# Sharp Bounds on the Distribution of the Treatment Effect and Their Statistical Inference*

Yanqin Fan and Sangsoo Park
Department of Economics
Vanderbilt University
VU Station B #351819
2301 Vanderbilt Place
Nashville, TN 37235-1819

This version: December, 2006

**Abstract**

In this paper, we propose nonparametric estimators of sharp bounds on the distribution of the treatement effect of a binary treatment and establish their asymptotic distributions. We point out the possible failure of the standard bootstrap with the same sample size and apply the fewer-than-$n$ bootstrap to making inferences on these bounds. The finite sample performance of the proposed estimators and the fewer-than-$n$ bootstrap confidence intervals is investigated via a simulation study. Finally we establish sharp bounds on the treatment effect distribution when covariates are available.

*We thank Jianqing Fan, Joel Horowitz, Chuck Manski, Per Mykland, Bryan Shepherd, Elie Tamer, and seminar participants in Department of Statistics at the University of Chicago and in Department of Economics at Northwestern University and University of North Carolina for helpful discussions. We also thank Chuck Manski and Jeff Smith for providing useful references. Y. Fan acknowledges financial support from the National Science Foundation.

# 1 Introduction

Evaluating the effect of a treatment or a program is important in diverse disciplines including social sciences and medical sciences. In medical sciences, randomized clinical trials are often used to evaluate the efficacy of a drug or a procedure in the treatment or prevention of disease. The central problem in the evaluation of a treatment is that any potential outcome that program participants would have received without the treatment is not observed. Because of this missing data problem, most work in the treatment effect literature has focused on the evaluation of various average treatment effects such as the mean of the treatment effect, see the recent book by Lee (2005) for discussion and references. However, empirical evidence strongly suggests that treatment effect heterogeneity prevails in many experiments and various interesting effects of the treatment are missed by the average treatment effects alone, see Djebbari and Smith (2004) who studied heterogeneous program impacts in social experiments such as PROGRESA; Black, Smith, Berger, and Noel (2003) who evaluated the Worker Profiling and Reemployment Services system; and Bitler, Gelbach, and Hoynes (2006) who studied Welfare Reform experiments. Other work focusing on treatment effect heterogeneity includes Heckman and Robb (1985), Manski (1990), Imbens and Rubin (1997), Lalonde (1995), Dehejia (1997), Heckman and Smith (1993), Heckman, Smith, and Clements (1997), Lechner (1999), Abadie, Angrist, and Imbens (2002).

When responses to treatment differ among otherwise observationally equivalent subjects, the entire distribution of the treatment effect or other features of the treatment effect than its mean may be of interest. Two approaches have been proposed in the literature to study the distribution of the treatment effect. The first one is the bounding approach originated in Manski (1997a). Assuming monotone treatment response, Manski (1997a) developed sharp bounds on the distribution of the treatment effect. In the second approach, restrictions are imposed on the dependence structure between the potential outcomes such that their joint distribution and the distribution of the treatment effect are identified, see, e.g., Heckman, Smith, and Clements (1997), Biddle, Boden, and Reville (2003), Carneiro, Hansen, and Heckman (2003), Aakvik, Heckman, and Vytlacil (2003), among others.

In this paper, we take the bounding approach and study the estimation and inference on sharp bounds on the distribution of the treatment effect, which are potentially useful when treatment effect is heterogeneous. Unlike Manski (1997a), we do not assume monotone treatment response. Instead, we assume the marginal distributions of the potential outcomes are identified, but their dependence structure is not. One prominent example of this is provided by ideal randomized experiments. In an ideal randomized experiment, participants of the experiment are randomly assigned to a treatment group and a control group. Because of random assignment, observations on the outcome of participants in the treatment group identify the distribution of the potential outcome with treatment and observations on the outcome of participants in the control group identify the distribution of the potential outcome without treatment, but the two independent random samples do not have any information on the dependence structure between the two potential outcomes. As a result, neither the joint distribution of the potential outcomes nor the distribution of the treatment effect (defined as the difference between the two potential outcomes) is identified.

Sharp bounds on the joint distribution of the potential outcomes with identified marginals are given by the Frechet-Hoeffding lower and upper bound distributions, see Heckman and Smith (1993), Heckman, Smith, and Clements (1997), and Manski (1997b) for their applications in program evaluation. For randomized experiments, Heckman, Smith, and Clements (1997) proposed nonparametric estimates of the Fréchet-Hoeffding distribution bounds and developed a test for the "common effect" model by testing the lower bound of the variance of the treatment effect. They also suggested an alternative test based on the difference between the quantile functions of the marginal distributions of the potential outcomes referred to as the quantile treatment effect (QTE), see Firpo (2005) or Section 2 for more references.

Sharp bounds on the distribution of the treatment effect—the difference between two potential outcomes with identified marginals—are known in the probability literature. A.N. Kolmogorov posed the question of finding sharp bounds on the distribution of a sum of two random variables with fixed marginal distributions. It was first solved by Makarov (1981) and later by Rüschendorf (1982) and Frank, Nelsen, and Schweizer (1987) using different techniques. Frank, Nelsen, and Schweizer (1987) showed that their proof based on copulas can be extended to more general functions than the sum. Sharp bounds on the respective distributions of a difference, a product, and a quotient of two random variables with fixed marginals can be found in Williamson and Downs (1990). More recently, Denuit, Genest, and Marceau (1999) extended the bounds for the sum to arbitrary dimensions and provided some applications in finance and risk management, see Embrechts, Hoeing, and Juri (2003) and McNeil, Frey, and Embrechts (2005) for more discussions and additional references.

By making use of the expressions in Williamson and Downs (1990), we propose nonparametric estimators of sharp bounds on the distribution of the treatment effect for randomized experiments and establish their asymptotic properties. More importantly, we develop asymptotically valid statistical methodologies for making inference on these bounds. We point out that the first order asymptotics and the standard bootstrap with the same sample size may not always be asymptotically valid. Instead, we apply the fewer-than-$n$ bootstrap (Bickel, Götze, and Zwet (1997) and Bickel and Sakov (2005)) to constructing confidence intervals for these sharp bounds. The finite sample performances of the first order asymptotics, the standard bootstrap with the same sample size, and the fewer-than-$n$ bootstrap are compared in a simulation study. Our results reveal that (i) when the coverage rate of asymptotic confidence intervals is low, both the standard bootstrap and the fewer-than-$n$ bootstrap correct for the low coverage rate and lead to confidence intervals with more accurate coverage rates; (ii) when the coverage rate of the standard bootstrap confidence intervals is low, the fewer-than-$n$ bootstrap corrects for the low coverage rate; (iii) when the coverage rate of the standard bootstrap is high, the fewer-than-$n$ bootstrap performs similarly to the standard bootstrap; (iv) overall the nonparametric estimators of the sharp bounds are very accurate, although the estimator of the lower bound tends to have a positive bias and the estimator of the upper bound tends to have a negative bias.

Given sharp bounds on the distribution of the treatment effect, we obtain bounds on the class of $D$-parameters introduced in Manski (1997a). One example of a $D$-parameter is any quantile of the treatment effect distribution. In addition, we obtain bounds on the class of $D_2$-parameters of the treatment effect distribution, see Stoye (2005) or Section 2 for the definition of a $D_2$-parameter. As pointed out in Stoye (2005), many inequality and risk measures are $D_2$-parameters. These results shed light on the relation and distinction between QTE and the quantile of the treatment effect distribution.

As an initial investigation of a unified approach to bounding or partially identifying the distribution of the treatment effect, this paper has focused on randomized experiments. Numerous extensions of the methodologies developed in this paper are possible and worthwhile. Of immediate concern is the incorporation of covariates into the analysis. We extend sharp bounds in Williamson and Downs (1990) to take into account the presence of covariates under the selection-on-observables assumption commonly used in the treatment effect literature, see, e.g., Rosenbaum and Rubin (1983a, b), Hahn (1998), Heckman, Ichimura, Smith, and Todd (1998), Dehejia and Wahba (1999), among others.

The rest of this paper is organized as follows. In Section 2, we review sharp bounds on the distribution of a difference of two random variables and provide bounds on parameters of the treatment effect distribution that respect either first or second order stochastic dominance.[1] In Section 3, we propose nonparametric estimators of the distribution bounds and establish their asymptotic properties. Section 4 describes the fewer-than-$n$ bootstrap procedure we use to construct confidence intervals for the distribution bounds. Results from a detailed simulation study are provided in Section 5. In Section 6, we summarize the asymptotic properties of nonparametric estimators of the distribution of a ratio of two random variables, a measure of the relative treatment effect. Section 7 provides sharp bounds on the treatment effect distribution when covariates are available. Section 8 concludes. Proofs are collected in Appendix A. Appendix B presents expressions for the sharp bounds on the distribution of the treatment effect for certain known marginal distributions.

Throughout the paper, we use $\Longrightarrow$ to denote weak convergence. All the limits are taken as the

[1] Horowitz and Manski (1995) first used the concept of 'respect stochastic dominance'. Manski (1997a) referred to parameters that respect first order stochastic dominance as $D$-parameters.

sample size goes to $\infty$.

# 2 Sharp Bounds on the Distribution of the Treatment Effect and its $D$-Parameters

The notation in this paper follows the convention in the treatment effect literature. We consider a binary treatment and use $Y_1$ to denote the potential outcome from receiving treatment and $Y_0$ the outcome without treatment. Let $F(y_1, y_0)$ denote the joint distribution of $Y_1, Y_0$ with marginals $F_1(\cdot)$ and $F_0(\cdot)$ respectively.

The characterization theorem of Sklar (1959) implies that there exists a copula[2] $C(u, v)$: $(u, v) \in [0, 1]^2$ such that $F(y_1, y_0) = C(F_1(y_1), F_0(y_0))$ for all $y_1, y_0$. Conversely, for any marginal distributions $F_1(\cdot), F_0(\cdot)$ and any copula function $C$, the function $C(F_1(y_1), F_0(y_0))$ is a bivariate distribution function with given marginal distributions $F_1, F_0$. This theorem provides the theoretical foundation for the widespread use of the copula approach in generating multivariate distributions from univariate distributions. For reviews, see Joe (1997) and Nelsen (1999). Since copulas connect multivariate distributions to marginal distributions, the copula approach provides a natural way to study the joint distribution of potential outcomes and the distribution of the treatment effect.

For $(u, v) \in [0, 1]^2$, let $C^L(u, v) = \max(u + v - 1, 0)$ and $C^U(u, v) = \min(u, v)$ denote the Fréchet-Hoeffding lower and upper bounds for a copula, i.e., $C^L(u, v) \le C(u, v) \le C^U(u, v)$. Then for any $(y_1, y_0)$, the following inequality holds:

$$C^L(F_1(y_1), F_0(y_0)) \le F(y_1, y_0) \le C^U(F_1(y_1), F_0(y_0)). \tag{1}$$

The bivariate distribution functions $C^L(F_1(y_1), F_0(y_0))$ and $C^U(F_1(y_1), F_0(y_0))$ are referred to as the Fréchet-Hoeffding lower and upper bounds for bivariate distribution functions with fixed marginal distributions $F_1$ and $F_0$. They are distributions of perfectly negatively dependent and perfectly positively dependent random variables respectively, see Nelsen (1999) for more discussions.

Heckman and Smith (1993), Heckman, Smith, and Clements (1997), and Manski (1997b) applied (1) in the context of program evaluation. Lee (2002) applied (1) to bound correlation coefficients in sample selection models. Fan (2006) developed valid statistical inference procedures for $C^L(F_1(y_1), F_0(y_0))$ and $C^U(F_1(y_1), F_0(y_0))$ based on two independent random samples from $F_1(y_1), F_0(y_0)$ respectively.

## 2.1 Sharp Bounds on the Distribution of the Treatment Effect

Let $\Delta = Y_1 - Y_0$ denote the treatment effect or outcome gain and $F_\Delta(\cdot)$ its distribution function. Given the marginals $F_1$ and $F_0$, sharp bounds on the distribution of $\Delta$ can be found in Williamson and Downs (1990).

[2] A copula is a bivariate distribution with uniform marginal distributions on $[0, 1]$.

**Lemma 2.1** *Let* $F^L(\delta) = \sup_y \max(F_1(y) - F_0(y-\delta), 0)$ *and* $F^U(\delta) = 1 + \inf_y \min(F_1(y) - F_0(y-\delta), 0)$. *Then* $F^L(\delta) \leq F_\Delta(\delta) \leq F^U(\delta)$.

We note the following alternative expressions for $F^L(\delta)$ and $F^U(\delta)$ :

$$F^L(\delta) = \max(\sup_y \{F_1(y) - F_0(y-\delta)\}, 0), \tag{2}$$

$$F^U(\delta) = 1 + \min(\inf_y \{F_1(y) - F_0(y-\delta)\}, 0). \tag{3}$$

At any given value of $\delta$, the bounds $(F^L(\delta), F^U(\delta))$ are informative on the value of $F_\Delta(\delta)$ as long as $[F^L(\delta), F^U(\delta)] \subset [0,1]$. Viewed as an inequality among all possible distribution functions, the sharp bounds $F^L(\delta)$ and $F^U(\delta)$ cannot be improved, because it is easy to show that if either $F_1$ or $F_0$ is the degenerate distribution at a finite value, then for all $\delta$, we have $F^L(\delta) = F_\Delta(\delta) = F^U(\delta)$. In fact, given any pair of distribution functions $F_1$ and $F_0$, the inequality: $F^L(\delta) \leq F_\Delta(\delta) \leq F^U(\delta)$ cannot be improved, that is, the bounds $F^L(\delta)$ and $F^U(\delta)$ for $F_\Delta(\delta)$ are point-wise best-possible, see Frank, Nelsen, and Schweizer (1987) for a proof of this for a sum of random variables and Williamson and Downs (1990) for a general operation on two random variables.

Lemma 2.1 implies that the treatment effect distribution $F_\Delta$ first order stochastically dominates $F^U$ and is first order stochastically dominated by $F^L$. Let $\succsim_{FSD}$ denote the first order stochastic dominance relation. Then

$$F^L \succsim_{FSD} F_\Delta \succsim_{FSD} F^U.$$

We note that unlike sharp bounds on the joint distribution of $Y_1, Y_0$, sharp bounds on the distribution of $\Delta$ are not reached at the Fréchet-Hoeffding lower and upper bounds for the distribution of $Y_1, Y_0$.

Let $Y_1', Y_0'$ be perfectly positively dependent and have the same marginal distributions as $Y_1, Y_0$ respectively. Let $\Delta' = Y_1' - Y_0'$. Then the distribution of $\Delta'$ is given by

$$F_{\Delta'}(\delta) = E1\{Y_1' - Y_0' \leq \delta\} = \int_0^1 1\{F_1^{-1}(u) - F_0^{-1}(u) \leq \delta\} du,$$

where $1\{\cdot\}$ is the indicator function the value of which is 1 if the argument is true, 0 otherwise. Similarly, let $Y_1'', Y_0''$ be perfectly negatively dependent and have the same marginal distributions as $Y_1, Y_0$ respectively. Let $\Delta'' = Y_1'' - Y_0''$. Then the distribution of $\Delta''$ is given by

$$F_{\Delta''}(\delta) = E1\{Y_1'' - Y_0'' \leq \delta\} = \int_0^1 1\{F_1^{-1}(u) - F_0^{-1}(1-u) \leq \delta\} du.$$

Interestingly, we show in the next lemma that there exists a second order stochastic dominance relation among the three distributions $F_\Delta, F_{\Delta'}, F_{\Delta''}$. Let $\succsim_{SSD}$ denote the second order stochastic dominance relation.

**Lemma 2.2** *Let $F_\Delta, F_{\Delta'}, F_{\Delta''}$ be defined as above. Then*

$$F_{\Delta'} \succsim_{SSD} F_\Delta \succsim_{SSD} F_{\Delta''}.$$

Theorem 1 in Stoye (2005) shows that $F_{\Delta'} \succsim_{SSD} F_\Delta$ is equivalent to $E\left[U(\Delta')\right] \leq E\left[U(\Delta)\right]$ or $E\left[U(Y_1' - Y_0')\right] \leq E\left[U(Y_1 - Y_0)\right]$ for every convex real-valued function $U$. Corollary 2.3 in Tchen (1980) implies the conclusion of Lemma 2.2, see also Cambanis, Simons, and Stout (1976).

## 2.2 Bounds on $D$-Parameters

The sharp bounds on the treatment effect distribution implies bounds on the class of "$D$-parameters" introduced in Manski (1997a), see also Manski (2003). One example of a "$D$-parameter" is any quantile of the distribution. Stoye (2005) introduced another class of parameters which measure the dispersion of a distribution, including the variance of the distribution. In this section, we show that sharp bounds can be placed on any dispersion or spread parameter of the treatment effect distribution in this class. For convenience, we restate the definitions of both classes of parameters from Stoye (2005). He refers to the class of "$D$-parameters" as the class of "$D_1$-parameters".

**Definition 2.1** *A population statistic $\theta$ is a $D_1$-parameter if it increases weakly with first-order stochastic dominance, that is,*

$$F \succsim_{FSD} G \text{ implies } \theta(F) \geq \theta(G).$$

Obviously if $\theta$ is a $D_1$-parameter, then Lemma 2.1 implies:

$$\theta(F^L) \geq \theta(F_\Delta) \geq \theta(F^U).$$

For example, taking $\theta$ as a quantile of the treatment effect distribution, we obtain immediately its sharp bounds from Lemma 2.1. In the following, we will use $G^{-1}(u)$ to denote the generalized inverse of a nondecreasing function $G$, that is, $G^{-1}(u) = \inf\{x | G(x) \geq u\}$. Then Lemma 2.1 implies: for $0 \leq q \leq 1$,

$$(F^U)^{-1}(q) \leq F_\Delta^{-1}(q) \leq (F^L)^{-1}(q).$$

For the quantile function of a distribution of a sum of two random variables, expressions for its sharp bounds in terms of quantile functions of the marginal distributions are first established in Makarov (1981). They can also be established via the duality theorem, see Schweizer and Sklar (1983). Using the same tool, one can establish the following expressions for sharp bounds on the quantile function of the distribution of the treatment effect, see Williamson and Downs (1990).

**Lemma 2.3** *For $0 \leq q \leq 1$, $(F^U)^{-1}(q) \leq F_\Delta^{-1}(q) \leq (F^L)^{-1}(q)$, where*

$$(F^L)^{-1}(q) = \begin{cases} \inf_{u\in[q,1]}[F_1^{-1}(u) - F_0^{-1}(u-q)] & \text{if } q \neq 0 \\ F_1^{-1}(0) - F_0^{-1}(1) & \text{if } q = 0, \end{cases}$$

$$(F^U)^{-1}(q) = \begin{cases} \sup_{u\in[0,q]}[F_1^{-1}(u) - F_0^{-1}(1+u-q)] & \text{if } q \neq 1 \\ F_1^{-1}(1) - F_0^{-1}(0) & \text{if } q = 1. \end{cases}$$

Like bounds on the distribution of the treatment effect, bounds on the quantile function of $\Delta$ are not reached at the Fréchet-Hoeffding bounds for the distribution of $(Y_1, Y_0)$. The following lemma provides simple expressions for the quantile functions of the treatment effect when the potential outcomes are either perfectly positively dependent or perfectly negatively dependent.

**Lemma 2.4** *For $q \in [0,1]$, we have (i) $F_{\Delta'}^{-1}(q) = \left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ if $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ is an increasing function of $q$; (ii) $F_{\Delta''}^{-1}(q) = \left[F_1^{-1}(q) - F_0^{-1}(1-q)\right]$.*

The proof of Lemma 2.4 follows that of Proposition 3.1 in Embrechts, Hoeing, and Juri (2003). In particular, they showed that for a real valued random variable $Z$ and a function $\varphi$ increasing and left continuous on the range of $Z$, it holds that the quantile of $\varphi(Z)$ at quantile level $q$ is given by $\varphi\left(F_Z^{-1}(q)\right)$, where $F_Z$ is the distribution function of $Z$. For (i), we note that $F_{\Delta'}^{-1}(q)$ equals the quantile of $\left[F_1^{-1}(U) - F_0^{-1}(U)\right]$, where $U$ is a uniform random variable on $[0,1]$. Let $\varphi(U) = F_1^{-1}(U) - F_0^{-1}(U)$. Then $F_{\Delta'}^{-1}(q) = \varphi(q) = F_1^{-1}(q) - F_0^{-1}(q)$ provided that $\varphi(U)$ is an increasing function of $U$. For (ii), let $\varphi(U) = F_1^{-1}(U) - F_0^{-1}(1-U)$. Then $F_{\Delta''}^{-1}(q)$ equals the quantile of $\varphi(U)$. Since $\varphi(U)$ is always increasing in this case, we get $F_{\Delta''}^{-1}(q) = \varphi(q)$.

Note that the condition in (i) is a necessary condition; without this condition, $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ can fail to be a quantile function. Doksum (1974) and Lehmann (1974) used $\left[F_1^{-1}(F_0(y_0)) - y_0\right]$ to measure treatment effect. Recently, $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ has been used to study treatment effect heterogeneity and is referred to as the quantile treatment effect (QTE), see e.g., Heckman, Smith, and Clements (1997), Abadie, Angrist, and Imbens (2002), Chen, Hong, and Tarozzi (2004), Chernozhukov and Hansen (2005), Firpo (2005), Imbens and Newey (2005), among others, for more discussion and references on the estimation of QTE. Manski (1997a) referred to QTE as $\Delta D$-parameters and the quantile of the treatment effect distribution as $D\Delta$-parameters. Assuming monotone treatment response, Manski (1997a) provided sharp bounds on the quantile of the treatment effect distribution.

It is interesting to note that Lemma 2.4 (i) shows that QTE equals the quantile function of the treatment effect only when the two potential outcomes are perfectly positively dependent AND QTE is increasing in $q$. Example 1 below illustrates a case where QTE is decreasing in $q$ and hence is not the same as the quantile function of the treatment effect even when the potential outcomes are perfectly positively dependent. In contrast to QTE, the quantile of the treatment effect distribution is not identified, but can be bounded, see Lemma 2.3. At any given quantile level $q$, the lower quantile bound $(F^U)^{-1}(q)$ is the minimum outcome gain (worst case) of at least $100 \times q$ percent of the population regardless of the dependence structure between the potential outcomes and should be useful to policy makers. For example, $(F^U)^{-1}(0.5)$ is the minimum gain of at least half of the population.

**Definition 2.2** *A population statistic $\theta$ is a $D_2$-parameter if it increases weakly with second order stochastic dominance, i.e.*

$$F \succsim_{SSD} G \text{ implies } \theta(F) \geq \theta(G).$$

If $\theta$ is a $D_2$-parameter, then Lemma 2.2 implies

$$\theta(F_{\Delta'}) \leq \theta(F_{\Delta}) \leq \theta(F_{\Delta''}).$$

Stoye (2005) defined the class of $D_2$-parameters in terms of mean-preserving spread. Since the mean of $\Delta$ is identified in our context, the two definitions lead to the same class of $D_2$-parameters. In contrast to $D_1$-parameters of the treatment effect distribution, bounds on $D_2$-parameters of the treatment effect distribution are reached when the potential outcomes are perfectly dependent on each other. One example of a $D_2$-parameter is the variance of the treatment effect $\Delta$. Using results in Cambanis, Simons, and Stout (1976), Heckman, Smith, and Clements (1997) provided bounds on the variance of $\Delta$ and proposed a test for the common effect model by testing the value of the lower bound on the variance of $\Delta$. Stoye (2005) presents many other examples of $D_2$-parameters, including many well-known inequality and risk measures.

## 2.3 An Illustrative Example: Example 1

In this subsection, we provide explicit expressions for sharp bounds on the distribution of the treatment effect and its quantiles when $Y_1 \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y_0 \sim N\left(\mu_0, \sigma_0^2\right)$. In addition, we provide explicit expressions for the distribution of the treatment effect and its quantiles when the potential outcomes are perfectly positively dependent, perfectly negatively dependent, and independent.

### 2.3.1 Distribution Bounds

Explicit expressions for bounds on the distribution of a sum of two random variables are available for the case where both random variables have the same distribution which includes the uniform, the normal, the Cauchy, and the exponential families, see Alsina (1981), Frank, Nelsen, and Schweizer (1987), and Denuit, Genest, and Marceau (1999). Using the alternative expressions in (2), we now derive sharp bounds on the distribution of $\Delta = Y_1 - Y_0$.

First consider the case $\sigma_1 = \sigma_0 = \sigma$. Let $\Phi(\cdot)$ denote the distribution function of the standard normal distribution. Simple algebra shows

$$\begin{aligned} \sup_y \{F_1(y) - F_0(y-\delta)\} &= 2\Phi\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right) - 1 \text{ for } \delta > \mu_1-\mu_0, \\ \inf_y \{F_1(y) - F_0(y-\delta)\} &= 2\Phi\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right) - 1 \text{ for } \delta < \mu_1-\mu_0. \end{aligned}$$

Hence,

$$F^L(\delta) = \begin{cases} 0, & \text{if } \delta < \mu_1-\mu_0 \\ 2\Phi\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right) - 1, & \text{if } \delta \geq \mu_1-\mu_0 \end{cases}, \tag{4}$$

$$F^U(\delta) = \begin{cases} 2\Phi\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right), & \text{if } \delta < \mu_1-\mu_0 \\ 1, & \text{if } \delta \geq \mu_1-\mu_0 \end{cases}. \tag{5}$$

When[3] $\sigma_1 \neq \sigma_0$, we get

$$\begin{aligned} \sup_y \{F_1(y) - F_0(y-\delta)\} &= \Phi\left(\frac{\sigma_1 s - \sigma_0 t}{\sigma_1^2-\sigma_0^2}\right) + \Phi\left(\frac{\sigma_1 t - \sigma_0 s}{\sigma_1^2-\sigma_0^2}\right) - 1, \\ \inf_y \{F_1(y) - F_0(y-\delta)\} &= \Phi\left(\frac{\sigma_1 s + \sigma_0 t}{\sigma_1^2-\sigma_0^2}\right) - \Phi\left(\frac{\sigma_1 t + \sigma_0 s}{\sigma_1^2-\sigma_0^2}\right) + 1, \end{aligned}$$

where $s = \delta - (\mu_1-\mu_0)$ and $t = \sqrt{s^2 + \left(\sigma_1^2-\sigma_0^2\right)\ln\left(\frac{\sigma_1^2}{\sigma_0^2}\right)}$. For any $\delta$, one can show that $\sup_y \{F_1(y) - F_0(y-\delta)\} > 0$ and $\inf_y \{F_1(y) - F_0(y-\delta)\} < 0$. As a result,

$$\begin{aligned} F^L(\delta) &= \Phi\left(\frac{\sigma_1 s - \sigma_0 t}{\sigma_1^2-\sigma_0^2}\right) + \Phi\left(\frac{\sigma_1 t - \sigma_0 s}{\sigma_1^2-\sigma_0^2}\right) - 1, \\ F^U(\delta) &= \Phi\left(\frac{\sigma_1 s + \sigma_0 t}{\sigma_1^2-\sigma_0^2}\right) - \Phi\left(\frac{\sigma_1 t + \sigma_0 s}{\sigma_1^2-\sigma_0^2}\right) + 1. \end{aligned}$$

For comparison purposes, we provide expressions for the distribution $F_\Delta$ in three special cases.

**Case I. Perfect positive dependence.** In this case, $Y_0$ and $Y_1$ satisfy $Y_0 = \mu_0 + \frac{\sigma_0}{\sigma_1}Y_1 - \frac{\sigma_0}{\sigma_1}\mu_1$.Therefore,

$$\Delta = \begin{cases} \left(\frac{\sigma_1-\sigma_0}{\sigma_1}\right)Y_1 + \left(\frac{\sigma_0}{\sigma_1}\mu_1 - \mu_0\right), & \text{if } \sigma_1 \neq \sigma_0 \\ \mu_1-\mu_0, \text{ if } \sigma_1 = \sigma_0. \end{cases}$$

If $\sigma_1 = \sigma_0$, then

$$F_\Delta(\delta) = \begin{cases} 0 \text{ and } \delta < \mu_1-\mu_0 \\ 1 \text{ and } \mu_1-\mu_0 \leq \delta \end{cases}. \tag{6}$$

[3]Frank, Nelsen, and Schweizer (1987) provided expressions for the sharp bounds on the distribution of a sum of two normal random variables. We believe there are typos in their expressions, as a direct application of their expressions to our case would lead to different expressions from ours. They are:

$$\begin{aligned} F^L(\delta) &= \Phi\left(\frac{-\sigma_1 s - \sigma_0 t}{\sigma_0^2-\sigma_1^2}\right) + \Phi\left(\frac{\sigma_0 s - \sigma_1 t}{\sigma_0^2-\sigma_1^2}\right) - 1, \\ F^U(\delta) &= \Phi\left(\frac{-\sigma_1 s + \sigma_0 t}{\sigma_0^2-\sigma_1^2}\right) + \Phi\left(\frac{\sigma_0 s + \sigma_1 t}{\sigma_0^2-\sigma_1^2}\right). \end{aligned}$$

If $\sigma_1 \neq \sigma_0$, then

$$F_\Delta(\delta) = \Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{|\sigma_1 - \sigma_0|}\right).$$

**Case II. Perfect negative dependence.** In this case, we have $Y_0 = \mu_0 - \frac{\sigma_0}{\sigma_1} Y_1 + \frac{\sigma_0}{\sigma_1}\mu_1$. Hence,

$$\begin{aligned}
\Delta &= \frac{\sigma_1 + \sigma_0}{\sigma_1} Y_1 - \left(\frac{\sigma_0}{\sigma_1}\mu_1 + \mu_0\right), \\
F_\Delta(\delta) &= \Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{\sigma_1 + \sigma_0}\right).
\end{aligned}$$

**Case III. Independence.** This yields

$$F_\Delta(\delta) = \Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{\sqrt{\sigma_1^2 + \sigma_0^2}}\right). \tag{7}$$

Figure 1 below plots the bounds on the distribution $F_\Delta$ (denoted by F_L and F_U) and the distribution $F_\Delta$ corresponding to perfect positive dependence, perfect negative dependence, and independence (denoted by F_PPD, F_PND, and F_IND respectively) of potential outcomes for the case $Y_1 \sim N(2, 2)$ and $Y_0 \sim N(1, 1)$. For notational compactness, we use $(F_1, F_0)$ to signify $Y_1 \sim F_1$ and $Y_0 \sim F_0$ throughout the rest of this paper.

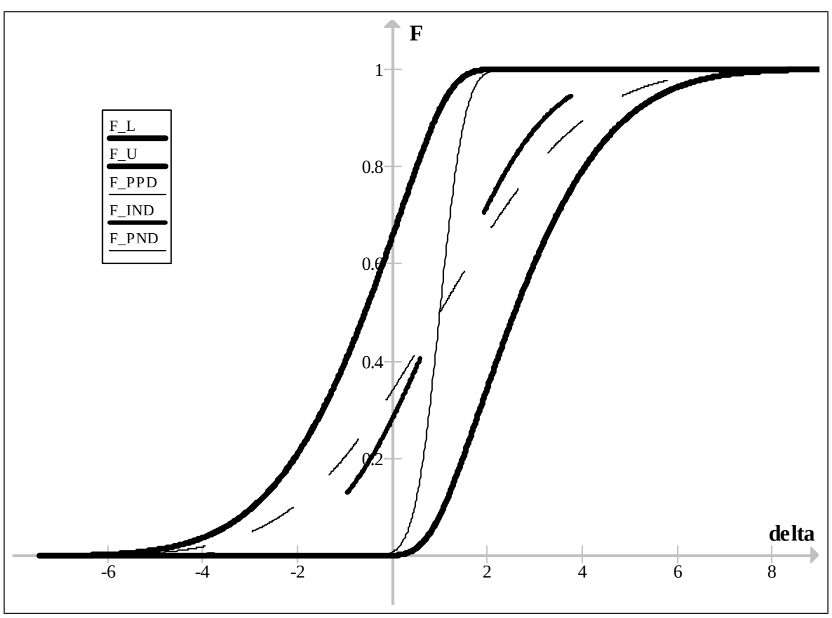


Figure 1. Bounds on the Distribution of the Treatment Effect: $(N(2,2), N(1,1))$

First we observe from Figure 1 that the bounds in this case are informative at all values of $\delta$ and are more informative in the tails of the distribution $F_\Delta$ than in the middle. In addition, Figure 1 indicates that the distribution of the treatment effect for perfectly positively dependent potential outcomes is most concentrated around its mean 1 implied by the second order stochastic relation F_PPD $\succsim_{SSD}$F_IND $\succsim_{SSD}$F_PND. In terms of the corresponding quantile functions, this implies that the quantile function corresponding to the perfectly positively dependent potential outcomes is flatter than the quantile functions corresponding to perfectly negatively dependent and independent potential outcomes, see Figure 2 below.

#### 2.3.2 Quantile Bounds

By inverting (4) and (5), we obtain the quantile bounds for the case $\sigma_1 = \sigma_0 = \sigma$:

$$\begin{aligned}
(F^L)^{-1}(q) &= \begin{cases} \text{any value in } (-\infty, \mu_1 - \mu_0] & \text{for } q = 0, \\ (\mu_1 - \mu_0) + 2\sigma\Phi^{-1}\left(\dfrac{1+q}{2}\right) & \text{otherwise;} \end{cases} \\
(F^U)^{-1}(q) &= \begin{cases} (\mu_1 - \mu_0) + 2\sigma\Phi^{-1}\left(\dfrac{q}{2}\right) & \text{for } q \in [0,1), \\ \text{any value in } [\mu_1 - \mu_0, \infty) & \text{for } q = 1. \end{cases}
\end{aligned}$$

When $\sigma_1 \neq \sigma_0$, there is no closed-form expression for the quantile bounds. But they can be computed numerically by either inverting the distribution bounds or using Lemma 2.3. We now derive the quantile function for the three special cases.

**Case I. Perfect positive dependence.** If $\sigma_1 = \sigma_0$, we get

$$F_\Delta^{-1}(q) = \begin{cases} \text{any value in } (-\infty, \mu_1 - \mu_0) & \text{for } q = 0, \\ \text{any value in } [\mu_1 - \mu_0, \infty) & \text{for } q = 1, \\ \text{undefined for } q \in (0,1). \end{cases}$$

When $\sigma_1 \neq \sigma_0$, we get

$$F_\Delta^{-1}(q) = (\mu_1 - \mu_0) + |\sigma_1 - \sigma_0|\,\Phi^{-1}(q) \ \text{ for } q \in [0,1].$$

Note that by definition, QTE is given by

$$F_1^{-1}(q) - F_0^{-1}(q) = (\mu_1 - \mu_0) + (\sigma_1 - \sigma_0)\Phi^{-1}(q)$$

which equals $F_\Delta^{-1}(q)$ only if $\sigma_1 > \sigma_0$, i.e., only if the condition of Lemma 2.4 (i) holds. If $\sigma_1 < \sigma_0$, $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ is a decreasing function of $q$ and hence can not be a quantile function.

**Case II. Perfect negative dependence.**

$$F_\Delta^{-1}(q) = (\mu_1 - \mu_0) + (\sigma_1 + \sigma_0)\,\Phi^{-1}(q) \ \text{ for } q \in [0,1].$$

**Case III. Independence.**

$$F_\Delta^{-1}(q) = (\mu_1 - \mu_0) + \sqrt{\sigma_1^2 + \sigma_0^2}\Phi^{-1}(q) \ \text{ for } q \in [0,1].$$

In Figure 2 below, we plot the quantile bounds for $\Delta$ (FL^{-1} and FU^{-1}) when $Y_1 \sim N(2,2)$ and $Y_0 \sim N(1,1)$ and the quantile functions of $\Delta$ when $Y_1$ and $Y_0$ are perfectly positively dependent, perfectly negatively dependent, and independent (F_PPD^{-1}, F_PND^{-1}, and F_IND^{-1} respectively).

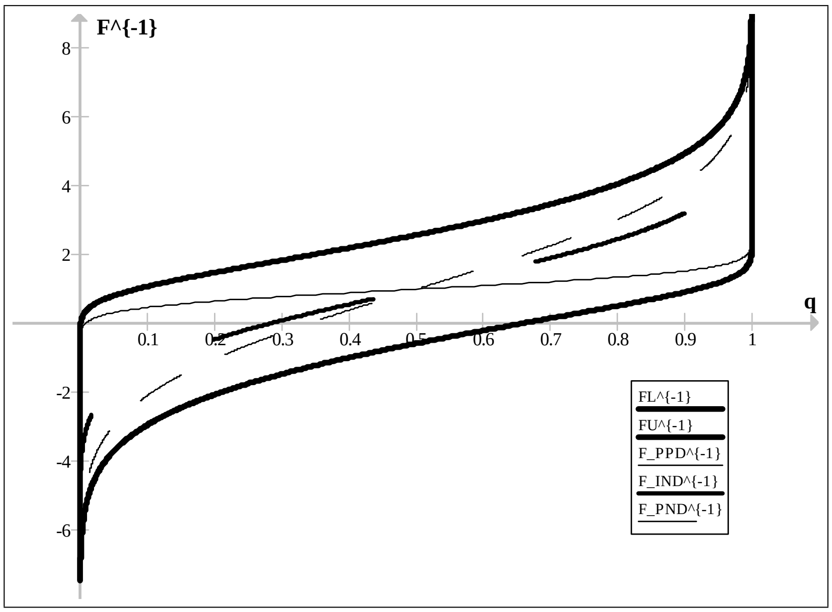


Figure 2. Bounds on the Quantile Function of the Treatment Effect: $(N(2,2), N(1,1))$

Again, Figure 2 reveals the fact that the quantile function of $\Delta$ corresponding to the case that $Y_1$ and $Y_0$ are perfectly positively dependent is flatter than that corresponding to all the other cases. Keeping in mind that in this case, $\sigma_1 > \sigma_0$, we conclude that the quantile function of $\Delta$ in the perfect positive dependence case is the same as QTE. Figure 2 leads to the conclusion that QTE is a conservative measure of the degree of heterogeneity of the treatment effect distribution.

# 3 Nonparametric Estimators and Their Asymptotic Properties

Suppose random samples $\{Y_{1i}\}_{i=1}^{n_1} \sim F_1$ and $\{Y_{0i}\}_{i=1}^{n_0} \sim F_0$ are available. Let $\mathcal{Y}_1$ and $\mathcal{Y}_0$ denote respectively the supports of $F_1$ and $F_0$. Note that the bounds in Lemma 2.1 can be written as

$$F^L(\delta) = \sup_{y \in \mathcal{R}} \{F_1(y) - F_0(y-\delta)\}, \ F^U(\delta) = 1 + \inf_{y \in \mathcal{R}} \{F_1(y) - F_0(y-\delta)\}, \tag{8}$$

since for any two distributions $F_1$ and $F_0$, it is always true that $\sup_{y \in \mathcal{R}} \{F_1(y) - F_0(y-\delta)\} \geq 0$ and $\inf_{y \in \mathcal{R}} \{F_1(y) - F_0(y-\delta)\} \leq 0$.

When $\mathcal{Y}_1 = \mathcal{Y}_0 = \mathcal{R}$, (8) suggests the following plug-in estimators of $F^L(\delta)$ and $F^U(\delta)$:

$$F_n^L(\delta) = \sup_{y\in\mathcal{R}}\{F_{1n}(y) - F_{0n}(y-\delta)\},\ F_n^U(\delta) = 1 + \inf_{y\in\mathcal{R}}\{F_{1n}(y) - F_{0n}(y-\delta)\}, \tag{9}$$

where $F_{1n}(\cdot)$ and $F_{0n}(\cdot)$ are the empirical distributions defined as

$$F_{kn}(y) = \frac{1}{n_k}\sum_{i=1}^{n_k} 1\{Y_{ki} \le y\},\ k = 1, 0.$$

When either $\mathcal{Y}_1$ or $\mathcal{Y}_0$ is not the whole real line, we derive alternative expressions for $F^L(\delta)$ and $F^U(\delta)$ which turn out to be convenient for both computational purposes and for asymptotic analysis. For illustration, we look at the case: $\mathcal{Y}_1 = \mathcal{Y}_0 = [0,1]$ in detail and provide results for the general case afterwards.

Suppose $\mathcal{Y}_1 = \mathcal{Y}_0 = [0,1]$. If $1 \ge \delta \ge 0$, then (8) implies

$$\begin{aligned}
& F^L(\delta) \\
= & \max\left\{\sup_{y\in[\delta,1]}\{F_1(y) - F_0(y-\delta)\},\ \sup_{y\in(-\infty,\delta)}\{F_1(y) - F_0(y-\delta)\},\ \sup_{y\in(1,\infty)}\{F_1(y) - F_0(y-\delta)\}\right\} \\
= & \max\left\{\sup_{y\in[\delta,1]}\{F_1(y) - F_0(y-\delta)\},\ \sup_{y\in(-\infty,\delta)} F_1(y),\ \sup_{y\in(1,\infty)}\{1 - F_0(y-\delta)\}\right\} \\
= & \max\left\{\sup_{y\in[\delta,1]}\{F_1(y) - F_0(y-\delta)\}, F_1(\delta), 1 - F_0(1-\delta)\right\} \\
= & \sup_{y\in[\delta,1]}\{F_1(y) - F_0(y-\delta)\},
\end{aligned} \tag{10}$$

and

$$\begin{aligned}
& F^U(\delta) \\
= & 1 + \min\left\{\inf_{y\in[\delta,1]}\{F_1(y) - F_0(y-\delta)\},\ \inf_{y\in(-\infty,\delta)}\{F_1(y) - F_0(y-\delta)\},\ \inf_{y\in(1,\infty)}\{F_1(y) - F_0(y-\delta)\}\right\} \\
= & 1 + \min\left\{\inf_{y\in[\delta,1]}\{F_1(y) - F_0(y-\delta)\},\ \inf_{y\in(-\infty,\delta)} F_1(y),\ \inf_{y\in(1,\infty)}\{1 - F_0(y-\delta)\}\right\} \\
= & 1 + \min\left\{\inf_{y\in[\delta,1]}\{F_1(y) - F_0(y-\delta)\}, 0\right\};
\end{aligned}$$

If $-1 \le \delta < 0$, then

$$\begin{aligned}
& F^L(\delta) \\
= & \max\left\{\sup_{y\in[0,1+\delta]}\{F_1(y) - F_0(y-\delta)\},\ \sup_{y\in(-\infty,0)}\{F_1(y) - F_0(y-\delta)\},\ \sup_{y\in(1+\delta,\infty)}\{F_1(y) - F_0(y-\delta)\}\right\} \\
= & \max\left\{\sup_{y\in[0,1+\delta]}\{F_1(y) - F_0(y-\delta)\},\ \sup_{y\in(-\infty,0)}\{-F_0(y-\delta)\},\ \sup_{y\in(1+\delta,\infty)}\{F_1(y) - 1\}\right\} \\
= & \max\left\{\sup_{y\in[0,1+\delta]}\{F_1(y) - F_0(y-\delta)\}, 0\right\},
\end{aligned} \tag{11}$$

and

$$
\begin{aligned}
& F^U(\delta) \\
= \;& 1+\min\left\{\inf_{y\in[0,1+\delta]}\{F_1(y)-F_0(y-\delta)\},\inf_{y\in(-\infty,0)}\{F_1(y)-F_0(y-\delta)\},\inf_{y\in(1+\delta,\infty)}\{F_1(y)-F_0(y-\delta)\}\right\} \\
= \;& 1+\min\left\{\inf_{y\in[0,1+\delta]}\{F_1(y)-F_0(y-\delta)\},\inf_{y\in(-\infty,0)}\{-F_0(y-\delta)\},\inf_{y\in(1+\delta,\infty)}\{F_1(y)-1\}\right\} \\
= \;& 1+\inf_{y\in[0,1+\delta]}\{F_1(y)-F_0(y-\delta)\}.
\end{aligned}
$$

Based on (10) and (11), we propose the following estimator of $F^L(\delta)$ :

$$
F_n^L(\delta)=\begin{cases} \sup_{y\in[\delta,1]}\{F_{1n}(y)-F_{0n}(y-\delta)\}, & \text{if } 1\geq\delta\geq 0; \\ \max\left\{\sup_{y\in[0,1+\delta]}\{F_{1n}(y)-F_{0n}(y-\delta)\},0\right\}, & \text{if } -1\leq\delta<0. \end{cases}
$$

Similarly, we propose the following estimator for $F^U(\delta)$ :

$$
F_n^U(\delta)=\begin{cases} 1+\min\left\{\inf_{y\in[\delta,1]}\{F_{1n}(y)-F_{0n}(y-\delta)\},0\right\}, & \text{if } 1\geq\delta\geq 0; \\ 1+\inf_{y\in[0,1+\delta]}\{F_{1n}(y)-F_{0n}(y-\delta)\}. & \text{if } -1\leq\delta<0. \end{cases}
$$

We now summarize the results for general supports $\mathcal{Y}_1$ and $\mathcal{Y}_0$. Suppose $\mathcal{Y}_1=[a,b]$ and $\mathcal{Y}_0=[c,d]$ for $a,b,c,d\in\overline{\mathcal{R}}\equiv\mathcal{R}\cup\{-\infty,+\infty\}$, $a<b, c<d$ with $F_1(a)=F_0(c)=0$ and $F_1(b)=F_0(d)=1$. It is easy to see that

$$
\begin{aligned}
F^L(\delta) &= F^U(\delta)=0, \text{ if } \delta\leq a-d \text{ and} \\
F^L(\delta) &= F^U(\delta)=1, \text{ if } \delta\geq b-c.
\end{aligned}
$$

For any $\delta\in[a-d,b-c]\bigcap\mathcal{R}$, let $\mathcal{Y}_\delta=[a,b]\bigcap[c+\delta,d+\delta]$. A similar derivation to the case $\mathcal{Y}_1=\mathcal{Y}_0=[0,1]$ leads to

$$
F^L(\delta)=\max\left\{\sup_{y\in\mathcal{Y}_\delta}\{F_1(y)-F_0(y-\delta)\},0\right\},
$$

$$
F^U(\delta)=1+\min\left\{\inf_{y\in\mathcal{Y}_\delta}\{F_1(y)-F_0(y-\delta)\},0\right\},
$$

which suggest the following plug-in estimators of $F^L(\delta)$ and $F^U(\delta)$:

$$
F_n^L(\delta)=\max\left\{\sup_{y\in\mathcal{Y}_\delta}\{F_{1n}(y)-F_{0n}(y-\delta)\},0\right\},
$$

$$
F_n^U(\delta)=1+\min\left\{\inf_{y\in\mathcal{Y}_\delta}\{F_{1n}(y)-F_{0n}(y-\delta)\},0\right\}.
$$

For any fixed $\delta$, the consistency of $F_n^L(\delta)$ and $F_n^U(\delta)$ is straightforward. In the rest of this section, we will establish the asymptotic distributions of $\sqrt{n_1}\left(F_n^L(\delta)-F^L(\delta)\right)$ and $\sqrt{n_1}\left(F_n^U(\delta)-F^U(\delta)\right)$. By using $F_n^L(\delta)$ and $F_n^U(\delta)$, we can provide bounds on effects of interest other than the average treatment effect including the proportion of people receiving the treatment who benefit from it, see Heckman, Smith, and Clements (1997) for discussion on some of these effects.

## 3.1 Asymptotic Distributions

Define

$$y_{\sup,\delta} = \arg\sup_{y\in\mathcal{Y}_\delta}\{F_1(y) - F_0(y-\delta)\}\,,\; y_{\inf,\delta} = \arg\inf_{y\in\mathcal{Y}_\delta}\{F_1(y) - F_0(y-\delta)\}\,,$$
$$M(\delta) = \sup_{y\in\mathcal{Y}_\delta}\{F_1(y) - F_0(y-\delta)\}\,,\; m(\delta) = \inf_{y\in\mathcal{Y}_\delta}\{F_1(y) - F_0(y-\delta)\}\,,$$
$$M_n(\delta) = \sup_{y\in\mathcal{Y}_\delta}\{F_{1n}(y) - F_{0n}(y-\delta)\}\,,\; m_n(\delta) = \inf_{y\in\mathcal{Y}_\delta}\{F_{1n}(y) - F_{0n}(y-\delta)\}\,.$$

Then

$$F_n^L(\delta) = \max\{M_n(\delta), 0\}\,,\; F_n^U(\delta) = 1 + \min\{m_n(\delta), 0\}\,.$$

We make the following assumptions.

**(A1)** (i) The two samples $\{Y_{1i}\}_{i=1}^{n_1}$ and $\{Y_{0i}\}_{i=1}^{n_0}$ are each i.i.d. and are independent of each other; (ii) $n_1/n_0 \to \lambda$ as $n_1 \to \infty$ with $0 < \lambda < \infty$.

**(A2)** The distribution functions $F_1$ and $F_0$ are twice differentiable with bounded density functions $f_1$ and $f_0$ on their supports.

**(A3)** (i) For every $\epsilon > 0$, $\sup_{y\in\mathcal{Y}_\delta:|y-y_{\sup,\delta}|\geq\epsilon}\{F_1(y) - F_0(y-\delta)\} < \{F_1(y_{\sup,\delta}) - F_0(y_{\sup,\delta}-\delta)\}$; (ii) $f_1(y_{\sup,\delta}) - f_0(y_{\sup,\delta}-\delta) = 0$ and $f_1'(y_{\sup,\delta}) - f_0'(y_{\sup,\delta}-\delta) < 0$.

**(A4)** (i) For every $\epsilon > 0$, $\inf_{y\in\mathcal{Y}_\delta:|y-y_{\inf,\delta}|\geq\epsilon}\{F_1(y) - F_0(y-\delta)\} > \{F_1(y_{\inf,\delta}) - F_0(y_{\inf,\delta}-\delta)\}$; (ii) $f_1(y_{\inf,\delta}) - f_0(y_{\inf,\delta}-\delta) = 0$ and $f_1'(y_{\inf,\delta}) - f_0'(y_{\inf,\delta}-\delta) > 0$.

The independence assumption of the two samples in (A1) is satisfied by data from ideal randomized experiments. (A2) imposes smoothness assumptions on the marginal distribution functions. (A3) and (A4) are identifiability assumptions. For a fixed $\delta \in [a-d, b-c]\bigcap\mathcal{R}$, (A3) requires the function $y \longmapsto \{F_1(y) - F_0(y-\delta)\}$ to have a well-separated interior maximum at $y_{\sup,\delta}$ on $\mathcal{Y}_\delta$, while (A4) requires the function $y \longmapsto \{F_1(y) - F_0(y-\delta)\}$ to have a well-separated interior minimum at $y_{\inf,\delta}$ on $\mathcal{Y}_\delta$. If $\mathcal{Y}_\delta$ is compact, then (A3) and (A4) are implied by (A2) and the assumption that the function $y \longmapsto \{F_1(y) - F_0(y-\delta)\}$ have a unique maximum at $y_{\sup,\delta}$ and a unique minimum at $y_{\inf,\delta}$ in the interior of $\mathcal{Y}_\delta$.

We first establish the asymptotic distributions of $M_n(\delta)$ and $m_n(\delta)$.

**Proposition 3.1** *Suppose (A1) and (A2) hold. For a given $\delta$, let*

$$\begin{aligned}\sigma_L^2 &= F_1(y_{\sup,\delta})\,[1 - F_1(y_{\sup,\delta})] + \lambda F_0(y_{\sup,\delta}-\delta)\,[1 - F_0(y_{\sup,\delta}-\delta)] \quad \textit{and} \\ \sigma_U^2 &= F_1(y_{\inf,\delta})\,[1 - F_1(y_{\inf,\delta})] + \lambda F_0(y_{\inf,\delta}-\delta)\,[1 - F_0(y_{\inf,\delta}-\delta)]\,.\end{aligned}$$

*Then (i) if (A3) also holds, then* $\sqrt{n_1}[M_n(\delta) - M(\delta)] \Longrightarrow N(0, \sigma_L^2)$; *(ii) if (A4) also holds, then* $\sqrt{n_1}[m_n(\delta) - m(\delta)] \Longrightarrow N(0, \sigma_U^2)$.

Following Fan (2006), we obtain immediately Theorem 3.2 below by using Proposition 3.1.

**THEOREM 3.2** *(i) Suppose (A1)-(A3) hold. Define* $\infty - \infty = \infty$*. For any* $\delta \in [a-d, b-c] \bigcap \mathcal{R}$*, if* $\min\{a-c, b-d\} < \delta$*, then* $\sqrt{n_1}[F_n^L(\delta) - F^L(\delta)] \Longrightarrow N(0, \sigma_L^2)$*; otherwise*

$$\begin{aligned} \sqrt{n_1}[F_n^L(\delta) - F^L(\delta)] \quad &\Longrightarrow \quad \begin{cases} N(0,\sigma_L^2), & \text{if } M(\delta) > 0; \\ \max\left\{N(0,\sigma_L^2), 0\right\} & \text{if } M(\delta) = 0; \end{cases} \\ \text{and } \Pr\left(F_n^L(\delta) = 0\right) \quad &\to \quad 1 \text{ if } M(\delta) < 0. \end{aligned}$$

*(ii) Suppose (A1), (A2), and (A4) hold. Define* $\infty - \infty = -\infty$*. For any* $\delta \in [a-d, b-c] \bigcap \mathcal{R}$*, if* $\delta < \max\{a-c, b-d\}$*, then* $\sqrt{n_1}[F_n^U(\delta) - F^U(\delta)] \Longrightarrow N(0, \sigma_U^2)$*; otherwise*

$$\begin{aligned} \sqrt{n_1}[F_n^U(\delta) - F^U(\delta)] \quad &\Longrightarrow \quad \begin{cases} N(0,\sigma_U^2), & \text{if } m(\delta) < 0; \\ \min\left\{N(0,\sigma_U^2), 0\right\} & \text{if } m(\delta) = 0; \end{cases} \\ \text{and } \Pr\left(F_n^U(\delta) = 1\right) \quad &\to \quad 1 \text{ if } m(\delta) > 0. \end{aligned}$$

Theorem 3.2 shows that the asymptotic distribution of $F_n^L(\delta)$ ($F_n^U(\delta)$) depends on the value of $M(\delta)$ ($m(\delta)$). For example, if $\delta$ is such that $M(\delta) > 0$ ($m(\delta) < 0$), then $F_n^L(\delta)$ ($F_n^U(\delta)$) is asymptotically normally distributed and inference on $F^L(\delta)$ ($F^U(\delta)$) is standard in the sense that both asymptotic normal theory and standard bootstrap with the same sample size are valid. On the other hand, if $\delta$ is such that $M(\delta) = 0$ ($m(\delta) = 0$), then the asymptotic distribution of $F_n^L(\delta)$ ($F_n^U(\delta)$) is truncated normal. Similar to Andrews (2000) and Fan (2006), it is straightforward to show that the standard bootstrap does not work in this case.

We point out that Theorem 3.2 presents the asymptotic distributions of $F_n^L(\delta)$ ($F_n^U(\delta)$) when $y_{\sup,\delta}$ ($y_{\inf,\delta}$) is a unique interior solution. As we demonstrate in Example 2 in the next subsection, when the supports of $F_1$ and $F_0$ are compact, there are often boundary solutions, i.e., $y_{\sup,\delta}$ or $y_{\inf,\delta}$ lie on the boundary of $\mathcal{Y}_\delta$. Moreover, it is also possible to have multiple values for $y_{\sup,\delta}$ and $y_{\inf,\delta}$, some in the interior and some on the boundary. It would be interesting and important to see if one can establish the asymptotic distributions of $F_n^L(\delta)$ and $F_n^U(\delta)$ to accommodate these possibilities. We'll explore this issue in future work. The following theorem, however, presents the rate of convergence of $F_n^L(\delta)$ and $F_n^U(\delta)$ in the general case. It follows from Sherman (2003).

**THEOREM 3.3** *Suppose the supports of* $F_1$ *and* $F_0$ *are compact. If (A1) holds and* $F_1$ *and* $F_0$ *are continuous on their supports, then* $|F_n^L(\delta) - F^L(\delta)| = O_p(n_1^{-1/2})$ *and* $|F_n^U(\delta) - F^U(\delta)| = O_p(n_1^{-1/2})$.

In practice, the supports of $F_1$ and $F_0$ may be unknown, but can be estimated by using the corresponding univariate order statistics in the usual way.

## 3.2 Two Examples

We present two examples to illustrate the various possibilities in Theorem 3.2. For the first example, the asymptotic distribution of $F_n^L(\delta)$ ($F_n^U(\delta)$) is normal for all $\delta$. For the second example, the asymptotic distribution of $F_n^L(\delta)$ $(F_n^U(\delta))$ is normal for some $\delta$ and non-normal for some other $\delta$. More examples can be found in Appendix B.

**Example 1 (Continued).** Let $Y_j \sim N\left(\mu_j, \sigma_j^2\right)$ for $j = 0, 1$ with $\sigma_1^2 \neq \sigma_0^2$. As shown in Section 2.3, $M(\delta) > 0$ and $m(\delta) < 0$ for all $\delta \in \mathcal{R}$. Moreover,

$$y_{\sup,\delta} = \frac{\sigma_1^2 s + \sigma_1\sigma_0 t}{\sigma_1^2 - \sigma_0^2} + \mu_1 \text{ and } y_{\inf,\delta} = \frac{\sigma_1^2 s - \sigma_1\sigma_0 t}{\sigma_1^2 - \sigma_0^2} + \mu_1$$

are unique interior solutions, where $s = \delta - (\mu_1 - \mu_0)$ and $t = \sqrt{s^2 + 2\left(\sigma_1^2 - \sigma_0^2\right)\ln\frac{\sigma_1}{\sigma_0}}$. Theorem 3.2 implies that the asymptotic distribution of $F_n^L(\delta)$ $(F_n^U(\delta))$ is normal for all $\delta \in \mathcal{R}$. Inferences can be made using asymptotic distributions or standard bootstrap with the same sample size.

**Example 2.** Consider the following family of distributions indexed by $a \in (0, 1)$. For brevity, we denote a member of this family by $C(a)$. If $X \sim C(a)$, then

$$F(x) = \begin{cases} \frac{1}{a}x^2 & \text{if } x \in [0, a] \\ 1 - \frac{(x-1)^2}{(1-a)} & \text{if } x \in [a, 1] \end{cases} \quad \text{and} \quad f(x) = \begin{cases} \frac{2}{a}x & \text{if } x \in [0, a] \\ \frac{2(1-x)}{(1-a)} & \text{if } x \in [a, 1] \end{cases}.$$

Suppose $Y_1 \sim C\left(\frac{1}{4}\right)$ and $Y_0 \sim C\left(\frac{3}{4}\right)$. The functional form of $F_1(y) - F_0(y-\delta)$ differs according to $\delta$. For $y \in \mathcal{Y}_\delta$, using the expressions for $F_1(y) - F_0(y-\delta)$ provided in Appendix B, one can find $y_{\sup,\delta}$ and $M(\delta)$. They are:

$$\begin{aligned} y_{\sup,\delta} &= \begin{cases} \frac{1+\delta}{2} & \text{if } -1 + \frac{1}{2}\sqrt{2} < \delta \leq 1 \\ \left\{0, \frac{1+\delta}{2}, 1+\delta\right\} & \text{if } \delta = -1 + \frac{1}{2}\sqrt{2} \\ \{0, 1+\delta\} & \text{if } -1 \leq \delta < -1 + \frac{1}{2}\sqrt{2} \end{cases}; \\ M(\delta) &= \begin{cases} 4(\delta+1)^2 - 1 & \text{if } -1 \leq \delta \leq -\frac{3}{4} \\ -\frac{4}{3}\delta^2 & \text{if } -\frac{3}{4} \leq \delta \leq -1 + \frac{1}{2}\sqrt{2} \\ -\frac{2}{3}(\delta-1)^2 + 1 & \text{if } -1 + \frac{1}{2}\sqrt{2} \leq \delta \leq 1 \end{cases}. \end{aligned}$$

Figure 3 below plots $y_{\sup,\delta}$ and $M(\delta)$ against $\delta$.

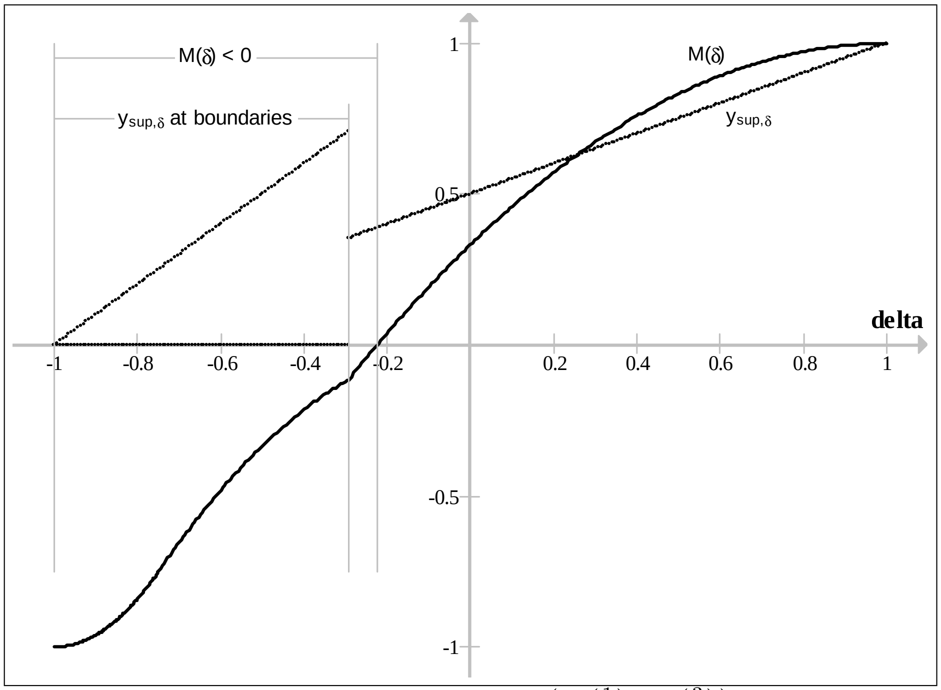

Figure 3. $M(\delta)$ and $y_{\sup,\delta}$: $\left(C\left(\frac{1}{4}\right), C\left(\frac{3}{4}\right)\right)$

Figure 4 below plots $F_1(y) - F_0(y-\delta)$ against $y \in [0,1]$ for a few selected values of $\delta$. When $\delta = -\frac{5}{8}$ (Figure 4(a)), the supremum occurs at the boundaries of $\mathcal{Y}_\delta$. When $\delta = -1 + \frac{\sqrt{2}}{2}$ (Figure 4(b)), $\{y_{\sup,\delta}\} = \left\{0, \frac{1+\delta}{2}, 1+\delta\right\}$, i.e., there are three values of $y_{\sup,\delta}$; one interior and two boundary solutions. When $\delta > -1 + \frac{\sqrt{2}}{2}$, $y_{\sup,\delta}$ becomes a unique interior solution. Figure 4(c) plots the case where the interior solution leads to a value 0 for $M(\delta)$ and Figure 4(d) a case where the interior solution corresponds to a positive value for $M(\delta)$.

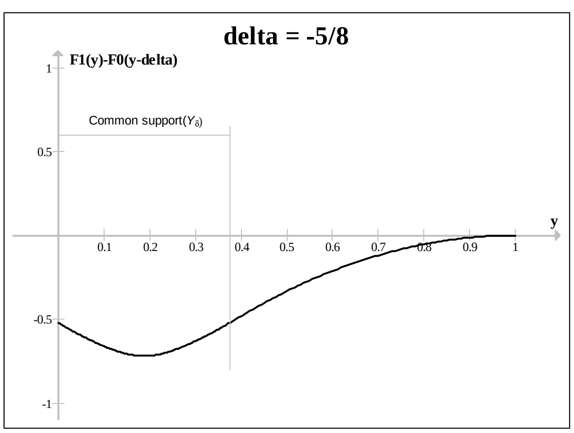

Figure 4(a). $[F_1(y) - F_0(y+5/8)]$

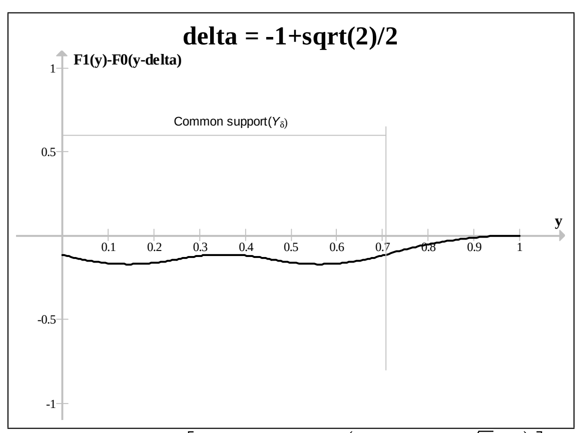

Figure 4(b). $\left[F_1(y) - F_0\left(y+1-\sqrt{2}/2\right)\right]$

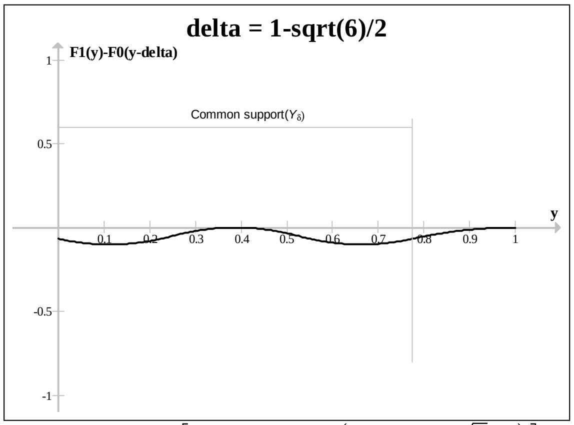


Figure 4(c). $\left[F_1(y) - F_0\left(y - 1 + \sqrt{6}/2\right)\right]$

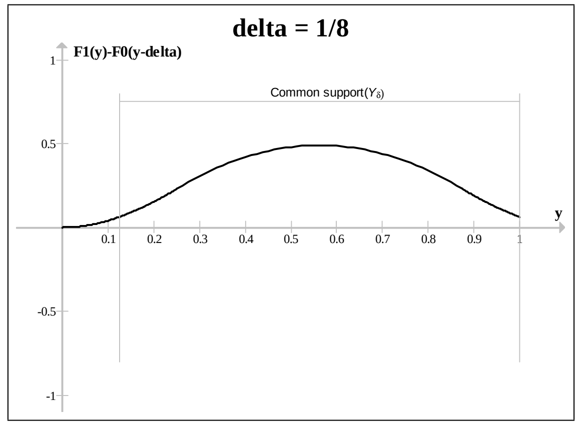


Figure 4(d). $[F_1(y) - F_0(y - 1/8)]$

In the simulation study in the next section, we focus on the case of a unique interior solution for $y_{\sup,\delta}$. Depending on the value of $\delta$, $M(\delta)$ can have different signs leading to different asymptotic distributions for $F_n^L(\delta)$. For example, when $\delta = 1 - \frac{\sqrt{6}}{2}$ (Figure 4(c)), $M(\delta) = 0$ and for $\delta > 1 - \frac{\sqrt{6}}{2}$, $M(\delta) > 0$. Since $M(\delta) = 0$ when $\delta = 1 - \frac{\sqrt{6}}{2}$, $y_{\sup,\delta} = 1 - \frac{\sqrt{6}}{4}$ is in the interior, and $f_1'(y_{\sup,\delta}) - f_0'(y_{\sup,\delta} - \delta) = -\frac{16}{3} < 0$, Theorem 3.2 implies that at $\delta = 1 - \frac{\sqrt{6}}{2}$,

$$\sqrt{n_1}[F_n^L(\delta) - F^L(\delta)] \Longrightarrow \max\left(N(0, \sigma_L^2), 0\right) \text{ where } \sigma_L^2 = \frac{(1+\lambda)}{4}.$$

When $\delta = \frac{1}{8}$ (Figure 4(d)),

$$y_{\sup,\delta} = \frac{9}{16},\ M(\delta) = \frac{47}{96} > 0,\ f_1'(y_{\sup,\delta}) - f_0'(y_{\sup,\delta} - \delta) = -\frac{16}{3} < 0.$$

Theorem 3.2 implies that when $\delta = \frac{1}{8}$,

$$\sqrt{n_1}[F_n^L(\delta) - F^L(\delta)] \Longrightarrow N(0, \sigma_L^2) \text{ where } \sigma_L^2 = (1+\lambda)\frac{7007}{36\,864}.$$

We now illustrate both possibilities for the upper bound $F^U(\delta)$. Suppose $Y_1 \sim C\left(\frac{3}{4}\right)$ and $Y_0 \sim C\left(\frac{1}{4}\right)$. Then using the expressions for $F_1(y) - F_0(y-\delta)$ provided in Appendix B, we obtain

$$y_{\inf,\delta} = \begin{cases} \frac{1+\delta}{2} & \text{if } -1 \le \delta \le 1 - \frac{\sqrt{2}}{2} \\ \left\{\delta, \frac{1+\delta}{2}, 1\right\} & \text{if } \delta = 1 - \frac{\sqrt{2}}{2} \\ \{\delta, 1\} & \text{if } 1 - \frac{1}{2}\sqrt{2} \le z \le 1 \end{cases};$$

$$m(\delta) = \begin{cases} \frac{2}{3}(\delta+1)^2 - 1 & \text{if } -1 \le \delta \le 1 - \frac{\sqrt{2}}{2} \\ \dfrac{4\delta^2}{3} & \text{if } 1 - \frac{\sqrt{2}}{2} \le \delta \le \frac{3}{4} \\ -4(1-\delta)^2 + 1 & \text{if } \frac{3}{4} \le \delta \le 1 \end{cases}.$$

The graphs of $y_{\inf,\delta}$ and $m(\delta)$ are:

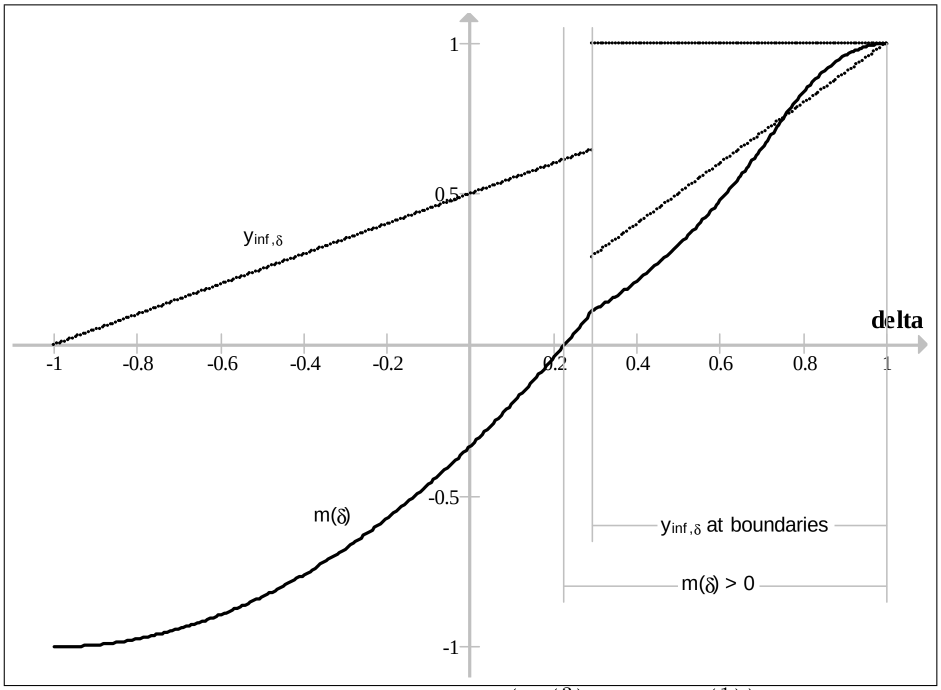


Figure 5. $m(\delta)$ and $y_{\inf,\delta}$: $\left(C\left(\frac{3}{4}\right), Y_0 \sim C\left(\frac{1}{4}\right)\right)$

Graphs of $F_1(y) - F_0(y-\delta)$ against $y$ for selective $\delta$'s are presented in Figure 6 below. Figures 6(a) and 6(b) illustrate two cases each having a unique interior minimum, but in Figure 6(a), $m(\delta)$ is negative and in Figure 6(b), $m(\delta)$ is 0. Figure 6(c) illustrates the case with multiple solutions: one interior minimizer and two boundary ones, while Figure 6(d) illustrates the case with two boundary minima.

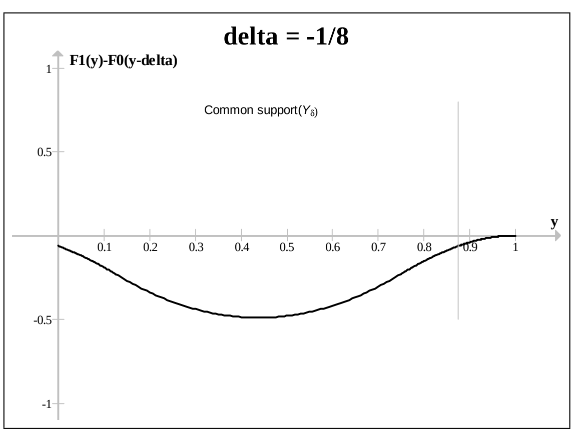


Figure 6(a). $[F_1(y) - F_0(y+1/8)]$

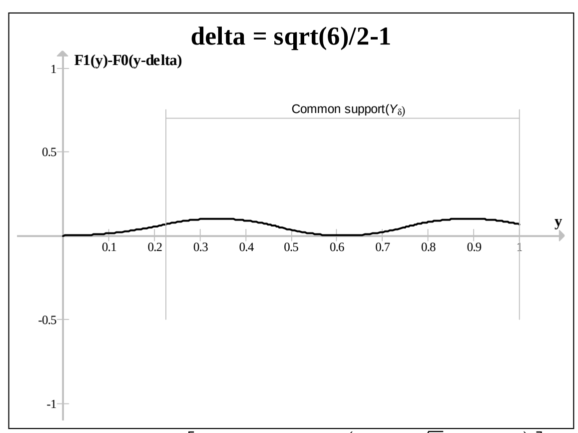


Figure 6(b). $\left[F_1(y) - F_0\left(y - \sqrt{6}/2 + 1\right)\right]$

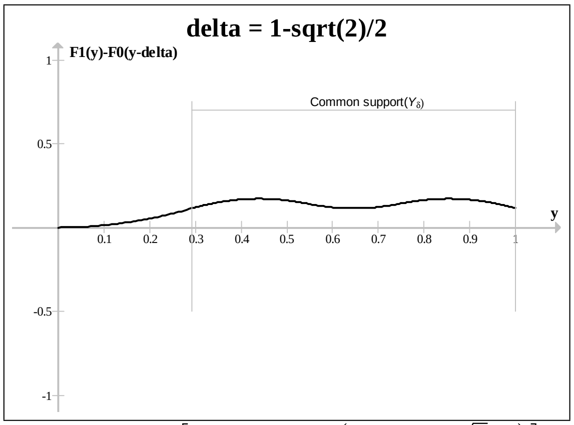


Figure 6(c). $\left[F_1(y) - F_0\left(y - 1 + \sqrt{2}/2\right)\right]$

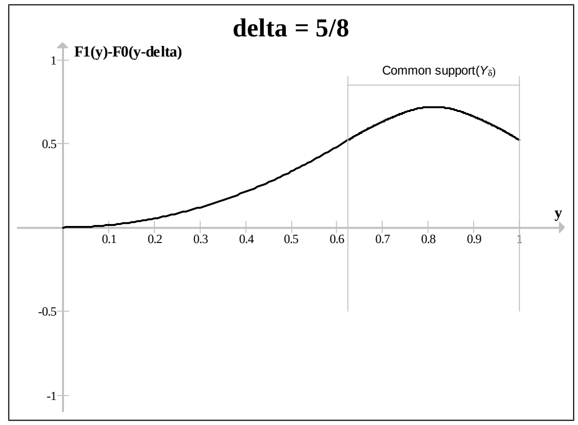


Figure 6(d). $[F_1(y) - F_0(y - 5/8)]$

In the simulation study, we considered the case with a unique interior solution corresponding to Figures 6(a) and (b). When $\delta = \dfrac{\sqrt{6}}{2} - 1$, we obtain $y_{\inf,\delta} = \dfrac{\sqrt{6}}{4}$, $m(\delta) = 0$, and $f_1'(y_{\inf,\delta}) - f_0'(y_{\inf,\delta} - \delta) = \dfrac{16}{3} > 0$. By Theorem 3.2, we get

$$\sqrt{n_1}[F_n^U(\delta) - F^U(\delta)] \Longrightarrow \min\left(N(0, \sigma_U^2), 0\right), \text{ where } \sigma_U^2 = \frac{1+\lambda}{4}.$$

When $\delta = -\dfrac{1}{8}$, we get $y_{\inf,\delta} = \frac{7}{16}$, $m(\delta) = -\frac{47}{96} < 0$, and $f_1'(y_{\inf,\delta}) - f_0'(y_{\inf,\delta} - \delta) = \dfrac{16}{3} > 0$. Hence

$$\sqrt{n_1}[F_n^U(\delta) - F^U(\delta)] \Longrightarrow N(0, \sigma_U^2) \text{ where } \sigma_U^2 = (1+\lambda)\,\frac{7007}{36\,864}.$$

# 4 Inference on the Sharp Bounds

Given Theorem 3.2, the same arguments in Fan (2006) show that the standard bootstrap with the same sample size is asymptotically invalid for $F^L(\delta)$ when $M(\delta) = 0$ (for $F^U(\delta)$ when $m(\delta) = 0$) and this bootstrap failure can be rectified by the fewer-than-$n$ bootstrap or subsampling. Alternatively, note that if $\delta$ is such that $M(\delta) = 0$, then $F^L(\delta) = 0$ and if $\delta$ is such that $m(\delta) = 0$, then $F^U(\delta) = 1$. The failure of the standard bootstrap (bootstrap with the same sample size) at such $\delta$ values follows from the bootstrap failure when parameters are at the boundary of the parameter space, see Andrews (2000).

Both subsampling and fewer-than-$n$ bootstrap have been explored in other contexts to rectify the failure of standard bootstrap, see Andrews (2000), Bickel, Götze, and van Zwet (1997), and Beran (1997) for discussion and references. Subsampling was first proposed by Wu (1990) and extended by Politis and Romano (1994), see Politis, Romano, and Wolf (1999) for more applications of subsampling. Bickel, Götze, and van Zwet (1997) provide numerous examples for which fewer-than-$n$ bootstrap works, while standard bootstrap fails.

In the next section, we investigate the performance of the fewer-than-$n$ bootstrap in constructing confidence intervals for $F^L(\delta)$ and $F^U(\delta)$ for $\delta$ values corresponding to $y_{\sup,\delta}$ ($y_{\inf,\delta}$) being an interior solution with $M(\delta) > 0$ and $M(\delta) = 0$ ($m(\delta) < 0$ and $m(\delta) = 0$). To implement the fewer-than-$n$ bootstrap, we need to choose the subsample size. We use the procedure suggested in Bickel and Sakov (2005). Let $m$ denote the subsample size and $\widehat{m}$ the value of $m$ chosen by the procedure in Bickel and Sakov (2005) (see below for a detailed description of this rule applied to our case). As shown by Bickel and Sakov (2005), $\widehat{m}$ has the desirable property that under general regularity conditions, when the standard bootstrap fails, $\widehat{m} \to \infty$ in probability and $\widehat{m}/n = o_p(1)$; and when the standard bootstrap works, $\widehat{m}/n = O_p(1)$. As a result, there is no loss in efficiency in using the fewer-than-$n$ bootstrap with this adaptive rule of choosing the subsample size. On the other hand, subsampling requires a strictly smaller subsample size, i.e., $m \to \infty$ and $m/n \to 0$.

We now describe this rule for the lower bound $F^L(\delta)$. For notational clarity, we consider the case $n_1 = n_0$. Let $\{Y_{1i}^*\}_{i=1}^m$ be i.i.d. from $F_{1n}(\cdot)$ and $\{Y_{0i}^*\}_{i=1}^m$ i.i.d. from $F_{0n}(\cdot)$ where $m \leq n$. Denote the bootstrap estimators of the sharp bounds by $F_{m,n}^{*L}(\delta)$ and $F_{m,n}^{*U}(\delta)$ and the bootstrap estimators of $\sigma_L^2$ and $\sigma_U^2$ by $\widehat{\sigma}_{m,L}^{2*}$ and $\widehat{\sigma}_{m,U}^{2*}$. Let $T_{m,n}^{*LT} = \sqrt{m}\left(F_{m,n}^{*L}(\delta) - F_n^L(\delta)\right)/\widehat{\sigma}_{m,L}^*$. To choose $m$, we follow the steps below.

**Step 1.** Consider a sequence of $m$'s of the form:

$$m_j = \left[q^j n\right] \text{ for } j = 0, 1, 2, \cdots,\ 0 < q < 1$$

where $[\gamma]$ denotes the largest integer $\leq \gamma$.

**Step 2.** For each $m_j$, let $L_{m_j,n}^*$ denote the empirical distribution of values of $T_{m,n}^{*LT}$ over a large number ($B$) of bootstrap repetitions.

**Step 3.** Let $\widehat{m} = \arg\min_{m_j}\left(\sup_x\left\{\left|L_{m_j,n}^*(x) - L_{m_{j+1},n}^*(x)\right|\right\}\right)$.

Once $\widehat{m}$ is chosen, the confidence intervals can be constructed in the usual way. For example, the $100 \times (1-\alpha)\,\%$ two-sided equal-tailed bootstrap confidence interval for $F^L(\delta)$ based on subsamples of size $\widehat{m}$ is

$$\left[F_n^L(\delta) - \frac{1}{n}\frac{c_{\widehat{m},(1-\alpha/2)}}{\hat{\sigma}_L}, F_n^L(\delta) - \frac{1}{n}\frac{c_{\widehat{m},\alpha/2}}{\hat{\sigma}_L}\right],$$

where $c_{m,\beta} = \inf\left\{x : L_{m,n}^*(x) \geq \beta\right\}$.

# 5 Simulation

In this section, we examine the finite sample accuracy of the nonparametric estimators of the treatment effect distribution bounds and investigate the coverage rates of the standard bootstrap and the fewer-than-$n$ bootstrap confidence intervals for the lower and upper bounds at different values of $\delta$. The data generating processes (DGP) being used in this simulation study are respectively Example 1 and Example 2 introduced in Sections 2 and 3. The detailed simulation design will be described in the subsections below.

## 5.1 Estimates of $F^L$ and $F^U$

### 5.1.1 Computation of $F_n^L$ and $F_n^U$

The quantile functions of $F_n^U$ and $F_n^L$ provide consistent estimators of the lower and upper bounds on the quantile function of $F_\Delta$. For $0 < q < 1$, Lemma 2.3 (the duality theorem) implies that the quantile bounds $(F_n^U)^{-1}(q)$ and $(F_n^L)^{-1}(q)$ can be computed as follows:

$$\begin{aligned}(F_n^L)^{-1}(q) &= \inf_{u\in[q,1]}\left[F_{1n}^{-1}(u) - F_{0n}^{-1}(u-q)\right],\\ (F_n^U)^{-1}(q) &= \sup_{u\in[0,q]}\left[F_{1n}^{-1}(u) - F_{0n}^{-1}(1+u-q)\right],\end{aligned}$$

where $F_{1n}^{-1}(\cdot)$ and $F_{0n}^{-1}(\cdot)$ represent the quantile functions of $F_{1n}(\cdot)$ and $F_{0n}(\cdot)$ respectively.

To estimate the distribution bounds, we compute the values of $(F_n^L)^{-1}(q)$ and $(F_n^U)^{-1}(q)$ at evenly spaced values of $q$ in $(0,1)$. One choice that leads to easily computed formulas for $(F_n^L)^{-1}(q)$ and $(F_n^U)^{-1}(q)$ is $q = r/n_1$ for $r = 1,\ldots,n_1$, as one can show that

$$(F_n^L)^{-1}(r/n_1) = \min_{l=r,\ldots,(n_1-1)}\ \min_{s=j,\ldots,k}[Y_{1(l+1)} - Y_{0(s)}], \tag{12}$$

where $j = \left[n_0\left(\frac{l-r}{n_1}\right)\right] + 1$ and $k = \left[n_0\left(\frac{l-r+1}{n_1}\right)\right]$, and

$$(F_n^U)^{-1}(r/n_1) = \max_{l=0,\ldots,(r-1)}\ \max_{s=j',\ldots,k'}[Y_{1(l+1)} - Y_{0(s)}], \tag{13}$$

where $j' = \left[n_0\left(\frac{n_1+l-r}{n_1}\right)\right] + 1$ and $k' = \left[n_0\left(\frac{n_1+l-r+1}{n_1}\right)\right]$. In the case where $n_1 = n_0 = n$, (12) and (13) simplify:

$$\begin{aligned}(F_n^L)^{-1}(r/n) &= \min_{l=r,\ldots,(n-1)}[Y_{1(l+1)} - Y_{0(l-r+1)}],\\ (F_n^U)^{-1}(r/n) &= \max_{l=0,\ldots,(r-1)}[Y_{1(l+1)} - Y_{0(n+l-r+1)}].\end{aligned}$$

The empirical distribution of $(F_n^L)^{-1}(r/n_1)$, $r = 1,...,n_1$ provides an estimate of the lower bound distribution and the empirical distribution of $(F_n^U)^{-1}(r/n_1)$, $r = 1,...,n_1$ provides an estimate of the upper bound distribution.

### 5.1.2 The Simulation Design and Results

The DGPs being used in this experiment are: (i) $F_1 = N(2,1)$ and $F_0 = N(1,1)$; (ii) $F_1 = N(2,2)$ and $F_0 = N(1,1)$; (iii) $F_1 = C(1/4)$ and $F_0 = C(3/4)$; (iv) $F_1 = C(3/4)$ and $F_0 = C(1/4)$. For each set of marginal distributions, random samples of sizes $n_1 = n_0 = n = 1,000$ are drawn and $F_n^L$ and $F_n^U$ are computed. This is repeated for 500 times. Below we present four graphs. In each graph, we plotted $F_n^L$ and $F_n^U$ randomly chosen from the 500 estimates, the averages of 500 $F_n^L$s and $F_n^U$s, and the simulation variances of $F_n^L$ and $F_n^U$ multiplied by $n$. Each graph consists of 8 curves. The true distribution bounds $F^L$ and $F^U$ are denoted as F^L and F^U, respectively. Their estimates ($F_n^L$ and $F_n^U$) are Fn^L and Fn^U. The lines denoted by avg(Fn^L) and avg(Fn^U) show the averages of 500 $F_n^L$s and $F_n^U$s. The simulation variances of $F_n^L$ and $F_n^U$ multiplied by $n$ are denoted as n*var(Fn^L) and n*var(Fn^U).

Figures 7(a) and (b) correspond to DGP (i) and (ii), while Figures 8(a) and (b) correspond to DGP (iii) and (iv). In all cases, we observe that Fn^L and avg(Fn^L) are very close to F^L at all points of its support (the same holds true for F^U). In fact, these curves are barely distinguishable from each other. The largest variance in all cases for all values of $\delta$ is less than 0.0005.

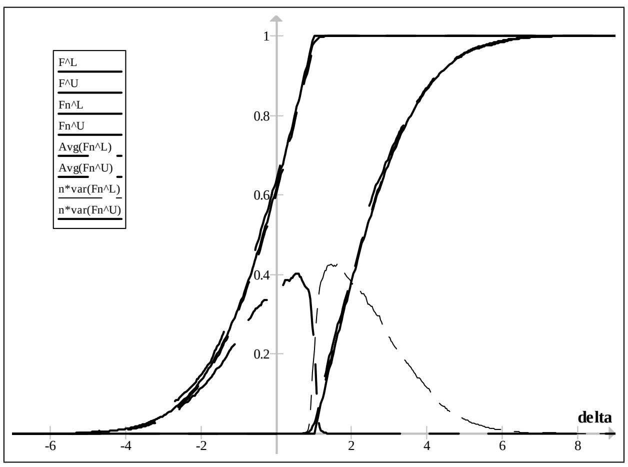


Figure 7(a). Estimates of the Distribution Bounds: $(N(2,1), N(1,1))$

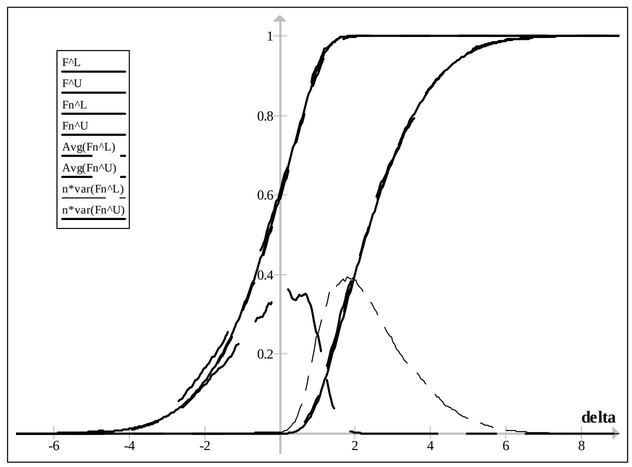


Figure 7(b). Estimates of the Distribution Bounds: $(N(2,2), N(1,1))$

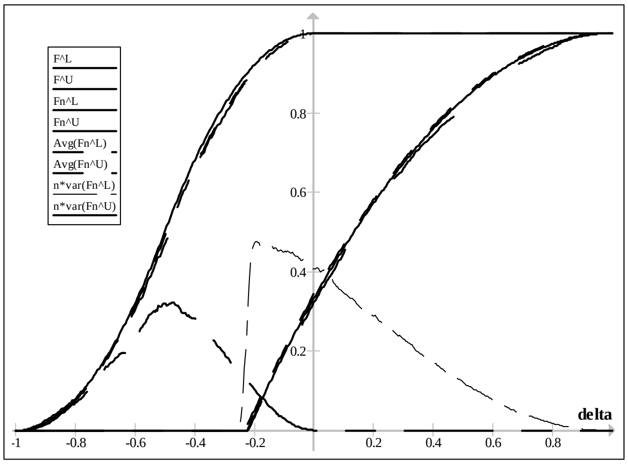


Figure 8(a). Estimates of the Distribution Bounds: $(C(1/4), C(3/4))$

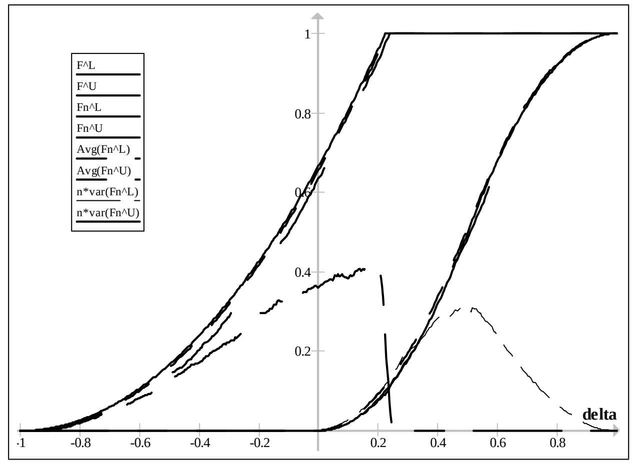


Figure 8(b). Estimates of the Distribution Bounds: $(C(3/4), C(1/4))$

## 5.2 Coverage Rates

### 5.2.1 Computation

Construction of the confidence intervals requires estimation of the variances $\sigma_L^2$ and $\sigma_U^2$ which depend on $y_{\sup .\delta}$ and $y_{\inf .\delta}$. Based on

$$F_n^L(\delta) = \max\{M_n(\delta), 0\} \text{ and } F_n^U(\delta) = 1 + \min\{m_n(\delta), 0\},$$

we now describe a method for computing $M_n(\delta)$, $m_n(\delta)$ and the corresponding $y_{\sup,\delta}$, $y_{\inf,\delta}$.

Suppose we know $\mathcal{Y}_\delta$. To compute $M_n(\delta)$ or $m_n(\delta)$, we just need to consider $Y_{1i} \in \mathcal{Y}_\delta$ and $Y_{0i} \in \mathcal{Y}_\delta - \delta$. If $\mathcal{Y}_\delta$ is unknown, we can estimate it by

$$\mathcal{Y}_{\delta n} = \left[Y_{1(1)}, Y_{1(n_1)}\right] \bigcap \left[Y_{0(1)} + \delta, Y_{0(n_0)} + \delta\right],$$

where $\left\{Y_{1(i)}\right\}_{i=1}^{n_1}$ and $\left\{Y_{0(i)}\right\}_{i=1}^{n_0}$ are the order statistics of $\{Y_{1i}\}_{i=1}^{n_1}$ and $\{Y_{0i}\}_{i=1}^{n_0}$ respectively (in ascending order). In the discussion below, $\mathcal{Y}_\delta$ can be replaced by $\mathcal{Y}_{\delta n}$ if $\mathcal{Y}_\delta$ is unknown.

We define a subset of the order statistics $\{Y_{1i}\}_{i=1}^{n_1}$ denoted as $\left\{Y_{1(i)}\right\}_{i=r_1}^{s_1}$ as follows:

$$r_1 = \arg\min_i \left[\left\{Y_{1(i)}\right\}_{i=1}^{n_1} \bigcap \mathcal{Y}_\delta\right] \text{ and } s_1 = \arg\max_i \left[\left\{Y_{1(i)}\right\}_{i=1}^{n_1} \bigcap \mathcal{Y}_\delta\right].$$

In words, $Y_{1(r_1)}$ is the smallest value of $\left\{Y_{1(i)}\right\}_{i=1}^{n_1} \bigcap \mathcal{Y}_\delta$ and $Y_{1(s_1)}$ is the largest. Then,

$$\begin{aligned} M_n(\delta) &= \max_i \left\{\frac{i}{n_1} - F_{0n}\left(Y_{1(i)} - \delta\right)\right\} \text{ for } i \in \{r_1, r_1+1, \cdots, s_1\} \text{ and} \\ m_n(\delta) &= \min_i \left\{\frac{i}{n_1} - F_{0n}\left(Y_{1(i)} - \delta\right)\right\} \text{ for } i \in \{r_1, r_1+1, \cdots, s_1\}. \end{aligned}$$

To estimate $\sigma_L^2$ and $\sigma_U^2$, we use the following method. Define two sets $I_M$ and $I_m$ such that

$$I_M = \left\{ i : i = \arg\max_i \left\{ \frac{i}{n_1} - F_{0n}\left(Y_{1(i)} - \delta\right) \right\} \right\} \text{ and}$$
$$I_m = \left\{ i : i = \arg\min_i \left\{ \frac{i}{n_1} - F_{0n}\left(Y_{1(i)} - \delta\right) \right\} \right\}.$$

Then the estimators $\sigma_{Ln}^2$ and $\sigma_{Un}^2$ can be defined as

$$\sigma_{Ln}^2 = \frac{i}{n_1}\left(1 - \frac{i}{n_1}\right) + \lambda F_{0n}\left(Y_{1(i)} - \delta\right)\left(1 - F_{0n}\left(Y_{1(i)} - \delta\right)\right) \text{ and}$$
$$\sigma_{Un}^2 = \frac{j}{n_1}\left(1 - \frac{j}{n_1}\right) + \lambda F_{0n}\left(Y_{1(j)} - \delta\right)\left(1 - F_{0n}\left(Y_{1(j)} - \delta\right)\right),$$

for $i \in I_M$ and $j \in I_m$. Since $I_M$ or $I_m$ may not be singleton, we may have multiple estimates of $\sigma_{Ln}^2$ or $\sigma_{Un}^2$. In the simulation, we experimented with different ways of selecting $\sigma_{Ln}^2$ or $\sigma_{Un}^2$ and the results are very similar.

### 5.2.2 The Simulation Design and Results

We looked at pointwise coverage rates of two-sided equal-tailed confidence intervals for the lower and upper bounds separately at deliberately chosen points. The true marginal distributions and the values of $\delta$ used in the simulation are summarized in Table 1.

Table 1: DGPs Used in the Simulation

| | Estimators for | Marginal Distributions $F_1$ | $F_0$ | $\delta$: $\delta_1$ | $\delta_2$ | $\delta_3$ |
|---|---|---|---|---|---|---|
| Example 1 | $F^L(\delta)$ | $N(2,2)$ | $N(1,1)$ | 1.3 | 2.6 | 4.5 |
| | $F^U(\delta)$ | $N(2,2)$ | $N(1,1)$ | $-2.4$ | $-0.6$ | 0.7 |
| Example 2 | $F^L(\delta)$ | $C\left(\frac{1}{4}\right)$ | $C\left(\frac{3}{4}\right)$ | $\frac{1}{8}$ | $1 - \frac{\sqrt{6}}{2}$ | · |
| | $F^U(\delta)$ | $C\left(\frac{3}{4}\right)$ | $C\left(\frac{1}{4}\right)$ | $-\frac{1}{8}$ | $\frac{\sqrt{6}}{2} - 1$ | · |

For Example 1, both $Y_1$ and $Y_0$ are normally distributed. As shown in Section 3.2, $M(\delta) > 0$ and $m(\delta) < 0$ for all three values of $\delta$. Hence both the first order asymptotics and standard bootstrap work for all $\delta$'s. The values of $\delta$ are chosen such that $F^L(\delta_1) \approx F^U(\delta_1) \approx 0.15$, $F^L(\delta_2) \approx F^U(\delta_2) \approx 0.5$, and $F^L(\delta_3) \approx F^U(\delta_3) \approx 0.85$ to see the effect of the relative position of $\delta$ on the coverage rates. For Example 2, $M(\delta_1) > 0$ and $m(\delta_1) < 0$ while $M(\delta_2) = m(\delta_2) = 0$ for both $F^L(\cdot)$ and $F^U(\cdot)$. Hence the standard bootstrap works for $\delta_1$ but not for $\delta_2$.

For each DGP described in Table 1, we generated random samples of the same size $n$ from $F_1$ and $F_0$ respectively. The sample sizes are $n = 1,000, 2,000, 4,000$ and the number of simulations was 1000. To select the number of bootstrap repetitions $B$, we followed Davidson and Mackinnon (2004; p. 163-165) by choosing $B$ such that $\alpha(B+1)$ is an integer. Specifically, we used $B = 999$ for $\alpha = 0.05$. For Example 1, we constructed confidence intervals for $F^L(\delta)$ and $F^U(\delta)$ for each $\delta$

by three methods. The first used the asymptotic distributions of $F_n^L(\delta)$ and $F_n^U(\delta)$. In particular,

$$\sqrt{n}\left[F_n^L(\delta) - F^L(\delta)\right] \Longrightarrow N\left(0, \sigma_L^2\right),$$

where

$$\begin{aligned}
F^L(\delta) &= \Phi\left(\sqrt{2}(\delta-1) - \sqrt{(\delta-1)^2 + \ln 2}\right) - \Phi\left(\delta - 1 - \sqrt{2(\delta-1)^2 + 2\ln 2}\right), \\
\sigma_L^2 &= \Phi\left(\sqrt{2}(\delta-1) + \sqrt{(\delta-1)^2 + \ln 2}\right)\left[1 - \Phi\left(\sqrt{2}(\delta-1) + \sqrt{(\delta-1)^2 + \ln 2}\right)\right] \\
&\quad + \Phi\left(\delta - 1 + \sqrt{2(\delta-1)^2 + 2\ln 2}\right)\left[1 - \Phi\left(\delta - 1 + \sqrt{2(\delta-1)^2 + 2\ln 2}\right)\right];
\end{aligned}$$

$$\sqrt{n}\left[F_n^U(\delta) - F^U(\delta)\right] \Longrightarrow N\left(0, \sigma_U^2\right),$$

where

$$\begin{aligned}
F^U(\delta) &= 1 + \Phi\left(\sqrt{2}(\delta-1) + \sqrt{(\delta-1)^2 + \ln 2}\right) - \Phi\left(\delta - 1 + \sqrt{2(\delta-1)^2 + 2\ln 2}\right), \\
\sigma_U^2 &= \Phi\left(\sqrt{2}(\delta-1) - \sqrt{(\delta-1)^2 + \ln 2}\right)\left[1 - \Phi\left(\sqrt{2}(\delta-1) - \sqrt{(\delta-1)^2 + \ln 2}\right)\right] \\
&\quad + \Phi\left(\delta - 1 - \sqrt{2(\delta-1)^2 + 2\ln 2}\right)\left[1 - \Phi\left(\delta - 1 - \sqrt{2(\delta-1)^2 + 2\ln 2}\right)\right].
\end{aligned}$$

We denote the corresponding results by 'Asymptotics' in Table 2 below. The second method used the standard bootstrap and the results are denoted by '$n$-bootstrap' in Table 2. Finally, we used the 'fewer-than-$n$-bootstrap' confidence intervals. In the 'fewer-than-$n$-bootstrap', we used $q = 0.95$. Here only one value for $q$ was used, because the 'fewer-than-$n$ bootstrap' was only used for comparison purposes (the standard bootstrap works for this case). For Example 2, we used the standard bootstrap ('$n$-bootstrap' in Table 3) and the 'fewer-than-$n$-bootstrap' with two values for $q$: 0.75 and 0.95.

First, we discuss the coverage rates for normal distributions in Table 2. Clearly the coverage rates depend critically on the location of $\delta$. For $\delta_2$, all three methods lead to confidence intervals with very accurate coverage rates for both $F^L$ and $F^U$. The coverage rates at $\delta_1$ and $\delta_3$ depend on the methods being used. Although in theory all three methods are asymptotically valid, in finite samples, confidence intervals based on asymptotic normal critical values often substantially under cover the true values at $\delta_1$ and/or $\delta_3$. For example, the coverage rates of confidence intervals based on normal critical values for $F^L(\delta)$ at $\delta = \delta_1$ and $\delta_3$ are respectively .927 and .937 when $n = 1,000$ and .935 and .936 when $n = 4,000$. On the other hand, the standard bootstrap leads to coverage rates of .942 and .950 when $n = 1,000$ and .945 and .953 when $n = 4,000$, supporting the asymptotic refinement of the standard bootstrap over asymptotic normality in this case. Interestingly, the fewer-than-$n$ bootstrap overall provides slightly better coverage rates than the standard bootstrap especially when the standard bootstrap under covers.

Table 2: Coverage Rates: $(N(2,2), N(1,1))$

| | | $F^L(\delta)$ | | | $F^U(\delta)$ | | |
|---|---|---|---|---|---|---|---|
| $n$ | Method | $\delta_1$ | $\delta_2$ | $\delta_3$ | $\delta_1$ | $\delta_2$ | $\delta_3$ |
| $1,000$ | Asymptotics | .927 | .944 | .937 | .927 | .942 | .927 |
| | $n$-bootstrap | .942 | .954 | .950 | .950 | .953 | .939 |
| $q = 0.95$ | Fewer-than-$n$ bootstrap | .948 | .949 | .948 | .952 | .951 | .942 |
| $2,000$ | Asymptotics | .942 | .944 | .934 | .939 | .941 | .926 |
| | $n$-bootstrap | .949 | .944 | .946 | .946 | .952 | .937 |
| $q = 0.95$ | Fewer-than-$n$ bootstrap | .941 | .944 | .952 | .949 | .950 | .939 |
| $4,000$ | Asymptotics | .935 | .951 | .936 | .947 | .947 | .926 |
| | $n$-bootstrap | .945 | .957 | .953 | .951 | .952 | .936 |
| $q = 0.95$ | Fewer-than-$n$ bootstrap | .944 | .957 | .952 | .951 | .952 | .939 |

For Example 2, the relative performance of the $n$-bootstrap and the fewer-than-$n$ bootstrap at $\delta_1$ is the same as that for the normal distributions in the sense that when the $n$-bootstrap confidence intervals undercover, the fewer-than-$n$ bootstrap confidence intervals correct for this and provide better coverage rates regardless of the value of $q$. At $\delta_2$, the $n$-bootstrap is asymptotically invalid, but leads to coverage rates higher than .95 for almost all sample sizes, while the fewer-than-$n$ bootstrap produces coverage rates that are slightly better than the $n$-bootstrap, but not by much[4]. On the other hand, it may be argued that it is more important to correct undercoverage than overcoverage. Both examples demonstrate that it is exactly when either the first order asymptotics or the standard bootstrap under cover the true bounds, the fewer-than-$n$ bootstrap improves on their coverage rates.

Table 3: Coverage Rates: $\left(C\left(\frac{1}{4}\right), C\left(\frac{3}{4}\right)\right)$ for $F^L$; $\left(C\left(\frac{3}{4}\right), C\left(\frac{1}{4}\right)\right)$ for $F^U$

| | | $F^L(\delta)$ | | $F^U(\delta)$ | |
|---|---|---|---|---|---|
| $n$ | Method | $\delta_1$ | $\delta_2$ | $\delta_1$ | $\delta_2$ |
| $1,000$ | $n$-bootstrap | .941 | .961 | .951 | .958 |
| | Fewer-than-$n$ bootstrap ($q = 0.75$) | .943 | .963 | .951 | .960 |
| | Fewer-than-$n$ bootstrap ($q = 0.95$) | .945 | .963 | .947 | .962 |
| $2,000$ | $n$-bootstrap | .951 | .970 | .947 | .959 |
| | Fewer-than-$n$ bootstrap ($q = 0.75$) | .944 | .971 | .946 | .959 |
| | Fewer-than-$n$ bootstrap ($q = 0.95$) | .951 | .969 | .946 | .959 |
| $4,000$ | $n$-bootstrap | .947 | .963 | .946 | .963 |
| | Fewer-than-$n$ bootstrap ($q = 0.75$) | .949 | .964 | .947 | .965 |
| | Fewer-than-$n$ bootstrap ($q = 0.95$) | .949 | .962 | .951 | .961 |

Tables 4 and 5 below present the bias and RMSE of $F_n^L(\delta)$ and $F_n^U(\delta)$ for the values of $\delta$ used to evaluate coverage rates. As expected, as the sample size $n$ increases, both the bias and the MSE of the lower and upper bound estimators decrease regardless of the values of $\delta$ for both examples.

[4] Our simulation results show that in this case, the average widths of the fewer-than-$n$ bootstrap confidence intervals are often shorter than the $n$-bootstrap confidence intervals.

Also for both examples, the lower bound estimator $F_n^L(\delta)$ is biased upward and the upper bound estimator $F_n^U(\delta)$ is biased downward for all sample sizes and for all values of $\delta$ considered.

Table 4: MSE and Bias: $(N(2,2), N(1,1))$

| | | $F^L(\delta)$ | | | $F^U(\delta)$ | | |
|---|---|---|---|---|---|---|---|
| $n$ | Statistics | $\delta_1$ | $\delta_2$ | $\delta_3$ | $\delta_1$ | $\delta_2$ | $\delta_3$ |
| $1,000$ | $\sqrt{MSE}$ | .0209 | .0194 | .0118 | .0123 | .0198 | .0215 |
| | Bias | .0094 | .0070 | .0046 | $-.0054$ | $-.0082$ | $-.0107$ |
| $2,000$ | $\sqrt{MSE}$ | .0143 | .0135 | .0083 | .0086 | .0138 | .0149 |
| | Bias | .0064 | .0040 | .0033 | $-.0030$ | $-.0052$ | $-.0077$ |
| $4,000$ | $\sqrt{MSE}$ | .0102 | .0094 | .0060 | .0062 | .0097 | .0103 |
| | Bias | .0045 | .0028 | .0022 | $-.0022$ | $-.0034$ | $-.0053$ |

Table 5: MSE and Bias: $\left(C\left(\frac{1}{4}\right), C\left(\frac{3}{4}\right)\right)$ for $F^L$; $\left(C\left(\frac{3}{4}\right), C\left(\frac{1}{4}\right)\right)$ for $F^U$

| | | $F^L(\delta)$ | | $F^U(\delta)$ | |
|---|---|---|---|---|---|
| $n$ | Statistics | $\delta_1$ | $\delta_2$ | $\delta_1$ | $\delta_2$ |
| $1,000$ | $\sqrt{MSE}$ | .0202 | .0216 | .0204 | .0221 |
| | Bias | .0080 | .0147 | $-.0087$ | $-.0155$ |
| $2,000$ | $\sqrt{MSE}$ | .0139 | .0149 | .0144 | .0153 |
| | Bias | .0044 | .0101 | $-.0057$ | $-.0104$ |
| $4,000$ | $\sqrt{MSE}$ | .0098 | .0102 | .0100 | .0103 |
| | Bias | .0033 | .0069 | $-.0033$ | $-.0069$ |

# 6 Estimation and Inference on the Distribution of the Relative Treatment Effect

When the potential outcomes are almost surely positive, an alternative measure of the treatment effect is the relative risk defined as the ratio of the two potential outcomes. Let $R = \frac{Y_1}{Y_0}$. A value of $R$ larger than 1 indicates effectiveness of the treatment and a value of $R$ smaller than 1 indicates ineffectiveness of the treatment. Williamson and Downs (1990) showed that the sharp bounds on the distribution of $R$ are:

$$\begin{aligned} F_R^L(\delta) &= \sup_y \max(F_1(y) - F_0(y/\delta), 0) \text{ and} \\ F_R^U(\delta) &= 1 + \inf_y \min(F_1(y) - F_0(y/\delta), 0). \end{aligned}$$

Let $\mathcal{Y}_1 = [a,b]$ and $\mathcal{Y}_0 = [c,d]$ for $a,b,c,d \in \mathcal{R}_+ \bigcup \{0,\infty\}$ denote the supports of $Y_1$ and $Y_0$ respectively. Define $\mathcal{Y}_{\delta,R} = [a,b] \bigcap [\delta c, \delta d]$ for $\delta \in \left[\frac{a}{d}, \frac{b}{c}\right] \bigcap \mathcal{R}_+$ with obvious definitions of $\frac{a}{d}$ and $\frac{b}{c}$

when one or more of $a, b, c, d \in \{0, \infty\}$. Then it can be shown that

$$
\begin{aligned}
F_R^L(\delta) &= \max\left\{\sup_{y\in\mathcal{Y}_{\delta,R}} \left[F_1(y) - F_0(y/\delta)\right], 0\right\} \equiv \max\left(M_R\left(\delta\right), 0\right) \text{ and} \\
F_R^U(\delta) &= 1 + \min\left\{\inf_{y\in\mathcal{Y}_{\delta,R}} \left[F_1(y) - F_0(y/\delta)\right], 0\right\} \equiv 1 + \min(m_R\left(\delta\right), 0),
\end{aligned}
$$

where

$$
\begin{aligned}
M_R\left(\delta\right) &= F_1(y_{\sup,\delta R}) - F_0(y_{\sup,\delta R}/\delta) \text{ and} \\
m_R\left(\delta\right) &= F_1(y_{\inf,\delta R}) - F_0(y_{\inf,\delta R}/\delta)
\end{aligned}
$$

in which

$$
\begin{aligned}
y_{\sup,\delta R} &= \arg\sup_{y\in\mathcal{Y}_{\delta,R}} \left(F_1(y) - F_0(y/\delta)\right) \text{ and} \\
y_{\inf,\delta R} &= \arg\inf_{y\in\mathcal{Y}_{\delta,R}} \left(F_1(y) - F_0(y/\delta)\right).
\end{aligned}
$$

Consistent estimators of $F_R^L(\delta)$ and $F_R^U(\delta)$ are:

$$
\begin{aligned}
F_{nR}^L(\delta) &= \max\left\{\sup_{y\in\mathcal{Y}_{\delta,R}} \left(F_{1n}(y) - F_{0n}(y/\delta)\right), 0\right\} \text{ and} \\
F_{nR}^U(\delta) &= 1 + \min\left\{\inf_{y\in\mathcal{Y}_{\delta,R}} \left(F_{1n}(y) - F_{0n}(y/\delta)\right), 0\right\}.
\end{aligned}
$$

To provide the asymptotic distributions of $F_{nR}^L(\delta)$ and $F_{nR}^U(\delta)$, we modify (A3) and (A4) to (A3R) and (A4R) below.

**(A3R)** (i) For every $\epsilon > 0$, $\sup_{y\in\mathcal{Y}_{\delta,R}:|y-y_{\sup,\delta R}|\geq\epsilon} \{F_1(y) - F_0(y/\delta)\} < \{F_1(y_{\sup,\delta R}) - F_0(y_{\sup,\delta R}/\delta)\}$;
(ii) $f_1\left(y_{\sup,\delta R}\right) - \frac{1}{\delta} f_0\left(y_{\sup,\delta R}/\delta\right) = 0$ and $f_1'\left(y_{\sup,\delta R}\right) - \frac{1}{\delta^2} f_0'\left(y_{\sup,\delta R}/\delta\right) < 0$.

**(A4R)** (i) For every $\epsilon > 0$, $\inf_{y\in\mathcal{Y}_{\delta,R}:|y-y_{\inf,\delta R}|\geq\epsilon} \{F_1(y) - F_0(y/\delta)\} > \{F_1(y_{\inf,\delta R}) - F_0(y_{\inf,\delta R}/\delta)\}$;
(ii) $f_1\left(y_{\inf,\delta R}\right) - \frac{1}{\delta} f_0\left(y_{\inf,\delta R}/\delta\right) = 0$ and $f_1'\left(y_{\inf,\delta R}\right) - \frac{1}{\delta^2} f_0'\left(y_{\inf,\delta R}/\delta\right) > 0$.

**THEOREM 6.1** *(i) Suppose (A1),(A2) and (A3R) hold. Define* $\frac{0}{0} = \frac{\infty}{\infty} = \infty$*. For any* $\delta \in \left[\frac{a}{d}, \frac{b}{c}\right] \bigcap \mathcal{R}_+$*, if* $\min\left\{\frac{a}{c}, \frac{b}{d}\right\} < \delta$*, then* $\sqrt{n_1}[F_{nR}^L(\delta) - F_R^L(\delta)] \Longrightarrow N(0, \sigma_{LR}^2)$*; otherwise*

$$
\begin{aligned}
\sqrt{n_1}[F_{nR}^L(\delta) - F_R^L(\delta)] &\Longrightarrow \begin{cases} N(0,\sigma_{LR}^2), & \textit{if } M_R(\delta) > 0; \\ \max\left\{N(0,\sigma_{LR}^2), 0\right\} & \textit{if } M_R(\delta) = 0; \end{cases} \\
\textit{and } \Pr\left(F_{nR}^L\left(\delta\right) = 0\right) &\to 1 \textit{ if } M_R(\delta) < 0,
\end{aligned}
$$

*where*

$$
\sigma_{LR}^2 = F_1(y_{\sup,\delta R})\left[1 - F_1(y_{\sup,\delta R})\right] + \lambda F_0(y_{\sup,\delta R}/\delta)\left[1 - F_0(y_{\sup,\delta R}/\delta)\right].
$$

*(ii) Suppose (A1), (A2), and (A4R) hold. Define* $\frac{0}{0} = \frac{\infty}{\infty} = 0$. *For any* $\delta \in \left[\frac{a}{d}, \frac{b}{c}\right] \bigcap \mathcal{R}_{+}$, *if* $\max\left\{\frac{a}{c}, \frac{b}{d}\right\} > \delta$, *then* $\sqrt{n_1}[F_{nR}^U(\delta) - F_R^U(\delta)] \Longrightarrow N(0, \sigma_{UR}^2)$; *otherwise*

$$\sqrt{n_1}[F_{nR}^U(\delta) - F_R^U(\delta)] \Longrightarrow \begin{cases} N(0, \sigma_{UR}^2), & \text{if } m_R(\delta) < 0; \\ \min\left\{N(0, \sigma_{UR}^2), 0\right\} & \text{if } m_R(\delta) = 0; \end{cases}$$
$$\text{and } \Pr\left(F_{nR}^U(\delta) = 1\right) \rightarrow 1 \text{ if } m_R(\delta) > 0,$$

*where*

$$\sigma_{UR}^2 = F_1(y_{\inf,\delta R})\left[1 - F_1(y_{\inf,\delta R}/\delta)\right] + \lambda F_0(y_{\inf,\delta R}/\delta)\left[1 - F_0(y_{\inf,\delta R}/\delta)\right].$$

The proof of Theorem 6.1 is similar to that of Theorem 3.2 and is thus omitted. Like Theorem 3.2, Theorem 6.1 implies that in general, the standard asymptotics and bootstrap may fail to provide valid inference on the sharp bounds $F_R^L(\delta)$ and $F_R^U(\delta)$. Instead the fewer-than-$n$ bootstrap or subsampling should be used to make inferences on these bounds.

# 7 Sharp Bounds on the Distribution of Treatment Effect With Covariates

In many applications, observations on a vector of covariates for individuals in the treatment and control groups are available. In this section, we extend our study on sharp bounds to take into account these covariates. For notational compactness, we let $n = n_1 + n_0$ so that there are $n$ individuals altogether. For $i = 1, ..., n$, let $X_i$ denote the observed vector of covariates and $D_i$ the binary variable indicating participation; $D_i = 1$ if individual $i$ belongs to the treatment group and $D_i = 0$ if individual $i$ belongs to the control group. Let

$$Y_i = Y_{1i} D_i + Y_{0i}(1 - D_i)$$

denote the observed outcome for individual $i$. We have a random sample $\{Y_i, X_i, D_i\}_{i=1}^n$. In the literature on program evaluation with selection-on-observables, the following two assumptions are often used to evaluate the effect of treatment or program, see e.g., Rosenbaum and Rubin (1983a,b), Hahn (1998), Heckman, Ichimura, Smith, and Todd (1998), Dehejia and Wahba (1999), and Hirano, Imbens, and Ridder (2000), to name only a few.

**(C1)** Let $(Y_1, Y_0, D, X)$ have a joint distribution. For all $x \in \mathcal{X}$ (the support of $X$), $(Y_1, Y_0)$ is jointly independent of $D$ conditional on $X = x$.

**(C2)** For all $x \in \mathcal{X}$, $0 < p(x) < 1$, where $p(x) = P(D = 1|x)$.

In the following, we present sharp bounds on the distribution of $\Delta$ under (C1) and (C2). For any fixed $x \in \mathcal{X}$, Lemma 2.1 provides sharp bounds on the conditional distribution of $\Delta$ given

$X = x$:

$$F^L(\delta|x) \le F_\Delta(\delta|x) \le F^U(\delta|x),$$

where

$$\begin{aligned} F^L(\delta|x) &= \sup_y \max(F_1(y|x) - F_0(y-\delta|x), 0), \\ F^U(\delta|x) &= 1 + \inf_y \min(F_1(y|x) - F_0(y-\delta|x), 0). \end{aligned}$$

Here, we use $F_\Delta(\cdot|x)$ to denote the conditional distribution function of $\Delta$ given $X = x$. The other conditional distributions are defined similarly. Conditions (C1) and (C2) allow the identification of the conditional distributions $F_1(y|x)$ and $F_0(y|x)$ appearing in the sharp bounds on $F_\Delta(\delta|x)$. To see this, note that

$$\begin{aligned} F_1(y|x) &= P(Y_1 \le y|X = x) = P(Y_1 \le y|X = x, D = 1) \\ &= P(Y \le y|X = x, D = 1), \end{aligned} \tag{14}$$

where (C1) is used to establish the second equality. Similarly, we get

$$F_0(y|x) = P(Y \le y|X = x, D = 0). \tag{15}$$

Given the random sample $\{Y_i, X_i, D_i\}_{i=1}^n$, nonparametric estimators of the bounds $F^L(\delta|x), F^U(\delta|x)$ can be easily constructed from nonparametric estimators of $F_1(y_1|x)$ and $F_0(y_0|x)$. Their asymptotic properties extend directly from those of $F^L(\delta), F^U(\delta)$ established in Section 3.

Sharp bounds on the unconditional distribution of $\Delta$ follow from those of the conditional distribution:

$$E\left(F^L(\delta|X)\right) \le F_\Delta(\delta) = E\left(F_\Delta(\delta|X)\right) \le E\left(F^U(\delta|X)\right).$$

Let $\widehat{F}_1(y_1|x)$ and $\widehat{F}_0(y_0|x)$ denote nonparametric estimators of $F_1(y_1|x)$ and $F_0(y_0|x)$ respectively. The bounds $E\left(F^L(\delta|X)\right), E\left(F^U(\delta|X)\right)$ can be estimated respectively by

$$\frac{1}{n}\sum_{i=1}^n \max\left(\sup_y \left\{\widehat{F}_1(y|X_i) - \widehat{F}_0(y-\delta|X_i)\right\}, 0\right)$$

and

$$1 + \frac{1}{n}\sum_{i=1}^n \min\left(\inf_y \left\{\widehat{F}_1(y|X_i) - \widehat{F}_0(y-\delta|X_i)\right\}, 0\right).$$

For the sake of space, we will present a complete asymptotic theory for these estimators in a separate paper.

# 8 Conclusion

This paper is the first to develop nonparametric estimation and inference procedures for sharp bounds on the distribution of a difference between two random variables. In the context of program evaluation or evaluation of a binary treatment, the difference between the two potential outcomes measures the program effect or effect of the treatment and hence plays an important role. We have also extended our results to a ratio of two random variables, a measure of the relative treatment effect. As we mentioned in the Introduction, sharp bounds on the distribution of a sum of random variables are important in finance and risk management. The results developed in this paper are directly applicable to a sum of two random variables by redefining the second random variable.

Much work remains to be done. In terms of the sharp bounds, those in this paper are the worst bounds in the sense that they do not make use of any prior information on the possible dependence between the potential outcomes. When such information is available, these bounds can be tightened. In a companion paper, we explore sharp bounds taking account of dependence information such as values of dependence measures of the potential outcomes. The focus on randomized experiments in this paper allows the identification of the marginal distributions. In cases where the marginal distributions themselves are not identifiable but bounds on them can be placed (see, e.g., Manski (1994, 2003), Manski and Pepper (2000), Shaikh and Vytlacil (2005), Blundell, Gosling, Ichimura, and Meghir (2006), Honore and Lleras-Muney (2006)), we can also place bounds on the treatment effect distribution. In terms of statistical inference, we looked at inference on the sharp bounds themselves. Confidence intervals on the true distribution instead of its bounds may be constructed using the methodologies developed recently in Horowitz and Manski (2000), Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2004), and Romano and Shaikh (2006). These results will be reported in separate papers.

# Appendix A: Technical Proofs

**Proof of Proposition 3.1:** Since the proofs of (i) and (ii) are similar, we provide a proof for (i) only. Let

$$Q_n(y,\delta) = F_{1n}(y) - F_{0n}(y-\delta),\ Q(y,\delta) = F_1(y) - F_0(y-\delta).$$

Define

$$\widehat{y}_{\sup,\delta} = \arg\sup_y Q_n(y,\delta).$$

Then $M_n(\delta) = Q_n(\widehat{y}_{\sup,\delta},\delta)$ and $M(\delta) = Q(y_{\sup,\delta},\delta)$. Let $\overline{M}_n(\delta) = Q_n(y_{\sup,\delta},\delta)$. Obviously, $\sqrt{n_1}\left(\overline{M}_n(\delta) - M(\delta)\right) \Longrightarrow N(0,\sigma_L^2)$. We will complete the proof of (i) in three steps:

1. We show that $\widehat{y}_{\sup,\delta} - y_{\sup,\delta} = o_p(1)$;
2. We show that $\widehat{y}_{\sup,\delta} - y_{\sup,\delta} = O_p(n_1^{-1/3})$;
3. $\sqrt{n_1}\left(M_n(\delta) - M(\delta)\right)$ has the same limiting distribution as $\sqrt{n_1}\left(\overline{M}_n(\delta) - M(\delta)\right)$.

**Proof of 1.** By the classical Glivenko-Cantelli theorem, the sequences $\sup_y |F_{1n}(y) - F_1(y)|$ and $\sup_y |F_{0n}(y-\delta) - F_0(y-\delta)|$ converge in probability to zero. Consequently, the sequence $\sup_y |[F_{1n}(y) - F_{0n}(y-\delta)] - [F_1(y) - F_0(y-\delta)]|$ also converges in probability to zero. This and A3(i) imply that the sequence $\widehat{y}_{\sup,\delta}$ converges in probability to $y_{\sup,\delta}$, see e.g., Theorem 5.7 in van der Vaart (1998).

**Proof of 2.** We use Theorem 3.2.5 in van der Vaart and Wellner (1996) to establish the rate of convergence for $\widehat{y}_{\sup,\delta}$. Given (A2), the map: $y \mapsto Q(y,\delta)$ is twice differentiable and has a maximum at $y_{\sup,\delta}$. By (A3), the first condition of Theorem 3.2.5 in van der Vaart and Wellner (1996) is satisfied with $\alpha = 2$. To check the second condition of Theorem 3.2.5 in van der Vaart and Wellner (1996), we consider the centered process:

$$\begin{aligned}\sqrt{n_1}(Q_n - Q)(\cdot,\delta) &= \sqrt{n_1}(F_{1n} - F_1)(\cdot) - \sqrt{n_1}(F_{0n} - F_0)(\cdot - \delta) \\ &\equiv G_{n1}(\cdot) - \frac{\sqrt{n_1}}{\sqrt{n_0}} G_{n0}(\cdot - \delta).\end{aligned}$$

For any $\eta > 0$,

$$\begin{aligned}&E \sup_{|y - y_{\sup,\delta}| < \eta} |\sqrt{n_1}(Q_n - Q)(y,\delta) - \sqrt{n_1}(Q_n - Q)(y_{\sup,\delta},\delta)| \\ &\leq E \sup_{|y - y_{\sup,\delta}| < \eta} |G_{n1}(y) - G_{n1}(y_{\sup,\delta})| + \frac{\sqrt{n_0}}{\sqrt{n_1}} E \sup_{|y - y_{\sup,\delta}| < \eta} |G_{n0}(y-\delta) - G_{n0}(y_{\sup,\delta} - \delta)|.\end{aligned}$$

Note that the envelope function of the class of functions

$$\{I\{(-\infty, y]\} - I\{(-\infty, y_{\sup,\delta}\} : y \in [y_{\sup,\delta} - \eta, y_{\sup,\delta} + \eta]\}$$

is bounded by $I\{(y_{\sup,\delta} - \eta, y_{\sup,\delta} + \eta)\}$ which has a squared $L_2$-norm bounded by $2\left[\sup_y f_1(y)\right]\eta$. Since the class of functions $I\{Y_{1i} \leq \cdot\}$ has a finite uniform entropy integral, Lemma 19.38 in van der Vaart (1998) implies:

$$E \sup_{|y - y_{\sup,\delta}| < \eta} |G_{n1}(y) - G_{n1}(y_{\sup,\delta})| \lesssim \eta^{1/2}. \tag{A.1}$$

Similarly, we can show that

$$E \sup_{|y-y_{\sup,\delta}|<\eta} \left|G_{n0}\left(y-\delta\right) - G_{n0}\left(y_{\sup,\delta}-\delta\right)\right| \lesssim \eta^{1/2}. \tag{A.2}$$

Consequently,

$$E \sup_{|y-y_{\sup,\delta}|<\eta} |\sqrt{n_1}(Q_n - Q)(y,\delta) - \sqrt{n_1}(Q_n - Q)(y_{\sup,\delta},\delta)| \lesssim \eta^{1/2}.$$

Hence the second condition of Theorem 3.2.5 in van der Vaart and Wellner (1996) is satisfied leading to the rate of $n_1^{-1/3}$.

**Proof of 3.** For a fixed $\delta$, we get

$$\begin{aligned}
&\sqrt{n_1}\left(M_n(\delta) - M(\delta)\right) \\
&= \sqrt{n_1}\left(F_{1n}(\widehat{y}_{\sup,\delta}) - F_{0n}(\widehat{y}_{\sup,\delta}-\delta)\right) - \sqrt{n_1}\left(F_1(y_{\sup,\delta}) - F_0(y_{\sup,\delta}-\delta)\right) \\
&= \sqrt{n_1}(Q_n - Q)(\widehat{y}_{\sup,\delta},\delta) + \sqrt{n_1}\left(F_1(\widehat{y}_{\sup,\delta}) - F_0(\widehat{y}_{\sup,\delta}-\delta)\right) - \sqrt{n_1}\left(F_1(y_{\sup,\delta}) - F_0(y_{\sup,\delta}-\delta)\right) \\
&= \sqrt{n_1}(Q_n - Q)(y_{\sup,\delta},\delta) + \sqrt{n_1}\left[F_1(\widehat{y}_{\sup,\delta}) - F_0(\widehat{y}_{\sup,\delta}-\delta) - F_1(y_{\sup,\delta}) - F_0(y_{\sup,\delta}-\delta)\right] + o_p(1) \\
&= \sqrt{n_1}\left(\overline{M}_n(\delta) - M(\delta)\right) + \frac{1}{2}\sqrt{n_1}\left\{f_1'(y^*_{\sup,\delta}) - f_0'(y^*_{\sup,\delta}-\delta)\right\}\left(\widehat{y}_{\sup,\delta} - y_{\sup,\delta}\right)^2 + o_p(1) \\
&= \sqrt{n_1}\left(\overline{M}_n(\delta) - M(\delta)\right) + o_p(1),
\end{aligned}$$

where $y^*_{\sup,\delta}$ lies between $\widehat{y}_{\sup,\delta}$ and $y_{\sup,\delta}$ and we have used stochastic equicontinuity of the process:$\sqrt{n_1}(Q_n - Q)(\cdot,\delta)$ and the first order condition for $\sup_y \{F_1(y) - F_0(y-\delta)\}$.

$\square$

## Appendix B: Functional Forms of $y_{\sup,\delta}$, $y_{\inf,\delta}$, $M(\delta)$ and $m(\delta)$ for Some Known Marginal Distributions

Denuit, Genest, and Marceau (1999) provided the distribution bounds for a sum of two random variables when they both follow shifted Exponential distributions or both follow shifted Pareto distributions. Below, we augment their results with explicit expressions for $y_{\sup,\delta}$, $y_{\inf,\delta}$, $M(\delta)$ and $m(\delta)$ which may help us understand the asymptotic behavior of the nonparametric estimators of the distribution bounds when the true marginals are either shifted Exponential or shifted Pareto.

First, we present some expressions used in Example 2.

**Example 2 (continued).** In Example 2, we considered the family of distributions denoted by $C(a)$ with $a \in (0,1)$. If $X \sim C(a)$, then

$$F(x) = \begin{cases} \dfrac{1}{a}x^2 & \text{if } x \in [0,a] \\ 1 - \dfrac{(x-1)^2}{(1-a)} & \text{if } x \in [a,1] \end{cases} \quad \text{and} \quad f(x) = \begin{cases} \dfrac{2}{a}x & \text{if } x \in [0,a] \\ \dfrac{2(1-x)}{(1-a)} & \text{if } x \in [a,1] \end{cases}.$$

Suppose $Y_1 \sim C(a_1)$ and $Y_0 \sim C(a_0)$. We now provide the functional form of $F_1(y) - F_0(y-\delta)$.

1. Suppose $\delta < 0$. Then $\mathcal{Y}_\delta = [0, 1+\delta]$.

   (a) If $a_0 + \delta \leq 0 < a_1 \leq 1 + \delta$, then

   $$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) & \text{if } 0 \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) & \text{if } a_1 \leq y \leq 1+\delta \end{cases};$$

   (b) If $0 \leq a_0 + \delta \leq a_1 \leq 1 + \delta$, then

   $$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} & \text{if } 0 \leq y \leq a_0+\delta \\ \frac{y^2}{a_1} - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) & \text{if } a_0+\delta \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) & \text{if } a_1 \leq y \leq 1+\delta \end{cases};$$

   (c) If $a_0 + \delta \leq 0 \leq 1 + \delta \leq a_1$, then

   $$F_1(y) - F_0(y-\delta) = \frac{y^2}{a_1} - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \quad \text{if } 0 \leq y \leq 1+\delta;$$

   (d) If $0 \leq a_0 + \delta < 1 + \delta \leq a_1$, then

   $$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} & \text{if } 0 \leq y \leq a_0+\delta \\ \frac{y^2}{a_1} - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) & \text{if } a_0+\delta \leq y \leq 1+\delta \end{cases};$$

   (e) If $0 < a_1 \leq a_0 + \delta \leq 1 + \delta$, then

   $$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} & \text{if } 0 \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \frac{(y-\delta)^2}{a_0} & \text{if } a_1 \leq y \leq a_0+\delta \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) & \text{if } a_0+\delta \leq y \leq 1+\delta \end{cases}.$$

2. Suppose $\delta \geq 0$. Then $\mathcal{Y}_\delta = [\delta, 1]$.

   (a) If $\delta < a_0 + \delta \leq a_1 < 1$, then
   (i) if $a_1 \neq a_0$ and $\delta \neq 0$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} \text{ if } \delta \leq y \leq a_0 + \delta \\ \frac{y^2}{a_1} - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_0 + \delta \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_1 \leq y \leq 1 \end{cases};$$

   (ii) if $a_1 = a_0 = a$ and $\delta = 0$, then

$$F_1(y) - F_0(y-\delta) = 0 \text{ for all } y \in [0,1].$$

   (b) If $\delta \leq a_1 \leq a_0 + \delta \leq 1$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} \text{ if } \delta \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \frac{(y-\delta)^2}{a_0} \text{ if } a_1 \leq y \leq a_0 + \delta \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_0 + \delta \leq y \leq 1 \end{cases};$$

   (c) If $\delta \leq a_1 < 1 \leq a_0 + \delta$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} \text{ if } \delta \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \frac{(y-\delta)^2}{a_0} \text{ if } a_1 \leq y \leq 1 \end{cases};$$

   (d) If $a_1 < \delta < a_0 + \delta \leq 1$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \frac{(y-\delta)^2}{a_0} \text{ if } \delta \leq y \leq a_0 + \delta \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_0 + \delta \leq y \leq 1 \end{cases};$$

   (e) If $a_1 < \delta < 1 \leq a_0 + \delta$, then

$$F_1(y) - F_0(y-\delta) = \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \frac{(y-\delta)^2}{a_0} if \delta \leq y \leq 1.$$

**(Shifted) Exponential marginals.** The marginal distributions are:

$$\begin{aligned} F_1(y) &= 1 - \exp\left(-\tfrac{y-\theta_1}{\alpha_1}\right) \text{ for } y \in [\theta_1, \infty) \text{ and} \\ F_0(y) &= 1 - \exp\left(-\tfrac{y-\theta_0}{\alpha_0}\right) \text{ for } y \in [\theta_0, \infty), \text{ where } \alpha_1, \theta_1, \alpha_0, \theta_0 > 0. \end{aligned}$$

Let $\delta_c = (\theta_1 - \theta_0) - \min\{\alpha_1, \alpha_0\}(\ln \alpha_1 - \ln \alpha_0)$.

1. Suppose $\alpha_1 < \alpha_0$.

(a) If $\delta \leq \delta_c$,

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = 0, \\
\text{where } M(\delta) &= \left(\left(\frac{\alpha_0}{\alpha_1}\right)^{\frac{\alpha_1}{\alpha_1-\alpha_0}} - \left(\frac{\alpha_0}{\alpha_1}\right)^{\frac{\alpha_0}{\alpha_1-\alpha_0}}\right)\exp\left(-\frac{\delta-(\theta_1-\theta_0)}{\alpha_1-\alpha_0}\right) < 0, \\
\text{and } y_{\inf,\delta} &= \frac{\alpha_0\alpha_1(\ln\alpha_1-\ln\alpha_0)+\alpha_1\theta_0-\alpha_0\theta_1+\alpha_1\delta}{\alpha_1-\alpha_0} \text{ (an interior solution).} \\
F^U(\delta) &= 1+\min\{m(\delta), 0\} = 1+m(\delta), \\
\text{where } m(\delta) &= \min\left\{\exp\left(-\tfrac{\max\{\theta_1-(\delta+\theta_0),0\}}{\alpha_0}\right) - \exp\left(-\tfrac{\max\{\theta_0+\delta-\theta_1,0\}}{\alpha_1}\right), 0\right\} \\
\text{and } y_{\sup,\delta} &= \max\{\theta_1, \theta_0+\delta\} \text{ or } \infty \text{ (boundary solution).}
\end{aligned}$$

(b) If $\delta > \delta_c$,

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta) > 0, \\
\text{where } M(\delta) &= 1-\exp\left(-\tfrac{\delta+\theta_0-\theta_1}{\alpha_1}\right) \text{ and } y_{\inf,\delta} = \theta_0+\delta. \\
F^U(\delta) &= 1+\min\{m(\delta), 0\} = 1 \\
\text{since } m(\delta) &= 0 \text{ and } y_{\sup,\delta} = \infty.
\end{aligned}$$

2. Suppose $\alpha_1 = \alpha_0 = \alpha$. Then

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= \begin{cases} 0 \text{ if } \delta \leq \theta_1-\theta_0 \\ 1-\exp\left(-\tfrac{\delta-(\theta_1-\theta_0)}{\alpha}\right) > 0 \text{ if } \delta > \theta_1-\theta_0 \end{cases} \\
\text{and } y_{\inf,\delta} &= \begin{cases} \infty \text{ if } \delta < \theta_1-\theta_0 \\ \text{any point in } \mathcal{R} \text{ if } \delta = \theta_1-\theta_0 \\ \theta_0+\delta \text{ if } \delta > \theta_1-\theta_0 \end{cases}. \\
F^U(\delta) &= 1+\min\{m(\delta), 0\} = 1+m(\delta), \\
\text{where } m(\delta) &= \begin{cases} \exp\left(-\tfrac{\theta_1-(\delta+\theta_0)}{\alpha}\right)-1 < 0 \text{ if } \delta < \theta_1-\theta_0 \\ 0 \text{ if } \delta \geq \theta_1-\theta_0 \end{cases} \\
\text{and } y_{\sup,\delta} &= \begin{cases} \theta_1 \text{ if } \delta < \theta_1-\theta_0 \\ \text{any point in } \mathcal{R} \text{ if } \delta = \theta_1-\theta_0 \\ \infty \text{ if } \delta > \theta_1-\theta_0 \end{cases}.
\end{aligned}$$

3. Suppose $\alpha_1 > \alpha_0$.

(a) If $\delta < \delta_c$,

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = 0 \text{ , since } M(\delta) = 0 \text{ and } y_{\inf,\delta} = \infty. \\
F^U(\delta) &= 1+\min\{m(\delta), 0\} = 1-m(\delta), \\
\text{where } m(\delta) &= \exp\left(-\frac{\theta_1-(\delta+\theta_0)}{\alpha_0}\right)-1 < 0,\ \ y_{\sup,\delta} = \theta_1.
\end{aligned}$$

(b) If $\delta \geq \delta_c$,

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= \max\left\{\exp\left(-\tfrac{\max\{\theta_1-(\delta+\theta_0),0\}}{\alpha_0}\right) - \exp\left(-\tfrac{\max\{\theta_0+\delta-\theta_1,0\}}{\alpha_1}\right), 0\right\} \\
\text{and } y_{\inf,\delta} &= \max\{\theta_1, \theta_0+\delta\} \text{ or } \infty \text{ (boundary solution).} \\
F^U(\delta) &= 1+\min\{m(\delta), 0\} = 1+m(\delta), \\
\text{where } m(\delta) &= \left(\left(\frac{\alpha_0}{\alpha_1}\right)^{\frac{\alpha_1}{\alpha_1-\alpha_0}} - \left(\frac{\alpha_0}{\alpha_1}\right)^{\frac{\alpha_0}{\alpha_1-\alpha_0}}\right)\exp\left(-\frac{\delta-(\theta_1-\theta_0)}{\alpha_1-\alpha_0}\right) < 0 \\
\text{and } y_{\sup,\delta} &= \frac{\alpha_0\alpha_1(\ln\alpha_1-\ln\alpha_0)+\alpha_1\theta_0-\alpha_0\theta_1+\alpha_1\delta}{\alpha_1-\alpha_0} \text{ (an interior solution).}
\end{aligned}$$

**(Shifted) Pareto marginals.** The marginal distributions are:

$$\begin{aligned}
F_1(y) &= 1-\left(\frac{\lambda_1}{\lambda_1+y-\theta_1}\right)^{\alpha} \text{ for } y \in [\theta_1, \infty) \text{ and} \\
F_0(y) &= 1-\left(\frac{\lambda_0}{\lambda_0+y-\theta_0}\right)^{\alpha} \text{ for } y \in [\theta_0, \infty), \text{ where } \alpha, \lambda_1, \theta_1, \lambda_0, \theta_0 > 0.
\end{aligned}$$

Define

$$\delta_c = (\theta_1-\theta_0) - (\max\{\lambda_1, \lambda_0\})^{\frac{\alpha}{\alpha+1}}\left(\lambda_1^{\frac{1}{\alpha+1}} - \lambda_0^{\frac{1}{\alpha+1}}\right).$$

1. Suppose $\lambda_1 < \lambda_0$.

(a) If $\delta \leq \delta_c$, then

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= \left(\lambda_0^{\frac{\alpha}{\alpha+1}} - \lambda_1^{\frac{\alpha}{\alpha+1}}\right)\left(\frac{\lambda_1^{\frac{\alpha}{\alpha+1}} - \lambda_0^{\frac{\alpha}{\alpha+1}}}{\delta-\lambda_0+\lambda_1-\theta_1+\theta_0}\right)^{\alpha} > 0 \\
\text{and } y_{\inf,\delta} &= \frac{(\delta+\theta_0-\lambda_0)\lambda_1^{\frac{\alpha}{\alpha+1}} + (\lambda_1-\theta_1)\lambda_0^{\frac{\alpha}{\alpha+1}}}{\lambda_1^{\frac{\alpha}{\alpha+1}} - \lambda_0^{\frac{\alpha}{\alpha+1}}} \text{ (an interior solution).} \\
F^U(\delta) &= 1+\min\{m(\delta), 0\} = 1+m(\delta), \\
\text{where } m(\delta) &= \min\left\{\left(\frac{\lambda_0}{\lambda_0+\max\{\theta_1-\delta-\theta_0, 0\}}\right)^{\alpha} - \left(\frac{\lambda_1}{\lambda_1+\max\{\theta_0+\delta-\theta_1, 0\}}\right)^{\alpha}, 0\right\} \\
\text{and } y_{\sup,\delta} &= \max\{\theta_1, \theta_0+\delta\} \text{ or } \infty \text{ (boundary solution).}
\end{aligned}$$

(b) If $\delta > \delta_c$, then

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= 1-\left(\frac{\lambda_1}{\lambda_1+\theta_0+\delta-\theta_1}\right)^{\alpha} \geq 0 \text{ and } y_{\inf,\delta} = \theta_0+\delta. \\
F^U(\delta) &= 1+\min\{m(\delta), 0\} = 1 \\
\text{since } m(\delta) &= 0 \text{ and } y_{\sup,\delta} = \infty.
\end{aligned}$$

2. Suppose $\lambda_1 = \lambda_0 = \lambda$. Then

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= \begin{cases} 0 \text{ if } \delta \le \theta_1 - \theta_0 \\ 1 - \left(\dfrac{\lambda}{\lambda + \delta - (\theta_1 - \theta_0)}\right)^{\alpha} \ge 0 \text{ if } \delta > \theta_1 - \theta_0 \end{cases} \\
\text{and } y_{\inf,\delta} &= \begin{cases} \infty \text{ if } \delta < \theta_1 - \theta_0 \\ \text{any point in } \mathcal{Y} \text{ if } \delta = \theta_1 - \theta_0 \\ \theta_0 + \delta \text{ if } \delta > \theta_1 - \theta_0 \end{cases}. \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 + m(\delta), \\
\text{where } m(\delta) &= \begin{cases} \left(\dfrac{\lambda}{\lambda - \delta + (\theta_1 - \theta_0)}\right)^{\alpha} - 1 \text{ if } \delta < \theta_1 - \theta_0 \\ 0 \text{ if } \delta \ge \theta_1 - \theta_0 \end{cases} \\
\text{and } y_{\sup,\delta} &= \begin{cases} \theta_1 \text{ if } \delta < \theta_1 - \theta_0 \\ \text{any point in } \mathcal{Y} \text{ if } \delta = \theta_1 - \theta_0 \\ \infty \text{ if } \delta > \theta_1 - \theta_0 \end{cases}.
\end{aligned}$$

3. Suppose $\lambda_1 > \lambda_0$.

   (a) If $\delta < \delta_c$, then

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = 0 \text{ since } M(\delta) = 0 \text{ and } y_{\inf,\delta} = \infty. \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 + m(\delta), \\
\text{where } m(\delta) &= \left(\frac{\lambda_0}{\lambda_0 + \theta_1 - \delta - \theta_0}\right)^{\alpha} - 1 \le 0 \text{ and } y_{\sup,\delta} = \theta_1.
\end{aligned}$$

   (b) If $\delta \ge \delta_c$, then

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= \max\left\{\left(\frac{\lambda_0}{\lambda_0 + \max\{\theta_1 - \delta - \theta_0, 0\}}\right)^{\alpha} - \left(\frac{\lambda_1}{\lambda_1 + \max\{\theta_0 + \delta - \theta_1, 0\}}\right)^{\alpha}, 0\right\} \\
\text{and } y_{\inf,\delta} &= \max\{\theta_1, \theta_0 + \delta\} \text{ or } \infty \text{ (boundary solution)}. \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 + m(\delta), \\
\text{where } m(\delta) &= \left(\lambda_0^{\frac{\alpha}{\alpha+1}} - \lambda_1^{\frac{\alpha}{\alpha+1}}\right)\left(\frac{\lambda_1^{\frac{\alpha}{\alpha+1}} - \lambda_0^{\frac{\alpha}{\alpha+1}}}{\delta - \lambda_0 + \lambda_1 - \theta_1 + \theta_0}\right)^{\alpha} < 0 \\
\text{and } y_{\sup,\delta} &= \frac{(\delta + \theta_0 - \lambda_0)\lambda_1^{\frac{\alpha}{\alpha+1}} + (\lambda_1 - \theta_1)\lambda_0^{\frac{\alpha}{\alpha+1}}}{\lambda_1^{\frac{\alpha}{\alpha+1}} - \lambda_0^{\frac{\alpha}{\alpha+1}}} \text{ (an interior solution)}.
\end{aligned}$$

## References


[1] Aakvik, A. , J. Heckman, and E. Vytlacil (2003). "Treatment Effects for Discrete Outcomes When Responses to Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs." Forthcoming in *Journal of Econometrics*.

[2] Abadie, A., J. Angrist, and G. Imbens (2002). "Instrumental Variables Estimation of Quantile Treatment Effects." *Econometrica* 70, 91-117.

[3] Alsina, C. (1981). "Some Functional Equations in the Space of Uniform Distribution Functions." *Equationes Mathematicae* 22, 153-164.

[4] Andrews, D. W. K. (2000). "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space." *Econometrica* 68, 399-405.

[5] Beran, R. (1997). "Diagnosing Bootstrap Success." *Ann. Inst. Statist. Math.* 49, 1-24.

[6] Bickel, P. J., F. Götze and W. R. Zwet (1997). "Resampling Fewer than $n$ Observations: Gains, Losses, and Remedies for Losses." *Statistica Sinica* 7, 1-31.

[7] Bickel, P. J., F. and A. Sakov (2005). "On the Choice of $m$ in the $m$ out of $n$ Bootstrap and its Application to Confidence Bounds for Extreme Percentiles." Working paper.

[8] Biddle, J., L. Boden and R. Reville (2003). "A Method for Estimating the Full Distribution of a Treatment Effect, With Application to the Impact of Workfare Injury on Subsequent Earnings." Mimeo.

[9] Bitler, M., J. Gelbach, and H. W. Hoynes (2006). "What Mean Impact Miss: Distributional Effects of Welfare Reform Experiments." Forthcoming in *American Economic Review*.

[10] Black, D. A. , J. A. Smith, M. C. Berger and B. J. Noel (2003). "Is the Threat of Reemployment Services More Effective Than the Services Themselves? Experimental Evidence From the UI System." *American Economic Review* 93(3), 1313-1327.

[11] Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2006). "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds." Forthcoming in *Econometrica*.

[12] Cambanis, S., G. Simons and W. Stout (1976), "Inequalities for $\mathcal{E}k(X,Y)$ when the Marginals are Fixed." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 36, 285-294.

[13] Carneiro, P. , K. T. 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(2), 361-422.

[14] Chen, X., H. Hong, and A. Tarozzi (2004). "Semiparametric Efficiency in GMM Models of Nonclassical Measurement Errors, Missing Data and Treatment Effects." Working paper.

[15] Chernozhukov, V. and C. Hansen (2005). "An IV Model of Quantile Treatment Effects." *Econometrica* 73, 245-261.

[16] Chernozhukov, V., H. Hong and E. Tamer (2004). "Inference on Parameter Sets in Econometric Models." Working paper.

[17] Davidson, R. and J. G. Mackinnon (2004). *Econometric Theory and Method*. Oxford University Press.

[18] Dehejia, R. (1997). "A Decision-theoretic Approach to Program Evaluation." Ph.D. Dissertation, Department of Economics, Harvard University.

[19] Dehejia, R. and S. Wahba (1999). "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94, 1053-1062.

[20] Denuit, M. , C. Genest, and E. Marceau (1999). "Stochastic Bounds on Sums of Dependent Risks." *Insurance: Mathematics and Economics* 25, 85-104.

[21] Djebbari, H. and J. A. Smith (2004). "Heterogeneous Program Impacts in PROGRESA." Mimeo.

[22] Doksum, K. (1974). "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case." *Annals of Statistics* 2, 267-277.

[23] Embrechts, P., A. Hoeing, and A. Juri (2003). "Using Copulae to Bound the Value-at-Risk for Functions of Dependent Risks." *Finance & Stochastics* 7(2), 145-167.

[24] Fan, Y. (2006). "Statistical Inference on the Frechet-Hoeffding Distribution Bounds." Mimeo.

[25] Firpo, S. (2005). "Efficient Semiparametric Estimation of Quantile Treatment Effects." Forthcoming in *Econometrica*.

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

[27] Hahn, J. (1998). "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects." *Econometrica* 66, 315-331.

[28] Heckman, J. and R. Robb (1985). "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*. New York: Cambridge University Press.

[29] Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998). "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66, 1017-1098.

[30] Heckman, J. and J. Smith (1993). "Assessing The Case For Randomized Evaluation of Social Programs," in *Measuring Labour Market Measures: Evaluating the Effects of Active Labour Market Policies*, ed. by K. Jensen and P. K. Madsen. Copenhagen: Danish Ministry of Labor, 35-96.

[31] 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.

[32] Hirano, K., G. W. Imbens, and G. Ridder (2000). "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *NBER Technical Working Papers* 0251, National Bureau of Economic Research, Inc.

[33] Honore, Bo E. and A. Lleras-Muney (2006). "Bounds in Competing Risks Models and the War on Cancer." *Econometrica* 74, 1675-1698.

[34] Horowitz, J. L. and C. F. Manski (1995). "Identification and Robustness with Contaminated and Corrupted Data." *Econometrica* 63, 281-302.

[35] Horowitz, J. L. and C. F. Manski (2000). "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data." *Journal of the American Statistical Association* 95, 77-84.

[36] Imbens, G. W. and C. F. Manski (2004). "Confidence Intervals For Partially Identified Parameters." *Econometrica* 72, 1845–1857.

[37] Imbens, G. W. and W. Newey (2005). "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity." Working Paper.

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

[39] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, London.

[40] Lalonde, R. (1995). "The Promise of Public-Sector Sponsored Training Programs." *Journal of Economic Perspectives* 9, 149-168.

[41] Lechner, M. (1999). "Earnings and Employment Effects of Continuous Off-the-job Training in East Germany After Unification." *Journal of Business and Economic Statistics* 17, 74-90.

[42] Lee, L. F. (2002). "Correlation Bounds for Sample Selection Models with Mixed Continuous, Discrete and Count Data Variables." Manuscript, The Ohio State University.

[43] Lee, M.-J. (2005). *Micro-Econometrics for Policy, Program, and Treatment Effects*. Oxford University Press.

[44] Lehmann, E. L. (1974). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day Inc., San Francisco, California.

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

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

[47] Manski, C. F. (1994). "The Selection Problem," in *Advances in Econometrics, Sixth World Congress* Vol 1, Editor C. Sims, Cambridge University Press.

[48] Manski, C. F. (1997a). "Monotone Treatment Effect." *Econometrica* 65, 1311-1334.

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

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

[51] Manski, C. F. and J. Pepper (2000). "Monotone Instrumental Variables: With Application to the Returns to Schooling." *Econometrica* 68, 997-1010.

[52] McNeil, A., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton Series in Finance, Springer Boston.

[53] Nelsen, R. B. (1999). *An Introduction to Copulas*. Springer, New York.

[54] Politis, D. N. and J. P. Romano (1994). "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions." *Annals of Statistics* 22, 2031-2050.

[55] Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling*. Springer-Verlag, New York.

[56] Romano, J. and A. M. Shaikh (2006). "Inference for Identifiable Parameters in Partially Identified Econometric Models." Working Paper.

[57] Rosenbaum, P. R. and D. B. Rubin (1983a). "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome." *Journal of the Royal Statistical Society, Series B* 45, 212-218.

[58] Rosenbaum, P. R. and D. B. Rubin (1983b). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70, 41-55.

[59] Rüschendorf, L. (1982). "Random Variables With Maximum Sums." *Advances in Applied Probability* 14, 623-632.

[60] Schweizer, B. and A. Sklar (1983). *Probabilistic Metric Spaces*. North-Holland, New York.

[61] Shaikh, A. M. and E. Vytlacil (2005). "Threshold Crossing Models and Bounds on Treatment Effects: A Nonparametric Analysis." Working Paper.

[62] Sherman, R. P. (2003). "Some Asymptotic Results for Bounds Estimation." Working paper.

[63] Sklar A. (1959). "Fonctions de réartition à n dimensions et leures marges," *Publications de l'Institut de Statistique de L'Université de Paris* 8, 229-231.

[64] Stoye, J. (2005). "Partial Identification of Spread Parameters." Working paper.

[65] Tchen, A. H. (1980). "Inequalities for Distributions with Given Marginals." *Annals of Probability* 8, 814-827.

[66] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

[67] van der Vaart , A. W. and Jon A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.

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

[69] Wu, C. F. J. (1990). "On the Asymptotic Properties of the Jackknife Histogram." *Annals of Statistics* 18, 1438-1452.