Sharp IV bounds on average treatment effects on the treated and other populations under endogeneity and noncompliance

Martin Huber[1], Lukas Laffers[2], and Giovanni Mellace[3]

[1]University of Fribourg, Dept. of Economics, [2]Matej Bel University, Dept. of Mathematics, [3]University of Southern Denmark, Dept. of Business and Economics

Abstract

In the presence of an endogenous binary treatment and a valid binary instrument, causal effects are (nonparametrically) point identified only for the subpopulation of compliers, given that the treatment is monotone in the instrument. With the exception of the entire population, causal inference for further subpopulations has been widely ignored in econometrics. Therefore, we invoke treatment monotonicity and/or dominance assumptions on the mean potential outcomes across subpopulations to derive sharp bounds on the average treatment effects on the treated, who often bear considerable policy relevance, as well as on other groups (non-treated, entire population, compliers, always takers, and never takers). Furthermore, we use our methods to assess the educational impact of a school voucher program in Colombia on various subpopulations and also discuss testable implications of our assumptions.

Keywords: instrument, noncompliance, principal stratification, nonparametric bounds.

JEL classification: C14, C31, C36.

We have benefited from comments by Sylvain Chabé-Ferret, Philipp Denter, Michael Lechner, Fabbrizia Mealli, Christoph Rothe, Anthony Strittmatter, seminar/conference participants in St. Gallen (research seminar), Tor Vergata (lunch seminar), Engelberg (labor seminar 2011), and Vancouver (SOLE 2011), and four anonymous referees. Addresses for correspondence: Martin Huber (martin.huber@unifr.ch), Department of Economics, University of Fribourg, Bd. de Pérolles 90, CH-1700 Fribourg; Lukáš Lafférs (lukas.laffers@gmail.com), Department of Mathematics, Matej Bel University, Tajovského 40, SK-97411 Banská Bystrica; Giovanni Mellace (giome@sam.sdu.dk), Department of Business and Economics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M.

1 Introduction

Endogeneity of the (binary) treatment variable and noncompliance to the treatment assignment in randomized experiments are widespread phenomena in the evaluation of treatment effects, see for instance Bloom (1984). Given a valid instrumental variable (IV) that is randomly assigned and has no direct effect on the mean potential outcomes and (weakly) positive monotonicity of the treatment in the instrument, Imbens and Angrist (1994) (see also Angrist, Imbens, and Rubin, 1996) show that the average treatment effect (ATE) is only identified in the subpopulation of compliers. The latter correspond to those whose the treatment status is equal to (i.e. reacts on) the instrument if both the treatment and the instrument are binary.

Whether the LATE is a relevant parameter heavily depends on the empirical context and has been controversially discussed in the literature, see for instance Imbens (2009), Deaton (2010), and Heckman and Urzúa (2010). Typically, researchers would like to identify the ATEs on the treated or the entire population. Note that these parameters are themselves weighted averages of the ATEs on several subpopulations, including the always takers (always treated irrespective of the instrument) and the never takers (never treated irrespective of the instrument). Maybe due to the fact that in a nonparametric framework, point identification is generally not feasible for the never takers, always takers, the treated, and the entire population (unless the complier share is 100 %), groups other than the compliers have (apart from the entire population) been widely ignored in the econometric literature.[1]

The main contribution of this paper is to derive nonparametric bounds on ATEs of populations that are potentially more policy relevant than the LATE on the compliers, which may not be externally valid. In particular, we also consider the treated population, which is of major interest in the program evaluation literature to assess the program effects on actual participants. In contrast to the commonly invoked full independence between the instrument and the potential outcomes/treatment states, we only assume mean independence between the instrument and the potential outcomes (within subpopulations) as well as the subpopulations. Moreover, we discuss the identifying power of (i) monotonicity of the treatment in the

[1] An exception is Frölich and Lechner (2010) who also point identify the ATEs on the always takers and never takers. To this end, they invoke both IV and selection on observables (or conditional independence, see for instance Imbens, 2004) assumptions. However, this identification strategy stands in contrast to virtually all other IV applications, where an instrument is used exactly for the reason that no other source of identification (such as selection on observables) is available.

instrument and/or (ii) mean dominance of the potential outcomes of one subpopulation over the others. Monotonicity and dominance, either w.r.t. the mean or to the entire distribution (i.e., stochastic dominance), have also been considered in a different context, namely under non-random sample selection and attrition, see for instance Zhang and Rubin (2003), Lechner and Melly (2007), Blundell, Gosling, Ichimura and Meghir (2007), Zhang, Rubin and Mealli (2008), and Lee (2009), and Huber and Mellace (2013a). We use the principal stratification framework suggested by Frangakis and Rubin (2002) to derive sharp bounds for the ATEs on the always takers, never takers, the treated, the non-treated, and the entire population.[2] As a further contribution, we find testable implications of the IV mean independence within subpopulations and mean dominance when monotonicity is invoked.

Partial identification of economic parameters in general goes back to Manski (1989, 1994) and Robins (1989). Previous work on nonparametric bounds under treatment endogeneity, which is the problem considered here, has almost exclusively focused on the ATE in the entire population,[3] but neglected further populations. E.g., Manski (1990) bounds the ATE by solely relying on independence between the mean potential outcomes and the instrument.[4] Considering binary outcomes, Balke and Pearl (1997) provide sharp bounds for the ATE under full (rather than mean) independence between the instrument and the potential values of the treatment (given the instrument) and the outcome (given the treatment) with and without monotonicity (see also Dawid, 2003) of the treatment in the instrument. Shaikh and Vytlacil (2011) bound the ATE on the entire population in the binary outcome case under monotonicity, too, and assume the treatment effect to be either weakly positive or weakly negative for all individuals (while the direction is a priori not restricted). See Bhattacharya, Shaikh and Vytlacil (2008) for an application. Cheng and Small (2006) extend the results for binary outcomes to three treatments (in contrast to the standard binary treatment framework considered here) under particular forms of (one-sided) noncompliance. Richardson and Robins (2010) is the only study apart from ours that also bounds the effects on further populations (compliers, defiers, never takers, and always takers). They assume full independence and a

[2] In addition, Appendix A.8 provides the bounds for the treated subpopulations receiving and not receiving the instrument.

[3] For the derivation of semiparametric bounds on the ATE on the entire population, see Chiburis (2010) and the references therein.

[4] As it is the aim of this paper to provide bounds for further populations, we also need to assume that the proportions of the subpopulations are independent of the instrument, otherwise the bounds on those populations might differ for different values of the instrument.

binary outcome, but do not consider monotonicity or any form of mean dominance.

In contrast to much of the epidemiologic literature, Heckman and Vytlacil (2001) and Kitagawa (2009) allow for both discrete and continuous outcomes. Kitagawa (2009) partially identifies the potential outcome distributions for the entire population under (various forms of) full independence between the instrument and potential treatments/outcomes as well as monotonicity and derives bounds on the ATE. Also Heckman and Vytlacil (2001) assume full independence of the instrument, but invoke a nonparametric threshold crossing model characterizing the treatment choice instead of monotonicity for deriving the bounds on the ATE. However, by the results of Vytlacil (2002), both approaches are equivalent. One interesting finding of Heckman and Vytlacil (2001) and Kitagawa (2009) is that the width of their bounds is the same as those of Manski (1990), given that the monotonicity/threshold crossing model assumptions are satisfied. The present work adds to the literature on nonparametric bounds under endogeneity by considering more populations and an extended set of identifying assumptions than any of the previous studies.

The identifying power of monotonicity and mean dominance is demonstrated in an empirical application to Colombia's "Programa de Ampliación de Cobertura de la Educación Secundaria", which provided pupils from low income families with vouchers for private secondary schooling. Using experimental data previously analyzed by Angrist, Bettinger, Bloom, King and Kremer (2002), we aim at assessing the program's impact on the educational achievement of various subpopulations. In particular, we find (in addition to the point identified complier effect) a significantly positive ATE on the treated which lies within reasonably tight bounds. This is an interesting result because it suggests that this and similar interventions have a positive effect on the participants, who are likely more policy relevant than the latent population of compliers.

The remainder of this paper is organized as follows. Section 2 characterizes the endogeneity/noncompliance problem based on principal stratification. Section 3 discusses the identifying assumptions and derives bounds on the ATEs for various populations. Section 4 briefly presents the estimators. In Section 5, we consider an empirical application to experimental education data. Section 6 concludes.

2 Using principal stratification to characterize noncompliance

Suppose that we want to estimate the effect of a binary treatment $D \in \{1, 0\}$ (e.g., a training activity) on an outcome Y (e.g., labor market success such as employment or earnings) evaluated at some point in time after the treatment. We use the experimental framework to motivate the problems of endogeneity and noncompliance. Assume that individuals are randomly assigned into treatment or non-treatment according to the binary assignment variable $Z \in \{1, 0\}$, which will serve as instrument. Denote by $D_i(z)$ the potential treatment state for $Z = z$ and by $Y_i(d)$ the potential outcome (see for instance Rubin, 1974) of individual i under treatment $D = d$. Throughout the discussion we will rule out interference between individuals or general equilibrium effects of the treatment by invoking the "Stable Unit Treatment Value assumption" (SUTVA), see for instance Rubin (1990). The SUTVA is formalized in Assumption 1:

Assumption 1:

$Y_i(d) \perp d_j \quad \text{and} \quad D_i(z) \perp z_j \ \forall j \neq i, d \in \{0, 1\}$ (SUTVA).

Even under Assumption 1, the individual effect $Y_i(1) - Y_i(0)$ can never be evaluated as individual i is either treated or not treated, but cannot be observed in both states. I.e., the observed outcome $Y_i = D_i \cdot Y_i(1) + (1 - D_i) \cdot Y_i(0)$. However, under particular assumptions aggregate parameters such as the average treatment effect (ATE) $\Delta = E[Y(1)] - E[Y(0)]$ can be identified. E.g., assume that compliance in an experiment is perfect such that $D_i(1) = 1$ and $D_i(0) = 0$ for all individuals i. In this case and under successful randomization, $E[Y|Z = 1] - E[Y|Z = 0] = E[Y|D = 1] - E[Y|D = 0] = E[Y(1)] - E[Y(0)] = \Delta$, where the first equality follows from perfect compliance and the second from random assignment. I.e., the ATE is identified because all individuals are compliers. However, if post-assignment complications occur such that $D_i(z) \neq z$ for some z and some individuals i, selection bias may flaw the validity of the evaluation in spite of the randomization of the assignment. This is due to the potential threat that individuals systematically select themselves into the treatment according to their potential outcomes.

Using the principal stratification framework advocated by Frangakis and Rubin (2002), the population can be divided into four principal strata, denoted by T, according to the choice of D

as a reaction of Z. Angrist, Imbens and Rubin (1996) refer to the four groups as (i) compliers, who react on the instrument in the intended way by taking the treatment when $Z = 1$ and abstaining from it when $Z = 0$, (ii) always takers, who are always treated irrespective of the assignment, (iii) never takers, who are never treated irrespective of the assignment, and (iv) defiers, who are treated when not assigned, but not treated when assigned. Table 1 visualizes this definition.

Table 1: Principal strata

Principal strata (T)	D(1)	D(0)	Notion
11	1	1	Always takers
10	1	0	Compliers
01	0	1	Defiers
00	0	0	Never takers

It is obvious that we cannot directly observe the principal stratum an individual belongs to as either $D(1)$ or $D(0)$ is known. Therefore, without the imposition of further assumptions, neither the principal strata proportions nor the distribution of Y within any stratum are identified. To see this, note that the observed values of Z and D generate four observed subgroups which are all mixtures of two principal strata. This implies that any individual i with a particular combination of Z_i, D_i may belong to two principal strata, see Table 2.

Table 2: Observed subgroups and principal strata

Observed subgroups	principal strata
$\{i : Z_i = 1, D_i = 1\}$	subject i belongs either to 11 or to 10
$\{i : Z_i = 1, D_i = 0\}$	subject i belongs either to 01 or to 00
$\{i : Z_i = 0, D_i = 1\}$	subject i belongs either to 11 or to 01
$\{i : Z_i = 0, D_i = 0\}$	subject i belongs either to 10 or to 00

As second identifying restriction maintained throughout the paper, we will assume Z to be independent of (i) the mean potential outcomes within principal strata and (ii) of the strata proportions, which has also been considered in Frölich (2007):

Assumption 2:

(i) $E(Y(d)|T = t, Z = 1) = E(Y(d)|T = t, Z = 0) = E(Y(d)|T = t)$ for $d \in \{0, 1\}$ and $t \in \{11, 10, 01, 00\}$ (mean independence within principal strata),

(ii) $\Pr(T = t|Z = 1) = \Pr(T = t|Z = 0) = \Pr(T = t)$ for $t \in \{11, 10, 01, 00\}$ (unconfounded strata proportions).

Assumption 2 (i) postulates that mere assignment does not have any direct effect on the mean potential outcomes within any stratum other than through the treatment,[5] i.e. mean independence within principal strata.[6] Taking assignment to a training as an example, it rules out that the average labor market success given T changes as a reaction to merely being assigned. I.e., what should matter is whether the training is actually received. By Assumption 2 (ii), the proportion of any stratum conditional on the instrument is equal to its unconditional proportion in the entire population. This holds for instance under random assignment, where Z is fully independent of the joint distribution of $(D(1), D(0))$ and thus, of T. Alternatively to the unconditional validity of Assumption 2, one may assume that it only holds conditional on some observed pre-assignment variables X. This is closely related to the framework of Frölich (2007) who shows point identification of the LATE under a conditionally valid instrument (given X). In the further discussion, conditioning on X will be kept implicit, such that all results either refer to the experimental framework or to an analysis within cells of X.

Unfortunately, even under Assumptions 1 and 2, point identification is not obtained. Let $\pi_t = \Pr(T = t)$ denote a particular proportion and $P_{d|z} \equiv \Pr(D = d|Z = z)$ the observed treatment probability conditional on assignment status. Under Assumption 2 (ii), which ensures that the strata proportions conditional on the instrument are equal to the unconditional strata proportions, the relation between the observed $P_{d|z}$ and the latent π_t is as displayed in Table 3. Likewise, any observed conditional mean outcome is a mixture of the mean out-

Table 3: Observed probabilities and principal strata proportions

Observed cond. treatment prob.	princ. strata proportions
$P_{1\|1} = \Pr(D = 1\|Z = 1)$	$\pi_{11} + \pi_{10}$
$P_{0\|1} = \Pr(D = 0\|Z = 1)$	$\pi_{01} + \pi_{00}$
$P_{1\|0} = \Pr(D = 1\|Z = 0)$	$\pi_{11} + \pi_{01}$
$P_{0\|0} = \Pr(D = 0\|Z = 0)$	$\pi_{10} + \pi_{00}$

comes of two strata. E.g.,

[5]However, in contrast to the full independence considered in Imbens and Angrist (1994) and Angrist et al. (1996), it may affect higher moments.

[6]For the case that mean independence is not satisfied, Flores and Flores-Lagunes (2013) derive bounds on the LATE in the presence of an invalid instrument. In contrast, we will assume the instrument to be valid (in the sense that it satisfies Assumption 2) throughout the discussion.

$$
\begin{aligned}
E(Y|Z=1,D=1) &= \frac{\pi_{11}}{\pi_{11}+\pi_{10}} \cdot E(Y|Z=1,D=1,T=11) \\
&+ \frac{\pi_{10}}{\pi_{11}+\pi_{10}} \cdot E(Y|Z=1,D=1,T=10), \\
&= \frac{\pi_{11}}{\pi_{11}+\pi_{10}} \cdot E(Y|D=1,T=11) + \frac{\pi_{10}}{\pi_{11}+\pi_{10}} \cdot E(Y|D=1,T=10), \\
&= \frac{\pi_{11}}{\pi_{11}+\pi_{10}} \cdot E(Y(1)|T=11) + \frac{\pi_{10}}{\pi_{11}+\pi_{10}} \cdot E(Y(1)|T=10),
\end{aligned}
$$

where the second equality follows from Assumption 2 (i) and the third from the fact that the treatment is unconfounded within each stratum consisting of individuals with identical (non-)compliance behavior.

Thus, point identification of causal effects would require us to invoke further assumptions. E.g., under monotonicity of D in Z and effect homogeneity, the ATE on the entire population is identified. Albeit used in much of the IV literature, effect homogeneity is a very unattractive assumption given the rich empirical evidence on effect heterogeneity in the field of treatment evaluation. Under monotonicity and effect heterogeneity, the LATE on the compliers is identified, but this effect may be "too local" to be of policy interest. Fortunately, assumptions as monotonicity and mean dominance also bear identifying power for further populations and may yield informative bounds, as discussed in the next section.

3 Assumptions and interval identification

3.1 Mean independence within principal strata without further assumptions

The partial identification of ATEs on various populations will be based on bounding the mean potential outcomes $E(Y(1)|T=t)$, $E(Y(0)|T=t)$, with $t \in \{11, 10, 00, 01\}$. To this end, we assume that the support $\mathcal{Y}$ of the outcome variable Y is bounded, i.e., $\mathcal{Y} = [y^{LB}, y^{UB}]$ with $-\infty < y^{LB} < y^{UB} < \infty$, and that Y is continuous over $\mathcal{Y}$ (see Appendix A.5 for discrete outcomes). Boundedness of Y rules out infinite upper or lower bounds on the mean potential outcomes and thus, on the ATE in any population. We will refer to the bounds on any ATE as being informative if its identification region does not coincide with (i.e. is tighter than) $[y^{LB} - y^{UB}, y^{UB} - y^{LB}]$.

Partial identification is obtained in three steps. In the first step, we derive sharp bounds on the principal strata proportions using Assumption 2 (ii). As one can express three out of four proportions as a function of the remaining one, we only need to bound the latter. Therefore, all bounds are computed as functions of the defier proportion, but choosing any other principal stratum would entail the same results. The second step (which is mostly discussed in the appendix) gives the bounds on the mean potential outcomes and the ATEs conditional on the defier proportion. It makes use of the fact that each observed conditional mean outcome is a mixture of the mean potential outcomes of two principal strata, with the mixing probabilities corresponding to the relative principal strata proportions:

$$\begin{aligned}
E(Y|Z=1,D=1) &= \frac{\pi_{11}}{\pi_{11}+\pi_{10}} \cdot E(Y(1)|T=11) + \frac{\pi_{10}}{\pi_{11}+\pi_{10}} \cdot E(Y(1)|T=10), && (1)\\
E(Y|Z=0,D=1) &= \frac{\pi_{11}}{\pi_{11}+\pi_{01}} \cdot E(Y(1)|T=11) + \frac{\pi_{01}}{\pi_{11}+\pi_{01}} \cdot E(Y(1)|T=01), && (2)\\
E(Y|Z=0,D=0) &= \frac{\pi_{10}}{\pi_{00}+\pi_{10}} \cdot E(Y(0)|T=10) + \frac{\pi_{00}}{\pi_{00}+\pi_{10}} \cdot E(Y(0)|T=00), && (3)\\
E(Y|Z=1,D=0) &= \frac{\pi_{01}}{\pi_{00}+\pi_{01}} \cdot E(Y(0)|T=01) + \frac{\pi_{00}}{\pi_{00}+\pi_{01}} \cdot E(Y(0)|T=00). && (4)
\end{aligned}$$

Given the defier proportion (and thus, the mixing probabilities), the results of Horowitz and Manski (1995) (see Section 3.2 and Proposition 4 therein) provide us with sharp bounds on the mean potential outcomes within each of equations (1) to (4), whereas Assumption 2 (i) allows further tightening these bounds across equations. Using an approach inspired by Kitagawa (2009), we derive sharp bounds on the mean potential outcomes and the ATEs under Assumption 2 (i) and conditional on the defier proportion. Finally, taking the supremum (infimum) of the ATEs in the second step over admissible defier proportions that satisfy Assumption 2 (ii) yields the sharp upper (lower) bounds on the ATEs (see Appendix A.1.2).[7] As the bounds are continuous in π_{01} (as shown in Appendix A.1.4) and $\mathcal{P}^*$ is an interval (as shown in Lemma 1), the optima are attained by the extreme value theorem.

Concerning the bounds on the defier proportion, note that under Assumptions 1 and 2, Table 3 provides us with the following equations:

$$\begin{aligned}
\pi_{11} = P_{1|0} - \pi_{01} &\quad\Rightarrow\quad \pi_{01} \leq P_{1|0}, && (5)\\
\pi_{00} = P_{0|1} - \pi_{01} &\quad\Rightarrow\quad \pi_{01} \leq P_{0|1},\\
\pi_{10} = P_{1|1} - P_{1|0} + \pi_{01} &\quad\Rightarrow\quad \pi_{01} \geq P_{1|0} - P_{1|1},
\end{aligned}$$

[7]It is worth noting that without further restrictions, it is generally not possible that one particular value of π_{01} jointly optimizes the bounds on all ATEs considered.

and thus, the defier proportion must lie in the following set

$$\pi_{01} \in \mathcal{P} = [\max(0, P_{1|0} - P_{1|1}), \min(P_{1|0}, P_{0|1})]. \tag{6}$$

Note that these bounds are valid *outer* bounds, but they need not be sharp. Sharp bounds on the proportion of defiers have to be constructed based on the joint distribution of (Y, D, Z), rather than (D, Z) alone. Under Assumptions 1 and 2, the condition $\pi_{01} = 0$ leads to testable implications as studied in Huber and Mellace (2013b). However, in some cases zero may not be in the identified set of the defier share. In Appendix A.1.3, we show that admissibility of π_{01} is equivalent to checking four moment inequalites that generalize those inequalities outlined in Huber and Mellace (2013b). Furthermore, Appendix A.1.1 presents a linear programming procedure for constructing sharp bounds on π_{01} in the case of a discrete Y.[8] Under discreteness, linear programming can also be used for constructing sharp bounds on various types of ATEs, see Freyberger and Horowitz (2013) and Laffférs (2013).

We denote by $\pi_t^{\min}$ and $\pi_t^{\max}$ the sharp lower and upper bounds of $\pi_t,\ t = 11, 10, 01, 00$, respectively, and by $\mathcal{P}^*$ the *sharp* identified set for π_{01} under Assumptions 1 and 2. In Appendix A.1.1, we show that $\mathcal{P}^*$ is an interval. If $\mathcal{P} = \mathcal{P}^*$, then the remaining strata proportions can be bounded by substituting (6) into (5). It is easy to see that either $\pi_{01}^{\min} = 0$ or $\pi_{10}^{\min} = 0$ and either $\pi_{11}^{\min} = 0$ or $\pi_{00}^{\min} = 0$. The outer bounds $\mathcal{P}$ are equivalent to those derived in Richardson and Robins (2010) (equation (6) of Section 3.1, page 9). In contrast to their paper we only assume mean independence within principal strata in Assumption 2, rather than full independence.

In order to bound the ATEs on the four populations, we introduce some additional notation. We define $\bar{Y}_{z,d} = E(Y|Z = z, D = d)$ to be the conditional mean of Y given $Z = z$ and $D = d$. Furthermore, $F_{Y_{z,d}}(y) = \Pr(Y \leq y|Z = z, D = d)$ denotes the conditional cdf of Y given $Z = z$ and $D = d$. Let $q_{z,d}^t$ denote the share of individuals belonging to stratum $T = t$ in the observed subgroup with $Z = z$ and $D = d$. If necessary, we will denote by $q_{z,d}^{t,\pi_{01}^{\max}}$ and $q_{z,d}^{t,\pi_{01}^{\min}}$, the value of $q_{z,d}^t$ when π_{01} is equal to $\pi_{01}^{\max}$ or $\pi_{01}^{\min}$, respectively. Let $F_{Y_{z,d}}^{-1}(q_{z,d}^t) = \inf\{y :\ F_{Y_{z,d}}(y) \geq q_{z,d}^t\}$[9] be the conditional quantile function of Y given $Z = z$ and $D = d$. We can then define the lower and upper bounds of $E(Y|Z = z, D = d, T = t)$, which by

[8]We thank an anonymous referee for pointing out the difference between $\mathcal{P}$ and $\mathcal{P}^*$ and for suggesting the linear programming tool.

[9]We define $F_{Y_{z,d}}^{-1}(0) \equiv y^{LB}$ and $F_{Y_{z,d}}^{-1}(1) \equiv y^{UB}$.

Assumption 2 corresponds to $E(Y(d)|T=t)$ (see Section 2), as $\bar{Y}_{z,d}(\min|q^t_{z,d}) = E(Y|Z = z, D = d, Y \leq F^{-1}_{Y_{z,d}}(q^t_{z,d}))$ and $\bar{Y}_{z,d}(\max|q^t_{z,d}) \equiv E(Y|Z = z, D = d, Y \geq F^{-1}_{Y_{z,d}}(1-q^t_{z,d}))$, respectively. Finally, the ATEs on the various principal strata, the treated, the non-treated, and the entire population are denoted by $\Delta_t \equiv E(Y(1)-Y(0)|T=t)$ with $t \in \{11, 10, 01, 00\}$, $\Delta_{D=d} \equiv E(Y(1)-Y(0)|D=d)$ with $d \in \{1, 0\}$, and $\Delta \equiv E(Y(1)-Y(0))$, respectively. The superscripts "^{UB}" and "^{LB}" denote the sharp upper and lower bounds on the respective parameters, where sharpness of the bounds on some parameter $\tilde{\Delta}$ is defined as follows:

Definition 1 *Given the knowledge of the distribution of the observed data, $\tilde{\Delta}^{LB}$ and $\tilde{\Delta}^{UB}$ are sharp if $[\tilde{\Delta}^{LB}, \tilde{\Delta}^{UB}]$ is the shortest interval such that, for every $\tilde{\Delta} \in [\tilde{\Delta}^{LB}, \tilde{\Delta}^{UB}]$, we can construct principal strata proportions $\Pr(T|Z) : T = 11, 10, 01, 00,\ Z = 1, 0$ and potential outcome distributions $f(Y(1), Y(0)|T, Z) : T = 11, 10, 01, 00,\ Z = 1, 0$ that satisfy the imposed assumptions.*

Considering the ATE on the compliers (Δ_{10}), if $\pi_{01} = P_{1|1} - P_{1|0} \notin \mathcal{P}^*$, then the upper and lower bounds are, respectively,

$$\begin{aligned}
\Delta^{UB}_{10} &= \sup_{\pi_{01}\in\mathcal{P}^*} \Bigg[\frac{P_{1|1}\cdot\bar{Y}_{1,1} - (P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}), \bar{Y}_{0,1}(\min|q^{11}_{0,1})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \\
&\quad - \frac{P_{0|0}\cdot\bar{Y}_{0,0} - (P_{0|1}-\pi_{01})\cdot\min\left(\bar{Y}_{0,0}(\max|q^{00}_{0,0}), \bar{Y}_{1,0}(\max|q^{00}_{1,0})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \Bigg], \\
\Delta^{LB}_{10} &= \inf_{\pi_{01}\in\mathcal{P}^*} \Bigg[\frac{P_{1|1}\cdot\bar{Y}_{1,1} - (P_{1|0}-\pi_{01})\cdot\min\left(\bar{Y}_{1,1}(\max|q^{11}_{1,1}), \bar{Y}_{0,1}(\max|q^{11}_{0,1})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \\
&\quad - \frac{P_{0|0}\cdot\bar{Y}_{0,0} - (P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}), \bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \Bigg],
\end{aligned} \tag{7}$$

where $q^{11}_{1,1} = \frac{P_{1|0}-\pi_{01}}{P_{1|1}}$ (the share of always takers among those with $Z=1$ and $D=1$), $q^{11}_{0,1} = \frac{P_{1|0}-\pi_{01}}{P_{1|0}}$ (the share of always takers among those with $Z=0$ and $D=1$), $q^{00}_{1,0} = \frac{P_{0|1}-\pi_{01}}{P_{0|1}}$ (the share of never takers among those with $Z=0$ and $D=1$), and $q^{00}_{0,0} = \frac{P_{0|1}-\pi_{01}}{P_{0|0}}$ (the share of never takers among those with $Z=0$ and $D=0$). The proofs of the sharpness of these bounds as well as of any other bounds proposed below are provided in the appendix. If $\pi_{01} = P_{1|1} - P_{1|0} \in \mathcal{P}^*$, then $\pi_{10} = 0$ and the bounds are uninformative. Therefore, $\Delta^{UB}_{10} = y^{UB} - y^{LB}$ and $\Delta^{LB}_{10} = y^{LB} - y^{UB}$. In Appendix A.1.4, we show that $\Delta^{UB}(\pi_{01})$ and $\Delta^{LB}(\pi_{01})$ are continuous in π_{01}.

Four points are worth noting concerning the derivation of these bounds. Firstly and as already mentioned, they make use of Proposition 4 of Horowitz and Manski (1995), which in general only holds for continuous outcomes. Fortunately, it is easy to show that their results can also be applied to discrete outcomes after a modification of the trimming function, see Appendix A.5 for further details. Secondly, (7) has to be optimized w.r.t. admissible defier proportions, given by $\mathcal{P}^*$. Thirdly, mean independence within strata (Assumption 2 (i)) gives rise to the maximum and minimum operators. Note that in the first (third) line in (7) one computes the upper (lower) bound of the compliers' mean potential outcome under treatment by subtracting the lower (upper) bound of the mean potential outcome of the always takers. As their lower (upper) bound under treatment is not affected by the value of Z due to mean independence, the lower (upper) bound is the maximum (minimum) of the always takers' lower (upper) bounds for $Z = 1$ and $Z = 0$. An analogous result holds for lines 2 and 4 w.r.t. the bounds on the potential mean outcomes under non-treatment of the never takers. Finally, these bounds are informative only if $P_{1|0} < P_{1|1}$. This is equivalent to $\pi_{10} > \pi_{01}$, saying that the share of compliers is larger than the share of defiers. The reason is that if $P_{1|0} \geq P_{1|1}$, then $\pi_{01}^{\min} = P_{1|0} - P_{1|1} > 0$, which implies that $\pi_{10}^{\min} = 0$ (so that the non-existence of compliers cannot be ruled out) and $\Delta_{10}^{UB} = y^{UB} - y^{LB}$, $\Delta_{10}^{LB} = y^{UB} - y^{LB}$.

In a symmetric way one obtains the sharp upper and lower bounds on the ATE on the defiers, Δ_{01}. If $\pi_{01} = 0 \notin \mathcal{P}^*$:

$$\begin{aligned}
\Delta_{01}^{UB} &= \sup_{\pi_{01}\in\mathcal{P}^*}\left[\frac{P_{1|0}\cdot\bar{Y}_{0,1} - (P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q_{1,1}^{11}), \bar{Y}_{0,1}(\min|q_{0,1}^{11})\right)}{\pi_{01}}\right. \\
&- \left.\frac{P_{0|1}\cdot\bar{Y}_{1,0} - (P_{0|1}-\pi_{01})\cdot\min\left(\bar{Y}_{0,0}(\max|q_{0,0}^{00}), \bar{Y}_{1,0}(\max|q_{1,0}^{00})\right)}{\pi_{01}}\right], \\
\Delta_{01}^{LB} &= \inf_{\pi_{01}\in\mathcal{P}^*}\left[\frac{P_{1|0}\cdot\bar{Y}_{0,1} - (P_{1|0}-\pi_{01})\cdot\min\left(\bar{Y}_{1,1}(\max|q_{1,1}^{11}), \bar{Y}_{0,1}(\max|q_{0,1}^{11})\right)}{\pi_{01}}\right. \\
&- \left.\frac{P_{0|1}\cdot\bar{Y}_{1,0} - (P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q_{0,0}^{00}), \bar{Y}_{1,0}(\min|q_{1,0}^{00})\right)}{\pi_{01}}\right],
\end{aligned} \tag{8}$$

For the same reason mentioned above these bounds are informative only if $P_{1|0} > P_{1|1}$, i.e., if there are more defiers than compliers. Therefore, without imposing further assumptions, the bounds are informative either for the defiers or for the compliers, but never for both populations. Furthermore, unless $P_{1|1} - P_{1|0} = 0$, either positive (if $P_{1|1} - P_{1|0} > 0$) or negative (if $P_{1|0} - P_{1|1} > 0$) monotonicity of D in Z can be consistent with the data, but not both at the same time.

Concerning the always takers, note that their outcomes are only observed under treatment. The shares of always takers in the observed groups with $Z=1, D=1$ and $Z=0, D=1$ are, respectively, $\pi_{11}/(\pi_{11}+\pi_{10}) = (P_{1|0}-\pi_{01})/P_{1|1}$ and $\pi_{11}/(\pi_{11}+\pi_{01}) = (P_{1|0}-\pi_{01})/P_{1|0}$. Therefore, we can bound the upper and lower values of the mean potential outcome under treatment for this population by $\min\left(\bar{Y}_{1,1}(\max|q_{1,1}^{11,\pi_{01}^{\max}}), \bar{Y}_{0,1}(\max|q_{0,1}^{11,\pi_{01}^{\max}})\right)$ and $\max\left(\bar{Y}_{1,1}(\min|q_{1,1}^{11,\pi_{01}^{\max}}), \bar{Y}_{0,1}(\min|q_{0,1}^{11,\pi_{01}^{\max}})\right)$, respectively. As already discussed, the intuition for the optimization over different values of the instrument is that Z does not have a direct effect on the mean potential outcomes. Therefore, the set of admissible potential outcomes for $D=1$ is the intersection of possible values under $Z=0$ and $Z=1$.

Since the outcomes of the always takers are never observed under non-treatment, we have to rely on the upper and lower bounds in the support of Y, y^{UB} and y^{LB}. The sharp upper and lower bounds for the ATE on the always takers Δ_{11}, are:

$$\begin{aligned}
\Delta_{11}^{UB} &= \min\left(\bar{Y}_{1,1}(\max|q_{1,1}^{11,\pi_{01}^{\max}}), \bar{Y}_{0,1}(\max|q_{0,1}^{11,\pi_{01}^{\max}})\right) - y^{LB}, \\
\Delta_{11}^{LB} &= \max\left(\bar{Y}_{1,1}(\min|q_{1,1}^{11,\pi_{01}^{\max}}), \bar{Y}_{0,1}(\min|q_{0,1}^{11,\pi_{01}^{\max}})\right) - y^{UB}.
\end{aligned} \tag{9}$$

It is easy to see [10] that $\pi_{01}^{\max}$ maximizes the upper bound and minimizes the lower bound of Δ_{11} w.r.t. π_{01}, so that $q_{z,1}^{11,\pi_{01}^{\max}} = \max\left(0, \frac{P_{1|0}-P_{0|1}}{P_{1|z}}\right)$. Similarly as for the compliers and defiers, these bounds are only informative if $P_{1|0} > P_{0|1} \Rightarrow \pi_{11} > \pi_{00}$, i.e., if the share of always takers is larger than the share of never takers. The sampling process constraints the identification region of either the average treatment effect on the always takers or of the one on the never takers. Once again if $P_{1|0} < P_{0|1}$ then $\pi_{01}^{\max} = P_{1|0}$, which implies that $\pi_{11}^{\min} = 0$.

Using a symmetric argument as for the always takers, the sharp upper and lower bounds on the ATE of the never takers, Δ_{00}, are, respectively:

$$\begin{aligned}
\Delta_{00}^{UB} &= y^{UB} - \max\left(\bar{Y}_{1,0}(\min|q_{1,0}^{00,\pi_{01}^{\max}}), \bar{Y}_{0,0}(\min|q_{0,0}^{00,\pi_{01}^{\max}})\right), \\
\Delta_{00}^{LB} &= y^{LB} - \min\left(\bar{Y}_{1,0}(\max|q_{1,0}^{00,\pi_{01}^{\max}}), \bar{Y}_{0,0}(\max|q_{0,0}^{00,\pi_{01}^{\max}})\right).
\end{aligned} \tag{10}$$

$\pi_{01}^{\max}$ maximizes the upper bound and minimizes the lower bound of Δ_{00} w.r.t. π_{01}, such that $q_{z,0}^{00,\pi_{01}^{\max}} = \max\left(0, \frac{P_{0|1}-P_{1|0}}{P_{0|z}}\right)$. The bounds are informative if $P_{1|0} < P_{0|1}$, i.e., if there are more never takers than always takers in the population. Similar as before if $P_{1|0} > P_{0|1}$,

[10] Both $\bar{Y}_{1,1}(\max|q_{1,1}^{11})$ and $\bar{Y}_{0,1}(\max|q_{0,1}^{11})$ are increasing function of π_{01}. Similarly, both $\bar{Y}_{1,1}(\min|q_{1,1}^{11})$ and $\bar{Y}_{0,1}(\min|q_{0,1}^{11})$ are increasing functions of π_{01}.

$\pi_{01}^{\max} = P_{0|1}$ which implies that $\pi_{00}^{\min} = 0$.[11]

The identification results presented so far refer to latent strata defined by Z and D, populations that are not directly observed in the data. However, in the program evaluation literature, most attention seems to be devoted to the (observed) population receiving the treatment, see for instance Heckman, LaLonde and Smith (1999), which generally appears more policy relevant than latent groups. As a major contribution of this paper, we therefore also derive sharp bounds on the ATEs on the treated, as well as the non-treated and the entire population. The discussion below shows that for doing so, it suffices to establish the sharp bounds on $E(Y(1))$ and $E(Y(0))$, which for continuous outcomes are given by

$$\begin{aligned}
E(Y(1))^{UB} &= (P_{0|1} - \pi_{01}^{\min}) \cdot y^{UB} - (P_{1|0} - \pi_{01}^{\min}) \cdot \max\left(\bar{Y}_{1,1}(\min|q_{1,1}^{11,\pi_{01}^{\min}}), \bar{Y}_{0,1}(\min|q_{0,1}^{11,\pi_{01}^{\min}})\right) \\
&+ P_{1|0} \cdot \bar{Y}_{0,1} + P_{1|1} \cdot \bar{Y}_{1,1}, \qquad (11) \\
E(Y(1))^{LB} &= (P_{0|1} - \pi_{01}^{\min}) \cdot y^{LB} - (P_{1|0} - \pi_{01}^{\min}) \cdot \min\left(\bar{Y}_{1,1}(\max|q_{1,1}^{11,\pi_{01}^{\min}}), \bar{Y}_{0,1}(\max|q_{0,1}^{11,\pi_{01}^{\min}})\right) \\
&+ P_{1|0} \cdot \bar{Y}_{0,1} + P_{1|1} \cdot \bar{Y}_{1,1},
\end{aligned}$$

and

$$\begin{aligned}
E(Y(0))^{UB} &= (P_{1|0} - \pi_{01}^{\min}) \cdot y^{UB} - (P_{0|1} - \pi_{01}^{\min}) \cdot \max\left(\bar{Y}_{0,0}(\min|q_{0,0}^{00,\pi_{01}^{\min}}), \bar{Y}_{1,0}(\min|q_{1,0}^{00,\pi_{01}^{\min}})\right) \\
&+ P_{0|0} \cdot \bar{Y}_{0,0} + P_{0|1} \cdot \bar{Y}_{1,0}, \qquad (12) \\
E(Y(0))^{LB} &= (P_{1|0} - \pi_{01}^{\min}) \cdot y^{LB} - (P_{0|1} - \pi_{01}^{\min}) \cdot \min\left(\bar{Y}_{0,0}(\max|q_{0,0}^{00,\pi_{01}^{\min}}), \bar{Y}_{1,0}(\max|q_{1,0}^{00,\pi_{01}^{\min}})\right) \\
&+ P_{0|0} \cdot \bar{Y}_{0,0} + P_{0|1} \cdot \bar{Y}_{1,0},
\end{aligned}$$

respectively, see Appendix A.1.5. If Y is not continuous these bounds, and thus the one on the treated, non-treated, and the entire population, have to be optimized w.r.t. admissible defier proportions π_{01}, given by $\mathcal{P}^*$.

Considering the ATE on the treated, $\Delta_{D=1} = E(Y(1) - Y(0)|D = 1)$, note that because $E(Y(1)|D = 1) = E(Y|D = 1)$ is identified from the data, we only need to bound $E(Y(0)|D = 1)$. Solving $E(Y(0)) = \Pr(D = 1) \cdot E(Y(0)|D = 1) + \Pr(D = 0) \cdot E(Y(0)|D = 0)$ for $E(Y(0)|D = 1)$ gives $E(Y(0)|D = 1) = \frac{E(Y(0)) - \Pr(D=0) \cdot E(Y|D=0)}{\Pr(D=1)}$. Letting $E(Y(0))^{UB}$ and $E(Y(0))^{LB}$ denote the sharp upper and lower bounds for $E(Y(0))$, respectively, it therefore

[11] This demonstrates that it is generally not possible to have a value of π_{01} that jointly optimizes the bounds on the average treatment effects of all principal strata.

follows that the sharp upper and lower bounds on the ATE on the treated $\Delta_{D=1}$ are given by

$$\Delta^{UB}_{D=1} = E(Y|D=1) - \frac{E(Y(0))^{LB} - \Pr(D=0)\cdot E(Y|D=0)}{\Pr(D=1)},$$
$$\Delta^{LB}_{D=1} = E(Y|D=1) - \frac{E(Y(0))^{UB} - \Pr(D=0)\cdot E(Y|D=0)}{\Pr(D=1)}.$$

Since

$$\Pr(D=0)\cdot E(Y|D=0) = \Pr(Z=0)\cdot P_{0|0}\cdot \bar{Y}_{0,0} + \Pr(Z=1)\cdot P_{0|1}\cdot \bar{Y}_{1,0}$$

and

$$\begin{aligned}
P_{0|0}\cdot \bar{Y}_{0,0} + P_{0|1}\cdot \bar{Y}_{1,0} &= \Pr(Z=0)\cdot P_{0|0}\cdot \bar{Y}_{0,0} + \Pr(Z=1)\cdot P_{0|1}\cdot \bar{Y}_{1,0} \\
&+ \Pr(Z=1)\cdot P_{0|0}\cdot \bar{Y}_{0,0} + \Pr(Z=0)\cdot P_{0|1}\cdot \bar{Y}_{1,0}, \\
&= \Pr(D=0)\cdot E(Y|D=0) + \Pr(Z=1)\cdot P_{0|0}\cdot \bar{Y}_{0,0} + \Pr(Z=0)\cdot P_{0|1}\cdot \bar{Y}_{1,0},
\end{aligned}$$

the bounds on $\Delta_{D=1}$ correspond to

$$\begin{aligned}
\Delta^{UB}_{D=1} &= E(Y|D=1) - \frac{(P_{1|0} - \pi^{\min}_{01})\cdot y^{LB} + \Pr(Z=1)\cdot P_{0|0}\cdot \bar{Y}_{0,0} + \Pr(Z=0)\cdot P_{0|1}\cdot \bar{Y}_{1,0}}{\Pr(D=1)} \\
&+ \frac{(P_{0|1} - \pi^{\min}_{01})\cdot \min\left(\bar{Y}_{0,0}(\max|q^{00,\pi^{\min}_{01}}_{0,0}), \bar{Y}_{1,0}(\max|q^{00,\pi^{\min}_{01}}_{1,0})\right)}{\Pr(D=1)}, \qquad (13) \\
\Delta^{LB}_{D=1} &= E(Y|D=1) - \frac{(P_{1|0} - \pi^{\min}_{01})\cdot y^{UB} + \Pr(Z=1)\cdot P_{0|0}\cdot \bar{Y}_{0,0} + \Pr(Z=0)\cdot P_{0|1}\cdot \bar{Y}_{1,0}}{\Pr(D=1)} \\
&+ \frac{(P_{0|1} - \pi^{\min}_{01})\cdot \max\left(\bar{Y}_{0,0}(\min|q^{00,\pi^{\min}_{01}}_{0,0}), \bar{Y}_{1,0}(\min|q^{00,\pi^{\min}_{01}}_{1,0})\right)}{\Pr(D=1)}.
\end{aligned}$$

The bounds of the ATE on the non-treated, $\Delta_{D=0} = E(Y(1)-Y(0)|D=0)$, are obtained in a symmetric way. As $E(Y(0)|D=0) = E(Y|D=0)$ and $E(Y(1)|D=0) = \frac{E(Y(1)) - \Pr(D=1)\cdot E(Y|D=1)}{\Pr(D=0)}$, they are

$$\Delta^{UB}_{D=0} = \frac{E(Y(1))^{UB} - \Pr(D=1)\cdot E(Y|D=1)}{\Pr(D=0)} - E(Y|D=0),$$
$$\Delta^{LB}_{D=0} = \frac{E(Y(1))^{LB} - \Pr(D=1)\cdot E(Y|D=1)}{\Pr(D=0)} - E(Y|D=0).$$

Furthermore, since

$$\Pr(D=1)\cdot E(Y|D=1)=\Pr(Z=0)\cdot P_{1|0}\cdot\bar{Y}_{0,1}+\Pr(Z=1)\cdot P_{1|1}\cdot\bar{Y}_{1,1}$$

and

$$\begin{aligned}
P_{1|1}\cdot\bar{Y}_{1,1}+P_{1|0}\cdot\bar{Y}_{0,1} &= \Pr(Z=0)\cdot P_{1|0}\cdot\bar{Y}_{0,1}+\Pr(Z=1)\cdot P_{1|1}\cdot\bar{Y}_{1,1}\\
&+ \Pr(Z=1)\cdot P_{1|0}\cdot\bar{Y}_{0,1}+\Pr(Z=0)\cdot P_{1|1}\cdot\bar{Y}_{1,1}\\
&= \Pr(D=1)\cdot E(Y|D=1)+\Pr(Z=1)\cdot P_{1|0}\cdot\bar{Y}_{0,1}+\Pr(Z=0)\cdot P_{1|1}\cdot\bar{Y}_{1,1},
\end{aligned}$$

the bounds on $\Delta_{D=0}$ are given by

$$\begin{aligned}
\Delta^{UB}_{D=0} &= \frac{(P_{0|1}-\pi^{\min}_{01})\cdot y^{UB}+\Pr(Z=1)\cdot P_{1|0}\cdot\bar{Y}_{1,0}+\Pr(Z=0)\cdot P_{1|1}\cdot\bar{Y}_{1,1}}{\Pr(D=0)} \qquad (14)\\
&- \frac{(P_{1|0}-\pi^{\min}_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11,\pi^{\min}_{01}}_{1,1}),\bar{Y}_{0,1}(\min|q^{11,\pi^{\min}_{01}}_{0,1})\right)}{\Pr(D=0)}-E(Y|D=0),\\
\Delta^{LB}_{D=0} &= \frac{(P_{0|1}-\pi^{\min}_{01})\cdot y^{LB}+\Pr(Z=1)\cdot P_{1|0}\cdot\bar{Y}_{1,0}+\Pr(Z=0)\cdot P_{1|1}\cdot\bar{Y}_{1,1}}{\Pr(D=0)}\\
&- \frac{(P_{1|0}-\pi^{\min}_{01})\cdot\min\left(\bar{Y}_{1,1}(\max|q^{11,\pi^{\min}_{01}}_{1,1}),\bar{Y}_{0,1}(\max|q^{11,\pi^{\min}_{01}}_{0,1})\right)}{\Pr(D=0)}-E(Y|D=0).
\end{aligned}$$

Interestingly, the bounds on $\Delta_{D=1}$ and $\Delta_{D=0}$ are always informative despite the fact that either the bounds for the compliers or the defiers and either the bounds for the always takers or the never takers are not informative.

Finally, the bounds for the ATE on the entire population $\Delta=E(Y(1)-Y(0))$ are directly obtained from the bounds on $E(Y(1))$ and $E(Y(0))$ in (11) and (12):

$$\begin{aligned}
\Delta^{UB} &= E(Y(1)^{UB}-Y(0)^{LB})\\
&= P_{1|0}\cdot\bar{Y}_{0,1}-(P_{1|0}-\pi^{\min}_{01})\cdot y^{LB}+P_{1|1}\cdot\bar{Y}_{1,1} \qquad (15)\\
&- (P_{1|0}-\pi^{\min}_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11,\pi^{\min}_{01}}_{1,1}),\bar{Y}_{0,1}(\min|q^{11,\pi^{\min}_{01}}_{0,1})\right)\\
&+ (P_{0|1}-\pi^{\min}_{01})\cdot\min\left(\bar{Y}_{0,0}(\max|q^{00,\pi^{\min}_{01}}_{0,0}),\bar{Y}_{1,0}(\max|q^{00,\pi^{\min}_{01}}_{1,0})\right)\\
&- P_{0|1}\cdot\bar{Y}_{1,0}+(P_{0|1}-\pi^{\min}_{01})\cdot y^{UB}-P_{0|0}\cdot\bar{Y}_{0,0}\ ,
\end{aligned}$$

and

$$
\begin{aligned}
\Delta^{LB} &= E(Y(1)^{LB} - Y(0)^{UB}) \\
&= P_{1|0} \cdot \bar{Y}_{0,1} - (P_{1|0} - \pi_{01}^{\min}) \cdot y^{UB} + P_{1|1} \cdot \bar{Y}_{1,1} \qquad (16) \\
&- (P_{1|0} - \pi_{01}^{\min}) \cdot \min\left(\bar{Y}_{1,1}(\max|q_{1,1}^{11,\pi_{01}^{\min}}), \bar{Y}_{0,1}(\max|q_{0,1}^{11,\pi_{01}^{\min}})\right) \\
&+ (P_{0|1} - \pi_{01}^{\min}) \cdot \max\left(\bar{Y}_{0,0}(\min|q_{0,0}^{00,\pi_{01}^{\min}}), \bar{Y}_{1,0}(\min|q_{1,0}^{00,\pi_{01}^{\min}})\right) \\
&- P_{0|1} \cdot \bar{Y}_{1,0} + (P_{0|1} - \pi_{01}^{\min}) \cdot y^{LB} - P_{0|0} \cdot \bar{Y}_{0,0} \, .
\end{aligned}
$$

Again, these bounds are always informative no matter whether the bounds on the effect in some of the principal strata are not informative. In Appendix A.1.8 we show that it is possible to order the bounds on $\Delta_{D=1}$, $\Delta_{D=0}$, and Δ with respect to their tightness. In particular, if $\Delta^{UB}_{D=1} < \Delta^{UB}_{D=0}$ (or equivalently, $\Delta^{LB}_{D=1} > \Delta^{LB}_{D=0}$), the bounds on $\Delta_{D=1}$ are tighter than those on Δ, which are in turn tighter than those on $\Delta_{D=0}$. The order is reversed if $\Delta^{UB}_{D=1} > \Delta^{UB}_{D=0}$ (or equivalently $\Delta^{LB}_{D=1} < \Delta^{LB}_{D=0}$).

Furthermore, note that Δ^{LB}, Δ^{UB} might be narrower than the IV bounds derived by Manski (1990). The reason is that we assume mean independence within strata and unconfounded strata proportions (see Assumption 2), whereas Manski imposes the weaker mean independence of the potential outcomes in the entire population: $E(Y(d)|Z = 1) = E(Y(d)|Z = 0)$ for $d \in \{0, 1\}$. In contrast, our bounds may be wider than those of Kitagawa (2009), who invokes the stronger assumption of full independence of the instrument and the potential treatment states and outcomes. A formal comparison between the various bounds is given in Appendices A.1.6 and A.1.7.

Without imposing additional restrictions, the bounds derived in this section are likely to be very wide for most populations. Therefore, they are often not helpful for obtaining meaningful results in applications. For this reason we subsequently introduce further assumptions that appear plausible in many empirical problems and might entail considerably tighter bounds.

3.2 Monotonicity

This subsection shows how assuming monotonicity of the treatment in the instrument in addition to Assumption 1 increases identifying power. (Weak) monotonicity of D in Z implies that the treatment state under $Z = 1$ is at least as high as under $Z = 0$ for all individuals.

Assumption 3:

$\Pr(D(1) \geq D(0)) = 1$ (monotonicity).

As the potential treatment state never decreases in the instrument, the existence of the defiers (stratum 01) is ruled out. A symmetric result is obtained by assuming $\Pr(D(0) \geq D(1)) = 1$ which implies that stratum 10 does not exist. Note that assuming $\Pr(D(1) \geq D(0)) = 1$ (positive monotonicity) is only consistent with the data if $P_{1|1} - P_{1|0} > 0$, otherwise stratum 01 must necessarily exist. Similarly, $\Pr(D(0) \geq D(1)) = 1$ (negative monotonicity) requires that $P_{1|0} - P_{1|1} > 0$, see Table 3. Even though these are necessary conditions for the respective monotonicity assumption, they are not sufficient. Due to the symmetry of positive and negative monotonicity, we will only focus on Assumption 3 (positive monotonicity) in the subsequent discussion.

In their seminal paper on the identification of the local average treatment effect (LATE), Imbens and Angrist (1994) (see also Angrist, Imbens, and Rubin, 1996) show that Δ_{10} is point identified under Assumptions 1, 2, and 3. I.e., the bounds collapse to a single point given that π_{01} is equal to zero:

$$\begin{aligned}
\Delta_{10} &= \left(\frac{P_{1|1}}{P_{1|1}-P_{1|0}} \cdot \bar{Y}_{1,1} - \frac{P_{1|0}}{P_{1|1}-P_{1|0}} \cdot \bar{Y}_{0,1}\right) - \left(\frac{P_{0|0}}{P_{1|1}-P_{1|0}} \cdot \bar{Y}_{0,0} - \frac{P_{0|1}}{P_{1|1}-P_{1|0}} \cdot \bar{Y}_{1,0}\right) \\
&= \frac{(P_{1|1} \cdot \bar{Y}_{1,1} + P_{0|1} \cdot \bar{Y}_{1,0}) - (P_{1|0} \cdot \bar{Y}_{0,1} + P_{0|0} \cdot \bar{Y}_{0,0})}{P_{1|1}-P_{1|0}} \\
&= \frac{\Pr(D=1|Z=1) \cdot E(Y|Z=1,D=1) + \Pr(D=0|Z=1) \cdot E(Y|Z=1,D=0)}{\Pr(D=1|Z=1) - \Pr(D=1|Z=0)} \\
&- \frac{\Pr(D=1|Z=0) \cdot E(Y|Z=0,D=1) + \Pr(D=0|Z=0) \cdot E(Y|Z=0,D=0)}{\Pr(D=1|Z=1) - \Pr(D=1|Z=0)} \\
&= \frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)}. \qquad (17)
\end{aligned}$$

The last equality gives the well known result that the ATE on the compliers is just the ratio of two differences in conditional expectations, namely the intention to treat effect divided by the share of compliers. Under monotonicity, the observed subgroup with $Z = 0$ and $D = 1$ consists of always takers only and therefore, $\bar{Y}_{0,1}$ immediately gives the mean potential outcome under treatment for the always takers. Thus, an optimization of the kind $\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}), \bar{Y}_{0,1}\right)$ and $\min\left(\bar{Y}_{1,1}(\max|q^{11}_{1,1}), \bar{Y}_{0,1}\right)$ (with $\pi_{01} = 0$) as it was used for the bounds in section 3.1 is not required here. Note, however, that this comparison gives a testable implication for the identifying assumptions. If it is satisfied, $\bar{Y}_{1,1}(\min|q^{11}_{1,1}) \leq \bar{Y}_{0,1} \leq \bar{Y}_{1,1}(\max|q^{11}_{1,1})$, otherwise Z

has a direct effect on the outcomes of the always takers. Similarly, $\bar{Y}_{1,0}$ is the mean potential outcome under non-treatment for the never takers. Therefore, another testable implication is $\bar{Y}_{0,0}(\min|q^{00}_{0,0}) \leq \bar{Y}_{1,0} \leq \bar{Y}_{0,0}(\max|q^{00}_{0,0})$. We refer to Huber and Mellace (2013b) for a joint test of these implications.

In the absence of defiers, the bounds for the always takers and never takers (Δ_{11} and Δ_{00}) simplify to

$$\begin{aligned} \Delta_{11}^{UB} &= \bar{Y}_{0,1} - y^{LB}, \\ \Delta_{11}^{LB} &= \bar{Y}_{0,1} - y^{UB}, \end{aligned} \tag{18}$$

and

$$\begin{aligned} \Delta_{00}^{UB} &= y^{UB} - \bar{Y}_{1,0}, \\ \Delta_{00}^{LB} &= y^{LB} - \bar{Y}_{1,0}. \end{aligned} \tag{19}$$

These bound are sharp because $E(Y|D=1,T=11)$ and $E(Y|D=0,T=00)$ are now point identified by $\bar{Y}_{0,1}$ and $\bar{Y}_{1,0}$ (if mean independence within strata holds). However, monotonicity does not impose any restrictions on the distributions of $Y|D=0,T=11$ and $Y|D=1,T=00$ so that the worst case bounds y^{LB}, y^{UB} have to be assumed.

As in the last section, the bounds on the ATEs on the treated ($\Delta_{D=1}$), the non-treated ($\Delta_{D=0}$), and the entire population (Δ) can be expressed as functions of the bounds on $E(Y(1))$ and $E(Y(0))$. In the appendix we show that under monotonicity,

$$\begin{aligned} E(Y(1))^{UB} &= P_{0|1} \cdot y^{UB} + P_{1|1} \cdot \bar{Y}_{1,1}, \\ E(Y(1))^{LB} &= P_{0|1} \cdot y^{LB} + P_{1|1} \cdot \bar{Y}_{1,1}, \\ E(Y(0))^{UB} &= P_{1|0} \cdot y^{UB} + P_{0|0} \cdot \bar{Y}_{0,0}, \\ E(Y(0))^{LB} &= P_{1|0} \cdot y^{LB} + P_{0|0} \cdot \bar{Y}_{0,0}, \end{aligned}$$

so that the bounds on the various populations are given by

$$\begin{aligned} \Delta_{D=1}^{UB} &= E(Y|D=1) - \frac{P_{1|0} \cdot y^{LB} + \Pr(Z=1) \cdot (P_{0|0} \cdot \bar{Y}_{0,0} + P_{0|1} \cdot \bar{Y}_{1,0})}{\Pr(D=1)}, \\ \Delta_{D=1}^{LB} &= E(Y|D=1) - \frac{P_{1|0} \cdot y^{UB} + \Pr(Z=1) \cdot (P_{0|0} \cdot \bar{Y}_{0,0} + P_{0|1} \cdot \bar{Y}_{1,0})}{\Pr(D=1)}, \end{aligned} \tag{20}$$

$$\Delta^{UB}_{D=0} = \frac{P_{0|1}\cdot y^{UB} + \Pr(Z=0)\cdot(P_{1|0}\cdot \bar{Y}_{1,0} + P_{1|1}\cdot \bar{Y}_{1,1})}{\Pr(D=0)} - E(Y|D=0), \qquad (21)$$
$$\Delta^{LB}_{D=0} = \frac{P_{0|1}\cdot y^{LB} + \Pr(Z=0)\cdot(P_{1|0}\cdot \bar{Y}_{1,0} + P_{1|1}\cdot \bar{Y}_{1,1})}{\Pr(D=0)} - E(Y|D=0),$$

and

$$\Delta^{UB} = P_{0|1}\cdot y^{UB} + P_{1|1}\cdot \bar{Y}_{1,1} - P_{1|0}\cdot y^{LB} - P_{0|0}\cdot \bar{Y}_{0,0}, \qquad (22)$$
$$\Delta^{LB} = P_{0|1}\cdot y^{LB} + P_{1|1}\cdot \bar{Y}_{1,1} - P_{1|0}\cdot y^{UB} - P_{0|0}\cdot \bar{Y}_{0,0}.$$

Balke and Pearl (1997), Heckman and Vytlacil (2001), and Kitagawa (2009) show that under monotonicity, their bounds on the ATE in the entire population coincide with the bounds of Manski (1990), who only invokes mean independence in the entire population. I.e., Assumption 3 does, if it is satisfied, not bring any additional identifying power for Δ. Interestingly, this is also the case for our bounds on the ATEs on the entire population, the treated, and the non-treated. As all these bounds are optimized at $\pi_{01}^{\min} = \max(0, P_{1|0} - P_{1|1})$ (at least for continuous outcomes), it follows under a satisfaction of monotonicity that $P_{1|1} - P_{1|0} \geq 0$ and therefore, $\pi_{01}^{\min} = 0$. For this reason, imposing the monotonicity assumption, which amounts to setting $\pi_{01} = \pi_{01}^{\min}$, does not further tighten the bounds if a defier proportion of zero is already a priori consistent with the data.

3.3 Mean dominance

Mean dominance or the stronger stochastic dominance assumption have been used in the sample selection framework by Zhang and Rubin (2003), Lechner and Melly (2007), Blundell et al. (2007), Zhang et al. (2008), and Huber and Mellace (2013a). We will show that mean dominance also bears identifying power in the IV framework.

Assumption 4:

$E[Y(d)|T=10] \geq E[Y(d)|T=t] \quad \forall\ d \in \{0,1\},\ t \in \{11,00\}$ (mean dominance).

Assumption 4 states that the mean potential outcomes of the compliers under treatment and non-treatment are at least as high as those of the always and never takers. Note that the particular mean dominance assumption considered here is only one out of many possible relations between the potential outcomes of various strata. Its plausibility has to be judged in the light of the empirical application and theoretical considerations. In Section 5, we present an example where the compliers are likely to have weakly higher mean potential educational outcomes than both the always takers and the never takers due to plausibly being more able and/or motivated on average. As discussed in the next subsection, mean dominance has testable implications if it is jointly assumed with monotonicity. In the application presented in Section 5 we will test Assumption 4 and show that it is not rejected at any conventional significance level.

The bounds on the ATEs in the various principal strata as well as among the treated, non-treated, and entire populations are provided in the appendix. Under Assumptions 1, 2, and 4, Appendix A.3.2 outlines the moment inequalities that provide necessary and sufficient conditions for the defiers' proportion π_{01} being the identified set, denoted by $\mathcal{P}^{**}$. The construction of the latter based on linear linear programming is presented in Appendix A.3.1. Note that in general $\mathcal{P} \subseteq \mathcal{P}^{*} \subseteq \mathcal{P}^{**}$. Finally, we derive the bounds on the various ATEs as well as the mean potential outcomes in Appendix A.3.3.

3.4 Monotonicity and mean dominance

We subsequently derive the bounds under both monotonicity (Assumption 3) and mean dominance (Assumption 4). Since Δ_{10} is point identified under Assumptions 1 to 3, Assumption 4 does not bring any further improvement w.r.t. the compliers. For all other populations, the bounds become tighter when invoking both assumptions.

The upper and lower bounds of the ATE on the always takers are now

$$\begin{aligned} \Delta_{11}^{UB} &= \bar{Y}_{0,1} - y^{LB}, \\ \Delta_{11}^{LB} &= \bar{Y}_{0,1} - \left(\frac{P_{0|0}}{P_{1|1} - P_{1|0}} \cdot \bar{Y}_{0,0} - \frac{P_{0|1}}{P_{1|1} - P_{1|0}} \cdot \bar{Y}_{1,0} \right). \end{aligned} \tag{23}$$

As under mean dominance, the upper bound of the always takers' mean potential outcome under non-treatment cannot be higher than the compliers' upper bound under non-treatment. Furthermore, monotonicity implies that the latter is point identified by $\frac{P_{0|0}}{P_{1|1}-P_{1|0}} \cdot \bar{Y}_{0,0} - \frac{P_{0|1}}{P_{1|1}-P_{1|0}} \cdot \bar{Y}_{1,0}$. Again, Δ_{11}^{LB} is sharp by Lemma 1 in Appendix A.3. Similarly, the bounds for the never takers tighten to

$$
\begin{aligned}
\Delta_{00}^{UB} &= \left(\frac{P_{1|1}}{P_{1|1}-P_{1|0}}\cdot\bar{Y}_{1,1}-\frac{P_{1|0}}{P_{1|1}-P_{1|0}}\cdot\bar{Y}_{0,1}\right)-\bar{Y}_{1,0}, \qquad (24)\\
\Delta_{00}^{LB} &= y^{LB}-\bar{Y}_{1,0}.
\end{aligned}
$$

By the monotonicity assumption, $\frac{P_{1|1}}{P_{1|1}-P_{1|0}}\cdot\bar{Y}_{1,1}-\frac{P_{1|0}}{P_{1|1}-P_{1|0}}\cdot\bar{Y}_{0,1}$ is the compliers' mean potential outcome under treatment. Under mean dominance, this is an upper bound for the never takers' mean potential outcome under treatment.

Under both monotonicity and mean dominance, the upper bounds of $E(Y(0))$ and $E(Y(1))$ become

$$
\begin{aligned}
E(Y(1))^{UB} &= P_{0|1}\cdot\left(\frac{P_{1|1}\cdot\bar{Y}_{1,1}-P_{1|0}\cdot\bar{Y}_{0,1}}{P_{1|1}-P_{1|0}}\right)+P_{1|1}\cdot\bar{Y}_{1,1},\\
E(Y(0))^{UB} &= P_{1|0}\cdot\left(\frac{P_{0|0}\cdot\bar{Y}_{0,0}-P_{0|1}\cdot\bar{Y}_{1,0}}{P_{1|1}-P_{1|0}}\right)+P_{0|0}\cdot\bar{Y}_{0,0},
\end{aligned}
$$

while the lower bounds are equivalent to those under monotonicity alone.

Therefore, the bounds on the ATEs on the treated, non-treated, and the entire population are given by

$$
\begin{aligned}
\Delta_{D=1}^{UB} &= E(Y|D=1)-\frac{P_{1|0}\cdot y^{LB}+\Pr(Z=1)\cdot(P_{0|0}\cdot\bar{Y}_{0,0}+P_{0|1}\cdot\bar{Y}_{1,0})}{\Pr(D=1)}, \qquad (25)\\
\Delta_{D=1}^{LB} &= E(Y|D=1)-\frac{P_{1|0}\cdot\left(\frac{P_{0|0}\cdot\bar{Y}_{0,0}-P_{0|1}\cdot\bar{Y}_{1,0}}{P_{1|1}-P_{1|0}}\right)+\Pr(Z=1)\cdot(P_{0|0}\cdot\bar{Y}_{0,0}+P_{0|1}\cdot\bar{Y}_{1,0})}{\Pr(D=1)},
\end{aligned}
$$

$$
\begin{aligned}
\Delta_{D=0}^{UB} &= \frac{P_{0|1}\cdot\left(\frac{P_{1|1}\cdot\bar{Y}_{1,1}-P_{1|0}\cdot\bar{Y}_{0,1}}{P_{1|1}-P_{1|0}}\right)+\Pr(Z=0)\cdot(P_{1|0}\cdot\bar{Y}_{1,0}+P_{1|1}\cdot\bar{Y}_{1,1})}{\Pr(D=0)}-E(Y|D=0),\\
\Delta_{D=0}^{LB} &= \frac{P_{0|1}\cdot y^{LB}+\Pr(Z=0)\cdot(P_{1|0}\cdot\bar{Y}_{1,0}+P_{1|1}\cdot\bar{Y}_{1,1})}{\Pr(D=0)}-E(Y|D=0), \qquad (26)
\end{aligned}
$$

and

$$
\begin{aligned}
\Delta^{UB} &= P_{0|1}\cdot\left(\frac{P_{1|1}\cdot\bar{Y}_{1,1}-P_{1|0}\cdot\bar{Y}_{0,1}}{P_{1|1}-P_{1|0}}\right)+P_{1|1}\cdot\bar{Y}_{1,1}-P_{1|0}\cdot y^{LB}-P_{0|0}\cdot\bar{Y}_{0,0}, \qquad (27)\\
\Delta^{LB} &= P_{0|1}\cdot y^{LB}+P_{1|1}\cdot\bar{Y}_{1,1}-P_{1|0}\cdot\left(\frac{P_{0|0}\cdot\bar{Y}_{0,0}-P_{0|1}\cdot\bar{Y}_{1,0}}{P_{1|1}-P_{1|0}}\right)-P_{0|0}\cdot\bar{Y}_{0,0}.
\end{aligned}
$$

As a final remark it is worth noting that under Assumptions 1 to 3, Assumption 4 (mean dominance) is testable. Recall that the always takers' mean potential outcome is identified by $\bar{Y}_{0,1}$. Therefore, mean dominance of the compliers can be tested by comparing $\bar{Y}_{0,1}$ and $\bar{Y}_{1,1}$, which also encounters compliers and, therefore, has to dominate. Equivalently, $\bar{Y}_{1,0}$ is the never takers' mean potential outcome under non-treatment and must be dominated by $\bar{Y}_{0,0}$, which contains never takers and compliers. The intuition is that since the mean potential outcome of the always takers (never takers) is not affected by Z under mean independence within strata, the observed mean outcome consisting of both compliers and always takers (never takers) dominates the observed mean outcome of the always takers (never takers) only. The respective null hypotheses to be tested are $\bar{Y}_{1,1} \geq \bar{Y}_{0,1}$ and $\bar{Y}_{0,0} \geq \bar{Y}_{1,0}$. See Section 5 for an application of mean dominance tests.

4 Estimation

Under Assumptions 1 to 3 or 1 to 4, estimators of the bounds can be constructed by using the sample analogs of the bounds derived under the various assumptions, which is straightforward. To this end, we define the following sample parameters:

$$
\begin{aligned}
\hat{P}_{1|1} &= \frac{\sum_{i=1}^n D_i \cdot Z_i}{\sum_{i=1}^n Z_i}, \quad \hat{P}_{0|1} = 1 - \frac{\sum_{i=1}^n D_i \cdot Z_i}{\sum_{i=1}^n Z_i}, \quad \hat{P}_{1|0} = \frac{\sum_{i=1}^n D_i \cdot (1-Z_i)}{\sum_{i=1}^n (1-Z_i)}, \\
\hat{P}_{0|0} &= 1 - \frac{\sum_{i=1}^n D_i \cdot (1-Z_i)}{\sum_{i=1}^n (1-Z_i)}, \quad \hat{\bar{Y}}_{1,1} = \frac{\sum_{i=1}^n Y_i \cdot D_i \cdot Z_i}{\sum_{i=1}^n D_i \cdot Z_i}, \quad \hat{\bar{Y}}_{0,1} = \frac{\sum_{i=1}^n Y_i \cdot D_i \cdot (1-Z_i)}{\sum_{i=1}^n D_i \cdot (1-Z_i)}, \\
\hat{\bar{Y}}_{1,0} &= \frac{\sum_{i=1}^n Y_i \cdot (1-D_i) \cdot Z_i}{\sum_{i=1}^n (1-D_i) \cdot Z_i}, \quad \hat{\bar{Y}}_{0,0} = \frac{\sum_{i=1}^n Y_i \cdot (1-D_i) \cdot (1-Z_i)}{\sum_{i=1}^n (1-D_i) \cdot (1-Z_i)}, \\
&\hat{\bar{Y}}_{z,d}(\max|q_{z,d}^t) = \frac{\sum_{i=1}^n Y_i \cdot I\{D_i = d\} \cdot I\{Z_i = z\} \cdot I\{Y \geq \hat{y}_{1-q_{z,d}^t}\}}{\sum_{i=1}^n I\{D_i = d\} \cdot I\{Z_i = z\} \cdot I\{Y \geq \hat{y}_{1-q_{z,d}^t}\}}, \\
&\hat{\bar{Y}}_{z,d}(\min|q_{z,d}^t) = \frac{\sum_{i=1}^n Y_i \cdot I\{D_i = d\} \cdot I\{Z_i = z\} \cdot I\{Y \leq \hat{y}_{q_{z,d}^t}\}}{\sum_{i=1}^n I\{D_i = d\} \cdot I\{Z_i = z\} \cdot I\{Y \leq \hat{y}_{q_{z,d}^t}\}}, \\
\hat{y}_{q_{z,d}^t} &= \min\left\{y : \frac{\sum_{i=1}^n D_i \cdot Z_i \cdot I\{Y_i \leq y\}}{\sum_{i=1}^n D_i \cdot Z_i} \geq q_{z,d}^t\right\}, \quad \hat{y}^{LB} = \min(Y), \quad \hat{y}^{UB} = \max(Y)
\end{aligned}
$$

where $I\{\cdot\}$ is the indicator function. Using these expressions instead of the population parameters in the various formulas for the bounds immediately yields feasible estimators. $\sqrt{n}$-consistency and asymptotic normality of these estimators follow immediately from the results of Lee (2009) and its discussion is, therefore, omitted.

Under Assumptions 1 and 2 or 1, 2, and 4, however, estimation is non-standard due to the presence of min/max and sup/inf operators. For example, the upper bound on the compliers under Assumptions 1 and 2 is constructed in two steps. First, the sharp upper bound given π_{01} is obtained as the minimum of the four possible combinations of the pairs $\big(\bar{Y}_{1,1}(\min|q^{11}_{1,1}), \bar{Y}_{0,1}(\min|q^{11}_{0,1})\big)$ and $\big(\bar{Y}_{0,0}(\max|q^{00}_{0,0}), \bar{Y}_{1,0}(\max|q^{00}_{1,0})\big)$, which are both functions of Z. In the second step, the upper bound is derived by taking the sup of the bound over π_{01}.

More general, denote by $\Delta_t^{LB}(\pi_{01}, z, z')$, $\Delta_t^{UB}(\pi_{01}, z, z')$[12] the upper and lower bounds of any Δ_t conditional on π_{01}, $Z = z$ in the first min (max) operator and $Z = z'$ in the second one. To simplify the exposition, we define

$$v = \begin{cases} 1 & \text{if } z = 1, z' = 1 \\ 2 & \text{if } z = 1, z' = 0 \\ 3 & \text{if } z = 0, z' = 1 \\ 4 & \text{if } z = 0, z' = 0 \end{cases}.$$

This allows rewriting $\Delta_t^{LB}(\pi_{01}, z, z')$, $\Delta_t^{UB}(\pi_{01}, z, z')$ as $\Delta_t^{LB}(\pi_{01}, v)$, $\Delta_t^{UB}(\pi_{01}, v)$. Then, the identification region of Δ_t is obtained by optimizing over admissible values of $\pi_{01} \in \mathcal{P}^*$ and $v \in V = \{1, 2, 3, 4\}$:

$$\inf_{\pi_{01}\in\mathcal{P}^*}\{\max_{v\in V}[\Delta_t^{LB}(\pi_{01}, v)]\} \leq \Delta_t \leq \sup_{\pi_{01}\in\mathcal{P}^*}\{\min_{v\in V}[\Delta_t^{UB}(\pi_{01}, v)]\}.$$

Hirano and Porter (2012) show that for parameters that are non-differentiable functionals of the data (such as min/max and sup/inf operators), asymptotically unbiased estimators do not exist. Therefore, the sample analog estimators of $\inf_{\pi_{01}\in\mathcal{P}^*}\{\max_{v\in V}[\Delta_t^{LB}(\pi_{01}, v)]\}$ and $\sup_{\pi_{01}\in\mathcal{P}^*}\{\min_{v\in V}[\Delta_t^{UB}(\pi_{01}, v)]\}$ may suffer from substantial finite sample bias and standard asymptotics as well as the bootstrap are not consistent for the estimation of confidence intervals. However, note that the biases induced by optimizing over the defier proportion and the instrument go in opposite directions. Taking the supremum (infimum) of the upper (lower) bounds over π_{01} yields overly conservative inference, while optimizing over Z produces bounds and confidence intervals that are too tight.[13] For this reason, we ignore the first source of

[12] This section as well as in Appendix A.7 focuses on bounds that contain two min/max operators. If the bound contains only one min (max) operator, these expressions have to be replaced by $\Delta_t^{LB}(\pi_{01}, z)$, $\Delta_t^{UB}(\pi_{01}, z)$. Estimation is then analogous except that $v = z$ and can therefore only take two (rather than four) values.

[13] By optimizing over admissible defiers proportions $\mathcal{P}^*$ (or $\mathcal{P}^{**}$) we ignore the fact that $\mathcal{P}^*$ is unknown and

bias due to π_{01}, but account for the second one due to Z by applying the method proposed in Chernozhukov, Lee and Rosen (2013) and also used in Chen and Flores (2012). In this way we obtain conservative (and half-median-unbiased) point estimates and confidence intervals for the bounds. The method is described in Appendix A.7.

5 Application

We apply the methods outlined in the last sections to a school voucher experiment that was conducted within Colombia's "Programa de Ampliación de Cobertura de la Educación Secundaria" (PACES) in order to evaluate the program's impact on the educational achievement of various subpopulations. The PACES program targeted low income families in Colombia and provided more than 125,000 pupils with vouchers covering somewhat more than half the cost of private secondary schooling. Its goals were, among others, to increase net enrollment rates in secondary education and to raise quality compared to a public only educational system, see King, Rawlings, Gutierrez, Pardo and Torres (1997). We use a subsample of the data previously analyzed by Angrist et al. (2002) which consists of 1201 pupils in the capital Bogotá whose average age was 12 years when they had applied for private school vouchers in 1995. After randomly (not) being offered a voucher the applicants were re-interviewed in the second half of 1998 to measure the outcome variables of interest such as the highest grade completed and whether grades had to be repeated.

Table 4: Observed strata proportions

Conditional treatment probability	estimate	standard error
$P_{1\|1} = \Pr(D = 1\|Z = 1)$	0.561	(0.020)
$P_{0\|1} = \Pr(D = 0\|Z = 1)$	0.439	(0.020)
$P_{1\|0} = \Pr(D = 1\|Z = 0)$	0.056	(0.010)
$P_{0\|0} = \Pr(D = 0\|Z = 0)$	0.944	(0.010)

Let Z denote the random assignment indicator, Y a dummy for never repeating a grade or the highest grade completed, respectively, and D whether private schooling was actually received. As shown in Table 4, compliance with the school voucher assignment was not perfect. Only 56.1 % of the 629 pupils offered a school voucher actually went to private schools, while 43.9 % did not. 94.4 % of the 583 pupils that were randomized out did not receive private

needs to be estimated in practice. Developing a statistical inference procedure that would account for the sampling distribution of $\mathcal{P}^*$ is, however, beyond the scope of the present paper, which focuses on identification.

schooling, but 5.6 % attended private schools anyway. Table 5 reports the estimated bounds on the strata proportions without monotonicity and the respective point estimates under Assumption 3 (monotonicity). In our application, $\mathcal{P}$, the outer bounds on the proportion of defiers based on the distribution of (D, Z) (given by (6)) coincide with the sharp bounds $\mathcal{P}^*$ based on the distribution of (Y, D, Z) under mean independence within principal strata, which are obtained by linear programming as discussed in Appendix A.1.1. They also coincide with the sharp bounds $\mathcal{P}^{**}$ under the additional assumption of mean dominance, which are presented in A.3.1.

Table 5: Estimated (bounds on the) proportions of latent strata

Latent strata	Bounds without monotonicity	Proportions under monotonicity
Always takers	[0.000, 0.056]	0.056
Compliers	[0.505, 0.561]	0.505
Never takers	[0.383, 0.439]	0.439
Defiers	[0.000, 0.056]	-

We estimate bounds on the ATEs of the compliers, the always takers, the never takers, the treated, and the total population under mean independence within strata, mean dominance, and/or monotonicity. We do not consider defiers, because $\hat{P}_{1|1} > \hat{P}_{0|1}$, implies that the bounds for the defiers are not informative when only invoking mean independence within strata. Furthermore, defiers are ruled out under monotonicity (and under both monotonicity and mean dominance). Note that also the bounds for the always takers are not informative under mean independence within strata alone, because $\hat{P}_{1|0} < \hat{P}_{0|1}$ such that the share of always takers is smaller than the share of never takers. However, under monotonicity and/or mean dominance, informative bounds can be obtained for this stratum.

Whenever optimization over the defier proportion is required,[14] we use an equidistant grid of 100 values between the minimum (0) and maximum (0.056) possible shares. Under mean independence within strata and/or mean dominance (without monotonicity), we apply the Chernozhukov et al. (2013) procedure (see the last section) for estimation and inference (using a nominal significance level of 5%) using 5000 bootstraps and 200000 simulations.[15] Under

[14]This concerns the compliers and –due to the discreteness of the outcomes– also the treated and the entire population under Assumptions 1 and 2, and all populations considered under Assumptions 1,2, and 4.

[15]We are indebted to Xuan Chen and Carlos Flores for providing us with their Matlab code implementing the Chernozhukov et al. (2013) procedure and for their helpful advice about its implementation. As we have to estimate 100 variance-covariance matrices for each bound when optimizing over π_{01} (one for each value of the grid), some of them are close to being singular. To overcome this problem we use the Matlab function "mchol" by Brian Borchers (downloaded on Feb 06th 2013 from http://infohost.nmt.edu/˜borchers/ldlt.html) for regularization.

monotonicity (with and without mean dominance), which implies that standard asympotics apply to the bounds, we compute the 95% confidence intervals for the ATEs (rather than the bounds) based on the method described in Imbens and Manski (2004):

$$\left(\hat{\Delta}_t^{LB} - 1.645 \cdot \hat{\sigma}_t^{LB}, \hat{\Delta}_t^{UB} + 1.645 \cdot \hat{\sigma}_t^{UB}\right),$$

where $\hat{\Delta}_t^{LB}, \hat{\Delta}_t^{UB}$ are the estimated bounds in stratum t and $\hat{\sigma}_t^{LB}, \hat{\sigma}_t^{UB}$ denote their respective estimated standard errors,[16] obtained from 5000 bootstrap replications. Concerning the worst case bounds y^{UB} and y^{LB}, note that the binary outcome "never repeating a grade" is naturally bounded between 0 and 1. For the highest grade completed, we take the maximum and minimum values observed in the data, which are 11 and 5 years of schooling, respectively.

Table 6: ATE estimates on "never repeating a grade" and confidence intervals

Assumptions	Compliers	Always takers	Never takers	Treated	Entire pop.
Assumptions 1 and 2 only	[0.072, 0.208]	[-1.000, 1.000]	[-0.783, 0.379]	[0.044, 0.263]	[-0.261, 0.253]
	(0.000, 0.281)	Not informative	(-0.837, 0.426)	(-0.010, 0.333)	(-0.302, 0.286)
Mean dominance	[0.071, 0.207]	[-0.923, 0.966]	[-0.785, 0.350]	[0.045, 0.266]	[-0.261, 0.255]
	(0.005, 0.250)	(-1.000, 0.981)	(-0.811, 0.401)	(-0.009, 0.333)	(-0.302, 0.287)
Monotonicity	0.118	[-0.156, 0.844]	[-0.684, 0.316]	[0.070, 0.245]	[-0.249, 0.245]
	(0.032, 0.203)	(-0.263, 0.951)	(-0.730, 0.363)	(0.010, 0.313)	(-0.293, 0.277)
Both	0.118	[-0.011, 0.844]	[-0.684, 0.289]	[0.095, 0.245]	[-0.241, 0.234]
	(0.032, 0.203)	(-0.138, 0.951)	(-0.730, 0.340)	(0.025, 0.313)	(-0.287, 0.270)

Note: Bounds in square brackets and confidence intervals in round brackets. Confidence intervals are based on 5000 bootstraps. The number of simulations for the half-median-unbiased estimators is 200000.

Table 6 presents the results for the outcome "never repeating a grade" after the school voucher assignment under the various assumptions. The bounds of the ATE estimates are given in square brackets, the 95% confidence intervals are in round brackets. When only invoking Assumptions 1 and 2, the bounds are not informative for the always takers and quite wide for the never takers and the entire population. For the treated, the estimated interval is positive, but the lower bound is not significantly different from zero. For the compliers, the set is significantly positive (on a nominal level of 5%) and suggests that private schooling decreases the probability to repeat a class by 7 to 21 percentage points. This result suggests that mean independence within strata might have considerable identifying power in applications even when other restrictions such as monotonicity do not appear plausible.

Mean dominance slightly narrows the bounds for the always takers, which are now informative, but all in all, the gains in identification are if anything modest. In contrast, mono-

[16] The confidence intervals apply to cases where the distance between the upper and lower bound of the effect is bounded away from zero, see Stoye (2009). Under point identification (as for the compliers under monotonicity), the conventional two-sided confidence intervals are to be used: $\left(\hat{\Delta}_t - 1.96 \cdot \hat{\sigma}_t, \hat{\Delta}_t + 1.96 \cdot \hat{\sigma}_t\right)$, where $\hat{\Delta}_t$, $\hat{\sigma}_t$ denote the point estimate of the effect and the estimated standard error.

tonicity of D in Z (such that defiers are ruled out) entails point identification of the ATE on the compliers. The positive and significant estimate implies that grade repetition is reduced by roughly 12 percentage points when attending a private school. Also the identification region of the ATE on the treated, which (from a policy perspective) often represent the most interesting population, is now significantly positive. When invoking both monotonicity and mean dominance, the lower bound for the treated is tightened further. The identification regions for the always takers, never takers, and the entire population shrink somewhat, too, but still include the possibility of a zero effect.

Table 7 shows the estimates for the outcome "highest grade completed". Under Assumptions 1 and 2 alone, the estimated set for the ATE on the compliers positive and almost significant. It suggests that attending a private school increases the highest grade completed on average by 0.15 to 0.56 years for this population. Mean dominance does little to shrink the complier bounds. Under monotonicity, the point estimate suggests that schooling is on average raised by a third of a year. When invoking both assumptions, also the ATE on the treated (between 0.29 and 0.76 years) is significantly positive. Even the identification region for the always takers is larger than zero, but the effect is not significant at the 5 % level. All in all, our results support the conclusion of Angrist et al. (2002) that pupils going to private schools benefited from higher educational attainment. We find economically important positive effects on the likelihood not to repeat grades and on the highest grade completed among the compliers, but also among the treated population. The latter result is particularly relevant, because it suggests that the program increases the outcomes of those actually participating, a group that is most likely of more policy interest than the latent population of compliers.

Table 7: ATE estimates on "highest grade completed" and confidence intervals

Assumptions	Compliers	Always takers	Never takers	Treated	Entire pop.
Assumptions 1 and 2 only	[0.149, 0.562]	[-6.000, 6.000]	[-2.512, 4.019]	[-0.353, 0.760]	[-1.026, 1.984]
	(-0.008, 0.721)	Not informative	(-2.627, 4.248)	(-0.562, 0.952)	(-1.155, 2.143)
Mean dominance	[0.155, 0.549]	[-2.007, 3.012]	[-2.513, 1.019]	[0.037, 0.810]	[-0.866, 0.734]
	(-0.001, 0.640)	(-2.624, 3.037)	(-2.576, 1.121)	(-0.084, 0.988)	(-0.984, 0.806)
Monotonicity	0.326	[-3.188, 2.813]	[-2.251, 3.749]	[-0.287, 0.760]	[-1.002, 1.968]
	(0.126, 0.526)	(-3.457, 2.967)	(-2.362, 3.987)	(-0.506, 0.940)	(-1.133, 2.127)
Both	0.326	[0.115, 2.813]	[-2.251, 0.773]	[0.289, 0.760]	[-0.818, 0.661]
	(0.126, 0.526)	(-0.112, 2.967)	(-2.362, 0.889)	(0.122, 0.940)	(-0.943, 0.746)

Note: Bounds in square brackets and confidence intervals in round brackets. Confidence intervals are based on 5000 bootstraps. The number of simulations for the half-median-unbiased estimators is 200000.

As mentioned in Section 3.4, mean dominance of the compliers' potential outcomes has testable implications if monotonicity holds. We therefore bootstrap the sample analogs of $E[Y(1)|T = 10] - E[Y(1)|T = 11]$ and $E[Y(0)|T = 10] - E[Y(0)|T = 00]$ (with $E[Y(1)|T =$

$10] = \frac{P_{1|1} \cdot \bar{Y}_{1,1} - P_{1|0} \cdot \bar{Y}_{0,1}}{P_{1|1} - P_{1|0}}$, $E[Y(0)|T = 10] = \frac{P_{0|0} \cdot \bar{Y}_{0,0} - P_{0|1} \cdot \bar{Y}_{1,0}}{P_{0|0} - P_{0|1}}$, $E[Y(1)|T = 11] = \bar{Y}_{0,1}$, and $E[Y(0)|T = 00] = \bar{Y}_{1,0}$) to test whether the respective mean potential outcome of the compliers dominates that of the always takers under treatment and that of the never takers under non-treatment.[17] Table 8 reports the mean potential outcomes of the various populations and the p-values of the tests. The results strongly support the mean dominance of the compliers over the always takers under treatment and the mean dominance of the compliers over the never takers under non-treatment.

However, strictly speaking we also have to test whether the compliers dominate the always takers under non-treatment and the never takers under treatment, respectively. Even though this is infeasible (because always takers are never observed under non-treatment just as never takers under treatment), the mean potential outcomes provide indirect evidence that these assumptions are most likely satisfied. First of all, the hypothesis that the mean potential outcome of the compliers under non-treatment dominates the mean potential outcome of the always takers under treatment cannot be rejected for either outcome. I.e., if the ATE on the always takers is either positive or at least not negative by a sufficiently large amount, the mean potential outcome of the always takers under non-treatment cannot be larger than that of the compliers. Furthermore, the never takers can only have a higher mean potential outcome under treatment than the compliers if the ATE on the former is substantially larger than that on the latter (as the mean potential outcome of the never takers under non-treatment is considerably lower than that of the compliers for both outcome variables). In this case, however, it seems irrational of the never takers not to take the treatment such that this scenario appears unlikely.

Given the results of the tests, the question arises under which circumstances it seems plausible that the compliers' mean educational achievement dominates those of the always and never takers. Suppose that the private schooling decision is a function of (monetary and non-monetary) costs and utility coming from educational achievement. Economic theory suggests that rational households should send their children to private schools only if the

[17] Of course, this approach tests mean dominance conditional on the satisfaction of Assumptions 2 and 3 and is otherwise a joint test of all three assumptions. Huber and Mellace (2013b) suggest tests (i) for Assumptions 2 and 3 alone and (ii) (as also Kitagawa, 2013) for full independence of the instrument and potential treatments/outcomes and Assumption 3. For the outcome "never repeating a grade", using the method of Chen and Szroeter (2014) (with a normal smoothing function and $\sqrt{n/2\log(\log(n))}$ as tuning parameter for the selection of binding moments) to test (i) and (ii) (with two equidistant probability measures) yields p-values of 1.000 and 0.999, respectively. For "highest grade completed", the respective p-values are 0.993 and 0.851. Therefore, our data provide no evidence for a violation of (i) or (ii).

Table 8: Mean potential outcomes and mean dominance tests

	never repeating a grade	highest grade completed
$E[Y(1)\|T=10]$	0.973	8.024
$E[Y(0)\|T=10]$	0.855	7.698
$E[Y(1)\|T=11]$	0.844	7.813
$E[Y(0)\|T=00]$	0.684	7.251
p-value for $H_0 : E[Y(1)\|T=10] \geq E[Y(1)\|T=11]$	0.964	0.975
p-value for $H_0 : E[Y(0)\|T=10] \geq E[Y(0)\|T=00]$	0.997	0.998
p-value for $H_0 : E[Y(0)\|T=10] \geq E[Y(1)\|T=11]$	0.559	0.211

Note: p-values of mean dominance tests are based on 1999 bootstraps.

expected utility is at least as high as the costs. Always taker households may get a relatively higher utility from education, e.g., because the parents are themselves better educated and, therefore, appreciate education more than the compliers. Furthermore, they may represent the more wealthy households (as they send their children to private schools even without vouchers) such that their relative costs for schooling are lower. This might again be correlated with parental education. Both increased utility and lower relative costs will give relatively more pupils with lower potential outcomes –related to lower ability and/or motivation– the chance to receive private schooling. This line of argumentation is supported by the data, which also contain information on father's and mother's education and the possession of phone, which may be regarded as a proxy for wealth. The means of these variables (which were measured before the assignment) are higher among always takers than among compliers and the differences are significant at the 10 % level.[18]

In contrast, mean parental education and possessing a phone does not significantly differ between the never takers and compliers. Given that they face similar utilities (for a particular level of education) and relative costs as the compliers, it is plausible that the never taker households did not respond to the vouchers because their kids were probably less motivated and/or able and for this reason their expected returns to private schooling were too small. This suggests that the never takers' ATE (and the mean potential outcomes) is lower than those of the compliers, as never taker households were not even willing to pay less than half of the cost of private schooling (recall that the vouchers did not cover the entire expenses).

[18]The test statistics are available from the authors upon request.

6 Conclusion

This paper sheds light on the question of what can be learnt about the average treatment effects (ATE) on various populations under endogeneity/noncompliance when a valid instrumental variable (IV) is at hand that satisfies mean independence within strata and ignorable assignment. Since the work by Imbens and Angrist (1994) it is well known that a local ATE (LATE) on the compliers (who take the treatment if instrumented, but do not otherwise) is point identified under monotonicity of the treatment in the instrument. Even though point identification is not feasible for other groups, we show that informative bounds can be obtained for the always takers (treated irrespective of the instrument), the never takers (not treated irrespective of the instrument), the treated, the non-treated, and the entire population. We also investigate the identifying power of mean dominance of the potential outcomes of the compliers over those of the always takers and never takers.

The main contribution is the derivation of sharp bounds on the ATE of various populations under monotonicity, mean dominance, and under both assumptions. We also present an application to Colombia's "Programa de Ampliación de Cobertura de la Educación Secundaria", which provided pupils from low income families with vouchers for private secondary schooling, using experimental data previously analyzed by Angrist et al. (2002). We find (on top of the complier effect) a significantly positive ATE on the educational achievement of the treated population, a group of major policy interest. As valuable "by-products" of our identification results we also obtain testable implications of the validity of the instrument and of mean dominance, respectively.

A Appendix

A.1 Mean independence within principal strata without further assumptions

A.1.1 Identified set for the proportion of defiers based on linear programming

To shorten notation, let $f(y(1), y(0)|t, z)$ denote the probability density of the counterfactual outcomes $(Y(1), Y(0))$ evaluated at $(y(1), y(0))$ given $T = t$ and $Z = z$ and let $f_Y(y)$ denote the probability density function of the observed outcome Y evaluated at y. Assumption 2 together with compatibility with the probability distribution of observed variables (Y, D, Z) translates into the following restrictions.

Assumption 2 (i)

$$E(Y(d)|T = t, Z = 1) = E(Y(d)|T = t, Z = 0), \quad \forall t \in \{11, 10, 01, 00\}, \forall d \in \{1, 0\} \tag{A.1}$$

We rewrite (A.1) in terms of $f(y(1), y(0)|t, z)$:

$$\begin{aligned} \int y(1) f(y(1), y(0)|t, 1) dy(1) dy(0) &= \int y(1) f(y(1), y(0)|t, 0) dy(1) dy(0), \\ \int y(0) f(y(1), y(0)|t, 1) dy(1) dy(0) &= \int y(0) f(y(1), y(0)|t, 0) dy(1) dy(0). \end{aligned} \tag{A.2}$$

Assumption 2 (ii) together with compatibility with the distribution of (Y, D, Z)

For $D = 1$ and $Z = 1$, we have

$$\begin{aligned} f_Y(y|Z = 1, D = 1) \Pr(D = 1|Z = 1) &= f_{Y(1)}(y|11, 1) \Pr(T = 11|Z = 1) + f_{Y(1)}(y|10, 1) \Pr(T = 10|Z = 1) \\ &= f_{Y(1)}(y|11, 1) \Pr(T = 11) + f_{Y(1)}(y|10, 1) \Pr(T = 10), \end{aligned}$$

where $f_Y(y|Z = z, D = d)$ denotes the conditional density function of Y given $D = d$ and $Z = z$, $f_{Y(d)}(y|t, z)$ stands for the conditional density function of a potential outcome $Y(d)$ given $T = t$ and $Z = z$. The second equation follows from the Assumption 2 (ii). Similarly,

$$\begin{aligned} f_Y(y|Z = 0, D = 1) \Pr(D = 1|Z = 0) &= f_{Y(1)}(y|11, 0) \Pr(T = 11) + f_{Y(1)}(y|01, 0) \Pr(T = 01), \\ f_Y(y|Z = 1, D = 0) \Pr(D = 0|Z = 1) &= f_{Y(0)}(y|01, 1) \Pr(T = 01) + f_{Y(0)}(y|00, 1) \Pr(T = 00), \\ f_Y(y|Z = 0, D = 0) \Pr(D = 0|Z = 0) &= f_{Y(0)}(y|10, 0) \Pr(T = 10) + f_{Y(0)}(y|00, 0) \Pr(T = 00) \end{aligned}$$

This implies that the following equations must hold for all $y \in \mathcal{Y}$:

$$
\begin{aligned}
P_{1|1} \cdot f_Y(y|Z=1,D=1) &= (P_{1|0}-\pi_{01}) \int f(y,y(0)|11,1)\,dy(0) + (P_{1|1}-P_{1|0}+\pi_{01}) \int f(y,y(0)|10,1)\,dy(0), \\
P_{1|0} \cdot f_Y(y|Z=0,D=1) &= (P_{1|0}-\pi_{01}) \int f(y,y(0)|11,0)\,dy(0) + \pi_{01} \int f(y,y(0)|01,0)\,dy(0), \qquad \text{(A.3)} \\
P_{0|1} \cdot f_Y(y|Z=1,D=0) &= \pi_{01} \int f(y(1),y|01,1)\,dy(1) + (P_{0|1}-\pi_{01}) \int f(y(1),y|00,1)\,dy(1), \\
P_{0|0} \cdot f_Y(y|Z=0,D=0) &= (P_{1|1}-P_{1|0}+\pi_{01}) \int f(y(1),y|10,0)\,dy(1) + (P_{0|1}-\pi_{01}) \int f(y(1),y|00,0)\,dy(1).
\end{aligned}
$$

If there exists a proper density function $f(y(1),y(0)|t,z)$ for all $t \in \{11,10,01,00\}$, $z \in \{1,0\}$ which satisfies (A.2) and (A.3), then $\pi_{01} \in \mathcal{P}^*$, where $\mathcal{P}^*$ denotes the sharp identified set for the share of defiers.

With discrete Y, the problem is equivalent to the non-emptiness of the linearly constrained feasible set. Consider a discrete y with support $\mathcal{Y}^* = \{y_1, y_2, \dots y_k\}$ and let $h_t^z(y_i,y_j) = \Pr(Y(1)=y_i, Y(0)=y_j|T=t, Z=z)$. We denote the set of conditional distributions by $\{h_t^z\}$. Then equations (A.2) can be rewritten as

$$
\begin{aligned}
\sum_{i=1}^{k}\sum_{j=1}^{k} y_i h_t^1(y_i,y_j) &= \sum_{i=1}^{k}\sum_{j=1}^{k} y_i h_t^0(y_i,y_j), \qquad \text{(A.4)} \\
\sum_{i=1}^{k}\sum_{j=1}^{k} y_j h_t^1(y_i,y_j) &= \sum_{i=1}^{k}\sum_{j=1}^{k} y_j h_t^0(y_i,y_j).
\end{aligned}
$$

Similarly, equations (A.3) become

$$
\begin{aligned}
\Pr(Y=y_i, D=1|Z=1) &= (P_{1|0}-\pi_{01}) \sum_{j=1}^{k} h_{11}^1(y_i,y_j) + (P_{1|1}-P_{1|0}+\pi_{01}) \sum_{j=1}^{k} h_{10}^1(y_i,y_j), \ \forall i=1,\dots k, \\
\Pr(Y=y_i, D=1|Z=0) &= (P_{1|0}-\pi_{01}) \sum_{j=1}^{k} h_{10}^0(y_i,y_j) + \pi_{01} \sum_{j=1}^{k} h_{01}^0(y_i,y_j), \ \forall i=1,\dots k, \qquad \text{(A.5)} \\
\Pr(Y=y_j, D=0|Z=1) &= \pi_{01} \sum_{i=1}^{k} h_{01}^1(y_i,y_j) + (P_{0|1}-\pi_{01}) \sum_{i=1}^{k} h_{00}^1(y_i,y_j), \ \forall j=1,\dots k, \\
\Pr(Y=y_j, D=0|Z=0) &= (P_{1|1}-P_{1|0}+\pi_{01}) \sum_{i=1}^{k} h_{10}^0(y_i,y_j) + (P_{0|1}-\pi_{01}) \sum_{i=1}^{k} h_{00}^0(y_i,y_j), \ \forall j=1,\dots k.
\end{aligned}
$$

At the same time, h_t^z must be a proper probability distribution for all $t \in \{11,10,01,00\}$ and $z \in \{1,0\}$:

$$
\begin{aligned}
\sum_{i=1}^{k}\sum_{j=1}^{k} h_t^z(y_i,y_j) &= 1, \qquad \forall t \in \{11,10,01,00\},\ \forall z \in \{1,0\}, \qquad \text{(A.6)} \\
h_t^z(y_i,y_j) &\geq 0, \qquad \forall i=1,\dots k,\ \forall j=1,\dots k,\ \forall t \in \{11,10,01,00\},\ \forall z \in \{1,0\}.
\end{aligned}
$$

We conclude that the identified set for the share of defiers is a collection of points $\pi_{01} \in \mathcal{P}$

for which a feasible solution exists to a set of constraints given by (A.4)-(A.6) (formally $\mathcal{P}^* = \{\pi_{01} \in \mathcal{P} : \exists \ \{h_t^z\}$ that solves (A.4)-(A.6)$\}$), which can be checked using a linear programming tool.

Lemma 1 *The identified set $\mathcal{P}^*$ for π_{01} is an interval.*

Proof of Lemma 1. Consider the smallest and the largest admissible values $\pi_{01}^{\min}$ and $\pi_{01}^{\max}$ and the corresponding sets of conditional distributions $\{h_t^{z,\min}\}$ and $\{h_t^{z,\max}\}$ that satisfy $(A.4) - (A.6)$. For a fixed $0 < \lambda < 1$, let $\pi_{01}^{\lambda} = \lambda \pi_{01}^{\min} + (1-\lambda)\pi_{01}^{\max}$.

Consider $h_t^{z,\lambda}(y_i, y_j) = \theta^t h_t^{z,\min}(y_i, y_j) + (1-\theta^t) h_t^{z,\max}(y_i, y_j)$ for all $t \in \{11, 10, 01, 00\}$, $z \in \{1, 0\}$, and $i, j \in \{1, \cdots, k\}$, where

$$\begin{aligned}
\theta^{11} &= \frac{\lambda(P_{1|0} - \pi_{01}^{\min})}{\lambda(P_{1|0} - \pi_{01}^{\min}) + (1-\lambda)(P_{1|0} - \pi_{01}^{\max})}, \\
\theta^{10} &= \frac{\lambda(P_{1|1} - P_{1|0} + \pi_{01}^{\min})}{\lambda(P_{1|1} - P_{1|0} + \pi_{01}^{\min}) + (1-\lambda)(P_{1|1} - P_{1|0} + \pi_{01}^{\max})}, \\
\theta^{01} &= \frac{\lambda \pi_{01}^{\min}}{\lambda \pi_{01}^{\min} + (1-\lambda)\pi_{01}^{\max}}, \\
\theta^{00} &= \frac{\lambda(P_{0|1} - \pi_{01}^{\min})}{\lambda(P_{0|1} - \pi_{01}^{\min}) + (1-\lambda)(P_{0|1} - \pi_{01}^{\max})}.
\end{aligned}$$

Then, $\{h_t^{z,\lambda}\}$ solves (A.4) and (A.6) because these constraints do not depend on π_{01}. It is left to prove that the set of distributions $\{h_t^{z,\lambda}\}$ satisfies (A.5) for π_{01}^{λ}. We demonstrate it for the first equality in (A.5) and the validity of the remaining equalities follows similarly. We would like to show that $\forall i \in \{1, \ldots, k\}$:

$$\Pr(Y = y_i, D = 1 | Z = 1) = (P_{1|0} - \pi_{01}^{\lambda}) \sum_{j=1}^{k} h_{11}^{1,\lambda}(y_i, y_j) + (P_{1|1} - P_{1|0} + \pi_{01}^{\lambda}) \sum_{j=1}^{k} h_{10}^{1,\lambda}(y_i, y_j).$$

Consider the first term and the second term on the right hand side of this equation separately:

$$
\begin{aligned}
&(P_{1|0}-\pi_{01}^{\lambda})\sum_{j=1}^{k}h_{11}^{1,\lambda}(y_i,y_j)=\\
&\Big[P_{1|0}-(\lambda\pi_{01}^{\min}+(1-\lambda)\pi_{01}^{\max})\Big]\sum_{j=1}^{k}\Big[\theta^{11}h_{11}^{1,\min}(y_i,y_j)+(1-\theta^{11})h_{11}^{1,\max}(y_i,y_j)\Big]=\\
&=\lambda(P_{1|0}-\pi_{01}^{\min})\sum_{j=1}^{k}h_{11}^{1,\min}(y_i,y_j)+(1-\lambda)(P_{1|0}-\pi_{01}^{\max})\sum_{j=1}^{k}h_{11}^{1,\max}(y_i,y_j),\\
&(P_{1|1}-P_{1|0}+\pi_{01}^{\lambda})\sum_{j=1}^{k}h_{10}^{1,\lambda}(y_i,y_j)=\\
&=\Big[P_{1|1}-P_{1|0}+(\lambda\pi_{01}^{\min}+(1-\lambda)\pi_{01}^{\max})\Big]\sum_{j=1}^{k}\Big[\theta^{10}h_{10}^{1,\min}(y_i,y_j)+(1-\theta^{10})h_{10}^{1,\max}(y_i,y_j)\Big]=\\
&=\lambda(P_{1|1}-P_{1|0}+\pi_{01}^{\min})\sum_{j=1}^{k}h_{10}^{1,\min}(y_i,y_j)+(1-\lambda)(P_{1|1}-P_{1|0}+\pi_{01}^{\max})\sum_{j=1}^{k}h_{10}^{1,\max}(y_i,y_j).
\end{aligned}
$$

Summing and rearranging the two terms results in

$$
\begin{aligned}
&(P_{1|0}-\pi_{01}^{\lambda})\sum_{j=1}^{k}h_{11}^{1,\lambda}(y_i,y_j)+(P_{1|1}-P_{1|0}+\pi_{01}^{\lambda})\sum_{j=1}^{k}h_{10}^{1,\lambda}(y_i,y_j)=\\
&\lambda\left[(P_{1|0}-\pi_{01}^{\min})\sum_{j=1}^{k}h_{11}^{1,\min}(y_i,y_j)+(P_{1|1}-P_{1|0}+\pi_{01}^{\min})\sum_{j=1}^{k}h_{10}^{1,\min}(y_i,y_j)\right]+\\
&(1-\lambda)\left[(P_{1|0}-\pi_{01}^{\max})\sum_{j=1}^{k}h_{11}^{1,\max}(y_i,y_j)+(P_{1|1}-P_{1|0}+\pi_{01}^{\max})\sum_{j=1}^{k}h_{10}^{1,\max}(y_i,y_j)\right]=\\
&\lambda\cdot\Pr(Y=y_i,D=1|Z=1)+(1-\lambda)\cdot\Pr(Y=y_i,D=1|Z=1)=\Pr(Y=y_i,D=1|Z=1),
\end{aligned}
$$

where the last equality follows from the fact that $\{h_t^{z,\min}\}$ and $\{h_t^{z,\max}\}$ satisfy (A.5) for $\pi_{01}^{\min}$ and $\pi_{01}^{\max}$, respectively. We have therefore shown that $\{h_t^{z,\lambda}\}$ satisfies (A.4)-(A.6) for π_{01}^{λ} so that $\mathcal{P}^*=\left[\pi_{01}^{\min},\pi_{01}^{\max}\right]$. We note that the proof easily extends to the case with a continuous Y, where the sums would be replaced by integrals. ■

In a similar manner, one can obtain sharp bounds on various ATEs or potential outcomes using the linear programming tool whenever Y is discrete. Instead of searching for a *feasible* solution to (A.4)-(A.6), we would search for a solution that minimizes (or maximizes) a linear functional of a set of conditional distributions $\{h_t^z\}$ corresponding to the object of our interest (e.g. $\Delta_t,\Delta_{D=1},\Delta,E(Y(d))$).

A.1.2 Proof of the validity and the sharpness of the bounds on the ATEs within principal strata

We begin with a lemma that we will make use of in the proof.

Lemma 2 *Let W be a random variable that is distributed as a two components mixture:*

$$f(w) = p \cdot f_1(w) + (1-p) \cdot f_2(w) \quad p \in [0,1],$$

so that $W = pW_1 + (1-p)W_2$, where f_1, f_2 and f are probability density functions of W_1, W_2 and W respectively. Denote by $E(W_1)$ and $E(W_2)$ the expected values of the first and second component, respectively, and assume that $E(W_2) \geq E(W_1)$. Then, regardless of the identifying restrictions placed on the joint distribution of (W, W_1, W_2),

1. *$E(W)$ is the sharp upper bound for $E(W_1)$,*
2. *$E(W)$ is the sharp lower bound for $E(W_2)$.*

Proof of Lemma 2. First of all, we need to show that $E(W_1) \leq E(W) \leq E(W_2)$. To see this, note that

$$E(W) = p \cdot E(W_1) + (1-p) \cdot E(W_2). \tag{A.7}$$

This implies

$$E(W_2) = E(W) + (p \cdot E(W_2) - p \cdot E(W_1)).$$

Since $E(W_2) \geq E(W_1)$ by assumption, $p \cdot E(W_2) - p \cdot E(W_1)$ cannot be negative, which implies that $E(W) \leq E(W_2)$. In a symmetric manner one can show that $E(W_1) \leq E(W)$. Finally, we need to show that $E(W)$ is the respective sharp upper bound for $E(W_2)$ and the sharp lower bound for $E(W_1)$. We only demonstrate the latter part as the proof for the former is symmetric. Let ψ_2 be a generic member of the identification region of $E(W_1)$, denoted by $\Psi_2 : \psi_2 \geq E(W) \geq E(W_1)$. Clearly, $E(W) \in \Psi_2$. From (A.7) we have

$$E(W_1) \leq \frac{E(W) - (1-p) \cdot \psi_2}{p}.$$

Suppose there exists an element of Ψ_2, denoted by ψ_2^*, such that

$$E(W_1) \leq \frac{E(W) - (1-p) \cdot \psi_2^*}{p} \leq E(W). \tag{A.8}$$

From (A.8) it must hold that

$$\psi_2^* \leq E(W). \tag{A.9}$$

Since we have already shown that $E(W_2) \geq E(W)$, the only admissible element of Ψ_2 that satisfies (A.9) is $E(W)$, which shows that the latter is the sharp lower bound of $E(W_1)$. ■

The proof consists of two steps. In the first step, we show that the bounds on the ATEs are valid. Secondly, we show that these bounds are sharp.

Validity

We start by deriving the valid bounds on the mean potential outcomes within principal strata for π_{01} fixed, which contain the respective actual mean potential outcomes with probability one under the imposed assumptions. For the sake of brevity, we omit conditioning on π_{01} in the mean potential outcomes and write $E(Y(d)|T=t)$ rather than $E(Y(d)|T=t)(\pi_{01})$ (for $d=1,0,\ t=11,10,01,00$). Secondly, we use our bounds on the mean potential outcomes given π_{01} to construct bounds on $\Delta_t^{UB}(\pi_{01})$ and $\Delta_t^{LB}(\pi_{01})$, the ATEs within principal strata given π_{01}. Thirdly, since π_{01} is unknown, we obtain the upper and lower bounds on the ATEs within principal strata by optimizing over π_{01}: $\Delta_t^{UB} = \sup_{\pi_{01}\in\mathcal{P}^*} \Delta_t^{UB}(\pi_{01})$ and $\Delta_t^{LB} = \sup_{\pi_{01}\in\mathcal{P}^*} \Delta_t^{LB}(\pi_{01})$ (for $t=11,10,01,00$).

Assume, for the moment, that π_{01} is fixed (albeit omitted in the notation in the subsequent discussion). Lemma 1 of Imai (2008), which applies Proposition 4 in Horowitz and Manski (1995) to the case that the upper and lower bounds of the mixing probabilities of the principal strata are known, together with Imai's Proposition 1, which shows the sharpness of the bounds on the mean potential outcomes (and the ATE), implies that $E(Y(1)|Z=1,T=11) \le \bar{Y}_{1,1}(\max|q_{1,1}^{11})$ and $E(Y(1)|Z=0,T=11) \le \bar{Y}_{0,1}(\max|q_{0,1}^{11})$. Under Assumption 2, $E(Y(1)|Z=1,T=11) = E(Y(1)|Z=0,T=11) = E(Y(1)|T=11)$ and therefore,

$$E(Y(1)|T=11) \le E(Y(1)|T=11)^{UB} = \min\left(\bar{Y}_{1,1}(\max|q_{1,1}^{11}), \bar{Y}_{0,1}(\max|q_{0,1}^{11})\right).$$

In a symmetric way one can show that

$$\begin{aligned}
E(Y(1)|T=11) &\ge E(Y(1)|T=11)^{LB} = \max\left(\bar{Y}_{1,1}(\min|q_{1,1}^{11}), \bar{Y}_{0,1}(\min|q_{0,1}^{11})\right),\\
E(Y(0)|T=00) &\le E(Y(0)|T=00)^{UB} = \min\left(\bar{Y}_{1,0}(\max|q_{1,0}^{00}), \bar{Y}_{0,0}(\max|q_{0,0}^{00})\right),\\
E(Y(0)|T=00) &\ge E(Y(0)|T=00)^{LB} = \max\left(\bar{Y}_{1,0}(\min|q_{1,0}^{00}), \bar{Y}_{0,0}(\min|q_{0,0}^{00})\right).
\end{aligned}$$

Our assumptions do not provide any restrictions on $Y(0)|T=11$ and $Y(1)|T=00$, so that

$$E(Y(0)|T=11)^{LB} = y^{LB} \le E(Y(0)|T=11) \le E(Y(0)|T=11)^{UB} = y^{UB},$$

and

$$E(Y(1)|T=00)^{LB} = y^{LB} \le E(Y(1)|T=00) \le E(Y(1)|T=00)^{UB} = y^{UB}.$$

By combining Lemma 2 (the valididity part) with equations (1) to (4), we obtain bounds on the mean potential outcomes that are valid under our assumptions:

$$
\begin{aligned}
\frac{P_{1|1}\cdot\bar{Y}_{1,1}-(P_{1|0}-\pi_{01})\cdot E(Y(1)|T=11)^{UB}}{P_{1|1}-P_{1|0}+\pi_{01}} &\leq E(Y(1)|T=10)\leq \frac{P_{1|1}\cdot\bar{Y}_{1,1}-(P_{1|0}-\pi_{01})\cdot E(Y(1)|T=11)^{LB}}{P_{1|1}-P_{1|0}+\pi_{01}},\\
\frac{P_{0|0}\cdot\bar{Y}_{0,0}-(P_{0|1}-\pi_{01})\cdot E(Y(0)|T=00)^{UB}}{P_{1|1}-P_{1|0}+\pi_{01}} &\leq E(Y(0)|T=10)\leq \frac{P_{0|0}\cdot\bar{Y}_{0,0}-(P_{0|1}-\pi_{01})\cdot E(Y(0)|T=00)^{LB}}{P_{1|1}-P_{1|0}+\pi_{01}},\\
\frac{P_{1|0}\cdot\bar{Y}_{0,1}-(P_{1|0}-\pi_{01})\cdot E(Y(1)|T=11)^{UB}}{\pi_{01}} &\leq E(Y(1)|T=01)\leq \frac{P_{1|0}\cdot\bar{Y}_{0,1}-(P_{1|0}-\pi_{01})\cdot E(Y(1)|T=11)^{LB}}{\pi_{01}},\\
\frac{P_{0|1}\cdot\bar{Y}_{1,0}-(P_{0|1}-\pi_{01})\cdot E(Y(0)|T=00)^{UB}}{\pi_{01}} &\leq E(Y(0)|T=01)\leq \frac{P_{0|1}\cdot\bar{Y}_{1,0}-(P_{0|1}-\pi_{01})\cdot E(Y(0)|T=00)^{LB}}{\pi_{01}}.
\end{aligned}
$$

We use the following compact notation to refer to these bounds, $E(Y(d)|T=t)^{LB}\leq E(Y(d)|T=t)\leq E(Y(d)|T=t)^{UB}, d=1,0,t=10,01$.

The bounds on the ATE within some stratum t for π_{10} fixed are given by

$$
\begin{aligned}
\Delta_t^{UB}(\pi_{01}) &= [E(Y(1)|T=t)^{UB}-E(Y(0)|T=t)^{LB}],\\
\Delta_t^{LB}(\pi_{01}) &= [E(Y(1)|T=t)^{LB}-E(Y(0)|T=t)^{UB}],
\end{aligned}
$$

where conditioning on π_{10} is now made explicit in the bounds.

Finally, since π_{01} is unknown, the bounds on the ATEs within strata over all admissible π_{01} (which are shown to be sharp further below) are obtained by the following optimization:

$$
\begin{aligned}
\Delta_t^{UB} &= \sup_{\pi_{01}\in\mathcal{P}^*}\Delta_t^{UB}(\pi_{01})=\sup_{\pi_{01}\in\mathcal{P}^*}[E(Y(1)|T=t)^{UB}-E(Y(0)|T=t)^{LB}],\\
\Delta_t^{LB} &= \inf_{\pi_{01}\in\mathcal{P}^*}\Delta_t^{LB}(\pi_{01})=\inf_{\pi_{01}\in\mathcal{P}^*}[E(Y(1)|T=t)^{LB}-E(Y(0)|T=t)^{UB}],
\end{aligned}
$$

which for $t=10,01$ are equivalent to the bounds provided in the main text.

Since $\bar{Y}_{1,1}(\max|q^{11}_{1,1})$, $\bar{Y}_{0,1}(\max|q^{11}_{0,1})$, $\bar{Y}_{0,0}(\max|q^{00}_{0,0})$, and $\bar{Y}_{1,0}(\max|q^{00}_{1,0})$ are increasing in π_{01}, $\sup_{\pi_{01}\in\mathcal{P}^*}\Delta_{11}^{UB}(\pi_{01})=\Delta_{11}^{UB}(\pi_{01}^{\max})$ and $\sup_{\pi_{01}\in\mathcal{P}^*}\Delta_{00}^{UB}(\pi_{01})=\Delta_{00}^{UB}(\pi_{01}^{\max})$. Moreover, since $\bar{Y}_{1,1}(\min|q^{11}_{1,1})$, $\bar{Y}_{0,1}(\min|q^{11}_{0,1})$, $\bar{Y}_{0,0}(\min|q^{00}_{0,0})$, and $\bar{Y}_{1,0}(\min|q^{00}_{1,0})$ are decreasing in π_{01}, $\inf_{\pi_{01}\in\mathcal{P}^*}\Delta_{11}^{LB}(\pi_{01})=\Delta_{11}^{LB}(\pi_{01}^{\max})$ and $\inf_{\pi_{01}\in\mathcal{P}^*}\Delta_{00}^{LB}(\pi_{01})=\Delta_{00}^{LB}(\pi_{01}^{\max})$.

Sharpness

To demonstrate the sharpness of the proposed bounds, one needs to show that for each $E(Y(d)|T=t)\in[E(Y(d)|T=t)^{LB},E(Y(d)|T=t)^{UB}]$, $d=1,0$, $t=11,10,01,00$, there exist distributions of T given Z and of $(Y(1),Y(0))$ given T and Z that are compatible with a data generating process that satisfies Assumption 2. As $E(Y(d)|T=t)^{LB}$ and $E(Y(d)|T=t)^{UB}$ are the smallest and largest values of the interval $[E(Y(d)|T=t)^{LB},E(Y(d)|T=t)^{UB}]$, it is sufficient to prove the existence at those two extremes, because the values of $E(Y(d)|T=t)$

inside this interval can be achieved as convex combinations of the probability distributions of T given Z and the joint distributions of $(Y(1), Y(0))$ given T and Z that generate $E(Y(d)|T = t)^{LB}$ and $E(Y(d)|T = t)^{UB}$. We show that such compatible distributions can indeed be found so that $E(Y(1)|T = t) = E(Y(1)|T = t)^{UB}, t = 10, 01, 00$, $E(Y(0)|T = t) = E(Y(0)|T = t)^{UB}, t = 11, 10, 01$, $E(Y(1)|T = 11) = E(Y(1)|T = 11)^{LB}$ and $E(Y(0)|T = 00) = E(Y(0)|T = 00)^{LB}$ hold under Assumption 2. To this end, let $h_t^z(y(1), y(0)) = f(y(1), y(0)|T = t, Z = z)$ be the conditional density of $(Y(1), Y(0))$ evaluated at $(y(1), y(0))$ given $T = t$ and $Z = z$. For the sake of brevity, we refer to $h_t^z(y(1), y(0))$ by h_t^z. Consider the following distributions of T given Z and $(Y(1), Y(0))$ given T and Z:

$$\begin{aligned}
\Pr(T = 11|Z = 1) &= \Pr(T = 11|Z = 0) = P_{1|0} - \pi_{01}, \\
\Pr(T = 00|Z = 1) &= \Pr(T = 00|Z = 0) = P_{0|1} - \pi_{01}, \\
\Pr(T = 10|Z = 1) &= \Pr(T = 10|Z = 0) = P_{1|1} - P_{1|0} + \pi_{01}, \\
\Pr(T = 01|Z = 1) &= \Pr(T = 01|Z = 0) = \pi_{01},
\end{aligned} \tag{A.10}$$

where $\pi_{01} \in \mathcal{P}^*$. This probability distribution satisfies Assumption 2(ii), as it does not depend on Z.

We will now introduce four different specifications for $h_t^z, t = 11, 10, 01, 00, Z = 1, 0$, (henceforth denoted as $\{h_t^z\}$) that are (i) proper probability density functions, (ii) satisfy Assumption 2, (iii) are compatible with the data generating process, and (iv) reach

Case 1 the lower bound on the ATE of $t = 11$ and the upper bound on the ATE of $t = 00$,

Case 2 the upper bound on the ATE of $t = 11$ and the lower bound on the ATE of $t = 00$,

Case 3 the upper bound on the ATEs of $t = 10$ and $t = 01$,

Case 4 the lower bound on the ATEs of $t = 10$ and $t = 01$.

Case 1

Let the function $I\{A\}$ stand for the indicator function of a set A. Consider the following $\{h_t^z\}$:

$$h_{11}^1 = \begin{cases} I\{y(0)=y^{UB}\}\cdot f_Y(y(1)|D=1,Z=1,Y\le F_{Y_{1,1}}^{-1}(q_{1,1}^{11})) & \text{if } \bar{Y}_{1,1}(\min|q_{1,1}^{11})\ge \bar{Y}_{0,1}(\min|q_{0,1}^{11}) \\ I\{y(0)=y^{UB}\}\cdot\Big\{\alpha_{11}^1 f_Y(y(1)|D=1,Z=1,Y\le F_{Y_{1,1}}^{-1}(q_{1,1}^{11}))+ & \text{if } \bar{Y}_{1,1}(\min|q_{1,1}^{11})< \bar{Y}_{0,1}(\min|q_{0,1}^{11}) \\ \qquad +(1-\alpha_{11}^1)f_Y(y(1)|D=1,Z=1,Y\ge F_{Y_{1,1}}^{-1}(1-q_{1,1}^{11}))\Big\} & \end{cases},$$

$$h_{11}^0 = \begin{cases} I\{y(0)=y^{UB}\}\cdot\Big\{\alpha_{11}^0 f_Y(y(1)|D=1,Z=0,Y\le F_{Y_{0,1}}^{-1}(q_{0,1}^{11}))+ & \text{if } \bar{Y}_{1,1}(\min|q_{1,1}^{11})\ge \bar{Y}_{0,1}(\min|q_{0,1}^{11}) \\ \qquad +(1-\alpha_{11}^0)f_Y(y(1)|D=1,Z=0,Y\ge F_{Y_{0,1}}^{-1}(1-q_{0,1}^{11}))\Big\} & \\ I\{y(0)=y^{UB}\}\cdot f_Y(y(1)|D=1,Z=0,Y\le F_{Y_{0,1}}^{-1}(q_{0,1}^{11})) & \text{if } \bar{Y}_{1,1}(\min|q_{1,1}^{11})< \bar{Y}_{0,1}(\min|q_{0,1}^{11}) \end{cases},$$

$$h_{00}^1 = \begin{cases} I\{y(1)=y^{UB}\}\cdot f_Y(y(0)|D=0,Z=1,Y\le F_{Y_{1,0}}^{-1}(q_{1,0}^{00})) & \text{if } \bar{Y}_{1,0}(\min|q_{1,0}^{00})\ge \bar{Y}_{0,0}(\min|q_{0,0}^{00}) \\ I\{y(1)=y^{UB}\}\cdot\Big\{\alpha_{00}^1 f_Y(y(0)|D=0,Z=1,Y\le F_{Y_{1,0}}^{-1}(q_{1,0}^{00}))+ & \text{if } \bar{Y}_{1,0}(\min|q_{1,0}^{00})< \bar{Y}_{0,0}(\min|q_{0,0}^{00}) \\ \qquad +(1-\alpha_{00}^1)f_Y(y(0)|D=0,Z=1,Y\ge F_{Y_{1,0}}^{-1}(1-q_{1,0}^{00}))\Big\} & \end{cases},$$

$$h_{00}^0 = \begin{cases} I\{y(1)=y^{UB}\}\cdot\Big\{\alpha_{00}^0 f_Y(y(0)|D=0,Z=0,Y\le F_{Y_{0,0}}^{-1}(q_{0,0}^{00}))+ & \text{if } \bar{Y}_{1,0}(\min|q_{1,0}^{00})\ge \bar{Y}_{0,0}(\min|q_{0,0}^{00}) \\ \qquad +(1-\alpha_{00}^0)f_Y(y(0)|D=0,Z=0,Y\ge F_{Y_{0,0}}^{-1}(1-q_{0,0}^{00}))\Big\} & \\ I\{y(1)=y^{UB}\}\cdot f_Y(y(0)|D=0,Z=0,Y\le F_{Y_{0,0}}^{-1}(q_{0,0}^{00})) & \text{if } \bar{Y}_{1,0}(\min|q_{1,0}^{00})< \bar{Y}_{0,0}(\min|q_{0,0}^{00}) \end{cases},$$

$$h_{10}^z = \begin{cases} (P_{1|1}-P_{1|0}+\pi_{01})^{-2}\cdot\big(P_{1|1}\cdot f_Y(y(1)|D=1,Z=1)-(P_{1|0}-\pi_{01})\cdot\int h_{11}^1 dy(0)\big) & \text{if } \pi_{01}>P_{1|0}-P_{1|1} \\ \qquad\cdot\big(P_{0|0}\cdot f_Y(y(0)|D=0,Z=0)-(P_{0|1}-\pi_{01})\cdot\int h_{00}^0 dy(1)\big) & \\ \text{arbitrary probability density function, (because } \pi_{10}=0) & \text{if } \pi_{01}=P_{1|0}-P_{1|1}. \end{cases},$$

$$h_{01}^z = \begin{cases} \pi_{01}^{-2}\cdot\big(P_{1|0}\cdot f_Y(y(1)|D=1,Z=0)-(P_{1|0}-\pi_{01})\cdot\int h_{11}^0 dy(0)\big) & \text{if } \pi_{01}>0 \\ \qquad\cdot\big(P_{0|1}\cdot f_Y(y(0)|D=0,Z=1)-(P_{0|1}-\pi_{01})\cdot\int h_{00}^1 dy(1)\big) & \\ \text{arbitrary probability density function, (because } \pi_{01}=0) & \text{if } \pi_{01}=0. \end{cases}.$$

where we set π_{01} to $\pi_{01}^{\max}$ and the parameters

$$\begin{aligned} \alpha_{11}^1 &= \frac{\bar{Y}_{1,1}(\max|q_{1,1}^{11})-\bar{Y}_{0,1}(\min|q_{0,1}^{11})}{\bar{Y}_{1,1}(\max|q_{1,1}^{11})-\bar{Y}_{1,1}(\min|q_{1,1}^{11})}, \\ \alpha_{11}^0 &= \frac{\bar{Y}_{0,1}(\max|q_{0,1}^{11})-\bar{Y}_{1,1}(\min|q_{1,1}^{11})}{\bar{Y}_{0,1}(\max|q_{0,1}^{11})-\bar{Y}_{0,1}(\min|q_{0,1}^{11})}, \\ \alpha_{00}^1 &= \frac{\bar{Y}_{1,0}(\max|q_{1,0}^{00})-\bar{Y}_{0,0}(\min|q_{0,0}^{00})}{\bar{Y}_{1,0}(\max|q_{1,0}^{00})-\bar{Y}_{1,0}(\min|q_{1,0}^{00})}, \\ \alpha_{00}^0 &= \frac{\bar{Y}_{0,0}(\max|q_{0,0}^{00})-\bar{Y}_{1,0}(\min|q_{1,0}^{00})}{\bar{Y}_{0,0}(\max|q_{0,0}^{00})-\bar{Y}_{0,0}(\min|q_{0,0}^{00})}, \end{aligned}$$

are chosen so that the mean independence within principal strata holds.

(i) Proper probability functions. We have to check that these functions are proper probability density functions. For h_{11}^z and h_{00}^z it is sufficient to verify that $0\le\alpha_t^z\le 1$ for $T=11,00$ and $z=1,0$, because then all functions h_{11}^z and h_{00}^z are products of two marginal probability density functions that are convex combinations of two proper marginal probability density functions. We show this for

α^1_{11} and the same argument extends to the other parameters. The parameter α^1_{11} only occurs in h^1_{11} if $\bar{Y}_{1,1}(\min|q^{11}_{1,1}) < \bar{Y}_{0,1}(\min|q^{11}_{0,1})$.

Suppose that the denominator of α^1_{11} is positive so that $\bar{Y}_{1,1}(\max|q^{11}_{1,1}) > \bar{Y}_{1,1}(\min|q^{11}_{1,1})$. As for the non-negativity of α^1_{11}, it is therefore sufficient to verify that the numerator is nonnegative. Assume the contrary, so that $\bar{Y}_{1,1}(\max|q^{11}_{1,1}) < \bar{Y}_{0,1}(\min|q^{11}_{0,1})$. Now we have that

$$\bar{Y}_{1,1}(\min|q^{11}_{1,1}) \leq \bar{Y}_{1,1} \leq \bar{Y}_{1,1}(\max|q^{11}_{1,1}) < \bar{Y}_{0,1}(\min|q^{11}_{0,1}) \leq \bar{Y}_{0,1} \leq \bar{Y}_{0,1}(\max|q^{11}_{0,1}), \tag{A.11}$$

where the weak inequalities follow from the definition of $\bar{Y}_{z,d}(\min|q^t_{z,d})$ and $\bar{Y}_{z,d}(\max|q^t_{z,d})$.

Because $\pi^{\max}_{01}$ is an admissible value for π_{01}, we know that there exists some set of proper probability densities $\{h^{*z}_t\}$ that satisfy Assumptions 1 and 2 and are compatible with the data. Consider the value of $E[Y(1)|T = 11]$ that is implied by $\{h^{*z}_t\}$, so that $E[Y(1)|T = 11] = \iint y(1)h^{*1}_{11}\, dy(0)\, dy(1) = \iint y(1)h^{*0}_{11}\, dy(0)\, dy(1)$. We have proven the validity of the bounds for $E[Y(1)|T = 11]$, so it must hold that

$$\max\left\{\bar{Y}_{1,1}(\min|q^{11}_{1,1}), \bar{Y}_{0,1}(\min|q^{11}_{0,1})\right\} \leq E[Y(1)|T = 11] \leq \min\left\{\bar{Y}_{1,1}(\max|q^{11}_{1,1}), \bar{Y}_{0,1}(\max|q^{11}_{0,1})\right\}.$$

This results in

$$\bar{Y}_{0,1}(\min|q^{11}_{0,1}) \leq E[Y(1)|T = 11] \leq \bar{Y}_{1,1}(\max|q^{11}_{1,1}),$$

which contradicts (A.11). We therefore conclude that $\bar{Y}_{0,1}(\min|q^{11}_{0,1}) \leq \bar{Y}_{1,1}(\max|q^{11}_{1,1})$ and consequently $0 \leq \alpha^1_{11}$. As $\bar{Y}_{1,0}(\min|q^{11}_{1,0}) > \bar{Y}_{1,1}(\min|q^{11}_{1,1})$ we obtain $\alpha^1_{11} \leq 1$.

In the case that $\bar{Y}_{1,1}(\max|q^{11}_{1,1}) = \bar{Y}_{1,1}(\min|q^{11}_{1,1})$, we immediately get that the upper bound on $E[Y(1)|T = 11]$ is smaller than the lower bound and this contradicts the admissibility of π_{01}.

Let us now consider h^1_{10}. It is a product of two marginal probability densities. Both the first and the second parenthesis in the expression for h^1_{10} integrate to $(P_{1|1} - P_{1|0} + \pi_{01})$, so it remains to verify that both marginal probability density functions are positive.

Consider the first parenthesis, that is $\left(P_{1|1} \cdot f_Y(y(1)|D = 1, Z = 1) - (P_{1|0} - \pi_{01}) \cdot \int h^1_{11} dy(0)\right)$, the marginal density of h^1_{10} with respect to $y(1)$. Assume that $\bar{Y}_{1,1}(\min|q^{11}_{1,1}) \geq \bar{Y}_{0,1}(\min|q^{11}_{0,1})$. Because $q^{11}_{1,1} = \frac{P_{1|0}-\pi_{01}}{P_{1|1}}$, it follows that the parenthesis, $\Big(P_{1|1} \cdot f_Y(y(1)|D = 1, Z = 1) - (P_{1|0} - \pi_{01}) \cdot f_Y(y(1)|D = 1, Z = 1, Y \leq F^{-1}_{Y_{1,1}}(q^{11}_{1,1}))\Big)$, is zero for $y(1) \leq F^{-1}_{Y_{1,1}}(q^{11}_{1,1})$ and equal to $P_{1|1} \cdot f_Y(y(1)|D = 1, Z = 1) \geq 0$ elsewhere.

Now assume $\bar{Y}_{1,1}(\min|q^{11}_{1,1}) < \bar{Y}_{0,1}(\min|q^{11}_{0,1})$.

- If $q^{11}_{1,1} < 0.5$ so that $F^{-1}_{Y_{1,1}}(q^{11}_{1,1}) < F^{-1}_{Y_{1,1}}(1 - q^{11}_{1,1})$, we inspect three cases:
  (i) $y(1) < F^{-1}_{Y_{1,1}}(q^{11}_{1,1})$: the second component of the marginal distribution of h^1_{11} is zero and

the fact that $0 \leq \alpha_{11}^1 \leq 1$ guarantees the non-negativity of the parenthesis,

(ii) $F_{Y_{1,1}}^{-1}(q_{1,1}^{11}) < y(1) < F_{Y_{1,1}}^{-1}(1 - q_{1,1}^{11})$: the marginal density of h_{11}^1 is zero, so the parenthesis is equal to $P_{1|1} f_Y(y(1)|Z = 1, D = 1)$,

(iii) $y(1) > F_{Y_{1,1}}^{-1}(1 - q_{1,1}^{11})$: the first component of the marginal distribution of h_{11}^1 is zero and the fact that $0 \leq \alpha_{11}^1 \leq 1$ guarantees the non-negativity of the parenthesis.

- If $q_{1,1}^{11} > 0.5$ so that $F_{Y_{1,1}}^{-1}(q_{1,1}^{11}) > F_{Y_{1,1}}^{-1}(1 - q_{1,1}^{11})$, we also inspect three cases:

(i) $y(1) < F_{Y_{1,1}}^{-1}(1 - q_{1,1}^{11})$: the second component of the marginal distribution of h_{11}^1 is zero and the fact that $0 \leq \alpha_{11}^1 \leq 1$ guarantees the non-negativity of the parenthesis

(ii) $F_{Y_{1,1}}^{-1}(1 - q_{1,1}^{11}) < y(1) < F_{Y_{1,1}}^{-1}(q_{1,1}^{11})$: both components of the marginal distribution of h_{11}^1 are equal to $f_Y(y(1)|Z = 1, D = 1)/q_{1,1}^{11}$ and therefore the whole parenthesis becomes 0.

(iii) $y(1) > F_{Y_{1,1}}^{-1}(q_{1,1}^{11})$: the first component of the marginal distribution of h_{11}^1 is zero and the fact that $0 \leq \alpha_{11}^1 \leq 1$ guarantees the non-negativity of the parenthesis.

The non-negativity of the second marginal of h_{10}^1 follows in a similar fashion.

(ii) Mean independence within principal strata

We have to show that for a given type $T = t$, the marginal probability densities of $\{h_t^z\}$ produce the same expected values of the potential outcomes across $z = 1, 0$. Let us consider h_{11}^1 and h_{11}^0. Both probability density functions lead to $E[Y(0)|T = 11] = y^{UB}$. If $\bar{Y}_{1,1}(\min|q_{1,1}^{11}) \geq \bar{Y}_{0,1}(\min|q_{0,1}^{11})$, then

$$\iint y(1) h_{11}^1 \, dy(1) \, dy(0) = \bar{Y}_{1,1}(\min|q_{1,1}^{11}),$$

and

$$\begin{aligned}
\iint y(1) h_{11}^0 \, dy(1) \, dy(0) &= \iint y(1) \Big\{ \alpha_{11}^0 f_Y(y(1)|D = 1, Z = 0, Y \leq F_{Y_{0,1}}^{-1}(q_{0,1}^{11})) + \\
&\quad + (1 - \alpha_{11}^0) f_Y(y(1)|D = 1, Z = 0, Y \geq F_{Y_{0,1}}^{-1}(1 - q_{0,1}^{11})) \, dy(1) \, dy(0) \Big\} \\
&= \alpha_{11}^0 \bar{Y}_{0,1}(\min|q_{0,1}^{11}) + (1 - \alpha_{11}^0) \bar{Y}_{0,1}(\max|q_{0,1}^{11}) \\
&= \alpha_{11}^0 \left(\bar{Y}_{0,1}(\min|q_{0,1}^{11}) - \bar{Y}_{0,1}(\max|q_{0,1}^{11}) \right) + \bar{Y}_{0,1}(\max|q_{0,1}^{11}) \\
&= \bar{Y}_{1,1}(\min|q_{1,1}^{11}).
\end{aligned}$$

Now if $\bar{Y}_{1,1}(\min|q_{1,1}^{11}) < \bar{Y}_{0,1}(\min|q_{0,1}^{11})$, we have that

$$\iint y(1) h_{11}^0 \, dy(1) \, dy(0) = \bar{Y}_{0,1}(\min|q_{0,1}^{11}),$$

and

$$\begin{aligned}
\iint y(1) h_{11}^1 \, dy(1) \, dy(0) &= \iint y(1) \Big\{ \alpha_{11}^1 f_Y(y(1)|D=1, Z=1, Y \leq F_{Y_{1,1}}^{-1}(q_{1,1}^{11})) + \\
&\quad + (1-\alpha_{11}^1) f_Y(y(1)|D=1, Z=1, Y \geq F_{Y_{1,1}}^{-1}(1-q_{1,1}^{11})) \, dy(1) \, dy(0) \Big\} \\
&= \alpha_{11}^1 \bar{Y}_{1,1}(\min|q_{1,1}^{11}) + (1-\alpha_{11}^1) \bar{Y}_{1,1}(\max|q_{1,1}^{11}) \\
&= \alpha_{11}^1 \left(\bar{Y}_{1,1}(\min|q_{1,1}^{11}) - \bar{Y}_{1,1}(\max|q_{1,1}^{11}) \right) + \bar{Y}_{1,1}(\max|q_{1,1}^{11}) \\
&= \bar{Y}_{0,1}(\min|q_{0,1}^{11}).
\end{aligned}$$

Mean independence of potential outcomes within principal strata for $T = 11, 00$ together with the definition of $\{h_t^z\}$ in (A.11) guarantees that the mean independence within principal strata must also hold for $T = 10, 01$.

(iii) Compatibility with the data generating process. From the definition of $\{h_t^z\}$ (A.11), it follows that $\{h_t^z\}$ are compatible with the data generating process:

$$\begin{aligned}
P_{1|1} \cdot f_Y(y(1)|Z=1, D=1) &= (P_{1|0} - \pi_{01}) \int h_{11}^1 \, dy(0) + (P_{1|1} - P_{1|0} + \pi_{01}) \int h_{10}^1 \, dy(0), \\
P_{1|0} \cdot f_Y(y(1)|Z=0, D=1) &= (P_{1|0} - \pi_{01}) \int h_{11}^0 \, dy(0) + \pi_{01} \int h_{01}^0 \, dy(0), \qquad \text{(A.12)} \\
P_{0|1} \cdot f_Y(y(0)|Z=1, D=0) &= \pi_{01} \int h_{01}^1 \, dy(1) + (P_{0|1} - \pi_{01}) \int h_{00}^1 \, dy(1), \\
P_{0|0} \cdot f_Y(y(0)|Z=0, D=0) &= (P_{1|1} - P_{1|0} + \pi_{01}) \int h_{10}^0 \, dy(1) + (P_{0|1} - \pi_{01}) \int h_{00}^0 \, dy(1).
\end{aligned}$$

The fact that $\{h_t^z\}$ are proper probability density functions and compatible with the data generating process and that they satisfy Assumption 2 extends to Cases 2 – 4.

(iv) The upper bound on the ATE of $T = 11$ and the lower bound on the ATE of $T = 00$ are attained. We have already shown in part (ii) that

$$\begin{aligned}
E(Y(0)|T=11) &= \iint y(0) h_{11}^z \, dy(1) \, dy(0) = y^{UB}, \qquad \text{(A.13)} \\
E(Y(1)|T=00) &= \iint y(1) h_{00}^z \, dy(1) \, dy(0) = y^{UB}, \\
E(Y(1)|T=11) &= \iint y(1) h_{11}^z \, dy(1) \, dy(0) = \max\left(\bar{Y}_{0,1}(\min|q_{0,1}^{11}), \bar{Y}_{1,1}(\min|q_{1,1}^{11}) \right),
\end{aligned}$$

hold for $z = 1, 0$. We can show in an analogous way that the following equation holds:

$$E(Y(0)|T=00) = \iint y(0) h_{00}^z \, dy(1) \, dy(0) = \max\left(\bar{Y}_{0,0}(\min|q_{0,0}^{00}), \bar{Y}_{1,0}(\min|q_{1,0}^{00}) \right).$$

The first equality signs in each of the previous four lines in (A.13) and (A.12) follow from (A.11). The respective second equations in the first and second lines follow from the fact that h_{11} is zero at

$y(0) \neq y^{UB}$ and that h_{00} is zero at $y(1) \neq y^{UB}$. We recall the definition of $\bar{Y}_{z,d}(\min|q^t_{z,d}) = E(Y|Z = z, D = d, Y \leq F^{-1}_{Y_{z,d}}(q^t_{z,d}))$ that motivates the respective second equations in third and fourth lines. We now inspect the remaining mean potential outcomes within principal strata. Because of mean independence within principal strata we obtain for both $z = 1$ and $z = 0$:

$$\begin{aligned}
E(Y(1)|T=10) &= \iint y(1)h^z_{10}\,dy(1)\,dy(0) = \\
&= \iint y(1)\frac{1}{(P_{1|1}-P_{1|0}+\pi_{01})^2}\cdot\left(P_{1|1}\cdot f_Y(y(1)|D=1,Z=1)-(P_{1|0}-\pi_{01})\cdot\int h^z_{11}dy(0)\right)\cdot \\
&\cdot \left(P_{0|0}\cdot f_Y(y(0)|D=0,Z=0)-(P_{0|1}-\pi_{01})\cdot\int h^z_{00}dy(1)\right)\,dy(1)\,dy(0) = \\
&= \frac{1}{(P_{1|1}-P_{1|0}+\pi_{01})}\int y(1)\left(P_{1|1}\cdot f_Y(y(1)|D=1,Z=1)-(P_{1|0}-\pi_{01})\cdot\int h^z_{11}dy(0)\right)dy(1)\,\cdot \\
&\cdot \int\frac{\left(P_{0|0}\cdot f_Y(y(0)|D=0,Z=0)-(P_{0|1}-\pi_{01})\cdot\int h^z_{00}dy(1)\right)}{P_{1|1}-P_{1|0}+\pi_{01}}\,dy(0) = \\
&= \frac{P_{1|1}\cdot\bar{Y}_{1,1}-(P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{0,1}(\min|q^{11}_{0,1}),\bar{Y}_{1,1}(\min|q^{11}_{1,1})\right)}{P_{1|1}-P_{1|0}+\pi_{01}}.
\end{aligned}$$

The first equation follows from (A.11). In the second equation we plug in h^z_{10}. In the third equation we use the fact that the part in the first parenthesis in the definition of h^z_{10} does not depend on $y(0)$ and that the second part does not depend on $y(1)$. The last equation follows from (A.13) and from the fact that the second integral is equal to one. We obtain the following equations in an analogous way

$$\begin{aligned}
E(Y(0)|T=10) &= \iint y(0)h^z_{10}\,dy(1)\,dy(0) = \frac{P_{0|0}\cdot\bar{Y}_{0,0}-(P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)}{P_{1|1}-P_{1|0}+\pi_{01}}, \\
E(Y(1)|T=01) &= \iint y(1)h^z_{01}\,dy(1)\,dy(0) = \frac{P_{1|0}\cdot\bar{Y}_{0,1}-(P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{0,1}(\min|q^{11}_{0,1}),\bar{Y}_{1,1}(\min|q^{11}_{1,1})\right)}{\pi_{01}}, \\
E(Y(0)|T=01) &= \iint y(0)h^z_{01}\,dy(1)\,dy(0) = \frac{P_{0|1}\cdot\bar{Y}_{1,0}-(P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)}{\pi_{01}}.
\end{aligned}$$

This demonstrates that $E(Y(1)|T = t)^{UB}, t = 10, 01, 00$, $E(Y(0)|T = t)^{UB}, t = 11, 10, 01$, $E(Y(1)|T = 11)^{LB}$, and $E(Y(0)|T = 00)^{LB}$ are sharp bounds. Therefore, we have found distributions of the potential outcomes $(Y(1), Y(0)))$ given T, Z and $\Pr(T = t|Z = z)$, i.e. strata probabilities given Z (which are uniquely determined by the value of $\pi_{01} = \pi^{\max}_{01}$), that achieve the lower bound on the ATE for $t = 11$ and the upper bound on the ATE for $t = 00$. In Appendix A.1.3 we show that if $\mathcal{P}^*$ is non-empty, then $\pi^{\max}_{01} = \min\{P_{1|0}, P_{0|1}\}$ must be admissible.

Sharpness of the bounds in Case 2 can be proven in an analogous way. Cases 3 and 4 follow similarly, but with the important difference that the bounds are suprema (infima) of the bounds conditional on π_{01}. In these cases, we also show that the suprema (infima) of $\mathcal{P}^*$ are attained.

Case 2 By setting π_{01} to $\pi_{01}^{\max}$, the distribution of the types to (A.10), h_t for $t \in \{11, 10, 01, 00\}$ to

$$h_{11}^1 = \begin{cases} I\{y(0) = y^{LB}\} \cdot f_Y(y(1)|D=1, Z=1, Y \geq F_{Y_{1,1}}^{-1}(1-q_{1,1}^{11})) & \text{if } \bar{Y}_{1,1}(\max|q_{1,1}^{11}) \leq \bar{Y}_{0,1}(\max|q_{0,1}^{11}) \\ I\{y(0) = y^{LB}\} \cdot \Big\{\alpha_{11}^1 f_Y(y(1)|D=1, Z=1, Y \leq F_{Y_{1,1}}^{-1}(q_{1,1}^{11}))+ & \text{if } \bar{Y}_{1,1}(\max|q_{1,1}^{11}) > \bar{Y}_{0,1}(\max|q_{0,1}^{11}) \\ \qquad +(1-\alpha_{11}^1) f_Y(y(1)|D=1, Z=1, Y \geq F_{Y_{1,1}}^{-1}(1-q_{1,1}^{11}))\Big\} & \end{cases},$$

$$h_{11}^0 = \begin{cases} I\{y(0) = y^{LB}\} \cdot \Big\{\alpha_{11}^0 f_Y(y(1)|D=1, Z=0, Y \leq F_{Y_{0,1}}^{-1}(q_{0,1}^{11}))+ & \text{if } \bar{Y}_{1,1}(\max|q_{1,1}^{11}) \leq \bar{Y}_{0,1}(\max|q_{0,1}^{11}) \\ \qquad +(1-\alpha_{11}^0) f_Y(y(1)|D=1, Z=0, Y \geq F_{Y_{0,1}}^{-1}(1-q_{0,1}^{11}))\Big\} & \\ I\{y(0) = y^{LB}\} \cdot f_Y(y(1)|D=1, Z=0, Y \geq F_{Y_{0,1}}^{-1}(1-q_{0,1}^{11})) & \text{if } \bar{Y}_{1,1}(\max|q_{1,1}^{11}) > \bar{Y}_{0,1}(\max|q_{0,1}^{11}) \end{cases},$$

$$h_{00}^1 = \begin{cases} I\{y(1) = y^{LB}\} \cdot f_Y(y(0)|D=0, Z=1, Y \geq F_{Y_{1,0}}^{-1}(1-q_{1,0}^{00})) & \text{if } \bar{Y}_{1,0}(\max|q_{1,0}^{00}) \leq \bar{Y}_{0,0}(\max|q_{0,0}^{00}) \\ I\{y(1) = y^{LB}\} \cdot \Big\{\alpha_{00}^1 f_Y(y(0)|D=0, Z=1, Y \leq F_{Y_{1,0}}^{-1}(q_{1,0}^{00}))+ & \text{if } \bar{Y}_{1,0}(\max|q_{1,0}^{00}) > \bar{Y}_{0,0}(\max|q_{0,0}^{00}) \\ \qquad +(1-\alpha_{00}^1) f_Y(y(0)|D=0, Z=1, Y \geq F_{Y_{1,0}}^{-1}(1-q_{1,0}^{00}))\Big\} & \end{cases},$$

$$h_{00}^0 = \begin{cases} I\{y(1) = y^{LB}\} \cdot \Big\{\alpha_{00}^0 f_Y(y(0)|D=0, Z=0, Y \leq F_{Y_{0,0}}^{-1}(q_{0,0}^{00}))+ & \text{if } \bar{Y}_{1,0}(\max|q_{1,0}^{00}) \leq \bar{Y}_{0,0}(\max|q_{0,0}^{00}) \\ \qquad +(1-\alpha_{00}^0) f_Y(y(0)|D=0, Z=0, Y \geq F_{Y_{0,0}}^{-1}(1-q_{0,0}^{00}))\Big\} & \\ I\{y(1) = y^{LB}\} \cdot f_Y(y(0)|D=0, Z=0, Y \geq F_{Y_{0,0}}^{-1}(1-q_{0,0}^{00})) & \text{if } \bar{Y}_{1,0}(\max|q_{1,0}^{00}) > \bar{Y}_{0,0}(\max|q_{0,0}^{00}) \end{cases},$$

and h_t^z for $t = 10, 01$ and $z = 1, 0$ to values as defined in $(A.11)$, where

$$\begin{aligned} \alpha_{11}^1 &= \frac{\bar{Y}_{1,1}(\max|q_{1,1}^{11}) - \bar{Y}_{0,1}(\max|q_{0,1}^{11})}{\bar{Y}_{1,1}(\max|q_{1,1}^{11}) - \bar{Y}_{1,1}(\min|q_{1,1}^{11})}, \\ \alpha_{11}^0 &= \frac{\bar{Y}_{0,1}(\max|q_{0,1}^{11}) - \bar{Y}_{1,1}(\max|q_{1,1}^{11})}{\bar{Y}_{0,1}(\max|q_{0,1}^{11}) - \bar{Y}_{0,1}(\min|q_{0,1}^{11})}, \\ \alpha_{00}^1 &= \frac{\bar{Y}_{1,0}(\max|q_{1,0}^{00}) - \bar{Y}_{0,0}(\max|q_{0,0}^{00})}{\bar{Y}_{1,0}(\max|q_{1,0}^{00}) - \bar{Y}_{1,0}(\min|q_{1,0}^{00})}, \\ \alpha_{00}^0 &= \frac{\bar{Y}_{0,0}(\max|q_{0,0}^{00}) - \bar{Y}_{1,0}(\max|q_{1,0}^{00})}{\bar{Y}_{0,0}(\max|q_{0,0}^{00}) - \bar{Y}_{0,0}(\min|q_{0,0}^{00})}, \end{aligned}$$

one obtains the lower bound on the ATE for $t = 11$ and the upper bound on the ATE for $t = 00$.

Case 3

For a given value of π_{01}, h_t for $t = 11, 10, 01, 00$ is set to

$$
\begin{aligned}
h^1_{11} &= \begin{cases} I\{y(0) = y^{LB}\} \cdot f_Y(y(1)|D=1, Z=1, Y \geq F^{-1}_{Y_{1,1}}(1 - q^{11}_{1,1})) & \text{if } \bar{Y}_{1,1}(\max|q^{11}_{1,1}) \leq \bar{Y}_{0,1}(\max|q^{11}_{0,1}) \\ I\{y(0) = y^{LB}\} \cdot \Big\{\alpha^1_{11} f_Y(y(1)|D=1, Z=1, Y \leq F^{-1}_{Y_{1,1}}(q^{11}_{1,1})) + & \text{if } \bar{Y}_{1,1}(\max|q^{11}_{1,1}) > \bar{Y}_{0,1}(\max|q^{11}_{0,1}) \\ \qquad +(1-\alpha^1_{11}) f_Y(y(1)|D=1, Z=1, Y \geq F^{-1}_{Y_{1,1}}(1 - q^{11}_{1,1}))\Big\} & \end{cases}, \\
h^0_{11} &= \begin{cases} I\{y(0) = y^{LB}\} \cdot \Big\{\alpha^0_{11} f_Y(y(1)|D=1, Z=0, Y \leq F^{-1}_{Y_{0,1}}(q^{11}_{0,1})) + & \text{if } \bar{Y}_{1,1}(\max|q^{11}_{1,1}) \leq \bar{Y}_{0,1}(\max|q^{11}_{0,1}) \\ \qquad +(1-\alpha^0_{11}) f_Y(y(1)|D=1, Z=0, Y \geq F^{-1}_{Y_{0,1}}(1 - q^{11}_{0,1}))\Big\} & \\ I\{y(0) = y^{LB}\} \cdot f_Y(y(1)|D=1, Z=0, Y \geq F^{-1}_{Y_{0,1}}(1 - q^{11}_{0,1})) & \text{if } \bar{Y}_{1,1}(\max|q^{11}_{1,1}) > \bar{Y}_{0,1}(\max|q^{11}_{0,1}) \end{cases}, \\
h^1_{00} &= \begin{cases} I\{y(1) = y^{UB}\} \cdot f_Y(y(0)|D=0, Z=1, Y \leq F^{-1}_{Y_{1,0}}(q^{00}_{1,0})) & \text{if } \bar{Y}_{1,0}(\min|q^{00}_{1,0}) \geq \bar{Y}_{0,0}(\min|q^{00}_{0,0}) \\ I\{y(1) = y^{UB}\} \cdot \Big\{\alpha^1_{00} f_Y(y(0)|D=0, Z=1, Y \leq F^{-1}_{Y_{1,0}}(q^{00}_{1,0})) + & \text{if } \bar{Y}_{1,0}(\min|q^{00}_{1,0}) < \bar{Y}_{0,0}(\min|q^{00}_{0,0}) \\ \qquad +(1-\alpha^1_{00}) f_Y(y(0)|D=0, Z=1, Y \geq F^{-1}_{Y_{1,0}}(1 - q^{00}_{1,0}))\Big\} & \end{cases}, \\
h^0_{00} &= \begin{cases} I\{y(1) = y^{UB}\} \cdot \Big\{\alpha^0_{00} f_Y(y(0)|D=0, Z=0, Y \leq F^{-1}_{Y_{0,0}}(q^{00}_{0,0})) + & \text{if } \bar{Y}_{1,0}(\min|q^{00}_{1,0}) \geq \bar{Y}_{0,0}(\min|q^{00}_{0,0}) \\ \qquad +(1-\alpha^0_{00}) f_Y(y(0)|D=0, Z=0, Y \geq F^{-1}_{Y_{0,0}}(1 - q^{00}_{0,0}))\Big\} & \\ I\{y(1) = y^{UB}\} \cdot f_Y(y(0)|D=0, Z=0, Y \leq F^{-1}_{Y_{0,0}}(q^{00}_{0,0})) & \text{if } \bar{Y}_{1,0}(\min|q^{00}_{1,0}) < \bar{Y}_{0,0}(\min|q^{00}_{0,0}) \end{cases} \\
h^z_{10} &= \begin{cases} (P_{1|1} - P_{1|0} + \pi_{01})^{-2} \cdot \big(P_{1|1} \cdot f_Y(y(1)|D=1, Z=1) - (P_{1|0} - \pi_{01}) \cdot \int h^1_{11} dy(0)\big) & \text{if } \pi_{01} > P_{1|0} - P_{1|1} \\ \qquad \cdot \big(P_{0|0} \cdot f_Y(y(0)|D=0, Z=0) - (P_{0|1} - \pi_{01}) \cdot \int h^0_{00} dy(1)\big) & \\ I\{y(0) = y^{LB}\} \cdot I\{y(1) = y^{UB}\} & \text{if } \pi_{01} = P_{1|0} - P_{1|1}. \end{cases}, \\
h^z_{01} &= \begin{cases} \pi_{01}^{-2} \cdot \big(P_{1|0} \cdot f_Y(y(1)|D=1, Z=0) - (P_{1|0} - \pi_{01}) \cdot \int h^0_{11} dy(0)\big) & \text{if } \pi_{01} > 0 \\ \qquad \cdot \big(P_{0|1} \cdot f_Y(y(0)|D=0, Z=1) - (P_{0|1} - \pi_{01}) \cdot \int h^1_{00} dy(1)\big) & \\ I\{y(0) = y^{LB}\} \cdot I\{y(1) = y^{UB}\} & \text{if } \pi_{01} = 0. \end{cases},
\end{aligned} \tag{A.14}
$$

where

$$
\begin{aligned}
\alpha^1_{11} &= \frac{\bar{Y}_{1,1}(\max|q^{11}_{1,1}) - \bar{Y}_{0,1}(\max|q^{11}_{0,1})}{\bar{Y}_{1,1}(\max|q^{11}_{1,1}) - \bar{Y}_{1,1}(\min|q^{11}_{1,1})}, \\
\alpha^0_{11} &= \frac{\bar{Y}_{0,1}(\max|q^{11}_{0,1}) - \bar{Y}_{1,1}(\max|q^{11}_{1,1})}{\bar{Y}_{0,1}(\max|q^{11}_{0,1}) - \bar{Y}_{0,1}(\min|q^{11}_{0,1})}, \\
\alpha^1_{00} &= \frac{\bar{Y}_{1,0}(\max|q^{00}_{1,0}) - \bar{Y}_{0,0}(\min|q^{00}_{0,0})}{\bar{Y}_{1,0}(\max|q^{00}_{1,0}) - \bar{Y}_{1,0}(\min|q^{00}_{1,0})}, \\
\alpha^0_{00} &= \frac{\bar{Y}_{0,0}(\max|q^{00}_{0,0}) - \bar{Y}_{1,0}(\min|q^{00}_{1,0})}{\bar{Y}_{0,0}(\max|q^{00}_{0,0}) - \bar{Y}_{0,0}(\min|q^{00}_{0,0})},
\end{aligned}
$$

entails the upper bounds on the ATE for $t = 10$ if π_{01} maximizes (7) and the upper bound on the ATE for $t = 01$ if π_{01} maximizes (8). Note that because the bounds conditional on π_{01} are continuous in π_{01} (as shown in Appendix A.1.4) and $\mathcal{P}^*$ is an interval (as shown in Lemma 1), we get that the maxima of $\mathcal{P}^*$ are attained by the extreme value theorem.

Case 4

Consider the following conditional densities

$$
h_{11}^{1} = \begin{cases} I\{y(0)=y^{UB}\}\cdot f_Y(y(1)|D=1,Z=1,Y\leq F^{-1}_{Y_{1,1}}(q^{11}_{1,1})) & \text{if } \bar{Y}_{1,1}(\min|q^{11}_{1,1}) \geq \bar{Y}_{0,1}(\min|q^{11}_{0,1}) \\ I\{y(0)=y^{UB}\}\cdot \Big\{\alpha^{1}_{11} f_Y(y(1)|D=1,Z=1,Y\leq F^{-1}_{Y_{1,1}}(q^{11}_{1,1}))+ & \text{if } \bar{Y}_{1,1}(\min|q^{11}_{1,1}) < \bar{Y}_{0,1}(\min|q^{11}_{0,1}) \\ \qquad +(1-\alpha^{1}_{11}) f_Y(y(1)|D=1,Z=1,Y\geq F^{-1}_{Y_{1,1}}(1-q^{11}_{1,1}))\Big\} & \end{cases},
$$

$$
h_{11}^{0} = \begin{cases} I\{y(0)=y^{UB}\}\cdot \Big\{\alpha^{0}_{11} f_Y(y(1)|D=1,Z=0,Y\leq F^{-1}_{Y_{0,1}}(q^{11}_{0,1}))+ & \text{if } \bar{Y}_{1,1}(\min|q^{11}_{1,1}) \geq \bar{Y}_{0,1}(\min|q^{11}_{0,1}) \\ \qquad +(1-\alpha^{0}_{11}) f_Y(y(1)|D=1,Z=0,Y\geq F^{-1}_{Y_{0,1}}(1-q^{11}_{0,1}))\Big\} & \\ I\{y(0)=y^{UB}\}\cdot f_Y(y(1)|D=1,Z=0,Y\leq F^{-1}_{Y_{0,1}}(q^{11}_{0,1})) & \text{if } \bar{Y}_{1,1}(\min|q^{11}_{1,1}) < \bar{Y}_{0,1}(\min|q^{11}_{0,1}) \end{cases},
$$

$$
h_{00}^{1} = \begin{cases} I\{y(1)=y^{LB}\}\cdot f_Y(y(0)|D=0,Z=1,Y\geq F^{-1}_{Y_{1,0}}(1-q^{00}_{1,0})) & \text{if } \bar{Y}_{1,0}(\max|q^{00}_{1,0}) \leq \bar{Y}_{0,0}(\max|q^{00}_{0,0}) \\ I\{y(1)=y^{LB}\}\cdot \Big\{\alpha^{1}_{00} f_Y(y(0)|D=0,Z=1,Y\leq F^{-1}_{Y_{1,0}}(q^{00}_{1,0}))+ & \text{if } \bar{Y}_{1,0}(\max|q^{00}_{1,0}) > \bar{Y}_{0,0}(\max|q^{00}_{0,0}) \\ \qquad +(1-\alpha^{1}_{00}) f_Y(y(0)|D=0,Z=1,Y\geq F^{-1}_{Y_{1,0}}(1-q^{00}_{1,0}))\Big\} & \end{cases},
$$

$$
h_{00}^{0} = \begin{cases} I\{y(1)=y^{LB}\}\cdot \Big\{\alpha^{0}_{00} f_Y(y(0)|D=0,Z=0,Y\leq F^{-1}_{Y_{0,0}}(q^{00}_{0,0}))+ & \text{if } \bar{Y}_{1,0}(\max|q^{00}_{1,0}) \leq \bar{Y}_{0,0}(\max|q^{00}_{0,0}) \\ \qquad +(1-\alpha^{0}_{00}) f_Y(y(0)|D=0,Z=0,Y\geq F^{-1}_{Y_{0,0}}(1-q^{00}_{0,0}))\Big\} & \\ I\{y(1)=y^{LB}\}\cdot f_Y(y(0)|D=0,Z=0,Y\geq F^{-1}_{Y_{0,0}}(1-q^{00}_{0,0})) & \text{if } \bar{Y}_{1,0}(\max|q^{00}_{1,0}) > \bar{Y}_{0,0}(\max|q^{00}_{0,0}) \end{cases},
$$

$$
h_{10}^{z} = \begin{cases} (P_{1|1}-P_{1|0}+\pi_{01})^{-2}\cdot \big(P_{1|1}\cdot f_Y(y(1)|D=1,Z=1)-(P_{1|0}-\pi_{01})\cdot \int h^{1}_{11}dy(0)\big) & \text{if } \pi_{01} > P_{1|0}-P_{1|1} \\ \qquad \cdot \big(P_{0|0}\cdot f_Y(y(0)|D=0,Z=0)-(P_{0|1}-\pi_{01})\cdot \int h^{0}_{00}dy(1)\big) & \\ I\{y(0)=y^{UB}\}\cdot I\{y(1)=y^{LB}\} & \text{if } \pi_{01} = P_{1|0}-P_{1|1}. \end{cases},
$$

$$
h_{01}^{z} = \begin{cases} \pi_{01}^{-2}\cdot \big(P_{1|0}\cdot f_Y(y(1)|D=1,Z=0)-(P_{1|0}-\pi_{01})\cdot \int h^{0}_{11}dy(0)\big) & \text{if } \pi_{01} > 0 \\ \qquad \cdot \big(P_{0|1}\cdot f_Y(y(0)|D=0,Z=1)-(P_{0|1}-\pi_{01})\cdot \int h^{1}_{00}dy(1)\big) & \\ I\{y(0)=y^{UB}\}\cdot I\{y(1)=y^{LB}\} & \text{if } \pi_{01} = 0, \end{cases},
$$

where

$$
\begin{aligned}
\alpha^{1}_{11} &= \frac{\bar{Y}_{1,1}(\max|q^{11}_{1,1}) - \bar{Y}_{1,0}(\min|q^{11}_{1,0})}{\bar{Y}_{1,1}(\max|q^{11}_{1,1}) - \bar{Y}_{1,1}(\min|q^{11}_{1,1})}, \\
\alpha^{0}_{11} &= \frac{\bar{Y}_{0,1}(\max|q^{11}_{0,1}) - \bar{Y}_{1,1}(\min|q^{11}_{1,1})}{\bar{Y}_{0,1}(\max|q^{11}_{0,1}) - \bar{Y}_{0,1}(\min|q^{11}_{0,1})}, \\
\alpha^{1}_{00} &= \frac{\bar{Y}_{1,0}(\max|q^{00}_{1,0}) - \bar{Y}_{0,0}(\max|q^{00}_{0,0})}{\bar{Y}_{1,0}(\max|q^{00}_{1,0}) - \bar{Y}_{1,0}(\min|q^{00}_{1,0})}, \\
\alpha^{0}_{00} &= \frac{\bar{Y}_{0,0}(\max|q^{00}_{0,0}) - \bar{Y}_{1,0}(\max|q^{00}_{1,0})}{\bar{Y}_{0,0}(\max|q^{00}_{0,0}) - \bar{Y}_{0,0}(\min|q^{00}_{0,0})},
\end{aligned}
$$

entails the lower bound on the ATE for $t=10$ if π_{01} minimizes (7) and the lower bound on the ATE for $t=01$ if π_{01} minimizes (8). The bounds conditional on π_{01} are continuous in π_{01} (as shown in Appendix A.1.4) and $\mathcal{P}^*$ is an interval (as shown in Lemma 1), so that the minima of $\mathcal{P}^*$ are attained by the extreme value theorem. This ends the proof. ■

A.1.3 Identified set for the proportion of defiers based on moment inequalities

From the construction of $\{h_t^z\}$ in Appendix A.1.2 we note that, if $\mathcal{P}^*$ is non-empty, so that there exists at least one value of π_{01} that is admissible, then $\pi_{01}^{\max} = \min\{P_{1|0}, P_{0|1}\}$ must also be admissible. The reason is that $\bar{Y}_{0,1}(\min|q_{0,1}^{11}), \bar{Y}_{0,1}(\min|q_{1,1}^{11}), \bar{Y}_{1,0}(\min|q_{1,0}^{00})$ and $\bar{Y}_{0,0}(\min|q_{0,0}^{00})$ are non-increasing in π_{01} and $\bar{Y}_{1,1}(\max|q_{1,1}^{11}), \bar{Y}_{0,1}(\max|q_{0,1}^{11}), \bar{Y}_{0,0}(\max|q_{0,0}^{00})$ and $\bar{Y}_{1,0}(\max|q_{1,0}^{00})$ are non-decreasing in π_{01}. Therefore if the following four inequalities hold for a particular value of π_{01}^*,

$$\begin{aligned}
\bar{Y}_{0,1}(\min|q_{0,1}^{11}) &\leq \bar{Y}_{1,1}(\max|q_{1,1}^{11}) \\
\bar{Y}_{1,1}(\min|q_{1,1}^{11}) &\leq \bar{Y}_{0,1}(\max|q_{0,1}^{11}) \\
\bar{Y}_{1,0}(\min|q_{1,0}^{00}) &\leq \bar{Y}_{0,0}(\max|q_{0,0}^{00}) \qquad \text{(A.15)} \\
\bar{Y}_{0,0}(\min|q_{0,0}^{00}) &\leq \bar{Y}_{1,0}(\max|q_{1,0}^{00})
\end{aligned}$$

then they must also hold for any $\pi_{01} > \pi_{01}^*$.

The inequalities (A.15) guarantee that $0 \leq \alpha_t^z \leq 1$ for $t = 11, 00$ and $Z = 0, 1$ and hence the possibility of constructing $\{h_t^z\}$ as outlined in (A.11). This makes any $\pi_{01} \leq \pi_{01}^{\max}$ admissible and therefore (A.15) provides a **sufficient condition** for $\pi_{01} \in \mathcal{P}^*$. But at the same time, for an admissible value of π_{01}, inequalities (A.15) must hold so that the upper bounds of $E[Y(1)|T = 11]$ and $E[Y(0)|T = 00]$ are larger than the lower bounds, such that (A.15) is a **necessary condition** for $\pi_{01} \in \mathcal{P}^*$. This shows that, given that $\pi_{01} \in \mathcal{P}$, (A.15) is a necessary and sufficient condition for the admissibility of π_{01} under Assumptions 1 and 2, that is $\pi_{01} \in \mathcal{P}^*$.

A.1.4 Continuity of bounds on the ATE for compliers and defiers in the share of defiers

Here we prove the continuity of the upper bound on the ATE for compliers conditional on π_{01}. The continuity of the other bounds on the ATE for other principal strata follows similarly. We make the dependence of quantiles on the value of π_{01} explicit. The quantiles $q_{z,d}^t$ are continuous functions in π_{01}, the trimmed means are continuous in quantiles and therefore it is sufficient to verify cases when the denominator in the analytic expression $E[Y(1)|T = 10]$ for is zero. Therefore we have to show that

$$\lim_{\pi_{01} \to (P_{1|0} - P_{1|1})^+} E[Y(1)|T = 10](\pi_{01}) = y^{UB},$$

for any admissible value of π_{01}.

Note that $q^{11}_{1,1}(P_{1|0}-P_{1|1})=1$ and $q^{11}_{0,1}(P_{1|0}-P_{1|1})=\frac{P_{1|1}}{P_{1|0}}$. If $\bar{Y}_{1,1}(\min|1)>\bar{Y}_{0,1}(\min|\frac{P_{1|1}}{P_{1|0}})$

$$\begin{aligned}
&\lim_{\pi_{01}\to(P_{1|0}-P_{1|1})^+}\frac{P_{1|1}\cdot\bar{Y}_{1,1}-(P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}(\pi_{01})),\bar{Y}_{0,1}(\min|q^{11}_{0,1}(\pi_{01}))\right)}{P_{1|1}-P_{1|0}+\pi_{01}}\\
=&\lim_{\pi_{01}\to(P_{1|0}-P_{1|1})^+}\frac{\bar{Y}_{1,1}(\min|q^{11}_{1,1}(\pi_{01}))-(P_{1|0}-\pi_{01})\frac{\partial\bar{Y}_{1,1}(\min|q^{11}_{1,1})}{\partial q^{11}_{1,1}}\frac{\partial q^{11}_{1,1}}{\partial\pi_{01}}}{1}\\
=&\ \bar{Y}_{1,1}-(P_{1|1})\left(\bar{Y}_{1,1}-y^{UB}\right)\left(-\frac{1}{P_{1|1}}\right)=y^{UB},
\end{aligned}$$

where the first equation follows from the l'Hospital rule and in the second we used the definition of the trimmed mean. If $\bar{Y}_{1,1}(\min|1)<\bar{Y}_{0,1}(\min|\frac{P_{1|1}}{P_{1|0}})$, then we get that $\bar{Y}_{1,1}(\min|1)=\bar{Y}_{1,1}(\max|1)<\bar{Y}_{0,1}(\min|\frac{P_{1|1}}{P_{1|0}})$, which contradicts the admissibility of $\pi_{01}=P_{1|0}-P_{1|1}$, because then the upper bound for $E[Y(1)|T=11]$ is smaller than its' lower bound.

Using the same argument we get that

$$\lim_{\pi_{01}\to(P_{1|0}-P_{1|1})^+}E[Y(0)|T=10](\pi_{01})=y^{LB},$$

and this leads to

$$\lim_{\pi_{01}\to(P_{1|0}-P_{1|1})^+}\Delta^{UB}_{10}(\pi_{01})=y^{UB}-y^{LB}.$$

The intuitive argument is that we could have chosen the parametrization differently. Instead of expressing the bounds in π_{01}, we could have expressed the bounds in π_{11} or π_{00} and in such case, we would not get zero in the denominator for the analytic expression for the bounds on compliers but for alwaystakers or nevertakers.

A.1.5 Proof of the validity and the sharpness of the bounds on the mean potential outcomes $E(Y(1))$, $E(Y(0))$

We only demonstrate sharpness of the upper bounds on $E(Y(1))$, $E(Y(0))$, as the proof for the lower bounds is symmetric and therefore omitted. The proof consists of three steps. In the first step, we show that for π_{01} fixed, the upper bounds on $E(Y(1))^{UB}$, $E(Y(0))^{UB}$ in expressions (11) and (12) in the main text are valid in the sense that they are weakly greater than $E(Y(1))$ and $E(Y(0))$, respectively, with probability one. In the second step, we prove that these upper bounds are sharp conditional on π_{01} by showing that there exists a distribution of T given Z and of $(Y(1),Y(0))$ given T and Z (for $T=11,10,01,00$ and $Z=1,0$) that is compatible with a data generating process satisfying Assumption 2 and at the same time yielding $E(Y(1))=E(Y(1))^{UB}$ and $E(Y(0))=E(Y(0))^{UB}$. Thirdly, since π_{01} is unknown, the sharp upper bounds over all admissible π_{01} are (by the virtue of the extreme value theorem and the continuity of the bounds in π_{01} as shown in Appendix A.1.4) obtained by taking the supremum of the conditional upper bounds w.r.t. π_{01}, which is shown to be achieved at $\pi_{01} = \pi_{01}^{\min}$.

We start the proof by showing that for π_{01} given,

$$\begin{aligned}
E(Y(1)) \leq E(Y(1))^{UB}(\pi_{01}) &= (P_{0|1} - \pi_{01}) \cdot y^{UB} - (P_{1|0} - \pi_{01}) \cdot \max\left(\bar{Y}_{1,1}(\min|q_{1,1}^{11}), \bar{Y}_{0,1}(\min|q_{0,1}^{11})\right) \\
&+ P_{1|0} \cdot \bar{Y}_{0,1} + P_{1|1} \cdot \bar{Y}_{1,1}, \\
E(Y(0)) \leq E(Y(0))^{UB}(\pi_{01}) &= (P_{1|0} - \pi_{01}) \cdot y^{UB} - (P_{0|1} - \pi_{01}) \cdot \max\left(\bar{Y}_{0,0}(\min|q_{0,0}^{00}), \bar{Y}_{1,0}(\min|q_{1,0}^{00})\right) \\
&+ P_{0|0} \cdot \bar{Y}_{0,0} + P_{0|1} \cdot \bar{Y}_{1,0},
\end{aligned}$$

so that the upper bounds are valid. For the sake of notational convenience we omit (π_{01}) in the subsequent discussion. Note that by the law of total probability,

$$E(Y(0)) = \pi_{11} \cdot E(Y(0)|T=11) + \pi_{10} \cdot E(Y(0)|T=10) + \pi_{01} \cdot E(Y(0)|T=01) + \pi_{00} \cdot E(Y(0)|T=00),$$

which, using equations (3) and (4), can be written as

$$\begin{aligned}
E(Y(0)) &= \pi_{11} \cdot E(Y(0)|T=11) + P_{0|0} \cdot \bar{Y}_{0,0} - \pi_{00} \cdot E(Y(0)|T=00) \\
&+ P_{0|1} \cdot \bar{Y}_{1,0} - \pi_{00} \cdot E(Y(0)|T=00) + \pi_{00} \cdot E(Y(0)|T=00) \\
&= \pi_{11} \cdot E(Y(0)|T=11) - \pi_{00} \cdot E(Y(0)|T=00) + P_{0|0} \cdot \bar{Y}_{0,0} + P_{0|1} \cdot \bar{Y}_{1,0} \qquad \text{(A.16)} \\
&= (P_{1|0} - \pi_{01}) \cdot E(Y(0)|T=11) - (P_{0|1} - \pi_{01}) \cdot E(Y(0)|T=00) + P_{0|0} \cdot \bar{Y}_{0,0} + P_{0|1} \cdot \bar{Y}_{1,0}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
E(Y(1)) &= \pi_{11} \cdot E(Y(1)|T=11) + \pi_{10} \cdot E(Y(1)|T=10) + \pi_{01} \cdot E(Y(1)|T=01) + \pi_{00} \cdot E(Y(1)|T=00) \\
&= (P_{0|1} - \pi_{01}) \cdot E(Y(1)|T=00) - (P_{1|0} - \pi_{11}) \cdot E(Y(1)|T=11) + P_{1|1} \cdot \bar{Y}_{1,1} + P_{1|0} \cdot \bar{Y}_{0,1},
\end{aligned} \qquad \text{(A.17)}$$

where we use equations (1) and (2) to go from the first to the second line.

It is easy to see that $E(Y(0))^{UB}$ is a valid upper bound for $E(Y(0))$:

$$\begin{aligned}
E(Y(0))^{UB} - E(Y(0)) &= (P_{1|0} - \pi_{01}) \cdot (y^{UB} - E(Y(0)|T=11)) \\
&+ (P_{0|1} - \pi_{01}) \cdot (E(Y(0)|T=00) - \max\left(\bar{Y}_{0,0}(\min|q_{0,0}^{00}), \bar{Y}_{1,0}(\min|q_{1,0}^{00})\right)) \geq 0,
\end{aligned}$$

because $(P_{1|0} - \pi_{01}) \geq 0$, $(P_{0|1} - \pi_{01}) \geq 0$, $y^{UB} \geq E(Y(0)|T=11)$, and, as already shown in Section A.1.2, $E(Y(0)|T=00) \geq E(Y(0)|T=00)^{LB} = \max\left(\bar{Y}_{0,0}(\min|q_{0,0}^{00}), \bar{Y}_{1,0}(\min|q_{1,0}^{00})\right)$. In a similar way one can prove the validity of $E(Y(1))^{UB}$.

To demonstrate sharpness we need to show that for each $E(Y(d)) \in [E(Y(d))^{LB}, E(Y(d))^{UB}], d = 1,0$ there exist principal strata proportions $\Pr(T = t|Z = z) : t = 11, 10, 01, 00, z = 1, 0$ and potential outcome distributions $f(y(1), y(0)|t, z) : T = 11, 10, 01, 00, Z = 1, 0$ that are compatible with a data generating process that satisfies Assumptions 2. As $E(Y(d))^{LB}, E(Y(d))^{UB}, d = 1, 0$ are the smallest and largest values of the interval $[E(Y(d))^{LB}, E(Y(d))^{UB}]$, it is sufficient to prove the existence at those two extremes. The reason is that the values of $E(Y(d)|T = t)$ inside this interval can be achieved as convex combinations of the potential outcome distributions $f(y(1), y(0)|t, z) : t = 11, 10, 01, 00, z = 1, 0$ that generate $E(Y(d)|T = t)^{LB}$ and $E(Y(d)|T = t)^{UB}$. For the upper bound, one therefore has to show that such distributions exist so that $E(Y(d)) = E(Y(d))^{UB}$.

To this end, reconsider (A.10) and (A.11), which have been shown to be consistent with the data generating process (see Section A.1.2), satisfy Assumption 2, and give

$$\int y(0)h^z_{11}\, dy(1) = E(Y(0)|T = 11) = y^{UB},$$

$$\int y(1)h^z_{00}\, dy(0) = E(Y(1)|T = 00) = y^{UB},$$

$$\int y(1)h^z_{11}\, dy(0) = E(Y(1)|T = 11) = \max\left(\bar{Y}_{0,1}(\min|q^{11}_{0,1}), \bar{Y}_{1,1}(\min|q^{11}_{1,1})\right),$$

and

$$\int y(0)h^z_{00}\, dy(1) = E(Y(0)|T = 00) = \max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}), \bar{Y}_{1,0}(\min|q^{00}_{1,0})\right).$$

Therefore, (A.10) and (A.11) imply $E(Y(1)) = E(Y(1))^{UB}$ and $E(Y(0)) = E(Y(0))^{UB}$, which proves that they are sharp for a given value of π_{01}.

Finally, to obtain the upper bound on $E(Y(0))$ over all admissible values of π_{01}, we show that $E(Y(0))^{UB}(\pi_{01})$, where we now make conditioning on π_{01} explicit again, is a decreasing function of π_{01} so that $E(Y(0))^{UB} = \sup_{\pi_{01}\in\mathcal{P}^*}[E(Y(0))^{UB}(\pi_{01})] = E(Y(0))^{UB}(\pi^{\min}_{01})$. In an analogous way one can demonstrate that $E(Y(1))^{UB} = \sup_{\pi_{01}\in\mathcal{P}^*}[E(Y(1))^{UB}(\pi_{01})] = E(Y(1))^{UB}(\pi^{\min}_{01})$, which is omitted for the sake of brevity. We first rewrite $E(Y(0))^{UB}(\pi_{01})$ as

$$\begin{aligned} E(Y(0))^{UB}(\pi_{01}) &= \min\left\{\begin{array}{c} (P_{1|0} - \pi_{01})\cdot y^{UB} - (P_{0|1} - \pi_{01})\cdot \bar{Y}_{0,0}(\min|q^{00}_{0,0}) + P_{0|0}\cdot\bar{Y}_{0,0} + P_{0|1}\cdot\bar{Y}_{1,0} \\ (P_{1|0} - \pi_{01})\cdot y^{UB} - (P_{0|1} - \pi_{01})\cdot \bar{Y}_{1,0}(\min|q^{00}_{1,0}) + P_{0|0}\cdot\bar{Y}_{0,0} + P_{0|1}\cdot\bar{Y}_{1,0} \end{array}\right\} \\ &= \min\left\{\begin{array}{c} E(Y(0))^{UB}_1(\pi_{01}) \\ E(Y(0))^{UB}_2(\pi_{01}) \end{array}\right\}. \end{aligned}$$

Note that if Y is a random variable with a continuous probability density function, the trimmed means $\bar{Y}_{z,0}(\min|q^{00}_{z,0})$ are differentiable in quantile $q^{00}_{z,0}$, because the quantile is differentiable in π_{01}.

We can therefore verify that both $E(Y(0))_1^{UB}(\pi_{01})$ and $E(Y(0))_2^{UB}(\pi_{01})$ are decreasing in π_{01}.

$$\frac{\partial E(Y(0))_1^{UB}}{\partial \pi_{01}} = (\bar{Y}_{0,0}(\min|q_{0,0}^{00}) - y^{UB}) - (P_{0|1} - \pi_{01}) \cdot \frac{\partial \bar{Y}_{0,0}(\min|q_{0,0}^{00})}{\partial \pi_{01}} < 0$$

and

$$\frac{\partial E(Y(0))_2^{UB}}{\partial \pi_{01}} = (\bar{Y}_{1,0}(\min|q_{1,0}^{00}) - y^{UB}) - (P_{0|1} - \pi_{01}) \cdot \frac{\partial \bar{Y}_{1,0}(\min|q_{1,0}^{00})}{\partial \pi_{01}} < 0.$$

The two inequalities are always satisfied, because $0 \leq (P_{0|1} - \pi_{01}) \leq 1$ and $\frac{\partial \bar{Y}_{1,0}(\min|q_{1,0}^{00})}{\partial \pi_{01}}$ (the marginal decrease in $\bar{Y}_{z,0}(\min|q_{z,0}^{00})$ due to a marginal increase in π_{01}), albeit always negative, cannot be larger in absolute terms than the difference of $\bar{Y}_{z,0}(\min|q_{z,0}^{00})$ and the largest possible value y^{UB}. Since the minimum of two monotonically decreasing functions is itself a monotonically decreasing function we conclude that $\sup_{\pi_{01}\in\mathcal{P}^*}[E(Y(0))^{UB}(\pi_{01})] = E(Y(0))^{UB}(\pi_{01}^{\min})$, because $\mathcal{P}^*$ is a compact set. This ends the proof. ■

A.1.6 Comparing our bounds to Manski (1990)

This section compares our bounds under mean independence within strata to those of Manski (1990). First, we write Manski's upper bound, denoted by Δ_{Ma}^{UB}, in our notation as

$$\begin{aligned}\Delta_{Ma}^{UB} &= \min(P_{1|1} \cdot \bar{Y}_{1,1} + P_{0|1} \cdot y^{UB}, P_{1|0} \cdot \bar{Y}_{0,1} + P_{0|0} \cdot y^{UB}) \\ &- \max(P_{0|0} \cdot \bar{Y}_{0,0} + P_{1|0} \cdot y^{LB}, P_{0|1} \cdot \bar{Y}_{1,0} + P_{1|1} \cdot y^{LB}).\end{aligned}$$

In order to see the differences between Manski's bounds and ours, consider the case where positive monotonicity is consistent with the data ($P_{1|1} > P_{1|0}$, which implies $P_{0|0} > P_{0|1}$). We have that[19]

$$\begin{aligned}P_{1|1} \cdot \bar{Y}_{1,1} + P_{0|1} \cdot y^{UB} &\leq P_{1|0} \cdot \bar{Y}_{0,1} + P_{0|0} \cdot y^{UB} \\ (P_{1|0} - P_{1|1}) \cdot y^{UB} &\leq (P_{1|0} - P_{1|1}) \cdot \left(\frac{P_{1|0}}{P_{1|0} - P_{1|1}} \cdot \bar{Y}_{0,1} - \frac{P_{1|1}}{P_{1|0} - P_{1|1}} \cdot \bar{Y}_{1,1}\right) \\ y^{UB} &\geq \frac{P_{1|0}}{P_{1|0} - P_{1|1}} \cdot \bar{Y}_{0,1} - \frac{P_{1|1}}{P_{1|0} - P_{1|1}} \cdot \bar{Y}_{1,1}.\end{aligned} \tag{A.18}$$

This is always satisfied since $\frac{P_{1|0}}{P_{1|0}-P_{1|1}} \cdot \bar{Y}_{0,1} - \frac{P_{1|1}}{P_{1|0}-P_{1|1}} \cdot \bar{Y}_{1,1}$, which is a weighted difference of two means, cannot be higher than y^{UB}, which is the largest value that Y can theoretically take. A symmetric argument can be used to show that $\max(P_{0|0} \cdot \bar{Y}_{0,0} + P_{1|0} \cdot y^{LB}, P_{0|1} \cdot \bar{Y}_{1,0} + P_{1|1} \cdot y^{LB}) = P_{0|0} \cdot \bar{Y}_{0,0} + P_{1|0} \cdot y^{LB}$.

Therefore, Δ_{Ma}^{UB} simplifies to

$$\Delta_{Ma}^{UB} = \left(P_{1|1} \cdot \bar{Y}_{1,1} + P_{0|1} \cdot y^{UB}\right) - \left(P_{0|0} \cdot \bar{Y}_{0,0} + P_{1|0} \cdot y^{LB}\right).$$

[19]Note that the sign of the inequality changes from the second to the third line because $P_{1|0} - P_{1|1}$ is negative.

On the other hand, since $\bar{Y}_{0,1}(\min|q_{0,1}^{11,0}) = \bar{Y}_{0,1}$ and $\bar{Y}_{1,0}(\min|q_{1,0}^{00,0}) = \bar{Y}_{1,0}$, our upper bound becomes

$$\begin{aligned}\Delta^{UB} &= \Big(P_{1|1}\cdot\bar{Y}_{1,1} + P_{1|0}\cdot\bar{Y}_{0,1} - P_{1|0}\cdot\max(\bar{Y}_{1,1}(\min|q_{1,1}^{11,0}),\bar{Y}_{0,1}) + P_{0|1}\cdot y^{UB}\Big)\\ &- \Big(P_{0|0}\cdot\bar{Y}_{0,0} + P_{0|1}\cdot\bar{Y}_{1,0} - P_{0|1}\cdot\min(\bar{Y}_{0,0}(\max|q_{0,0}^{00,0}),\bar{Y}_{1,0}) + P_{1|1}\cdot y^{LB}\Big).\end{aligned}$$

This implies that

$$\begin{aligned}\Delta^{UB} - \Delta^{UB}_{Ma} &= P_{1|0}\cdot(\max(\bar{Y}_{1,1}(\min|q_{1,1}^{11,0}),\bar{Y}_{0,1}) - \bar{Y}_{0,1})\\ &- P_{0|1}\cdot(\min(\bar{Y}_{0,0}(\max|q_{0,0}^{00,0}),\bar{Y}_{1,0}) - \bar{Y}_{1,0}) \le 0 \qquad (A.19)\end{aligned}$$

It is easy to see that $\Delta^{UB} = \Delta^{UB}_{Ma}$, unless either $\bar{Y}_{1,1}(\min|q_{1,1}^{11,0}) > \bar{Y}_{0,1}$ or $\bar{Y}_{0,0}(\max|q_{0,0}^{00,0}) < \bar{Y}_{1,0}$ (or both) is satisfied. As discussed in Section 3.2, any of these inequalities can be interpreted as a violation of Assumptions 2 and 3. This shows that Manski's and our bounds are the same unless we detect a violation of the LATE assumptions.

A.1.7 Comparing our bounds to Kitagawa (2009)

To compare our bounds to those proposed in Kitagawa (2009), we need to introduce some extra notation. Let

$$\begin{aligned}p_1(y) &= f(y, D=1|Z=1), \quad p_0(y) = f(y, D=0|Z=1),\\ q_1(y) &= f(y, D=1|Z=0), \quad q_0(y) = f(y, D=0|Z=0),\\ \delta_1 &= \int \max(p_1(y), q_1(y))dy, \quad \delta_0 = \int \max(p_0(y), q_0(y))dy,\\ \lambda_1 &= \int \min(p_1(y), q_1(y))dy, \quad \lambda_0 = \int \min(p_0(y), q_0(y))dy.\end{aligned}$$

Moreover, Kitagawa (2009) defines the α-th left- and right- trimming of a nonnegative integrable function $g : \mathcal{Y} \mapsto \mathbb{R}$. Let $q^{lt}_\alpha = \inf\left\{t : \int_{[y^{LB},t]} g(y)dy \ge \alpha\right\}$ and $q^{rt}_\alpha = \inf\left\{t : \int_{[t,y^{UB}]} g(y)dy \ge \alpha\right\}$. The α-th left- and right- trimming functions are

$$(g)^{lt}_\alpha(y) = g(y)\cdot I\{y > q^{lt}_\alpha\} + \left(\int_{[y^{LB},q^{lt}_\alpha]} g(y)dy - \alpha\right)\cdot I\{y = q^{lt}_\alpha\}, \qquad (A.20)$$

$$(g)^{rt}_\alpha(y) = g(y)\cdot I\{y < q^{rt}_\alpha\} + \left(\int_{[q^{rt}_\alpha,y^{UB}]} g(y)dy - \alpha\right)\cdot I\{y = q^{rt}_\alpha\}. \qquad (A.21)$$

Notice that those functions degenerate to zero for $\alpha \ge \int g(y)dy$, sum up to $g(y)$ and integrate to $\int g(y)dy - \alpha$.

Given this notation Kitagawa's bounds on Δ can be written in a compact way as

$$\begin{aligned}
\Delta_{Ki}^{UB} &= \min(1-\delta_1,\lambda_0)\cdot y^{UB} + \int y_1 \cdot (\max(p_1(y),q_1(y)) + (\min(p_1(y),q_1(y)))^{lt}_{\min(1-\delta_1,\lambda_0)})dy \\
&- \min(1-\delta_0,\lambda_1)\cdot y^{LB} - \int y_0 \cdot (\max(p_0(y),q_0(y)) + (\min(p_0(y),q_0(y)))^{rt}_{\min(1-\delta_0,\lambda_1)})dy, \\
\Delta_{Ki}^{LB} &= \min(1-\delta_1,\lambda_0)\cdot y^{LB} + \int y_1 \cdot (\max(p_1(y),q_1(y)) + (\min(p_1(y),q_1(y)))^{rt}_{\min(1-\delta_1,\lambda_0)})dy \\
&- \min(1-\delta_0,\lambda_1)\cdot y^{UB} - \int y_0 \cdot (\max(p_0(y),q_0(y)) + (\min(p_0(y),q_0(y)))^{lt}_{\min(1-\delta_0,\lambda_1)})dy.
\end{aligned}$$

For the comparison we will only consider the upper bound, a symmetric argument holds for the lower bound. For the same reason we only consider the case of $P_{1|1} > P_{1|0}$. Our upper bound on Δ can be rewritten as

$$\begin{aligned}
\Delta^{UB} &= P_{0|1}\cdot y^{UB} + \int y_1 \cdot (\max(p_1(y),q_1(y)) + \min(p_1(y),q_1(y)))dy \\
&- P_{1|0}\cdot \max(\bar{Y}_{1,1}(\min|q_{1,1}^{11,0}), \bar{Y}_{0,1}) \\
&- P_{1|0}\cdot y^{LB} - \int y_0 \cdot (\max(p_0(y),q_0(y)) + \min(p_0(y),q_0(y)))dy \\
&+ P_{0|1}\cdot \min(\bar{Y}_{0,0}(\max|q_{0,0}^{00,0}), \bar{Y}_{1,0}).
\end{aligned}$$

Therefore, the difference between Kitagawa's upper bound and ours is

$$\begin{aligned}
\Delta^{UB} - \Delta_{Ki}^{UB} &= (P_{0|1} - \min(1-\delta_1,\lambda_0))\cdot y^{UB} + \int y_1 \cdot \min(p_1(y),q_1(y))dy \qquad \text{(A.22)} \\
&- \int y_1 \cdot (\min(p_1(y),q_1(y)))^{lt}_{\min(1-\delta_1,\lambda_0)}dy - P_{1|0}\cdot \max(\bar{Y}_{1,1}(\min|q_{1,1}^{11,0}), \bar{Y}_{0,1}) \\
&- (P_{1|0} - \min(1-\delta_0,\lambda_1))\cdot y^{LB} - \int y_0 \cdot \min(p_0(y),q_0(y))dy \\
&+ \int y_0 \cdot (\min(p_0(y),q_0(y)))^{rt}_{\min(1-\delta_0,\lambda_1)}dy + P_{0|1}\cdot \min(\bar{Y}_{0,0}(\max|q_{0,0}^{00,0}), \bar{Y}_{1,0}) \geq 0
\end{aligned}$$

We now show that this inequality is always satisfied. First of all, we demonstrate that

$$(P_{0|1} - \min(1-\delta_1,\lambda_0))\cdot y^{UB} - (P_{1|0} - \min(1-\delta_0,\lambda_1))\cdot y^{LB} \geq 0$$

In order to do so, it is sufficient to show that $P_{1|0} - \min(1-\delta_0,\lambda_1) = P_{0|1} - \min(1-\delta_1,\lambda_0) \geq 0$ Kitagawa (2009) proves that $P_{1|1}+P_{1|0} = \delta_1+\lambda_1$ and similarly, one can show that $P_{0|0}+P_{0|1} = \delta_0+\lambda_0$. Thus, $P_{1|0}-\lambda_1 = P_{0|1}-(1-\delta_1)$ and $P_{0|1}-\lambda_0 = P_{1|0}-(1-\delta_0)$. Moreover, from $\delta_1+\lambda_1+\delta_0+\lambda_0 = 2$, we have $(1-\delta_1)-\lambda_0 = \lambda_1-(1+\delta_0)$, which implies $(1-\delta_1) > \lambda_0 \Rightarrow \lambda_1 > (1-\delta_0)$ and $(1-\delta_1) < \lambda_0 \Rightarrow \lambda_1 < (1-\delta_0)$. Therefore, $P_{1|0} - \min(1-\delta_0,\lambda_1) = P_{0|1} - \min(1-\delta_1,\lambda_0)$. Finally $P_{1|0} \geq \lambda_1$ by construction and $P_{1|0} \geq (1-\delta_0)$, since $P_{0|1} \geq \lambda_0$ by construction and $P_{0|1}-\lambda_0 = P_{1|0}-(1-\delta_0)$.

To end the proof, we need to show that

$$\left(\int y_1 \cdot \min(p_1(y), q_1(y))dy - \int y_1 \cdot (\min(p_1(y), q_1(y)))^{lt}_{\min(1-\delta_1,\lambda_0)}dy\right) - P_{1|0} \cdot \max(\bar{Y}_{1,1}(\min|q^{11,0}_{1,1}), \bar{Y}_{0,1}) \geq 0,$$

$$P_{0|1} \cdot \min(\bar{Y}_{0,0}(\max|q^{00,0}_{0,0}), \bar{Y}_{1,0}) - \left(\int y_0 \cdot \min(p_0(y), q_0(y))dy - \int y_0 \cdot (\min(p_0(y), q_0(y)))^{rt}_{\min(1-\delta_0,\lambda_1)}dy\right) \geq 0.$$

From the results of Kitagawa it follows that $\int y_1 \cdot \min(p_1(y), q_1(y))dy - \int y_1 \cdot (\min(p_1(y), q_1(y)))^{lt}_{\min(1-\delta_1,\lambda_0)}dy$ is the upper bound for $E(Y, D = 1|T = 11)$ while from our results it follows that $P_{1|0} \cdot \max(\bar{Y}_{1,1}(\min|q^{11,0}_{1,1}), \bar{Y}_{0,1})$ is the lower bound on the same parameter, therefore the first inequality is always satisfied. Similarly, $\int y_0 \cdot \min(p_0(y), q_0(y))dy - \int y_0 \cdot (\min(p_0(y), q_0(y)))^{rt}_{\min(1-\delta_0,\lambda_1)}dy$ is the lower bound for $E(Y, D = 0|T = 00)$, while $P_{0|1} \cdot \min(\bar{Y}_{0,0}(\max|q^{00,0}_{0,0}), \bar{Y}_{1,0})$ is the upper bound on the same parameter. This ends the proof.

An important result of Kitagawa (2009) is that under full independence of the instrument and the potential outcomes/treatment states, monotonicity does not hold if one or both of the following inequalities are violated for some $y \in \mathcal{Y}$:

$$p_1(y) \geq q_1(y), \quad q_0(y) \geq p_0(y) \quad \forall y \in \mathcal{Y}. \tag{A.23}$$

However, under the satisfaction of (A.23), Kitagawa's bounds are equivalent to Manski's (1990), which are in turn equivalent to ours, because (A.23) implies the satisfaction of Assumptions 2 and 3 (mean independence within strata and monotonicity). Therefore, only if (A.23) is violated, the inequality in (A.22) becomes strict (and turns into an equality otherwise).

A.1.8 Order of the bounds on the treated, non-treated and the entire population w.r.t. their tightness

First of all, note that $\Delta^{UB}_{D=1} > \Delta^{UB}_{D=0}$ immediately implies that $\Delta^{LB}_{D=1} < \Delta^{LB}_{D=0}$ since the bounds are symmetric. Moreover,

$$\begin{aligned}
\Delta^{UB}_{D=1} - \Delta^{UB} &= \Delta^{UB}_{D=1} - \Pr(D = 1) \cdot \Delta^{UB}_{D=1} - (1 - \Pr(D = 1)) \cdot \Delta^{UB}_{D=0}, \\
&= \Pr(D = 0) \cdot (\Delta^{UB}_{D=1} - \Delta^{UB}_{D=0}),
\end{aligned}$$

and

$$\begin{aligned}
\Delta^{UB}_{D=0} - \Delta^{UB} &= \Delta^{UB}_{D=0} - \Pr(D = 0) \cdot \Delta^{UB}_{D=0} - (1 - \Pr(D = 0)) \cdot \Delta^{UB}_{D=1}, \\
&= \Pr(D = 1) \cdot (\Delta^{UB}_{D=0} - \Delta^{UB}_{D=1}).
\end{aligned}$$

Therefore, $\Delta^{UB}_{D=1} < \Delta^{UB}_{D=0}$ implies $\Delta^{UB}_{D=1} < \Delta^{UB} < \Delta^{UB}_{D=0}$ and $\Delta^{LB}_{D=0} < \Delta^{LB} < \Delta^{LB}_{D=1}$. The converse is true if $\Delta^{UB}_{D=1} > \Delta^{UB}_{D=0}$. This ends the proof.

A.2 Monotonicity

A.2.1 Proof of the sharpness of the bounds for the always takers

Under monotonicity and Assumption 2, $E(Y(1)|T = 11)$ is identified by $\bar{Y}_{0,1}$. Since monotonicity does not impose any restrictions on the distribution of $Y(0)|T = 11$, $E(Y(1)|T = 11)^{UB} = y^{UB}$ and $E(Y(1)|T = 11)^{LB} = y^{LB}$. This implies that Δ^{UB}_{11} and Δ^{LB}_{11} are the sharp upper and lower bounds of Δ_{11}.

A.2.2 Proof of the sharpness of the bounds for the never takers

Under monotonicity and Assumption 2, $E(Y(0)|T = 00)$ is identified by $\bar{Y}_{1,0}$. Since monotonicity does not impose any restrictions on the distribution of $Y(1)|T = 00$, $E(Y(0)|T = 11)^{UB} = y^{UB}$ and $E(Y(0)|T = 11)^{LB} = y^{LB}$. This implies that Δ^{UB}_{00} and Δ^{LB}_{00} are the sharp upper and lower bounds of Δ_{00}.

A.2.3 Proof of the sharpness of the bounds on the mean potential outcomes $E(Y(1))$, $E(Y(0))$

Under monotonicity, $\pi_{11} = P_{1|0}$, $\pi_{00} = P_{0|1}$, $E(Y(1)|T = 11) = \bar{Y}_{0,1}$ and $E(Y(0)|T = 00) = \bar{Y}_{1,0}$, so that equations (A.16) and (A.17) simplify to

$$
\begin{aligned}
E(Y(0)) &= P_{1|0} \cdot E(Y(0)|T = 11) + P_{0|0} \cdot \bar{Y}_{0,0}, \\
E(Y(1)) &= P_{0|1} \cdot E(Y(1)|T = 00) + P_{1|1} \cdot \bar{Y}_{1,1}.
\end{aligned}
$$

The only unidentified element of $E(Y(0))$ is $E(Y(0)|T = 11)$. Therefore, its sharp upper and lower bounds are obtained by substituting $E(Y(0)|T = 11)$ with $E(Y(0)|T = 11)^{UB} = y^{UB}$ and $E(Y(0)|T = 11)^{LB} = y^{LB}$, respectively. Similarly, the sharp bounds upper and lower bounds on $E(Y(1))$ are obtained by substituting $E(Y(1)|T = 11)$ with $E(Y(1)|T = 11)^{UB} = y^{UB}$ and $E(Y(1)|T = 11)^{LB} = y^{LB}$, respectively.

A.3 Mean dominance

A.3.1 Identified set of the proportion of defiers under mean dominance based on linear programming

Assumption 4 together with Assumption 2 (i) can be rewritten as

$$E[Y(d)|T=10,Z=0] \geq E[Y(d)|T=t,Z=0] \quad \forall d \in \{0,1\},\ t \in \{11,00\},$$

and the corresponding additional restrictions on h (recall that $h_z^t(y_i,y_j) = \Pr(Y(1)=y_i, Y(0)=y_j|T=t,Z=z)$) are

$$\begin{aligned}
\sum_{i=1}^{k}\sum_{j=1}^{k} y_i h_0^{10}(y_i,y_j) &\geq \sum_{i=1}^{k}\sum_{j=1}^{k} y_i h_0^{11}(y_i,y_j), \\
\sum_{i=1}^{k}\sum_{j=1}^{k} y_i h_0^{10}(y_i,y_j) &\geq \sum_{i=1}^{k}\sum_{j=1}^{k} y_i h_0^{00}(y_i,y_j), \\
\sum_{i=1}^{k}\sum_{j=1}^{k} y_j h_0^{10}(y_i,y_j) &\geq \sum_{i=1}^{k}\sum_{j=1}^{k} y_j h_0^{11}(y_i,y_j), \\
\sum_{i=1}^{k}\sum_{j=1}^{k} y_j h_0^{10}(y_i,y_j) &\geq \sum_{i=1}^{k}\sum_{j=1}^{k} y_j h_0^{00}(y_i,y_j).
\end{aligned} \tag{A.24}$$

If there exists a solution to $(A.4)-(A.6)$ and $(A.24)$ for a given value of the defier share π_{01}, then such a π_{01} is admissible. Note that the sharp identified set for the share of defiers under Assumptions 1, 2, and 4 is an interval because the mechanism of the proof of Lemma 1 applies.

A.3.2 Identified set of the proportion of defiers under mean dominance based on moment inequalities

The inequalities are similar to (A.15), with the only difference that $\bar{Y}_{1,1}(\max|q_{1,1}^{11})$ is replaced by $\bar{Y}_{1,1}$ and $\bar{Y}_{0,0}(\max|q_{0,0}^{00})$ is replaced by $\bar{Y}_{0,0}$.

$$\begin{aligned}
\bar{Y}_{0,1}(\min|q_{0,1}^{11}) &\leq \bar{Y}_{1,1}, \\
\bar{Y}_{1,1}(\min|q_{1,1}^{11}) &\leq \bar{Y}_{0,1}(\max|q_{0,1}^{11}), \\
\bar{Y}_{1,0}(\min|q_{1,0}^{00}) &\leq \bar{Y}_{0,0}, \\
\bar{Y}_{0,0}(\min|q_{0,0}^{00}) &\leq \bar{Y}_{1,0}(\max|q_{1,0}^{00}).
\end{aligned} \tag{A.25}$$

The proof for necessity and sufficiency of (A.25) for the share of defiers π_{01} to be included in $\mathcal{P}^{**}$

is analogous to the one presented in Appendix A.1.3. A similar argument as in Appendix A.1.3 also applies to prove that $\pi_{01}^{\max} = \min\{P_{1|0}, P_{0|1}\} \in \mathcal{P}^{**}$.

A.3.3 Bounds on the ATEs under mean dominance only

Under mean dominance, the ATE on the compliers is bounded by

$$\begin{aligned}
\Delta_{10}^{UB} &= \sup_{\pi_{01}\in\mathcal{P}^{**}} \left[\frac{P_{1|1}\cdot \bar{Y}_{1,1} - (P_{1|0}-\pi_{01})\cdot \max\left(\bar{Y}_{1,1}(\min|q_{1,1}^{11}), \bar{Y}_{0,1}(\min|q_{0,1}^{11})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \right. \\
&- \left. \frac{P_{0|0}\cdot \bar{Y}_{0,0} - (P_{0|1}-\pi_{01})\cdot \min\left(\bar{Y}_{0,0}, \bar{Y}_{1,0}(\max|q_{1,0}^{00})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \right], \\
\Delta_{10}^{LB} &= \inf_{\pi_{01}\in\mathcal{P}^{**}} \left[\frac{P_{1|1}\cdot \bar{Y}_{1,1} - (P_{1|0}-\pi_{01})\cdot \min\left(\bar{Y}_{1,1}, \bar{Y}_{0,1}(\max|q_{0,1}^{11})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \right. \\
&- \left. \frac{P_{0|0}\cdot \bar{Y}_{0,0} - (P_{0|1}-\pi_{01})\cdot \max\left(\bar{Y}_{0,0}(\min|q_{0,0}^{00}), \bar{Y}_{1,0}(\min|q_{1,0}^{00})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \right].
\end{aligned} \tag{A.26}$$

The intuition of this result is that the compliers' mean potential outcome under treatment cannot be lower than that of the always takers, while under non-treatment it cannot be lower than the one of the never takers. Therefore, we now have the minimization problems $\min\left(\bar{Y}_{0,0}, \bar{Y}_{1,0}(\max|q_{1,0}^{00})\right)$ and $\min\left(\bar{Y}_{1,1}, \bar{Y}_{0,1}(\max|q_{0,1}^{11})\right)$, as $\bar{Y}_{0,0}$ is the lower bound on the compliers' mean potential outcome in the mixed population with the defiers and $\bar{Y}_{1,1}$ in the mixed group with the always takers. The sharpness of these bounds and all other bounds under mean dominance proposed below follows from the fact that they are special cases of the bounds derived in Section 3.1 and that we can apply Lemma 2 in Appendix A.3 to formally prove their sharpness.

The bounds for the defiers are the same as in Section 3.1, since we do not impose any mean dominance assumption w.r.t. the potential outcomes of this population. The bounds for the ATEs on the always takers and never takers are, respectively,

$$\begin{aligned}
\Delta_{11}^{UB} &= \min\left(\bar{Y}_{1,1}, \bar{Y}_{0,1}(\max|q_{0,1}^{11,\pi_{01}^{\max}})\right) - y^{LB}, \\
\Delta_{11}^{LB} &= \inf_{\pi_{01}\in\mathcal{P}^{**}} \left[\max\left(\bar{Y}_{1,1}(\min|q_{1,1}^{11}), \bar{Y}_{0,1}(\min|q_{0,1}^{11})\right)\right. \\
&- \left. \frac{P_{0|0}\cdot \bar{Y}_{0,0} - (P_{0|1}-\pi_{01})\cdot \max\left(\bar{Y}_{0,0}(\min|q_{0,0}^{00}), \bar{Y}_{1,0}(\min|q_{1,0}^{00})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \right],
\end{aligned} \tag{A.27}$$

and

$$\begin{aligned}
\Delta_{00}^{UB} &= \sup_{\pi_{01}\in\mathcal{P}^{**}} \left[\frac{P_{1|1}\cdot \bar{Y}_{1,1} - (P_{1|0}-\pi_{01})\cdot \max\left(\bar{Y}_{1,1}(\min|q_{1,1}^{11}), \bar{Y}_{0,1}(\min|q_{0,1}^{11})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \right. \\
&- \left. \max\left(\bar{Y}_{1,0}(\min|q_{1,0}^{00}), \bar{Y}_{0,0}(\min|q_{0,0}^{00})\right)\right], \\
\Delta_{00}^{LB} &= y^{LB} - \min\left(\bar{Y}_{1,0}(\max|q_{1,0}^{00,\pi_{01}^{\max}}), \bar{Y}_{0,0}\right).
\end{aligned} \tag{A.28}$$

$\bar{Y}_{1,1}$, $\bar{Y}_{0,0}$ are now the upper bounds on the mean potential outcome under treatment and the mean potential outcome under non-treatment for the always takers and the never takers, respectively. Moreover, mean dominance implies that the always takers' upper bound under non-treatment cannot be higher than the compliers' upper bound under non-treatment and that the never takers' upper bound under treatment cannot be higher than the compliers' upper bound under treatment. This is a considerable improvement over the bounds only invoking mean independence within strata, as it allows us to replace y^{UB} by observed quantities. It, however, requires optimization over all possible values of π_{01}.

Under mean dominance the sharp bounds on $E(Y(1))$ and $E(Y(0))$ become

$$\begin{aligned}
E(Y(0))^{UB} &= \sup_{\pi_{01}\in\mathcal{P}^{**}} \Big[(P_{1|0}-\pi_{01})\cdot \frac{P_{1|1}\cdot\bar{Y}_{1,1}-(P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}),\bar{Y}_{0,1}(\min|q^{11}_{0,1})\right)}{P_{1|1}-P_{1|0}+\pi_{01}}\\
&- (P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)+P_{0|0}\cdot\bar{Y}_{0,0}+P_{0|1}\cdot\bar{Y}_{1,0}\Big],\\
E(Y(0))^{LB} &= \inf_{\pi_{01}\in\mathcal{P}^{**}} \Big[(P_{1|0}-\pi_{01})\cdot y^{LB}-(P_{0|1}-\pi_{01})\cdot\min\left(\bar{Y}_{0,0},\bar{Y}_{1,0}(\max|q^{00}_{1,0})\right)\\
&+ P_{0|0}\cdot\bar{Y}_{0,0}+P_{0|1}\cdot\bar{Y}_{1,0}\Big],
\end{aligned}$$

and

$$\begin{aligned}
E(Y(1))^{UB} &= \sup_{\pi_{01}\in\mathcal{P}^{**}} \Big[(P_{0|1}-\pi_{01})\cdot \frac{P_{0|0}\cdot\bar{Y}_{0,0}-(P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)}{P_{1|1}-P_{1|0}+\pi_{01}}\\
&- (P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}),\bar{Y}_{0,1}(\min|q^{11}_{0,1})\right)+P_{1|0}\cdot\bar{Y}_{0,1}+P_{1|1}\cdot\bar{Y}_{1,1}\Big],\\
E(Y(1))^{LB} &= \inf_{\pi_{01}\in\mathcal{P}^{**}} \Big[(P_{0|1}-\pi_{01})\cdot y^{LB}-(P_{1|0}-\pi_{01})\cdot\min\left(\bar{Y}_{1,1},\bar{Y}_{0,1}(\max|q^{11}_{0,1})\right)\\
&+ P_{1|0}\cdot\bar{Y}_{0,1}+P_{1|1}\cdot\bar{Y}_{1,1}\Big].
\end{aligned}$$

Therefore, the bounds on the ATEs on the treated, non-treated, and the entire population correspond to

$$\begin{aligned}
\Delta^{UB}_{D=1} &= E(Y|D=1)-\inf_{\pi_{01}\in\mathcal{P}^{**}}\left[\frac{(P_{1|0}-\pi_{01})\cdot y^{LB}+\Pr(Z=1)\cdot P_{0|0}\cdot\bar{Y}_{0,0}+\Pr(Z=0)\cdot P_{0|1}\cdot\bar{Y}_{1,0}}{\Pr(D=1)}\right. \quad (A.29)\\
&- \left.\frac{(P_{0|1}-\pi_{01})\cdot\min\left(\bar{Y}_{0,0},\bar{Y}_{1,0}(\max|q^{00}_{1,0})\right)}{\Pr(D=1)}\right],\\
\Delta^{LB}_{D=1} &= E(Y|D=1)-\sup_{\pi_{01}\in\mathcal{P}^{**}}\left[\frac{(P_{1|0}-\pi_{01})\cdot\frac{P_{0|0}\cdot\bar{Y}_{0,0}-(P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)}{P_{1|1}-P_{1|0}+\pi_{01}}}{\Pr(D=1)}\right. \quad (A.30)\\
&+ \left.\frac{\Pr(Z=1)\cdot P_{0|0}\cdot\bar{Y}_{0,0}+\Pr(Z=0)\cdot P_{0|1}\cdot\bar{Y}_{1,0}}{\Pr(D=1)}-\frac{(P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)}{\Pr(D=1)}\right],
\end{aligned}$$

$$\Delta^{UB}_{D=0} = \sup_{\pi_{01}\in\mathcal{P}^{**}} \left[\frac{(P_{0|1}-\pi_{01})\cdot \frac{P_{1|1}\cdot\bar{Y}_{1,1}-(P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}),\bar{Y}_{0,1}(\min|q^{11}_{0,1})\right)}{P_{1|1}-P_{1|0}+\pi_{01}}}{\Pr(D=0)} \right. \tag{A.31}$$

$$+ \left. \frac{\Pr(Z=1)\cdot P_{1|0}\cdot\bar{Y}_{1,0}+\Pr(Z=0)\cdot P_{1|1}\cdot\bar{Y}_{1,1}}{\Pr(D=0)} - \frac{(P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}),\bar{Y}_{0,1}(\min|q^{11}_{0,1})\right)}{\Pr(D=0)}\right]$$

$$- E(Y|D=0),$$

$$\Delta^{LB}_{D=0} = \inf_{\pi_{01}\in\mathcal{P}^{**}} \left[\frac{(P_{0|1}-\pi_{01})\cdot y^{LB}+\Pr(Z=1)\cdot P_{1|0}\cdot\bar{Y}_{1,0}+\Pr(Z=0)\cdot P_{1|1}\cdot\bar{Y}_{1,1}}{\Pr(D=0)} \right. \tag{A.32}$$

$$- \left. \frac{(P_{1|0}-\pi_{01})\cdot\min\left(\bar{Y}_{1,1},\bar{Y}_{0,1}(\max|q^{11}_{0,1})\right)}{\Pr(D=0)}\right] - E(Y|D=0),$$

and

$$\Delta^{UB} = \sup_{\pi_{01}\in\mathcal{P}^{**}} \left[P_{1|0}\cdot\bar{Y}_{0,1} - (P_{1|0}-\pi_{01})\cdot y^{LB} + P_{1|1}\cdot\bar{Y}_{1,1} - (P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}),\bar{Y}_{0,1}(\min|q^{11}_{0,1})\right)\right.$$

$$+ \left(P_{0|1}-\pi_{01}\right)\cdot\min\left(\bar{Y}_{0,0},\bar{Y}_{1,0}(\max|q^{00}_{1,0})\right) - P_{0|1}\cdot\bar{Y}_{1,0} \tag{A.33}$$

$$+ \left. (P_{0|1}-\pi_{01})\cdot\frac{P_{1|1}\cdot\bar{Y}_{1,1}-(P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}),\bar{Y}_{0,1}(\min|q^{11}_{0,1})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} - P_{0|0}\cdot\bar{Y}_{0,0}\right],$$

$$\Delta^{LB} = \inf_{\pi_{01}\in\mathcal{P}^{**}} \left[P_{1|0}\cdot\bar{Y}_{0,1} + P_{1|1}\cdot\bar{Y}_{1,1}\right. \tag{A.34}$$

$$- (P_{1|0}-\pi_{01})\cdot\frac{P_{0|0}\cdot\bar{Y}_{0,0}-(P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)}{P_{1|1}-P_{1|0}+\pi_{01}}$$

$$- (P_{1|0}-\pi_{01})\cdot\min\left(\bar{Y}_{1,1},\bar{Y}_{0,1}(\max|q^{11}_{0,1})\right) + (P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)$$

$$- \left. P_{0|1}\cdot\bar{Y}_{1,0} + (P_{0|1}-\pi_{01})\cdot y^{LB} - P_{0|0}\cdot\bar{Y}_{0,0}\right].$$

Note that in contrast to the bounds of Section 3.1, mean dominance requires optimizing w.r.t. π_{01}, because any of the previously used y^{UB} has been substituted by the upper bounds on the mean potential outcomes of the compliers.

Lemma 2 shows that under mean dominance, the upper bounds of $E(Y(1)|Z=1,T=11)$ and $E(Y(0)|Z=0,T=00)$ are $\bar{Y}_{1,1}$ and $\bar{Y}_{0,0}$, respectively. Moreover, mean dominance implies that $E(Y(0)|T=11) \le E(Y(0)|T=10)$ and $E(Y(1)|T=00) \le E(Y(1)|T=10)$, thus $E(Y(0)|T=11)^{UB} \le E(Y(0)|T=10)^{UB}$ and $E(Y(1)|T=00)^{UB} \le E(Y(1)|T=10)^{UB}$. The bounds on the ATEs within the four principal strata are special cases of the bounds derived in Section 3.1 under the restrictions just mentioned. Therefore, all of them are sharp.

A.3.4 Proof of the sharpness of the bounds on the mean potential outcomes $E(Y(1))$, $E(Y(0))$

The proofs for the lower bounds are very similar to the ones provided in Appendix A.1.5 (with the only difference that $\bar{Y}_{1,1}(\max|q^{11}_{1,1})$ is replaced by $\bar{Y}_{1,1}$ and $\bar{Y}_{0,0}(\max|q^{00}_{0,0})$ is replaced by $\bar{Y}_{0,0}$ in the definitions of densities $\{h_t\}$ and coefficients $\{\alpha\}$) and therefore omitted. Hence, we only consider the upper bounds of $E(Y(1))$, $E(Y(0))$ for a given π_{01}:

$$\begin{aligned} E(Y(0))^{UB}(\pi_{01}) &= (P_{1|0}-\pi_{01})\cdot\frac{P_{1|1}\cdot\bar{Y}_{1,1}-(P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}),\bar{Y}_{0,1}(\min|q^{11}_{0,1})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \\ &- (P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)+P_{0|0}\cdot\bar{Y}_{0,0}+P_{0|1}\cdot\bar{Y}_{1,0}, \end{aligned}$$

and

$$\begin{aligned} E(Y(1))^{UB}(\pi_{01}) &= (P_{0|1}-\pi_{01})\cdot\frac{P_{0|0}\cdot\bar{Y}_{0,0}-(P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)}{P_{1|1}-P_{1|0}+\pi_{01}} \\ &- (P_{1|0}-\pi_{01})\cdot\max\left(\bar{Y}_{1,1}(\min|q^{11}_{1,1}),\bar{Y}_{0,1}(\min|q^{11}_{0,1})\right)+P_{1|0}\cdot\bar{Y}_{0,1}+P_{1|1}\cdot\bar{Y}_{1,1}. \end{aligned}$$

By assumption, $E(Y(0)|T=11)^{UB}\leq E(Y(0)|T=10)^{UB}=\frac{P_{0|0}\cdot\bar{Y}_{0,0}-(P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)}{P_{1|1}-P_{1|0}+\pi_{01}}$ and $E(Y(1)|T=00)^{UB}\leq E(Y(1)|T=10)^{UB}=\frac{P_{0|0}\cdot\bar{Y}_{0,0}-(P_{0|1}-\pi_{01})\cdot\max\left(\bar{Y}_{0,0}(\min|q^{00}_{0,0}),\bar{Y}_{1,0}(\min|q^{00}_{1,0})\right)}{P_{1|1}-P_{1|0}+\pi_{01}}$. Therefore, the bounds are valid.

To show that they are sharp, we need to show that for $T=11,10,01,00$ and $Z=1,0$, there exist distributions of T given Z and of $(Y(1),Y(0))$ given T and Z that are compatible with a data generating process satisfying Assumptions 2 and 4 and implying $E(Y(0))=E(Y(0))^{UB}$ and $E(Y(1))=E(Y(1))^{UB}$.

We reconsider (A.10) and (A.11) and change the marginal distribution of h^1_{11} and h^0_{11} with respect to $y(0)$ by replacing $I\{y(0)=y^{UB}\}$ with the marginal distributions of h^z_{10} with respect to $y(0)$. We also modify the marginal distribution of h^1_{00} and h^0_{00} with respect to $y(1)$ by replacing $I\{y(1)=y^{UB}\}$ by the marginal distributions of h^z_{10} with respect to $y(1)$.

This choice of $\Pr(T|Z)$ and $\{h^z_t\}$ satisfies Assumption 2 by construction. Note that $\iint y(0)h_{11}\,dy(1)\,dy(0)=\iint y(0)h_{10}\,dy(1)\,dy(0)$ and $\iint y(1)h_{00}\,dy(1)\,dy(0)=\iint y(1)h_{10}\,dy(1)\,dy(0)$ imply $E(Y(0)|T=11)=E(Y(0)|T=10)$ and $E(Y(1)|T=00)=E(Y(1)|T=10)$. Therefore, Assumption 4 is also satisfied. Since all equations in (A.12) are still valid because the marginal distributions that we replaced do not affect restriction (A.12), this distributional choice is compatible with the data generating process. In an analogous way as in A.1.5, it is easy to see that the distributions imply $E(Y(0))=E(Y(0))^{UB}$ and $E(Y(1))=E(Y(1))^{UB}$. This shows that the provided bounds are sharp for a given value of π_{01}. The sharp upper bounds over all admissible π_{01} are obtained by taking the supremum w.r.t. to π_{01}. Because of the continuity of the bounds in π_{01} (see Appendix A.1.4) for a compact set $\mathcal{P}^{**}$, the maximum is attained by the extreme value theorem.

A.4 Monotonicity and mean dominance

A.4.1 Proof of the sharpness of the bounds for always takers

Mean dominance implies that $E(Y(0)|T = 11) \leq E(Y(0)|T = 10) = \frac{P_{0|0} \cdot \bar{Y}_{0,0} - P_{0|1} \cdot \bar{Y}_{1,0}}{P_{1|1} - P_{1|0}}$. Thus, $E(Y(0)|T = 11)^{LB} = \frac{P_{0|0} \cdot \bar{Y}_{0,0} - P_{0|1} \cdot \bar{Y}_{1,0}}{P_{1|1} - P_{1|0}}$ and Δ_{11}^{LB} is the sharp lower bound of Δ_{11}.

A.4.2 Proof of the sharpness of the bounds for never takers

Mean dominance implies that $E(Y(1)|T = 00) \leq E(Y(1)|T = 10) = \frac{P_{1|1} \cdot \bar{Y}_{1,1} - P_{1|0} \cdot \bar{Y}_{0,1}}{P_{1|1} - P_{1|0}}$. Thus, $E(Y(1)|T = 00)^{LB} = \frac{P_{1|1} \cdot \bar{Y}_{1,1} - P_{1|0} \cdot \bar{Y}_{0,1}}{P_{1|1} - P_{1|0}}$ and Δ_{00}^{UB} is the sharp upper bound of Δ_{00}.

A.4.3 Proof of the sharpness of the bounds on the mean potential outcomes $E(Y(1))$, $E(Y(0))$

Mean dominance implies $E(Y(0)|T = 11)^{LB} = \frac{P_{0|0} \cdot \bar{Y}_{0,0} - P_{0|1} \cdot \bar{Y}_{1,0}}{P_{1|1} - P_{1|0}}$, $E(Y(1)|T = 00)^{LB} = \frac{P_{1|1} \cdot \bar{Y}_{1,1} - P_{1|0} \cdot \bar{Y}_{0,1}}{P_{1|1} - P_{1|0}}$. Monotonicity implies

$$\begin{aligned} E(Y(0)) &= P_{1|0} \cdot E(Y(0)|T = 11) + P_{0|0} \cdot \bar{Y}_{0,0}, \\ E(Y(1)) &= P_{0|1} \cdot E(Y(1)|T = 00) + P_{1|1} \cdot \bar{Y}_{1,1}. \end{aligned}$$

Therefore,

$$\begin{aligned} E(Y(0))^{LB} &= P_{1|0} \cdot \frac{P_{0|0} \cdot \bar{Y}_{0,0} - P_{0|1} \cdot \bar{Y}_{1,0}}{P_{1|1} - P_{1|0}} + P_{0|0} \cdot \bar{Y}_{0,0}, \\ E(Y(1))^{LB} &= P_{0|1} \cdot \frac{P_{1|1} \cdot \bar{Y}_{1,1} - P_{1|0} \cdot \bar{Y}_{0,1}}{P_{1|1} - P_{1|0}} + P_{1|1} \cdot \bar{Y}_{1,1}, \end{aligned}$$

are sharp.

A.5 Discrete outcomes

A.5.1 Identification

If the outcome Y is discrete, the bounds based on Proposition 4 in Horowitz and Manski (1995) are generally not valid. This is due to the presence of ties in the outcome, i.e. the occurrence of mass points with equal outcome values, which entails a non-unique quantile function such that a particular outcome value is observed at several ranks. The quantile function is required to construct (i) a distribution which is stochastically dominated by any feasible distribution that is consistent with the identification region of some mixture component (i.e., stratum) of interest to determine the lower bound of its mean outcome and (ii) a distribution which stochastically dominates any distribution consistent with the identification to determine the upper bound, respectively. In the presence of discrete outcomes we have to replace the non-unique quantile function, which gives equal ranks to all ties, by a modified version which accounts for ties.

To this end, we denote by $Y_{z,d}$ the outcome variable in the respective observed group and introduce the following trimming functions which are similar to the ones proposed by Kitagawa (2009)

$$\begin{aligned} Trim_{z,d}^{\min,t} &= \frac{I\{Y_{z,d} < F^{-1}(q_{z,d}^t)\} \cdot \Pr(Y_{z,d} < F^{-1}(q_{z,d}^t))}{1-q_{z,d}^t} \\ &+ \frac{I\{Y_{z,d} = F^{-1}(q_{z,d}^t)\} \cdot \big(\Pr(Y_{z,d} \geq F^{-1}(q_{z,d}^t)) - q_{z,d}^t\big)}{1-q_{z,d}^t} \\ Trim_{z,d}^{\max,t} &= \frac{I\{Y_{z,d} > F^{-1}(1-q_{z,d}^t)\} \cdot \Pr(Y_{z,d} > F^{-1}(1-q_{z,d}^t))}{q_{z,d}^t} \\ &+ \frac{I\{Y_{z,d} = F^{-1}(1-q_{z,d}^t)\} \cdot \big(\Pr(Y_{z,d} \leq F^{-1}(1-q_{z,d}^t)) - (1-q_{z,d}^t)\big)}{q_{z,d}^t} \end{aligned} \tag{A.35}$$

Let $Y_{z,d}^{\min,t} = Y_{z,d} \cdot Trim_{z,d}^{\min,t}$ and $Y_{z,d}^{\max,t} = Y_{z,d} \cdot Trim_{z,d}^{\min,t}$. Since $F_{Y_{z,d}^{\min,t}}(y)$ is stochastically dominated by any feasible distribution consistent with the identification region and $F_{Y_{z,d}^{\max,t}}(y)$ stochastically dominates any feasible distribution consistent with the identification region, we can define the lower and upper bounds of $E(Y(d)|Z = z, T = t)$ as $\bar{Y}_{z,d}(\min|q_{z,d}^t) = E(Y_{z,d}^{\min,t})$ and $\bar{Y}_{z,d}(\max|q_{z,d}^t) = E(Y_{z,d}^{\max,t})$, respectively.

A.6 Estimation

The trimming function can be easily estimated by its sample analog. A more intuitive and asymptotically equivalent estimator can be obtained as follows. Denote by $n_{z,d}$ the number of observations with $Z = z$ and $D = d$. Let $Y_{z,d}^{(1)}, \ldots, Y_{z,d}^{(n_{z,d})}$ be the order statistic of $Y_{z,d}$ which assigns increasing ranks at the ties. E.g., if there are two observations i and $j \in \{1, ..., n_{z,d}\}$ with the same outcome values y, the rank of the first one, denoted by the function $rank(\cdot)$, will be some integer i while the rank of the second observation will be $i + 1$. Define $\widetilde{q_{z,d}^t \cdot n_{z,d}}$ and $\widetilde{n_{z,d} - q_{z,d}^t} \cdot n_{z,d}$ to be the integer part of $q_{z,d}^t \cdot n_{z,d}$ and $n_{z,d} - q_{z,d}^t \cdot n_{z,d}$, respectively, if $q_{z,d}^t \cdot n_{z,d}$ and $n_{z,d} - q_{z,d}^t \cdot n_{z,d}$ are larger than 1, while they are 1 otherwise.

The only difference with Section 4 is that the trimmed means are now estimated by

$$\begin{aligned} \hat{\bar{Y}}_{z,d}(\max|q_{z,d}^t) &= \frac{\sum_{i=1}^n Y_i \cdot I\{D_i = d\} \cdot I\{Z_i = z\} \cdot I\{rank(Y_i) \geq n_{z,d} \widetilde{- q_{z,d}^t} \cdot n_{z,d}\}}{\sum_{i=1}^n I\{D_i = d\} \cdot I\{Z_i = z\} \cdot I\{rank(Y_i) \geq n_{z,d} \widetilde{- q_{z,d}^t} \cdot n_{z,d}\}}, \\ \hat{\bar{Y}}_{z,d}(\min|q_{z,d}^t) &= \frac{\sum_{i=1}^n Y_i \cdot I\{D_i = d\} \cdot I\{Z_i = z\} \cdot I\{rank(Y_i) \leq \widetilde{q_{z,d}^t \cdot n_{z,d}}\}}{\sum_{i=1}^n I\{D_i = d\} \cdot I\{Z_i = z\} \cdot I\{rank(Y_i) \leq \widetilde{q_{z,d}^t \cdot n_{z,d}}\}}. \end{aligned}$$

Since we just replace the quantile function by the modified rank function, this only results in using a different indicator function when computing the trimmed means. Note that if Y is continuous, defining the bounds in terms of the modified rank function is equivalent to defining them in terms of the regular quantile function, as each outcome value has a unique rank in this case. Therefore, the two approaches converge to each other as the number of support points of Y goes to infinity.

A.7 Inference based on Chernozhukov, Lee, and Rosen (2009)

Recall from Section 4 that $\Delta_t^{LB}(\pi_{01}, v)$, $\Delta_t^{UB}(\pi_{01}, v)$ are the conditional bounds on the ATE in some subpopulation t given π_{01} and v[20] and that the identification region of Δ_t (the unconditional ATE in the subpopulation) is obtained by optimizing over admissible values of $\pi_{01}\in\mathcal{P}$ and $v \in V = \{1, 2, 3, 4\}$:

$$\inf_{\pi_{01}\in\mathcal{P}}\{\max_{v\in V}[\Delta_t^{LB}(\pi_{01}, v)]\} \leq \Delta_t \leq \sup_{\pi_{01}\in\mathcal{P}}\{\min_{v\in V}[\Delta_t^{UB}(\pi_{01}, v)]\},$$

We subsequently discuss the estimation of the upper bound along with its confidence region (the proceeding for the lower bound is analogous) under Assumptions 1 and 2 or 1, 2, and 4, where non-differentiability complicates inference. We use the procedure of Chernozhukov et al. (2013) to obtain a half-median-unbiased estimator of $\min_{v\in V}[\Delta^{UB}(\pi_{01}, v)]$ conditional on π_{01}, see also the application of this method in Chen and Flores (2012).

The main idea is that instead of taking the minimum of the estimated upper bounds $\hat{\Delta}_t^{UB}(\pi_{01}, v)$ directly, one uses the following precision adjusted version, denoted by $\tilde{\Delta}_t^{UB}(\pi_{01}, p)$, which consists of the initial estimate plus $s(v)$, a measure of the precision of $\hat{\Delta}_t^{UB}(\pi_{01}, v)$, times an appropriate critical value $k(p)$:

$$\tilde{\Delta}_t^{UB}(\pi_{01}, p) = \min_{v\in V}[\hat{\Delta}_t^{UB}(\pi_{01}, v) + k(p) \cdot s(v)].$$

Define $\boldsymbol{\Delta}_t^{UB}(\pi_{01})$ to be the four dimensional vector with elements $\Delta_t^{UB}(\pi_{01}, v)$, $v = 1, 2, 3, 4$. As outlined below, $k(p)$ is a function of the sample size and the estimated variance-covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\Delta}}_t^{UB}(\pi_{01}) - \boldsymbol{\Delta}_t^{UB}(\pi_{01}))$, denoted by $\hat{\Omega}$. For $p = \frac{1}{2}$, the estimator $\tilde{\Delta}_t^{UB}(\pi_{01}, p)$ is half-median-unbiased, which implies that the estimate of the upper bound exceeds its true value with probability at least one half asymptotically. Finally, by taking the supremum of $\tilde{\Delta}_t^{UB}(\pi_{01}, p)$ over π_{01}, we obtain a conservative estimate for Δ_t^{UB}.

The following algorithm briefly sketches the estimation of Δ_t^{UB} along with its upper confidence band based on the precision adjustment.

[20] $v = \begin{cases} 1 & \text{if } z = 1, z' = 1 \\ 2 & \text{if } z = 1, z' = 0 \\ 3 & \text{if } z = 0, z' = 1 \\ 4 & \text{if } z = 0, z' = 0 \end{cases}.$

1. Estimate the vector $\hat{\mathbf{\Delta}}_t^{UB}(\pi_{01})$ by its sample analog. Estimate the variance-covariance matrix $\hat{\Omega}$ by bootstrapping B times.
2. Denoting by $\hat{g}(v)^\top$ the v-th row of $\hat{\Omega}^{\frac{1}{2}}$, estimate $\hat{s}(v) = \frac{\|\hat{g}(v)\|}{\sqrt{n}}$, where $\|\cdot\|$ is the Euclidean norm.
3. Simulate R draws, $H_1, \ldots, H_R$ from a $\mathcal{N}(\mathbf{0}, I_4)$, where $\mathbf{0}$ and I_4 are the null vector and the identity matrix of dimension 4, respectively.
4. Let $H_r^*(v) = \hat{g}(v)^\top Z_r / \|\hat{g}(v)\|$ for $r = 1, \ldots, R$.
5. Let $\tilde{k}(c)$ be the c-th quantile of $\min_{v \in V} H_r^*(v), r = 1, \ldots, R$, where $c = 1 - \frac{0.1}{\log(n)}$.
6. Compute the set estimator $\hat{V} = \{v \in V : \hat{\Delta}^{UB}(\pi_{01}, v) \leq \min_{v' \in V}\{[\hat{\Delta}_t^{UB}(\pi_{01}, v') + \tilde{k}(c) \cdot \hat{s}(v')] + 2 \cdot \tilde{k}(c) \cdot \hat{s}(v')\}\}$.
7. Estimate the critical value $\hat{k}(p)$ by the p-th quantile of $\min_{v \in \hat{V}} H_r^*(v), r = 1, \ldots, R$.
8. For half-median-unbiasedness, set $p = \frac{1}{2}$ and compute $\tilde{\Delta}_t^{UB}(\pi_{01}, \frac{1}{2}) = \min_{v \in V}[\hat{\Delta}_t^{UB}(\pi_{01}, v) + \hat{k}(\frac{1}{2}) \cdot \hat{s}(v)]$.
9. Estimate the upper bound Δ_t^{UB} by $\hat{\Delta}_t^{UB} = \sup_{\pi_{01} \in \mathcal{P}}[\tilde{\Delta}_t^{UB}(\pi_{01}, \frac{1}{2})]$.
10. To obtain the upper confidence band, estimate the half-median-unbiased lower bound given π_{01}, $\tilde{\Delta}_t^{LB}(\pi_{01}, \frac{1}{2})$.
11. Let $\Gamma = \max(0, \tilde{\Delta}_t^{UB}(\pi_{01}, \frac{1}{2}) - \tilde{\Delta}_t^{LB}(\pi_{01}, \frac{1}{2}))$, $\rho = \max(\tilde{\Delta}_t^{UB}(\pi_{01}, \frac{3}{4}) - \tilde{\Delta}_t^{UB}(\pi_{01}, \frac{1}{4}), \tilde{\Delta}_t^{LB}(\pi_{01}, \frac{3}{4}) - \tilde{\Delta}_t^{LB}(\pi_{01}, \frac{1}{4}))$ and $\tau = (\rho \cdot \log(n))^{-1}$. Compute $\hat{a} = 1 - \Phi(\tau \cdot \Gamma) \cdot \alpha$, where α is the chosen confidence level.
12. The upper confidence band for the estimate of Δ_t^{UB} is obtained by $\sup_{\pi_{01} \in \mathcal{P}}[\tilde{\Delta}_t^{UB}(\pi_{01}, \hat{a})]$.

A.8 Bounding the ATEs on the treated with $Z = 1$ and $Z = 0$

In the subsequent sections[21], we derive bounds on $\Delta_{D=1,Z=1} = E(Y(1) - Y(0)|D = 1, Z = 1)$, the ATE among those treated receiving the instrument (consisting of always takers and compliers), and on $\Delta_{D=1,Z=0} = E(Y(1) - Y(0)|D = 1, Z = 0)$, the ATE among those treated not receiving the instrument (consisting of defiers and never takers) under various assumptions. These parameters appear interesting in the context of experimentally evaluated encouragement designs aiming at increasing treatment participation (e.g. by advertisement). $\Delta_{D=1,Z=1}$ then gives the effect on the treated that have been encouraged ($Z = 1$) to take the treatment, whereas $\Delta_{D=1,Z=0}$ is the impact on those who are treated despite of not being encouraged ($Z = 0$). This allows assessing whether/how the ATE on the treated differs with and without encouragement (however, bearing in mind that the treated are not identical under $Z = 1$ and $Z = 0$), which may be relevant for deciding whether an encouragement should be provided to everybody or nobody in some population of interest.

[21] We would like to thank an anonymous referee for suggesting those parameters.

A.8.1 Mean independence within principal strata without further assumptions

To derive bounds on $\Delta_{D=1,Z=1} = E(Y(1)-Y(0)|D=1,Z=1)$, note that $E(Y(1)|D=1,Z=1) = E(Y|D=1,Z=1) = \bar{Y}_{1,1}$. Therefore, we only need to bound

$$\begin{aligned}
E(Y(0)|D=1,Z=1) &= \frac{\pi_{11}}{P_{1|1}} \cdot E(Y(0)|T=11) + \frac{\pi_{10}}{P_{1|0}} \cdot E(Y(0)|T=10) \\
&= \frac{(P_{1|0}-\pi_{01}) \cdot E(Y(0)|T=11) + P_{0|0} \cdot \bar{Y}_{0,0} - (P_{0|1}-\pi_{01}) \cdot E(Y(0)|T=00)}{P_{1|1}}.
\end{aligned}$$

The sharp bounds on $\Delta_{D=1,Z=1}$ are given by

$$\begin{aligned}
\Delta^{UB}_{D=1,Z=1} &= \bar{Y}_{1,1} - \frac{(P_{1|0}-\pi^{\min}_{01}) \cdot y^{LB} + P_{0|0} \cdot \bar{Y}_{0,0}}{P_{1|1}} \qquad \text{(A.36)} \\
&+ \frac{(P_{0|1}-\pi^{\min}_{01}) \cdot \min\Big(\bar{Y}_{0,0}(\max|q^{00,\pi^{\min}_{01}}_{0,0}), \bar{Y}_{1,0}(\max|q^{00,\pi^{\min}_{01}}_{1,0})\Big)}{P_{1|1}}, \\
\Delta^{LB}_{D=1,Z=1} &= \bar{Y}_{1,1} - \frac{(P_{1|0}-\pi^{\min}_{01}) \cdot y^{UB} + P_{0|0} \cdot \bar{Y}_{0,0}}{P_{1|1}} \qquad \text{(A.37)} \\
&+ \frac{(P_{0|1}-\pi^{\min}_{01}) \cdot \max\Big(\bar{Y}_{0,0}(\min|q^{00,\pi^{\min}_{01}}_{0,0}), \bar{Y}_{1,0}(\min|q^{00,\pi^{\min}_{01}}_{1,0})\Big)}{P_{1|1}}.
\end{aligned}$$

To see this, note that (A.10) and (A.11) imply $\Delta_{D=1,Z=1} = \Delta^{UB}_{D=1,Z=1}(\pi_{01})$. In an analogous manner as in Appendix A.1.5, one can show that $\sup_{\pi_{01}\in\mathcal{P}} \Delta^{UB}_{D=1,Z=1}(\pi_{01}) = \Delta^{UB}_{D=1,Z=1}(\pi^{\min}_{01})$.

Concerning $\Delta_{D=1,Z=0} = E(Y(1)-Y(0)|D=1,Z=0)$, note that $E(Y(1)|D=1,Z=0) = E(Y|D=1,Z=0) = \bar{Y}_{0,1}$, so that we only need to bound

$$\begin{aligned}
E(Y(0)|D=1,Z=0) &= \frac{\pi_{11}}{P_{1|0}} \cdot E(Y(0)|T=11) + \frac{\pi_{01}}{P_{1|0}} \cdot E(Y(0)|T=01) \\
&= \frac{(P_{1|0}-\pi_{01}) \cdot E(Y(0)|T=11) + P_{0|1} \cdot \bar{Y}_{1,0} - (P_{0|1}-\pi_{01}) \cdot E(Y(0)|T=00)}{P_{1|0}}.
\end{aligned}$$

In the same way as for $\Delta^{UB}_{D=1,Z=1}$, one can show that the sharp bounds on $\Delta_{D=1,Z=1}$ are given by

$$\begin{aligned}
\Delta^{UB}_{D=1,Z=0} &= \bar{Y}_{0,1} - \frac{(P_{1|0}-\pi^{\min}_{01}) \cdot y^{LB} + P_{0|1} \cdot \bar{Y}_{0,1}}{P_{1|0}} \qquad \text{(A.38)} \\
&+ \frac{(P_{0|1}-\pi^{\min}_{01}) \cdot \min\Big(\bar{Y}_{0,0}(\max|q^{00,\pi^{\min}_{01}}_{0,0}), \bar{Y}_{1,0}(\max|q^{00,\pi^{\min}_{01}}_{1,0})\Big)}{P_{1|0}}, \\
\Delta^{LB}_{D=1,Z=0} &= \bar{Y}_{0,1} - \frac{(P_{1|0}-\pi^{\min}_{01}) \cdot y^{UB} + P_{0|1} \cdot \bar{Y}_{1,0}}{P_{1|0}} \qquad \text{(A.39)} \\
&+ \frac{(P_{0|1}-\pi^{\min}_{01}) \cdot \max\Big(\bar{Y}_{0,0}(\min|q^{00,\pi^{\min}_{01}}_{0,0}), \bar{Y}_{1,0}(\min|q^{00,\pi^{\min}_{01}}_{1,0})\Big)}{P_{1|0}}.
\end{aligned}$$

A.8.2 Monotonicity

Under monotonicity, $\Delta_{D=1,Z=0} = \Delta_{11}$ because defiers do not exist. Concerning $\Delta_{D=1,Z=1}$, $\pi_{01} = 0$ and $E(Y(0)|T = 00) = \bar{Y}_{1,0}$ imply that the bounds simplify to

$$\Delta^{UB}_{D=1,Z=1} = \frac{P_{1|1} \cdot \bar{Y}_{1,1} + P_{0|1} \cdot \bar{Y}_{1,0} - P_{1|0} \cdot y^{LB} - P_{0|0} \cdot \bar{Y}_{0,0}}{P_{1|1}}, \tag{A.40}$$

$$\Delta^{LB}_{D=1,Z=1} = \frac{P_{1|1} \cdot \bar{Y}_{1,1} + P_{0|1} \cdot \bar{Y}_{1,0} - P_{1|0} \cdot y^{UB} - P_{0|0} \cdot \bar{Y}_{0,0}}{P_{1|1}}. \tag{A.41}$$

Sharpness follows immediately.

A.8.3 Mean dominance

Using a similar argument as the one used for the bounds under Assumption 2 alone, one can show that the following bounds are sharp.

$$\begin{aligned}\Delta^{UB}_{D=1,Z=1} &= \sup_{\pi_{01}\in\mathcal{P}} \Bigg[\bar{Y}_{1,1} - \frac{(P_{1|0} - \pi^{\min}_{01}) \cdot y^{LB} + P_{0|0} \cdot \bar{Y}_{0,0}}{P_{1|1}} \qquad (A.42)\\ &+ \frac{(P_{0|1} - \pi^{\min}_{01}) \cdot \min\left(\bar{Y}_{0,0}(\max|q^{00,\pi^{\min}_{01}}_{0,0}), \bar{Y}_{1,0}(\max|q^{00,\pi^{\min}_{01}}_{1,0})\right)}{P_{1|1}}\Bigg],\end{aligned}$$

$$\begin{aligned}\Delta^{LB}_{D=1,Z=1} &= \inf_{\pi_{01}\in\mathcal{P}} \Bigg[\bar{Y}_{1,1} - \frac{(P_{1|0} - \pi^{\min}_{01}) \cdot y^{UB} + P_{0|0} \cdot \bar{Y}_{0,0}}{P_{1|1}} \qquad (A.43)\\ &+ \frac{(P_{0|1} - \pi^{\min}_{01}) \cdot \max\left(\bar{Y}_{0,0}(\min|q^{00,\pi^{\min}_{01}}_{0,0}), \bar{Y}_{1,0}(\min|q^{00,\pi^{\min}_{01}}_{1,0})\right)}{P_{1|1}}\Bigg].\end{aligned}$$

$$\begin{aligned}\Delta^{UB}_{D=1,Z=0} &= \sup_{\pi_{01}\in\mathcal{P}} \Bigg[\bar{Y}_{0,1} - \frac{(P_{1|0} - \pi^{\min}_{01}) \cdot y^{LB} + P_{0|1} \cdot \bar{Y}_{0,1}}{P_{1|0}} \qquad (A.44)\\ &+ \frac{(P_{0|1} - \pi^{\min}_{01}) \cdot \min\left(\bar{Y}_{0,0}(\max|q^{00,\pi^{\min}_{01}}_{0,0}), \bar{Y}_{1,0}(\max|q^{00,\pi^{\min}_{01}}_{1,0})\right)}{P_{1|0}}\Bigg],\end{aligned}$$

$$\begin{aligned}\Delta^{LB}_{D=1,Z=0} &= \inf_{\pi_{01}\in\mathcal{P}} \Bigg[\bar{Y}_{0,1} - \frac{(P_{1|0} - \pi^{\min}_{01}) \cdot y^{UB} + P_{0|1} \cdot \bar{Y}_{1,0}}{P_{1|0}} \qquad (A.45)\\ &+ \frac{(P_{0|1} - \pi^{\min}_{01}) \cdot \max\left(\bar{Y}_{0,0}(\min|q^{00,\pi^{\min}_{01}}_{0,0}), \bar{Y}_{1,0}(\min|q^{00,\pi^{\min}_{01}}_{1,0})\right)}{P_{1|0}}\Bigg].\end{aligned}$$

A.8.4 Monotonicity and mean dominance

Under both monotonicity and mean dominance, $\Delta_{D=1,Z=0} = \Delta_{11}$ and $\Delta^{UB}_{D=1,Z=1}$ are the same as under monotonicity alone. Concerning $\Delta^{LB}_{D=1,Z=1}$, note that $E(Y(0)|T = 11)^{UB} \leq E(Y(0)|T = 10) =$

$\frac{P_{0|0} \cdot \bar{Y}_{0,0} - P_{0|1} \cdot \bar{Y}_{1,0}}{P_{1|1} - P_{1|0}}$. Therefore,

$$\Delta^{LB}_{D=1,Z=1} \quad = \quad \frac{P_{1|1} \cdot \bar{Y}_{1,1} + P_{0|1} \cdot \bar{Y}_{1,0} - P_{1|0} \cdot \frac{P_{0|0} \cdot \bar{Y}_{0,0} - P_{0|1} \cdot \bar{Y}_{1,0}}{P_{1|1} - P_{1|0}} - P_{0|0} \cdot \bar{Y}_{0,0}}{P_{1|1}}. \tag{A.46}$$

Sharpness follows immediately.

References

Angrist, J., Bettinger, E., Bloom, E., King, E. and Kremer, M.: 2002, Vouchers for private schooling in colombia: Evidence from a randomized natural experiment, The American Economic Review **92**, 1535–1558.

Angrist, J., Imbens, G. and Rubin, D.: 1996, Identification of causal effects using instrumental variables, Journal of American Statistical Association **91**, 444–472 (with discussion).

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

Bhattacharya, J., Shaikh, A. M. and Vytlacil, E.: 2008, Treatment effect bounds under monotonicity assumptions: An application to swan-ganz catheterization, The American Economic Review pp. 351–356.

Bloom, H. S.: 1984, Accounting for no-shows in experimental evaluation designs, Evaluation Review **8**, 225–246.

Blundell, R., Gosling, A., Ichimura, H. and Meghir, C.: 2007, Changes in the distribution of male and female wages accounting for employment composition using bounds, Econometrica **75**, 323–363.

Chen, L.-Y. and Szroeter, J.: 2014, Testing multiple inequality hypotheses: a smoothed indicator approach, Journal of Econometrics **178**, 678–693.

Chen, X. and Flores, C. A.: 2012, Bounds on treatment effects in the presence of sample selection and noncompliance: The wage effects of job corps, Working Paper .

Cheng, J. and Small, D. S.: 2006, Bounds on causal effects in three-arm trials with non-compliance, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **68**(815-836).

Chernozhukov, V., Lee, S. and Rosen, A.: 2013, Intersection bounds: Estimation and inference, Econometrica **81**, 667–737.

Chiburis, R. C.: 2010, Semiparametric bounds on treatment effects, Journal of Econometrics **159**.

Dawid, A. P.: 2003, Causal inference using influence diagrams: The problem of partial compliance (with discussion), in P. J. Green, N. L. Hjort and S. Richardson (eds), Highly Structured Stochastic Systems, Oxford University Press.

Deaton, A. S.: 2010, Instruments, randomization, and learning about development, Journal of Economic Literature **48**, 424–455.

Flores, C. A. and Flores-Lagunes, A.: 2013, Partial identification of local average treatment effects with an invalid instrument, Journal of Business and Economic Statistics **31**, 534–545.

Frangakis, C. E. and Rubin, D. B.: 2002, Principal stratification in causal inference, Biometrics **58**, 21–29.

Freyberger, J. and Horowitz, J.: 2013, Identification and shape restrictions in nonparametric instrumental variables estimation, Technical report, cemmap working paper, Centre for Microdata Methods and Practice.

Frölich, M.: 2007, Nonparametric iv estimation of local average treatment effects with covariates, Journal of Econometrics **139**, 35–75.

Frölich, M. and Lechner, M.: 2010, Combining matching and nonparametric iv estimation: Theory and an application to the evaluation of active labour market policies, University of St Gallen, Department of Economics Discussion Paper No. 2010-21 .

Heckman, J. J., LaLonde, R. and Smith, J.: 1999, The economics and econometrics of active labour market programs, in O. Ashenfelter and D. Card (eds), The Handbook of Labor Economics, North-Holland, New York, pp. 1865–2097.

Heckman, J. J. and Urzúa, S.: 2010, Comparing iv with structural models: What simple iv can and cannot identify, Journal of Econometrics **156**, 27–37.

Heckman, J. J. and Vytlacil, E.: 2001, Instrumental variables, selection models, and tight bounds on the average treatment effects, in M. Lechner and M. Pfeiffer (eds), Econometric Evaluation of Labour Market Policies, Center for European Economic Research, New York, pp. 1–15.

Hirano, K. and Porter, J. R.: 2012, Impossibility results for nondifferentiable functionals, Econometrica **80**, 1769–1790.

Horowitz, J. L. and Manski, C. F.: 1995, Identification and robustness with contaminated and corrupted data, Econometrica **63**, 281–302.

Huber, M. and Mellace, G.: 2013a, Sharp bounds on causal effects under sample selection, forthcoming in the Oxford Bulletin of Economics and Statistics .

Huber, M. and Mellace, G.: 2013b, Testing instrument validity for late identification based on inequality moment constraints, forthcoming in the Review of Economics and Statistics .

Imai, K.: 2008, Sharp bounds on the causal effects in randomized experiments with 'truncation-by-death', Statistics & Probability Letters **78**, 144–149.

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

Imbens, G. W.: 2009, Better late than nothing: Some comments on deaton (2009) and heckman and urzua (2009), NBER Working Paper No. 14896 .

Imbens, G. W. and Angrist, J.: 1994, Identification and estimation of local average treatment effects, Econometrica **62**, 467–475.

Imbens, G. W. and Manski, C. F.: 2004, Confidence intervals for partially identified parameters, Econometrica **72**(6), 1845–1857.

King, E., Rawlings, L., Gutierrez, M., Pardo, C. and Torres, C.: 1997, Colombia's targeted education voucher program: Features, coverage and participation, Working Paper No. 3, Series on Impact Evaluation of Education Reforms, Development Economics Research Group, The World Bank .

Kitagawa, T.: 2009, Identification region of the potential outcome distribution under instrument independence, CeMMAP working paper 30/09 .

Kitagawa, T.: 2014, A test for instrument validity, CeMMAP working paper 34/14 .

Laffërs, L.: 2013, A note on bounding average treatment effects, Economics Letters **120**(3), 424–428.

Lechner, M. and Melly, B.: 2007, Earnings effects of training programs, IZA Discussion Paper no. 2926 .

Lee, D. S.: 2009, Training, wages, and sample selection: Estimating sharp bounds on treatment effects, Review of Economic Studies **76**, 1071–1102.

Manski, C. F.: 1989, Anatomy of the selection problem, The Journal of Human Resources **24**, 343–360.

Manski, C. F.: 1990, Nonparametric bounds on treatment effects, American Economic Review, Papers and Proceedings **80**, 319–323.

Manski, C. F.: 1994, The selection problem, in C. Sims. (ed.), Advances in Econometrics: Sixth World Congress, Cambridge University Press, pp. 143–170.

Richardson, T. S. and Robins, J. M.: 2010, Analysis of the binary instrumental variable model, in R. Dechter, H. Geffner and J. Y. Halpern (eds), Heuristics, probability and causality: a tribute to Judea Pearl, College Publications, London, UK, pp. 415–440.

Robins, J.: 1989, The analysis of randomized and non-randomizedaids treatment trials using a new approach to causal inference in longitudinal studies, in L. Sechrest, H. Freeman and A. Mulley (eds), Health Service Research Methodology: A Focus on AIDS, U.S. Public Health Service, Washington, DC, pp. 113–159.

Rubin, D. B.: 1974, Estimating causal effects of treatments in randomized and nonrandomized studies, Journal of Educational Psychology **66**, 688–701.

Rubin, D. B.: 1990, Formal modes of statistical inference for causal effects, Journal of Statistical Planning and Inference **25**, 279–292.

Shaikh, A. M. and Vytlacil, E. J.: 2011, Partial identification in triangular systems of equations with binary dependent variables, Econometrica **79**(3), 949–955.

Stoye, J.: 2009, More on confidence intervals for partially identified parameters, Econometrica **77**, 1299–1315.

Vytlacil, E.: 2002, Independence, monotonicity, and latent index models: An equivalence result, Econometrica **70**, 331–341.

Zhang, J. and Rubin, D. B.: 2003, Estimation of causal effects via principal stratification when some outcome are truncated by death, Journal of Educational and Behavioral Statistics **28**, 353–368.

Zhang, J., Rubin, D. B. and Mealli, F.: 2008, Evaluating the effects of job training programs on wages through principal stratification, in D. Millimet, J. Smith and E. Vytlacil (eds), Advances in Econometrics: Modelling and Evaluating Treatment Effects in Econometrics, Vol. 21, Elsevier Science Ltd., pp. 117–145.