Identification of the average treatment effect when SUTVA is violated

by

Lukáš Laffers and Giovanni Mellace

Discussion Papers on Business and Economics
No. 3/2020

FURTHER INFORMATION
Department of Business and Economics
Faculty of Business and Social Sciences
University of Southern Denmark
Campusvej 55, DK-5230 Odense M
Denmark
E-mail: lho@sam.sdu.dk / http://www.sdu.dk/ivoe

ISSN 2596-8157

Identification of the average treatment effect when SUTVA is violated

Lukáš Lafférs[*] Giovanni Mellace[†]

March 3, 2020

Abstract

The stable unit treatment value assumption (SUTVA) ensures that only two potential outcomes exist and that one of them is observed for each individual. After providing new insights on SUTVA validity, we derive sharp bounds on the average treatment effect (ATE) of a binary treatment on a binary outcome as a function of the share of units, α, for which SUTVA is potentially violated. Then we show how to compute the maximum value of α such that the sign of the ATE is still identified. After decomposing SUTVA into two separate assumptions, we provide weaker conditions that might help sharpening our bounds. Furthermore, we show how some of our results can be extended to continuous outcomes. Finally, we estimate our bounds in two well known experiments, the U.S. Job Corps training program and the Colombian PACES vouchers for private schooling.

Keywords: SUTVA; Bounds; Average treatment effect; Sensitivity analysis.

JEL classification: C14, C21, C31.

[*] Matej Bel University, Dept. of Mathematics. E-mail: lukas.laffers@gmail.com, Web: http://www.lukaslaffers.com. Lafférs acknowledges support provided by the Slovak Research and Development Agency under the contract No. APVV-17-0329 and VEGA-1/0843/17.

[†] University of Southern Denmark, Dept. of Business and Economics, E-mail: giome@sam.sdu.dk, Web: http://sites.google.com/site/giovannimellace/

We have benefited from the feedback from seminar participants in Toulouse School of Economics, University of Fribourg, and University of Rome 3.

1 Introduction and literature review

The stable unit treatment value assumption (SUTVA) first appeared in Rubin (1980), but it had already been discussed in earlier studies. For example, Cox (1958) assumes no interference between units. SUTVA plays a central role in the identification of causal effects, as i) it ensures that there exist as many potential outcomes as the number of the value the treatment can take on (two for the binary case considered in this paper) and ii) only under SUTVA we can observe one of the potential outcomes for each unit. Although SUTVA is essential for the identification of causal effects, there is still some confusion about its implications. Moreover, many studies only implicitly assume SUTVA and rarely discuss the implications of possible violations.

However, SUTVA is not always plausibly satisfied. For example, it is violated in the presence of general equilibrium effects (Heckman et al. 1999) or peer-effects, or in the presence of externalities and spillover effects. Most of the literature has focused on either modeling general equilibrium effects (Heckman et al. 1999) or has dealt with other types of interaction effects (see, e.g., Horowitz and Manski 1995, Miguel and Kremer 2004, Huber and Steinmayr 2019, Forastiere et al. 2016). However, SUTVA is also violated if some unit has access to different versions of the treatment, which may result in a different value of the potential outcome (Imbens and Rubin 2015). For this reason, the recent literature on causal inference decomposes SUTVA into two components that are somehow equivalent to those two reasons that induce SUTVA violations (Cole and Frangakis 2009a, VanderWeele 2009a, Pearl 2010, Petersen 2011).

This paper contributes to the literature in several ways. First, we discuss another potential violation of SUTVA, namely the presence of measurement error in either the observed outcome or the treatment indicator. Then we start by providing identification results for the binary outcome case. In particular, we derive sharp bounds on the ATE, which are functions of the share of units for which SUTVA could potentially be violated (i.e., the observed outcome differs from the potential outcome). This allows us to perform a sensitivity analysis of the point identified ATE (under SUTVA). In particular, we show how to estimate the maximum share of units for which SUTVA can be violated without changing the conclusion about the sign of the ATE. In addition, we show how the bounds can be sharpened and the sensitivity analysis can be improved by using observable covariates.

We use our sensitivity analysis to evaluate the sensitivity of the ATE estimated in two well known experiments: the U.S. Job Corps training program, which was already studied in Lee (2009), and the Colombia vouchers for private school, which was first evaluated in Angrist et al. (2006). We find that the ATE of the random assignment (intention-to-treat) is very sensitive to SUTVA violations and that the maximum share of units for which SUTVA can be violated is very small but statistically different from zero in both experiments.

Finally, we decompose SUTVA in two separate assumptions and provide weaker alternative assumptions, which can help to narrow the bounds and generalize some of our results for continuous outcomes. The paper is organized as follows: in Section 2 we introduce some necessary notation and discuss potential reasons for SUTVA violations, in Section 3 we derive our bounds and provide the sensitivity analysis, in Section 4 we show the results of the empirical application, in Section 5 we show how we can norrow the bounds by decomposing SUTVA into two separate assumptions and Section 6 concludes. All proofs as well as potential extensions to continuous outcomes are provided in the appendix.

2 Setup and notation

For each individual, i, in the population, $\mathcal{I}$, we define:

- the observed binary outcome as $Y_i \in \mathcal{Y} = \{0, 1\}$,
- the observed binary treatment as $D_i \in \mathcal{D} = \{0, 1\}$, and
- the two potential outcomes, that which only exist when SUTVA is satisfied, as $(Y_i(0), Y_i(1)) \in \mathcal{Y} \times \mathcal{Y}$.

We can observe the probability distribution of (Y, D) while the joint distribution of the potential outcomes $(Y(0), Y(1))$ is not observable, as we can only observe at most one potential outcome for each individual. We are interested in the average treatment effect, $ATE = E[Y(1) - Y(0)]$, which is a functional of the joint distribution of $(Y(0), Y(1), Y, D)$, and represent the ATE in the hypothetical scenario where SUTVA is satisfied.

The literature contains several definitions of SUTVA, which is often only implicitly assumed. We define SUTVA as

Assumption 1: (SUTVA)

$$\forall d \in \mathcal{D}, \forall i \in \mathcal{I}: \quad \text{If } D_i = d \text{ then } Y_i(d) = Y_i.$$

This definition of SUTVA is equivalent to the one included in Rubin (1980) and allows us to relate observed and potential outcomes through the well known observational rule,

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0).$$

As already discussed in the introduction, SUTVA requires that:

(i) There are no interaction effects.

(ii) The treatment is exhaustive, so that there are no hidden versions of the treatment that may affect the potential outcomes.

(iii) Neither the treatment nor the observed outcomes are measured with error.

While (i) and (ii) have been extensively discussed as potential sources of SUTVA violations, (iii) is rarely considered in relation to SUTVA. However, if either the treatment status or the observed outcome are measured with error, Assumption 1 is likely violated. This is important, as measurement error issues are arguably more prevalent in empirical applications than the other two potential sources of SUTVA violation.

Note that we do not define the potential outcomes as an explicit function of the treatment status of other individuals nor of a hidden version of a treatment. One can consider the the way the potential outcomes are defined as a modeling choice. Imposing less structure does not enable us to distinguish between the different reasons for SUTVA violations but, in return, our results can be applied in general for all three different sources of violation. In Section 5 we impose more structure when modeling the potential outcomes, and this allows us to gain some further insights into the impacts of different sources of SUTVA violations on the identification of the ATE.

We will denote the joint probability distribution of $(Y(0), Y(1), Y, D)$ by π, formally,

$$\pi_{ij} = \Pr\left((Y(0), Y(1)) = m(j), (Y, D) = m(i)\right), \ \ \forall i, j \in \{1, 2, 3, 4\},$$

$$m(1) = (0, 0), \ m(2) = (0, 1), \ m(3) = (1, 0), \ m(4) = (1, 1)$$

and by $S_i = I\{D_i = d \implies Y_i(d) = Y_i\}$ an indicator function equal to 1 if for individual i Assumption 1 holds.

As illustrated in Figure 1 (in appendix B) under SUTVA it must hold that

$$\pi_{13} = \pi_{14} = \pi_{22} = \pi_{24} = \pi_{31} = \pi_{32} = \pi_{41} = \pi_{43} = 0. \tag{1}$$

3 Results

3.1 Illustration: Identification when SUTVA is satisfied

Under SUTVA, the observed joint probabilities of the outcome and the treatment can be rewritten in terms of the unobserved joint probability distribution, π, in the following way:

$$\begin{aligned}
p_{00} &\equiv \Pr(Y = 0, D = 0) = \pi_{11} + \pi_{12}, & E[Y(0)|D = 0] &= \frac{\pi_{33} + \pi_{34}}{\Pr(D = 0)}, \\
p_{01} &\equiv \Pr(Y = 0, D = 1) = \pi_{21} + \pi_{23}, & E[Y(0)|D = 1] &= \frac{\pi_{23} + \pi_{44}}{\Pr(D = 1)} \\
p_{10} &\equiv \Pr(Y = 1, D = 0) = \pi_{33} + \pi_{34}, & E[Y(1)|D = 0] &= \frac{\pi_{12} + \pi_{34}}{\Pr(D = 0)} \\
p_{11} &\equiv \Pr(Y = 1, D = 1) = \pi_{42} + \pi_{44}, & E[Y(1)|D = 1] &= \frac{\pi_{42} + \pi_{44}}{\Pr(D = 1)}.
\end{aligned}$$

Similarly, conditional on the treatment status, the observed mean outcome is equal to the mean potential outcome

$$\begin{aligned}
E[Y|D = 0] &= \frac{\pi_{33} + \pi_{34}}{\Pr(D = 0)} = E[Y(0)|D = 0], \\
E[Y|D = 1] &= \frac{\pi_{42} + \pi_{44}}{\Pr(D = 1)} = E[Y(1)|D = 0].
\end{aligned}$$

We can rewrite the mean potential outcomes as

$$\begin{aligned}
E[Y(0)] &= E[Y(0)|D = 1] \cdot \Pr(D = 1) + E[Y(0)|D = 0] \cdot \Pr(D = 0) \\
&= \pi_{23} + \pi_{44} + \pi_{33} + \pi_{34}, \\
E[Y(1)] &= E[Y(1)|D = 1] \cdot \Pr(D = 1) + E[Y(1)|D = 0] \cdot \Pr(D = 0) \\
&= \pi_{42} + \pi_{44} + \pi_{12} + \pi_{34}.
\end{aligned} \tag{2}$$

This implies that the ATE can be written as

$$E[Y(1)-Y(0)] = \pi_{42}+\pi_{12}-\pi_{23}-\pi_{33}. \tag{3}$$

If we assume that the treatment is exogenous, it is well known that the ATE is a function of only observable quantities and is therefore identified. We summarize this result in Lemma 1 after having formally defined exogeneity.

Assumption 2: (Exogenous Treatment Selection)

$$\forall d \in \mathcal{D}: \quad E[Y(d)|D=1] = E[Y(d)|D=0].$$

Lemma 1. *Under Assumptions 1 and 2, the ATE is identified.*

Proof of Lemma 1. Under Assumption 1, $E[Y(d)|D=d] = E[Y|D=d]$, and under Assumption 2, $E[Y(d)|D=1] = E[Y(d)|D=0]$, and hence $ATE = E[Y(1)-Y(0)] = E[Y|D=1]-E[Y|D=0]$ is identified from the data. □

3.2 (Point) identification when SUTVA is violated

When SUTVA does not hold, the observed probabilities become

$$\begin{aligned}
p_{00} &= \pi_{11}+\pi_{12}+\pi_{13}+\pi_{14}, \quad & E[Y(0)|D=0] &= \frac{\pi_{33}+\pi_{34}+\pi_{13}+\pi_{14}}{\Pr(D=0)},\\
p_{01} &= \pi_{21}+\pi_{23}+\pi_{22}+\pi_{24}, \quad & E[Y(0)|D=1] &= \frac{\pi_{23}+\pi_{44}+\pi_{24}+\pi_{43}}{\Pr(D=1)},\\
p_{10} &= \pi_{33}+\pi_{34}+\pi_{31}+\pi_{32}, \quad & E[Y(1)|D=0] &= \frac{\pi_{12}+\pi_{34}+\pi_{14}+\pi_{32}}{\Pr(D=0)},\\
p_{11} &= \pi_{42}+\pi_{44}+\pi_{41}+\pi_{43}, \quad & E[Y(1)|D=1] &= \frac{\pi_{42}+\pi_{44}+\pi_{22}+\pi_{24}}{\Pr(D=1)}.
\end{aligned} \tag{4}$$

The fundamental difference is that now the potential outcomes for a given observed treatment value are not identified from the data, so the observed $E[Y|D=d]$ does not need be equal to $E[Y(d)|D=d]$, i.e.,

$$\begin{aligned}
E[Y|D=0] &= \frac{\pi_{33}+\pi_{34}+\pi_{31}+\pi_{32}}{\Pr(D=0)} \neq \frac{\pi_{33}+\pi_{34}+\pi_{13}+\pi_{14}}{\Pr(D=0)} = E[Y(0)|D=0],\\
E[Y|D=1] &= \frac{\pi_{42}+\pi_{44}+\pi_{41}+\pi_{43}}{\Pr(D=1)} \neq \frac{\pi_{42}+\pi_{44}+\pi_{22}+\pi_{24}}{\Pr(D=1)} = E[Y(1)|D=1].
\end{aligned}$$

The mean potential outcomes are now given by

$$
\begin{aligned}
E[Y(0)] &= E[Y(0)|D=1]\cdot \Pr(D=1) + E[Y(0)|D=0]\cdot \Pr(D=0)\\
&= \pi_{23}+\pi_{44}+\pi_{24}+\pi_{43}+\pi_{33}+\pi_{34}+\pi_{13}+\pi_{14},\\
E[Y(1)] &= E[Y(1)|D=1]\cdot \Pr(D=1) + E[Y(1)|D=0]\cdot \Pr(D=0)\\
&= \pi_{42}+\pi_{44}+\pi_{22}+\pi_{24}+\pi_{12}+\pi_{34}+\pi_{14}+\pi_{32}.
\end{aligned}
$$

Therefore,

$$E[Y(1)-Y(0)] = \pi_{42}+\pi_{12}+\pi_{22}+\pi_{32}-\pi_{23}-\pi_{33}-\pi_{13}-\pi_{43}.$$

The ATE can still be identified but at the price of imposing strong additional assumptions. For illustration, we propose an example of a sufficient condition that guarantees identification.

Assumption 3: (Balanced bias)

$$
\begin{aligned}
&\Pr(Y=1,S=0|D=1)-\Pr(Y=0,S=0|D=1)\\
&=\Pr(Y=1,S=0|D=0)-\Pr(Y=0,S=0|D=0).
\end{aligned}
\tag{5}
$$

Assumption 3 states that the bias induced by the violation of SUTVA is the same in the treated and non-treated populations. The following lemma shows that this assumption guarantees that the naive ATE estimator $E[Y|D=1]-E[Y|D=0]$ still identifies the true ATE.

Lemma 2. *Under Assumptions 2 and 3, the ATE is identified.*

Proof. See Appendix A. □

3.3 Relaxing SUTVA

In this section we first derive sharp bounds on the ATE as a function of the share of units, α, for which SUTVA can potentially be violated. The sensitivity parameter $0 \leq \alpha \leq 1$ can be directly interpreted as the maximum probability that SUTVA does not hold. First we assume that α is known, and then we show how to estimate the maximum value of α such that our bounds are able to identify the sign of the ATE. For a given α we assume that

Assumption 1α: (Known maximum SUTVA violation share)

$$\Pr(\forall d \in \mathcal{D} : D_i = d \implies Y_i(d) = Y_i) \geq 1 - \alpha.$$

This assumption, previously used to model measurement error in the observed outcomes or in the treatment by Lafférs (2019), implies that

$$\pi_{13} + \pi_{14} + \pi_{22} + \pi_{24} + \pi_{31} + \pi_{32} + \pi_{41} + \pi_{43} \leq \alpha.$$

Under Assumption 1α, the ATE is no longer point identified. We first provide its sharp bounds without imposing any further assumptions in the following lemma.

Lemma 3. *Under Assumption 1α, the sharp bounds on the ATE are as follows:*[1]

$$\begin{aligned} ATE &\in [ATE^{LB}, ATE^{UB}] \\ ATE^{LB} &= \max\{-p_{10} - p_{01} - \alpha, -1\}, \\ ATE^{UB} &= \min\{p_{00} + p_{11} + \alpha, 1\}. \end{aligned} \tag{6}$$

Proof. See Appendix A. □

The width of these bounds is $1 + 2\alpha$, and they are therefore not useful in practice. We extend this result to continuous outcomes in Appendix C. In order to obtain meaningful bounds we also need to assume that the treatment is exogenous (Assumption 2). The resulting bounds are presented in the following lemma.

[1] The dependence of ATE^{LB} and ATE^{UB} on α is suppressed for brevity.

Lemma 4. *Under Assumptions 1α and 2, the sharp bounds on the ATE are as follows:*

$$
\begin{aligned}
&ATE \in [ATE^{LB}, ATE^{UB}] \\
&\quad \text{if } p_{11} + p_{01} > p_{00} + p_{10}: \\
ATE^{LB} &= \frac{p_{11} - \min\{\max\{\alpha - p_{00}, 0\}, p_{11}\}}{p_{11} + p_{01}} - \frac{p_{10} + \min\{p_{00}, \alpha\}}{p_{00} + p_{10}}, \\
ATE^{UB} &= \frac{p_{11} + \min\{\max\{\alpha - p_{10}, 0\}, p_{01}\}}{p_{11} + p_{01}} - \frac{p_{10} - \min\{p_{10}, \alpha\}}{p_{00} + p_{10}}, \\
&\quad \text{if } p_{11} + p_{01} < p_{00} + p_{10}: \\
ATE^{LB} &= \frac{p_{11} - \min\{p_{11}, \alpha\}}{p_{11} + p_{01}} - \frac{p_{10} + \min\{\max\{\alpha - p_{11}, 0\}, p_{00}\}}{p_{00} + p_{10}}, \\
ATE^{UB} &= \frac{p_{11} + \min\{p_{01}, \alpha\}}{p_{11} + p_{01}} - \frac{p_{10} - \min\{\max\{\alpha - p_{01}, 0\}, p_{01}\}}{p_{00} + p_{10}}.
\end{aligned}
\tag{7}
$$

Proof. See Appendix A. □

The relationship between our bounds and α is visualized in Figure 2. In particular, it is important to notice that as α increases the width of our bounds becomes larger. This is not surprising as, intuitively, the larger the share of units for which SUTVA is violated the less we can learn about the ATE from the observed data.

In most applications it is very likely that α is unknown. If this is the case, we can use the results of Lemma 4 to detect the maximum share of units for which SUTVA can be violated that allows our bounds to identify the sign of the ATE. This is shown in the following lemma.

Lemma 5. *Under Assumptions 1α and 2, $ATE^{LB} \geq 0$ if and only if*

$$0 \leq \alpha \leq \alpha^{+} \equiv \min\{\Pr(D=1), \Pr(D=0)\} \cdot [E(Y|D=1) - E(Y|D=0)]$$

and $ATE^{UB} \leq 0$ if and only if

$$0 \leq \alpha \leq \alpha^{-} \equiv -\min\{\Pr(D=1), \Pr(D=0)\} \cdot [E(Y|D=1) - E(Y|D=0)].$$

Proof. See Appendix A. □

Lemma 5 shows that knowing whether either α^{+} or α^{-} is bigger than zero is useful. For example, an α^{+} bigger than zero implies a positive ATE if the fraction of individuals affected by SUTVA violations is smaller than α^+. Thus, it is interesting to test $H_0 : \alpha^- = 0$ and $H_0 : \alpha^+ = 0$. For example, if the latter is rejected it means that as soon as less than α^+ are subject to SUTVA violations the ATE is positive. Notice that under the Assumptions 1α and 2 $\alpha = 0$ implies $ATE = E[Y|D = 1] - E[Y|D = 0] > \alpha^+$. Thus, it is possible that the naive ATE estimator can be significantly different from 0, while α^+ is not.

3.4 Narrowing the bounds using covariates

Suppose that a set of covariates, $X_i \in \mathbf{X}$, is also available and that all our assumptions also hold conditional on X, such that $ATE = \int_{\mathcal{X}} ATE_x \Pr(X = x)dx$, where $ATE_x = E[Y(1) - Y(0)|X = x]$. Further assume that the treament is exogenous conditional on these covariates.

Assumption 2X: (Conditional Exogenous Treatment Selection)

$$\forall d \in \mathcal{D}, \forall x \in \mathcal{X} : \quad E[Y(d)|D = 1, X = x] = E[Y(d)|D = 0, X = x].$$

Lemma 6. *Under Assumptions 1α and 2X, the sharp bounds on the ATE are as follows:*

$$\begin{aligned} ATE &\in \left[\overline{ATE}^{LB}, \overline{ATE}^{UB}\right], \\ \overline{ATE}^{LB} &= \int_{\mathbf{X}} ATE_x^{LB} \Pr(X = x)dx, \\ \overline{ATE}^{UB} &= \int_{\mathbf{X}} ATE_x^{UB} \Pr(X = x)dx. \end{aligned} \tag{8}$$

If $p_{11|x} + p_{01|x} > p_{00|x} + p_{10|x}$:

$$\begin{aligned} ATE_x^{LB} &= \frac{p_{11|x} - \min\{\max\{\alpha - p_{00|x}, 0\}, p_{11|x}\}}{p_{11|x} + p_{01|x}} - \frac{p_{10|x} + \min\{p_{00|x}, \alpha\}}{p_{00|x} + p_{10|x}}, \\ ATE_x^{UB} &= \frac{p_{11|x} + \min\{\max\{\alpha - p_{10|x}, 0\}, p_{01|x}\}}{p_{11|x} + p_{01|x}} - \frac{p_{10} - \min\{p_{10|x}, \alpha\}}{p_{00|x} + p_{10|x}}, \end{aligned} \tag{9}$$

and if $p_{11|x} + p_{01|x} < p_{00|x} + p_{10|x}$:

$$\begin{aligned} ATE_x^{LB} &= \frac{p_{11|x} - \min\{p_{11|x}, \alpha\}}{p_{11|x} + p_{01|x}} - \frac{p_{10|x} + \min\{\max\{\alpha - p_{11|x}, 0\}, p_{00|x}\}}{p_{00} + p_{10}}, \\ ATE_x^{UB} &= \frac{p_{11|x} + \min\{p_{01|x}, \alpha\}}{p_{11|x} + p_{01|x}} - \frac{p_{10|x} - \min\{\max\{\alpha - p_{10|x}, 0\}, p_{01|x}\}}{p_{00|x} + p_{10|x}}. \end{aligned} \tag{10}$$

Furthermore, $\overline{ATE}^{LB} \geq ATE^{LB}$ *and* $\overline{ATE}^{UB} \leq ATE^{UB}$.

Proof. See Appendix A. □

In practice, including covariates might require dividing the sample into a finite number of groups depending on the predicted value of the outcome variable.[2] The choice of the number of groups depends on the problem at hand. The larger the number of groups the sharper are the resulting bounds, but at the same time the statistical uncertainty within each group increases.

When information about X is available, the maximum possible violation of SUTVA, $\alpha^{+}(\alpha^{-})$ that guarantees positive (negative) ATE are given in the following lemma

Lemma 7. *Under the Assumptions 1α and 2X,* $\overline{ATE}^{LB} \geq 0$ *if and only if*

$$0 \leq \alpha \leq \overline{\alpha}^{+} \equiv \int_{\mathbf{X}} \min\{\alpha_x^{+}, 0\} \Pr(X = x) dx$$

and $\overline{ATE}^{UB} \leq 0$ *if and only if*

$$0 \leq \alpha \leq \overline{\alpha}^{-} \equiv \int_{\mathbf{X}} \min\{\alpha_x^{-}, 0\} \Pr(X = x) dx,$$

where

$$\alpha_x^{+} \equiv \min\{\Pr(D = 1, X = x), \Pr(D = 0, X = x)\} \cdot [E(Y|D = 1, X = x) - E(Y|D = 0, X = x)]$$
$$\alpha_x^{-} \equiv -\alpha_x^{+}.$$

Proof. See Appendix A. □

We note that $\overline{\alpha}^{+} \leq \alpha^{+}$ (and similarly $\overline{\alpha}^{-} \geq \alpha^{+}$), because for some x the quantity $E(Y|D = 1, X = x) - E(Y|D = 0, X = x)$ may be negative even though $E(Y|D = 1) - E(Y|D = 0) \geq 0$.

3.5 Estimation and inference

The fact that the expressions for bounds α^{+} and α^{-} involve minimum and maximum operators gives rise to a non-standard inferential procedure, as no regular $\sqrt{n}$-consistent estimator exists (Hirano and Porter 2012) and analog estimators may be

[2] For example, Lee (2009) uses all available covariates to construct a single index that defines five groups depending on the predicted values of the outcome.

severely biased in small samples. For this reason, we suggest using the intersection bounds approach of Chernozhukov et al. (2013), which creates half-median unbiased point estimates and confidence intervals.[3] This method corrects for the small sample bias *before* the max/min operator is applied.

4 Empirical illustrations

We consider two empirical applications to illustrate the scope and usefulness of our results. In the first one we are interested in the effect of the random assignment to the U.S. Job Corps training program on the probability of employment four years after the assignment. As not everyone in the sample complied with the random assignment, we will focus on the intention-to-treat effect as in Lee (2009). Evaluations of this program have aroused considerable interest among policymakers and researchers during recent decades, which is hardly surprising given the high costs associated with the program. We use the same data from National Job Corps Study as Lee (2009). We refer the reader to Lee (2009) for an extensive data description.

Our second application looks at a school voucher experiment implemented in Colombia, namely the "programa de ampliacion de cobertura de la educacion secundaria" (PACES). We focus on the impact of being randomly assigned to the voucher covering approximately half of the cost of private secondary schooling, on the probability that low income pupils had to repeat a grade. We use data previously analyzed in Angrist et al. (2006).

4.1 The effect of Job Corps on employment

Table 1 provides the summary statistics.

The ATE bounds as a function of α are presented in Table 2 and visualized in Figure 2.

Under SUTVA and the exogenous treatment selection assumptions the impact of the assignment on the employment probability is 1.6%, which is significant at the

[3] Half-median unbiased means that the estimate of the upper(lower) bound exceeds (lies below) its true value with probability at least one half asymptotically.

$Y \setminus D$	offered training	not offered training
	$(D = 1)$	$(D = 0)$
working $(Y = 1)$	$p_{11} = 49.26\%$	$p_{10} = 31.63\%$
not working $(Y = 0)$	$p_{01} = 11.16\%$	$p_{00} = 7.94\%$
$n = 11146$	$\Pr(D = 1) = 60.43\%$	$\Pr(D = 0) = 39.57\%$

Table 1: Probability distribution of the *working after 202 weeks* indicator (Y) and the randomized assignment to Job corps indicator (D). Based on a data set from Lee (2009). Missing values were removed.

α	$[ATE^{LB}, ATE^{UB}]$
	(CB^{LB}, CB^{UB})
0	[0.016, 0.016]
	(0.001, 0.031)
0.01	[-0.009, 0.041]
	(-0.023, 0.055)
0.05	[-0.111, 0.142]
	(-0.124, 0.155)
0.1	[-0.219, 0.269]
	(-0.230, 0.282)
0.2	[-0.384, 0.521]
	(-0.394, 0.537)
0.5	[-0.881, 1]
	(-0.893, 1)

α^+	0.954%
(CB^l, CB^u)	(0.076%, 1.213%)

Table 2: Bounds on the ATE for different choices of α. The left table presents estimates of bounds on ATE together with 95% confidence bounds. On the right-hand side, α^+ is the estimated maximum value of α that still gives a positive ATE. All estimates are half-median unbiased and based on Chernozhukov et al. (2013) using 9999 bootstrap samples and 200000 replications.

95% confidence level. The minimum share of individuals for which SUTVA has to be satisfied to have a positive ATE, α^+, is 0.954%. Although statistically different from zero, α^+ is very small. This implies that we can only conclude that the effect is positive if we are willing assume that less than 1% of the individuals is affected by SUTVA violations.

4.2 The effect of school vouchers on never repeating a grade

Some relevant descriptive statistics are reported in Table 3. We refer to Angrist et al. (2006) for an extensive data description.

$Y \setminus D$	offered voucher $(D = 1)$	not offered voucher $(D = 0)$
never repeated a grade $(Y = 1)$	$p_{11} = 43.71\%$	$p_{10} = 37.30\%$
repeated a grade $(Y = 0)$	$p_{01} = 8.41\%$	$p_{00} = 10.57\%$
$n = 1201$	$\Pr(D = 1) = 52.12\%$	$\Pr(D = 0) = 47.88\%$

Table 3: Probability distribution of the outcome *never repeating a grade* (Y) and of the randomized treatment (*school vouchers offered*). Based on a dataset from Angrist et al. (2006). Missing values were removed.

Under Assumptions 1 and 2, the point identified ATE of the voucher offer on the probability of never repeating a grade is 6% and it is statistically significant at the 95% confidence level. The sign of the effect is confirmed if SUTVA is violated for no more than 3.03% of the population. This effect is more robust to SUTVA violations than in the previous example; however, the estimated α^+ is still very low.

Our results are summarized in Table 4 and visualized in Figure 3.

α	$[ATE^{LB}, ATE^{UB}]$ (CB^{LB}, CB^{UB})
	[0.060, 0.060] (0.009, 0.110)
0.01	[0.033, 0.092] (-0.014, 0.136)
0.05	[-0.050, 0.174] (-0.094, 0.215)
0.1	[-0.154, 0.278] (-0.192, 0.318)
0.2	[-0.348, 0.485] (-0.384, 0.528)
0.5	[-0.932, 1] (-0.969, 1)

α^+	3.03%
(CB^l, CB^u)	$(0.69\%, 5.08\%)$

Table 4: Bounds on the ATE for different choices of α. The left table presents estimates of bounds on ATE together with 95% confidence bounds. On the right-hand side, α^+ is the estimated maximum value of α that still gives a positive ATE. All estimates are half-median unbiased and based on Chernozhukov et al. (2013) using 9999 bootstrap samples and 200000 replications.

5 Extension: Decomposing SUTVA assumption

So far we have been completely agnostic about the mechanisms that can lead to SUTVA violation. However, in some applications it could be useful to consider them separately. In the epidemiology literature, the version of SUTVA we consider in this paper (Assumption 1) is known as the *consistency* assumption (Cole and Frangakis 2009b).

VanderWeele (2009b) propose a decomposition of this assumption into two components. They refer to the first component as *treatment-variation irrelevance* and to the second component as *consistency*. We will now consider their separation and propose alternative weaker assumptions, which can be used to derive bounds on the ATE that are sharper than the one we derived in Section 3.3.

To this end, we allow the potential outcomes of individual i to be a function of not only the treatment indicator, but also of a variable, $H_i \in \mathcal{H}$, which can represent different things. It can capture different dose or length of exposure to the treatment, it can be a function of the treatment indicator of other individuals or it can be a binary indicator that represents whether either the observed outcome or the treatment indicator is measured with error. In the latter case, the potential outcome itself is not affected by H, but H selects individuals affected by measurement error. Hereafter, for the sake of easy exposition, we will refer to H as "hidden treatment". Now we can define the potential outcomes as functions of both the observed and hidden treatments $Y(d,h)$. Depending on the application, the average treatment effect of interest can be defined in different ways since the potential outcomes also depend on H. For example, if there exists different version of the treatment, the quantity of interest could be the mean of the ATEs for different values of H:

$$ATE = \int_{\mathcal{H}} ATE(h) \Pr(H = h) dh,$$

where $ATE(h) = E[Y(1,h) - Y(0,h)]$.

VanderWeele (2009b) introduce the following assumptions that together are equivalent to Assumption 1 (SUTVA) above.

Assumption 1A: (Treatment-variation irrelevance assumption)

$$\forall d \in \mathcal{D}, \forall h, h' \in \mathcal{H}, \forall i \in \mathcal{I}: \quad D_i = d \implies Y_i(d,h) = Y_i(d,h'). \tag{11}$$

Assumption 1A implies that there are neither multiple versions of the treatment (e.g., different treatment intensities) nor interference between units; i. e.,

$$Y_i(d_i, \mathbf{d}_{-i}) = Y_i(d_i, \mathbf{d'}_{-i}), \forall \mathbf{d}_{-i}, \mathbf{d'}_{-i},$$

where $\mathbf{d}_{-i}$ stands for the vector of treatments of individuals other than i. Under Assumption 1A the notation $Y_i(d)$ is appropriate and the quantity $ATE = E(Y(1) - Y(0))$ is well defined.

Assumption 1B: (Consistency Assumption)

$$\forall d \in \mathcal{D}, \forall h \in \mathcal{H}, \forall i \in \mathcal{I}: \; D_i = d, \; H_i = h \implies Y_i(d,h) = Y_i. \tag{12}$$

This assumption states that the observed value of outcome Y_i is consistent with the potential outcome model formulation. A possible violation of this assumption is mismeasurement of the observed outcome or the treatment.

We note that Assumptions 1A and 1B imply the following condition:

$$\forall d \in \mathcal{D}, \forall h, h' \in \mathcal{H}, \forall i \in \mathcal{I}: \quad D_i = d, \; H_i = h \implies Y_i(d,h) = Y_i(d,h') = Y_i(d) = Y_i,$$

which it is equivalent to imposing SUTVA.

Figure 4 depicts the individual average treatment effects on and the support of the joint probability distribution of $(Y^{00}, Y^{01}, Y^{10}, Y^{11}, Y, D, H)$ for a binary hidden treatment, H. In most figures we use the notation $Y^{dh} = Y(d,h)$.

Both Assumptions 1A and 1B are support restrictions, and thus we can relax them separately. For example, this is important in applications where one is only concerned about measurement error and can safely impose Assumption 1B.

Assumption 1Aβ: (Relaxed Treatment-variation Irrelevance Assumption)

$$\Pr(\forall d \in \mathcal{D}, \forall h, h' \in \mathcal{H}: \; Y_i(d,h) = Y_i(d,h')) \geq 1 - \beta. \tag{13}$$

Assumption 1Bγ: (Relaxed Consistency Assumption)

$$\Pr(\forall d \in \mathcal{D}, \forall h \in \mathcal{H} : \ D_i = d, \ H_i = h \implies Y_i(d,h) = Y_i) \geq 1 - \gamma. \tag{14}$$

In addition, we impose the following assumption, which is satisfied under random treatment allocation:

Assumption 2H: (Exogenous Treatment Selection with Hidden Treatment)

$$\forall d \in \mathcal{D}, \ \forall h \in \mathcal{H} : \ E[Y(d,h)|D=1] = E[Y(d,h)|D=0].$$

The effects on the ATE of different relaxations are visualized using a simulated example in Figure 5. Figures 6 and 7 show joint probability distributions that maximize the ATE under different relaxations of SUTVA. All the identifying assumptions impose linear restrictions on the space of admissible joint probability distributions $(Y^{00}, Y^{01}, Y^{10}, Y^{11}, Y, D, H)$. On top of that, these distributions have to be compatible with the distribution of (Y, D), which is also a linear restriction. The bounds on the ATE are calculated using a linear programming procedure described in Laffers (2019). We note that there are recent advances in statistical inference of partially identified parameters that deal with random linear programs of such form (Kaido et al. 2019[4] or Hsieh et al. 2018). Subsampling approaches may be used on the lower and upper bounds separately, as described in Laffers (2019) or Demuynck (2015).

6 Conclusion

This paper discusses the Stable Unit Treatment Value Assumption (SUTVA) assumptions and the implications of its violations for the identification of the average treatment effect. We derive bounds on the ATE under the assumption that only at most a known fraction of individuals is affected by SUTVA. Moreover, we show how to estimate the maximum share of individuals that can be affected by SUTVA violation that still allows us to identify the sign of the ATE and illustrate our theoretical results with two empirical examples. Finally, following the epidemiology literature, we show how

[4] Which was implemented in Kaido et al. (2017).

decomposing SUTVA into two separate assumptions allows to distinguish between the different sources of SUTVA violation and potentially narrow our bounds.

Appendix

A Proofs

Proof of Lemma 2. The Assumption 2 together with (4) implies

$$\begin{aligned} ATE = E[Y(1)] - E[Y(0)] &= E[Y(1)|D=1] - E[Y(0)|D=0] \\ &= \frac{\pi_{42} + \pi_{44} + \pi_{22} + \pi_{24},}{\Pr(D=1)} - \frac{\pi_{33} + \pi_{34} + \pi_{13} + \pi_{14},}{\Pr(D=0)}. \end{aligned} \tag{A.1}$$

From (5) we can see that

$$E[Y|D=1] - E[Y|D=0] = \frac{\pi_{42} + \pi_{44} + \pi_{41} + \pi_{43},}{\Pr(D=1)} - \frac{\pi_{33} + \pi_{34} + \pi_{31} + \pi_{32}}{\Pr(D=0)}. \tag{A.2}$$

We note that under Assumption 3,

$$\frac{\pi_{41} + \pi_{43}}{\Pr(D=1)} - \frac{\pi_{22} + \pi_{24}}{\Pr(D=1)} = \frac{\pi_{31} + \pi_{32}}{\Pr(D=0)} - \frac{\pi_{13} + \pi_{14}}{\Pr(D=0)},$$

so that the equations (A.1) and (A.2) are equal. □

Proof of Lemma 3. We show the proof for the upper bound as the proof for the lower bound follows in an analogous way.

Let us further denote $ATE^s_{yd} = E[Y(1) - Y(0)|Y = y, D = d, S = s]$.

(i) Validity

$$
\begin{aligned}
ATE = &\Big[ATE^1_{00} \cdot \Pr(S=1|Y=0,D=0) + ATE^0_{00} \cdot \Pr(S=0|Y=0,D=0)\Big] \cdot p_{00} \\
&+ \Big[ATE^1_{01} \cdot \Pr(S=1|Y=0,D=1) + ATE^0_{01} \cdot \Pr(S=0|Y=0,D=1)\Big] \cdot p_{01} \\
&+ \Big[ATE^1_{10} \cdot \Pr(S=1|Y=1,D=0) + ATE^0_{10} \cdot \Pr(S=0|Y=1,D=0)\Big] \cdot p_{10} \\
&+ \Big[ATE^1_{11} \cdot \Pr(S=1|Y=1,D=1) + ATE^0_{11} \cdot \Pr(S=0|Y=1,D=1)\Big] \cdot p_{11} \\
\leq &[1 \cdot \Pr(S=1|Y=0,D=0)] + 0 \cdot \Pr(S=0|Y=0,D=0)] \cdot p_{00} \\
&+ [0 \cdot \Pr(S=1|Y=0,D=1)] + 1 \cdot \Pr(S=0|Y=0,D=1)] \cdot p_{01} \\
&+ [0 \cdot \Pr(S=1|Y=1,D=0)] + 1 \cdot \Pr(S=0|Y=1,D=0)] \cdot p_{10} \\
&+ [1 \cdot \Pr(S=1|Y=1,D=1)] + 0 \cdot \Pr(S=0|Y=1,D=1)] \cdot p_{11} \\
= &\Pr(S=1|Y=0,D=0) \cdot p_{00} + \Pr(S=1|Y=1,D=1) \cdot p_{11} \\
&+ \Pr(S=0|Y=0,D=1) \cdot p_{01} + \Pr(S=0|Y=1,D=0) \cdot p_{10} \\
\leq &p_{00} + p_{11} + \min\{p_{01} + p_{10}, \alpha\} = \min\{p_{00} + p_{11} + \alpha, 1\},
\end{aligned}
$$

Where the last inequality follows from the fact that $\Pr(S=0) \leq \alpha$.

(ii) Sharpness

Suppose that $\alpha < p_{01} + p_{10}$. Then there must exist constants $0 \leq \alpha_{01} \leq p_{01}$ and $0 \leq \alpha_{10} \leq p_{10}$, so that $\alpha = \alpha_{01} + \alpha_{10}$. The following specification for $\Pr(Y(0), Y(1), Y, D)$ is compatible with Assumption 1α and and with the distribution of (Y, D).

$$
\begin{aligned}
&\pi_{12} = p_{00},\ \pi_{22} = \alpha_{01},\ \pi_{32} = \alpha_{10},\ \pi_{42} = p_{11},\ \pi_{21} = p_{01} - \alpha_{01},\ \pi_{34} = p_{10} - \alpha_{10}, \\
&\pi_{11} = \pi_{13} = \pi_{14} = \pi_{23} = \pi_{24} = \pi_{31} = \pi_{33} = \pi_{41} = \pi_{43} = \pi_{44} = 0.
\end{aligned}
$$

Suppose now that $\alpha \geq p_{01} + p_{10}$.

$$
\begin{aligned}
&\pi_{12} = p_{00},\ \pi_{22} = p_{01},\ \pi_{32} = p_{10},\ \pi_{42} = p_{11}, \\
&\pi_{11} = \pi_{13} = \pi_{14} = \pi_{21} = \pi_{23} = \pi_{24} = \pi_{31} = \pi_{33} = \pi_{34} = \pi_{41} = \pi_{43} = \pi_{44} = 0.
\end{aligned}
$$

Figure 8 illustrates the sharpness part of the proof of Lemma 3, it depicts the compatible joint probability distributions that attains the lower and upper bound on ATE respectively.

□

Proof of Lemma 4. We show the proof for the upper bound and for $\pi_{11} + \pi_{01} > \pi_{00} + \pi_{10}$ as the proof for the lower bound and for $\pi_{11} + \pi_{01} < \pi_{00} + \pi_{10}$ follows in an analogous way.

(i) Validity

$$\begin{aligned}
ATE &= E[Y(1)-Y(0)] = E[Y(1)|D=1] - E[Y(0)|D=0] \\
&= \frac{\pi_{42}+\pi_{44}+\pi_{22}+\pi_{24}}{\Pr(D=1)} - \frac{\pi_{33}+\pi_{34}+\pi_{13}+\pi_{14}}{\Pr(D=0)} \\
&= \frac{p_{11}-\pi_{41}-\pi_{43}+\pi_{22}+\pi_{24}}{p_{11}+p_{01}} - \frac{p_{10}-\pi_{31}-\pi_{32}+\pi_{13}+\pi_{14}}{p_{00}+p_{10}} \\
&\leq \frac{p_{11}+\pi_{22}+\pi_{24}}{p_{11}+p_{01}} - \frac{\pi_{10}-\pi_{31}-\pi_{32}}{p_{00}+p_{10}} \\
&\leq \frac{p_{11}+\min\{\max\{\alpha-p_{10},0\},p_{01}\}}{p_{11}+p_{01}} - \frac{p_{10}-\min\{p_{10},\alpha\}}{p_{00}+p_{10}} = ATE^{UB}.
\end{aligned}$$

where the last inequality follows from inequalities $\pi_{31}+\pi_{32} \leq p_{10}$, $\pi_{22}+\pi_{24} \leq p_{01}$ and $\pi_{11}+\pi_{01} > \pi_{00}+\pi_{10}$.

(ii) Sharpness

Given that $\pi_{11}+\pi_{01} > \pi_{00}+\pi_{10}$, the following specification for $\Pr(Y(0),Y(1),Y,D)$ is compatible with Assumptions 1α, 2, with the distribution of (Y,D) and achieves the ATE^{UB}.

$$\begin{aligned}
&c_1 = \min\{p_{10},\alpha\}, \\
&c_2 = \min\{\max\{\alpha-p_{10},0\},p_{01}\}, \\
&\pi_{11} = p_{00} - p_{00}\frac{p_{11}+c_2}{p_{11}+p_{01}}, && \pi_{21} = p_{01} - c_2 - p_{01}\frac{p_{10}-c_1}{p_{00}+p_{10}}, \\
&\pi_{12} = p_{00}\frac{p_{11}+c_2}{p_{11}+p_{01}}, && \pi_{22} = c_2, \\
&\pi_{13} = 0, && \pi_{23} = p_{01}\frac{p_{10}-c_1}{p_{00}+p_{10}}, \\
&\pi_{14} = 0, && \pi_{24} = 0, \\
&\pi_{31} = c_1 - c_1\frac{p_{11}+c_2}{p_{11}+p_{01}}, && \pi_{41} = 0, \\
&\pi_{32} = c_1\frac{p_{11}+c_2}{p_{11}+p_{01}}, && \pi_{42} = p_{11} - p_{11}\frac{p_{10}-c_1}{p_{00}+p_{10}}, \\
&\pi_{33} = p_{10} - c_1 - (p_{10}-c_1)\frac{p_{11}+c_2}{p_{11}+p_{01}}, && \pi_{43} = 0, \\
&\pi_{34} = (p_{10}-c_1)\frac{p_{11}+c_2}{p_{11}+p_{01}}, && \pi_{44} = p_{11}\frac{p_{10}-c_1}{p_{00}+p_{10}}.
\end{aligned}$$

Straightforward manipulations show that the proposed specification is a proper probability distribution function.

□

Proof of Lemma 5. We only present the proof for $ATE^{LB} \geq 0$, as the the proof for $ATE^{UB} \leq 0$ is similar. Consider the case $\pi_{11}+\pi_{01} > \pi_{00}+\pi_{10}$. If $p_{00}+p_{11} \geq \alpha \geq p_{00}$, then

$$ATE^{LB} = \frac{p_{11}-(\alpha-p_{00})}{p_{11}+p_{01}} - 1,$$

so that $ATE^{LB} \geq 0$ would imply $p_{00} - p_{01} \geq \alpha$ which contradicts $\alpha \geq p_{00}$, so we have to have $\alpha \leq p_{00}$ and thus

$$ATE^{LB} = \frac{p_{11}}{p_{11}+p_{01}} - \frac{p_{10}+\alpha}{p_{00}+p_{10}} \geq 0 \iff \alpha \leq p_{11}\frac{p_{00}+p_{10}}{p_{11}+p_{01}} - p_{10} = \Pr(D=0)\left[E(Y|D=1) - E(Y|D=0)\right].$$

Similarly, for $\pi_{11} + \pi_{01} > \pi_{00} + \pi_{10}$ we get that for $ATE^{UB} \leq 0$ we have to have $\alpha \leq p_{11}$ and therefore

$$ATE^{UB} = \frac{p_{11}-\alpha}{p_{11}+p_{01}} - \frac{p_{10}}{p_{00}+p_{10}} \leq 0 \iff \alpha \leq p_{11} - p_{10}\frac{p_{11}+p_{01}}{p_{00}+p_{10}} = \Pr(D=1)\left[E(Y|D=1) - E(Y|D=0)\right],$$

which leads to the desired result.

□

Proof of Lemma 6. The proof is similar to the one or Proposition 1b in Lee (2009). The validity and sharpness of the bounds results from the application of Lemma 4 conditional on $X = x$. The second part follows from the fact that any ATE that is consistent with (Y, D, X) has to be consistent with (Y, D), that is ignoring the information about X cannot lead to a more informative result (sharpen the bounds). □

Proof of Lemma 7. Analoguous to the proof of Lemma 5 and hence ommited. □

B Figures

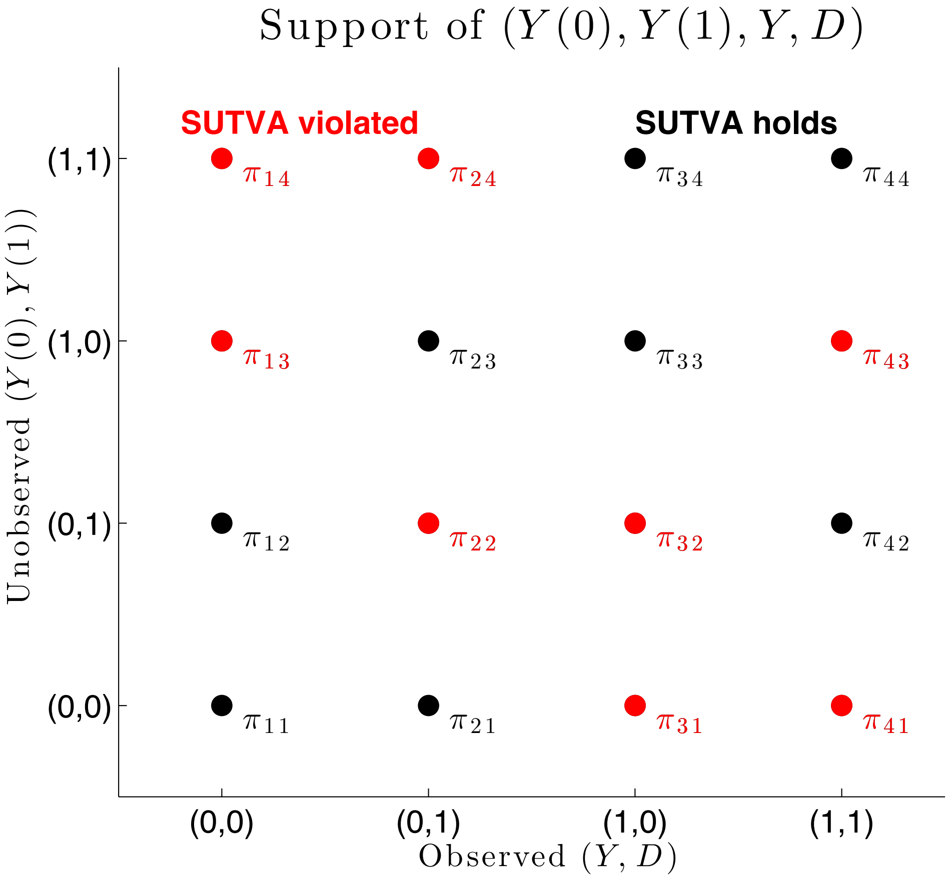


Figure 1: Support of the joint probability distribution of $(Y(0), Y(1), Y, D)$. Under SUTVA, the red points must have zero probability mass.

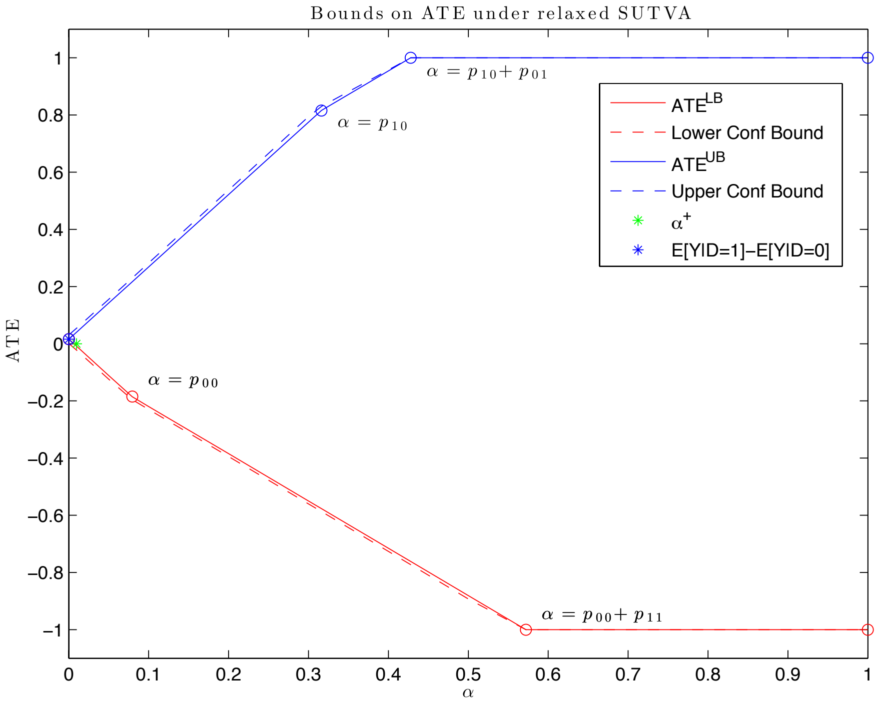


Figure 2: Sensitivity analysis to SUTVA assumption of the bounds on ATE of the assignment to job training on the probability of employment (Intention-to-Treat). All estimates are half-median unbiased and based on Chernozhukov et al. (2013).

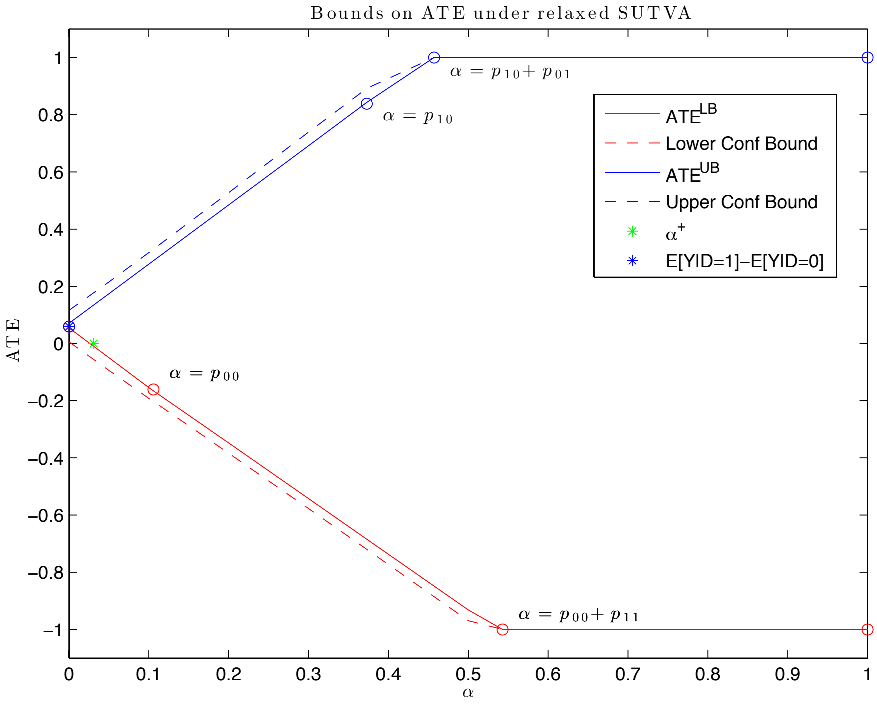


Figure 3: Sensitivity analysis to SUTVA assumption of the bounds on ATE of the school vouchers on the probability of never repeating a grade (Intention-to-Treat). All estimates are half-median unbiased and based on Chernozhukov et al. (2013).

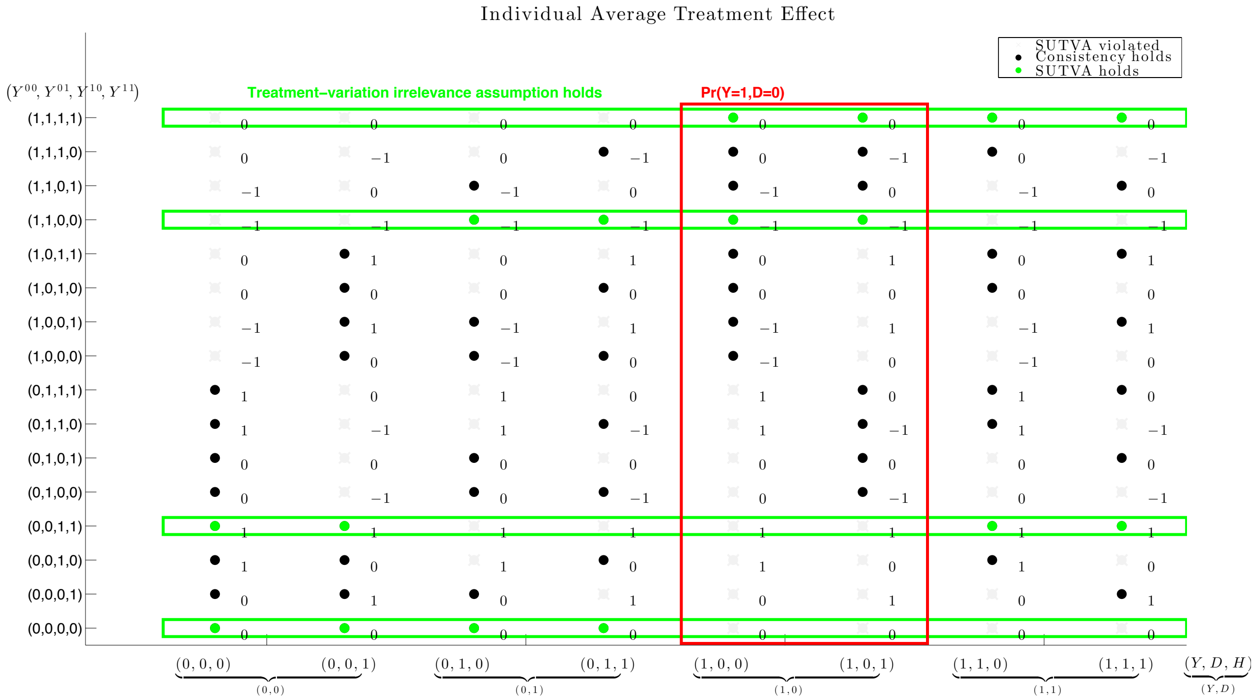
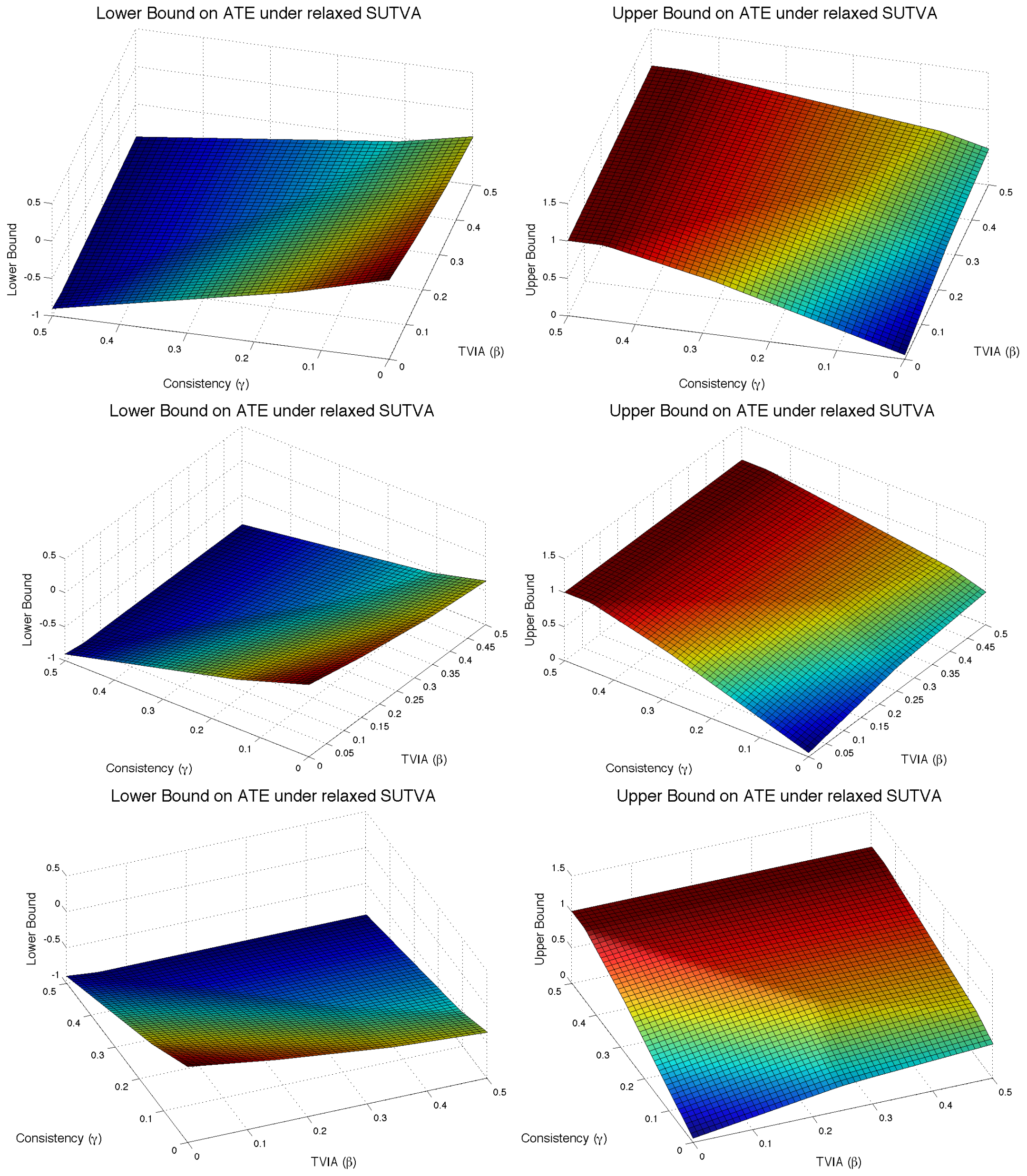


Figure 4: Individual Average Treatment Effect depicted on the support of the joint probability distribution of $(Y^{00}, Y^{01}, Y^{10}, Y^{11}, Y, D, H)$ for binary hidden treatment H. Note that only the proportions $\Pr(Y = y, D = d)$ are observed. Both Assumptions 1A (highlighted with green rectangles) and 1B (depicted by black dots) are restrictions on the support of $(Y^{00}, Y^{01}, Y^{10}, Y^{11}, Y, D, H)$.



Lower Bound on ATE under relaxed SUTVA

Upper Bound on ATE under relaxed SUTVA

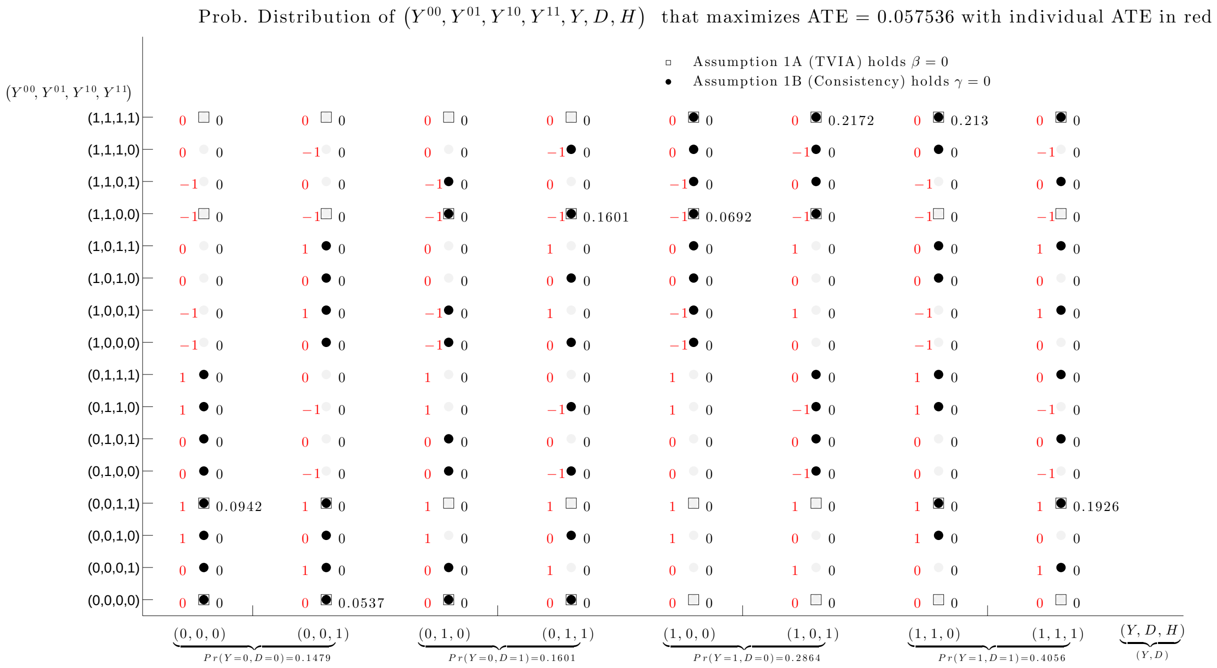
Lower Bound on ATE under relaxed SUTVA

Upper Bound on ATE under relaxed SUTVA

Lower Bound on ATE under relaxed SUTVA

Upper Bound on ATE under relaxed SUTVA

Figure 5: Lower and upper bounds on the ATE (viewed from different angles) under different relaxations of Assumptions 1A and 1B.



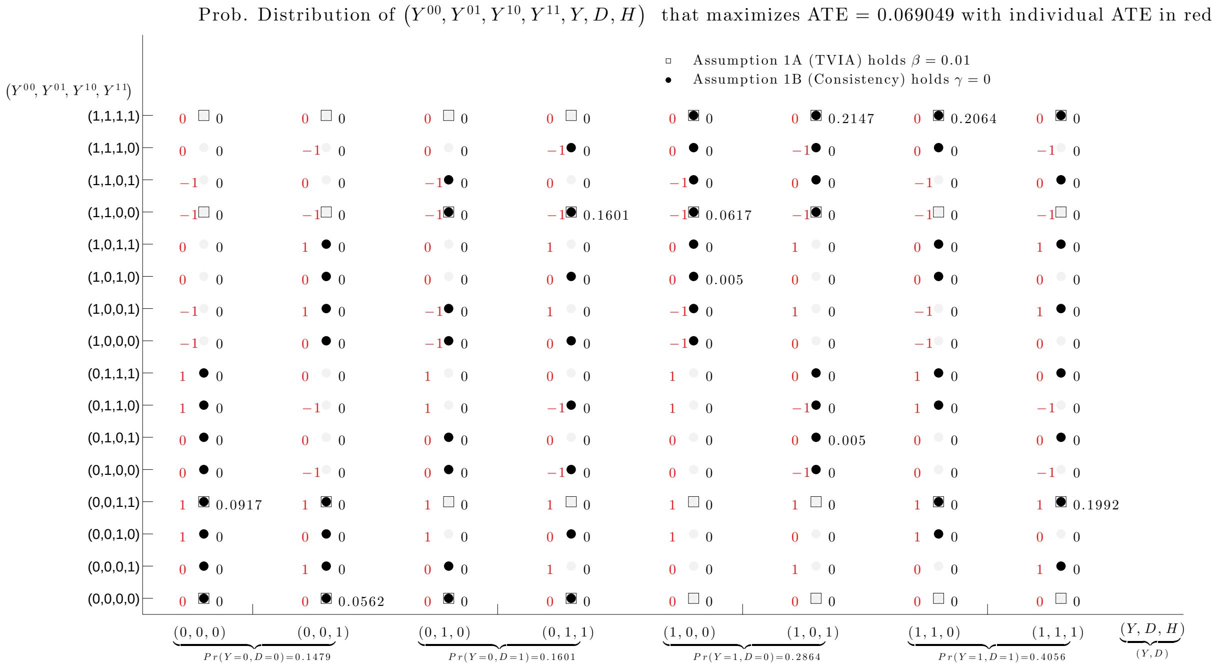
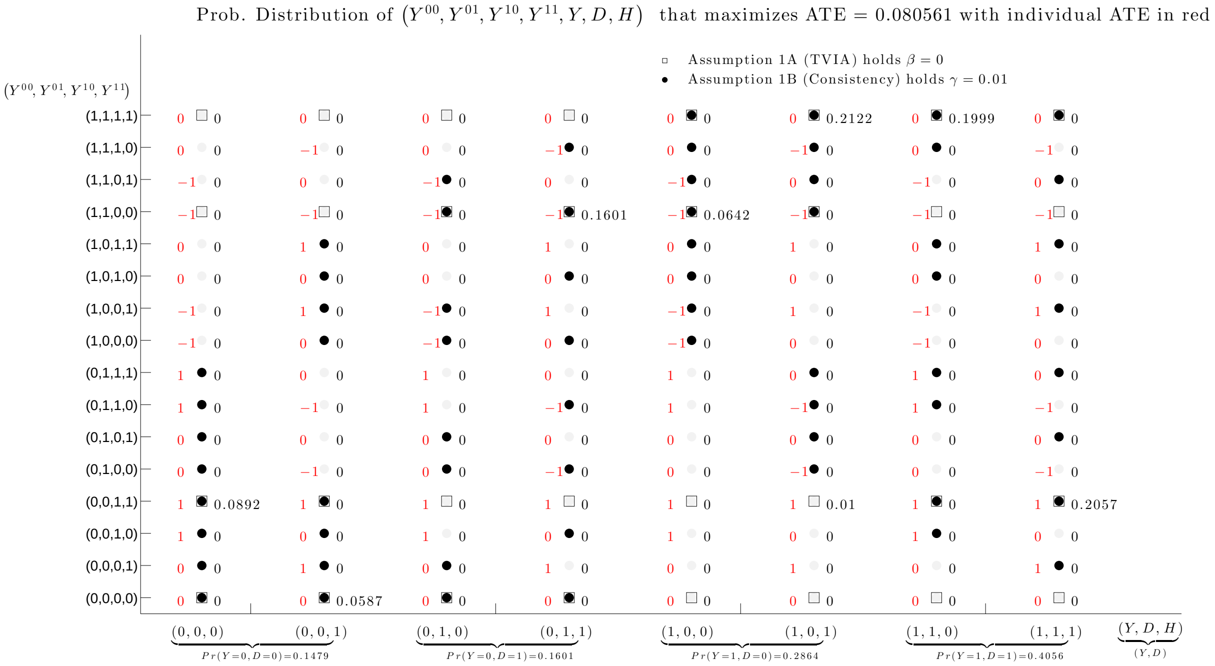


Figure 6: Joint probability distributions of $(Y^{00}, Y^{01}, Y^{10}, Y^{11}, Y, D, H)$ that maximize ATE under different relaxations of SUTVA.



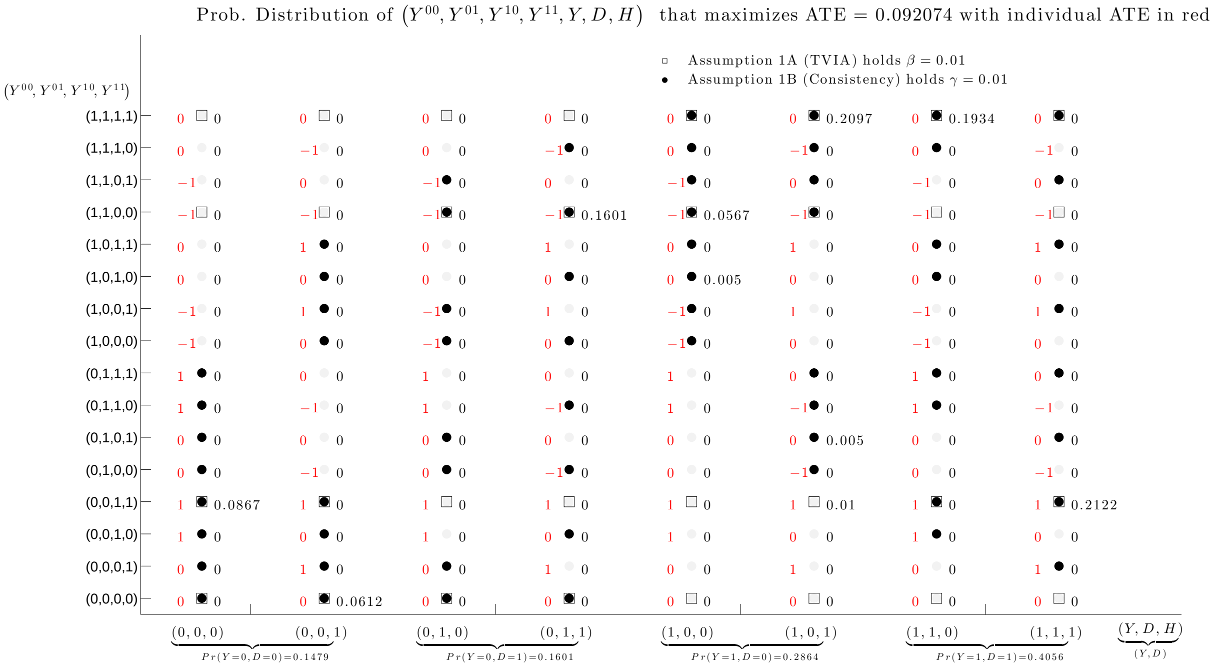


Figure 7: Joint probability distributions of $(Y^{00}, Y^{01}, Y^{10}, Y^{11}, Y, D, H)$ that maximize ATE under different relaxations of SUTVA.

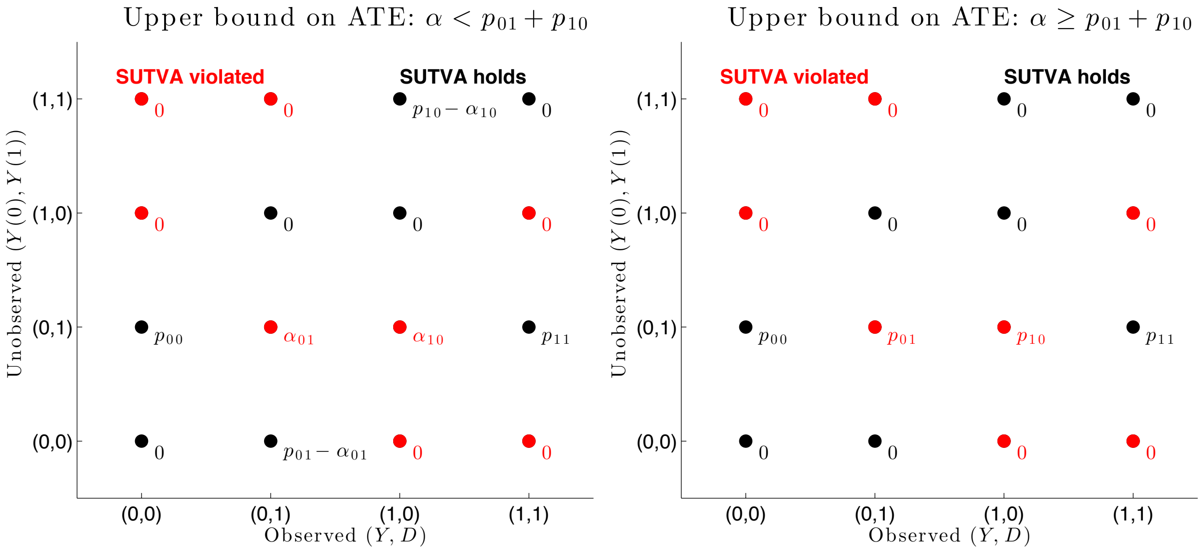


Figure 8: Visualization of the sharpness part of Lemma 3.

C Continuous outcome

Notation: $\forall d \in \mathcal{D} : \pi^d = \pi^d(y_0, y_1, y) = f(y_0, y_1, y|d)$, $p^d = \Pr(D = d)$

$$\begin{aligned} \forall y \in \mathcal{Y} : \iint \pi^1(y_0, y_1, y)\ dy_0\ dy_1 &= f_Y(y|D = 1) \\ \iint \pi^0(y_0, y_1, y)\ dy_0\ dy_1 &= f_Y(y|D = 0) \end{aligned} \tag{C.1}$$

$$\begin{aligned} \forall y_0, y_1, y \in \mathcal{Y} : \pi^1(y_0, y_1, y) I\{y_1 \neq y\} = 0 \\ \forall y_0, y_1, y \in \mathcal{Y} : \pi^0(y_0, y_1, y) I\{y_0 \neq y\} = 0 \end{aligned} \tag{C.2}$$

$$\begin{aligned} \iiint y_1 \pi^1\ dy_0\ dy_1\ dy &= \iiint y_1 \pi^0\ dy_0\ dy_1\ dy \\ \iiint y_0 \pi^1\ dy_0\ dy_1\ dy &= \iiint y_0 \pi^0\ dy_0\ dy_1\ dy \end{aligned} \tag{C.3}$$

These restrictions state that π^d is compatible with the data (C.1), satisfy SUTVA assumption (C.2) and the Exogenous Treatment Selection assumption (C.3).

Given that $\pi^d \geq 0$ and t, conditions (C.2) can be rewritten as:

$$\iiint \pi^1(y_0, y_1, y) I\{y_1 \neq y\} + \pi^0(y_0, y_1, y) I\{y_0 \neq y\}\ dy_0\ dy_1\ dy = 0$$

and we can rewrite relaxed SUTVA (Assumption 1α) as

$$\iiint \pi^1(y_0, y_1, y) I\{y_1 \neq y\} + \pi^0(y_0, y_1, y) I\{y_0 \neq y\}\ dy_0\ dy_1\ dy \leq \alpha. \tag{C.4}$$

The $ATE = E[Y^1 - Y^0]$ can be rewritten in terms of π^d in the following way:

$$ATE = \iiint (y_1 - y_0)(\pi^1 p^1 + \pi^0 p^0)\ dy_0\ dy_1\ dy. \tag{C.5}$$

In order to to find meaningful bounds without the ETS assumption, we will need bounded support of the outcome, suppose now that $y \in \mathcal{Y} \subset [y_{\min}, y_{\max}]$.

Lemma 8. *Under Assumption 1α, the sharp bounds on the ATE are the following:*

$$\begin{aligned} ATE &\in [ATE^{LB}, ATE^{UB}] \\ ATE^{LB} &= \max\left\{ p^1\left(E[Y|D=1] - y_{\max}\right) + p^0\left(y_{\min} - E[Y|D=0]\right) - \alpha(y_{\max} - y_{\min}),\ -(y_{\max} - y_{\min}) \right\} \\ ATE^{UB} &= \min\left\{ p^1\left(E[Y|D=1] - y_{\min}\right) + p^0\left(y_{\max} - E[Y|D=0]\right) + \alpha(y_{\max} - y_{\min}),\ y_{\max} - y_{\min} \right\}. \end{aligned} \tag{C.6}$$

Proof of Lemma 8. We show the proof for the upper bound as the proof for the lower bound follows in an analogous way.

(i) Validity

$$
\begin{aligned}
ATE &= \iiint (y_1 - y_0)(\pi^1 p^1 + \pi^0 p^0)\; dy_0\; dy_1\; dy \\
&= p^1 \iiint (y_1 - y_0)\pi^1\; dy_0\; dy_1\; dy \\
&+ p^0 \iiint (y_1 - y_0)\pi^0\; dy_0\; dy_1\; dy \\
&= p^1 \iiint y_1\pi^1\; dy_0\; dy_1\; dy - p^1 \iiint y_0\pi^1\; dy_0\; dy_1\; dy \\
&+ p^0 \iiint y_1\pi^0\; dy_0\; dy_1\; dy - p^0 \iiint y_0\pi^0\; dy_0\; dy_1\; dy \\
&= p^1\; \iiint (y_1 - y_0)\left[\pi^1 I\{y_1 = y\} + \pi^1 I\{y_1 \neq y\}\right] dy_0\; dy_1\; dy \\
&+ p^0\; \iiint (y_1 - y_0)\left[\pi^0 I\{y_0 = y\} + \pi^0 I\{y_0 \neq y\}\right] dy_0\; dy_1\; dy \\
&\leq p^1\left(E[Y|D=1] - y_{\min}\right) \\
&+ p^1(y_{\max} - y_{\min}) \iiint \pi^1(y_0, y_1, y) I\{y_1 \neq y\}\; dy_0\; dy_1\; dy \\
&+ p^0\left(y_{\max} - E[Y|D=0]\right) \\
&+ p^0(y_{\max} - y_{\min}) \iiint \pi^0(y_0, y_1, y) I\{y_0 \neq y\}\; dy_0\; dy_1\; dy \\
&= p^1\left(E[Y|D=1] - y_{\min}\right) + p^0\left(y_{\max} - E[Y|D=0]\right) + \alpha(y_{\max} - y_{\min})
\end{aligned}
\tag{C.7}
$$

(ii) Sharpness

The following specification for π^d is compatible with Assumption 1α, with the distribution of (Y, D) and achieves the ATE^{UB}. Note that for $\alpha \leq p^1 E[Y|D=1] + p^0 E[Y|D=0]$ there exists α_0, α_1 such that $\alpha_0 \leq p^0 E[Y|D=0]$, $\alpha_1 \leq p^1 E[Y|D=1]$ and $\alpha = \alpha_0 + \alpha_1$. For $\alpha \leq p^1 E[Y|D=1] + p^0 E[Y|D=0]$:

$$
\begin{aligned}
\pi^0(y_0, y_1, y) &= ((1-\alpha_0) I\{y_0 = y\} + \alpha_0 I\{y_0 = y_{\min}\}) \cdot I\{y_1 = y_{\max}\} \cdot f_Y(y|D=0), \\
\pi^1(y_0, y_1, y) &= I\{y_0 = y_{\min}\} \cdot ((1-\alpha_1) I\{y_1 = y\} + \alpha_1 I\{y_1 = y_{\max}\}) \cdot f_Y(y|D=1),
\end{aligned}
\tag{C.8}
$$

and for $\alpha > p^1 E[Y|D=1] + p^0 E[Y|D=0]$ we set $\alpha_0 = p^0 E[Y|D=0]$ and $\alpha_1 = p^1 E[Y|D=1]$.

□

References

Angrist, J., E. Bettinger, and M. Kremer (2006): "Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia," *The American Economic Review*, 847–862.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

Cole, S. R. and C. E. Frangakis (2009a): "The consistency statement in causal inference: a definition or an assumption?" *Epidemiology*, 20, 3–5.

——— (2009b): "The consistency statement in causal inference: a definition or an assumption?" *Epidemiology*, 20, 3–5.

Cox, D. R. (1958): *Planning of experiments.*, Wiley.

Demuynck, T. (2015): "Bounding average treatment effects: A linear programming approach," *Economics Letters*, 137, 75–77.

Forastiere, L., E. M. Airoldi, and F. Mealli (2016): "Identification and estimation of treatment and interference effects in observational studies on networks," *arXiv preprint arXiv:1609.06245*.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The economics and econometrics of active labor market programs," *Handbook of labor economics*, 3, 1865–2097.

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

Horowitz, J. L. and C. F. Manski (1995): "Identification and robustness with contaminated and corrupted data," *Econometrica: Journal of the Econometric Society*, 281–302.

Hsieh, Y.-W., X. Shi, and M. Shum (2018): "Inference on estimators defined by mathematical programming," *Available at SSRN 3041040*.

Huber, M. and A. Steinmayr (2019): "A framework for separating individual-level treatment effects from spillover effects," *Journal of Business & Economic Statistics*, 1–15.

Imbens, G. W. and D. B. Rubin (2015): *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press.

Kaido, H., F. Molinari, and J. Stoye (2019): "Confidence intervals for projections of partially identified parameters," *Econometrica*, 87, 1397–1432.

Kaido, H., F. Molinari, J. Stoye, and M. Thirkettle (2017): "Calibrated Projection in MATLAB: Users' Manual," *arXiv preprint arXiv:1710.09707*.

Lafférs, L. (2019): "Bounding average treatment effects using linear programming," *Empirical Economics*, 57, 727–767.

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

Miguel, E. and M. Kremer (2004): "Worms: identifying impacts on education and health in the presence of treatment externalities," *Econometrica*, 72, 159–217.

Pearl, J. (2010): "On the consistency rule in causal inference: axiom, definition, assumption, or theorem?" *Epidemiology*, 21, 872–875.

Petersen, M. L. (2011): "Compound treatments, transportability, and the structural causal model: the power and simplicity of causal graphs," *Epidemiology*, 22, 378–381.

Rubin, D. B. (1980): "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment," *Journal of the American Statistical Association*, 75, 591–593.

VanderWeele, T. J. (2009a): "Concerning the consistency assumption in causal inference," *Epidemiology*, 20, 880–883.

——— (2009b): "Concerning the consistency assumption in causal inference," *Epidemiology*, 20, 880–883.