Bounding the Distribution of Treatment Effects*

Sergio Firpo
Departmento de Economia, PUC-Rio
firpo@econ.puc-rio.br

Geert Ridder
Department of Economics, USC
ridder@usc.edu

January 2007

Abstract

We use bounds on the distribution function of the sum of two random variables with known marginal distributions obtained by Makarov (1981) to bound the distribution function of the individual treatment effect. Although the Makarov bounds are pointwise sharp if we only consider the marginals, they can be improved by using the fact that we know the mean (and only the mean) of the treatment effect distribution, i.e. the Average Treatment Effect. We propose a procedure that uses this additional information to obtain more informative bounds. We also show that if the treatment effect varies with observable covariates, averaging over these covariates improves the Makarov bounds. We can combine the two types of additional information by using the known conditional treatment effect means to improve the conditional Makarov bounds and by subsequently averaging these conditional bounds. The (improved) Makarov bounds on the cdf of the treatment effect distribution yield bounds on the quantiles of that distribution. Without covariates we can use the ATE to directly obtain bounds on these quantiles, that turn out to be identical to the bounds obtained by using the ATE to improve the cdf bounds. By this equivalence we can use the average improved conditional Makarov bounds that use information on the conditional ATE to improve the bounds on the unconditional quantiles. Bounds on the conditional quantiles cannot be averaged to obtain bounds on the unconditional quantiles. We illustrate the qualitative features of the bounds for normal and dichotomous outcome distributions.

1 Introduction

The key problem when estimating the effect of a treatment or intervention is that we cannot observe both the treated and non-treated outcomes for the same unit in the population. As a consequence, we can only identify treatment parameters that depend on the marginal distribution of the treated and control outcomes. This is one of the reasons that most studies focus on parameters as the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT), since the mean, either for the whole population or for some subpopulation, is the only functional of the distribution of the treatment effect that can be identified from the marginal distributions of the treated and control outcomes.

Many other estimands of interest in the treatment effect literature only depend on the marginal distributions of treated and non-treated outcomes. They are not, however, functionals

*We have benefited from comments of Rustam Ibragimov who pointed us to the Makarov bounds. Financial support from the NSF and the SSHRC is gratefully acknowledged.

of the distribution of individual treatment effects. For instance, we may be interested in learning how a program may increase the inequality of outcomes. If we choose some inequality measure, say the variance, then the effect of the program on the variance can be identified from difference in variances of the marginal distributions of the treated and control outcomes. Such approach has been used in some studies as in Imbens and Rubin (1997), Abadie, Angrist Imbens (2002), Firpo (2005, 2007), Abadie (2002, 2003).

There are other parameters that require an estimate of the distribution of individual treatment effects. An example is the median (or other quantile) of the treatment effect distribution. If we estimate this by the difference of the medians of the treated and non-treated outcome distributions, then because the difference of quantiles is not equal to the quantile of the difference, this parameter is not equal to the median of the individual treatment effect distribution. Another estimand that requires knowledge of the treatment effect distribution is the fraction of the population that benefits from the intervention. Heckman, Smith and Clements (1997, HSC henceforth) propose focusing on many other parameters of the distribution of treatment effects. In general, if our goal is to assess the effect of an intervention on social welfare and not individual welfare, then the marginal distributions suffice.[1] [2]

Parameters that depend on the distribution of individual treatment effects require knowledge of the joint distribution of treated and non-treated outcomes. If the treatment effect is the same for all members of the population or of subpopulations characterized by a vector of observable variables, this (conditional) joint distribution is singular and the (conditional) distribution of individual treatment effects is degenerate. However, in most cases the observed (conditional) marginal distributions are not related in this way. In that case we can either introduce additional information that allows us to point identify the distribution of treatment effects, or we can as e.g. HSC derive bounds on the distribution of treatment effects.

Bounds on the cumulative distribution function (c.d.f.) of the sum of two random variables with known marginal distributions were first obtained by Makarov (1981) and the generalization to the difference is trivial. However if we know the marginal distributions we also know the mean of the distribution of the treatment effect distribution, the Average Treatment Effect. Only the first moment is identified from the marginal distributions, because higher order moments require knowledge of the joint distribution. We show that this additional knowledge can be used to improve the bounds on the c.d.f. and on the quantiles of the treatment effect distribution. The improvement procedure for the c.d.f. differs from that for the quantiles. Because improved bounds on the c.d.f. can also be used to obtain (improved) bounds on the quantiles, it is important to note that these two approaches give identical bounds, so that the choice of procedure is mainly a matter of convenience. Obviously if one is interested in bounds on a quantile, then it is computationally harder to improve the c.d.f. bounds for all values of the individual treatment effect and to subsequently invert these improved bounds. However, one may wish to pay that computational cost if the data contain covariates that are correlated with the outcomes.

Covariates that are correlated with the outcomes can be used to narrow the bounds on the c.d.f. and on the quantiles This is because the average bounds are in general more informative than the bounds on the average, i.e. the unconditional, outcome distributions. The availability

[1] In the social welfare literature, this is due the "anonymity principle", which is considered a desirable property of social welfare function.

[2] Manski(1997) considers the distribution of the mixture of the treated and non-treated outcome, a distribution that requires knowledge of the joint distribution of the treated and non-treated outcome.

of covariates raises a number of issues. First, with covariates we can recover the conditional mean of the treatment effect distribution given the covariates. This conditional mean function is used to improve the conditional Makarov bounds and the resulting improved bounds are averaged to obtain the unconditional bounds. Second, because the average of the conditional quantiles is not equal to the unconditional quantile (that relation does hold for the conditional c.d.f.), it does not make much sense to use the conditional ATE to improve the bounds on the conditional quantiles. Instead we invert the improved (using the conditional ATE) bounds on the conditional c.d.f. to obtain bounds on the unconditional quantiles. The fact that this gives identical bounds if there are no covariates, shows that this is an efficient way to incorporate conditional mean information in bounds on the unconditional quantiles. Third, if (some of) the covariates are continuous the bounds are obtained by averaging nonparametric estimates of the improved conditional bounds. To do inference, e.g. to compute confidence intervals based on estimates of the bounds as in Imbens and Manski (2004), we need to derive the asymptotic distribution of the estimates of the bounds. In this paper we only consider the case that the covariates are discrete or discretized.

As already noted we are not the first to focus on the distribution of individual treatment effects. Heckman and Smith (1993, 1998) and in particular Heckman, Smith and Clements (1997) derive bounds on the distribution of the treatment effect. They also consider additional restrictions that narrow these bounds and in some cases these assumptions are sufficient to point identify the treatment effect distribution. Aakvik, Heckman and Vytlacil (2005), Carneiro, Hansen and Heckman (2003), and Wu and Perloff (2006)) also introduce additional restrictions.

In the next section we introduce the notation and discuss the HSC bounds and subsequent restrictions that narrow these bounds. The remainder of the paper is organized as follows. In Section 3 we introduce the Makarov bounds on the distribution of the individual treatment effect. We show that averaging over covariates that are related to the outcomes improves these bounds. In Section 4 we show how bounds on the quantiles of the treatment effect distribution can be improved by using the mean of the treatment effect distribution, the Average Treatment Effect (ATE). In Section 5 we introduce a similar procedure to improve bounds on the cdf of the treatment effect distribution. We show that bounds on the quantiles obtained by inverting these bounds on the cdf are identical to those obtained in Section 4. We also show that the conditional (on covariates) bounds on the (conditional) cdf can be improved by using the conditional ATE. Bounds on the (unconditional) cdf of the treatment effect can be obtained by averaging these conditional bounds and bounds on the unconditional quantiles by inverting these unconditional cdf bounds. Throughout we illustrate the properties of the bounds for normal and dichotomous outcome distributions.

2 Distribution of Individual Treatment Effects

We consider the joint distribution (Y, D, X) where Y is the observed outcome; D is a binary treatment variable that equals one if the unit is treated and zero otherwise; X is a vector of covariates. We assume first that the support of (Y, D, X) is $\mathcal{Y} \times \{0, 1\} \times \mathcal{X}$ where $\mathcal{X}$ is a finite subset of $\mathbb{R}^k$.

The outcome Y can be written as $Y = \Delta \cdot D + Y_0$, where $\Delta \equiv Y_1 - Y_0$ is the individual treatment effect and Y_1 and Y_0 are respectively the outcome if the unit is treated and not treated. The variable of interest is Δ, the individual treatment effect. Every unit has either $D = 1$ or $D = 0$, so that we never observe Δ but we do observe $Y_1|D_i = 1$ or $Y_0|D_i = 0$.

Therefore, without further assumptions it is impossible to identify the distribution of Δ with a cdf that we denote by G.

The marginal distributions of Y_1 and Y_0 are identified if treatment is assigned randomly, so that$(Y_1, Y_0) \perp\!\!\!\perp D$, or is ignorable, so that $(Y_1, Y_0) \perp\!\!\!\perp D|X$ (Rubin , 1977).[3] We call the identified marginal distribution of Y_1, F_1 and the identified marginal distribution of Y_0, F_0.

Heckman, Smith and Clements (1997) derive bounds on the distribution of individual treatment effects by considering extreme values of parameters defined on this distribution, e.g. quantiles, if the joint distribution of $Y(0), Y(1)$ ranges over all possible joint distributions with given marginals. In the construction of the set of joint distributions with given marginals they use an idea of Whitt (1976). This idea can be directly applied to the empirical distributions of $Y(0)$ and $Y(1)$ if these have the same number of points of support. In that case the set of joint distributions is the set pairings obtained by permuting the values of $Y(1)$.

To see how this works consider the case where the number of treated and control units is the same and equals 2, so that the outcome distributions are dichotomous with points of support y_{01}, y_{02} for Y_0, both with probability 1/2, and y_{11}, y_{12} for Y_1 also with probability 1/2. In that case there are only two possible joint distributions with two corresponding distribution of individual treatment effects: Distribution 1 has points of support $y_{11} - y_{01}, y_{12} - y_{02}$ with probability 1/2 each and distribution 2 has support $y_{12} - y_{01}, y_{11} - y_{02}$ each point with probability also 1/2. In general the number of joint distributions and hence of distributions of Δ is $N!$. The parameter of interest is computed for each of these distributions and the extreme points give the bounds.

This method of finding the bounds is computationally intensive. If there are covariates, the computational burden increases, as we need to repeat the procedure for every value of the covariate vector. It is also not clear how to assess the sampling variation in the bounds. Note also that if the number of treated units is for instance smaller than the number of controls, then there is a need for grouping of controls. In that case the mean of the distributions obtained by permutation is not equal to the average treatment effect, and that "bias" persists in large samples. Finally, although HSC suggest that these rearrangement bounds are consistent for the continuous population bounds they do not give a formal proof.

HSC also consider restrictions that allow for narrower bounds or even point identification, and much of the literature has followed their lead. Two types of assumptions have been used: (i) assumptions on the joint distribution of the potential outcomes, and (ii) assumptions on treatment assignment (or combinations of these two types of assumptions. An example of the first type of assumptions is the assumption that the treatment effect is independent of the non-treated outcome, that is $\Delta \perp\!\!\!\perp Y_0$. HSC and Wu and Perloff (2006) show that under this assumption one can use deconvolution to point identify the treatment effect distribution. An example of an assumption of the second type is that treatment assignment maximizes the gain from participation so that $D = I(Y(1) > Y(0))$. This is the Roy model (Roy, 1951 and Heckman and Honoré, 1990). If in addition we assume separability between covariates X and unobservables U and independence of X and U, then (under some support conditions) G is point identified. Aakvik, Heckman and Vytlacil (2005) and Carneiro, Hansen and Heckman (2003) consider a sample selection model in which treatment assignment is not determined only by Δ. If the other assumptions on the independence of observables and unobservables are maintained, and in addition an assumption (one factor structure) on the joint distribution of the unobservables in the treatment assignment and the two potential outcomes is made, then

[3] $A \perp\!\!\!\perp B$ reads A is independent of B.

the joint distribution of the outcomes and hence of Δ is identified.

If one is not interested in the whole distribution G but in a parameter of that distribution, then the derivation of the bounds may be simpler. An example is the variance of Δ. Bounds on the variance are obtained from bounds on the correlation of $Y(0)$ and $Y(1)$. Fan (2005) derives sharp bounds on this correlation.

In this paper we consider the bounds on the cdf of the difference of two random variables with known marginal distributions, as first obtained by Makarov (1981). As we are concerned with bounds on the distribution of the individual treatment effect Δ, the known marginal distributions have cdf F_0 and F_1. Although the Makarov bounds are pointwise sharp, in the sense that for any value of the individual treatment effect there exist joint distributions (that depend on the value of the individual treatment effect) of the potential outcomes with the given marginals, such that the c.d.f. of the difference reaches the upper and lower bound at that particular value, they can be improved by using the fact that we know ATE. We are to our knowledge the first to suggest a procedure that uses this additional information to obtain more informative bounds. The mean is the only feature of the treatment effect distribution that is identified from the marginal outcome distributions, because higher order moments require knowledge of the joint distribution. We also show that if the treatment effect varies with observable covariates, averaging over these covariates improves the Makarov bounds. We can combine the two types of additional information by using the known conditional treatment effect means to improve the conditional Makarov bounds and by subsequently averaging these conditional bounds. The (improved) Makarov bounds on the cdf of the treatment effect distribution yield bounds on the quantiles of that distribution. Without covariates we can use the ATE to directly obtain bounds on the quantiles of the treatment effect distribution, that turn out to be identical to the bounds obtained by using the ATE to improve the cdf bounds. This equivalence means that we can use the average improved conditional Makarov bounds that use information on the conditional ATE to improve the bounds on the unconditional quantiles. Bounds on the conditional quantiles cannot be averaged to obtain bounds on the unconditional quantiles.

3 Makarov Bounds on the Distribution of Individual Treatment effects

Bounds on the sum and by extension on the difference of the distribution of two random variables were first derived by Makarov (1981) (see also Frank, Nelsen, and Schweizer, 1987). We extend these bounds to the case that we observe conditional marginal distributions of the random variables in the difference. If the conditional cdf of Y_0 and Y_1 given $X = x$ are given by $F_0(y|x)$ and $F_1(y|x)$ respectively with $x \in \mathbb{X}$, the support of the distribution of X, then we have the following bounds

THEOREM 1 (MAKAROV, 1981) *Let G denote the cdf of $Y_1 - Y_0$, then*

$$G_{ML}(y) \equiv \mathbb{E}\left[\sup_t \max\{F_1(t|X) - F_0(t-y|X)_-, 0\}\right] \leq G(y) \leq$$

$$\mathbb{E}\left[\inf_t \min\{F_1(t|X) - F_0(t-y|X)_- + 1, 1\}\right] \equiv G_{MU}(y) \quad (1)$$

with $F_0(.)_-$ the function of left-hand limits of the cdf.

Proof: First consider the lower bound. We have for all y_0, y_1 with $y_0 + y_1 = y$ and using the Bonferroni inequality

$$G(y|x) = \Pr(Y_1 + (-Y_0) \leq y|X = x) \geq \Pr(Y_1 \leq y_1, -Y_0 \leq y_0|X = x) \geq$$

$$\max\{\Pr(Y_1 \leq y_1|X = x) + \Pr(-Y_0 \leq y_0|X = x) - 1, 0\} = \max\{F_1(y_1|x) - F_0(-y_0|x)_-, 0\}$$

Hence of we define $t \equiv y_1, y \equiv y_1 + y_0$

$$G(y) \geq \mathbb{E}\left[\sup_t \max\{F_1(t|X) - F_0(t - y|X), 0\}\right]$$

For the upper bound we have

$$1 - G(y|x) = \Pr(Y_1 + (-Y_0) > y|X = x) \geq \Pr(Y_1 > y_1, -Y_0 > y_0|X = x) \geq$$

$$\max\{\Pr(Y_1 > y_1|X = x) + \Pr(-Y_0 > y_0|X = x) - 1, 0\}$$

Taking the opposite on both sides of the equation, adding 1, substituting $t \equiv y_1, y \equiv y_1 + y_0$, and taking the expectation gives

$$G(y) \leq \mathbb{E}\left[\inf_t \min\{F_1(t|X) - F_0(t - y|X)_- + 1, 1\}\right]$$

We show that the bounds are themselves cdf-s. Consider the lower bound for $G(y|x)$

$$L(y|x) = \sup_t \max\{F_1(t|x) - F_0(t - y|x), 0\}$$

Now if $y' \geq y$, then for all t

$$\max\{F_1(t|x) - F_0(t - y'|x)_-, 0\} \geq \max\{F_1(t|x) - F_0(t - y|x)_-, 0\}$$

so that $L(y'|x) \geq L(y|x)$. Next we show that $L(y|x)$ is right continuous. Consider a sequence $y_n \downarrow y$. First the sequence $L(y_n|x)$ is decreasing and bounded from below, so that it has a limit. Obviously $y_n - y < \varepsilon$ iff $(t - y) - (t - y_n) < \varepsilon$ independent of t. Hence for all $\delta > 0$ and n large enough

$$F_0(t - y_n|x) \geq F_0((t - y)|x)_- - \delta$$

Note that $t - y_n \uparrow t - y$ and that $F_0(.)_-$ is left-continuous. Using this inequality we have for all t

$$F_1(t|x) - F_0(t - y|x)_- \leq F_1(t|x) - F_0(t - y_n|x)_- \leq F_1(t|x) - F_0((t - y)|x)_- + \delta$$

Taking the sup over t from right to left we obtain

$$L(y|x) \leq L(y_n|x) \leq L(y|x) + \delta$$

Taking the limit we obtain, because δ is arbitrary, that $\lim_{n\to\infty} L(y_n|x) = L(y|x)$, so that the lower bound is right-continuous. Note that

$$L(y|x) \geq F_1(y/2|x) - F_0(-y/2|x)$$

so that $\lim_{y\to\infty} L(y|x) = 1$. Taking the expectation over X we conclude that the lower bound is indeed a cdf (by dominated convergence limits and expectations can be interchanged). The proof that the upper bound is also a cdf is analogous. □

Frank, Nelsen, and Schweizer (1987) show that these bounds are pointwise sharp in the following sense: for all y there is a joint distribution of Y_0, Y_1 with marginal cdf-s equal to F_0, F_1 such that the cdf of $Y_1 - Y_0$ at y is equal to $G_{ML}(y)$. There is also (a different) joint distribution with the same marginals such that the cdf of $Y_1 - Y_0$ at y is equal to $G_{MU}(y)$. This result is derived for the unconditional distributions, but it applies directly if we have distributions conditional on $X = x$.

For some distributions the bounds have a closed form.

Example 1: Difference of normals with the same variance.
Consider

$$Y_k \sim N(\mu_k, \sigma^2) \qquad k = 0, 1$$

Define the ATE by $\theta = \mu_1 - \mu_0$. The lower bound on the cdf of the treatment effect is

$$\begin{aligned} G_{ML}(y) &= 0 && \text{if } y < \theta \\ &= 2\Phi\left(\frac{y-\theta}{2\sigma}\right) - 1 && \text{if } y \geq \theta \end{aligned}$$

The corresponding density is

$$\begin{aligned} g_{ML}(y) &= 0 && \text{if } y < \theta \\ &= \frac{1}{\sigma}\phi\left(\frac{y-\theta}{2\sigma}\right) && \text{if } y \geq \theta \end{aligned}$$

Note that this is the density of a halfnormal distribution with begin point θ. Hence the mean of the lower bound distribution is

$$\theta + \sigma\frac{2\sqrt{2}}{\sqrt{\pi}} > \theta$$

Hence mean of the lower bound distribution is strictly larger that the mean of the distribution of $Y_1 - Y_0$. The upper bound is

$$\begin{aligned} G_{MU}(y) &= 2\Phi\left(\frac{y-\theta}{2\sigma}\right) && \text{if } y < \theta \\ &= 1 && \text{if } y \geq \theta \end{aligned}$$

The corresponding density is

$$\begin{aligned} g_{MU}(y) &= \frac{1}{\sigma}\phi\left(\frac{y-\theta}{2\sigma}\right) && \text{if } y < \theta \\ &= 0 && \text{if } y \geq \theta \end{aligned}$$

which is the density of a halfnormal distribution distribution with end point θ, so that the mean of the lower bound distribution is equal to

$$\theta - \sigma\frac{2\sqrt{2}}{\sqrt{\pi}} < \theta.$$

By inverting the bounds we obtain bounds on the quantiles of the treatment effect distribution. If we denote the upper and lower bounds for the p quantile by $q_{ML}(p)$ and $q_{MU}(p)$, respectively[4],

[4]The subscripts L, U indicate the fact that the upper bound on the p quantile is obtained by inverting the lower Makarov bound and the lower bound by inverting the upper Makarov bound.

then

$$q_{MU}(p) = \theta + 2\sigma\Phi^{-1}\left(\frac{p}{2}\right)$$

and

$$q_{ML}(p) = \theta + 2\sigma\Phi^{-1}\left(\frac{p+1}{2}\right)$$

and these are best possible bounds on the p quantile of the distribution of the treatment effect, if we do not use the known ATE.

The Makarov bounds on the cdf are not informative if $y = \theta$, because at that point the lower bound on the cdf is 0 and the upper bound is 1. The bounds on the cdf are more informative if y is much larger or smaller than the ATE. Interestingly the length of the bound on the p quantile is proportional to

$$\Phi^{-1}\left(\frac{p+1}{2}\right) - \Phi^{-1}\left(\frac{p}{2}\right)$$

and this is minimal if $p = \frac{1}{2}$, i.e. the bounds for the median are most informative. Note that for the median the bounds are symmetric around θ, but that they are asymmetric for other quantiles. □

If there are covariates that are correlated with the outcomes, then Theorem 1 suggests to compute the average (with respect to the covariate distribution) bounds. Theorem 2 states that the average bounds are narrower than the bounds derived from the average, i.e. unconditional, outcome distributions.

THEOREM 2

$$\sup_t \max\{\mathbb{E}[F_1(t|X)] - \mathbb{E}[F_0(t-y|X)_-], 0\} \leq \mathbb{E}\left[\sup_t \max\{F_1(t|X) - F_0(t-y|X)_-, 0\}\right] \quad (2)$$

$$\mathbb{E}\left[\inf_t \min\{F_1(t|X) - F_0(t-y|X)_- + 1, 1\}\right] \leq \inf_t \min\{\mathbb{E}[F_1(t|X)] - \mathbb{E}[F_0(t-y|X)_-] + 1, 1\} \quad (3)$$

Proof: For all $x \in \mathbb{X}$ and all $s \in \Re$

$$\sup_t \max\{F_1(t|X=x) - F_0(t-y|X=x)_-, 0\} \geq \max\{F_1(s|X=x) - F_0(s-y|X=x)_-, 0\}$$

Hence for all $x \in \mathbb{X}$ and all $s \in \Re$

$$\sup_t \max\{F_1(t|X=x) - F_0(t-y|X=x)_-, 0\} \geq F_1(s|X=x) - F_0(s-y|X=x)_-$$

and

$$\sup_t \max\{F_1(t|X=x) - F_0(t-y|X=x)_-, 0\} \geq 0$$

Averaging over the distribution of X gives that for all $s \in \Re$

$$\mathbb{E}\left[\sup_t \max\{F_1(t|X) - F_0(t-y|X)_-, 0\}\right] \geq \mathbb{E}[F_1(s|X)] - \mathbb{E}[F_0(s-y|X)_-]$$

and

$$\mathbb{E}\left[\sup_t \max\{F_1(t|X) - F_0(t-y|X)_-, 0\}\right] \geq 0$$

Hence for all $s \in \Re$

$$\mathbb{E}\left[\sup_t \max\{F_1(t|X) - F_0(t-y|X)_-, 0\}\right] \geq \max\{\mathbb{E}[F_1(s|X)] - \mathbb{E}[F_0(s-y|X)_-], 0\}$$

so that we obtain (2). The proof of inequality (3) is analogous. □

The theorem shows that the average Makarov bounds are more informative than the Makarov bounds on the average distribution. This means that even in a randomized experiment covariate information can be useful in narrowing the bounds on the cdf and the quantiles of the treatment effect distribution. The next example illustrates the role of averaging for normal outcome distributions.

Example 2: Mixture of normals with the same variance.
We assume that the outcome distributions are a two-component mixture of normal distributions

$$Y_{kl} \sim N(\mu_{kl}, \sigma^2) \qquad k = 0, 1,\ l = 1, 2$$

Define $\theta_l = \mu_{1l} - \mu_{0l}, l = 1, 2$ and without loss of generality we assume $\theta_1 < \theta_2$. The fraction in the population with ATE θ_l is $r_l, l = 1, 2$ with $r_1 + r_2 = 1$. The average lower Makarov bound as in (1) is

$$\begin{aligned} G_{ML}(y) &= 0 && \text{if } y < \theta_1 \\ &= r_1\left(2\Phi\left(\frac{y-\theta_1}{2\sigma}\right) - 1\right) && \text{if } \theta_1 \leq y < \theta_2 \\ &= 2\left(r_1\Phi\left(\frac{y-\theta_1}{2\sigma}\right) + r_2\Phi\left(\frac{y-\theta_2}{2\sigma}\right)\right) - 1 && \text{if } y \geq \theta_2 \end{aligned}$$

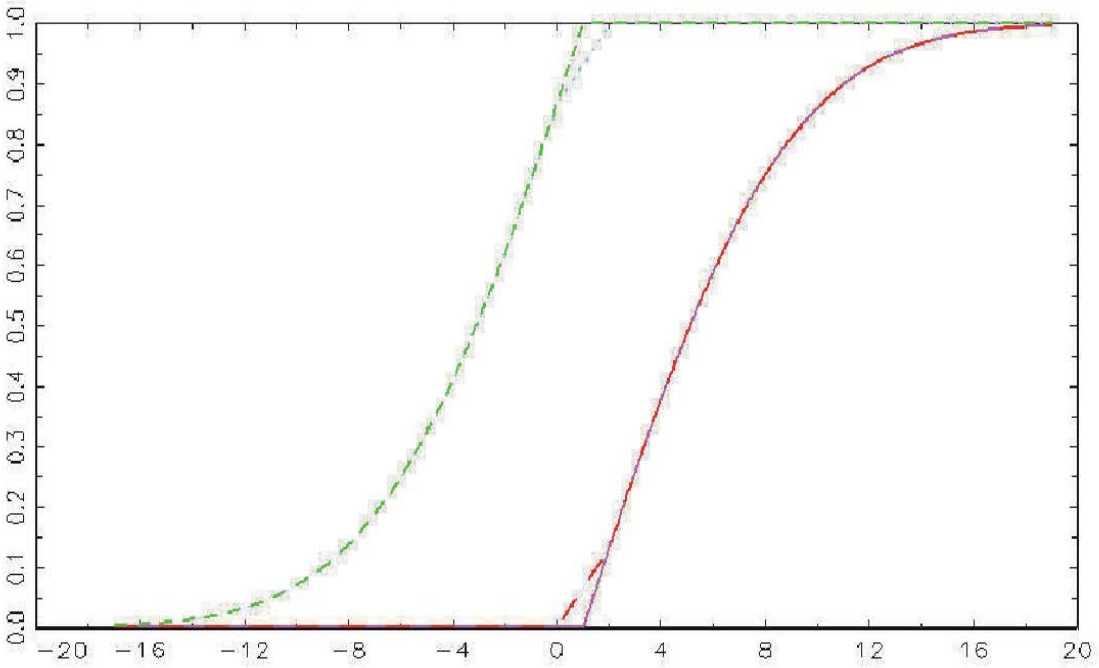
and the average upper Makarov bound is

$$\begin{aligned} G_{MU}(y) &= 2\left(r_1\Phi\left(\frac{y-\theta_1}{2\sigma}\right) + r_2\Phi\left(\frac{y-\theta_2}{2\sigma}\right)\right) && \text{if } y < \theta_1 \\ &= r_1 + 2r_2\Phi\left(\frac{y-\theta_2}{2\sigma}\right) && \text{if } \theta_1 \leq y < \theta_2 \\ &= 1 && \text{if } y \geq \theta_2 \end{aligned}$$

The average outcome distribution is a mixture of normals and the bounds do not have a closed form solution in that case. We consider an example with $\sigma = 3$,$\mu_{01} = \mu_{11} = 1$ so that $\theta_1 = 0$ and $\mu_{02} = 0, \mu_{12} = 2$ so that $\theta_1 = 2$. Also $r_1 = r_2 = .5$ so that the ATE is 1. Note that averaging improves the bounds in the interval $[0, 2] = [\theta_1, \theta_2]$, but not outside this interval. In this example the variation in μ explains a fraction $.25/9 = .028$ of the variation in the outcomes. □

Example 3: Dummy outcome distributions.
The Makarov bounds also apply of the outcome distribution is discrete. Here we consider the special case that both Y_0 and Y_1 are 0-1 variables with $\Pr(Y_0 = 1) = p_0$ and $\Pr(Y_1 = 1) = p_1$. In this case it is possible to derive the bounds on the distribution of $Y_1 - Y_0$ directly, because this distribution has only three points of support -1,0,1. We verify that these bounds are equal

Fig 1. *Average Makarov bounds and Makarov bounds for the average population: Normal two-point mixture with* $\theta_1 = 0, \theta_2 = 2$.



to the Makarov bounds. Also in this case the bounds are uniformly sharp so that knowledge of the ATE $p_1 - p_0$ does not change the bounds.

We first obtain bounds on $\Pr(Y_1 - Y_0 = -1) = \Pr(Y_0 = -1, Y_1 = 1)$. Because of this equality we have the bounds (use the Bonferroni inequality for the lower bound)

$$\max\{p_0 - p_1, 0\} \leq \Pr(Y = -1) \leq \min\{1 - p_1, p_0\}$$

The bounds are summarized in Figure 2. Above the 45 degree line the ATE is positive and below it the ATE is negative. Note that for instance, if the ATE is positive, then less than half of the population is harmed (negative treatment effect) by the treatment. Actually, for the bounds for $\Pr(Y = 1)$ imply that if the treatment effect is negative, then less than half of the population benefits (positive treatment effect) from the treatment.

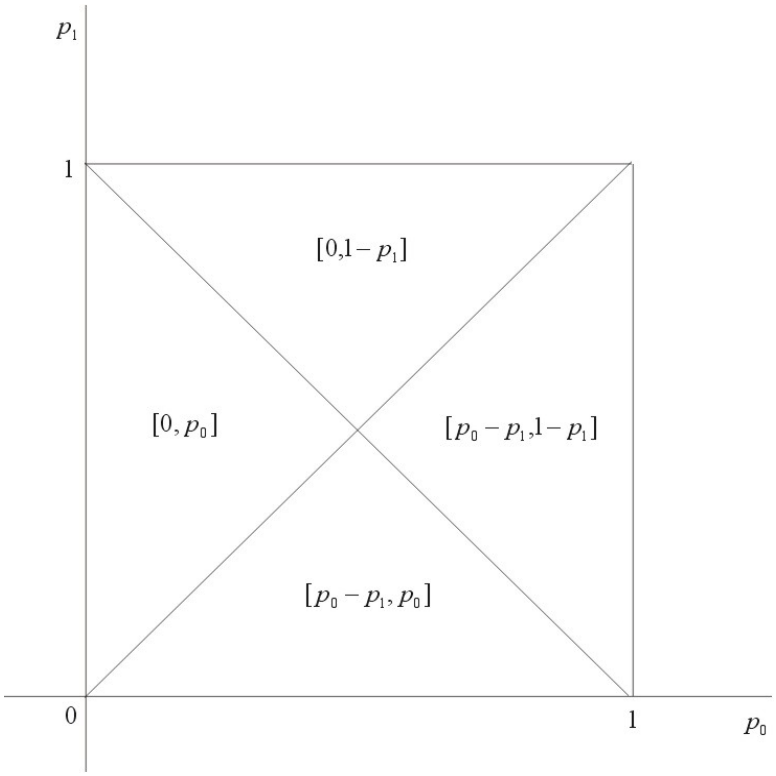
To show that these bounds are indeed Makarov bounds we define

$$U(t) = \min\{F_1(t) - F_0(t+1)_- + 1, 1\}$$

and

$$L(t) = \max\{F_1(t) - F_0(t+1)_-, 0\}$$

Fig 2. *Bounds on* $\Pr(Y = -1)$ *for 0-1 outcome variables*



so that

$$\begin{aligned} U(t) &= 1 && t \le -1 \\ &= p_0 && -1 < t < 0 \\ &= \min\{p_0 + 1 - p_1, 1\} \ge p_0 && t = 0 \\ &= 1 - p_1 && 0 < t < 1 \\ &= 1 && t \ge 1 \end{aligned}$$

and

$$\begin{aligned} L(t) &= 0 && t < 0 \\ &= \max\{p_0 - p_1, 0\} && t = 0 \\ &= 0 && t > 0 \end{aligned}$$

By taking the inf and sup over t we get the upper and lower Makarov bounds that are equal to the bounds derived before.

For $\Pr(Y = 1) = \Pr(Y_0 = 0, Y_1 = 1)$ we obtain

$$\max\{p_1 - p_0, 0\} \le \Pr(Y = 1) \le \min\{p_1, 1 - p_0\}$$

Because $\Pr(Y = 1) = 1 - \Pr(Y \le 0)$ we can show that these bounds are Makarov bounds by substitution of the Makarov bounds on $\Pr(Y \le 0)$. Finally, because $\Pr(Y = 0) = 1 - \Pr(Y =$

$-1) - \Pr(Y = 1)$ we have the following bounds

$$1 - \min\{1 - p_1, p_0\} - \min\{p_1, 1 - p_0\} \leq \Pr(Y = 0) \leq 1 - \max\{p_0 - p_1, 0\} - \max\{p_1 - p_0, 0\}$$

which are again equal to the Makarov bounds.

Let Y^* be a random variable with values -1,0,1 and

$$\begin{aligned} \Pr(Y^* = -1) &= \max\{p_0 - p_1, 0\} \\ \Pr(Y = 0) &= 1 - \max\{p_0 - p_1, 0\} - \max\{p_1 - p_0, 0\} \\ \Pr(Y^* = 1 &= \max\{p_1 - p_0, 0\} \end{aligned}$$

This distribution is inside the Makarov bounds and equal to the lower bound on the probability $\Pr(Y = -1)$ and $\Pr(Y = 1)$ and the upper bound on the probability of $\Pr(Y = 0)$. Moreover,

$$\mathrm{E}(Y^*) = p_1 - p_0$$

so that the mean is equal to the ATE. Hence, there is a joint distribution of Y_0, Y_1 with the given marginals that coincides with the lower, upper, and lower bounds of the probabilities. This distribution is the same for all values of Y. An analogous construction gives a distribution with mean equal to the ATE that coincides with the upper, lower and upper bounds of the probabilities $\Pr(Y = y)$. This implies that the Makarov bounds in this case are best possible.

The construction of the bounding distributions is similar to that used in the next section, be it that here we do not obtain an improvement. The case of dummy outcomes is special and the results do not extend to the general discrete case.□

4 Improved Bounds on the Quantiles of the Distribution of Treatment Effects

We argued that the Makarov bounds are pointwise sharp, because for any value of the individual treatment effect there is a joint distribution of Y_0, Y_1 such that the cdf of $Y_1 - Y_0$ is equal to the lower Makarov bound, and the same holds for the upper bound. However, in Example 1 the mean of the lower bound distribution is larger than $\mu_1 - \mu_0$ while the mean of the upper bound distribution is larger than $\mu_1 - \mu_0$. In general the Makarov bounds do do incorporate the mean of the treatment effect distribution which is known from the marginal means of Y_0 and Y_1. Actually the mean is the only feature of the treatment effect distribution that can be recovered from the marginal outcome distributions. Formally, for any non-constant function h the moment $\mathbb{E}[h(Y_1 - Y_0)]$ is a functional of the marginal distributions of Y_1 and Y_0 if and only if h is the identity function (or a linear transformation of the identity function with known coefficients). As far as we know the fact that the Makarov bounds do not incorporate the information on the mean treatment effect has not been noticed in the literature on bounds on the distribution of sums of random variables. If we observe covariates that are correlated with the outcomes, then the conditional Makarov bounds do not account for the information in the conditional treatment effect $\mathbb{E}[Y_1 - Y_0|X = x]$.

In this section we show how we can use the ATE to improve the Makarov bounds on the quantiles of the treatment effect distribution. In Section 4 we discuss an analogous procedure to improve the Makarov bounds on the cdf. The latter bounds can be inverted to obtain improved bounds on the quantiles. It turns out that the resulting bounds are the same as the improved

bounds on the quantiles derived in this section. We discuss the improvement procedure for the quantiles separately for expository purposes. A drawback of improving the quantiles is that with covariates we obtain bounds on the conditional quantiles, and there is no obvious way to obtain bounds on the unconditional quantiles from these bounds. However, the improved bounds on the conditional cdf can be averaged to obtain bounds on the unconditional cdf that can be inverted to obtain bounds on the unconditional quantiles. For this reason omit the covariates in the derivation of the improved bounds on the quantiles.

The derivation of the improved bounds is based on the fact that if the distribution of Z_1 first order stochastically dominates that of Z_2 then for any increasing function ϕ, $E(\phi(Z_1)) \geq E(\phi(Z_2))$ provided that the expectations exist (e.g. Shaked and Shanthikumar, 1994, p. 4). In the sequel we assume that all (conditional) expectations exist. We fix the index $0 < p < 1$ of the quantile of the distribution of $Y_1 - Y_0$ that will be bounded. We denote this quantile by $q_{Y_1-Y_0}(p)$. Initially we have the bounds

$$q_{MU}(p) \leq q_{Y_1-Y_0}(p) \leq q_{LU}(p)$$

with $q_{Mk}(p) = \inf\{y|G_{Mk}(y) \geq p\}$ for $k = L, U$. The labelling of the bounds seems counterintuitive, but the subscripts indicate that the lower bound is the p quantile of the upper Makarov bound on the cdf (see Figure 2). These bounds are pointwise, i.e. for a given p, sharp in the same sense as the Makarov bounds, i.e. there is a joint distribution of Y_0, Y_1 with marginal cdf-s equal to F_0, F_1 such that the cdf of $Y_1 - Y_0$ has p quantile $q_{MU}(p)$. This joint distribution is obtained by the construction of Frank, Nelsen, and Schweizer(1987) with $y = q_{MU}(p)$. The upper bound is also sharp in this sense.

In general the upper and lower bound distributions will not have a mean equal to $\mathrm{E}(Y_1 - Y_0)$. The question is whether this additional information gives narrower bounds on the p quantile $q_{Y_1-Y_0}(p)$. To investigate this we first construct distributions that have p quantile equal to $q_{MU}(p)$ and $q_{ML}(p)$ and are have the largest and smallest cdf, in the sense of first order stochastic domination, with these p quantiles, and that also are within the Makarov bounds. The largest cdf G_{pL} is

$$\begin{aligned} G_{pL}(y) &= G_{ML}(y) \quad && y < q_{MU}(p) \\ &= p && q_{MU}(p) \leq y < q_{ML}(p) \\ &= G_{ML}(y) && y \geq q_{ML}(p) \end{aligned}$$

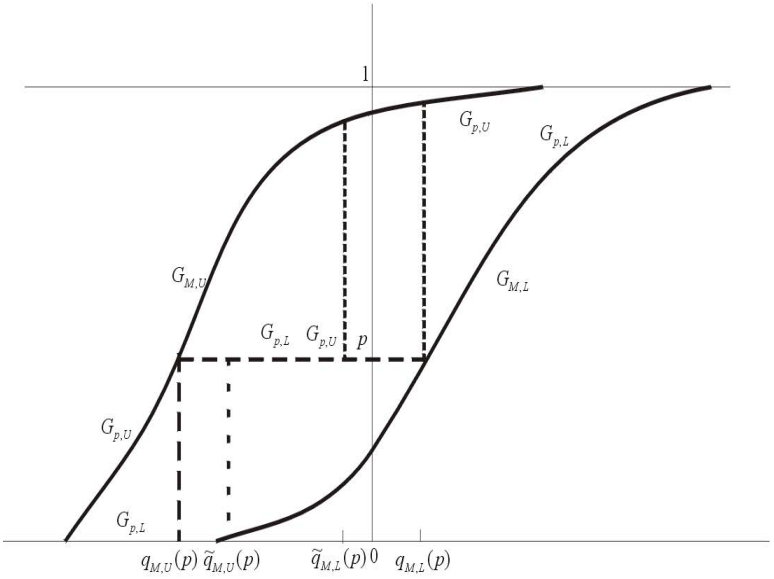
The smallest he cdf G_{pU} is

$$\begin{aligned} G_{pU}(y) &= G_{MU}(y) \quad && y < q_{MU}(p) \\ &= p && q_{MU}(p) \leq y < q_{ML}(p) \\ &= G_{MU}(y) && y \geq q_{ML}(p) \end{aligned}$$

The construction is illustrated in Figure 3. The labelling of these cdf-s indicates the Makarov bound that they must follow to achieve the objective of having a mean greater than the ATE for G_{pL}, i.e. G_{ML}, and less than the ATE for G_{pU}, i.e. G_{MU}. The next step is to compute the mean of the distributions with cdf G_{pL} and G_{pU}. The corresponding distributions are mixed discrete-continuous. For the continuous part we use the fact that for any continuous random variable Z with support $\underline{z}, \overline{z}$ and cdf F_Z

$$\mathrm{E}(Z) = \int_0^{\overline{z}} (1 - F_Z(z))\mathrm{d}z - \int_{\underline{z}}^0 F_Z(z)\mathrm{d}z$$

Fig 3. *Largest cdf G_{pL} and smallest cdf G_{pU} consistent with Makarov bounds on the p quantile*



First, we consider G_{pL}. The corresponding distribution has an atom at $q_{MU}(p)$ and the probability of this point is $p - G_{ML}(q_{MU}(p))$. To compute the mean we take the average of the mean of the distribution truncated from above at $q_{MU}(p)$, the mean of the distribution truncated from below at $q_{ML}(p)$, and $q_{MU}(p)$ with weights $G_{ML}(q_{MU}(p))$, $1-p$, and $p - G_{ML}(q_{MU}(p))$, respectively. Note that the interval $[q_{MU}(p), q_{LU}(p)]$ has probability 0 in the distribution with cdf G_{pL}. The truncated means are

$$\frac{1}{G_{ML}(q_{MU}(p))}\left(\int_0^{\max\{q_{MU}(p),0\}}(1-G_{ML}(y))\mathrm{d}y - \int_{-\infty}^{\min\{q_{MU}(p),0\}} G_{ML}(y)\mathrm{d}y\right)$$

and

$$\frac{1}{1-p}\left(\int_{\max\{q_{ML}(p),0\}}^{\infty}(1-G_{ML}(y))\mathrm{d}y - \int_{\min\{q_{ML}(p),0\}}^{0} G_{ML}(y)\mathrm{d}y\right)$$

Hence

$$\mu_{pL} = (p-G_{ML}(q_{MU}(p)))q_{MU}(p) + \int_0^{\max\{q_{MU}(p),0\}}(1-G_{ML}(y))\mathrm{d}y - \int_{-\infty}^{\min\{q_{MU}(p),0\}} G_{ML}(y)\mathrm{d}y +$$

$$\int_{\max\{q_{ML}(p),0\}}^{\infty}(1-G_{ML}(y))\mathrm{d}y - \int_{\min\{q_{ML}(p),0\}}^{0} G_{ML}(y)\mathrm{d}y$$

An analogous argument gives

$$\mu_{pU} = (G_{MU}(q_{ML}(p))-p)q_{ML}(p) + \int_0^{\max\{q_{MU}(p),0\}}(1-G_{MU}(y))\mathrm{d}y - \int_{-\infty}^{\min\{q_{MU}(p),0\}} G_{MU}(y)\mathrm{d}y +$$

$$\int_{\max\{q_{ML}(p),0\}}^{\infty}(1-G_{MU}(y))\mathrm{d}y - \int_{\min\{q_{ML}(p),0\}}^{0} G_{MU}(y)\mathrm{d}y$$

Now we distinguish between two cases: (i) $\mu_{pL} \geq \mathbb{E}(Y_1 - Y_0)$, and (ii) $\mu_{pL} < \mathbb{E}(Y_1 - Y_0)$. In case (i) there exists a cdf within the Makarov bounds that goes through the lower bound on the p quantile and has a mean equal to $\mathbb{E}(Y_1 - Y_0)$. In case (ii) such a cdf does not exist. The only way to obtain a cdf of a distribution with a mean equal to the ATE is to move the lower bound on the p quantile to the right, i.e. to move the atom in the G_{pL} distribution to the right. Because the mean of the G_{ML} distribution is larger than the ATE, there is a lower bound $\tilde{q}_{MU}(p)$ such that the corresponding distribution has mean equal to the ATE. An analogous argument gives the expression for the improved upper bound and the condition for improvement.

The adjusted bounds on $q_{Y_1-Y_0}(p)$ are

$$\tilde{q}_{MU}(p) \leq q_{Y_1-Y_0}(p) \leq \tilde{q}_{ML}(p)$$

with $\tilde{q}_{MU}(p) \geq q_{MU}(p)$ the solution to

$$(p - G_{ML}(\tilde{q}_{MU}(p)))\tilde{q}_{MU}(p) - (p - G_{ML}(q_{MU}(p)))q_{MU}(p) + \int_{\max\{q_{MU}(p),0\}}^{\max\{\tilde{q}_{MU}(p),0\}} (1 - G_{ML}(y))\mathrm{d}y$$

$$- \int_{\min\{q_{MU}(p),0\}}^{\min\{\tilde{q}_{MU}(p),0\}} G_{ML}(y)\mathrm{d}y = \max\{\mathrm{E}(Y_1 - Y_0) - \mu_{pL}, 0\}$$

and $\tilde{q}_{ML}(p) \leq q_{ML}(p)$ the solution to

$$(G_{MU}(q_{ML}(p)) - p)q_{ML}(p) - (G_{MU}(\tilde{q}_{ML}(p)) - p)\tilde{q}_{ML}(p) + \int_{\max\{q_{MU}(p),0\}}^{\max\{q_{MU}(p),0\}} (1 - G_{MU}(y))\mathrm{d}y -$$

$$\int_{\min\{\tilde{q}_{MU}(p),0\}}^{\min\{q_{MU}(p),0\}} G_{MU}(y)\mathrm{d}y = \max\{\mu_{pU} - \mathrm{E}(Y_1 - Y_0), 0\}$$

Note that the information on the mean only gives improved bounds if the right-hand side of the equations is positive.

Because the lower bound $\tilde{q}_{MU}(p)$ is the p quantile of a cdf with mean $\mathrm{E}(Y_1 - Y_0)$ that is within the Makarov bounds and that by construction is the largest cdf within these bounds with p quantile $\tilde{q}_{MU}(p)$ and mean $\mathrm{E}(Y_1 - Y_0)$, the adjusted lower bound is best possible for the p quantile. The same argument establishes that the adjusted upper bound is best possible.

Example 1, continued: Difference of normals with the same variance.

As discussed in Section 3 the Makarov bounds on the p quantile are

$$q_{M,U}(p) = \theta + 2\sigma\Phi^{-1}\left(\frac{p}{2}\right)$$

and

$$q_{M,L}(p) = \theta + 2\sigma\Phi^{-1}\left(\frac{p+1}{2}\right)$$

and these are best possible, if we do not use the known ATE. Can these bounds be improved if we know the ATE? We assume without loss of generality that $\theta \geq 0$. This implies that $q_{ML}(p) \geq \theta \geq 0$. Also we consider the case that $q_{MU}(p) \leq 0$ which is equivalent to

$$\frac{\theta}{\sigma} \leq -2\Phi^{-1}\left(\frac{p}{2}\right)$$

p	$H_L(p)$	$H_U(p)$
.10	1.14	1.43
.20	0.78	1.08
.30	0.46	0.76
.40	0.15	0.46
.50	-.16	.16
,60	-.46	-.15
.70	-.76	-.46
.80	-1.08	-.78
.90	-1.43	-1.14

This restrictions yields simple explicit expressions for the improved bounds. In most cases the ATE is small relative to the standard deviation of the outcome distribution, so that this seems to be the relevant case. If the ratio of the ATE and the standard deviation of the outcome distribution is less than .2, this case covers up to the 92-th percentile and if this ratio is less than .1 up it covers up to the 96-th percentile.

If the restriction is met we have

$$\mu_{pL} = p\theta + \sigma\left(2p\Phi^{-1}\left(\frac{p}{2}\right) - 2(1-p)\Phi^{-1}\left(\frac{p+1}{2}\right) + 4\phi\left(\Phi^{-1}\left(\frac{p+1}{2}\right)\right)\right)$$

and

$$\mu_{pU} = (1-p)\theta + \sigma\left(2(1-p)\Phi^{-1}\left(\frac{p+1}{2}\right) + 2p\Phi^{-1}\left(\frac{p}{2}\right) - 4\phi\left(\Phi^{-1}\left(\frac{p}{2}\right)\right)\right)$$

We define

$$H_L(p) = \frac{1}{p}\left(2(1-p)\Phi^{-1}\left(\frac{p+1}{2}\right) + 2p\Phi^{-1}\left(\frac{p}{2}\right) - 4\phi\left(\Phi^{-1}\left(\frac{p}{2}\right)\right)\right)$$

and

$$H_U(p) = \frac{1}{1-p}\left(2(1-p)\Phi^{-1}\left(\frac{p+1}{2}\right) + 2p\Phi^{-1}\left(\frac{p}{2}\right) - 4\phi\left(\Phi^{-1}\left(\frac{p}{2}\right)\right)\right)$$

It can be shown that $H_L(p) \leq H_U(p)$ and that the lower bound $q_{MU}(p)$ on the p quantile can be improved iff

$$\frac{\theta}{\sigma} \geq H_L(p)$$

The upper bound $q_{ML}(p)$ can be improved if

$$\frac{\theta}{\sigma} \leq H_U(p)$$

The table gives $H_L(p)$ and $H_U(p)$ for selected values of p.

Now assume that $\frac{\theta}{\sigma} = .10$. In that case both the upper and lower bound on the median of the treatment effect distribution can be improved. For lower quantiles only the upper bound is improved by knowledge of the ATE while for lower quantiles only the lower bound is improved. As an illustration consider $\theta = 1$ and $\sigma = 10$. Then the Makarov bounds on the median are $[-12.48, 14.48]$. The improved lower bound is $-12.48 + 2.56 = -9.92$. The improved upper bound is $14.48 - .56 = 13.92$. □

5 Improved Bounds on the Distribution Function of the Treatment Effect Distribution

We can also use knowledge of the ATE to improve the bounds on the cdf of the treatment effect distribution. For instance,in the case that the outcome is valued positively, it is important to know which fraction of the population is harmed by the intervention, i.e. $G(0)$. As noted above we can average the improved bounds. Hence we improve the conditional Makarov bounds.

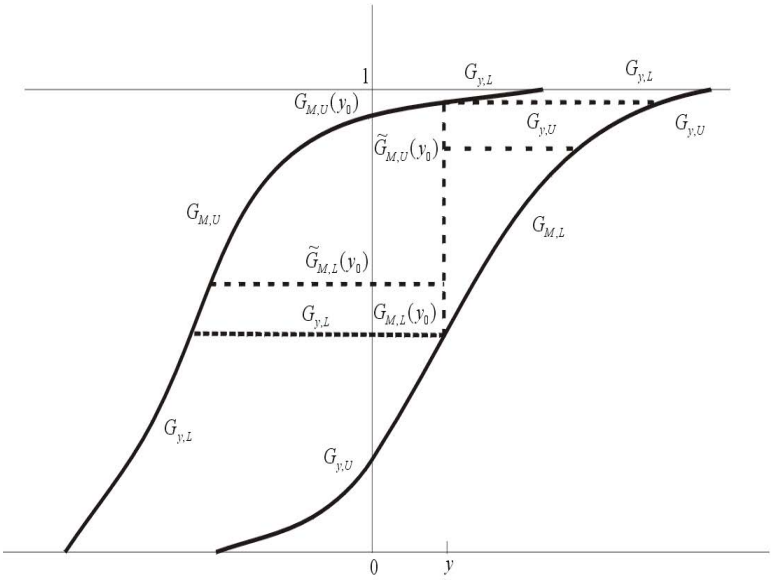
The conditional Makarov bounds on the cdf of the treatment effect distribution in y_0 given $X = x$ are

$$G_{ML}(y_0|x) \equiv \sup_t \max\{F_1(t|x) - F_0(t - y_0|x)_-, 0\} \leq G(y_0|x) \leq$$

$$\inf_t \min\{F_1(t|x) - F_0(t - y_0|x)_- + 1, 1\} \equiv G_{MU}(y_0|x)$$

As in the case of bounds on a quantile we note that the bounds do not take into account the known (conditional) ATE. For the upper bound $G_{ML}(y_0|x)$ we construct the largest, in the sense of first order stochastic domination, cdf G_{y_0U} inside the Makarov bounds that passes through this upper bound. For the lower bound we find the smallest cdf G_{y_0L} inside the Makarov bounds that passes through the upper bound on the conditional cdf. We have

Fig 4. *Smallest cdf G_{y_0L} and largest cdf G_{y_0U} consistent with Makarov bounds on the cdf at y_0*



$$\begin{aligned} G_{y_0U}(y|x) &= G_{ML}(y|x) \qquad y < y_0 \\ &= G_{MU}(y_0|x) \qquad y_0 \leq y < G_{ML}^{-1}(G_{MU}(y_0|x)|x)) \\ &= G_{ML}(y|x) \qquad y \geq G_{ML}^{-1}(G_{MU}(y_0|x)|x)) \end{aligned}$$

and

$$\begin{aligned} G_{y_0L}(y|x) &= G_{MU}(y|x) \qquad y < G_{MU}^{-1}(G_{ML}(y_0|x)|x)) \\ &= G_{ML}(y_0|x) \qquad G_{MU}^{-1}(G_{ML}(y_0|x)|x)) \leq y < y_0 \\ &= G_{MU}(y|x) \qquad y \geq y_0 \end{aligned}$$

The distribution $G_{y_0U}(y|x)$ has an atom at y_0 and gives 0 probability to the interval $(y_0, G_{ML}^{-1}(G_{MU}(y_0|x)|x))]$. The truncated means of the continuous parts of the distribution are

$$\frac{1}{1-G_{MU}(y_0|x)}\left(\int_{\max\{G_{ML}^{-1}(G_{MU}(y_0|x)|x)),0\}}^{\infty}(1-G_{ML}(y|x))\mathrm{d}y - \int_{\min\{G_{ML}^{-1}(G_{ML}(y_0|x)|x)),0\}}^{0} G_{ML}(y|x)\mathrm{d}y\right)$$

and

$$\frac{1}{G_{ML}(y_0|x)}\left(\int_{0}^{\max\{y_0,0\}}(1-G_{ML}(y|x))\mathrm{d}y - \int_{-\infty}^{\min\{y_0,0\}} G_{ML}(y|x)\mathrm{d}y\right)$$

Hence we have

$$\mu_{y_0U} = (G_{MU}(y_0|x) - G_{ML}(y_0|x))y_0 + \int_{\max\{G_{ML}^{-1}(G_{MU}(y_0|x)|x)),0\}}^{\infty}(1-G_{ML}(y|x))\mathrm{d}y$$

$$-\int_{\min\{G_{ML}^{-1}(G_{ML}(y_0|x)|x)),0\}}^{0} G_{ML}(y|x)\mathrm{d}y + \int_{0}^{\max\{y_0,0\}}(1-G_{ML}(y|x))\mathrm{d}y - \int_{-\infty}^{\min\{y_0,0\}} G_{ML}(y|x)\mathrm{d}y$$

The distribution $G_{y_0L}(y|x)$ has an atom at y_0 and gives 0 probability to the interval $(G_{MU}^{-1}(G_{ML}(y_0|x)|x)), y_0]$. The truncated means of the continuous parts of the distribution are

$$\frac{1}{1-G_{MU}(y_0|x)}\left(\int_{\max\{y_0,0\}}^{\infty}(1-G_{MU}(y|x))\mathrm{d}y - \int_{\min\{y_0,0\}}^{0} G_{MU}(y|x)\mathrm{d}y\right)$$

and

$$\frac{1}{G_{ML}(y_0|x)}\left(\int_{0}^{\max\{G_{MU}^{-1}(G_{ML}(y_0|x)|x)),0\}}(1-G_{MU}(y|x))\mathrm{d}y - \int_{-\infty}^{\min\{G_{MU}^{-1}(G_{ML}(y_0|x)|x)),0\}} G_{MU}(y|x)\mathrm{d}y\right)$$

Hence we have

$$\mu_{y_0L} = (G_{MU}(y_0|x) - G_{ML}(y_0|x))y_0 + \int_{\max\{y_0,0\}}^{\infty}(1-G_{MU}(y|x))\mathrm{d}y - \int_{\min\{y_0,0\}}^{0} G_{MU}(y|x)\mathrm{d}y +$$

$$\int_{0}^{\max\{G_{MU}^{-1}(G_{ML}(y_0|x)|x)),0\}}(1-G_{MU}(y|x))\mathrm{d}y - \int_{-\infty}^{\min\{G_{MU}^{-1}(G_{ML}(y_0|x)|x)),0\}} G_{MU}(y|x)\mathrm{d}y$$

The improved bounds are

$$\widetilde{G}_{ML}(y_0|x) \leq G(y_0|x) \leq \widetilde{G}_{MU}(y_0|x)$$

with $\widetilde{G}_{ML}(y_0|x) \geq G_{ML}(y_0|x)$ the solution to

$$(\widetilde{G}_{ML}(y_0|x) - G_{ML}(y_0|x))y_0 + \int_{\max\{G_{MU}^{-1}(G_{ML}(y_0|x)|x)),0\}}^{\max\{G_{MU}^{-1}(\widetilde{G}_{ML}(y_0|x)|x)),0\}}(1-G_{MU}(y|x))\mathrm{d}y -$$

$$\int_{\min\{G_{MU}^{-1}(\widetilde{G}_{ML}(y_0|x)|x)),0\}}^{\min\{G_{MU}^{-1}(G_{ML}(y_0|x)|x)),0\}} G_{MU}(y|x)\mathrm{d}y = \max\{\mu_{y_0L} - \mathbb{E}(Y_1 - Y_0), 0\}$$

and $\widetilde{G}_{MU}(y_0|x) \leq G_{MU}(y_0|x)$ the solution to

$$(\widetilde{G}_{MU}(y_0|x) - G_{MU}(y_0|x))y_0 + \int_{\max\{G_{ML}^{-1}(\widetilde{G}_{MU}(y_0|x)|x)),0\}}^{\max\{G_{ML}^{-1}(G_{MU}(y_0|x)|x)),0\}} (1 - G_{ML}(y|x))\mathrm{d}y$$

$$- \int_{\max\{G_{ML}^{-1}(\widetilde{G}_{MU}(y_0|x)|x)),0\}}^{\max\{G_{ML}^{-1}(G_{MU}(y_0|x)|x)),0\}} G_{ML}(y|x)\mathrm{d}y = \max\{\mathbb{E}(Y_1 - Y_0) - \mu_{y_0U}, 0\}$$

Example 1, continued: Difference of normals with the same variance.

Figures 5-7 give the Makarov bounds and the ATE improved bounds for normal outcome distributions. The common standard deviation is $\sigma = 3$ and the ATE-s are 0, 1 and 2. As noted before the bounds on the cdf can be inverted to obtain bounds on the quantiles. These bounds are identical to the bounds that derived in Section 3.

Note that if θ increases the improvement is larger for the upper bound on the cdf or lower bound on the quantiles. For the largest value of the ATE the fraction that is hurt by the intervention is between 0 and .4 for the improved bound but between 0 and .6 for the original Makarov bound. Hence for the improved bound we can conclude that a majority of the population benefits from the intervention. □

Example 2, continued: Mixture of normals with the same variance.

The setup is as before and we use the same parameter values. Figure 8 gives the improved bounds that use the conditional ATE-s, i.e. the conditional Makarov bounds are improved and averaged over the subpopulations. □

REFERENCES

Aakvik, A., J. Heckman, and E. Vytlacil, (2005) "Estimating Treatment Effects for Discrete Outcomes when Responses to Treatment Vary: an application to Norwegian vocational rehabilitation programs," *Journal of Econometrics*, 125, 15-51.

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

Abadie, A., (2003), "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231-263.

Abadie, A., J. Angrist, and G. Imbens, (2002), "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91-117.

Carneiro, P., K. Hansen, and J. Heckman, (2003), "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44, 361-422.

Crump, R., J. Hotz, , G. Imbens, and O. Mitnik, (2006), "Nonparametric Tests for Treatment Effect Heterogeneity," *NBER Technical Working Paper,* No. 324.

Fan, Y., (2005), "Sharp Correlation Bounds and Their Applications," typescript, Department of Economics,Vanderbilt University..

Firpo, S., (2005), "Inequality Treatment Effects," typescript, Department of Economics, The University of British Columbia.

Firpo, S., (2007), "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259-276.

Frank M., R. Nelsen, and B. Schweizer, (1987), "Best-Possible Bounds for the Distribution of a Sum–a Problem of Kolmogorov," *Probability Theory and Related Fields*, 74, 199–211.

Fréchet, M., (1951), "Sur les Tableaux de Corrélation dont les Marges Sont Données," *Annals University Lyon*, Series A, 53-77.

Heckman, J., and J. Smith, (1993), "Assesing the Case for Randomized Evaluation of Social Programs," in Jensen, K and P. Madsen (Eds.) *Measuring Labour Measures: Evaluating the Effects of Active Labour Market Policy Initiatives*, Copenhagen, Ministry Labour.

Heckman, J., and J. Smith, (1998), "Evaluating the Welfare State," in S. Strom (Ed.) *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial*, Econometric Soceity Monograph Series, Cambridge, Cambridge University Press.

Heckman, J., J. Smith, and N. Clements, (1997), "Making the Most out of Programme Evaluations and Social Experiments Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487-535.

Hoeffding, W., (1940), "Masstabinvariate Korrelations-Theorie," *Scriftenreite Math. Inst. Univ. Berlin*, 5, 181–233.

Hotz, J., G. Imbens and J. Klerman, (2006), "Evaluating the Differential Effects of Alternative Welfare-to-Work Training Components: A Re-Analysis of the California GAIN Program,", *Journal of Labor Economics*, 24, 521-566.

Imbens, G., and C. Manski, (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845-1857.

Imbens, G., and D. Rubin, (1997), "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies*, 64, 555–574.

Makarov, G., (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.

Manski, C., (1997), "The Mixing Problem in Programme Evaluation," *Review of Economic Studies*, 64, 537–553.

Riccio, J., D. Friedlander, and S. Freedman, (1994), "GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program," *Report on California's Grater Avenues for Independence (GAIN) Program*, Manpower Demonstration Research Corporation (MDRC), September.

Shaked, M., and J.G. Shanthikumar, (1994), Stochastic Orders and Their Applications, Academic Press, San Diego.

Whitt, W., (1976), "Bivariate Distributions With Given Marginals," *The Annals of Statistics*, 4, 1280-1289.

Wu, X., and J. Perloff, (2006), "Information-Theoretic Deconvolution Approximation of Treatment Effect Distribution," typescript, Department of Agricultural Economics, Texas A&M University.

Fig 5. *Makarov bounds and improved Makarov bounds: Normal outcome distributions* $\theta = 0, \sigma = 3$

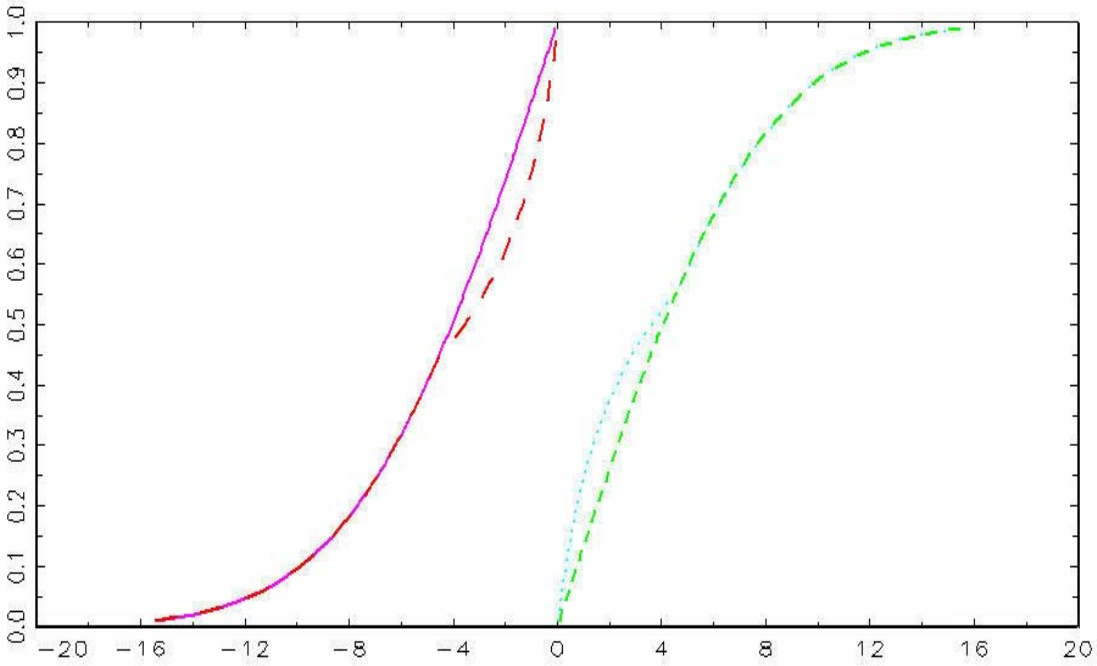


Fig 6. *Makarov bounds and improved Makarov bounds: Normal outcome distributions* $\theta = 1, \sigma = 3$

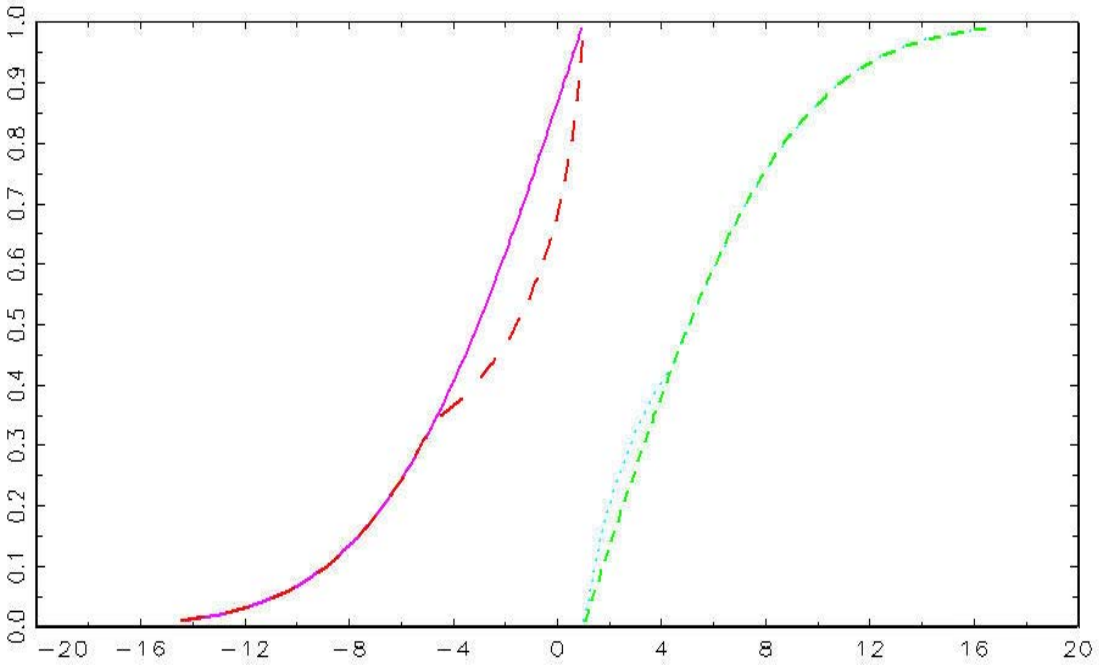


Fig 7. *Makarov bounds and improved Makarov bounds: Normal outcome distributions* $\theta = 2, \sigma = 3$

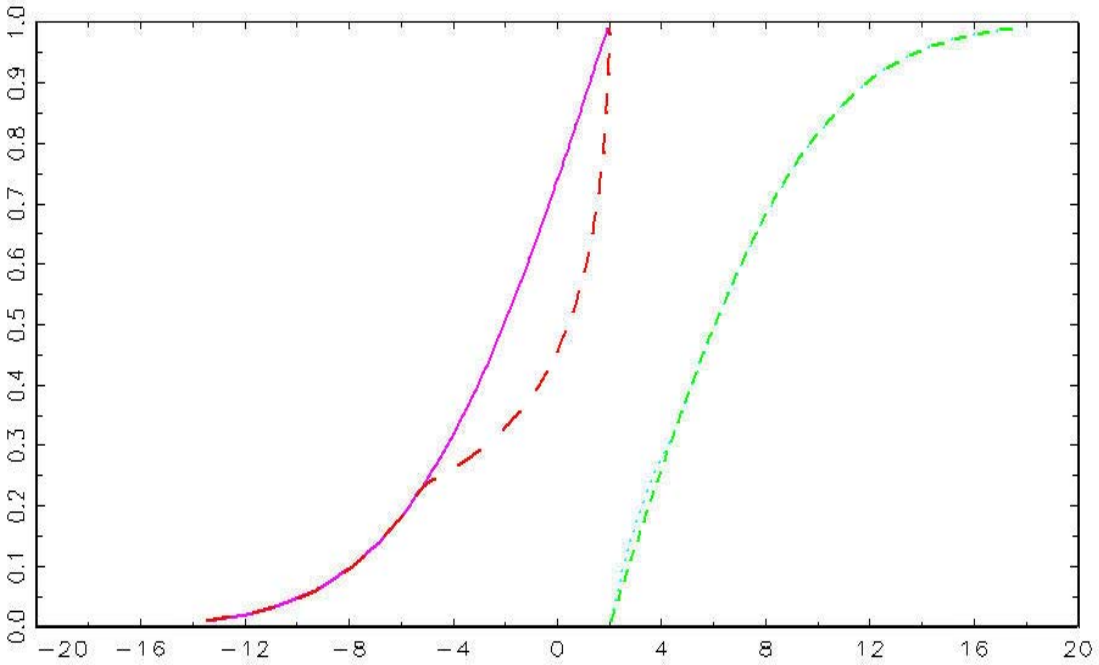


Fig 8. *Average improved Makarov bounds and Makarov bounds for the average population: Normal two-point mixture with* $\theta_1 = 0, \theta_2 = 2$.