

**Textos para
Discussão**

201

Setembro
de 2008



Bounds on Functionals of the Distribution
of
Treatment Effects

SERGIO FIRPO
GEERT RIDDER

Os artigos dos *Textos para Discussão da Escola de Economia de São Paulo da Fundação Getúlio Vargas* são de inteira responsabilidade dos autores e não refletem necessariamente a opinião da FGV-EESP. É permitida a reprodução total ou parcial dos artigos, desde que creditada a fonte.

Escola de Economia de São Paulo da Fundação Getúlio Vargas FGV-EESP
www.fgvsp.br/economia

Bounds on Functionals of the Distribution of Treatment Effects*

Sergio Firpo

Escola de Economia de São Paulo, FGV-Brazil[†]

Geert Ridder

Department of Economics, University of Southern California [‡]

September 22, 2008

Abstract

Bounds on the distribution function of the sum of two random variables with known marginal distributions obtained by Makarov (1981) can be used to bound the cumulative distribution function (c.d.f.) of individual treatment effects. Identification of the distribution of individual treatment effects is important for policy purposes if we are interested in functionals of that distribution, such as the proportion of individuals who gain from the treatment and the expected gain from the treatment for these individuals. Makarov bounds on the c.d.f. of the individual treatment effect distribution are pointwise sharp, i.e. they cannot be improved in any single point of the distribution. We show that the Makarov bounds are not uniformly sharp. Specifically, we show that the Makarov bounds on the region that contains the c.d.f. of the treatment effect distribution in two (or more) points can be improved, and we derive the smallest set for the c.d.f. of the treatment effect distribution in two (or more) points. An implication is that the Makarov bounds on a functional of the c.d.f. of the individual treatment effect distribution are not best possible.

Keywords: Treatment effects, bounds, social welfare.

JEL-code: C31

*Financial support for this research was generously provided through NSF grant SES 0452590 (Ridder) and by CNPq (Firpo). We thank Rustam Ibragimov, Guido Imbens, Yanqin Fan and seminar and conference participants at Harvard Econometrics Lunch, IZA-Bonn, 2007 North American Winter Meeting of the Econometric Society and UFRJ-Brazil for comments.

[†]Rua Itapeva 474, São Paulo, Brazil, 01332-000. Electronic correspondence: sergio.firpo@fgv.br.

[‡]Kaprilian Hall, Los Angeles, CA 90089. Electronic correspondence: ridder@usc.edu, <http://www-rcf.usc.edu/~ridder/>.

1 Introduction

The key problem when estimating the effect of a treatment or intervention on a population is that we cannot observe both the treated and non-treated outcomes for a unit in the population, but at most either its treated or non-treated outcome. As a consequence, we can only identify treatment effect parameters that depend on the marginal distributions of the treated and control outcomes and, in general, not parameters that depend on the distribution of individual treatment effects. The only exception is the mean of the individual treatment effect distribution, the Average Treatment Effect (ATE), which, given linearity of expectations, can be identified from the marginal distributions of treated and control outcomes.

Under the assumption that the social welfare function (SWF) is a functional of the distribution of outcomes, gains or losses in social welfare due to an intervention can be measured as the difference of functionals on the marginal distributions of treated and non-treated outcomes. For instance, we may be interested in the effect of a program on the inequality of outcomes in the population. If we choose some inequality measure, say the variance, then the effect of the program on the variance is equal to the difference of the variances of the marginal distributions of the treated and control outcomes. Such an approach has been used, for example, in Imbens and Rubin (1997), Abadie, Angrist Imbens (2002), Abadie (2002, 2003) and Firpo (2007). Therefore, if our goal is to assess the effect of an intervention on social welfare and not individual welfare, then the marginal outcome distributions suffice.

There are some other treatment effect parameters that are defined as functionals of the distribution of individual treatment effects. Examples of functionals of the distribution of individual treatment effects are the fraction of the population that benefits from a program, the total and average gains of those who benefit from the program, the fraction of the population that has gains or losses in a specific range, and the median (or other quantile) of the treatment effect distribution.¹ Heckman, Smith and Clements (1997) discuss a number of parameters that depend on the distribution of individual treatment effects.

We show that a general reason why we should be interested in functionals of the distribution of individual treatment effects is that individuals in a population may be loss averse. Loss aversion has been shown to be a feature of individual preferences if an individual faces an uncertain outcome (e.g. Tversky and Kahneman (1991) and the large literature on non-expected utility). With loss aversion at the individual level a utilitarian social welfare function will exhibit aversion to redistribution. As a consequence the social welfare function depends on the distribution of individual treatment effects.

Point identification of parameters that depend on the distribution of individual treatment effects requires knowledge of the joint distribution of treated and non-treated outcomes, as the marginal themselves do not contain enough infor-

¹Because the difference of quantiles is not equal to the quantile of the difference, we cannot estimate this parameter by the difference of the medians of the treated and non-treated outcome distributions.

mation to identify the distribution of the difference. 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 by a simple location shift. 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. Heckman, Smith and Clements (1997) 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. Fan and Park (2007) were the first to apply these bounds to the distribution of treatment effects with an emphasis on the statistical inference for these bounds.

This paper will disregard inference completely and will focus instead on the nature of the Makarov bounds. An important property of a bound is whether it is sharp or best possible. Our results show that Makarov bounds are *pointwise* but not *uniformly* sharp. This implies that Makarov bounds on functionals of the distribution of individual treatment effects are in general not sharp. In the case of a scalar parameter bounds are defined by a set of restrictions on the parameter. Assume for simplicity that these restrictions imply that the parameter is in a closed connected interval. A lower bound on the parameter is sharp if every parameter value that satisfies the restrictions is not smaller than the bound and the bound itself satisfies all the restrictions. In the case that we bound a function defined on some domain the definition of a sharp bound is not as simple. Again the bounds are defined by a set of restrictions. In our case we consider all c.d.f. of a distribution of $Y_1 - Y_0$ where Y_0, Y_1 have a joint distribution with given marginal distributions. If the bounding functions satisfy all the restrictions we call them uniformly sharp. This corresponds to the usual definition of sharpness for a scalar parameter. The Fréchet (1951) (see also Hoeffding (1940)) bounds on the joint distribution of two random variables with given marginal distributions are uniformly sharp. It is however possible that the bounding functions do not satisfy all the restrictions. This is the case with the Makarov bounds. In that case it is possible that the bounds are best possible in a point (and every point) of the domain. This occurs if there is a function that satisfies all the restrictions and is equal to the bounding function at that point (and such a function exists at every point)². We call such a bound pointwise sharp. The Makarov bounds are pointwise, but not uniformly sharp.

If a bound is uniformly sharp, then the joint bound on the set of function values in two (or more) points on the domain derived from the uniformly sharp bound and possible other restrictions like monotonicity is also sharp. This is not true if the bounds are pointwise sharp.

In this paper, we show that a joint bound of c.d.f. points using the Makarov bounds is not best possible. Moreover, we derive more informative joint bounds,

²The supporting function cannot intersect the bound, but touches the bound at the point.

i.e. a smaller region, for the c.d.f. of the individual treatment effect in two (or more) points. This result is not at odds with the sharpness of the Makarov bounds in a single point, because the projections of the smaller higher-dimensional region coincides with the one-dimensional Makarov bounds. Bounds on the treatment effect c.d.f. in two (or more) points imply bounds on functions of the treatment effect c.d.f. in those points. We consider linear functionals of the treatment effect c.d.f. and derive conditions under which the bounds on this functional can be improved.

A second contribution of this paper is that we show that if the outcomes are correlated with covariates, then averaging the bounds obtained from the conditional (on these covariates) outcome distributions gives bounds that are more informative than the bounds obtained from the unconditional outcome distributions. This result holds both for the one-dimensional pointwise Makarov bounds and for the improved higher dimensional regions. Hence, even if treatment is randomly assigned it is useful to have covariates that are correlated with the outcomes in order to improve the bounds on (functionals of) the distribution of the individual treatment effects.

There is a small literature on bounds on the treatment effect c.d.f. in a point for given marginal outcome distributions. None of it considers bounds on the c.d.f. in two or more points or bounds on functionals of the c.d.f. We already mentioned Fan and Park (2007) who use the pointwise sharp Makarov bounds. Most papers introduce additional restrictions, as a factor structure or rank preservation that narrow the bounds or even lead to point identification of the treatment effect distribution. In chronological order contributions can be found in Heckman and Smith (1993, 1998) and in particular Heckman, Smith and Clements (1997), Aakvik, Heckman and Vytlačil (2005), Carneiro, Hansen and Heckman (2003), and Wu and Perloff (2006). Djebbari and Smith (2008) use the Heckman-Smith-Clements bounds³ in an empirical study of the distribution of treatment effects in a conditional cash transfer program in Mexico.

The plan of the paper is as follows. In section 2 we show that if individuals are loss averse then the social welfare function is a functional of the distribution of individual treatment effects. In section 3 we discuss the Makarov bounds on the cdf of treatment effects and we introduce the concepts of pointwise and uniformly sharp bounds. Section 4 establishes that the Makarov bounds are pointwise, but in general not uniformly sharp. In section 5 we show that averaging over covariates that are correlated with the outcomes improves the bounds. In section 6 we obtain higher dimensional Makarov bounds and we derive a necessary condition for a vector of function values to be compatible with a treatment effect distribution. We then use this necessary condition to show that the higher dimensional Makarov bounds are in general not sharp and we derive improved bounds. In section 7 we use these improved bounds to obtain improved bounds on functionals of the treatment effect distribution. Section 8 concludes.

³These bounds are derived from the Fréchet-Hoeffding bounds on the joint distribution of treated and non-treated outcomes and are not best possible.

2 Welfare and the distribution of treatment effects

Consider an intervention with potential outcomes Y_{0i} and Y_{1i} for individual i of the population. The individual has a vector of characteristics X_i . An experiment is performed in a randomly selected sample from this population and treatment assignment T_i in the sample is either random or unconfounded given X . Hence, if the sample is large we can identify $F_0(\cdot|x)$ and $F_1(\cdot|x)$ for all $x \in \mathcal{X}$ with \mathcal{X} the support of the distribution of X . Let us define $D_i = Y_{1i} - Y_{0i}$ and assume that all i know Y_{0i} , their non-treated or status quo outcome, but not necessarily Y_{1i} , their treated outcome at the time of treatment. In general Y_{1i} can be thought of as a function of X_i and ε_i , where X_i is a vector of characteristics that is known to the individual and ε_i is a random term that may or may not be known to individual i at the time of the intervention. We assume that Y_{0i} is known to the individual at the time of the intervention even if i undergoes the intervention. If i undergoes the intervention, then Y_{0i} is not known to the econometrician or the social planner. The vector X_i is observed irrespective of treatment assignment (and not affected by that).

Note that the treated and control outcomes are treated asymmetrically. Often individuals can predict their outcome under the status quo accurately but not necessarily their outcome under the treatment. We consider both the case that Y_{1i} is known at the time of the treatment and the case that this outcome is not known at that time. Moreover, we consider two types of preferences. The first type corresponds to expected utility in the case that Y_{1i} is unknown to i at the time of the treatment. The second type assumes that individuals are loss averse, as introduced by Tversky and Kahneman (1991) and extensively discussed by Rabin (1998). For the second type of preferences we need the distribution of D at the time of treatment. We also consider the utilitarian social welfare functions corresponding to these individual preferences.

The social welfare functions are our main focus in this section. They are the same irrespective which assumption we make on knowledge of Y_1 at the time of treatment (we consider the unknown Y_1 case only for expositional purposes). If we start from individual preferences that exhibit loss aversion we obtain a social welfare function that has redistribution aversion. In particular, if we fix the average benefit of an intervention, i.e. the Average Treatment Effect (ATE), then society will prefer an intervention that spreads the gains evenly in the population over an intervention that achieves the same ATE with large losses for some and slightly larger gains for others in the population. Easterlin (2008) discusses the relevance of the distribution of gains and losses for social welfare in a transition economy.

First, we assume that Y_{1i} is not known (but Y_{0i} is) to i at the time of the treatment. Both outcomes are net of the private cost of treatment and non-treatment. If the utility of outcome Y is $u(Y)$ with u concave if i is risk averse, then the expected utility of treatment for i is $\mathbb{E}[u(Y_{1i})|Y_{0i}, X_i]$ and the expected utility of the status quo is $u(Y_{0i})$. The utilitarian social welfare that sums these

individual preferences over all members of the population is

$$W_1 = \mathbb{E}[\mathbb{E}[u(Y_1)|Y_0, X]] = \mathbb{E}[u(Y_1)]$$

and

$$W_0 = \mathbb{E}[u(Y_0)]$$

Both individual and social welfare only depend on the marginal distributions of Y_0 and Y_1 (given X). If we assume that both Y_{0i} and Y_{1i} are known to i at the time of treatment then the individual utilities of treatment and non-treatment are $u(Y_{0i})$ and $u(Y_{1i})$, respectively. Therefore, the utilitarian social welfare function assigns W_1 and W_0 to treatment and non-treatment, which are the same values as in the case that Y_{1i} is not known at the time of treatment. The obvious conclusion is that utilitarian social welfare depends only on the marginal outcome distributions and the distribution of D does not play a role⁴.

As Tversky and Kahneman (1991) have pointed out, individual preferences are actually not as in the standard expected utility theory. In (cumulative) prospect theory preferences exhibit the so-called “framing effect”, because people tend to think of possible outcomes relative to a reference value. In the simple potential outcome model the natural choice for the reference value is the status quo outcome Y_{0i} . Moreover, individuals have different risk attitudes towards gains $Y_{1i} - Y_{0i} > 0$ and losses $Y_{1i} - Y_{0i} < 0$ and the disutility of a loss is in general larger than the utility of an equal gain. This is called *loss aversion*. If we denote the valuation function of gains/losses by $v(Y_1 - Y_0) = v^+(Y_1 - Y_0)1(Y_1 - Y_0 > 0) + v^-(Y_1 - Y_0)1(Y_1 - Y_0 \leq 0)$ then the utility of non-treatment is 0 (essentially a normalization) and the utility of treatment is⁵

$$\mathbb{E} [v^+(Y_{1i} - Y_{0i})1(Y_{1i} - Y_{0i} > 0)|Y_{0i}, X_i] + \mathbb{E} [v^-(Y_{1i} - Y_{0i})1(Y_{1i} - Y_{0i} \leq 0)|Y_{0i}, X_i].$$

Hence the utilitarian social welfare is $V_0 = 0$ and

$$V_1 = \mathbb{E} [v^+(Y_1 - Y_0)1(Y_1 - Y_0 > 0)] + \mathbb{E} [v^-(Y_1 - Y_0)1(Y_1 - Y_0 \leq 0)].$$

Two interesting particular cases of valuation functions are the following. If $v^+(Y_1 - Y_0) = 1$ and $v^-(Y_1 - Y_0) = -1$ then V_1 is the difference of the fractions of the population with a positive and a negative treatment effect respectively. This majority parameter is mentioned by Heckman, Smith and Clements (1997). If

$$v^+(Y_1 - Y_0) = v^-(Y_1 - Y_0) = u(Y_1) - u(Y_0),$$

then the expected utility and loss aversion social welfare functions are the same (up to normalization).

In general, the valuation functions are such that $v^+(0) = v^-(0) = 0$, $v^+(z) \geq 0$ for $z \geq 0$, $v^-(z) \leq 0$ for $z < 0$, and $v^+(z) < -v^-(-z)$, where the final

⁴We could also let the utility function depend on X , i.e., consider $u(Y, X)$. In that case knowledge of the conditional distributions of Y_0 and Y_1 given X is sufficient.

⁵In cumulative prospect theory we would also weight the cdf of gains and losses with a weighting function. This is not essential for our argument.

condition is loss aversion. Also, it is often assumed that v^+ is concave and v^- is convex, and both are increasing in z .

Now, suppose we want compare two possible treatments, A and B . Both treatments have the same ATE, $\mathbb{E}[Y_1] - \mathbb{E}[Y_0]$. However, for treatment A every individual has a gain equal to the ATE and for treatment B some individuals have a large loss while an equal fraction of the population has a gain that exceeds the (opposite of the) loss by the ATE. It is obvious that treatment A is preferred over treatment B if individuals are loss averse. Treatment A does not involve any redistribution of gains while under treatment B gains and losses are unequal. Therefore we can say that a social welfare function derived from loss averse individual preferences shows *redistribution aversion*.

The loss aversion social welfare function is also relevant if individual treatment effects are nonnegative, i.e. if all individuals benefit from the treatment. Individuals may still use the status quo outcome as a reference. As a consequence, society may prefer less variation in the distribution of individual gains.

Although we derived the social welfare function on the assumption that individuals use the known non-treated outcome as a reference, the analysis is also relevant in the case that treated individuals only learn Y_1 (so that the reference value is unknown) and control individuals only learn Y_0 . Let us first assume that the identified F_0 and F_1 are known. For individuals who are in the control group the expected utility under loss aversion is as above (the expectation is over F_1). For individuals who are treated the expected utility is of the same form except that the expectation is over F_0 . The social welfare function does not change. If individuals only learn Y_0 or Y_1 and not their marginal distributions, then the social planner may still care about the distribution of gains and use the redistribution averse social welfare function. Of course, because only F_0 and F_1 are identified the social planner can only prefer treatment A over treatment B if the lower bound on the social welfare of A exceeds the upper bound on the social welfare of B . If the bounds overlap the social planner has to use some criterion to rank the treatments, e.g. the largest lower bound.

The social welfare function that assumes that individual preferences exhibit loss aversion depends in general on the distribution of the individual treatment effect D . By partial integration we find

$$V_1 = \int_0^\infty v^{+'}(z) \cdot (1 - G(z)) dz - \int_{-\infty}^0 v^{-'}(z) \cdot G(z) dz$$

where $v^{+'}(z)$, $v^{-'}(z)$ are the derivatives of $v^+(z)$, $v^-(z)$ and G is the cdf of D . As noted, the functions v^+ and v^- are in general nonlinear increasing functions. An obvious specification is a linear spline with nodes $0 = d_0 < d_1 < \dots < d_K = \infty$ and $v_0^+ = 0$, $v_k^+ > 0$

$$v^+(z) = \sum_{k=1}^K [v_k^+(z - d_{k-1}) + v_{k-1}^+ d_k] 1(d_{k-1} \leq z < d_k)$$

and a similar specification for v^- . With this specification the derivatives of v^+

and v^- are step functions so that

$$\int_0^\infty v^{+'}(z)(1 - G(z))dz = \sum_{k=1}^K v_k^+ \int_{d_{k-1}}^{d_k} (1 - G(z))dz$$

In section 7 we consider bounds on integrals $\int_{d_{k-1}}^{d_k} (1 - G(z))dz$.

3 Pointwise and uniformly sharp bounds on the distribution of treatment effects

Let \mathcal{G} be a set of distribution functions on \mathfrak{R} , i.e. a set of non-decreasing and right-continuous functions on \mathfrak{R} that are 0 in $-\infty$ and 1 in ∞ . All distribution functions in \mathcal{G} satisfy a set of restrictions. In this paper the restriction is that each $G \in \mathcal{G}$ is the c.d.f. of $D = Y_1 - Y_0$ for given marginal c.d.f. of Y_1 , denoted by F_1 , and Y_0 , denoted by F_0 , but unspecified joint distribution of Y_0, Y_1 . We are interested in bounds on the distribution functions in \mathcal{G} , which is the set of distributions of individual treatment effects for given marginal outcome distributions. We often have a vector of covariates X with a distribution with support \mathcal{X} that are correlated with Y_1 and Y_0 , so that in the statement above we can replace the treatment effect distribution by the conditional treatment effect distribution given X and the outcome distributions by conditional outcome distributions. The bounds on the distribution of the treatment effect are obtained by averaging the conditional bounds over the distribution of X . Sometimes it is convenient to ignore the fact that we are dealing with conditional distributions and only to introduce the covariates in the final result. In general, averaging makes the bounds more informative.

An upper and lower bound on $G(d)$ for $G \in \mathcal{G}$ was derived by Makarov (1981) (see also Frank, Nelsen, and Schweizer, 1987). Note that this is a bound for the c.d.f. in a single point. We extend the Makarov bound to the case that we observe conditional marginal distributions of the outcomes $F_0(\cdot|x)$ and $F_1(\cdot|x)$.

Theorem 3.1 (Makarov, 1981) *Let the conditional c.d.f. of $Y_0|X$ and $Y_1|X$ be $F_0(\cdot|X)$ and $F_1(\cdot|X)$ and $G \in \mathcal{G}$ be the c.d.f. of $D = Y_1 - Y_0$, then for $-\infty < d < \infty$*

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

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

with $F_0(\cdot)_-$ the function of left-hand limits of the c.d.f.. The bounds $G_{ML}(d), G_{MU}(d)$ are c.d.f., i.e. non-decreasing, right-continuous, and 0 and 1 for $d \downarrow -\infty$ and $d \uparrow \infty$, respectively.

Proof:See Appendix B.

Are the Makarov bounds on the treatment effect c.d.f. the best possible bounds? The answer to this question depends on our definition of best possible bounds. Because we are bounding a function we can consider bounds in each point d or joint bounds for all d . If we consider bounds in a single point d , then the relevant notion is *pointwise sharpness*. To keep the notation simple the discussion is for unconditional c.d.f. but it applies directly to conditional c.d.f.

Definition 3.1 (Pointwise sharp bounds on a c.d.f.) *Let \mathcal{G} be a set of c.d.f. and let $G_L(d) \leq G(d) \leq G_U(d)$ for all $d \in \mathfrak{R}$ and $G \in \mathcal{G}$. We say that G_L is a pointwise sharp lower bound on \mathcal{G} , if for all d_0 there is a c.d.f. $G_{d_0L} \in \mathcal{G}$ such that $G_L(d_0) = G_{d_0L}(d_0)$. G_U is a pointwise sharp upper bound on \mathcal{G} , if for all d_0 there is a c.d.f. $G_{d_0U} \in \mathcal{G}$ such that $G_U(d_0) = G_{d_0U}(d_0)$.⁶*

Note that the c.d.f. that supports the lower bound may depend on d_0 . If for all d_0 the supporting c.d.f. does not depend on d_0 we call the lower bound *uniformly sharp*. A uniformly sharp upper bound is defined analogously.

Definition 3.2 (Uniformly sharp bounds on a c.d.f.) *Let \mathcal{G} be a set of c.d.f. and let $G_L(d) \leq G(d) \leq G_U(d)$ for all $d \in \mathfrak{R}$ and $G \in \mathcal{G}$. We say that G_L is a uniformly sharp lower bound on \mathcal{G} , if $G_L \in \mathcal{G}$. G_U is a uniformly sharp upper bound on \mathcal{G} , if $G_U \in \mathcal{G}$.*

It should be noted that if the bounds are uniformly sharp they have all the properties of the set \mathcal{G} . If they are pointwise sharp, the bounds will have some but not all properties of \mathcal{G} .

Theorem 3.2 *If \mathcal{G} is a set of c.d.f. with pointwise sharp bounds G_L, G_U , then G_L, G_U are non-decreasing, right-continuous and 0 and 1 at $-\infty$ and ∞ , respectively, i.e. the bounds are themselves c.d.f.⁷*

Proof:See Appendix B.

4 Makarov bounds are pointwise sharp

We are now able to answer the question whether the Makarov bounds are best possible. Frank, Nelsen, and Schweitzer (FNS) (1981) construct for any d_0 joint distributions H_{d_0L} and H_{d_0U} of Y_0, Y_1 such that

$$G_{ML}(d_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{v+d_0} dH_{d_0L}(u, v) \quad G_{MU}(d_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{v+d_0} dH_{d_0U}(u, v)$$

⁶Although we define these concepts for sets of c.d.f. they apply to any set of functions on \mathfrak{R} .

⁷The same argument as in Theorem 2.1. can be used to find bounds on the distribution of the sum of three or more random variables with given (marginal) distributions (see Kreinovich and Ferson (2005)). These bounds are in general not a c.d.f. and hence are not pointwise sharp.

i.e. these joint distributions support the lower and upper bounds. It is instructive to show the result of the construction, because it clearly illustrates that the supporting c.d.f. are local, i.e. they depend on d_0 , so that the Makarov bounds are pointwise sharp. To keep the notation simple we consider the case that the marginal c.d.f. F_0 and F_1 are strictly increasing on the respective supports. We only consider the supporting c.d.f. G_{d_0L} of the lower Makarov bound in d_0 . Instead of the joint c.d.f. of Y_1, Y_0 , H_{d_0L} , we consider that of $Y_1, -Y_0$, \tilde{H}_{d_0L} that has a simpler form. Define $u_0 = F_1^{-1}(G_{ML}(d_0))$, $v_0 = d_0 - u_0$, and $v_1 = -F_0^{-1}(1 - G_{ML}(d_0))$ where $v_1 \leq v_0$. The supporting c.d.f. is obtained from

$$\begin{aligned} \tilde{H}_{d_0L}(u, v) &= F_1(u) && u < u_0, v > v_0 \\ &= \min\{F_1(u), 1 - F_0(-v)\} && u \leq u_0, v \leq v_0 \\ &= \min\{1 - F_0(-v), G_{ML}(d_0)\} && u > u_0, v \leq d_0 - u \\ &= \max\{F_1(u) - F_0(-v), G_{ML}(d_0)\} && u \geq u_0, v \geq v_0 \text{ or } u \geq d_0 - v, v_1 \leq v < v_0 \\ &= 1 - F_0(-v) && u \geq d_0 - v, v < v_1 \end{aligned}$$

The regions are as in Figure 1. Using this figure it is easily checked that the c.d.f. has the correct marginal distributions $F_0(y)$ and $1 - F_1(-y)$.

The joint distribution of $Y_1, -Y_0$ is singular, because all probability is concentrated on two curves

$$\mathcal{S}_1 = \{(u, v) | v = -F_0^{-1}(1 - F_1(u)), u \leq u_0\}$$

and

$$\mathcal{S}_2 = \{(u, v) | v = -F_0^{-1}(F_1(u) - G_{ML}(d_0)), u > u_0\}$$

If $G_{ML}(d_0) > 0$ (we only consider this case; if $G_{ML}(d_0) = 0$ the analysis is slightly different), the curve \mathcal{S}_1 is increasing in u and is equal to v_1 if $u = u_0$. The curve \mathcal{S}_2 is ∞ if $u = u_0$ and converges to v_1 as $u \rightarrow \infty$. Moreover if \tilde{u} minimizes $F_1(u) - F_0(-(d_0 - u))$ (the minimand need not be unique), then $F_1(\tilde{u}) - F_0(-(d_0 - \tilde{u})) = G_{ML}(d_0)$ so that

$$d_0 - \tilde{u} = -F_0^{-1}(F_1(\tilde{u}) - G_{ML}(d_0))$$

and we conclude that \mathcal{S}_2 touches the line $u + v = d_0$ at all minimands \tilde{u} . The same argument shows that \mathcal{S}_2 cannot be below the line $u + v = d_0$. The two curves are drawn in Figure 1 for the case that there is a unique minimand \tilde{u} .

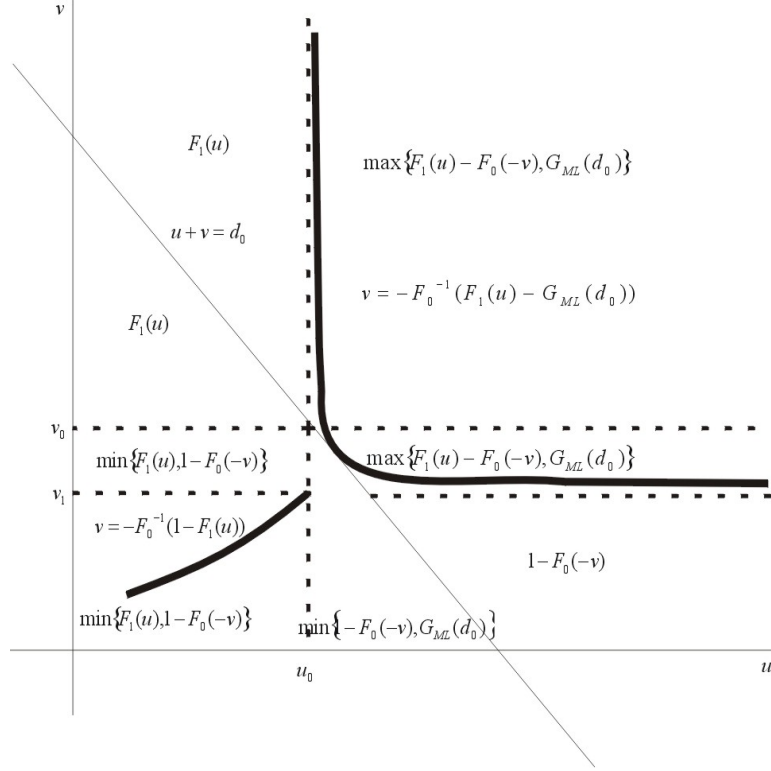
The c.d.f. $G_{d_0L}(d)$ that supports the lower Makarov bound in d_0 is obtained by computing the probability mass in the set $\{(u, v) | u + v \leq d\}$. For $d \leq d_0$

$$\begin{aligned} G_{d_0L}(d) &= G_{ML}(d_0) && u_0 + v_1 \leq d \leq d_0 \\ &= F_1(u(d)) && d < u_0 + v_1 \end{aligned} \tag{2}$$

with $u(d)$ the solution to

$$F_1(u(d)) = 1 - F_0(-(d - u(d))) \tag{3}$$

Fig 1. Definition and support of G_{d_0L} if there is a unique minimand.



i.e. \mathcal{S}_1 intersects $u + v = d$ at $u = u(d), v = d - u(d)$. Note that $G_{d_0L}(d)$ is constant if $u_0 + v_1 \leq d \leq d_0$, because the line $u + v = u_0 + v_1$ intersects \mathcal{S}_1 at $u = u_0, v = v_1$ and

$$G_{d_0L}(u_0 + v_1) = \tilde{H}_{d_0L}(u_0, v_1) = F_1(u_0) = G_{ML}(d_0)$$

For $d > d_0$

$$G_{d_0L}(d) = F_1(u_2(d)) - F_0(-(d - u_1(d))) \quad (4)$$

where $u_1(d) \leq u_2(d)$ are the two solutions to

$$u = d + F_0^{-1}(F_1(u) - G_{ML}(d_0))$$

i.e. the two points of intersection of \mathcal{S}_2 and $d - u$.

Note that for d_0, d_1 $G_{d_0L}(d) = G_{d_1L}(d)$ for all $d \in \mathfrak{R}$ iff $G_{ML}(d_0) = G_{ML}(d_1)$. Therefore

Theorem 4.1 *The lower Makarov bound G_{ML} is uniformly sharp on a set where G_{ML} is constant. The same holds for the upper Makarov bound. On sets where the bounds are not constant the bounds are pointwise, but not uniformly sharp.*

Although according to Theorem 2.2 pointwise sharp bounds are c.d.f. they need not have all the properties of \mathcal{G} . We show this in an example which we will use as an illustration throughout this paper.

Example 1: Difference of normals with the same variance.

Consider

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

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

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

The corresponding density is

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

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

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

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

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

The corresponding density is

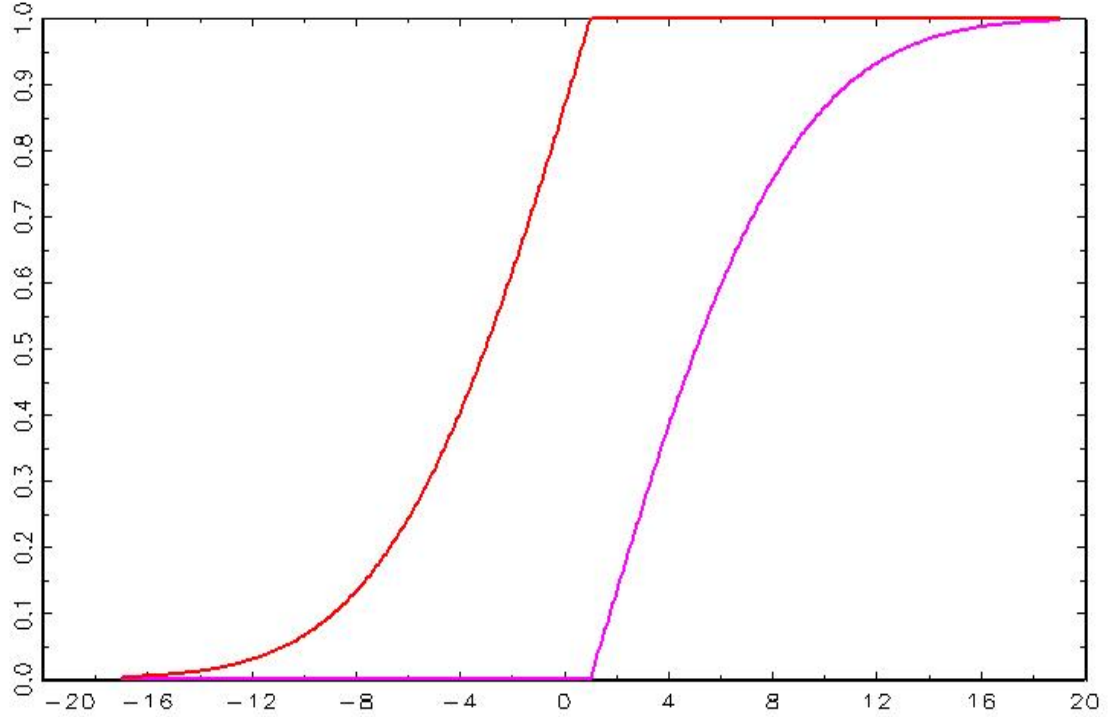
$$\begin{aligned} g_{MU}(d) &= \frac{1}{\sigma} \phi\left(\frac{d-\theta}{2\sigma}\right) && \text{if } d < \theta \\ &= 0 && \text{if } d \geq \theta \end{aligned}$$

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

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

The bounds are drawn in Figure 2 for $\theta = 1$ and $\sigma = 3$. Note that the bounds are not informative if $d = \theta$.

Fig 2. *Makarov bounds on the treatment effect c.d.f.: Normal outcome distributions with equal variance $\theta = 1, \sigma = 3$.*



It is also illustrative to give the supporting c.d.f. that passes through the lower bound $G_{ML}(d_0)$. For $d \leq d_0$ from (3)⁸

$$u(d) = \frac{d + \mu_0 + \mu_1}{2}$$

Also

$$u_0 = \mu_1 + \sigma\Phi^{-1}(G_{ML}(d_0)) \quad v_0 = d_0 - u_0 \quad v_1 = -\mu_0 - \sigma\Phi^{-1}(1 - G_{ML}(d_0))$$

⁸This holds for all symmetric outcome distributions.

Therefore

$$\begin{aligned}
G_{d_0L}(d) &= \Phi\left(\frac{d-\theta}{2\sigma}\right) && d < u_0 + v_1 \\
&= 2\Phi\left(\frac{d_0-\theta}{2\sigma}\right) - 1 && u_0 + v_1 \leq d \leq d_0 \\
&= \Phi\left(\frac{u_2(d)-\mu_1}{\sigma}\right) - \Phi\left(\frac{-(d-u_1(d))-\mu_0}{\sigma}\right) && d > d_0
\end{aligned}$$

with $u_1(d) < u_2(d)$ the solutions to

$$u = d + \mu_0 + \sigma\Phi^{-1}\left(\Phi\left(\frac{u-\mu_1}{\sigma}\right) - G_{ML}(d_0)\right)$$

□.

The conclusion is that although all c.d.f. in the set of treatment effect distributions have mean θ , the c.d.f. that correspond to the lower and upper Makarov bounds have a mean that is strictly larger and smaller than θ . Hence they do not have all the properties of the set of c.d.f. that they bound. The Makarov bounds are envelopes of the c.d.f. that support them, i.e. the c.d.f. in (2) and (4). These envelopes need not have a mean equal to θ .

5 Averaging over covariates

The conditional on X Makarov bounds on the conditional treatment effect distribution in a point d are pointwise sharp. If we average these conditional pointwise sharp bounds over X we obtain pointwise sharp bounds on the unconditional treatment effect distribution. To see this we construct the supporting joint c.d.f. conditional on X as in the previous section where we substitute conditional outcome distributions for unconditional ones. Averaging this supporting conditional joint c.d.f. over X we obtain the unconditional joint c.d.f. that has marginal distributions equal to the given (unconditional) outcome distributions of Y_0 and Y_1 . The distribution of $Y_1 - Y_0$ derived from this average supporting c.d.f. has a c.d.f. that is equal to the lower or upper average Makarov bounds in d , depending on which supporting c.d.f. we use.

The pointwise sharp average Makarov bounds improve on the bounds derived from the average, i.e. unconditional, outcome distributions.

Theorem 5.1 *Averaging over covariates gives tighter bounds, that is,*

$$\sup_t \max\{\mathbb{E}[F_1(t|X)] - \mathbb{E}[F_0(t-d|X)]_-, 0\} \leq \mathbb{E}\left[\sup_t \max\{F_1(t|X) - F_0(t-d|X)_-, 0\}\right] \quad (5)$$

$$\mathbb{E}\left[\inf_t \min\{F_1(t|X) - F_0(t-d|X)_- + 1, 1\}\right] \leq \inf_t \min\{\mathbb{E}[F_1(t|X)] - \mathbb{E}[F_0(t-d|X)]_- + 1, 1\} \quad (6)$$

Proof: See Appendix B.

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 c.d.f.. The next example illustrates the role of averaging for normal outcome distributions.

Example 2: Conditional normal outcome distributions.

The conditional outcome distributions are

$$Y_k|X \sim N(\alpha_k + \beta_k X, \sigma^2) \quad k = 0, 1$$

i.e. they are obtained from linear regression models with normal errors with the same variance that does not depend on X . The ATE given X is $\theta(X) = \alpha_1 - \alpha_0 + (\beta_1 - \beta_0)X$. The conditional lower Makarov bound is

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

and the conditional upper Makarov bound is

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

Hence the average lower bound is

$$\mathbb{E}[G_{ML}(d|X)] = \mathbb{E}\left[I(d \geq \theta(X))\left(2\Phi\left(\frac{d - \theta(X)}{2\sigma}\right) - 1\right)\right]$$

and the average upper bound is

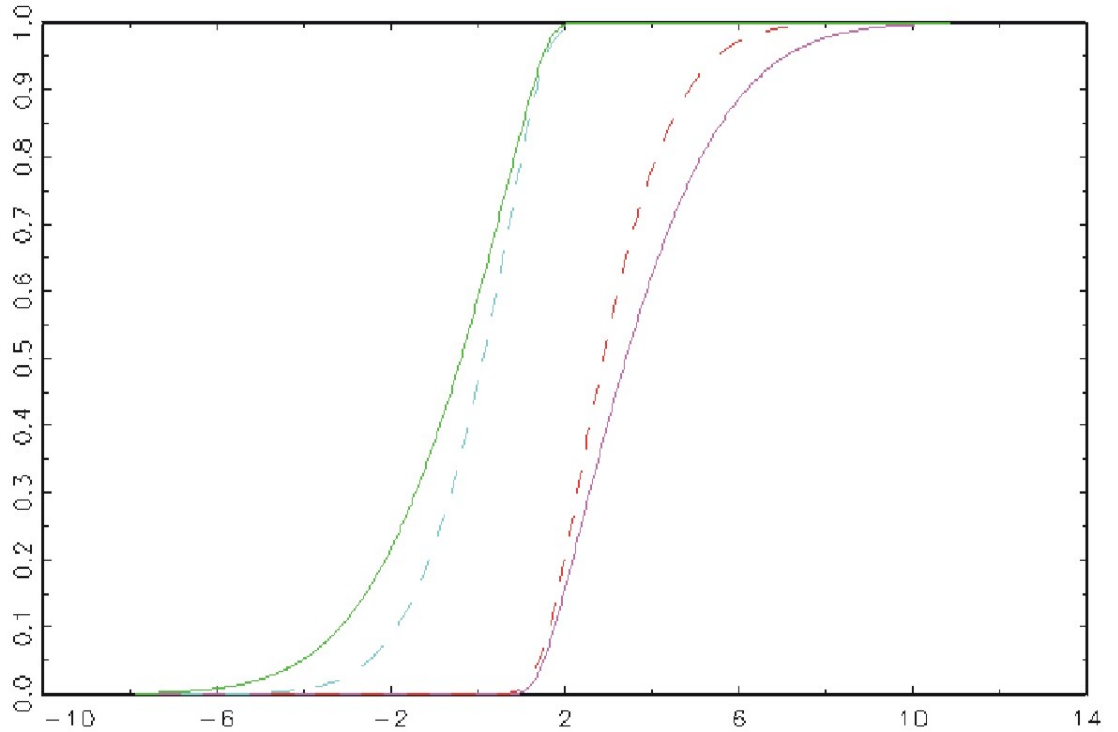
$$\mathbb{E}[G_{MU}(d|X)] = \mathbb{E}\left[2I(d \leq \theta(X))\Phi\left(\frac{d - \theta(X)}{2\sigma}\right) + I(d > \theta(X))\right]$$

If X is itself normally distributed then the unconditional outcome distributions are normal

$$Y_k \sim N(\alpha_k + \beta_k \mu_X, \beta_k^2 \sigma_X^2 + \sigma^2)$$

The Makarov bounds for normal outcome distributions with different variances have an explicit expression that is given in Appendix A. In Figure 3 we plot the average bounds (dashed line) and the bounds for the average (solid line) population for $\alpha_0 = 0, \alpha_1 = 1, \beta_0 = 1, \beta_1 = 1.5, \sigma = 1$. The mean and standard deviation of X 1 and .8, respectively. The implied R^2 in the two outcome distributions are .39 (control) and .59 (treatment). Note that the average bounds show that less than half of the population has a negative treatment effect, but that the bounds on the average outcome distributions do not allow such a conclusion.

Fig 3. Average Makarov bounds and Makarov bounds for the average population: conditional Normal outcome distributions with Normally distributed covariate.



6 Bounds on the distribution function of treatment effects in two points

6.1 A necessary condition for being compatible with a treatment effect c.d.f. in two points

Because the Makarov bounds are pointwise, but not uniformly sharp, the region that these bounds imply for the vector of values of the treatment effect c.d.f. in a vector of points is not necessarily best possible. Let $d_1 < \dots < d_K$ be K ordered real numbers. We are interested in obtaining bounds on the set of K -vectors $\mathcal{B}(d_1, \dots, d_K) = \{((G(d_1) \cdots G(d_K))', G \in \mathcal{G})\}$ with as before \mathcal{G} the set of c.d.f. of treatment effect distributions for given (conditional) outcome distributions. To keep the notation simple we consider unconditional outcome distributions and the case $K = 2$. Because $G(d_1) \leq G(d_2)$ and both $G(d_1)$ and

$G(d_2)$ are within the Makarov bounds we have that

$$\mathcal{B}(d_1, d_2) \subseteq \mathcal{M}(d_1, d_2) = \{(G(d_1), G(d_2)) | G_{ML}(d_1) \leq G(d_1) \leq G_{MU}(d_1), G_{ML}(d_2) \leq G(d_2) \leq G_{MU}(d_2), G(d_1) \leq G(d_2)\}$$

The set $\mathcal{M}(d_1, d_2)$ is drawn in the bottom panel of Figure 5. It is the region bounded by the extreme points A, B, C, D, E . For obvious reasons we call $\mathcal{M}(d_1, d_2)$ the two-dimensional Makarov bounds on $G(d_1), G(d_2)$. The analysis is somewhat different for the case that $d_1 < d_2$ are ‘close’ in the sense that $G_{MU}(d_1) \geq G_{ML}(d_2)$. If d_1, d_2 are not close in this sense, the two-dimensional Makarov bounds are a rectangle, because the monotonicity restriction is not binding. Because we are interested in functionals of the treatment effect c.d.f. that can be approximated by the value of that functional in a finite (but possibly large) number of points on the support of the treatment effect c.d.f. the case that $G_{MU}(d_k) \geq G_{ML}(d_{k+1})$ is the most relevant case.

The two-dimensional Makarov bounds \mathcal{M} on the treatment effect c.d.f. in $d_1 < d_2$ contain \mathcal{B} . The two-dimensional Makarov bounds are sharp if and only if $\mathcal{M} = \mathcal{B}$. Therefore they are not best possible, if we can find points in \mathcal{M} that are not in \mathcal{B} . To establish that a point, e.g. point C in Figure 5 is in \mathcal{B} , we would have to construct a joint c.d.f. of Y_0, Y_1 with given marginal distributions, such that the c.d.f. of $Y_1 - Y_0$, i.e. the supporting c.d.f. G_C , satisfies $G_C(d_1) = G_{MU}(d_1)$ and $G_C(d_2) = G_{MU}(d_2)$. A simpler procedure is to find necessary conditions for the existence of a supporting c.d.f. G_C . If these conditions do not hold in C , then $C \notin \mathcal{B}$. The same is true for all points in \mathcal{M} where the necessary conditions do not hold. Therefore, the set \mathcal{B} is strictly smaller than \mathcal{M} and by eliminating all points where the necessary condition does not hold, we obtain the maximal reduction relative to the necessary condition. We have been unable to show that our necessary condition for membership of \mathcal{B} is also sufficient. So strictly speaking we cannot call our improved bounds sharp.

To derive the necessary condition for $C \in \mathcal{B}$, we note that if $G_C \in \mathcal{G}$, then $G_{ML}(d) \leq G_C(d) \leq G_{MU}(d)$ for all d and the corresponding treatment effect distribution has mean $\mathbb{E}(Y_1) - \mathbb{E}(Y_0)$. In addition, if G_C exists it is larger than the smallest c.d.f. $G_{ML} \leq G_{CK} \leq G_{MU}$ and smaller than the largest c.d.f. $G_{ML} \leq G_{CG} \leq G_{MU}$ with $G_{CK}(d_1) = G_{CG}(d_1) = G_{MU}(d_1)$ and $G_{CK}(d_2) = G_{CG}(d_2) = G_{MU}(d_2)$. A c.d.f. F is smaller than a c.d.f. G if G first-order stochastically dominates F . Of course, this implies that the mean of the distribution of G cannot be smaller than the mean of the distribution of F . Combining these observations we conclude that if G_C exists, then the mean of G_{CK} is not greater than $\mathbb{E}(Y_1) - \mathbb{E}(Y_0)$ and the mean of G_{CG} not smaller than $\mathbb{E}(Y_1) - \mathbb{E}(Y_0)$. If this necessary condition does not hold then $C \notin \mathcal{B}$. We show how to check the necessary condition and find the smallest set in \mathcal{M} where this condition is satisfied.

As a first step in the derivation of the necessary condition we derive the stochastically smallest distribution G_{d_0K} that is within the Makarov bounds and passes through $G_{ML}(d_0)$ and the stochastically largest distribution G_{d_0G} that

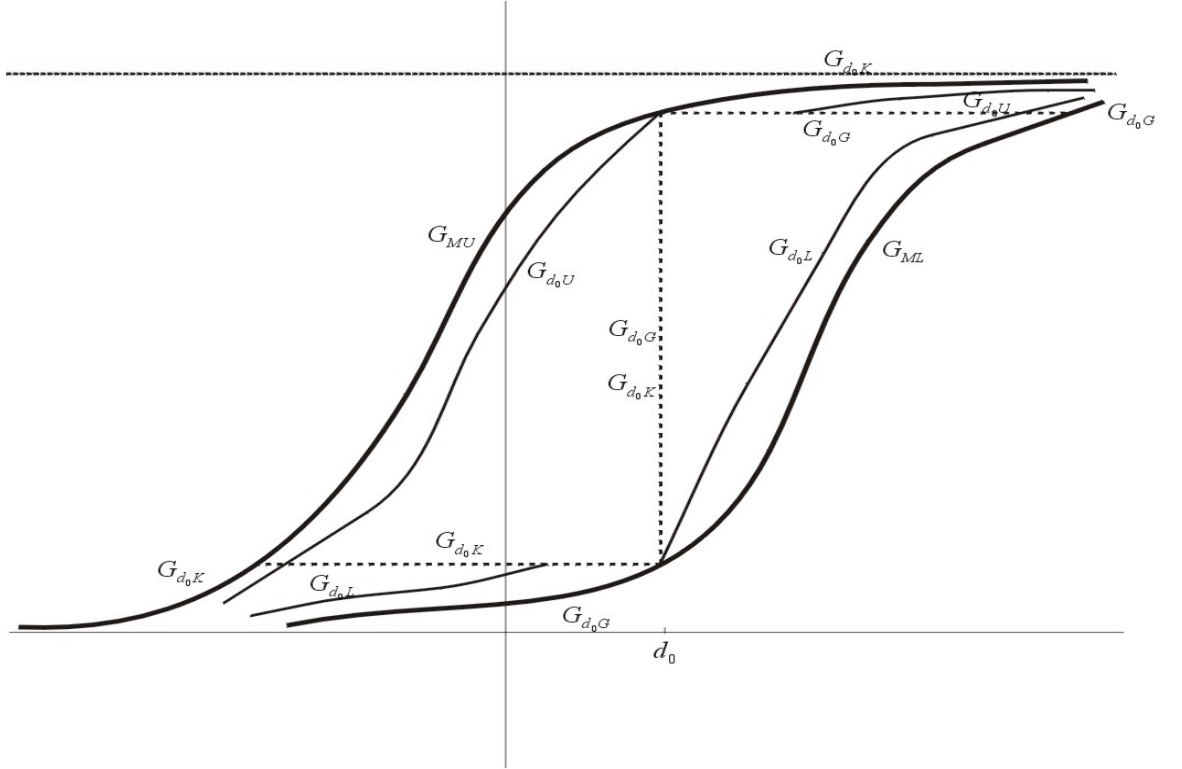
is within the Makarov bounds and passes through $G_{MU}(d_0)$. The construction is illustrated in Figure 4.

$$\begin{aligned}
G_{d_0K}(d) &= G_{MU}(d) & d < G_{MU}^{-1}(G_{ML}(d_0)) \\
&= G_{ML}(d_0) & G_{MU}^{-1}(G_{ML}(d_0)) \leq d < d_0 \\
&= G_{MU}(d) & d \geq d_0
\end{aligned} \tag{7}$$

$$\begin{aligned}
G_{d_0G}(d) &= G_{ML}(d) & d < d_0 \\
&= G_{MU}(d_0) & d_0 \leq d < G_{ML}^{-1}(G_{MU}(d_0)) \\
&= G_{ML}(d) & d \geq G_{ML}^{-1}(G_{MU}(d_0))
\end{aligned} \tag{8}$$

Note that $G_{d_0K}(d_0) = G_{MU}(d_0) > G_{ML}(d_0) = G_{d_0K}(d_0)_-$. However, G_{d_0K} is

Fig 4. The smallest c.d.f. G_{d_0K} through the lower Makarov bound, the largest c.d.f. G_{d_0G} through the upper Makarov bound and the c.d.f. G_{d_0L} , G_{d_0U} that support the bounds.



smaller than all c.d.f. that have $G_{d_0K'}(d_0) = G_{ML}(d_0)$ and calling G_{d_0K} the smallest c.d.f. that passes through $G_{ML}(d_0)$ is appropriate.

Because G_{d_0K} is the smallest c.d.f. within the Makarov bounds that passes through $G_{ML}(d_0)$, it is first-order stochastically dominated by the c.d.f. G_{d_0L} that supports the lower bound $G_{ML}(d_0)$. Because this distribution has a mean equal to $\mathbb{E}(Y_1) - \mathbb{E}(Y_0)$ we conclude that the mean of the distribution of G_{d_0K} cannot be larger than $\mathbb{E}(Y_1) - \mathbb{E}(Y_0)$. In the same way the mean of the distribution of G_{d_0G} cannot be smaller than $\mathbb{E}(Y_1) - \mathbb{E}(Y_0)$. Therefore the necessary condition is met for the Makarov bounds in a single point, because the largest c.d.f. corresponding to $G_{ML}(d_0)$ and the smallest corresponding to $G_{MU}(d_0)$ are the Makarov bounds G_{ML} and G_{MU} . Note that G_{d_0K} and G_{d_0G} are mixed discrete-continuous distributions with a support that is the union of two disjoint sets and an atom in d_0 .

Stochastically smallest and largest c.d.f. that are within the Makarov bounds and pass through a particular point can also be constructed in the two-dimensional case.

Lemma 6.1 *Let $d_1 < d_2$ be such that $G_{MU}(d_1) \geq G_{ML}(d_2)$. The mean of the smallest c.d.f. that passes through B,C,D, and E and is within the Makarov bounds is smaller than or equal to $\mathbb{E}(Y_1) - \mathbb{E}(Y_0)$. The mean of the largest c.d.f. that passes through A,B,D and E and is within the Makarov bounds is larger than or equal to $\mathbb{E}(Y_1) - \mathbb{E}(Y_0)$.*

Proof: See Appendix B.

Consider a point in $\mathcal{B}(d_1, d_2)$ which is equal to $G(d_1), G(d_2)$ for some $G \in \mathcal{G}$. The c.d.f. G first-order stochastically dominates the smallest c.d.f. that passes through $G(d_1)$ and $G(d_2)$ and is within the Makarov bounds, $G_{d_1d_2K}$, and it is first-order stochastically dominated by the largest c.d.f. that passes through $G(d_1)$ and $G(d_2)$ and is within the Makarov bounds, $G_{d_1d_2G}$. Hence a necessary condition for $(G(d_1), G(d_2)) \in \mathcal{B}(d_1, d_2)$ is that $G_{d_1d_2K}$ has a mean that does not exceed the ATE and $G_{d_1d_2G}$ has a mean that is not smaller than the ATE.

Theorem 6.1 *If $(G(d_1), G(d_2)) \in \mathcal{B}(d_1, d_2)$ for some $G \in \mathcal{G}$, then the mean of the distribution with c.d.f. $G_{d_1d_2K}$ is less than or equal to $\mathbb{E}(Y_1) - \mathbb{E}(Y_0)$ and the mean of the distribution with c.d.f. $G_{d_1d_2G}$ is greater than or equal to $\mathbb{E}(Y_1) - \mathbb{E}(Y_0)$.*

Lemma 6.1 implies that B,D, and E are in $\mathcal{B}(d_1, d_2)$. However, it is not obvious that A and C are in this set. To decide this we construct the smallest c.d.f. that passes through A (see Figure 5). We only need to consider the smallest c.d.f. because the largest c.d.f. that passes through A is equal to the lower Makarov bound and has a mean that is larger than or equal to the ATE. If $d_1 < d_2$ are close so that $G_{MU}(d_1) \geq G_{ML}(d_2)$, the smallest c.d.f. that passes through A is

$$\begin{aligned}
G_{AK}(d) &= G_{MU}(d) & d < G_{MU}^{-1}(G_{ML}(d_1)) \\
&= G_{ML}(d_1) & G_{MU}^{-1}(G_{ML}(d_1)) \leq d < d_1 \\
&= G_{ML}(d_2) & d_1 \leq d < d_2 \\
&= G_{MU}(d) & d \geq d_2
\end{aligned} \tag{9}$$

We show that this c.d.f. can have a mean that is larger than the ATE and in that case $A \notin \mathcal{B}(d_1, d_2)$. For C the smallest c.d.f. that passes through this point and is within the Makarov bounds is the c.d.f. of the upper Makarov bound with a mean that is smaller than or equal to the ATE. The largest c.d.f. within the Makarov bounds that passes through C is (if $d_1 < d_2$ are close as defined above)

$$\begin{aligned}
G_{CG}(d) &= G_{ML}(d) & d < d_1 \\
&= G_{MU}(d_1) & d_1 \leq d < d_2 \\
&= G_{MU}(d_2) & d_2 \leq d < G_{ML}^{-1}(G_{MU}(d_2)) \\
&= G_{ML}(d) & d \geq G_{ML}^{-1}(G_{MU}(d_2))
\end{aligned} \tag{10}$$

and this c.d.f. may have a mean that is less than or equal to the ATE, and in that case $C \notin \mathcal{B}(d_1, d_2)$.

We compute the mean of the distribution in (9) by subdividing the support in the interval $(-\infty, G_{MU}^{-1}(G_{ML}(d_1)))$, the point d_1 , the point d_2 and the interval $[d_2, \infty)$. The distribution corresponding to the c.d.f. assigns positive probability to these points and intervals and zero probability elsewhere. By partial integration we find

$$\begin{aligned}
\mu_{AK} &= \int_{-\infty}^{G_{MU}^{-1}(G_{ML}(d_1))} sdG_{MU}(s) + \int_{d_2}^{\infty} sdG_{MU}(s) + \\
&\quad d_1[G_{ML}(d_2) - G_{ML}(d_1)] + d_2[G_{MU}(d_2) - G_{ML}(d_2)] =
\end{aligned} \tag{11}$$

$$\begin{aligned}
&G_{ML}(d_1)G_{MU}^{-1}(G_{ML}(d_1)) - \int_{-\infty}^{G_{MU}^{-1}(G_{ML}(d_1))} G_{MU}(s)ds + \int_{d_2}^{\infty} (1 - G_{MU}(s))ds \\
&+ d_1[G_{ML}(d_2) - G_{ML}(d_1)] + d_2[1 - G_{ML}(d_2)]
\end{aligned}$$

An analogous argument gives the mean of G_{CG}

$$\begin{aligned}
\mu_{CG} &= \int_{-\infty}^{d_1} sdG_{ML}(s) + \int_{G_{ML}^{-1}(G_{MU}(d_2))}^{\infty} s dg_{ML}(s) + \\
&\quad d_1[G_{MU}(d_1) - G_{ML}(d_1)] + d_2[G_{MU}(d_2) - G_{MU}(d_1)] =
\end{aligned} \tag{12}$$

$$\begin{aligned}
&G_{ML}^{-1}(G_{MU}(d_2))(1 - G_{MU}(d_2)) + \int_{G_{ML}^{-1}(G_{MU}(d_2))}^{\infty} (1 - G_{ML}(s))ds - \int_{-\infty}^{d_1} G_{ML}(s)ds \\
&+ d_1G_{MU}(d_1) + d_2[G_{MU}(d_2) - G_{MU}(d_1)]
\end{aligned}$$

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

Because the density g_{MU} is the density of a halfnormal distribution with endpoint θ , we can use the truncated normal mean formula⁹ to derive

$$\int_{-\infty}^b \text{sd}G_{MU}(s) = \begin{cases} 2\theta\Phi\left(\frac{b-\theta}{2\sigma}\right) - 4\sigma\phi\left(\frac{b-\theta}{2\sigma}\right) & \text{if } b < \theta \\ \theta - 4\sigma\phi(0) & \text{if } b \geq \theta \end{cases}$$

and

$$\int_a^\theta \text{sd}G_{MU}(s) = 2\theta\left(\frac{1}{2} - \Phi\left(\frac{a-\theta}{2\sigma}\right)\right) + 4\sigma\left(\phi\left(\frac{a-\theta}{2\sigma}\right) - \phi(0)\right) \quad \text{if } a < \theta$$

In this example $G_{MU}(d_1) \geq G_{ML}(d_2)$ iff $d_2 \geq d_1 \geq \theta$ or $d_1 \leq d_2 \leq \theta$ or $d_1 < \theta < d_2$ and

$$\frac{d_1}{\sigma} - 2\Phi^{-1}\left(\Phi\left(\frac{d_2-\theta}{2\sigma}\right) - \frac{1}{2}\right) \geq \frac{\theta}{\sigma}$$

This restriction is assumed to hold in the rest of the example.

Upon substitution of the integrals above in (11) we obtain the mean of the smallest distribution that passes through A. If $d_1 < d_2 \leq \theta$, then because $G_{ML}(d_1) = G_{ML}(d_2) = 0$, so that in the truncated mean formula $b = G_{MU}^{-1}(G_{ML}(d_1)) = -\infty$ and $a = d_2$

$$\mu_{AK} = 2\theta\left(\frac{1}{2} - \Phi\left(\frac{d_2-\theta}{2\sigma}\right)\right) + 4\sigma\left(\phi\left(\frac{d_2-\theta}{2\sigma}\right) - \phi(0)\right) + 2d_2\Phi\left(\frac{d_2-\theta}{2\sigma}\right).$$

Thus, because $d_2 - \theta \leq 0$, we have that $\mu_{AK} \leq \theta$ since

$$\mu_{AK} - \theta = 4\sigma\left(\phi\left(\frac{d_2-\theta}{2\sigma}\right) - \phi(0) + \left(\frac{d_2-\theta}{2\sigma}\right) \cdot \Phi\left(\frac{d_2-\theta}{2\sigma}\right)\right) \leq 0.$$

Therefore if $d_1 \leq d_2 < \theta$, then $A \in \mathcal{B}(d_1, d_2)$.

If $\theta \leq d_1 < d_2$, we have $G_{MU}(d_2) = 1$, $b = G_{MU}^{-1}(G_{ML}(d_1)) = \theta + 2\sigma\Phi^{-1}\left(\Phi\left(\frac{d_1-\theta}{2\sigma}\right) - \frac{1}{2}\right) \leq \theta$ and $a = d_2 > \theta$. Thus

$$\begin{aligned} \mu_{AK} &= 2\theta\left[\Phi\left(\frac{d_1-\theta}{2\sigma}\right) - \frac{1}{2}\right] - 4\sigma\phi\left(\Phi^{-1}\left(\Phi\left(\frac{d_1-\theta}{2\sigma}\right) - \frac{1}{2}\right)\right) \\ &\quad + 2d_1\left[\Phi\left(\frac{d_2-\theta}{2\sigma}\right) - \Phi\left(\frac{d_1-\theta}{2\sigma}\right)\right] + 2d_2\left[1 - \Phi\left(\frac{d_2-\theta}{2\sigma}\right)\right] \end{aligned}$$

If for example, $\theta = 1, \sigma = 3$ and $d_1 = 1.5, d_2 = 2.5$, then $\mu_{AK} = 1.3814 > 1 = \theta$ so that $A \notin \mathcal{B}(d_1, d_2)$.

⁹If Y has a normal distribution with mean μ and variance σ^2 , then

$$\mathbb{E}(Y|a \leq Y \leq b) = \mu + \sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

Finally, if $d_1 < \theta < d_2$ then because $G_{ML}(d_1) = 0$, $G_{MU}(d_2) = 1$ and in the truncated mean formula $b = G_{MU}^{-1}(G_{ML}(d_1)) = -\infty$ and $a = d_2 > \theta$ (so that the truncated means are 0)

$$\mu_{AK} = d_1 \left[2\Phi\left(\frac{d_2 - \theta}{2\sigma}\right) - 1 \right] + 2d_2 \left[1 - \Phi\left(\frac{d_2 - \theta}{2\sigma}\right) \right]$$

If, for example, $\theta = 1$, $\sigma = 3$ and $d_1 = -1$ and $d_2 = 2$, then $\mu_{AK} = 1.60290 > 1 = \theta$ so that $A \notin \mathcal{B}(d_1, d_2)$.

The density g_{ML} is the density of halfnormal distribution with support $[\theta, \infty)$ and again using the truncated normal mean formula

$$\int_a^\infty \text{sd}g_{ML}(s) = \begin{cases} \theta + 4\sigma\phi(0) & \text{if } a \leq \theta \\ 2\theta \left(1 - \Phi\left(\frac{a-\theta}{2\sigma}\right)\right) + 4\sigma\phi\left(\frac{a-\theta}{2\sigma}\right) & \text{if } a > \theta \end{cases}$$

and

$$\int_\theta^b \text{sd}g_{ML}(s) = 2\theta \left(\Phi\left(\frac{b-\theta}{2\sigma}\right) - \frac{1}{2} \right) + 4\sigma \left(\phi(0) - \phi\left(\frac{b-\theta}{2\sigma}\right) \right)$$

If we substitute these expressions in (10) we obtain an expression for μ_{CG} . We distinguish between the cases that $d_1 < d_2 \leq \theta$, that $\theta < d_1 \leq d_2$, and that $d_1 < \theta < d_2$. We maintain the restrictions that ensure that $G_{MU}(d_1) \geq G_{ML}(d_2)$.

If $\theta < d_1 \leq d_2$, we have $G_{MU}(d_2) = 1$, $G_{MU}(d_1) = 1$, $a = G_{ML}^{-1}(G_{MU}(d_2)) = \infty$, $b = d_1$, so that

$$\mu_{CG} = 2\theta \left[\Phi\left(\frac{d_1 - \theta}{2\sigma}\right) - \frac{1}{2} \right] + 4\sigma \left[\phi(0) - \phi\left(\frac{d_1 - \theta}{2\sigma}\right) \right] + 2d_1 \left[1 - \Phi\left(\frac{d_1 - \theta}{2\sigma}\right) \right]$$

and therefore

$$\frac{\mu_{CG} - \theta}{2\sigma} = 2 \left[\phi(0) - \phi\left(\frac{d_1 - \theta}{2\sigma}\right) \right] + 2 \left(\frac{d_1 - \theta}{2\sigma} \right) - 2 \left(\frac{d_1 - \theta}{2\sigma} \right) \Phi\left(\frac{d_1 - \theta}{2\sigma}\right) \geq 0$$

Hence if $\theta < d_1 < d_2$, then $C \in \mathcal{B}(d_1, d_2)$.

If $d_1 < d_2 \leq \theta$, we have $G_{ML}(d_1) = 0$ and in the truncated means $a = G_{ML}^{-1}(G_{MU}(d_2)) = \theta + 2\sigma\Phi^{-1}\left(\Phi\left(\frac{d_2 - \theta}{2\sigma}\right) + \frac{1}{2}\right)$, $b = d_1 < \theta$, so that

$$\begin{aligned} \mu_{CG} &= 2\theta \left[\frac{1}{2} - \Phi\left(\frac{d_2 - \theta}{2\sigma}\right) \right] + 4\sigma\phi\left(\Phi^{-1}\left(\Phi\left(\frac{d_2 - \theta}{2\sigma}\right) + \frac{1}{2}\right)\right) \\ &\quad + 2d_2 \left[\Phi\left(\frac{d_2 - \theta}{2\sigma}\right) - \Phi\left(\frac{d_1 - \theta}{2\sigma}\right) \right] + 2d_1\Phi\left(\frac{d_1 - \theta}{2\sigma}\right) \end{aligned}$$

If for example, $\theta = 1$, $\sigma = 3$ and $d_1 = -0.5$, $d_2 = 0.5$, then $\mu_{CG} = 0.6186 < 1 = \theta$ and $C \notin \mathcal{B}(d_1, d_2)$.

Finally, if $d_1 < \theta < d_2$ then $G_{ML}(d_1) = 0$, $G_{MU}(d_2) = 1$, $a = G_{ML}^{-1}(G_{MU}(d_2)) = \infty$, and $b = d_1 < \theta$, so that

$$\mu_{CG} = 2d_1\Phi\left(\frac{d_1 - \theta}{2\sigma}\right) + d_2 \left[1 - 2\Phi\left(\frac{d_1 - \theta}{2\sigma}\right) \right]$$

If, for example, $\theta = 1$, $\sigma = 3$ and $d_1 = -1$, $d_2 = 2$, then $\mu_{CG} = -0.2166 < 1 = \theta$ and $C \notin \mathcal{B}(d_1, d_2)$. \square

6.2 More informative bounds on the treatment effect c.d.f. in two points

The example shows that μ_{AK} can be larger and μ_{CG} can be smaller than the ATE so that either A or C (or both) are not in $\mathcal{B}(d_1, d_2)$. By continuity, if e.g. $A \notin \mathcal{B}(d_1, d_2)$, then the points in a neighborhood of A are also not in that set. We will determine the (largest) subset of $\mathcal{M}(d_1, d_2)$ that is not in $\mathcal{B}(d_1, d_2)$. That subset is drawn in Figure 5, i.e. the region bounded by A, F and G . If $C \notin \mathcal{B}(d_1, d_2)$, then the largest subset of $\mathcal{M}(d_1, d_2)$ that is not in $\mathcal{B}(d_1, d_2)$ is bounded by I, H , and C in Figure 6.

Theorem 6.2 *If the smallest c.d.f. $G_{ML} \leq G_{AK} \leq G_{MU}$ that passes through $G_{ML}(d_1)$ and $G_{ML}(d_2)$ has a mean $\mu_{AK} > \mathbb{E}(Y_1) - \mathbb{E}(Y_0)$ then all points in $\mathcal{M}(d_1, d_2)$ below the convex curve $G_2 = P(G_1)$ defined by*

$$\begin{aligned} \mathbb{E}(Y_1) - \mathbb{E}(Y_0) &= \int_{-\infty}^{G_{MU}^{-1}(G_1)} sg_{MU}(s) ds + d_1[\min\{G_2, G_{MU}(d_1)\} - G_1] \quad (13) \\ &\quad + \int_{d_1}^{G_{MU}^{-1}(G_2)} sg_{MU}(s) ds + d_2[G_{MU}(d_2) - G_2] + \int_{d_2}^{\infty} sg_{MU}(s) ds \\ &= G_{MU}^{-1}(G_1)G_1 - \int_{-\infty}^{G_{MU}^{-1}(G_1)} G_{MU}(s) ds \\ +1(G_2 > G_{MU}(d_1)) &\left[G_{MU}^{-1}(G_2) - d_1 G_{MU}(d_1) - \int_{d_1}^{G_{MU}^{-1}(G_2)} G_{MU}(s) ds \right] \\ &\quad + d_1[\min\{G_2, G_{MU}(d_1)\} - G_1] + d_2[1 - G_2] + \int_{d_2}^{\infty} (1 - G_{MU}(s)) ds \end{aligned}$$

(where we adopt the convention that an integral is 0 if the upper integration limit is smaller than the lower integration limit and $1(\cdot)$ is the indicator function) are not in $\mathcal{B}(d_1, d_2)$.

If the largest c.d.f. $G_{ML} \leq G_{CG} \leq G_{MU}$ that passes through $G_{MU}(d_1)$ and $G_{MU}(d_2)$ has a mean $\mu_{CG} < \mathbb{E}(Y_1) - \mathbb{E}(Y_0)$ then all points in $\mathcal{M}(d_1, d_2)$ above the concave curve $H_2 = Q(H_1)$ defined by

$$\begin{aligned} \mathbb{E}(Y_1) - \mathbb{E}(Y_0) &= \int_{-\infty}^{d_1} sg_{ML}(s) ds + d_1[H_1 - G_{ML}(d_1)] + \int_{G_{ML}^{-1}(H_1)}^{d_2} sg_{ML}(s) ds + \\ &\quad d_2[H_2 - \max\{H_1, G_{ML}(d_2)\}] + \int_{G_{ML}^{-1}(H_2)}^{\infty} sg_{ML}(s) ds = \\ - \int_{-\infty}^{d_1} G_{ML}(s) ds &+ d_1 H_1 + 1(G_{ML}(d_2) > H_1) \left[d_2 G_{ML}(d_2) - G_{ML}^{-1}(H_1) H_1 - \int_{G_{ML}^{-1}(H_1)}^{d_2} G_{ML}(s) ds \right] + \\ d_2 [H_2 - \max\{H_1, G_{ML}(d_2)\}] &+ G_{ML}^{-1}(H_2) (1 - H_2) + \int_{G_{ML}^{-1}(H_2)}^{\infty} (1 - G_{ML}(s)) ds \end{aligned} \quad (14)$$

are not in $\mathcal{B}(d_1, d_2)$. The set $\mathcal{C}(d_1, d_2)$ bounded by $\mathcal{M}(d_1, d_2)$ and the curves (13) and (14) is convex.

Proof: See Appendix B.

If $G_2 \leq G_{MU}(d_1)$ the curve P has an explicit expression

$$P(G_1) = \frac{-(\mathbb{E}(Y_1) - \mathbb{E}(Y_0)) + d_2 + \int_{d_2}^{\infty} (1 - G_{MU}(s)) ds - d_1 G_1 + G_{MU}^{-1}(G_1) G_1 - \int_{-\infty}^{G_{MU}^{-1}(G_1)} G_{MU}(s) ds}{d_2 - d_1}$$

and the same is true for Q if $H_1 \geq G_{ML}(d_2)$

$$Q(H_1) = \frac{-(\mathbb{E}(Y_1) - \mathbb{E}(Y_0)) + d_2 H_2 - \int_{G_{ML}^{-1}(H_2)}^{\infty} (1 - G_{ML}(s)) ds + G_{ML}^{-1}(H_2) (1 - H_2) - \int_{-\infty}^{d_1} G_{ML}(s) ds}{d_2 - d_1}$$

Theorem 6.2 defines a subset $\mathcal{C}(d_1, d_2)$ of $\mathcal{M}(d_1, d_2)$ that contains $\mathcal{B}(d_1, d_2)$. If the mean of the smallest c.d.f. that passes through A is larger than the ATE and/or the largest c.d.f. that passes through C is smaller than the ATE, then $\mathcal{C}(d_1, d_2)$ is a strict subset of $\mathcal{M}(d_1, d_2)$ and we have bounds that are more informative than the two-dimensional Makarov bounds.

It follows directly from the construction that $G_{\tilde{F}K}(d_1) \leq G_{\tilde{H}G}(d_2)$ so that in Figures 5 and 6 G is below I . This implies that the projection of $\mathcal{C}(d_1, d_2)$ are the original Makarov bounds in d_1 and d_2 , respectively. In other words, although $\mathcal{C}(d_1, d_2)$ may be smaller than $\mathcal{M}(d_1, d_2)$, the projections are equal to the Makarov bounds in a single point.

All results until now hold also for the conditional (on X) bounds. We now show that the specific shape of the improved bounds implies that averaging over X makes them more informative. By Theorem 6.2 $\mathcal{C}(d_1, d_2)$ is bounded by the one-dimensional Makarov bounds (vertical and horizontal bounds), the curves (13) and (14), and the 45 degree line. If the bounds are obtained from conditional outcome distributions, then it follows from Theorem 5.1 that the horizontal and vertical lower bounds cannot decrease if we average, that the horizontal and vertical upper bounds cannot increase if we average. Finally, by Jensen's inequality the convex curve (13) cannot decrease and the concave curve (14) cannot increase if we average. Together with the observation that the 45 degree line is unaffected by averaging, we have

Theorem 6.3 *Let $\mathcal{C}(d_1, d_2)(X)$ be the convex set defined in Theorem 6.2 as derived from the conditional outcome distributions, then $\mathbb{E}[\mathcal{C}(d_1, d_2)(X)]$ cannot be larger than $\mathcal{C}(d_1, d_2)$ that is derived from the unconditional outcome distributions.*

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

We found that for $\theta = 1, \sigma = 3$ and $d_1 = 1.5, d_2 = 2.5$ $A \notin \mathcal{B}(d_1, d_2)$. For these values $\mu_{CG} = 1.4834 > 1$ so that $C \in \mathcal{B}(d_1, d_2)$. Therefore we only have a more informative lower bound. This bound is drawn in Figure 7.

For $d_1 = -0.5, d_2 = 0.5$, we found that $C \notin \mathcal{B}(d_1, d_2)$. However, $\mu_{AK} = .5166$ so that $A \in \mathcal{B}(d_1, d_2)$ and we only have a more informative upper bound that is drawn in Figure 8.

Finally, for $d_1 = -1$ and $d_2 = 2$, $A \notin \mathcal{B}(d_1, d_2)$ and $C \notin \mathcal{B}(d_1, d_2)$. Both the lower and upper Makarov bound can be improved and the more informative bounds are in Figure 9. \square

7 Bounds on functions of the distribution of treatment effects

The bounds $\mathcal{C}(d_1, d_2)$ on the c.d.f. of the treatment effect distribution in d_1 and d_2 imply bounds on functions of the treatment effect c.d.f. in $d_1 < d_2$. Here we consider linear functions of $G(d_1), G(d_2)$

$$B(G(d_1), G(d_2)) = b_1 G(d_1) + b_2 G(d_2)$$

If $b_1 = 1, b_2 = -1$ this function is equal to the interval probability $G(d_2) - G(d_1)$. Another parameter that can be approximated by B is the total net gain for those individuals whose net gain is between 0 and C .

$$\int_0^C (1 - G(s)) ds$$

If we divide the integration region in two intervals $[0, c)$ and $[c, C]$, then an approximation is

$$C - cG(c/2) - (C - c)G((c + C)/2)$$

If we pick $d_1 = c/2, d_2 = (c + C)/2, b_1 = c, b_2 = C - c$ we obtain bounds on the total gain from bounds on B . In the sequel we can, without loss of generality, assume that $b_2 > 0$.

Manski (1997b), (2003) introduces the concept of a D parameter which is some increasing functional of a c.d.f. where the c.d.f. are ordered according to first-order stochastic dominance. The linear functional that we consider is a D parameter iff $b_1, b_2 \geq 0$. An interval probability, and in general the linear functional with $b_1 \geq 0$ and $b_2 < 0$, is not a D parameter, but it can be expressed a difference of D parameters. Manski derives bounds for D parameters and differences of D parameters. These bounds are different from ours, because he assumes that outcomes are weakly increasing in the level of treatment. We do not make his assumption, in particular we do not assume that everybody benefits from the treatment.

In Figure 10 we draw the set $\mathcal{C}(d_1, d_2)$. In the sequel we use the notation $G_1 \equiv G(d_1)$ and $G_2 \equiv G(d_2)$. The bounds on $B(G_1, G_2)$ depend on whether $\mu_{AK} \gtrless \theta$ and $\mu_{CG} \gtrless \theta$. $B(G_1, G_2)$ is minimal in D, E or A if $\mu_{AK} \leq \theta$ and in D, E, a point at which B touches $P(G_1)$, F or G if $\mu_{AK} > \theta$. The latter is a direct consequence of the convexity of $P(G_1)$. If $\mu_{AK} \leq \theta$ the point at which $B(G_1, G_2)$ is minimal is determined by the slope of $B(G_1, G_2)$, i.e. $-\frac{b_1}{b_2}$. If

$\mu_{AK} > \theta$, then the point at which $B(G_1, G_2)$ is minimal is determined by the slope of $B(G_1, G_2)$ and the slope of $P(G_1)$ in F and G. The upper bound of $B(G_1, G_2)$ is determined in a similar way. Therefore we define

$$P'_F \equiv P'(P^{-1}(G_{ML}(d_2)))$$

$$P'_G \equiv P'(G_{ML}(d_1))$$

$$Q'_H \equiv Q'(G_{MU}(d_1))$$

$$Q'_I \equiv Q'(Q^{-1}(G_{MU}(d_2)))$$

The bounds on $B(G_1, G_2)$ that we denote by $B_L \leq B_U$ are given in the following theorem.

Theorem 7.1 *If $\mu_{AK} \leq \theta$, then the lower bound on $B(G_1, G_2)$ is*

$$\begin{aligned} B_L &= b_1 G_{MU}(d_1) + b_2 G_{MU}(d_1) && \text{if } b_1 < -b_2 \\ &= b_1 G_{ML}(d_2) + b_2 G_{ML}(d_2) && \text{if } -b_2 \leq b_1 < 0 \\ &= b_1 G_{ML}(d_1) + b_2 G_{ML}(d_2) && \text{if } b_1 > 0 \end{aligned}$$

If $\mu_{AK} > \theta$, then the lower bound on $B(G_1, G_2)$ is

$$\begin{aligned} B_L &= b_1 G_{MU}(d_1) + b_2 G_{MU}(d_1) && \text{if } b_1 < -b_2 \\ &= b_1 G_{ML}(d_2) + b_2 G_{ML}(d_2) && \text{if } -b_2 \leq b_1 < 0 \\ &= b_1 P^{-1}(G_{ML}(d_2)) + b_2 G_{ML}(d_2) && \text{if } 0 < b_1 \leq -b_2 P'_F \\ &= b_1 \tilde{G}_1 + b_2 P(\tilde{G}_1) && \text{if } -b_2 P'_F < b_1 < -b_2 P'_G \\ &= b_1 G_{ML}(d_1) + b_2 P(G_{ML}(d_1)) && \text{if } b_1 \geq -b_2 P'_G \end{aligned}$$

where \tilde{G}_1 is the unique solution to

$$P'(\tilde{G}_1) = -\frac{b_1}{b_2}$$

If $\mu_{CG} \geq \theta$ then the upper bound is

$$\begin{aligned} B_U &= b_1 G_{ML}(d_1) + b_2 G_{MU}(d_2) && \text{if } b_1 < 0 \\ &= b_1 G_{MU}(d_1) + b_2 G_{MU}(d_2) && \text{if } b_1 > 0 \end{aligned}$$

and if $\mu_{CG} < \theta$

$$\begin{aligned} B_U &= b_1 G_{ML}(d_1) + b_2 G_{MU}(d_2) && \text{if } b_1 < 0 \\ &= b_1 Q^{-1}(G_{MU}(d_2)) + b_2 G_{MU}(d_2) && \text{if } 0 < b_1 < -b_2 Q'_H \\ &= b_1 \tilde{H}_1 + b_2 Q(\tilde{H}_1) && \text{if } -b_2 Q'_H < b_1 \leq -b_2 Q'_I \\ &= b_1 G_{MU}(d_1) + b_2 Q(G_{MU}(d_1)) && \text{if } b_1 \geq -b_2 Q'_I \end{aligned}$$

where \tilde{H}_1 is the unique solution to

$$Q'(\tilde{H}_1) = -\frac{b_1}{b_2}$$

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

We consider bounds on the functions

$$B_1(G_1, G_2) = G_2 - G_1$$

and

$$B_2(G_1, G_2) = G_1 + G_2$$

If $\theta = 1, \sigma = 3$ and $d_1 = 1.5, d_2 = 2.5$ the bound in Figure 7 implies that

$$0 \leq B_1(G_1, G_2) \leq .934$$

with no improvement over the Makarov bounds. For $B_2(G_1, G_2)$

$$.298 \leq B_2(G_1, G_2) \leq 2$$

and this improves on the Makarov bounds that are

$$.263 \leq B_2(G_1, G_2) \leq 2$$

For $d_1 = -0.5, d_2 = 0.5$ we obtain from the bound that is drawn in Figure 8

$$0 \leq B_1(G_1, G_2) \leq .934$$

with no improvement and

$$0 \leq B_2(G_1, G_2) \leq 1.703$$

where the upper bound improves on the Makarov bound that is 1.737. Finally, for $d_1 = -1$ and $d_2 = 2$ the bound are in Figure 9 gives

$$0 \leq B_1(G_1, G_2) \leq 1$$

which is noninformative and

$$.617 \leq B_2(G_1, G_2) \leq 1.630$$

which improves considerably on the Makarov bounds

$$.132 \leq B_2(G_1, G_2) \leq 1.739$$

□

8 Conclusion

If a function is not non-parametrically identified we may be able to bound the set to which it belongs. Bounds on sets of functions can be best possible just as bounds on sets of finite dimensional parameters. If we can establish that these bounds are best possible, it may be that the bounds are pointwise or uniformly

sharp with the latter implying the former, but the former not implying the latter. Uniformly sharp bounds are members of the set that is being bounded. Pointwise sharp bounds share some of the properties of the set, but not all. This fact implies that K dimensional bounds on the value of the function in K points may not be best possible. We consider bounds on the set of treatment effect c.d.f. with given marginal outcome distributions. The Makarov bounds on this set are pointwise sharp but in general¹⁰ not uniformly sharp, because their mean is in general not equal to the Average Treatment Effect. We have shown that this allows us to narrow the higher dimensional Makarov bounds. Because the set bounded by the improved bounds is convex, it is straightforward to use these bounds obtain bounds on linear functionals. In some cases the improved higher dimensional bounds narrow the bounds on the functionals substantially. We give explicit expressions for the bounds on the set and on linear functionals for $K = 2$. These expressions can be generalized to arbitrary K . Moreover, because the set is convex, averaging over covariates that are correlated with the outcomes will narrow the bounds even further.

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.

¹⁰They are uniformly sharp if the outcomes are dichotomous.

DJEBBARI, H, AND J. SMITH, (2008), “Heterogeneous Impacts in PROGRESA,” *Journal of Econometrics*, 145, 64-80.

EASTERLIN, R., (2008), “Lost in Transition: Life Satisfaction on the Road to Capitalism,” Working Paper, No. 3409, IZA, Bonn

FAN, Y., S. PARK, (2007), “Sharp Bounds on the Distribution of the Treatment Effect and Their Statistical Inference,” typescript, Department of Economics, Vanderbilt University.

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 Society 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), “Masstabinvariante Korrelations-Theorie,” *Schriftenreihe Math. Inst. Univ. Berlin*, 5, 181–233.

IMBENS, G., AND D. RUBIN, (1997), “Estimating Outcome Distributions for Compliers in Instrumental Variables Models,” *Review of Economic Studies*, 64, 555–574.

KREINOVICH, V. AND S. FERSON, (2005), “Computing Best-Possible Bounds for the Distribution of a Sum of Several Variables is NP-Hard,” Working paper, Department of Computer Science, University of Texas at El Paso.

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

MANSKI, C., (1997b), "Monotone Treatment Response," *Econometrica*, 65, 1311-1334.

MANSKI, C., (2003), *Partial Identification of Probability Distributions* (Springer Series in Statistics), Springer-Verlag, New York.

RABIN, M., (1998), "Psychology and Economics," *Journal of Economic Literature* 36, 11-46.

TVERSKY, A., AND D. KAHNEMAN, (1991), "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics* 106, pp.1039-61.

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

A Makarov bounds on the treatment effect distribution if the marginal outcome distributions are normal with unequal variances

If $Y_k \sim N(\mu_k, \sigma_k^2)$, $k = 0, 1$, then with $\theta = \mu_1 - \mu_0$

$$\begin{aligned}
 G_{ML}(d) &= \Phi \left(\frac{-\sigma_1(d - \theta) + \sigma_0 \sqrt{(d - \theta)^2 + 2(\sigma_0^2 - \sigma_1^2) \ln \frac{\sigma_0}{\sigma_1}}}{\sigma_0^2 - \sigma_1^2} \right) - \\
 &\quad \Phi \left(\frac{-\sigma_0(d - \theta) + \sigma_1 \sqrt{(d - \theta)^2 + 2(\sigma_0^2 - \sigma_1^2) \ln \frac{\sigma_0}{\sigma_1}}}{\sigma_0^2 - \sigma_1^2} \right) \\
 G_{MU}(d) &= \Phi \left(\frac{-\sigma_1(d - \theta) - \sigma_0 \sqrt{(d - \theta)^2 + 2(\sigma_0^2 - \sigma_1^2) \ln \frac{\sigma_0}{\sigma_1}}}{\sigma_0^2 - \sigma_1^2} \right) - \\
 &\quad \Phi \left(\frac{-\sigma_0(d - \theta) - \sigma_1 \sqrt{(d - \theta)^2 + 2(\sigma_0^2 - \sigma_1^2) \ln \frac{\sigma_0}{\sigma_1}}}{\sigma_0^2 - \sigma_1^2} \right) + 1
 \end{aligned}$$

B Proofs

Proof of Theorem 3.1:

First consider the lower bound. We have for all v, u with $v + u = d$ and using the Bonferroni inequality

$$\begin{aligned}
 G(d|x) &= \Pr(Y_1 + (-Y_0) \leq d | X = x) \geq \Pr(Y_1 \leq u, -Y_0 \leq v | X = x) \geq \\
 &\max\{\Pr(Y_1 \leq u | X = x) + \Pr(-Y_0 \leq v | X = x) - 1, 0\} = \max\{F_1(u|x) - F_0(-v|x)_-, 0\}
 \end{aligned}$$

Hence if we define $t \equiv u, d \equiv u + v$

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

For the upper bound we have

$$\begin{aligned}
 1 - G(d|x) &= \Pr(Y_1 + (-Y_0) > d | X = x) \geq \Pr(Y_1 > u, -Y_0 > v | X = x) \geq \\
 &\max\{\Pr(Y_1 > u | X = x) + \Pr(-Y_0 > v | X = x) - 1, 0\}
 \end{aligned}$$

Taking the opposite on both sides of the equation, adding 1, substituting $t \equiv u, d \equiv u + v$, and taking the expectation gives

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

We show that the bounds are themselves c.d.f. Consider the lower bound for $G(d|x)$

$$G_{ML}(d|x) = \sup_t \max\{F_1(t|x) - F_0(t-d|x)_-, 0\}$$

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

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

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

$$F_0(t-d_n|x) \geq F_0(t-d|x)_- - \delta$$

because $t-d_n \uparrow t-d$ and $F_0(\cdot)_-$ is the left-hand limit. Using this inequality we have for all t

$$F_1(t|x) - F_0(t-d|x)_- \leq F_1(t|x) - F_0(t-d_n|x)_- \leq F_1(t|x) - F_0((t-d)|x)_- + \delta$$

Taking the sup over t from right to left we obtain

$$G_{ML}(d|x) \leq G_{ML}(d_n|x) \leq G_{ML}(d|x) + \delta$$

Taking the limit we obtain, because δ is arbitrary, that $\lim_{n \rightarrow \infty} G_{ML}(d_n|x) = G_{ML}(d|x)$, so that the lower bound is right-continuous. Note that

$$G_{ML}(d|x) \geq F_1(d/2|x) - F_0(-d/2|x)$$

so that $\lim_{d \rightarrow \infty} G_{ML}(d|x) = 1$. Taking the expectation over X we conclude that the lower bound is indeed a c.d.f. (by dominated convergence limits and expectations can be interchanged). The proof that the upper bound is also a c.d.f. is analogous. \square

Proof of Theorem 3.2:

Let G_L be decreasing in d_0 , so that for some $d' < d_0$ $G_L(d') > G_L(d_0)$. The supporting c.d.f. G_{d_0L} is such that $G_{d_0L}(d_0) = G_L(d_0)$. Therefore $G_{d_0L}(d') < G_L(d')$ which implies that $G_{d_0L} \notin \mathcal{G}$. In the same way we show that the lower bound is 0 and 1 at $-\infty$ and ∞ , respectively. If G_L is discontinuous at d_0 and not right-continuous, then $G_L(d_0) < G_L(d_0)_+$. The supporting c.d.f. G_{d_0L} satisfies $G_{d_0L}(d_0) = G_L(d_0)$ and because $G_{d_0L} \in \mathcal{G}$ also $G_{d_0L}(d) \geq G_L(d_0)_+$ for $d > d_0$. Therefore the supporting c.d.f. is not right-continuous in d_0 , a contradiction. We prove in the same way that G_U is a c.d.f. \square

Proof of Theorem 5.1:

For all $x \in \mathcal{X}$ and all $s \in \mathfrak{R}$

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

Hence for all $x \in \mathcal{X}$ and all $s \in \mathfrak{R}$

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

and

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

Averaging over the distribution of X gives that for all $s \in \mathcal{R}$

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

and

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

Hence for all $s \in \mathcal{R}$

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

so that we obtain (5), because by dominated convergence $\mathbb{E}[F_0(s - d|X)_-] = \mathbb{E}[F_0(s - d|X)]_-$. The proof of inequality (6) is analogous. \square

Proof of Lemma 6.1:

We only prove the first part of the lemma. The proof of the second part is analogous. First, consider C. The top panel of Figure 5 draws the smallest c.d.f. that is within the Makarov bounds and passes through C (labeled by C). Note that it is just G_{MU} that has a mean that cannot be larger than the ATE. Next, consider B. The smallest c.d.f., labeled by B, is the smallest c.d.f. that passes through the lower bound $G_{ML}(d_1)$ on $G(d_1)$. Because this corresponds to the construction for the one-dimensional case as in (7), such a c.d.f. has a mean that does not exceed the ATE. The smallest c.d.f. that passes through E is the smallest c.d.f. that passes through the lower bound $G_{ML}(d_2)$ on $G(d_2)$ and therefore it is like (7) and has a mean that is not larger than the ATE. Finally, the smallest c.d.f. that passes through D is stochastically dominated by a c.d.f. that is the smallest c.d.f. that is within the Makarov bounds and passes through the lower Makarov bound at $d = G_{ML}^{-1}(G_{MU}(d_1))$, again a c.d.f. as in (7) and has a mean that cannot be larger than the ATE, so that c.d.f. labeled D has a mean that cannot exceed the ATE. \square

Proof of Theorem 6.2:

First we show that the smallest c.d.f. that passes through F and G, G_{FK} and G_{GK} respectively, have a mean equal to the ATE. Moreover F is to the left of E and G is below B. We have $G_{FK}(d_2)_- = G_{ML}(d_2)$ and $G_{ML}(d_1) < G_{FK}(d_1)_- \leq G_{EK}(d_1)$ with $G_{ML} \leq G_{EK} \leq G_{MU}$ the smallest c.d.f. that passes through E. $G_{FK}(d_1)_- > G_{ML}(d_1)$ because the smallest c.d.f. that passes through A has a mean strictly larger than the ATE and $G_{FK}(d_1)_- \leq G_{EK}(d_1)$ because G_{EK} is

the smallest c.d.f. that passes through E and by Lemma 6.1 has a mean that is less than or equal to the ATE. Because the mean of G_{FK} is decreasing in $G_{FK}(d_1)_-$ there is a value $G_{ML}(d_1) < G_{FK}(d_1)_- \leq G_{EK}(d_1)$ such that the mean of G_{FK} is equal to the ATE. For G we have $G_{GK}(d_1)_- = G_{ML}(d_1)$ and $G_{ML}(d_2) < G_{GK}(d_2)_- \leq G_{MU}(d_2)$. Note that if $G_{GK}(d_2)_- = G_{ML}(d_2)$ then $G_{GK} = G_{AK}$ with a mean that is strictly larger than the ATE. If $G_{GK}(d_2)_- = G_{MU}(d_2)$ then $G_{GK} = G_{BK}$ and by Lemma 6.1 G_{BK} has a mean less than or equal to the ATE. Because the mean of G_{GK} is decreasing in $G_{GK}(d_2)_-$ there is a value of $G_{GK}(d_2)_-$ such that the mean of G_{GK} is equal to the ATE. Note that $G_{GK}(d_2)_-$ can be larger than $G_{MU}(d_1)$ and in that case G_{GK} has continuous support on three disjoint sets.

We now find an expression for the curve that connects F and G. The smallest c.d.f. that passes through \bar{F} on the curve is determined by $G_1 \equiv G_{\bar{F}K}(d_1)_-$ and $G_2 \equiv G_{\bar{F}K}(d_2)_-$ with $G_{FK}(d_1)_- \leq G_1 \leq G_{GK}(d_1)_-$ and $G_{FK}(d_2)_- \leq G_2 \leq G_{GK}(d_2)_-$. The curve is the solution to the equation that sets the mean of $G_{\bar{F}K}$ equal to the ATE. The expression for the mean is different if $G_2 > G_{MU}(d_1)$, because there is an additional disjoint interval in the support. The expression is given in (13). The distribution corresponding to $G_{\bar{F}K}$ is continuous up to $G_{MU}^{-1}(G_1)$, has an atom at d_1 , possibly is continuous between d_1 and $G_{MU}^{-1}(G_2)$ if $G_{MU}(d_1) < G_2$, has an atom at d_2 , and finally a continuous part on the interval $[d_2, \infty)$.

The derivative is

$$P'(G_1) = -\frac{G_{MU}^{-1}(G_1) - d_1}{\max\{G_{MU}^{-1}(G_2), d_1\} - d_2} \leq 0$$

because $G_1 \leq G_{MU}(d_1)$ and $G_2 \leq G_{MU}(d_2)$. The second derivative if $G_2 \leq G_{MU}(d_1)$ is

$$P''(G_1) = \frac{1}{(d_2 - d_1)g_{MU}(G_{MU}^{-1}(G_1))} \geq 0$$

and if $G_2 > G_{MU}(d_1)$

$$P''(G_1) = \frac{1}{(d_2 - G_{MU}^{-1}(G_2))g_{MU}(G_{MU}^{-1}(G_1))} + \frac{P'(G_1)^2}{(d_2 - G_{MU}^{-1}(G_2))g_{MU}(G_{MU}^{-1}(G_2))} \geq 0$$

so that the curve is convex. The extreme points F and G in Figure 5 are found by setting $G_2 = G_{ML}(d_2)$ (and solving for G_1) for F, and by setting $G_1 = G_{ML}(d_1)$ (and solving for G_2) for G.

If μ_{CG} strictly smaller than the ATE, then C is not in $\mathcal{B}(d_1, d_2)$. The same reasoning as above shows that we can find c.d.f. marked by H and I in Figure 6 that have a mean equal to the ATE. The c.d.f. corresponding to \tilde{H} on the curve in Figure 6 is characterized by $H_1 \equiv G_{\tilde{H}G}(d_1)$ and $H_2 \equiv G_{\tilde{H}G}(d_2)$. The corresponding distribution is continuous up to d_1 , has an atom at d_1 , if $H_1 < G_{ML}(d_2)$, a continuous part between $G_{ML}^{-1}(H_1)$ and d_2 , an atom at d_2 , and finally a continuous part on $[G_{ML}^{-1}(H_2), \infty)$. Hence the curve as in (14).

The derivative is

$$Q'(H_1) = -\frac{\min\{G_{ML}^{-1}(H_1), d_2\} - d_1}{G_{ML}^{-1}(H_2) - d_2} \leq 0$$

because $H_1 \geq G_{ML}(d_1)$ and $H_2 \geq G_{ML}(d_2)$. The second derivative is if $G_{ML}(d_2) \leq H_1$

$$Q''(H_1) = -\frac{Q'(H_1)^2}{G_{ML}^{-1}(H_2)g_{ML}(G_{ML}^{-1}(H_2))} \leq 0$$

and if $G_{ML}(d_2) > H_1$

$$Q''(H_1) = -\frac{1}{G_{ML}^{-1}(H_2)g_{ML}(G_{ML}^{-1}(H_1))} - \frac{Q'(H_1)^2}{G_{ML}^{-1}(H_2)g_{ML}(G_{ML}^{-1}(H_2))} \leq 0$$

so that the curve is concave. \square

Fig 5. The smallest c.d.f. that passes through A, B, C, D, E .

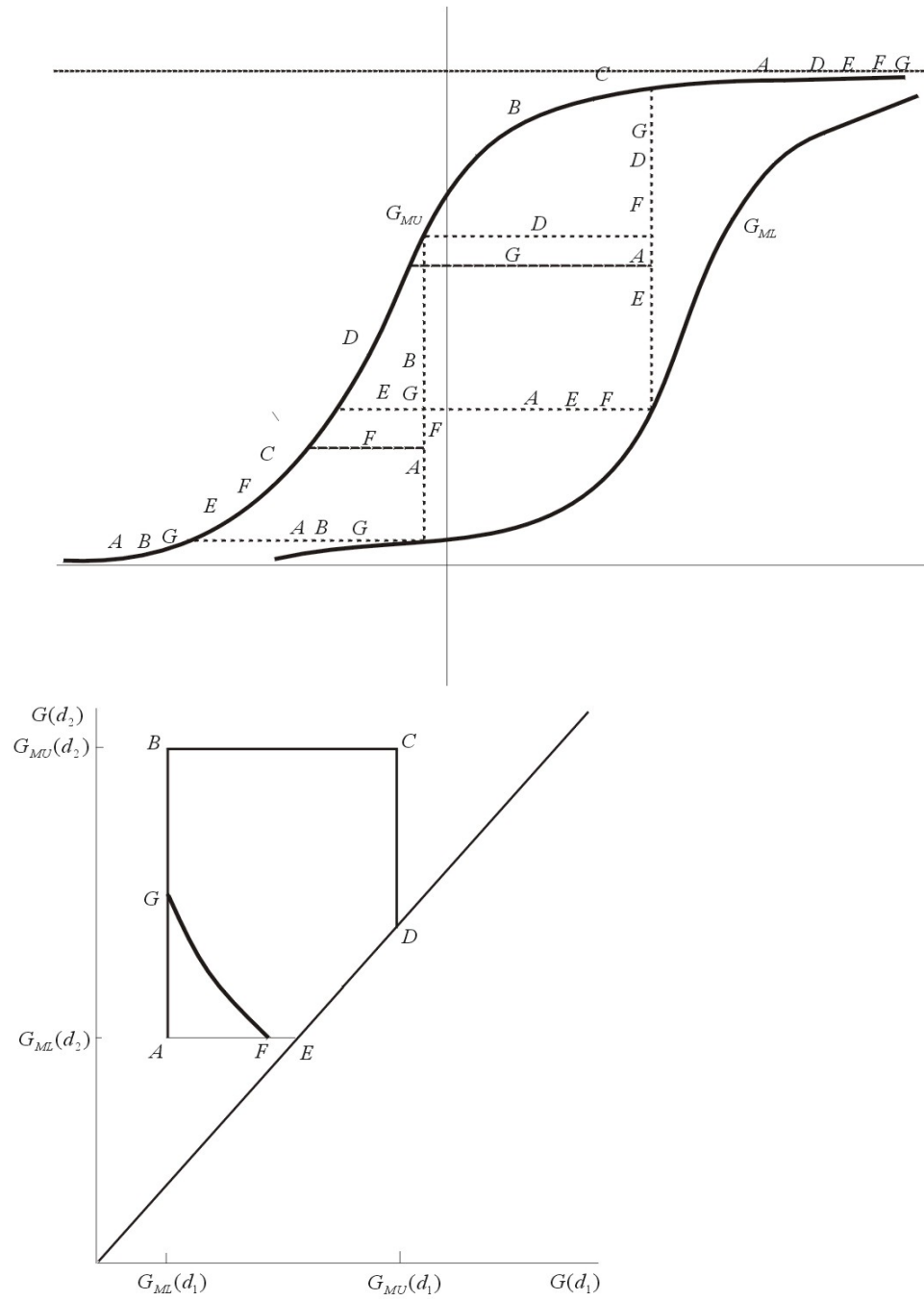


Fig 6. The largest c.d.f. that passes through A, B, C, D, E .

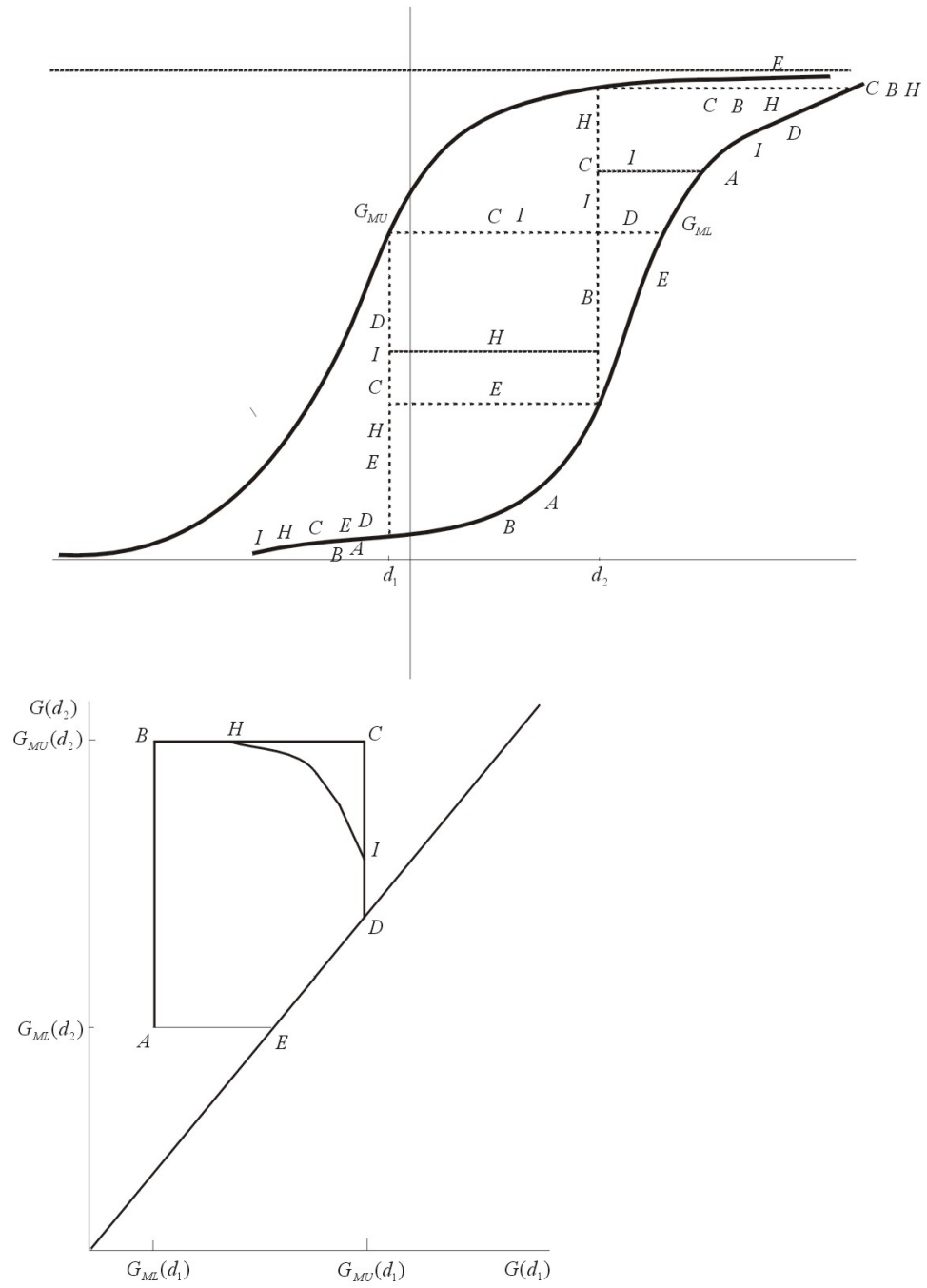


Fig 7. Improved lower bound on $G(d_1)$ and $G(d_2)$ for $d_1 = 1.5$, $d_2 = 2.5$.

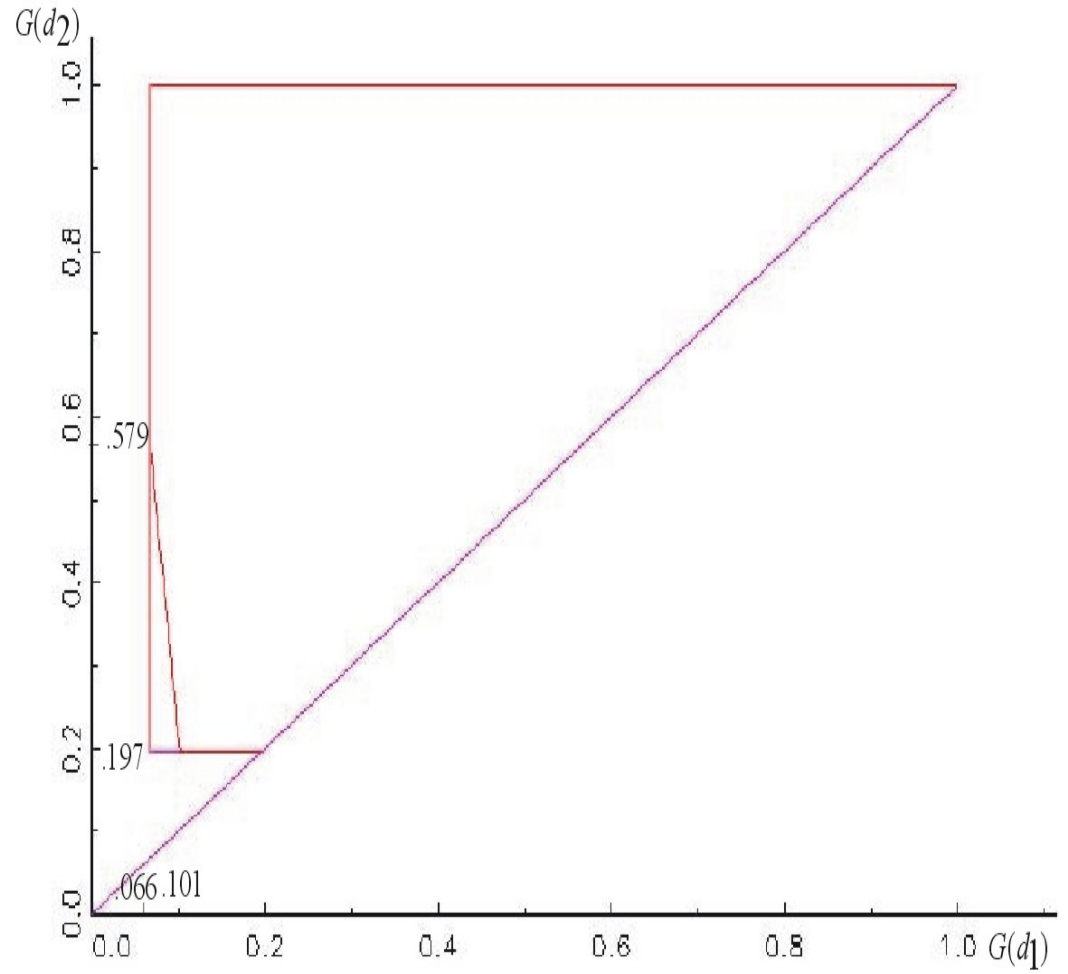


Fig 8. Improved upper bound on $G(d_1)$ and $G(d_2)$ for $d_1 = -.5$, $d_2 = .5$.

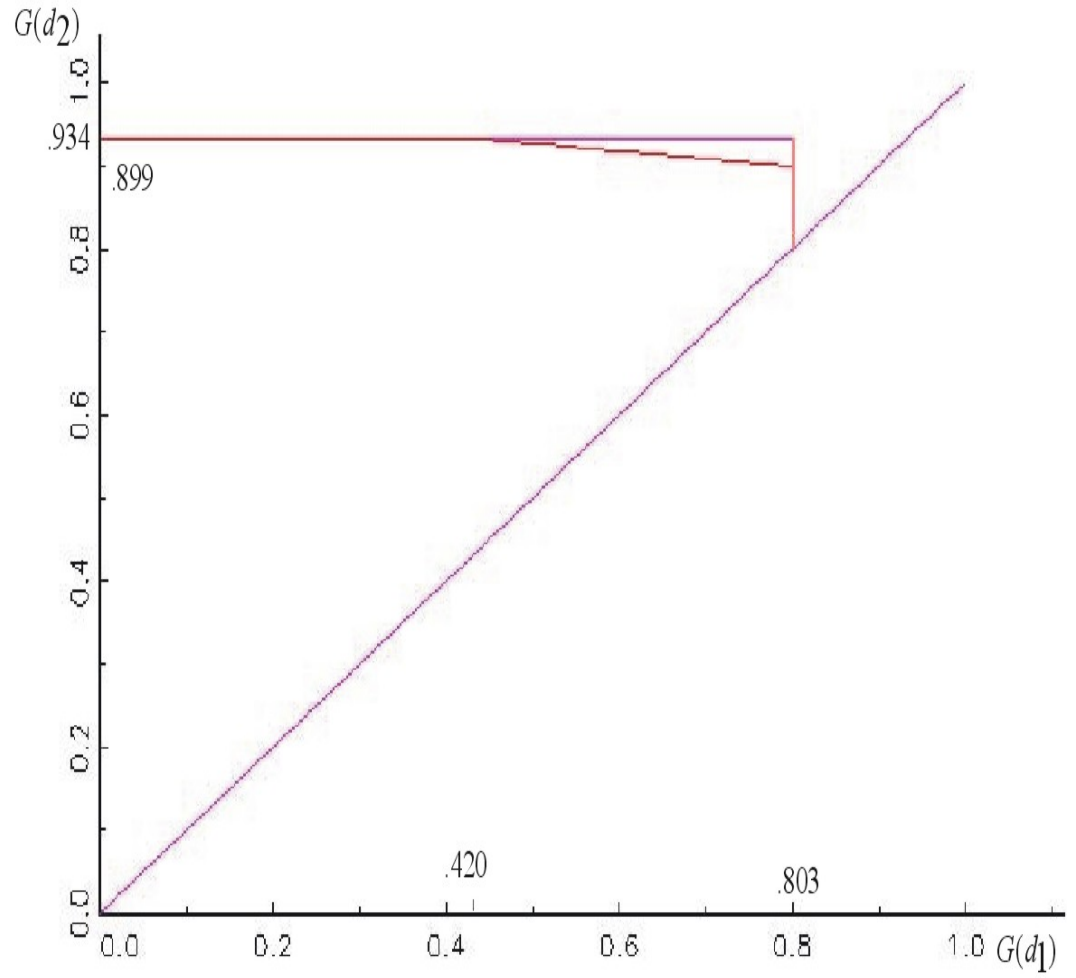


Fig 9. Improved upper and lower bound on $G(d_1)$ and $G(d_2)$ for $d_1 = -1$, $d_2 = 2$.

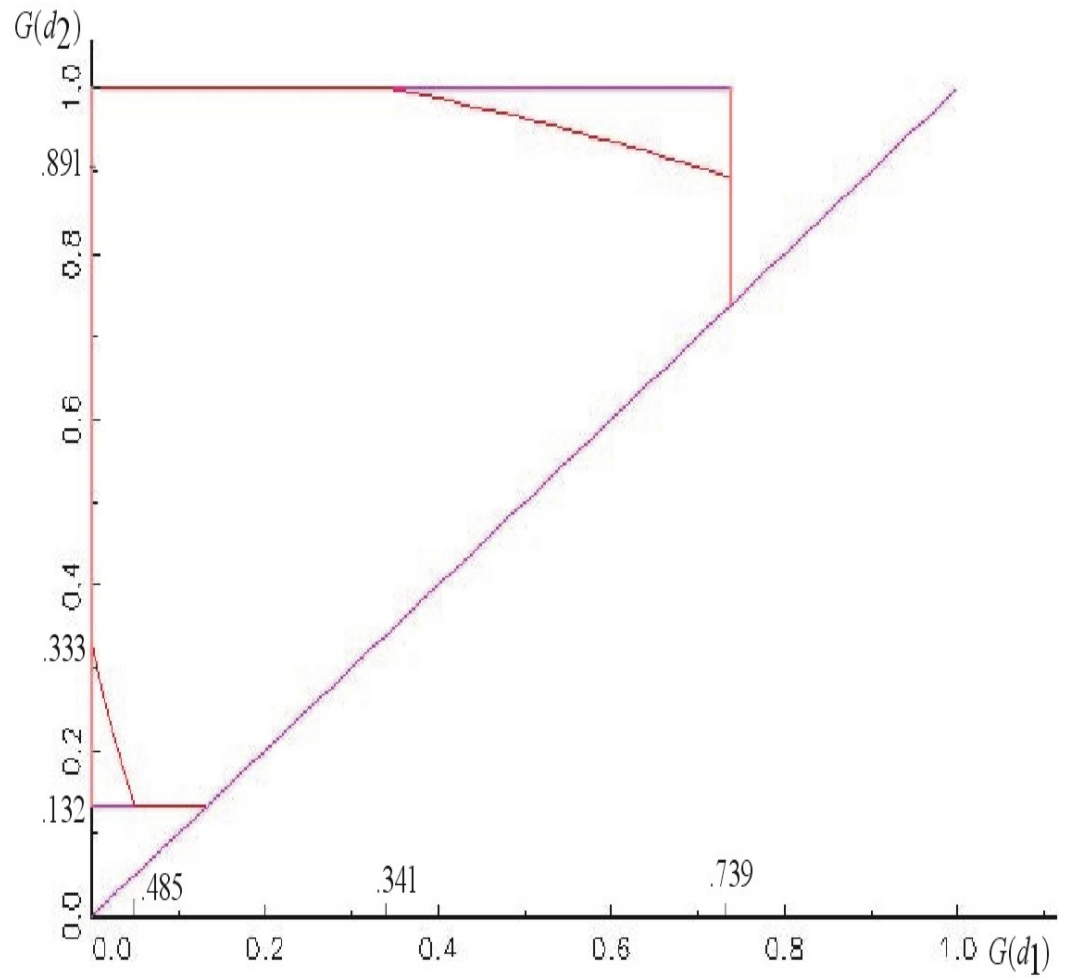


Fig 10. *Bounds on the sum and difference of the treatment effect c.d.f. in two points*

