^{-1}(1-\hat{\alpha})\}}{\sum_{i=1}^n D_i Z_i \mathbb{1}\{Y_i \leq \hat{F}_{Y|D=1,Z=1}^{-1}(1-\hat{\alpha})\}} - \frac{\sum_{i=1}^n Y_i (1-D_i)(1-Z_i) \mathbb{1}\{Y_i > \hat{F}_{Y|D=0,Z=0}^{-1}(\hat{\gamma})\}}{\sum_{i=1}^n (1-D_i)(1-Z_i) \mathbb{1}\{Y_i > \hat{F}_{Y|D=0,Z=0}^{-1}(\hat{\gamma})\}}, \\ \widehat{\text{UB}}_3 &= \frac{\sum_{i=1}^n Y_i D_i Z_i \mathbb{1}\{Y_i > \hat{F}_{Y|D=1,Z=1}^{-1}(\hat{\alpha})\}}{\sum_{i=1}^n D_i Z_i \mathbb{1}\{Y_i > \hat{F}_{Y|D=1,Z=1}^{-1}(\hat{\alpha})\}} - \frac{\sum_{i=1}^n Y_i (1-D_i)(1-Z_i) \mathbb{1}\{Y_i \leq \hat{F}_{Y|D=0,Z=0}^{-1}(1-\hat{\gamma})\}}{\sum_{i=1}^n (1-D_i)(1-Z_i) \mathbb{1}\{Y_i \leq \hat{F}_{Y|D=0,Z=0}^{-1}(1-\hat{\gamma})\}}.\end{aligned}$$

Proposition 2 in Lee (2009) proves that these estimate bounds are consistent.

In order to obtain the confidence intervals for bounds described in Corollary 2, we have to know the asymptotic distributions of the estimated bound values. We apply Proposition 3 in Lee (2009) to derive the asymptotic behaviors of those estimated bound values. For the estimated bounds on  $\delta_{1a}$ ,  $\sqrt{N}(\widehat{\text{LB}}_1 - \text{LB}_1) \xrightarrow{d} N(0, V_{\text{LB}_1} + V_{C_1})$ ,  $\sqrt{N}(\widehat{\text{UB}}_1 - \text{UB}_1) \xrightarrow{d} N(0, V_{\text{UB}_1} + V_{C_1})$ , where

$$\begin{aligned}V_{\text{LB}_1} &= \frac{V(Y \mid D=1, Z=1, Y \leq F_{Y|D=1,Z=1}^{-1}(\alpha))}{\mathbb{E}[DZ]\alpha} \\ &\quad + \frac{(F_{Y|D=1,Z=1}^{-1}(\alpha) - \mathbb{E}[Y \mid D=1, Z=1, Y \leq F_{Y|D=1,Z=1}^{-1}(\alpha)])^2(1-\alpha)}{\mathbb{E}[DZ]\alpha} \\ &\quad + \left( \frac{F_{Y|D=1,Z=1}^{-1}(\alpha) - \mathbb{E}[Y \mid D=1, Z=1, Y \leq F_{Y|D=1,Z=1}^{-1}(\alpha)]}{\alpha} \right)^2 V_{\alpha_1}, \\ V_{\text{UB}_1} &= \frac{V(Y \mid D=1, Z=1, Y > F_{Y|D=1,Z=1}^{-1}(1-\alpha))}{\mathbb{E}[DZ]\alpha} \\ &\quad + \frac{(F_{Y|D=1,Z=1}^{-1}(1-\alpha) - \mathbb{E}[Y \mid D=1, Z=1, Y > F_{Y|D=1,Z=1}^{-1}(1-\alpha)])^2(1-\alpha)}{\mathbb{E}[DZ]\alpha} \\ &\quad + \left( \frac{F_{Y|D=1,Z=1}^{-1}(1-\alpha) - \mathbb{E}[Y \mid D=1, Z=1, Y > F_{Y|D=1,Z=1}^{-1}(1-\alpha)]}{\alpha} \right)^2 V_{\alpha_1}, \\ V_{\alpha_1} &= \alpha^2 \left[ \frac{1-p_a/\alpha}{\mathbb{E}[Z]p_a/\alpha} + \frac{1-p_a}{(1-\mathbb{E}[Z])p_a} \right], \\ V_{C_1} &= \frac{V(Y \mid D=1, Z=0)}{\mathbb{E}[D(1-Z)]}.\end{aligned}$$

The estimators of bounds on  $\delta_{0n}$  satisfy  $\sqrt{N}(\widehat{\text{LB}}_2 - \text{LB}_2) \xrightarrow{d} N(0, V_{\text{LB}_2} + V_{C_2})$ ,  $\sqrt{N}(\widehat{\text{UB}}_2 - \text{UB}_2) \xrightarrow{d}$



$N(0, V_{\text{UB}_2} + V_{C_2})$ , where

$$\begin{aligned}
V_{\text{LB}_2} &= \frac{V(Y \mid D = 0, Z = 0, Y > F_{Y|D=0, Z=0}^{-1}(1 - \gamma))}{\mathbb{E}[(1 - D)(1 - Z)]\gamma} \\
&+ \frac{(F_{Y|D=0, Z=0}^{-1}(1 - \gamma) - \mathbb{E}[Y \mid D = 0, Z = 0, Y > F_{Y|D=0, Z=0}^{-1}(1 - \gamma)])^2(1 - \gamma)}{\mathbb{E}[(1 - D)(1 - Z)]\gamma} \\
&+ \left( \frac{F_{Y|D=0, Z=0}^{-1}(1 - \gamma) - \mathbb{E}[Y \mid D = 0, Z = 0, Y > F_{Y|D=0, Z=0}^{-1}(1 - \gamma)]}{\gamma} \right)^2 V_{\gamma_1}, \\
V_{\text{UB}_2} &= \frac{V(Y \mid D = 0, Z = 0, Y \leq F_{Y|D=0, Z=0}^{-1}(\gamma))}{\mathbb{E}[(1 - D)(1 - Z)]\gamma} \\
&+ \frac{(F_{Y|D=0, Z=0}^{-1}(\gamma) - \mathbb{E}[Y \mid D = 0, Z = 0, Y \leq F_{Y|D=0, Z=0}^{-1}(\gamma)])^2(1 - \gamma)}{\mathbb{E}[(1 - D)(1 - Z)]\gamma} \\
&+ \left( \frac{F_{Y|D=0, Z=0}^{-1}(\gamma) - \mathbb{E}[Y \mid D = 0, Z = 0, Y \leq F_{Y|D=0, Z=0}^{-1}(\gamma)]}{\gamma} \right)^2 V_{\gamma_1}, \\
V_{\gamma_1} &= \gamma^2 \left[ \frac{1 - p_n/\gamma}{\mathbb{E}[(1 - Z)]p_n/\gamma} + \frac{1 - p_n}{\mathbb{E}[Z]p_n} \right], \\
V_{C_2} &= \frac{V(Y \mid D = 0, Z = 1)}{\mathbb{E}[(1 - D)Z]}.
\end{aligned}$$

Finally, the estimators of bounds on  $\delta_{1c} + \theta_{0c} = \delta_{0c} + \theta_{1c}$  have the asymptotic distributions as

$\sqrt{N}(\widehat{LB}_3 - LB_3) \xrightarrow{d} N(0, V_{LB_3} + V_{LB_4})$ ,  $\sqrt{N}(\widehat{UB}_3 - UB_3) \xrightarrow{d} N(0, V_{UB_3} + V_{UB_4})$ , where

$$\begin{aligned}
V_{LB_3} &= \frac{V(Y | D=1, Z=1, Y \leq F_{Y|D=1, Z=1}^{-1}(1-\alpha))}{\mathbb{E}[DZ](1-\alpha)} \\
&\quad + \frac{(F_{Y|D=1, Z=1}^{-1}(1-\alpha) - \mathbb{E}[Y | D=1, Z=1, Y \leq F_{Y|D=1, Z=1}^{-1}(1-\alpha)])^2 \alpha}{\mathbb{E}[DZ](1-\alpha)} \\
&\quad + \left[ \frac{F_{Y|D=1, Z=1}^{-1}(1-\alpha) - \mathbb{E}[Y | D=1, Z=1, Y \leq F_{Y|D=1, Z=1}^{-1}(1-\alpha)]}{(1-\alpha)} \right]^2 V_{\alpha_2}, \\
V_{LB_4} &= \frac{V(Y | D=0, Z=0, Y > F_{Y|D=0, Z=0}^{-1}(\gamma))}{\mathbb{E}[(1-D)(1-Z)](1-\gamma)} \\
&\quad + \frac{(F_{Y|D=0, Z=0}^{-1}(\gamma) - \mathbb{E}[Y | D=0, Z=0, Y > F_{Y|D=0, Z=0}^{-1}(\gamma)])^2 \gamma}{\mathbb{E}[(1-D)(1-Z)](1-\gamma)} \\
&\quad + \left[ \frac{F_{Y|D=0, Z=0}^{-1}(\gamma) - \mathbb{E}[Y | D=0, Z=0, Y > F_{Y|D=0, Z=0}^{-1}(\gamma)]}{1-\gamma} \right]^2 V_{\gamma_2}, \\
V_{UB_3} &= \frac{V(Y | D=1, Z=1, Y > F_{Y|D=1, Z=1}^{-1}(\alpha))}{\mathbb{E}[DZ](1-\alpha)} \\
&\quad + \frac{(F_{Y|D=1, Z=1}^{-1}(\alpha) - \mathbb{E}[Y | D=1, Z=1, Y > F_{Y|D=1, Z=1}^{-1}(\alpha)])^2 \alpha}{\mathbb{E}[DZ](1-\alpha)} \\
&\quad + \left( \frac{F_{Y|D=1, Z=1}^{-1}(\alpha) - \mathbb{E}[Y | D=1, Z=1, Y > F_{Y|D=1, Z=1}^{-1}(\alpha)]}{1-\alpha} \right)^2 V_{\alpha_2}, \\
V_{UB_4} &= \frac{V(Y | D=0, Z=0, Y \leq F_{Y|D=0, Z=0}^{-1}(1-\gamma))}{\mathbb{E}[(1-D)(1-Z)](1-\gamma)} \\
&\quad + \frac{(F_{Y|D=0, Z=0}^{-1}(1-\gamma) - \mathbb{E}[Y | D=0, Z=0, Y \leq F_{Y|D=0, Z=0}^{-1}(1-\gamma)])^2 \gamma}{\mathbb{E}[(1-D)(1-Z)](1-\gamma)} \\
&\quad + \left[ \frac{F_{Y|D=0, Z=0}^{-1}(1-\gamma) - \mathbb{E}[Y | D=0, Z=0, Y \leq F_{Y|D=0, Z=0}^{-1}(1-\gamma)]}{1-\gamma} \right]^2 V_{\gamma_2}, \\
V_{\alpha_2} &= (1-\alpha)^2 \left[ \frac{1-p_a/(1-\alpha)}{\mathbb{E}[Z]p_a/(1-\alpha)} + \frac{1-p_a}{(1-\mathbb{E}[Z])p_a} \right], \\
V_{\gamma_2} &= (1-\gamma)^2 \left[ \frac{1-p_n/(1-\gamma)}{\mathbb{E}[(1-Z)]p_n/(1-\gamma)} + \frac{1-p_n}{\mathbb{E}[Z]p_n} \right].
\end{aligned}$$

For all variances derived in the asymptotic distributions, we can estimate them as their sample analogs.

Since we care about the true value of the average controlled effects instead of their identified sets, we apply the method proposed by [Imbens and Manski \(2004\)](#) to construct the confidence intervals for identified bounds in Corollary 2. The 95% confidence intervals are constructed as  $[\widehat{LB}_i - \bar{C}_i \hat{\sigma}_{LB_i} / \sqrt{N}, \widehat{UB}_i + \bar{C}_i \hat{\sigma}_{UB_i} / \sqrt{N}]$ ,  $i = 1, 2, 3$ , where  $\hat{\sigma}_{LB_i}$  and  $\hat{\sigma}_{UB_i}$  are estimated standard

**Table 7:** Estimated identified sets and confidence sets under  $A_3$ 

Parameter	Estimated Bounds	Confidence Sets
$\delta_{1a}$	[0.0557, 0.7155]	[-0.0834, 1.0184]
$\delta_{0n}$	[0.0829, 0.2505]	[-0.0135, 0.3445]
$\delta_{1c} + \theta_{0c}$	[-0.7140, 1.9242]	[-0.9048, 2.4797]

deviations of the corresponding lower bounds and upper bounds, and  $\bar{C}_i$  satisfies

$$\Phi \left( \bar{C}_i + \sqrt{N} \frac{\widehat{UB}_i - \widehat{LB}_i}{\max(\hat{\sigma}_{LB_i}, \hat{\sigma}_{UB_i})} \right) - \Phi(-\bar{C}_i) = 0.95.$$

We revisit the empirical study in [Bursztyn, González, and Yanagizawa-Drott \(2020\)](#) with imposing the assumptions of random assignment and monotonicity. Using the estimation and inference methods in this section, we estimate the identified sets and construct uniform confidence intervals for the parameters  $\delta_{1a}$ ,  $\delta_{0n}$ , and  $\delta_{1c} + \theta_{0c}$ . The results are presented in [Table 7](#).