Identification of Positive Treatment Effects in Randomized Experiments with Non-Compliance*

Aleksey Tetenov†

February 18, 2012

Abstract

I derive sharp nonparametric lower bounds on some parameters of the full distribution of individual treatment effects from experimental data with non-compliance. These parameters are: $P(Y(1) - Y(0) > c)$, the proportion of individuals whose benefit from treatment exceeds c, $E(Y(1) - Y(0) - c)^+$, the sum of positive treatment effects exceeding c, and $E|Y(1) - Y(0)|$, the total absolute value of individual effects. I show that sharp bounds on these parameters for $c = 0$ depend only on the distributions of observed outcomes in assigned treatment and control groups and observing compliance does not improve them.

Keywords: bounds; partial identification; intent-to-treat; public goods

JEL codes: C14; C21; C26; H43

*I am grateful to Kei Hirano, Chuck Manski, Franco Peracchi, Joerg Stoye and seminar participants at Univ. of Virginia, NYU and Univ. of Arizona for helpful comments.

†Collegio Carlo Alberto, Moncalieri (Torino), Italy. aleksey.tetenov@carloalberto.org

1 Introduction

I derive sharp lower bounds on parameters of the population distribution of individual treatment effects from experimental data with non-compliance. The first parameter of interest is the sum of positive treatment effects exceeding a given constant c, $E\left(Y\left(1\right)-Y\left(0\right)-c\right)^{+}$. If individuals could choose the treatment that is best for them and must pay an additional cost c for treatment 1, then the population welfare when both choices are available equals $E\max\left(Y\left(0\right),Y\left(1\right)-c\right)$. If only treatment 0 were available, the population welfare would be $EY\left(0\right)$. Then

$$E\max\left(Y\left(0\right),Y\left(1\right)-c\right)-EY\left(0\right)=E\left(Y\left(1\right)-Y\left(0\right)-c\right)^{+} \tag{1}$$

is the welfare gained by making both choices available. If making the choice of treatment 1 available to the population of interest carries a fixed cost C, then it is welfare optimal to do so when $E\left(Y\left(1\right)-Y\left(0\right)-c\right)^{+}>C$. If access to treatment 1 is a pure public good with $c=0$, then the parameter of interest becomes $E\left(Y\left(1\right)-Y\left(0\right)\right)^{+}$.

The second parameter of interest is

$$P\left(Y\left(1\right)-Y\left(0\right)>c\right), \tag{2}$$

the proportion of population for whom additional benefit from treatment 1 exceeds c or the proportion that would use treatment 1 if this choice was made available.

The third parameter is the total absolute size of treatment effects:

$$E\left|Y\left(1\right)-Y\left(0\right)\right|=E\max\left(Y\left(0\right),Y\left(1\right)\right)-E\min\left(Y\left(0\right),Y\left(1\right)\right). \tag{3}$$

This is the width of the bounds in the mixing problem of Manski (1997a), capturing the difference in welfare between the best and the worst possible assignments of individuals to the two treatments.

An attractive feature of the bounds on $E\left(Y\left(1\right)-Y\left(0\right)\right)^{+}$, $P\left(Y\left(1\right)>Y\left(0\right)\right)$, and $E\left|Y\left(1\right)-Y\left(0\right)\right|$ is that they are simple functions only of the cdfs of observed outcomes

among individuals randomized to treatment and control groups, even if it is also known which individuals complied with assigned treatment and which didn't. While this looks like "intent-to-treat" analysis, the bounds are for parameters of the distribution of the effects of actual treatment. This makes the bounds particularly useful for analyzing experiments in which compliance is difficult to monitor or where compliance data is not reliable. Unlike bounds on average effects, they are not sensitive to the specification of the range of feasible outcomes, which makes them reasonable even for outcomes with a wide range of feasible values (e.g., earnings, longevity) without additional theoretical or shape restrictions.

The paper makes contributions to two strands of research on identification of treatment effects. First, to the literature on partial identification in experiments with noncompliance, which primarily focused on bounding the average treatment effect, e.g. Manski (1996, 2003, 2007), Balke and Pearl (1997), Kitagawa (2009). Second, to the literature on partial identification of parameters of the joint distribution of treatment outcomes given the knowledge of the marginal distributions, e.g. Heckman, Smith and Clements (1997), Firpo and Ridder (2008), Fan and Park (2010), Fan and Zhu (2009).

The paper proceeds as follows: I first describe the setting and the assumptions made, then review general results for bounding parameters of a joint distribution with known marginals, then apply them to randomized experiments with perfect or unobserved compliance, and then study the additional identification power of data on compliance.

2 The setting

The environment of this paper is a standard Neyman-Rubin potential outcomes framework. There is a set of treatments $\mathcal{D}$. I consider two cases: $\mathcal{D} = \{0, 1\}$ (two treatments) and $\mathcal{D} = [0, 1]$, denoting different potential "intensity" of treatment 1. The potential outcome of each treatment d (if the individual were actually to receive treatment d) is denoted by $Y(d), d \in \mathcal{D}$. It is a maintained assumption embedded in the framework that treatment response does not depend on treatments assigned or received by other

individuals or on the process by which the treatments are assigned. Treatment actually received by an individual in an experiment is denoted by $D \in \mathcal{D}$. Observed outcome for an individual receiving treatment D will be $Y = Y(D)$.

Binary variable $Z \in \{0, 1\}$ denotes the treatment to which individual is assigned in an experiment (a binary instrument). Even when considering $\mathcal{D} = [0, 1]$, the assignment is assumed to be binary. The treatment received by an individual is denoted by $D = D(Z)$. The pair of variables $(D(0), D(1))$ describes potential compliance of an individual with randomized assignment in an experiment. To simplify notation, the analysis of this paper is performed for a population of observationally identical individual (with no observable covariates). If the assumptions are valid conditional on covariate values, sharp population could be obtained by integrating conditional bounds across covariates as in Fan and Zhu (2009).

Observable experimental data identifies the probability distribution $P(Y, Z, D)$ if compliance is observed or just the distribution $P(Y, Z)$ if compliance is not observable.

The assumption that experimental treatment assignment Z is independent of potential outcomes $Y(d)$ and potential compliance with assigned treatment $D(z)$ is maintained throughout the paper:

Assumption R (Randomization):

$$Z \perp (D(0), D(1), Y(d), d \in \mathcal{D}). \tag{4}$$

This assumption is empirically refutable for some distributions $P(Y, Z, D)$. Kitagawa (2009 and 2010) studies the full identification power of Assumption R for the joint distribution of $(Y(0), Y(1))$ when $\mathcal{D} = \{0, 1\}$ and constructs a test of consistency between the observable data and Assumption R.

The condition that potential outcomes are bounded is maintained for simplicity, it could be weakened by imposing restrictions on tails of the distribution of outcomes.

Assumption B (Boundedness):

$$|Y(d)| < B, d \in \mathcal{D} \text{ for some } B \in \mathbb{R}. \tag{5}$$

The value B could be chosen to be arbitrarily large, it is not used substantively for deriving the bounds on parameters considered in this paper (the restrictions on the range of Y would play a substantive role, however, for upper bounds on the parameters or for bounding average treatment effects, see for example Manski 2003, 2007).

I also assume throughout the paper that treatment assignment Z has a non-negative effect on the intensity of treatment $D(Z)$ received by each individual:

Assumption M (Monotonicity):

$$D(1) \geq D(0). \tag{6}$$

For $\mathcal{D} = \{0, 1\}$ this assumption is equivalent to the condition that there are no "deniers" who always take the opposite of assigned treatment, so $P(D(0) = 1, D(1) = 0) = 0$. Imbens and Angrist (1994), and Angrist, Imbens and Rubin (1996) used it to identify the average treatment effect in a subpopulation of "compliers" $(D(0) = 0, D(1) = 1)$. Assumption M is also empirically refutable for some distributions $P(Y, Z, D)$ if Assumption R is maintained. If $P(D = 1|Z = 0) + P(D = 0|Z = 1) > 1$, for example, then necessarily $P(D(0) = 1, D(1) = 0) > 0$.

Results for treatment with varying intensity $\mathcal{D} = [0, 1]$ will rely on an additional assumption that each individual's response to treatment is monotonic in $d \in \mathcal{D}$. It is weaker than the *monotone treatment response* assumption in Manski (1997b), allowing treatment response to be positive for some individuals and negative for others.

Assumption H (Heterogeneous Monotone Treatment Response):

For each individual, $Y(d)$ is a weakly increasing or a weakly decreasing function of d.

Clearly, Assumption H is trivially satisfied for $\mathcal{D} = \{0, 1\}$, so it will be maintained throughout the paper both for $\mathcal{D} = \{0, 1\}$ and for $\mathcal{D} = [0, 1]$.

3 Bounds on parameters of the joint distribution of two random variables

Derivation of the bounds for treatment effects relies on two general results for bounding parameters of the joint distribution of two random variables with known marginals, which I review in this section. Let W_0 and W_1 be two arbitrary random variables with known marginal distribution functions G_0 and G_1 and unknown joint distribution.

Sharp bounds on $E\left(W_1 - W_0 - c\right)^+$ and $E\left|W_1 - W_0\right|$ could be obtained from a result due to Cambanis, Simons and Stout (1976). They show that if $k\left(x, y\right)$ is a right-continuous quasi-antitone (sub-modular) function, that is

$$\forall x \leq x', y \leq y' : k\left(x, y\right) + k\left(x', y'\right) \leq k\left(x, y'\right) + k\left(x', y\right), \tag{7}$$

and $G_0\left(x\right), G_1\left(y\right)$ are marginal distributions of random variables W_0, W_1 with unknown joint distribution $H\left(.,.\right)$, then $E_H k\left(W_0, W_1\right)$ is minimized/maximized at the Fréchet-Hoeffding upper and lower bounds on the joint distribution of $\left(W_0, W_1\right)$:

$$\begin{aligned} H_+\left(x, y\right) &\equiv \min\left\{G_0\left(x\right), G_1\left(y\right)\right\}, \\ H_-\left(x, y\right) &\equiv \max\left\{G_0\left(x\right) + G_1\left(y\right) - 1, 0\right\} \end{aligned} \tag{8}$$

(joint distributions with maximal positive and negative dependence). Sharp bound on $Ek\left(W_0, W_1\right)$ could be explicitly evaluated as a function of the quantiles of G_0 and G_1:

$$\begin{aligned} E_H k\left(W_0, W_1\right) &\geq E_{H_+} k\left(W_0, W_1\right) = \int_0^1 k({G_0}^{-1}\left(u\right), G_1^{-1}\left(u\right))du, \\ E_H k\left(W_0, W_1\right) &\leq E_{H_-} k\left(W_0, W_1\right) = \int_0^1 k({G_0}^{-1}\left(u\right), G_1^{-1}\left(1-u\right))du. \end{aligned} \tag{9}$$

This result has been applied to identification of other parameters of the treatment effect distribution (e.g., variance, correlation coefficient, inequality measures) by Heckman, Smith and Clements (1997), and by Fan and Zhu (2009) when full marginal distributions are point-identified.

Functions $(W_1 - W_0 - c)^+$ and $|W_1 - W_0|$ and quasi-antitone in (W_0, W_1) are continuous. We could therefore apply the previous result to derive sharp lower bounds on their expectations over the set of all joint distributions of (W_0, W_1) with given marginals G_0 and G_1:

$$E\left(W_1 - W_0 - c\right)^+ \geq \int_0^1 (G_1^{-1}(u) - G_0^{-1}(u) - c)^+ du = \int_{\mathbb{R}} \left(G_0(t-c) - G_1(t)\right)^+ dt, \quad (10)$$

$$E\left|W_1 - W_0\right| \geq \int_0^1 |G_1^{-1}(u) - G_0^{-1}(u)| du = \int_{\mathbb{R}} |G_0(t) - G_1(t)| dt. \quad (11)$$

Sharp lower bound on $P(W_1 - W_0 > c)$ cannot be obtained from the result of Cambanis, Simons and Stout since functions $\mathrm{I}[W_1 - W_0 > c]$ are neither sub-modular nor super-modular. Instead, it immediately follows from the sharp bounds on the distribution function of the sum of random variables with known marginals derived by Makarov (1981).

From Williamson and Downs (1990, Theorem 2) we get the following expression for the upper Makarov bound on the distribution of $W_1 - W_0$:

$$P(W_1 - W_0 \leq c) \leq 1 + \inf_t \left\{\min\left(G_1(t) - G_0(t-c), 0\right)\right\}. \quad (12)$$

Then the sharp lower bound on $P(W_1 - W_0 > c)$ is equal to

$$\begin{aligned} P(W_1 - W_0 > c) &\geq \sup_{t\in\mathbb{R}} \left\{\max\left(G_0(t-c) - G_1(t), 0\right)\right\} \\ &= \sup_{t\in\mathbb{R}} \left\{G_0(t-c) - G_1(t)\right\}. \end{aligned} \quad (13)$$

In particular, $P(W_1 > W_0) \geq \sup\{G_0(t) - G_1(t)\}$.

4 Bounds with full or unobserved compliance

First, let's consider two cases in which bounds on parameters depend only on the observed distribution of outcomes in assigned treatment and control groups. Denote the distribution function of the random variable $Y(D(0))$, which is the observed outcome

if an individual is randomized to treatment 0, by $F_0(\cdot)$ and the distribution function of $Y(D(1))$ by $F_1(\cdot)$. Proposition 1 establishes sharp lower bounds in the simple cases of full compliance and of completely unobserved compliance.

Proposition 1 *If Assumptions R, B, M and H hold and either*
a) there is full compliance, so $D(0) = 0$ *and* $D(1) = 1$, *or*
b) only $P(Y, Z)$ *is observed (the received treatment* D *is unobserved) and* $c \geq 0$,
then the following bounds are sharp:

$$E(Y(1) - Y(0) - c)^+ \geq \int_{\mathbb{R}} (F_0(t-c) - F_1(t))^+ dt, \tag{14}$$

$$P(Y(1) - Y(0) > c) \geq \sup_{t \in \mathbb{R}} \{F_0(t-c) - F_1(t)\}, \tag{15}$$

$$\text{and } E|Y(1) - Y(0)| \geq \int_{\mathbb{R}} |F_0(t) - F_1(t)| dt. \tag{16}$$

Proof: a) If there is full compliance with randomized treatment, that is:

$$P(D(Z) = z | Z = z) = 1, z \in \{0, 1\}, \tag{17}$$

then the experimental data point-identifies the marginal distributions of $Y(0)$ and $Y(1)$, which are equal to the observed distributions F_0, F_1 of $Y|Z = 0$ and $Y|Z = 1$, but places no other restrictions on the joint distribution of $(Y(0), Y(1))$. The bounds on parameters of the distribution of individual treatment effects $Y(1) - Y(0)$ then follow immediately from the results in the previous section.

b) First, I will show that for $c \geq 0$ Assumptions M and H imply that for each individual

$$[Y(1) - Y(0) - c]^+ \geq [Y(D(1)) - Y(D(0)) - c]^+. \tag{18}$$

If $D(1) = D(0)$ or if $Y(D(1)) - Y(D(0)) - c \leq 0$, then the inequality is satisfied because the right hand side equals zero.

If $Y(D(1)) - Y(D(0)) - c > 0$, then $Y(D(1)) - Y(D(0)) > 0$ because $c \geq 0$. This implies $D(1) \neq D(0)$, and hence, due to Assumption M, $D(1) > D(0)$. Then it follows from Assumption H that this individual's response to treatment must be non-

decreasing: $Y(1) \geq Y(D(1)) > Y(D(0)) \geq Y(0)$ and therefore $Y(1) - Y(0) - c \geq Y(D(1)) - Y(D(0)) - c$ and the inequality (18) is also satisfied.

It could be similarly shown that for $c \leq 0$,

$$[Y(1) - Y(0) - c]^{-} \geq [Y(D(1)) - Y(D(0)) - c]^{-}. \tag{19}$$

Since both (18) and (19) are true for $c = 0$, also $|Y(1) - Y(0)| \geq |Y(D(1)) - Y(D(0))|$.

Integrating these inequalities shows that for $c \geq 0$:

$$\begin{aligned} E[Y(1) - Y(0) - c]^{+} &\geq E[Y(D(1)) - Y(D(0)) - c]^{+}, \\ P[Y(1) - Y(0) > c] &\geq P[Y(D(1)) - Y(D(0)) > c], \\ \text{and } E|Y(1) - Y(0)| &\geq E|Y(D(1)) - Y(D(0))|. \end{aligned} \tag{20}$$

The experimental data point-identifies the distributions of variables $Y(D(0))$ and $Y(D(1))$, so the lower bounds on parameters of their joint distribution from the previous section apply and extend to $Y(0)$ and $Y(1)$ by the virtue of (20).

To see that the bounds are sharp, take a joint distribution of $Y(D(0))$ and $Y(D(1))$ that attains one of the lower bounds, then suppose that for all individuals $D(0) = 0$, $D(1) = 1$, $Y(0) = Y(D(0))$, and $Y(d) = Y(D(1))$ for $d > 0$. This distribution of latent variables attains the same lower bound for functions of $Y(1) - Y(0)$ as for functions of $Y(D(1)) - Y(D(1))$ and is consistent with the observed distribution of (Y, Z).

□

5 Bounds with observable compliance

The bounds in the previous section remain correct if compliance with assigned treatments is observable. Observation of D, however, may reduce the set of joint distributions of $(Y(0), Y(1))$ consistent with the data and may sharpen the bounds. Below I derive the sharp bounds for the binary treatment case $\mathcal{D} = \{0, 1\}$ when variables (Y, Z, D) are observable.

As before, let F_0, F_1 be the observed distributions of $Y|Z=0$ and $Y|Z=1$.

Conditional distributions of potential outcomes $Y(0)$ and $Y(1)$ in the subpopulation of compliers, which will be denoted by

$$F_{C,d}(t) = P(Y(d) \leq t | D(0) < D(1)), d \in \{0,1\} \tag{21}$$

are point-identified by the data. Abadie (2002, eqs. 5, 6) shows that they are equal to

$$F_{C,0}(t) = P_C^{-1}[P(Y \leq t, D=0|Z=0) - P(Y \leq t, D=0|Z=1)], \tag{22}$$

$$\text{and } F_{C,1}(t) = P_C^{-1}[P(Y \leq t, D=1|Z=1) - P(Y \leq t, D=1|Z=0)], \tag{23}$$

where

$$P_C = P(D=1|Z=1) - P(D=1|Z=0) \tag{24}$$

is the proportion of compliers in the population. Also, Abadie shows that the difference between the two cdfs any point is proportional to the difference between the cdfs of observed outcomes in assigned treatment and control groups:

$$F_{C,1}(t) - F_{C,0}(t) = P_C^{-1} \cdot (F_1(t) - F_0(t)). \tag{25}$$

Proposition 2 *If $D = \{0,1\}$ and Assumptions R, B, and M hold, then the following bounds are sharp given the observed distribution $P(Y,Z,D)$:*

$$E(Y(1) - Y(0) - c)^+ \geq P_C \int_{\mathbb{R}} (F_{C,0}(t-c) - F_{C,1}(t))^+ dt, \tag{26}$$

$$P(Y(1) - Y(0) > c) \geq P_C \sup_{t \in \mathbb{R}} \{F_{C,0}(t-c) - F_{C,1}(t)\}, \tag{27}$$

$$\textit{and } E|Y(1) - Y(0)| \geq P_C \int_{\mathbb{R}} |F_{C,0}(t) - F_{C,1}(t)| dt. \tag{28}$$

Proof: The observable data point-identifies the proportion of population that falls into each of three principal strata based on potential compliance, including the proportion of compliers P_C given in (24). Since each of the population parameters on the left hand side is an expectation of some function of $(Y(1) - Y(0))$, each is bounded below by the

sum of lower bounds for each of the three strata multiplied by the probability of each type.

For the compliers, the data point-identifies the marginal distributions of potential outcomes of both treatments $F_{C,0}, F_{C,1}$, but places no restrictions on their joint distribution. The sharp lower bounds on the values of all parameters follow from Proposition 1 with F_0, F_1 replaced by $F_{C,0}, F_{C,1}$.

For always-takers and never-takers, the data point identifies the distribution of potential outcomes of one treatment, but places no restrictions on the distribution of the other treatment. Hence distributions of latent outcomes for which $Y(1) = Y(0)$ for all non-compliers are consistent with the data. Such distributions attain the lower bound for non-compliers on all the parameters of interest, which is equal to zero.

□

Substituting (25) into the expressions for the bounds for $c = 0$ shows that observability of actually received treatment D does not yield sharper lower bounds on $E(Y(1) - Y(0))^+$, $P(Y(1) > Y(0))$, and $E|Y(1) - Y(0)|$.

Corollary 3 *If $D = \{0, 1\}$ and Assumptions R, B, and M hold, then the following bounds are sharp given the observed distribution $P(Y, Z, D)$:*

$$E(Y(1) - Y(0) - c)^+ \geq \int_{\mathbb{R}} (F_0(t) - F_1(t))^+ dt, \tag{29}$$

$$P(Y(1) - Y(0) > c) \geq \sup_{t \in \mathbb{R}} \{F_0(t) - F_1(t)\}, \tag{30}$$

$$\text{and } E|Y(1) - Y(0)| \geq \int_{\mathbb{R}} |F_0(t) - F_1(t)| dt. \tag{31}$$

The following example shows that for $c > 0$ lower bounds in Proposition 2 (with observable D) could be higher than those in Proposition 1.

Example: Consider the following probability distribution of latent compliance and

outcome variables:

Proportion:	$D(0)$	$D(1)$	$Y(0)$	$Y(1)$
0.25	0	1	0	2
0.25	0	1	1	2
0.5	1	1	1	1

Let $c = 1$, then $P(Y(1) - Y(0) > 1) = 0.25$ and $E(Y(1) - Y(0) - 1)^+ = 0.25$.

For bounds in Proposition 1,

$$\begin{aligned} F_0(t) &= P(Y(D(0)) \leq t) = 0.25\mathrm{I}[t \geq 0] + .75\mathrm{I}[t \geq 1], \\ F_1(t) &= P(Y(D(1)) \leq t) = 0.5\mathrm{I}[t \geq 1] + .5\mathrm{I}[t \geq 2]. \end{aligned}$$

The lower bound on $P(Y(1) - Y(0) > 1)$ equals $\sup\{F_0(t-1) - F_1(t)\} = 0$, the lower bound on $E(Y(1) - Y(0) - 1)^+$ is also equal to zero.

Instead, for bounds in Proposition 2, $P_C = 0.5$ and

$$\begin{aligned} F_{C,0}(t) &= 2[P(Y \leq t, D = 0|Z = 0) - P(Y \leq t, D = 0|Z = 1)] \\ &= 0.5\mathrm{I}[t \geq 0] + .5\mathrm{I}[t \geq 1], \\ F_{C,1}(t) &= 2[P(Y \leq t, D = 1|Z = 1) - P(Y \leq t, D = 1|Z = 0)] \\ &= \mathrm{I}[t \geq 2]. \end{aligned}$$

Then the lower bound on $P(Y(1) - Y(0) > 1)$ equals $0.5 \sup\{F_{C,0}(t-1) - F_{C,1}(t)\} = 0.25$, the lower bound on $E(Y(1) - Y(0) - 1)^+$ from Proposition 2 is also equal to 0.25. For $c = 1$ information on compliance with assigned treatments sharpens the parameter bounds, while for $c = 0$ the bounds from Propositions 1 and 2 are identical.

References

[1] Abadie, A. (2002), "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association,* 97, 284-292.

[2] Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association,* 91, 444-455.

[3] Balke, A., and Pearl, J. (1997), "Bounds on Treatment Effects From Studies With Imperfect Compliance," *Journal of the American Statistical Association,* 92, 1171-1176.

[4] Cambanis, S., Simons, G., and Stout, W. (1976), "Inequalities for Ek(X,Y) when the Marginals are Fixed," *Probability Theory and Related Fields,* 36, 285-294.

[5] Heckman, J., Smith, J., and Clements, N. (1997), "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *Review of Economic Studies,* 64, 487–535.

[6] Fan, Y., and Park, S. (2010), "Sharp Bounds on the Distribution of the Treatment Effect and Their Statistical Inference," *Econometric Theory,* 26, 931-951.

[7] Fan, Y., and Zhu, D. (2009), "Partial Identification and Confidence Sets for Functionals of the Joint Distribution of Potential Outcomes," Working paper, Vanderbilt University and HSBC School of Business.

[8] Firpo, S., and Ridder, G. (2008), "Bounds on Functionals of the Distribution of Treatment Effects," Working paper, FGV Brazil.

[9] Imbens, G.W., and Angrist, J.D. (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica,* 62, 467-475.

[10] Kitagawa, T. (2009), "Identification Region of the Potential Outcome Distributions under Instrument Independence," Working paper, University College London.

[11] Kitagawa, T. (2010), "Testing for Instrument Independence in the Selection Model," Working paper, University College London.

[12] Makarov, G.D. (1981), "Estimates for the Distribution Function of a Sum of Two Random Variables when the Marginal Distributions are Fixed," *Theory of Probability and its Applications*, 26, 803-806.

[13] Manski, C.F. (1990), "Nonparametric Bounds on Treatment Effects," *American Economic Review Papers and Proceedings,* 80, 319-323.

[14] Manski, C.F. (1996), "Learning about Treatment Effects from Experiments with Random Assignment of Treatments," *Journal of Human Resources,* 31, 709-733.

[15] Manski, C.F. (1997a), "The Mixing Problem in Programme Evaluation," *Review of Economic Studies,* 64, 537-553.

[16] Manski, C.F. (1997b), "Monotone Treatment Response," *Econometrica*, 65, 1311-1334.

[17] Manski, C.F. (2003), *Partial Identification of Probability Distributions,* Springer-Verlag, New York.

[18] Manski, C.F. (2007), *Identification for Prediction and Decision,* Harvard University Press, Cambridge, Massachusetts.

[19] Williamson, R.C., and Downs, T. (1990), "Probabilistic Arithmetic. I. Numerical Methods for Calculating Convolutions and Dependency Bounds," *International Journal of Approximate Reasoning,* 4, 89-158.