Partial identification of the distribution of treatment effects and its confidence sets

Fan, Yanqin and Park, Sang Soo

Vanderbilt University, University of North Carolina at Chapel Hill

2009

PARTIAL IDENTIFICATION OF THE DISTRIBUTION OF TREATMENT EFFECTS AND ITS CONFIDENCE SETS

Yanqin Fan and Sang Soo Park

ABSTRACT

In this paper, we study partial identification of the distribution of treatment effects of a binary treatment for ideal randomized experiments, ideal randomized experiments with a known value of a dependence measure, and for data satisfying the selection-on-observables assumption, respectively. For ideal randomized experiments, (i) we propose nonparametric estimators of the sharp bounds on the distribution of treatment effects and construct asymptotically valid confidence sets for the distribution of treatment effects; (ii) we propose bias-corrected estimators of the sharp bounds on the distribution of treatment effects; and (iii) we investigate finite sample performances of the proposed confidence sets and the bias-corrected estimators via simulation.

Nonparametric Econometric Methods
Advances in Econometrics, Volume 25, 3–70

ISSN: 0731-9053/doi:10.1108/S0731-9053(2009)0000025004

1. INTRODUCTION

Evaluating the effect of a treatment or a social program is important in diverse disciplines including the social and medical sciences. 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 treatment effects. See Lee (2005), Abbring and Heckman (2007), Heckman and Vytlacil (2007a, 2007b) for discussions 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 (2008) 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 the welfare effect of the change from Aid to Families with Dependent Children (AFDC) to Temporary Assistance for Needy Families (TANF) programs. 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), and Abadie, Angrist, and Imbens (2002).

When responses to treatment differ among otherwise observationally equivalent subjects, the entire distribution of the treatment effects or other features of the treatment effects than its mean may be of interest. Two general approaches have been proposed in the literature to study the distribution of treatment effects. In the first approach, the distribution of treatment effects is partially identified, see Manski (1997), Fan and Park (2007a), Fan and Wu (2007), Fan (2008), and Firpo and Ridder (2008). Assuming monotone treatment response, Manski (1997) developed sharp bounds on the distribution of treatment effects, while (i) assuming the availability of ideal randomized data,[1] Fan and Park (2007a) developed estimation and inference tools for the sharp bounds on the distribution of treatment effects and (ii) assuming that data satisfy the selection-on-observables or the strong ignorability assumption, Fan and Park (2007a) and Firpo and Ridder (2008) established sharp bounds on the distribution of treatment effects and Fan (2008) proposed nonparametric estimators of the sharp bounds and constructed asymptotically valid confidence sets (CSs) for the distribution of treatment effects. In the context of switching regimes models, Fan and Wu (2007) studied partial identification and inference for conditional distributions of treatment effects. In the second approach, restrictions are imposed on the dependence structure between the potential outcomes such that distributions of the treatment effects are point identified, see, for example, Heckman et al. (1997), Biddle, Boden, and Reville (2003), Carneiro, Hansen, and Heckman (2003), Aakvik, Heckman, and Vytlacil (2005), and Abbring and Heckman (2007), among others. In addition to the distribution of treatment effects, Fan and Park (2007c) studied partial identification of and inference for the quantile of treatment effects for randomized experiments; Fan and Zhu (2009) investigated partial identification of and inference for a general class of functionals of the joint distribution of potential outcomes including the correlation coefficient between the potential outcomes and many commonly used inequality measures of the distribution of treatment effects under the selection-on-observables assumption. Firpo and Ridder (2008) also presented some partial identification results for functionals of the distribution of treatment effects under the selection-on-observables assumption.

The objective of this paper is threefold. First, this paper provides a review of existing results on partial identification of the distribution of treatment effects in Fan and Park (2007a) and establishes similar results for randomized experiments when the value of a dependence measure between the potential outcomes such as Kendall's τ is known. Second, this paper relaxes two strong assumptions used in Fan and Park (2007a) to derive the asymptotic distributions of nonparametric estimators of sharp bounds on the distribution of treatment effects and constructs asymptotically valid CSs for the distribution of treatment effects. Third, as evidenced in the simulation results presented in Fan and Park (2007a), the simple plug-in nonparametric estimators of the sharp bounds on the distribution of treatment effects tend to have upward/downward bias in finite samples. In this paper, we confirm this analytically and construct bias-corrected estimators of these bounds. We present an extensive simulation study of finite sample performances of the proposed CSs and of the bias-corrected estimators. The issue of constructing CSs for the distribution of treatment effects belongs to the recently fast growing area of inference for partially identified parameters, see for example, Imbens and Manski (2004), Bugni (2007), Canay (2007), Chernozhukov, Hong, and Tamer (2007), Galichon and Henry (2006), Horowitz and Manski (2000), Romano and Shaikh (2008), Stoye (2008), Rosen (2008), Soares (2006), Beresteanu and Molinari (2006), Andrews and Guggenberger (2007), Andrews and Soares (2007), Fan and Park (2007b), and Moon and Schorfheide (2007). Like Fan and Park

(2007c), we follow the general approach developed in Andrews and Guggenberger (2005a, 2005b, 2005c, 2007) for nonregular models.

The rest of this paper is organized as follows. In Section 2, we review sharp bounds on the distribution of treatment effects and related results for randomized experiments in Fan and Park (2007a). In Section 3, we present improved bounds when additional information is available. In Section 4, we first revisit the nonparametric estimators of the distribution bounds proposed in Fan and Park (2007a) and their asymptotic properties. Motivated by the restrictive nature of the unique, interior assumption of the sup and inf in Fan and Park (2007a), we then provide asymptotic properties of the estimators with a weaker assumption. Section 5 constructs asymptotically valid CSs for the bounds and the true distribution of treatment effects under much weaker assumptions than those in Fan and Park (2007a). Section 6 provides bias-corrected estimators of the sharp bounds in Fan and Park (2007a). Results from an extensive simulation study are provided in Section 7. Section 8 concludes. Some technical proofs are collected in Appendix A. Appendix B presents expressions for the sharp bounds on the distribution of treatment effects in Fan and Park (2007a) for certain known marginal distributions.

Throughout the paper, we use $\Rightarrow$ to denote weak convergence. All the limits are taken as the sample size goes to ∞.

2. SHARP BOUNDS ON THE DISTRIBUTION OF TREATMENT EFFECTS AND BOUNDS ON ITS *D*-PARAMETERS FOR RANDOMIZED EXPERIMENTS

In this section, we review the partial identification results in Fan and Park (2007a). Consider a randomized experiment with a binary treatment and continuous outcomes. Let Y_1 denote the potential outcome from receiving the treatment and Y_0 the potential outcome without receiving the 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. It is well known that with randomized data, the marginal distribution functions $F_1(\cdot)$ and $F_0(\cdot)$ are identified, but the joint distribution function $F(y_1, y_0)$ is not identified. 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 treatment effects when the marginal distributions are identified.

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, that is, $C^{L}(u, v) \leq C(u, v) \leq C^{U}(u, v)$. Then for any (y_1, y_0), the following inequality holds:

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

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.

For randomized experiments, the marginals F_1 and F_0 are identified and Eq. (1) partially identifies $F(y_1, y_0)$. See Heckman and Smith (1993), Heckman et al. (1997), Manski (1997b), and Fan and Wu (2007) for applications of Eq. (1) in the context of program evaluation. Lee (2002) used Eq. (1) to bound correlation coefficients in sample selection models.

2.1. Sharp Bounds on the Distribution of Treatment Effects

Let $\Delta = Y_1 - Y_0$ denote the individual treatment effect and $F_{\Delta}(\cdot)$ its distribution function. For randomized experiments, the marginals F_1 and F_0 are identified. Given F_1 and F_0, sharp bounds on the distribution of Δ can be found in Williamson and Downs (1990).

Lemma 1. Let

$$F^{L}(\delta) = \max\left(\sup_{y}\{F_1(y) - F_0(y - \delta)\}, 0\right) \text{ and}$$

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

Then $F^{L}(\delta) \leq F_{\Delta}(\delta) \leq F^{U}(\delta)$.

At any given value of δ, the bounds $(F^{\mathrm{L}}(\delta), F^{\mathrm{U}}(\delta))$ are informative on the value of $F_{\Delta}(\delta)$ as long as $[F^{\mathrm{L}}(\delta), F^{\mathrm{U}}(\delta)] \subset [0,1]$ in which case, we say $F_{\Delta}(\delta)$ is partially identified. Viewed as an inequality among all possible distribution functions, the sharp bounds $F^{\mathrm{L}}(\delta)$ and $F^{\mathrm{U}}(\delta)$ 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 δ, we have $F^{\mathrm{L}}(\delta) = F_{\Delta}(\delta) = F^{\mathrm{U}}(\delta)$. In fact, given any pair of distribution functions F_1 and F_0, the inequality: $F^{\mathrm{L}}(\delta) \leq F_{\Delta}(\delta) \leq F^{\mathrm{U}}(\delta)$ cannot be improved, that is, the bounds $F^{\mathrm{L}}(\delta)$ and $F^{\mathrm{U}}(\delta)$ 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.

Let $\gtrsim_{\mathrm{FSD}}$ and $\gtrsim_{\mathrm{SSD}}$ denote the first-order and second-order stochastic dominance relations, that is, for two distribution functions G and H,

$$G \gtrsim_{\mathrm{FSD}} H \text{ iff } G(x) \leq H(x) \text{ for all } x$$

$$G \gtrsim_{\mathrm{SSD}} H \text{ iff } \int_{-\infty}^{x} G(v)dv \leq \int_{-\infty}^{x} H(d)dv \text{ for all } x$$

Lemma 1 implies: $F^{\mathrm{L}} \gtrsim_{\mathrm{FSD}} F_{\Delta} \gtrsim_{\mathrm{FSD}} F^{\mathrm{U}}$. We note that unlike sharp bounds on the joint distribution of Y_1, Y_0, sharp bounds on the distribution of Δ 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 Δ' 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 Δ'' 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 $F_{\Delta}, F_{\Delta'}, F_{\Delta''}$.

Lemma 2. Let $F_{\Delta}, F_{\Delta'}, F_{\Delta''}$ be defined as above. Then $F_{\Delta'} \gtrsim_{\mathrm{SSD}} F_{\Delta} \gtrsim_{\mathrm{SSD}} F_{\Delta''}$.

Theorem 1 in Stoye (2008b), see also Tesfatsion (1976), shows that $F_{\Delta'} \gtrsim_{\text{SSD}} F_{\Delta}$ is equivalent to $E[U(\Delta')] \leq E[U(\Delta)]$ or $E[U(Y_1' - Y_0')] \leq E[U(Y_1 - Y_0)]$ for every convex real-valued function U. Corollary 2.3 in Tchen (1980) implies the conclusion of Lemma 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 "*D*-parameters" is any quantile of the distribution. Stoye (2008b) 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 (2008b). He refers to the class of "*D*-parameters" as the class of "D_1-parameters."

Definition 1. A population statistic θ is a D_1-parameter, if it increases weakly with first-order stochastic dominance, that is, $F \gtrsim_{\text{FSD}} G$ implies $\theta(F) \geq \theta(G)$.

Obviously if θ is a D_1-parameter, then Lemma 1 implies: $\theta(F^{\text{L}}) \geq \theta(F_{\Delta}) \geq \theta(F^{\text{U}})$. In general, the bounds $\theta(F^{\text{L}}), \theta(F^{\text{U}})$ on a D_1-parameter may not be sharp, as the bounds in Lemma 1 are point-wise sharp, but not uniformly sharp, see Firpo and Ridder (2008) for a detailed discussion on this issue. In the special case where θ is a quantile of the treatment effect distribution, the bounds $\theta(F^{\text{L}}), \theta(F^{\text{U}})$ are known to be sharp and can be expressed in terms of the quantile functions of the marginal distributions of the potential outcomes. Specially, let $G^{-1}(u)$ denote the generalized inverse of a nondecreasing function G, that is, $G^{-1}(u) = \inf\{x | G(x) \geq u\}$. Then Lemma 1 implies: for $0 \leq q \leq 1, (F^{\text{U}})^{-1}(q) \leq F_{\Delta}^{-1}(q) \leq (F^{\text{L}})^{-1}(q)$ and the bounds are known to be sharp. 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 treatment effects, see Williamson and Downs (1990).

Lemma 3. For $0 \leq q \leq 1, (F^{\mathrm{U}})^{-1}(q) \leq F_{\Delta}^{-1}(q) \leq (F^{\mathrm{L}})^{-1}(q)$, where

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

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

Like sharp bounds on the distribution of treatment effects, sharp bounds on the quantile function of Δ 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 treatment effects when the potential outcomes are either perfectly positively dependent or perfectly negatively dependent.

Lemma 4. For $q \in [0,1]$, we have (i) $F_{\Delta'}^{-1}(q) = [F_1^{-1}(q) - F_0^{-1}(q)]$ if $[F_1^{-1}(q) - F_0^{-1}(q)]$ is an increasing function of q; (ii) $F_{\Delta''}^{-1}(q) = [F_1^{-1}(q) - F_0^{-1}(1-q)]$.

The proof of Lemma 4 follows that of the proof 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 φ increasing and left continuous on the range of Z, it holds that the quantile of $\varphi(Z)$ at quantile level q is given by $\varphi(F_Z^{-1}(q))$, where F_Z is the distribution function of Z. For (i), we note that $F_{\Delta'}^{-1}(q)$ equals the quantile of $[F_1^{-1}(U) - F_0^{-1}(U)]$, 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, $[F_1^{-1}(q) - F_0^{-1}(q)]$ can fail to be a quantile function. Doksum (1974) and Lehmann (1974) used $[F_1^{-1}(F_0(y_0)) - y_0]$ to measure treatment effects. Recently, $[F_1^{-1}(q) - F_0^{-1}(q)]$ has been used to study treatment effects heterogeneity and is referred to as the quantile treatment effects (QTE), see for example, Heckman et al. (1997), Abadie et al. (2002), Chernozhukov and Hansen (2005), Firpo (2007), Firpo and Ridder (2008), and Imbens and Newey (2005), among others, for more discussion and references on the estimation of QTE. Manski (1997a) referred to QTE as ΔD-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 4 (i) shows that QTE equals the quantile function of the treatment effects 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 effects 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 3. At any given quantile level, the lower quantile bound $(F^{\mathrm{U}})^{-1}(q)$ is the smallest outcome gain (worst case) regardless of the dependence structure between the potential outcomes and should be useful to policy makers. For example, $(F^{\mathrm{U}})^{-1}(0.5)$ is the minimum gain of at least half of the population.

Definition 2. A population statistic θ is a D_2-parameter, if it increases weakly with second-order stochastic dominance, that is, $F \gtrsim_{\mathrm{SSD}} G$ implies $\theta(F) \geq \theta(G)$.

If θ is a D_2-parameter, then Lemma 2 implies $\theta(F_{\Delta'}) \leq \theta(F_{\Delta}) \leq \theta(F_{\Delta''})$. Stoye (2008) defined the class of D_2-parameters in terms of mean-preserving spread. Since the mean of Δ 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, the above bounds on D_2-parameters of the treatment effect distribution are reached when the potential outcomes are perfectly dependent on each other and they are known to be sharp. For a general functional of F_{Δ}, Firpo and Ridder (2008) investigated the possibility of obtaining its bounds that are tighter than the bounds implied by F^{L}, F^{U}. Here we point out that for the class of D_2-parameters of F_{Δ}, their sharp bounds are available. One example of D_2-parameters is the variance of the treatment effect Δ. Using results in Cambanis et al. (1976), Heckman et al. (1997) provided sharp bounds on the variance of Δ for randomized experiments and proposed a test for the common effect model by testing the value of the lower bound of the variance of Δ. Stoye (2008) 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 treatment effects and its quantiles when $Y_1 \sim N(\mu_1, \sigma_1^2)$ and

$Y_0 \sim N(\mu_0, \sigma_0^2)$. In addition, we provide explicit expressions for the distribution of treatment effects and its quantiles when the potential outcomes are perfectly positively dependent, perfectly negatively dependent, and independent.

2.3.1. Distribution Bounds

Explicit expressions for sharp 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 et al. (1987), and Denuit, Genest, and Marceau (1999). Using Lemma 1, 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

$$\sup_y \{F_1(y) - F_0(y-\delta)\} = 2\Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{2\sigma}\right) - 1 \text{ for } \delta > \mu_1 - \mu_0,$$

$$\inf_y \{F_1(y) - F_0(y-\delta)\} = 2\Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{2\sigma}\right) - 1 \text{ for } \delta < \mu_1 - \mu_0$$

Hence,

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

$$F^{\mathrm{U}}(\delta) = \begin{cases} 2\Phi\left(\dfrac{\delta - (\mu_1 - \mu_0)}{2\sigma}\right) & \text{if } \delta < \mu_1 - \mu_0 \\ 1, & \text{if } \delta \geq \mu_1 - \mu_0 \end{cases} \tag{3}$$

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

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

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

where $s = \delta - (\mu_1 - \mu_0)$ and $t = \sqrt{s^2 + (\sigma_1^2 - \sigma_0^2)\ \ln(\sigma_1^2/\sigma_0^2)}$. For any δ, 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,

$$F^{\mathrm{L}}(\delta) = \Phi\left(\frac{\sigma_1 s - \sigma_0 t}{\sigma_1^2 - \sigma_0^2}\right) + \Phi\left(\frac{\sigma_1 t - \sigma_0 s}{\sigma_1^2 - \sigma_0^2}\right) - 1$$

$$F^{\mathrm{U}}(\delta) = \Phi\left(\frac{\sigma_1 s + \sigma_0 t}{\sigma_1^2 - \sigma_0^2}\right) + \Phi\left(\frac{\sigma_1 t + \sigma_0 s}{\sigma_1^2 - \sigma_0^2}\right) + 1$$

For comparison purposes, we provide expressions for the distribution F_Δ in three special cases.

Case I. Perfect positive dependence. In this case, Y_0 and Y_1 satisfy $Y_0 = \mu_0 + (\sigma_0/\sigma_1)Y_1 - (\sigma_0/\sigma_1)\mu_1$. Therefore,

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

If $\sigma_1 = \sigma_0$, then

$$F_\Delta(\delta) = \begin{cases} 0 \text{ and } \delta < \mu_1 - \mu_0 \\ 1 \text{ and } \mu_1 - \mu_0 \leq \delta \end{cases} \tag{4}$$

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 $Y_0 = \mu_0 - (\sigma_0/\sigma_1)Y_1 + (\sigma_0/\sigma_1)\mu_1$. Hence,

$$\Delta = \frac{\sigma_1 + \sigma_0}{\sigma_1} Y_1 - \left(\frac{\sigma_0}{\sigma_1}\mu_1 + \mu_0\right)$$

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

Case III. Independence. This yields

$$F_\Delta(\delta) = \Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{\sqrt{\sigma_1^2 + \sigma_0^2}}\right) \tag{5}$$

Fig. 1 below plots the bounds on the distribution F_Δ (denoted by F_L and F_U) and the distribution F_Δ 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.

First, we observe from Fig. 1 that the bounds in this case are informative at all values of δ and are more informative in the tails of the distribution F_Δ than in the middle. In addition, Fig. 1 indicates that the distribution of the treatment effects for perfectly positively dependent potential outcomes is most concentrated around its mean 1 implied by the second-order stochastic

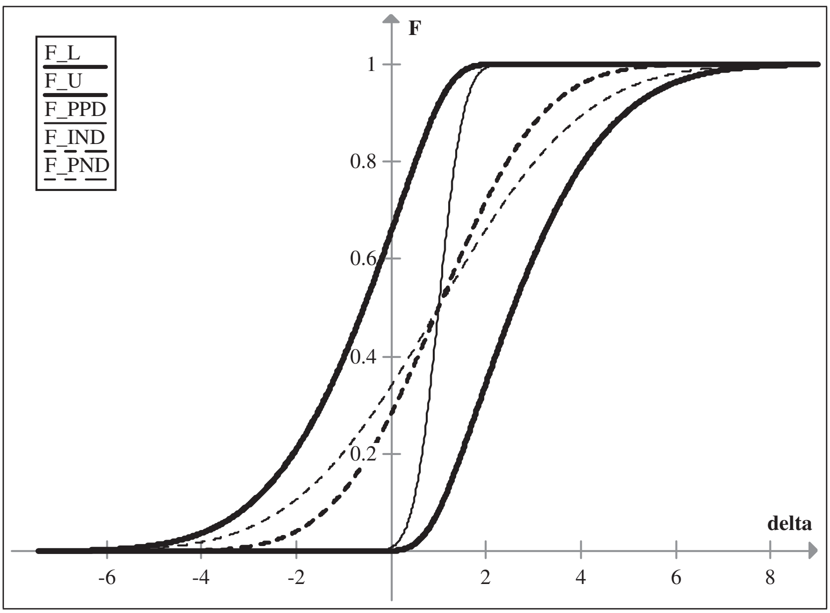


Fig. 1. Bounds on the Distribution of the Treatment Effect: (N(2,2), N(1,1)).

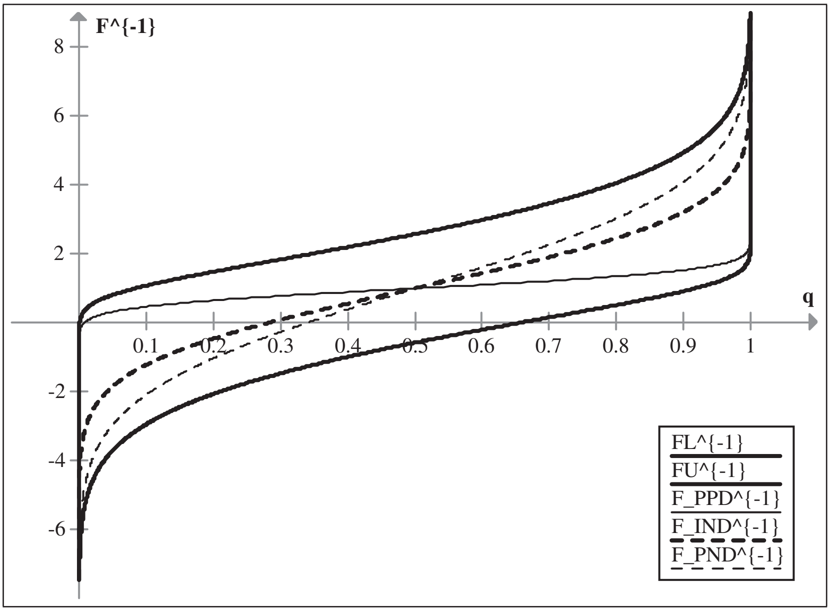


Fig. 2. Bounds on the Quantile Function of the Treatment Effect: (N(2,2), N(1,1)).

relation F_PPD$\gtrsim_{\text{SSD}}$F_IND$\gtrsim_{\text{SSD}}$F_PPD. 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 Fig. 2 above.

2.3.2. Quantile Bounds

By inverting Eqs. (2) and (3), we obtain the quantile bounds for the case $\sigma_1 = \sigma_0 = \sigma$:

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

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

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 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 $\sigma_1 > \sigma_0$, that is, only if the condition of Lemma 4 (i) holds. If $\sigma_1 < \sigma_0$, $[F_1^{-1}(q) - F_0^{-1}(q)]$ is a decreasing function of q and hence cannot 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 Fig. 2 below, we plot the quantile bounds for Δ (FL^{−1} and FU^{−1}) when $Y_1 \sim N(2, 2)$ and $Y_0 \sim N(1, 1)$ and the quantile functions of Δ 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).

Again, Fig. 2 reveals the fact that the quantile function of Δ 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 Δ in the perfect positive dependence case is the same as QTE. Fig. 2 leads to the conclusion that QTE is a conservative measure of the degree of heterogeneity of the treatment effect distribution.

3. MORE ON SHARP BOUNDS ON THE JOINT DISTRIBUTION OF POTENTIAL OUTCOMES AND THE DISTRIBUTION OF TREATMENT EFFECTS

For randomized experiments, Eq. (1) and Lemma 1, respectively, provide sharp bounds on the joint distribution of potential outcomes and the distribution of treatment effects. When additional information is available, these bounds are no longer sharp. In this section, we consider two types of additional information. One is the availability of a known value of a dependence measure between the potential outcomes and the other is the availability of covariates ensuring the validity of the selection-on-observables assumption.

3.1. Randomized Experiments with a Known Value of Kendall's τ

In this subsection, we first review sharp bounds on the joint distribution of the potential outcomes Y_1, Y_0 when the value of a dependence measure such as Kendall's τ between the potential outcomes is known. Then we point out how this information can be used to tighten the bounds on the distribution of Δ presented in Lemma 1. We provide details for Kendall's τ and point out relevant references for other measures including Spearman's ρ.

To begin, we introduce the notation used in Nelsen, Quesada-Molina, Rodriguez-Lallena, and Ubeda-Flores (2001). Let (X_1, Y_1), (X_2, Y_2), and (X_3, Y_3) be three independent and identically distributed random vectors of dimension 2 whose joint distribution is H. Kendall's τ and Spearman's ρ are defined as:

$$\tau = \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - \Pr[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

$$\rho = 3\{\Pr[(X_1 - X_2)(Y_1 - Y_3) > 0] - \Pr[(X_1 - X_2)(Y_1 - Y_3) < 0]\}$$

For any $t \in [-1,1]$, let $\mathcal{T}_t$ denote the set of copulas with a common value t of Kendall's τ, that is,

$$\mathcal{T}_t = \{C | C \text{ is a copula such that } \tau(C) = t\}$$

Let $\underline{T}_t$ and $\bar{T}_t$ denote, respectively, the point-wise infimum and supremum of $\mathcal{T}_t$. The following result presents sharp bounds on the joint distribution of the potential outcomes Y_1, Y_0. It can be found in Nelsen et al. (2001).

Lemma 5. Suppose that the value of Kendall's τ between Y_1 and Y_0 is t. Then

$$\underline{T}_t(F_1(y_1), F_0(y_0)) \leq F(y_1, y_0) \leq \bar{T}_t(F_1(y_1), F_0(y_0))$$

where, for any $(u, v) \in [0,1]^2$;

$$\underline{T}_t(u, v) = \max\left(0, u + v - 1, \frac{1}{2}\left[(u + v) - \sqrt{(u - v)^2 + 1 - t}\right]\right)$$

$$\overline{T}_t(u, v) = \min\left(u, v, \frac{1}{2}\left[(u + v - 1) + \sqrt{(u + v - 1)^2 + 1 + t}\right]\right)$$

As shown in Nelsen et al. (2001),

$$\begin{aligned} \underline{T}_t(u, v) &= C^{\mathrm{L}}(u, v) \quad \text{if } t \in [-1, 0] \\ \underline{T}_t(u, v) &\geq C^{\mathrm{L}}(u, v) \quad \text{if } t \in [0, 1] \end{aligned} \tag{6}$$

and

$$\bar{T}_t(u, v) = C^{\mathrm{U}}(u, v) \quad \text{if } t \in [0, 1]$$

$$\bar{T}_t(u, v) \leq C^{\mathrm{U}}(u, v) \quad \text{if } t \in [-1, 0]$$

Hence, for any fixed (y_1, y_0), the bounds $[\underline{T}_t(F_1(y_1), F_0(y_0)), \bar{T}_t(F_1(y_1), F_0(y_0))]$ are in general tighter than the bounds in Eq. (1) unless $t = 0$. The lower bound on $F(y_1, y_0)$ can be used to tighten bounds on the distribution of treatment effects via the following result in Williamson and Downs (1990).

Lemma 6. Let $\underline{C}_{XY}$ denote a lower bound on the copula C_{XY} and F_{X+Y} denote the distribution function of $X + U$. Then

$$\sup_{x+y=z} \underline{C}_{XY}(F(x), G(y)) \leq F_{X+Y}(z) \leq \inf_{x+y=z} \underline{C}^d_{XY}(F(x), G(y))$$

where $\underline{C}^d_{XY}(u, v) = u + v - \underline{C}_{XY}(u, v)$.

Let $Y_1 = X$ and $Y_0 = -Y$ in Lemma 6. By using Lemma 5 and the duality theorem, we can prove the following proposition. AU :3

Proposition 1. Suppose the value of Kendall's τ between Y_1 and Y_0 is t. Then

(i) $\sup_x \underline{T}_{-t}(F_1(x), 1 - F_0(x-\delta)) \leq F_\Delta(\delta) \leq \inf_x \underline{T}^d_{-t}(F_1(x), 1 - F_0(x-\delta))$, where

$$\underline{T}_{-t}(u, v) = \max\left\{0, u + v + 1, \frac{1}{2}\left[(u+v) - \sqrt{(u-v)^2 + 1 + t}\right]\right\}$$

$$\underline{T}^d_{-t}(u, v) = \max\left\{u + v, 1, \frac{1}{2}\left[(u+v) + \sqrt{(u-v)^2 + 1 + t}\right]\right\}$$

(ii) $\sup_{\underline{T}^d_{-t}(u,v)=q}[F_1^{-1}(u) - F_0^{-1}(1-v)] \leq F_\Delta^{-1}(q) \leq \inf_{\underline{T}_{-t}(u,1-v)=q}[F_1^{-1}(u) - F_0^{-1}(1-v)]$.

Proposition 1 and Eq. (6) imply that the bounds in Proposition 1 (i) are sharper than those in Lemma 1 if $t \in [-1, 0]$ and are the same as those in Lemma 1 if $t \in [0, 1]$. This implies that if the potential outcomes Y_1 and Y_0 are positively dependent in the sense of having a nonnegative Kendall's τ, then the information on the value of Kendall's τ does not improve the bounds on the distribution of treatment effects. On contrary, if they are negatively dependent on each other, then knowing the value of Kendall's τ will in general improve the bounds.

Remark 1. If instead of Kendall's τ, the value of Spearman's ρ between the potential outcomes is known, one can also establish tighter bounds on $F_\Delta(z)$ by using Theorem 4 in Nelsen et al. (2001) and Lemma 6.

Remark 2. Other dependence information that may be used to tighten bounds on the joint distribution of potential outcomes and thus the distribution of treatment effects include known values of the copula function of the potential outcomes at certain points, see Nelsen and Ubeda-Flores (2004) and Nelsen, Quesada-Molina, Rodriguez-Lallena, and Ubeda-Flores (2004).

3.2. Selection-on-Observables

In many applications, observations on a vector of covariates for individuals in the treatment and control groups are available. In this subsection, we extend sharp bounds for randomized experiments in Lemma 1 to take into account these covariates. For notational compactness, we let $n = n_1 + n_0$ so that there are n individuals altogether. For $i = 1, \ldots, n$, 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 a treatment or a program, see for example, Rosenbaum and Rubin (1983), Hahn (1998), Heckman, Ichimura, Smith, and Todd (1998), Dehejia and Wahba (1999), and Hirano, Imbens, and Ridder (2003), 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 joint distribution of potential outcomes and the distribution of Δ under (C1) and (C2). For any fixed $x \in \mathcal{X}$, Eq. (1) provides sharp bounds on the conditional joint distribution of Y_1, Y_0 given $X = x$:

$$C^{\mathrm{L}}(F_1(y_1|x), F_0(y_0|x)) \leq F(y_1, y_0|x) \leq C^{\mathrm{U}}(F_1(y_1|x), F_0(y_0|x))$$

and Lemma 1 provides sharp bounds on the conditional distribution of Δ given $X = x$:

$$F^{\mathrm{L}}(\delta|x) \leq F_\Delta(\delta|x) \leq F^{\mathrm{U}}(\delta|x)$$

where

$$F^{\mathrm{L}}(\delta|x) = \sup_y \max(F_1(y|x) - F_0(y - \delta|x), 0)$$

$$F^{\mathrm{U}}(\delta|x) = 1 + \inf_y \min(F_1(y|x) - F_0(y - \delta|x), 0)$$

Here, we use $F_\Delta(\cdot|x)$ to denote the conditional distribution function of Δ 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 $F(y_1, y_0|x)$ and $F_\Delta(\delta|x)$. To see this, note that

$$\begin{aligned} F_1(y|x) &= P(Y_1 \leq y|X = x) = P(Y_1 \leq y|X = x, D = 1) \\ &= P(Y \leq y|X = x, D = 1) \end{aligned} \tag{7}$$

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

$$F_0(y|x) = P(Y \leq y|X = x, D = 0) \tag{8}$$

Sharp bounds on the unconditional joint distribution of Y_1, Y_0 and the unconditional distribution of Δ follow from those of the conditional distributions:

$$E[C^{\mathrm{L}}(F_1(y_1|X), F_0(y_0|X))] \leq F(y_1, y_0) \leq C^{\mathrm{U}}(F_1(y_1|X), F_0(y_0|X))$$

$$E(F^{\mathrm{L}}(\delta|X)) \leq F_\Delta(\delta) = E(F_\Delta(\delta|X)) \leq E(F^{\mathrm{U}}(\delta|X))$$

We note that if X is independent of (Y_1, Y_0), then the above bounds on $F(y_1, y_0)$ and $F_\Delta(\delta)$ reduce, respectively, to those in Eq. (1) and Lemma 1. In general, X is not independent of (Y_1, Y_0) and the above bounds are tighter than those in Eq. (1) and Lemma 1, see Fan (2008) for a more detailed discussion on the sharp bounds with covariates. Under the selection on observables assumption, Fan and Zhu (2009) established sharp bounds on a general class of functionals of the joint distribution $F(y_1, y_0)$ including the correlation coefficient between the potential outcomes and the class of D_2-parameters of the distribution of treatment effects.

4. NONPARAMETRIC ESTIMATORS OF THE SHARP BOUNDS AND THEIR ASYMPTOTIC PROPERTIES FOR RANDOMIZED EXPERIMENTS

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[4] of F_1 and F_0. Note that the bounds in Lemma 1 can be written as:

$$F^{\mathrm{L}}(\delta) = \sup_{y \in \mathcal{R}}\{F_1(y) - F_0(y - \delta)\}, F^{\mathrm{U}}(\delta) = 1 + \inf_{y \in \mathcal{R}}\{F_1(y) - F_0(y - \delta)\} \tag{9}$$

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 $\mathcal{Y}_1 = \mathcal{Y}_0 = \mathcal{R}$, Eq. (9) suggests the following plug-in estimators of $F^{\mathrm{L}}(\delta)$ and $F^{\mathrm{U}}(\delta)$:

$$F_n^{\mathrm{L}}(\delta) = \sup_{y \in \mathcal{R}}\{F_{1n}(y) - F_{0n}(y - \delta)\}, F_n^{\mathrm{U}}(\delta) = 1 + \inf_{y \in \mathcal{R}}\{F_{1n}(y) - F_{0n}(y - \delta)\} \tag{10}$$

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

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

When either $\mathcal{Y}_1$ or $\mathcal{Y}_0$ is not the whole real line, we derive alternative expressions for $F^{\mathrm{L}}(\delta)$ and $F^{\mathrm{U}}(\delta)$ 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 the results for the general case afterwards.

Suppose $\mathcal{Y}_1 = \mathcal{Y}_0 = [0, 1]$. If $1 \geq \delta \geq 0$, then Eq. (9) implies:

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

and

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

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

$$\begin{aligned} F^{\mathrm{L}}(\delta) = & \max\left\{ \sup_{y\in[0,1+\delta]}\{F_1(y) - F_0(y-\delta)\}, \sup_{y\in(-\infty,0)}\{F_1(y) - F_0(y-\delta)\}, \right. \\ & \left. \sup_{y\in(1+\delta,\infty)}\{F_1(y) - F_0(y-\delta)\} \right\} \\ = & \max\left\{ \sup_{y\in[0,1+\delta]}\{F_1(y) - F_0(y-\delta)\}, \sup_{y\in(-\infty,0)}\{-F_0(y-\delta)\}, \right. \\ & \left. \sup_{y\in(1+\delta,\infty)}\{F_1(y) - 1)\} \right\} \\ = & \max\left\{ \sup_{y\in[0,1+\delta]}\{F_1(y) - F_0(y-\delta)\}, 0 \right\} \end{aligned} \tag{12}$$

and

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

Based on Eqs. (11) and (12), we propose the following estimator of $F^{\mathrm{L}}(\delta)$:

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

Similarly, we propose the following estimator for $E^{\mathrm{U}}(\delta)$:

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

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 $a,b,c,d \in \bar{\mathcal{R}} \equiv \mathcal{R} \cup \{-\infty, +\infty\}, a<b, c<d$ with $F_1(a) = F_0(c) = 0$ and $F_1(b) = F_0(d) = 1$. It is easy to see that

$$F^{\mathrm{L}}(\delta) = F^{\mathrm{U}}(\delta) = 0, \quad \text{if } \delta \le a-d \quad \text{and} \quad F^{\mathrm{L}}(\delta) = F^{\mathrm{U}}(\delta) = 1, \quad \text{if } \delta \ge b-c$$

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

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

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

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

$$F_n^{\mathrm{L}}(\delta) = \max\left\{ \sup_{y \in \mathcal{Y}_\delta} \{F_{1n}(y) - F_{0n}(y-\delta)\}, 0 \right\} \tag{13}$$

$$F_n^{\mathrm{U}}(\delta) = 1 + \min\left\{ \inf_{y \in \mathcal{Y}_\delta} \{F_{1n}(y) - F_{0n}(y-\delta)\}, 0 \right\} \tag{14}$$

By using $F_n^{\mathrm{L}}(\delta)$ and $F_n^{\mathrm{U}}(\delta)$, we can estimate bounds on effects of interest other than the average treatment effects including the proportion of people receiving the treatment who benefit from it, see Heckman et al. (1997) for discussion on some of these effects. In the rest of this section, we review the asymptotic distributions of $\sqrt{n_1}(F_n^{\mathrm{L}}(\delta) - F^{\mathrm{L}}(\delta))$ and $\sqrt{n_1}(F_n^{\mathrm{U}}(\delta) - F^{\mathrm{U}}(\delta))$ established in Fan and Park (2007a), provide two numerical examples to demonstrate the restrictiveness of two assumptions used in Fan and Park (2007a), and then establish asymptotic distributions of $\sqrt{n_1}(F_n^{\mathrm{L}}(\delta) - F^{\mathrm{L}}(\delta))$ and $\sqrt{n_1}(F_n^{\mathrm{U}}(\delta) - F^{\mathrm{U}}(\delta))$ with much weaker assumptions.

4.1. Asymptotic Distributions of $F_n^{\mathrm{L}}(\delta), F_n^{\mathrm{U}}(\delta)$

Define

$$\mathcal{Y}_{\sup,\delta} = \arg\sup_{y \in \mathcal{Y}_\delta} \{F_1(y) - F_0(y-\delta)\}, \quad \mathcal{Y}_{\inf,\delta} = \arg\inf_{y \in \mathcal{Y}_\delta} \{F_1(y) - F_0(y-\delta)\}$$

$$M(\delta) = \sup_{y \in \mathcal{Y}_\delta} \{F_1(y) - F_0(y-\delta)\}, \quad m(\delta) = \inf_{y \in \mathcal{Y}_\delta} \{F_1(y) - F_0(y-\delta)\}$$

$$M_n(\delta) = \sup_{y \in \mathcal{Y}_\delta}\{F_{1n}(y) - F_{0n}(y-\delta)\}, \quad m_n(\delta) = \inf_{y \in \mathcal{Y}_\delta}\{F_{1n}(y) - F_{0n}(y-\delta)\}$$

Then

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

Fan and Park (2007a) assume that $\mathcal{Y}_{\sup,\delta}$ and $\mathcal{Y}_{\inf,\delta}$ are both singletons. Let $y_{\sup,\delta}$ and $y_{\inf,\delta}$ denote, respectively, the elements of $\mathcal{Y}_{\sup,\delta}$ and $\mathcal{Y}_{\inf,\delta}$. The following assumptions are used in Fan and Park (2007a).

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 $\varepsilon>0$, $\sup_{y\in\mathcal{Y}_\delta:|y-y_{\sup,\delta}|\geq\varepsilon}\{F_1(y) - F_0(y-\delta)\}<\{F_1(y_{\sup,\delta}) - F_0(y_{\sup,\delta}-\delta)\}$; (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 $\varepsilon>0, \inf_{y\in\mathcal{Y}_\delta:|y-y_{\inf,\delta}|\geq\varepsilon}\{F_1(y) - F_0(y-\delta)\}<\{F_1(y_{\inf,\delta}) - F_0(y_{\inf,\delta}-\delta)\}$; (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 $\delta \in [a-d, b-c]\cap\mathcal{R}$, (A3) requires the function $y \to \{F_1(y) - F_0(y-\delta)\}$ to have a well-separated interior maximum at $y_{\sup,\delta}$ on $\mathcal{Y}_\delta$, while (A4) requires the function $y \to \{F_1(y) - F_0(y-\delta)\}$ 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 $y \to \{F_1(y) - F_0(y-\delta)\}$ have a unique maximum at $y_{\sup,\delta}$ and a unique minimum at $y_{\inf,\delta}$ in the interior of $\mathcal{Y}_\delta$.

The following result is provided in Fan and Park (2007a).

Theorem 1. Define

$$\sigma_{\mathrm{L}}^2 = F_1(y_{\sup,\delta})[1-F_1(y_{\sup,\delta})] + \lambda F_0(y_{\sup,\delta}-\delta)[1-F_0(y_{\sup,\delta}-\delta)] \quad \text{and}$$

$$\sigma_{\mathrm{U}}^2 = F_1(y_{\inf,\delta})[1-F_1(y_{\inf,\delta})] + \lambda F_0(y_{\inf,\delta}-\delta)[1-F_0(y_{\inf,\delta}-\delta)]$$

(i) Suppose (A1)–(A3) hold. For any $\delta \in [a-d, b-c] \cap \mathcal{R}$

$$\sqrt{n_1}[F_n^{\mathrm{L}}(\delta) - F^{\mathrm{L}}(\delta)] \Rightarrow \begin{cases} N(0, \sigma_{\mathrm{L}}^2), & \text{if } M(\delta) > 0 \\ \max\{N(0, \sigma_{\mathrm{L}}^2), 0\} & \text{if } M(\delta) = 0 \end{cases}$$

$$\text{and } \Pr(F_n^{\mathrm{L}}(\delta) = 0) \to 1 \text{ if } M(\delta) < 0$$

(ii) Suppose (A1), (A2), and (A4) hold. For any $\delta \in [a-d, b-c] \cap \mathcal{R}$,

$$\sqrt{n_1}[F_n^{\mathrm{U}}(\delta) - F^{\mathrm{U}}(\delta)] \Rightarrow \begin{cases} N(0, \sigma_{\mathrm{U}}^2) & \text{if } m(\delta) > 0 \\ \min\{N(0, \sigma_{\mathrm{U}}^2), 0\} & \text{if } m(\delta) = 0 \end{cases}$$

$$\text{and } \Pr(F_n^{\mathrm{U}}(\delta) = 1) \to 1 \text{ if } m(\delta) > 0$$

Theorem 1 shows that the asymptotic distribution of $F_n^{\mathrm{L}}(\delta)(F_n^{\mathrm{U}}(\delta))$ depends on the value of $M(\delta)$ ($m(\delta)$). For example, if δ is such that $M(\delta) > 0$ ($m(\delta) < 0$), then $F_n^{\mathrm{L}}(\delta)$ ($F_n^{\mathrm{U}}(\delta)$) is asymptotically normally distributed, but if δ is such that $M(\delta) = 0$ ($m(\delta) = 0$), then the asymptotic distribution of $F_n^{\mathrm{L}}(\delta)(F_n^{\mathrm{U}}(\delta))$ is truncated normal.

Remark 3. Fan and Park (2007a) proposed the following procedure for computing the estimates $F_n^{\mathrm{L}}(\delta), F_n^{\mathrm{U}}(\delta)$ and estimates of σ_{L}^2 and σ_{U}^2 in Theorem 1. Suppose we know $\mathcal{Y}_\delta$. If $\mathcal{Y}_\delta$ is unknown, we can estimate it by:

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

where $\{Y_{1(i)}\}_{i=1}^{n_1}$ and $\{Y_{0(i)}\}_{i=1}^{n_0}$ are the order statistics of $\{Y_{1(i)}\}_{i=1}^{n_1}$ and $\{Y_{0(i)}\}_{i=1}^{n_0}$, 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 $\{Y_{1(i)}\}_{i=1}^{n_1}$ denoted as $\{Y_{1(i)}\}_{i=r_1}^{s_1}$ as follows:

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

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

$$M_n(\delta) = \max_i \left\{ \frac{i}{n_1} - F_{0n}(Y_{1(i)} - \delta) \right\} \quad \text{for } i \in \{r_1, r_1 + 1, \ldots, s_1\} \tag{15}$$

$$m_n(\delta) = \min_i \left\{ \frac{i}{n_1} - F_{0n}(Y_{1(i)} - \delta) \right\} \quad \text{for } i \in \{r_1, r_1 + 1, \ldots, s_1\} \tag{16}$$

The estimates $F_n^{\mathrm{L}}(\delta)$, $F_n^{\mathrm{U}}(\delta)$ are given by: $F_n^{\mathrm{L}}(\delta) = \max\{M_n(\delta), 0\}$, $F_n^{\mathrm{U}}(\delta) = 1 + \min\{m_n(\delta), 0\}$.

Define two sets I_M and I_m such that

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

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

Then the estimators σ^2_{Ln} and σ^2_{Un} can be defined as:

$$\sigma^2_{\mathrm{Ln}} = \frac{i}{n_1}\left(1 - \frac{i}{n_1}\right) + \lambda F_{0n}(Y_{1(i)} - \delta)(1 - F_{0n}(Y_{1(i)} - \delta)) \quad \text{and}$$

$$\sigma^2_{\mathrm{Un}} = \frac{j}{n_1}\left(1 - \frac{j}{n_1}\right) + \lambda F_{0n}(Y_{1(j)} - \delta)(1 - F_{0n}(Y_{1(j)} - \delta))$$

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 σ^2_{Ln} or σ^2_{Un}. In such a case, we may use $i = \min_k\{k \in I_M\}$ and $j = \min_k\{k \in I_m\}$.

Remark 4. Alternatively we can compute $F_n^{\mathrm{L}}(\delta), F_n^{\mathrm{U}}(\delta)$ as follows. Note that for $0 < q < 1$, Lemma 3 (the duality theorem) implies that the quantile bounds $(F_n^{\mathrm{U}})^{-1}(q)$ and $(F_n^{\mathrm{L}})^{-1}(q)$ can be computed by:

$$(F_n^{\mathrm{L}})^{-1}(q) = \inf_{u \in [q,1]} [F_{1n}^{-1}(u) - F_{0n}^{-1}(u - q)], (F_n^{\mathrm{U}})^{-1}(q)$$
$$= \sup_{u \in [0,q]} [F_{1n}^{-1}(u) - F_{0n}^{-1}(1 + u - q)]$$

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 $(F_n^{\mathrm{L}})^{-1}(q)$ and $(F_n^{\mathrm{U}})^{-1}(q)$ a evenly spaced values of q in (0, 1). One choice that leads to easily computed formulas for $(F_n^{\mathrm{L}})^{-1}(q)$ and

$(F_n^{\mathrm{U}})^{-1}(q)$ is $q=r/n_1$ for $r=1, \ldots, n_1$, as one can show that

$$(F_n^{\mathrm{L}})^{-1}\left(\frac{r}{n_1}\right) = \min_{l=r,\ldots,(n_1-1)} \min_{s=j,\ldots,k} [Y_{1(l+1)} - Y_{0(s)}] \tag{17}$$

where $j = [n_0((l-r)/n_1)] + 1$ and $k = [n_0((l-r+1)/n_1)]$, and

$$(F_n^{\mathrm{U}})^{-1}\left(\frac{r}{n_1}\right) = \max_{l=0,\ldots,(r-1)} \max_{s=j',\ldots,k'} [Y_{1(l+1)} - Y_{0(s)}] \tag{18}$$

where $j' = [n_0((n_1+l-r)/n_1)] + 1$ and $k' = [n_0((n_1+l-r+1)/n_1)]$. In the case where $\mathrm{n}_1 = \mathrm{n}_0 = \mathrm{n}$, Eqs. (17) and (18) simplify:

$$(F_n^{\mathrm{L}})^{-1}\left(\frac{r}{n}\right) = \min_{l=r,\ldots,(n-1)} [Y_{1(l+1)} - Y_{0(l-r+1)}]$$

$$(F_n^{\mathrm{U}})^{-1}\left(\frac{r}{n}\right) = \max_{l=0,\ldots,(r-1)} [Y_{1(l+1)} - Y_{0(n+l-r+1)}]$$

The empirical distribution of $(F_n^{\mathrm{L}})^{-1}(r/n_1), r=1,\ldots,n_1$, provides an estimate of the lower bound distribution and the empirical distribution of $(F_n^{\mathrm{U}})^{-1}(r/n_1), r=1,\ldots,n_1$, provides an estimate of the upper bound distribution. This is the approach we used in our simulations to compute $F_n^{\mathrm{L}}(\delta), F_n^{\mathrm{U}}(\delta)$.

4.2. Two Numerical Examples

We present two examples to illustrate the various possibilities in Theorem 1. For the first example, the asymptotic distribution of $F_n^{\mathrm{L}}(\delta)(F_n^{\mathrm{U}}(\delta))$ is normal for all δ. For the second example, the asymptotic distribution of $F_n^{\mathrm{L}}(\delta)(F_n^{\mathrm{U}}(\delta))$ is normal for some δ and nonnormal for some other δ. More examples can be found in Appendix B.

Example 1 (Continued). Let $Y_j \sim N(\mu_j, \sigma_j^2)$ 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,

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

are unique interior solutions, where $s = \delta - (\mu_1 - \mu_0)$ and $\sqrt{s^2 + 2(\sigma_1^2 - \sigma_0^2)\ln(\sigma_1/\sigma_0)}$. Theorem 3.2 implies that the asymptotic AU :4

distribution of $F_n^{\mathrm{L}}(\delta)(F_n^{\mathrm{U}}(\delta))$ 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

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

Suppose $Y_1 \sim C(1/4)$ and $Y_0 \sim C(3/4)$. The functional form of $F_1(y) - F_0(y-\delta)$ differs according to δ. 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:

$$y_{\sup,\delta} = \begin{cases} \dfrac{1+\delta}{2} & \text{if } -1 + \dfrac{1}{2}\sqrt{2} < \delta \le 1 \\ \left\{0, \dfrac{1+\delta}{2}, 1+\delta\right\} & \text{if } \delta = -1 + \dfrac{1}{2}\sqrt{2} \\ \{0, 1+\delta\} & \text{if } -1 \le \delta < -1 + \dfrac{1}{2}\sqrt{2} \end{cases}$$

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

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

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

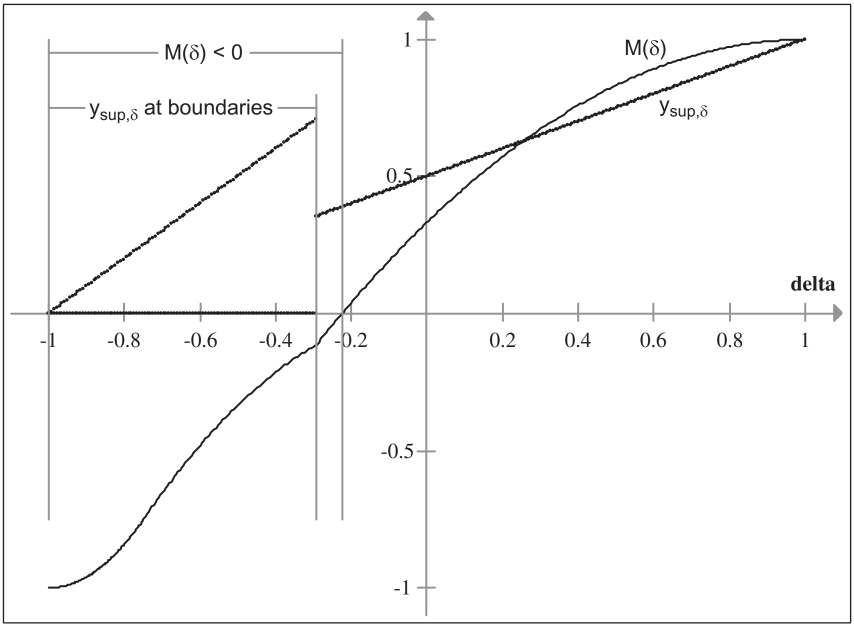


Fig. 3. $M(\delta)$ and $y_{\sup,\delta}$: $(C(1/4), C(3/4))$.

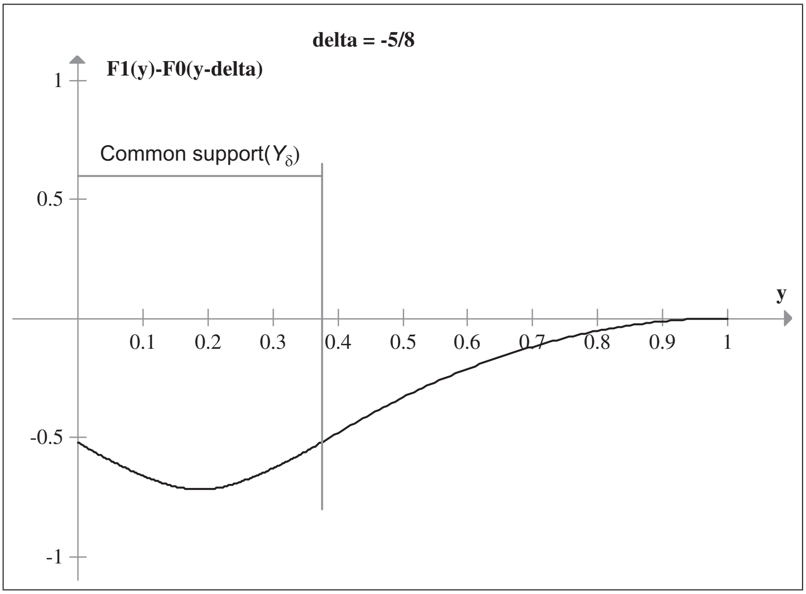
Fig. 4(d) a case where the interior solution corresponds to a positive value for M (δ).

Depending on the value of δ, $M(\delta)$ can have different signs leading to different asymptotic distributions for $F_n^{\mathrm{L}}(\delta)$. For example, when $\delta = 1 - (\sqrt{6}/2)$ (Fig. 4(c)), $M(\delta) = 0$ and for $\delta > 1 - (\sqrt{6}/2), M(\delta) > 0$. Since $M(\delta) = 0$ when $\delta = 1 - (\sqrt{6}/2), y_{\sup,\delta} = 1 - (\sqrt{6}/4)$ is in the interior, and $f_1'(y_{\sup,\delta})- \ f_0'(y_{\sup,\delta} - \delta) = -(16/3) < 0$, Theorem 3.2 implies that at $\delta = 1 - (\sqrt{6}/2)$,

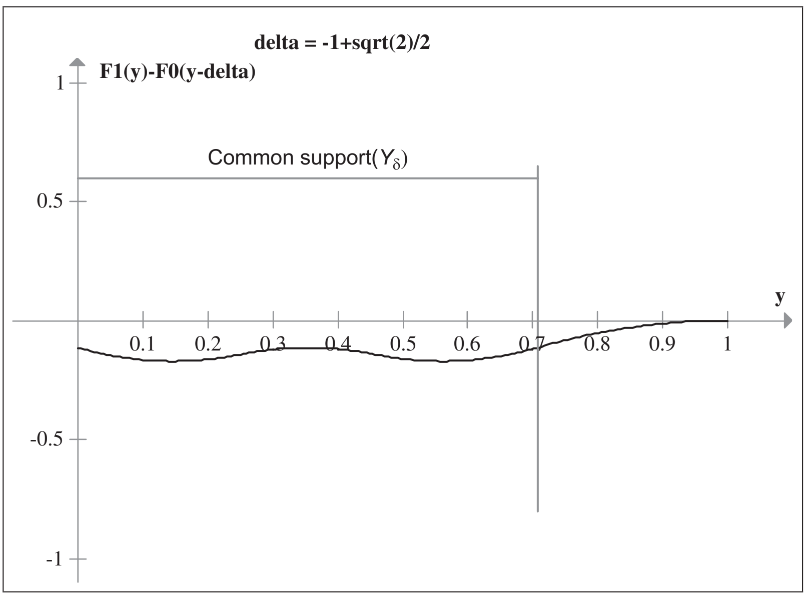
$$\sqrt{n_1}[F_n^{\mathrm{L}}(\delta) - F^{\mathrm{L}}(\delta)] \Rightarrow \max(N(0, \sigma_{\mathrm{L}}^2), 0) \quad \text{where} \quad \sigma_{\mathrm{L}}^2 = \frac{(1 + \lambda)}{4}$$

When $\delta = 1/8$ (Fig. 4(d)),

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

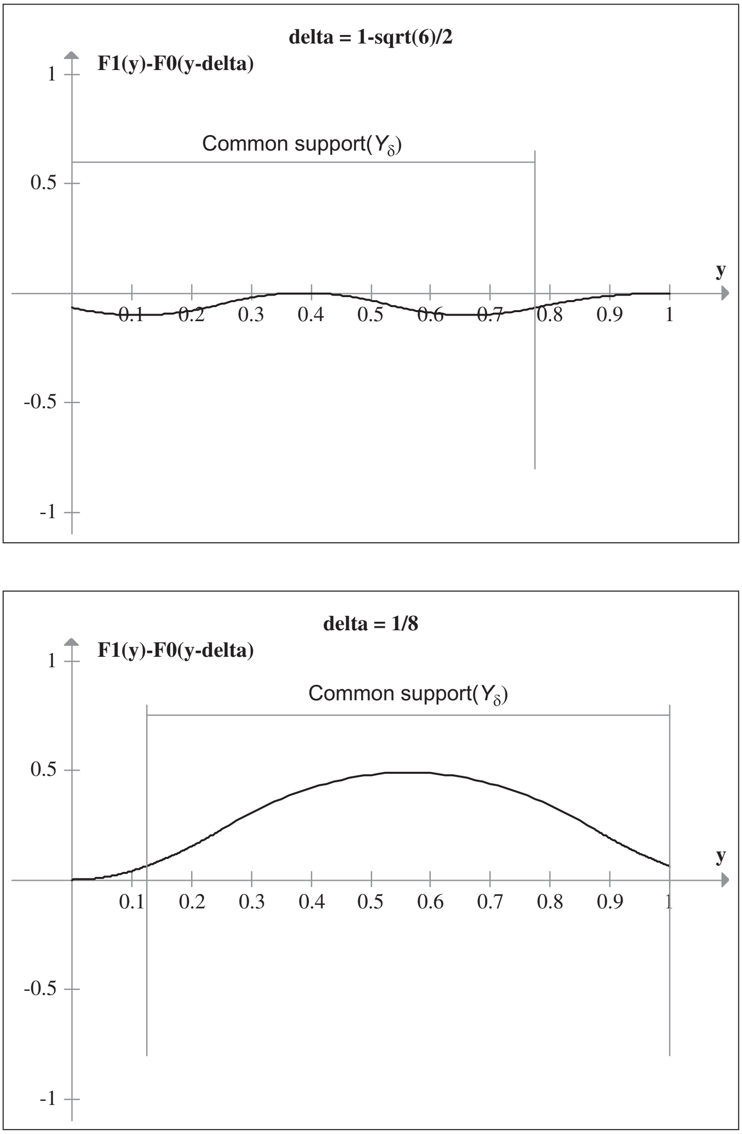


(a)



(b)

Fig. 4. (a) $[F_1(y) - F_0(y + 5/8)]$; (b) $[F_1(y) - F_0(y + 1 - \sqrt{2}/2)]$; (c) $[F_1(y) - F_0(y - 1 - \sqrt{6}/2)]$; and (d) $[F_1(y) - F_0(y - 1/8)]$. AU :1



(c)

(d)

Fig. 4. (*Continued*)

Theorem 3.2 implies that when $\delta = 1/8$,

$$\sqrt{n_1}[F_n^{\mathrm{L}}(\delta) - F^{\mathrm{L}}(\delta)] \Rightarrow N(0, \sigma_{\mathrm{L}}^2) \quad \text{where} \quad \sigma_{\mathrm{L}}^2 = (1+\lambda)\frac{7,007}{36,864}$$

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

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

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

Fig. 5 shows $y_{\inf,\delta}$ and $m(\delta)$.

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

4.3. Asymptotic Distributions of $F_n^{\mathrm{L}}(\delta), F_n^{\mathrm{U}}(\delta)$ Without (A3) and (A4)

As Example 2 illustrates, assumptions (A3) and (A4) may be violated. Figs. 4 or 6 provide us with cases where multiple interior maximizers or minimizers exist. In Fig. 6(b) and (c), there are two interior maximizers when $\delta = (\sqrt{6}/2) - 1$ or $\delta = 1 - (\sqrt{2}/2)$ with $a_1 = 3/4$ and $a_0 = 1/4$. When $\delta = (\sqrt{6}/2) - 1$, $M(\delta) = (\sqrt{6} - 2)^2/2$ and $\mathcal{Y}_{\sup,\delta} = \{((6-\sqrt{6})/4), ((3\sqrt{6}-6)/4)\}$. When $\delta = 1 - (\sqrt{2}/2), M(\delta) = ((2-\sqrt{2})^2)/2$ and $\mathcal{Y}_{\sup,\delta} = \{((\sqrt{2}+2)/4), ((6-3\sqrt{2})/4)\}$. Shown in Fig. 4(b) and (c) are

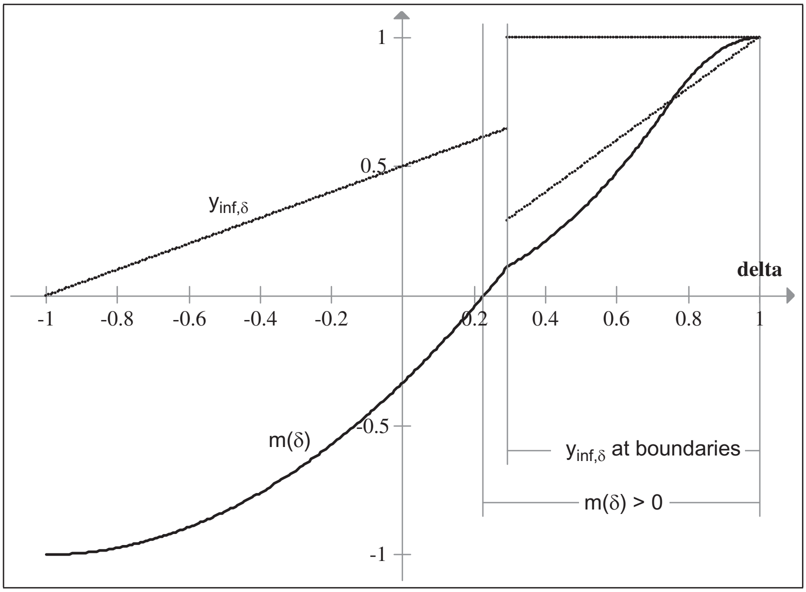


Fig. 5. $m\ (\delta)$ and $y_{\inf,\delta}: (C(3/4), Y_0 \sim C(1/4))$.

cases with multiple interior minimizers for $a_1 = 1/4$ and $a_0 = 3/4$. When $\delta = (\sqrt{2}/2) - 1, m(\delta) = -((2-\sqrt{2})^2/2)$ and $\mathcal{Y}_{\inf,\delta} = \{((2-\sqrt{2})/4), ((3\sqrt{2}-2)/4)\}$. When $\delta = 1 - (\sqrt{6}/2), m(\delta) = -(\sqrt{6}-2)^2/2$ and $\mathcal{Y}_{\inf,\delta} = \{((\sqrt{6}-2)/4), ((10-3\sqrt{6})/4)\}$.

We now dispense with assumptions (A3) and (A4). Recall that

$$\mathcal{Y}_{\sup,\delta} = \{y \in \mathcal{Y}_\delta : F_1(y) - F_0(y-\delta) = M(\delta)\}$$

$$\mathcal{Y}_{\inf,\delta} = \{y \in \mathcal{Y}_\delta : F_1(y) - F_0(y-\delta) = m(\delta)\}$$

For a given $b>0$, define

$$\mathcal{Y}^b_{\sup,\delta} = \{y \in \mathcal{Y}_\delta : F_1(y) - F_0(y-\delta) \geq M(\delta) - b\}$$

$$\mathcal{Y}^b_{\inf,\delta} = \{y \in \mathcal{Y}_\delta : F_1(y) - F_0(y-\delta) \leq m(\delta) + b\}$$

A3′. There exists $K>0$ and $0<\eta<1$ such that for all $y \in \mathcal{Y}^b_{\sup,\delta}$, for $b>0$ sufficiently small, there exists a $y_{\sup,\delta} \in \mathcal{Y}_{\sup,\delta}$ such that $y_{\sup,\delta} \leq y$ and $(y - y_{\sup,\delta}) \leq Kb^\eta$.

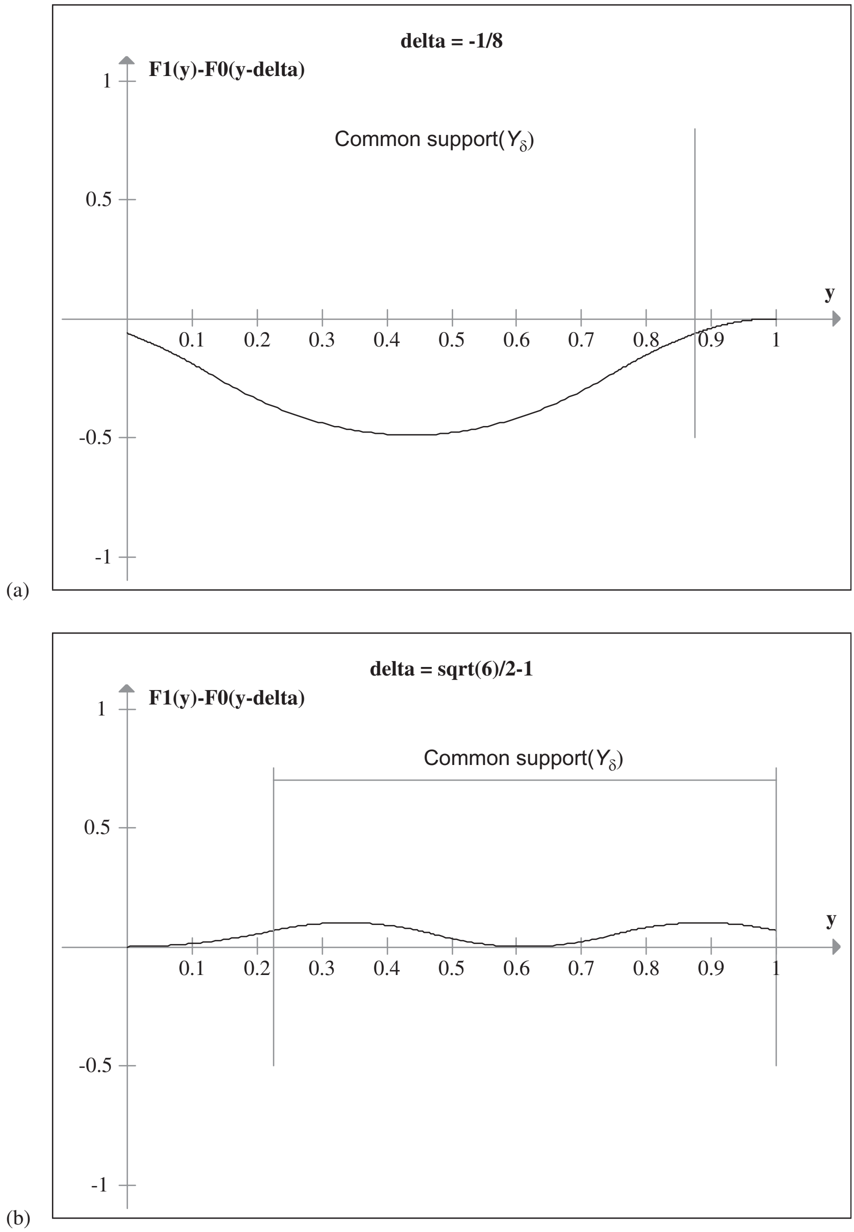


Fig. 6. (a) $[F_1(y) - F_0(y + 1/8)]$; (b) $[F_1(y) - F_0(y - \sqrt{6}/2 + 1)]$; (c) $[F_1(y) - F_0(y - 1 + \sqrt{2}/2)]$; and (d) $[F_1(y) - F_0(y - 5/8)]$.

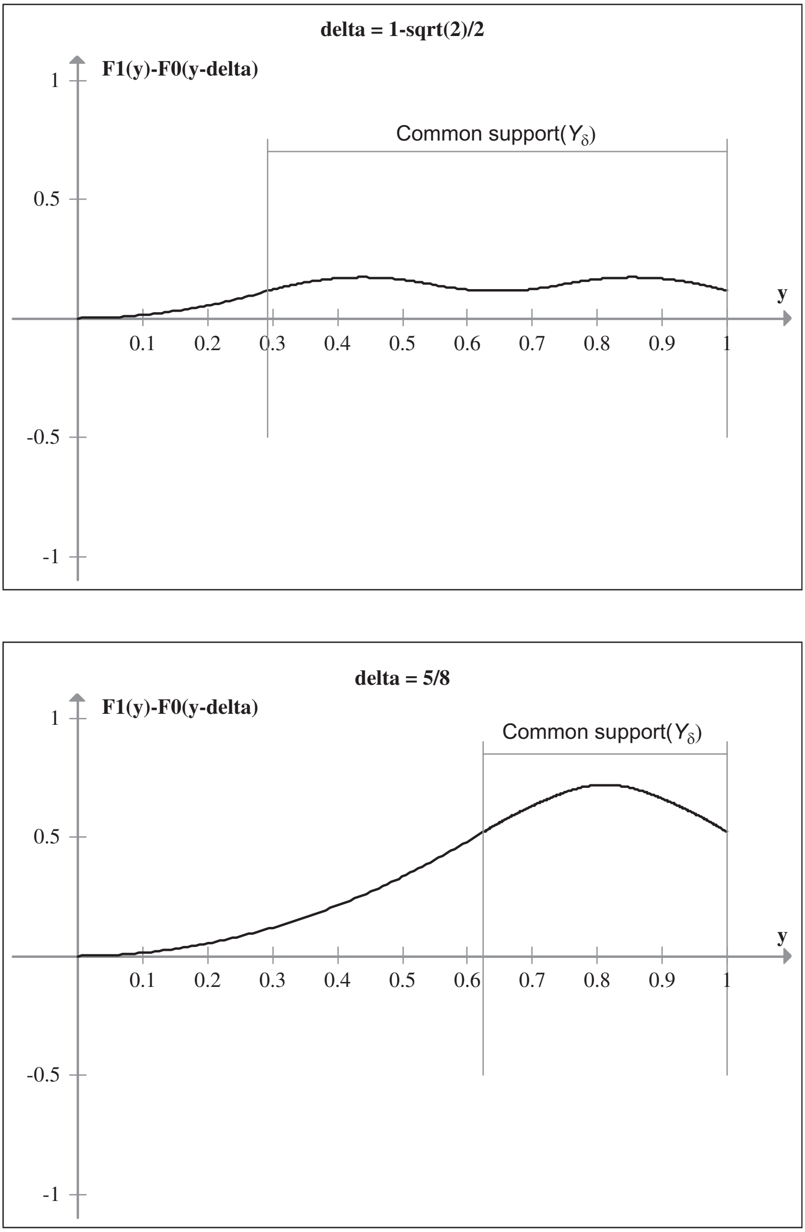


Fig. 6. (*Continued*)

A4′. There exists $K>0$ and $0<\eta<1$ such that for all $y \in \mathcal{Y}^{b}_{\inf,\delta}$ for $b>0$ sufficiently small, there exists a $y_{\inf,\delta} \in \mathcal{Y}_{\inf,\delta}$ such that $y_{\inf,\delta} \leq y$ and $(y - y_{\inf,\delta}) \leq Kb^{\eta}$.

Assumptions (A3)′ and (A4)′ adapt Assumption (1) in Galichon and Henry (2008). As discussed in Galichon and Henry (2008), they are very mild assumptions. By following the proof of Theorem 1 in Galichon and Henry (2008), we can show that under conditions stated in the theorem below,

$$\sqrt{n_1}[M_n(\delta) - M(\delta)] \Rightarrow \sup_{y \in \mathcal{Y}_{\sup,\delta}} G(y,\delta), \sqrt{n_1}[m_n(\delta) - m(\delta)] \Rightarrow \inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta)$$

where $\{G(y,\delta) : y \in \mathcal{Y}_{\delta}\}$ is a tight Gaussian process with zero mean. Thus the theorem below holds.

Theorem 2.

(i) Suppose (A1) and (A3)′ hold. For any $\delta \in [a-d, b-c] \cap \mathcal{R}$, we have

$$\sqrt{n_1}[F_n^{\mathrm{L}}(\delta) - F^{\mathrm{L}}(\delta)] \Rightarrow \begin{cases} \sup_{y \in \mathcal{Y}_{\sup,\delta}} G(y,\delta), & \text{if } M(\delta)>0 \\ \max\{\sup_{y \in \mathcal{Y}_{\sup,\delta}} G(y,\delta), 0\} & \text{if } M(\delta)=0 \end{cases}$$

$$\text{and} \quad \Pr(F_n^{\mathrm{L}}(\delta) = 0) \to 1 \ \text{ if } M(\delta)<0$$

where $\{G(y,\delta) : y \in \mathcal{Y}_{\delta}\}$ is a tight Gaussian process with zero mean.

(ii) Suppose (A1) and (A4)′ hold. For any $\delta \in [a-d, b-c] \cap \mathcal{R}$, we get

$$\sqrt{n_1}[F_n^{\mathrm{U}}(\delta) - F^{\mathrm{U}}(\delta)] \Rightarrow \begin{cases} \inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta), & \text{if } m(\delta)<0 \\ \min\{\inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta), 0\} & \text{if } m(\delta)=0 \end{cases}$$

$$\text{and} \quad \Pr(F_n^{\mathrm{U}}(\delta) = 1) \to 1 \ \text{ if } m(\delta)>0$$

When (A3) and (A4) hold, $\mathcal{Y}_{\sup,\delta}$ and $\mathcal{Y}_{\inf,\delta}$ are singletons and Theorem 2 reduces to Theorem 1.

5. CONFIDENCE SETS FOR THE DISTRIBUTION OF TREATMENT EFFECTS FOR RANDOMIZED EXPERIMENTS

5.1. Confidence Sets for the Sharp Bounds

First, we consider the lower bound. Let

$$G_n(y,\delta) = \sqrt{n_1}[F_{1n}(y) - F_1(y)] - \sqrt{n_1}[F_{0n}(y-\delta) - F_0(y-\delta)]$$

Then

$$\begin{aligned}
&\sqrt{n_1}[F_n^{\mathrm{L}}(\delta) - F^{\mathrm{L}}(\delta)] \\
&= \max\left\{ \sup_{y\in\mathcal{Y}_\delta}\{G_n(y,\delta) + \sqrt{n_1}[F_1(y) - F_0(y-\delta)]\}, 0 \right\} - \max\{\sqrt{n_1}M(\delta), 0\} \\
&\Rightarrow \max\left\{ \sup_{y\in\mathcal{Y}_\delta}[G(y,\delta) + h_{\mathrm{L}}(y,\delta)] + \min\{h_{\mathrm{L}}(\delta),0\}, -\max\{h_{\mathrm{L}}(\delta),0\} \right\} (\equiv W^1_{\mathrm{L},\delta}) \qquad (19) \\
&= \max\left\{ \sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta) + \min\{h_{\mathrm{L}}(\delta),0\}, -\max\{h_{\mathrm{L}}(\delta),0\} \right\} (\equiv W^2_{\mathrm{L},\delta}) \qquad (20)
\end{aligned}$$

where $h_{\mathrm{L}}(y,\delta) = \lim \sqrt{n_1}[F_1(y) - F_0(y-\delta) - M(\delta)] \le 0$ and $h_{\mathrm{L}}(\delta) = \lim[\sqrt{n_1}M(\delta)]$.

Define $h^*_{\mathrm{L}}(\delta) = \sqrt{n_1}M_n(\delta)I\{|M_n(\delta)| > b_n\}$ and

$$\begin{aligned}
h^*_{\mathrm{L}}(y,\delta) = \sqrt{n_1}[F_{1n}(y) - F_{0n}(y-\delta) - M_n(\delta)]I\{[F_{1n}(y) \\
- F_{0n}(y-\delta) - M_n(\delta)] < -b'_n\}
\end{aligned}$$

where b_n is a prespecified deterministic sequence satisfying $b_n \to 0$ and $\sqrt{n_1}b_n \to \infty$ and b'_n is a prespecified deterministic sequence satisfying $b'_n \ln\ln\, n_1 + (\sqrt{n_1}b')^{-1}\sqrt{\ln\ln\, n_1} \to 0$. In the simulations, we considered $b_n = cn_1^{-a}, 0<a<(1/2), c>0$ and $b'_n = c'n_1^{-(1-a')/2}, 0<a'<1, c'>0$. For such b'_n, we have

$$b'_n \ln\ln\, n_1 + (\sqrt{n_1}b'_n)^{-1}\sqrt{\ln\ \ln\ n_1} = c'\frac{\ln\ \ln\ n_1}{\sqrt{n_1^{1-a'}}} + \frac{1}{c'}\frac{\sqrt{\ln\ \ln\ n_1}}{\sqrt{n_1^{a'}}} \to 0$$

Based on Eqs. (19) and (20), we propose two bootstrap procedures to approximate the distribution of $\sqrt{n_1}[F_n^{\mathrm{L}}(\delta) - F^{\mathrm{L}}(\delta)]$. In the first procedure, we approximate the distribution of $W^1_{\mathrm{L},\delta}$ and in the second procedure, we approximate the distribution of $W^2_{\mathrm{L},\delta}$. Draw bootstrap samples with replacement from $\{Y_{1i}\}_{i=1}^{n_1}$ and $\{Y_{0i}\}_{i=1}^{n_0}$, respectively. Let $F^*_{1n}(y)$, $F^*_{0n}(y)$ denote the empirical distribution functions based on the bootstrap samples, respectively. Define

$$G^*_n(y,\delta) = \sqrt{n_1}[F^*_{1n}(y) - F_{1n}(y)] - \sqrt{n_1}[F^*_{0n}(y-\delta) - F_{0n}(y-\delta)]$$

In the first bootstrap approach, we use the distribution of the following random variable conditional on the original sample to approximate the quantiles of the limiting distribution of $\sqrt{n_1}[F^{\mathrm{L}}_n(\delta) - F^{\mathrm{L}}(\delta)]$:

$$W^{1*}_{\mathrm{L},\delta} = \max\left\{ \sup_{y\in\mathcal{Y}_\delta}\{G^*_n(y,\delta) + h^*_{\mathrm{L}}(y,\delta)\} + \min\{h^*_{\mathrm{L}}(\delta),0\}, -\max\{h^*_{\mathrm{L}}(\delta),0\} \right\}$$

In the second bootstrap approach, we estimate $\mathcal{Y}_{\sup,\delta}$ directly and approximate the distributions of $W_{\mathrm{L},\delta}$. Define

$$\mathcal{Y}_{n\sup,\delta} = \{y_i \in \{Y_{1i}\}_{i=1}^{n_1} \cup \{Y_{0i}\}_{i=1}^{n_0} : M_n(\delta) - (F_{n1}(y_i) - F_{n0}(y_i - \delta)) \le b'_n\}$$

Then the distribution of the following random variable conditional on the original sample can be used to approximate the quantiles of the limiting distribution of $\sqrt{n_1}[F^{\mathrm{L}}_n(\delta) - F^{\mathrm{L}}(\delta)]$:

$$W^{2*}_{\mathrm{L},\delta} = \max\left\{ \sup_{y\in\mathcal{Y}_{n\ \sup,\delta}} G^*_n(y,\delta), -h^*_{\mathrm{L}}(\delta) \right\} + \min\{h^*_{\mathrm{L}}(\delta),0\}$$

The upper bound can be dealt with similarly. Note that

$$\begin{aligned}
&\sqrt{n_1}[F^{\mathrm{U}}_n(\delta) - F^{\mathrm{U}}(\delta)] \\
&\quad\Rightarrow \min\left\{ \inf_{y\in\mathcal{Y}_\delta}\{G_n(y,\delta) + h_{\mathrm{U}}(y,\delta)\} + \max\{h_{\mathrm{U}}(\delta),0\}, -\min\{h_{\mathrm{U}}(\delta),0\} \right\} \\
&\quad\Rightarrow \min\left\{ \inf_{y\in\mathcal{Y}_\delta}[G(y,\delta) + h_{\mathrm{U}}(y,\delta)] + \max\{h_{\mathrm{U}}(\delta),0\}, -\min\{h_{\mathrm{U}}(\delta),0\} \right\} (\equiv W^1_{\mathrm{U},\delta}) \\
&\quad= \min\left\{ \inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta) + \max\{h_{\mathrm{U}}(\delta),0\}, -\min\{h_{\mathrm{U}}(\delta),0\} \right\} (\equiv W^2_{\mathrm{U},\delta})
\end{aligned}$$

where $h_{\mathrm{U}}(y,\delta) = \lim \sqrt{n_1}[F_1(y) - F_0(y-\delta) - m(\delta)] \ge 0$ and $h_{\mathrm{U}}(\delta) = \lim[\sqrt{n_1}m(\delta)]$.

Define $h_{\mathrm{U}}^{*}(\delta)=\sqrt{n_1}m_n(\delta)I\{|m_n(\delta)|>b_n\}$ and

$$h_{\mathrm{U}}^{*}(y,\delta)=\sqrt{n_1}[F_{1n}(y)-F_{0n}(y-\delta)-m_n(\delta)]I\{[F_{1n}(y)-F_{0n}(y-\delta)-m_n(\delta)]>b_n'\}$$

We propose to use the distribution of $W_{\mathrm{U},\delta}^{1*}$ or $W_{n\mathrm{U},\delta}^{2*}$ conditional on the original sample to approximate the quantiles of the distribution of $\sqrt{n_1}[F_n^{\mathrm{U}}(\delta)-F^{\mathrm{U}}(\delta)]$, where

$$W_{\mathrm{U},\delta}^{1*}=\min\left\{\inf_{y\in\mathcal{Y}_\delta}\{G_n^{*}(y,\delta)+h_{\mathrm{U}}^{*}(y,\delta)\}+\max\{h_{\mathrm{U}}^{*}(\delta),0\},-\min\{h_{\mathrm{U}}^{*}(\delta),0\}\right\}$$

$$W_{\mathrm{U},\delta}^{2*}=\min\left\{\inf_{y\in\mathcal{Y}_{n\inf\delta}}G_n^{*}(y,\delta),-h_{\mathrm{U}}^{*}(\delta)\right\}+\max\{h_{\mathrm{U}}^{*}(\delta),0\}$$

in which

$$\mathcal{Y}_{n\inf,\delta}=\{y_i\in\{Y_{1i}\}_{i=1}^{n_1}\cup\{Y_{0i}\}_{i=1}^{n_0}:m_n(\delta)-(F_{n1}(y_i)-F_{n0}(y_i-\delta))\geq -b_n'\}$$

Throughout the simulations presented in Section 7, we used $W_{\mathrm{L},\delta}^{2*}$ and $W_{\mathrm{U},\delta}^{2*}$.

5.2. Confidence Sets for the Distribution of Treatment Effects

For notational simplicity, we let $\theta_0=F_\Delta(\delta)$, $\theta_{\mathrm{L}}=F^{\mathrm{L}}(\delta)$, and $\theta_{\mathrm{U}}=F^{\mathrm{U}}(\delta)$. Also let $\Theta=[0,\,1]$. This subsection follows similar ideas to Fan and Park (2007c). Noting that

$$\theta_0=\arg\min_{\theta\in\Theta}\{(\theta_{\mathrm{L}}-\theta)_+^2+(\theta_{\mathrm{U}}-\theta)_-^2\}$$

where $(x)_-=\min\{x,\,0\}$ and $(x)_+=\max\,\{x,\,0\}$, we define the test statistic

$$T_n(\theta_0)=n_1(\hat\theta_L-\theta_0)_+^2+n_1(\hat\theta_{\mathrm{U}}-\theta_0)_-^2 \tag{21}$$

where $\hat\theta_{\mathrm{L}}=F_n^{\mathrm{L}}(\delta)$ and $\hat\theta_{\mathrm{U}}=F_n^{\mathrm{U}}(\delta)$. Then a $(1-\alpha)$ level CS for θ_0 can be constructed as,

$$\mathrm{CS}_n=\{\theta\in\Theta:T_n(\theta)\leq c_{1-\alpha}(\theta)\} \tag{22}$$

for an appropriately chosen critical value $c_{1-\alpha}(\theta)$.

To determine the critical value $c_{1-\alpha}(\theta)$, the limiting distribution of $T_n(\theta)$ under an appropriate local sequence is essential. We introduce some necessary notation. Let

$$h^{\mathrm{L}}(\theta_0)=-\lim_{n\to\infty}\sqrt{n}[\theta_{\mathrm{L}}-\theta_0]\quad\text{and}\quad h^{\mathrm{U}}(\theta_0)=\lim_{n\to\infty}\sqrt{n}[\theta_{\mathrm{U}}-\theta_0]$$

Then $h^{\mathrm{L}}(\theta_0) \geq 0, h^{\mathrm{U}}(\theta_0) \geq 0$, and $h^{\mathrm{L}}(\theta_0) + h^{\mathrm{U}}(\theta_0) = \lim_{n\to\infty}(\sqrt{n}\nabla)$, where $\nabla \equiv \theta_{\mathrm{U}} - \theta_{\mathrm{L}}$ is the length of the identified interval. As proposed in Fan and Park (2007c), we use the following shrinkage "estimators" of $h^{\mathrm{L}}(\theta_0)$ and $h^{\mathrm{U}}(\theta_0)$.

$$h^{\mathrm{L}*}(\theta_0) = -\sqrt{n}[\widehat{\theta}_{\mathrm{L}} - \theta_0] I\{[\theta_0 - \widehat{\theta}_{\mathrm{L}}] > b_n\}$$

$$h^{\mathrm{U}*}(\theta_0) = \sqrt{n}[\widehat{\theta}_{\mathrm{U}} - \theta_0] I\{[\widehat{\theta}_{\mathrm{U}} - \theta_0] > b_n\}$$

It remains to establish the asymptotic distribution of $T_n(\theta_0)$:

$$\begin{aligned} T_n(\theta_0) &= (\sqrt{n_1}[\widehat{\theta}_{\mathrm{L}} - \theta_{\mathrm{L}}] - \sqrt{n_1}[\theta_0 - \theta_L])_+^2 + (\sqrt{n_1}[\widehat{\theta}_{\mathrm{U}} - \theta_{\mathrm{U}}] + \sqrt{n_1}[\theta_{\mathrm{U}} - \theta_0])_-^2 \\ &\Rightarrow (W_{\mathrm{L},\delta} - h^{\mathrm{L}}(\theta_0))_+^2 + (W_{\mathrm{U},\delta} - h^{\mathrm{U}}(\theta_0))_-^2 \end{aligned}$$

Let

$$T_n^*(\theta_0) = (W_{\mathrm{L},\delta}^* - h^{\mathrm{L}}(\theta_0))_+^2 + (W_{\mathrm{U},\delta}^* - h^{\mathrm{U}}(\theta_0))_-^2$$

and $cv_{1-\alpha}^*(h^{\mathrm{L}}(\theta_0), h^{\mathrm{U}}(\theta_0))$ denote the $1-\alpha$ quantile of the bootstrap distribution of $T_n^*(\theta_0)$, where $W_{\mathrm{L},\delta}^*$ and $W_{\mathrm{U},\delta}^*$ are either $W_{\mathrm{L},\delta}^{1*}$ and $W_{\mathrm{U},\delta}^{1*}$ or $W_{\mathrm{L},\delta}^{2*}$ and $W_{\mathrm{U},\delta}^{2*}$ defined in the previous subsection. The following theorem holds for a $\underline{p} \in [0,1]$.

Theorem 3. Suppose (A1), (A3)′, and (A4)′ hold. Then, for $\alpha \in [0, \underline{p}]$,

$$\lim_{n_1\to\infty} \inf_{\theta_0\in[\theta_{\mathrm{L}},\theta_{\mathrm{U}}]} \Pr(\theta_0 \in \{\theta : T_n(\theta) \leq cv_{1-\alpha}^*(h^{\mathrm{L}*}(\theta), h^{\mathrm{U}*}(\theta))\}) \geq 1 - \alpha$$

The coverage rates presented in Section 7 are results of the confidence sets of Theorem 3 (i). The presence of $\underline{p}$ in Theorem 3 is due to the fact that AU :5 $T_n(\theta_0)$ is nonnegative and so is $cv_{1-\alpha}^*(h^{\mathrm{L}*}(\theta), h^{\mathrm{U}*}(\theta))$. In Appendix A, we show that one can take $\underline{p}$ as,

$$\underline{p} = 1 - \Pr\left[\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta) \leq 0, \inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta) \geq 0\right] \tag{23}$$

In actual implementation, $\underline{p}$ has to be estimated. We propose the following estimator $\hat{\underline{p}}$:

$$\hat{\underline{p}} = 1 - \frac{1}{B}\sum_{b=1}^{B} 1\left\{\sup_{y\in\mathcal{Y}_{n\sup,\delta}} G_n^{(b)}(y,\delta) \leq 0, \inf_{y\in\mathcal{Y}_{n\inf,\delta}} G_n^{(b)}(y,\delta) \geq 0\right\}$$

where $G_n^{(b)}(y,\delta)$ is $G_n^*(y,\delta)$ from bth bootstrap samples.

6. BIAS-CORRECTED ESTIMATORS OF SHARP BOUNDS ON THE DISTRIBUTION OF TREATMENT EFFECTS FOR RANDOMIZED EXPERIMENTS

In this section, we demonstrate that the plug-in estimators $F_n^{\mathrm{L}}(\delta)$, $F_n^{\mathrm{U}}(\delta)$ tend to have nonnegligible bias in finite samples. In particular, $F_n^{\mathrm{L}}(\delta)$ tends to be biased upward and $F_n^{\mathrm{U}}(\delta)$ tends to be biased downward. We show this analytically when (A3) and (A4) hold. In particular, when (A3) and (A4) hold, we provide closed-form expressions for the first-order asymptotic biases of $F_n^{\mathrm{L}}(\delta), F_n^{\mathrm{U}}(\delta)$ and use these expressions to construct bias-corrected estimators for $F^{\mathrm{L}}(\delta)$ and $F^{\mathrm{U}}(\delta)$. When (A3) and (A4) fail, we propose bootstrap bias-corrected estimators of the sharp bounds $F^{\mathrm{L}}(\delta)$ and $F^{\mathrm{U}}(\delta)$.

Recall

$$F_n^{\mathrm{L}}(\delta) = \max\{M_n(\delta), 0\} \quad \text{and} \quad F^{\mathrm{L}}(\delta) = \max\{M(\delta), 0\}$$

$$F_n^{\mathrm{U}}(\delta) = 1 + \min\{m_n(\delta), 0\} \quad \text{and} \quad F^{\mathrm{U}}(\delta) = 1 + \min\{m(\delta), 0\}$$

where under (A3) and (A4), we have

$$\sqrt{n_1}(M_n(\delta) - M(\delta)) \Rightarrow N(0, \sigma_{\mathrm{L}}^2) \quad \text{and} \quad \sqrt{n_1}(m_n(\delta) - m(\delta)) \Rightarrow N(0, \sigma_{\mathrm{U}}^2)$$

First, we consider the lower bound. Ignoring the second-order terms, we get:

$$\begin{aligned}
E[F_n^{\mathrm{L}}(\delta)] &= E[M_n(\delta) I_{\{M_n(\delta)\geq 0\}}] \\
&= E\left[\left\{M(\delta) + \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}} Z\right\} I_{\{M(\delta)+(\sigma_{\mathrm{L}}/\sqrt{n_1})Z\geq 0\}}\right] \quad \text{where} \quad Z \sim N(0,1) \\
&= M(\delta) E[I_{\{M(\delta)+(\sigma_{\mathrm{L}}/\sqrt{n_1})Z\geq 0\}}] + \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}} E[Z I_{\{M(\delta)+(\sigma_{\mathrm{L}}/\sqrt{n_1})Z\geq 0\}}] \\
&= M(\delta) E[I_{\{z\geq -(\sqrt{n_1}/\sigma_{\mathrm{L}})M(\delta)\}}] + \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}} E[Z I_{\{Z\geq -(\sqrt{n_1}/\sqrt{\sigma_{\mathrm{L}}})M(\delta)\}}] \\
&= M(\delta) \int_{-(\sqrt{n_1}/\sigma_{\mathrm{L}})M(\delta)}^{\infty} \phi(z)dz + \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}} \int_{-(\sqrt{n_1}/\sigma_{\mathrm{L}})M(\delta)}^{\infty} z\phi(z)dz \\
&= M(\delta)\left\{1 - \Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}} M(\delta)\right)\right\} \\
&\qquad - \frac{1}{\sqrt{2\pi}} \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}} \int_{-(\sqrt{n_1}/\sigma_{\mathrm{L}})M(\delta)}^{\infty} \exp\left(-\frac{z^2}{2}\right) d\left(-\frac{z^2}{2}\right) \\
&= M(\delta)\Phi\left(\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}} M(\delta)\right) + \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}} \phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}} M(\delta)\right)
\end{aligned}$$

Case I. Suppose $M(\delta) \geq 0$. Then ignoring second-order terms, we obtain

$$\begin{aligned} E[F_n^{\mathrm{L}}(\delta)] - F^{\mathrm{L}}(\delta) &= M(\delta)\Phi\left(\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right) + \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}}\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right) - M(\delta) \\ &= M(\delta)\left\{\Phi\left(\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right) - 1\right\} + \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}}\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right) \\ &= -M(\delta)\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right) + \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}}\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right) \\ &= \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}}\left\{\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right) - \frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right)\right\} \\ &> 0 \text{ (positive bias)} \end{aligned}$$

because

$$\lim_{x\to 0}\{\phi(-x) - x\Phi(-x)\} = \phi(0) = \frac{1}{\sqrt{2\pi}}$$

$$\begin{aligned} \lim_{x\to +\infty}\{\phi(-x) - x\Phi(-x)\} = \lim_{x\to -\infty}\{\phi(x) + x\Phi(x)\} &= \lim_{x\to -\infty}\frac{d}{dx}\left(\frac{\Phi(x)}{x^{-1}}\right) \\ &= -\lim_{x\to -\infty}\left(\frac{\Phi(x)}{x^{-2}}\right) = 0 \end{aligned}$$

$$\frac{d}{dx}\{\phi(-x) - x\Phi(-x)\} = -\Phi(-x) < 0 \text{ for all } x \in R_+ \cap \{0\}$$

Case II. Suppose $M\ (\delta) < 0$. Then ignoring second-order terms, we obtain

$$\begin{aligned} E[F_n^{\mathrm{L}}(\delta)] - F^{\mathrm{L}}(\delta) &= M(\delta)\Phi\left(\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right) + \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}}\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right) \\ &= \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}}\left\{\phi\left(\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right) + \frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\Phi\left(\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}M(\delta)\right)\right\} \\ &= \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}}\left\{\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}|M(\delta)|\right) - \frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}|M(\delta)|\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}|M(\delta)|\right)\right\} \\ &> 0 \text{ (positive bias)} \end{aligned}$$

Summarizing Case I and Case II, we obtain the first-order asymptotic bias of $F_n^{\mathrm{L}}(\delta)$:

$$E[F_n^{\mathrm{L}}(\delta)] - F^{\mathrm{L}}(\delta) = \frac{\sigma_{\mathrm{L}}}{\sqrt{n_1}}\left\{\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}|M(\delta)|\right) - \frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}|M(\delta)|\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}}}|M(\delta)|\right)\right\}$$

regardless of the sign of $M(\delta)$, an estimator of which is

$$\widehat{\mathrm{Bias}}_{\mathrm{L}} = \frac{\sigma_{\mathrm{L}n}}{\sqrt{n_1}}\left\{\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}n}}|M_n^*(\delta)|\right) - \frac{\sqrt{n_1}}{\sigma_{\mathrm{L}n}}|M_n^*(\delta)|\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}n}}|M_n^*(\delta)|\right)\right\}$$

where $M_n^*(\delta) = M_n(\delta)I\{|M_n(\delta)| > b_n\}$ in which $b_n \to 0$ and $\sqrt{n_1}b_n \to \infty$. We define the bias-corrected estimator of $F^{\mathrm{L}}(\delta)$ as,

$$\begin{aligned} F_{n\mathrm{BC}}^{\mathrm{L}}(\delta) &= \max\{F_n^{\mathrm{L}}(\delta) - \widehat{\mathrm{Bias}}_{\mathrm{L}}, 0\} \\ &= \max\left\{F_n^{\mathrm{L}}(\delta) - \frac{\sigma_{\mathrm{Ln}}}{\sqrt{n_1}}\left\{\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}n}}|M_n^*(\delta)|\right)\right.\right. \\ &\quad \left.\left. - \frac{\sqrt{n_1}}{\sigma_{\mathrm{L}n}}|M_n^*(\delta)|\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{L}n}}|M_n^*(\delta)|\right)\right\}, 0\right\} \\ &\leq F_n^{\mathrm{L}}(\delta) \end{aligned}$$

Now consider the upper bound. The following holds:

$$\begin{aligned} E[F_n^{\mathrm{U}}(\delta)] &= 1 + E[m_n(\delta)I_{\{m_n(\delta)\leq 0\}}] \\ &= 1 + E\left[\left\{m(\delta) + \frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}Z\right\}I_{\{m(\delta)+(\sigma_{\mathrm{U}}/\sqrt{n_1})Z\leq 0\}}\right] \\ &= 1 + m(\delta)E[I_{\{m(\delta)+(\sigma_{\mathrm{U}}/\sqrt{n_1})Z\leq 0\}}] + \frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}E[ZI_{\{m(\delta)+(\sigma_{\mathrm{U}}/\sqrt{n_1})Z\leq 0\}}] \\ &= 1 + m(\delta)\int_{-\infty}^{-(\sqrt{n_1}/\sigma_{\mathrm{U}})m(\delta)}\phi(z)dz + \frac{1}{\sqrt{2\pi}}\frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}\int_{-\infty}^{-(\sqrt{n_1}/\sigma_{\mathrm{U}})m(\delta)} z\exp\left(-\frac{z^2}{2}\right)dz \\ &= 1 + m(\delta)\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\sigma)\right) - \frac{1}{\sqrt{2\pi}}\frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}\int_{-\infty}^{-(\sqrt{n_1}/\sigma_{\mathrm{U}})m(\delta)}\exp\left(-\frac{z^2}{2}\right)d\left(-\frac{z^2}{2}\right) \\ &= 1 + m(\delta)\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) - \frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) \end{aligned}$$

Case I. Suppose $m(\delta) \leq 0$. Then ignoring second-order terms, we obtain

$$\begin{aligned} E[F_n^{\mathrm{U}}(\delta)] - F^{\mathrm{U}}(\delta) &= m(\delta)\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) - \frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) - m(\delta) \\ &= -m(\delta)\Phi\left(\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) - \frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) \\ &= -m(\delta)\Phi\left(\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) - \frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}\phi\left(\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) \\ &= -\frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}\left(\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}|m(\delta)|\right) - \frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}|m(\delta)|\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}|m(\delta)|\right)\right) \\ &< 0 \text{ (negative bias)} \end{aligned}$$

Case II. Suppose $m(\delta) > 0$. Then ignoring second-order terms, we obtain

$$\begin{aligned} E[F_n^{\mathrm{U}}(\delta)] - F^{\mathrm{U}}(\delta) &= m(\delta)\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) - \frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) \\ &= -\frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}\left(\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right) - \frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}m(\delta)\right)\right) \\ &< 0 \text{ (negative bias)} \end{aligned}$$

Therefore, the first-order asymptotic bias of $F_n^{\mathrm{U}}(\delta)$ is given by:

$$E[F_n^{\mathrm{U}}(\delta)] - F^{\mathrm{U}}(\delta) = -\frac{\sigma_{\mathrm{U}}}{\sqrt{n_1}}\left(\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}|m(\delta)|\right) - \frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}|m(\delta)|\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}}}|m(\delta)|\right)\right)$$

regardless of the sign of $m(\delta)$, an estimator of which is

$$\widehat{\mathrm{Bias}}_{\mathrm{U}} = -\frac{\sigma_{\mathrm{U}n}}{\sqrt{n_1}}\left(\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}n}}|m_n^*(\delta)|\right) - \frac{\sqrt{n_1}}{\sigma_{\mathrm{U}n}}|m_n^*(\delta)|\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}n}}|m_n^*(\delta)|\right)\right)$$

where $m_n^*(\delta) = m_n(\delta)I\{|m_n(\delta)| > b_n\}$. A bias corrected estimator of $F^{\mathrm{U}}(\delta)$ is defined as,

$$\begin{aligned} F_{n\mathrm{BC}}^{\mathrm{U}}(\delta) = \min\{F_n^{\mathrm{U}}(\delta) - \widehat{\mathrm{Bias}}, 1\} = \min\Bigg\{ & F_n^{\mathrm{U}}(\delta) + \frac{\sigma_{\mathrm{U}n}}{\sqrt{n_1}}\left(\phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}n}}|m_n^*(\delta)|\right)\right. \\ & \left. - \frac{\sqrt{n_1}}{\sigma_{\mathrm{U}n}}|m_n^*(\delta)|\Phi\left(-\frac{\sqrt{n_1}}{\sigma_{\mathrm{U}n}}|m_n^*(\delta)|\right)\right), 1\Bigg\} \geq F_n^{\mathrm{U}}(\delta) \end{aligned}$$

The bias-corrected estimators we just proposed depend on the validity of (A3) and (A4). Without these assumptions, the analytical expressions derived for the bias may not be correct. Instead, we propose the following bootstrap bias-corrected estimators. Define

$$\widehat{\text{Bias}(F_n^{\text{L}}(\delta))} = \frac{1}{B}\sum_{b=1}^{B}\frac{W_{\text{L},\delta}^{(b)}}{\sqrt{n_1}} \quad \text{and} \quad \widehat{\text{Bias}(F_n^{\text{U}}(\delta))} = \frac{1}{B}\sum_{b=1}^{B}\frac{W_{\text{U},\delta}^{(b)}}{\sqrt{n_1}}$$

where $W_{\text{L},\delta}^{(b)}(W_{\text{U},\delta}^{(b)})$ are $W_{\text{L},\delta}^{F*}(W_{\text{U},\delta}^{F*})$ or $W_{\text{L},\delta}^{*}(W_{\text{U},\delta}^{*})$ from bth bootstrap samples, where $W_{\text{L},\delta}^{F*}$, $W_{\text{U},\delta}^{F*}$, $W_{\text{L},\delta}^{*}$, and $W_{\text{U},\delta}^{*}$ are defined in the previous subsections. The bootstrap bias-corrected estimators of $F^{\text{L}}(\delta)$ and $F^{\text{U}}(\delta)$ are, respectively,

$$\widehat{F}_{n\text{BC}}^{\text{L}}(\delta) = \max\{F_n^{\text{L}}(\delta) - \widehat{\text{Bias}(F_n^{\text{L}}(\delta))}, 0\} \quad \text{and}$$
$$\widehat{F}_{n\text{BC}}^{\text{U}}(\delta) = \min\{F_n^{\text{U}}(\delta) - \widehat{\text{Bias}(F_n^{\text{U}}(\delta))}, 1\}$$

7. SIMULATION

In this section, we examine the finite sample accuracy of the nonparametric estimators of the treatment effect distribution bounds, investigate the coverage rates of the proposed CSs for the distribution of treatment effects at different values of δ, and the finite sample performance of the bootstrap bias-corrected estimators of the sharp bounds on the distribution of treatment effects. We focus on randomized experiments.

The data generating processes (DGP) used in this simulation study are, respectively, Example 1 and Example 2 introduced in Sections 2.3 and 4.2. The detailed simulation design will be described in Section 7.1 together with estimates F_n^{L} and F_n^{U}. Section 7.2 presents results on the coverage rates of the CSs for the distribution of treatment effects and Section 7.3 presents results on the bootstrap bias-corrected estimators.

7.1. The Simulation Design and Estimates F_n^{L} and F_n^{U}

The DGPs used in the simulations are: (i) (Case C1) $(F_1, F_0, \delta) = (C(1/4), C(3/4), (1/8))$; (ii) (Case C2) $(F_1, F_0, \delta) = (C(1/4), C(3/4), 1 - (\sqrt{6}/2))$; (iii) (Case C3) $(F_1, F_0, \delta) = (C(3/4), C(1/4), (\sqrt{6}/2) - 1)$; and (iv) (Case C4) $(F_1, F_0, \delta) = (C(3/4), C(1/4), -(1/8))$.

(Case C1) is aiming at the case where $M(\delta)>0$ with a singleton $\mathcal{Y}_{\sup,\delta}$ so that we have a normal asymptotic distribution for $\sqrt{n_1}(F_n^{\mathrm{L}}(\delta)-F^{\mathrm{L}}(\delta))$. The $m(\delta)$ for this case is greater than zero so $F^{\mathrm{U}}(\delta)=1$ and $\Pr(F_n^{\mathrm{U}}(\delta)=1)\to 1$. In this case, $\mathcal{Y}_{\inf,\delta}$ consists of two boundary points of $\mathcal{Y}_\delta$.

In (Case C2), $M(\delta)=0$ and $\mathcal{Y}_{\sup,\delta}$ is a singleton so we have a truncated normal asymptotic distribution for $\sqrt{n_1}(F_n^{\mathrm{L}}(\delta)-F^{\mathrm{L}}(\delta))$. The $m(\delta)$, however, is less than zero and has two interior maximizers. So the asymptotic distribution of $\sqrt{n_1}(F_n^{\mathrm{U}}(\delta)-F^{\mathrm{U}}(\delta))$ is $\sup_{y\in\mathcal{Y}_{\inf,\delta}}G(y,\delta)$.

(Case C3) is opposite to (Case C2). In (Case C3), $\sqrt{n_1}(F_n^{\mathrm{L}}(\delta)-F^{\mathrm{L}}(\delta))$ has an asymptotic distribution of $\sup_{y\in\mathcal{Y}_{\sup,\delta}}G(y,\delta)$ because $M(\delta)>0$ and $\mathcal{Y}_{\sup,\delta}$ has two interior points whereas $\sqrt{n_1}(F_n^{\mathrm{U}}(\delta)-F^{\mathrm{U}}(\delta))$ has a truncated normal asymptotic distribution since $m(\delta)=0$ and $\mathcal{Y}_{\inf,\delta}$ is a singleton.

Finally, (Case C4) is the opposite of (Case C1). In (Case C4), $M(\delta)<0$ so $\Pr(F_n^{\mathrm{L}}(\delta)=0)\to 1$ and $m(\delta)<0$ with $\mathcal{Y}_{\inf,\delta}$ being a singleton so $\sqrt{n_1}(F_n^{\mathrm{U}}(\delta)-F^{\mathrm{U}}(\delta))$ has a normal asymptotic distribution. Table 1 summarizes these DGPs.

We also generated DGPs for two normal marginal distributions. Table 2 summarizes the cases considered in the simulation. In all of these cases, $\sqrt{n_1}(F_n^{\mathrm{L}}(\delta)-F^{\mathrm{L}}(\delta))$ and $\sqrt{n_1}(F_n^{\mathrm{U}}(\delta)-F^{\mathrm{U}}(\delta))$ have asymptotic normal distributions but we include these DGPs in order to see the finite sample

Table 1. DGPs (Case C1)–(Case C4).

	(Case C1)	(Case C2)
(F_1, F_0, δ)	$\left(C(1/4), C(3/4), \frac{1}{8}\right)$	$\left(C(1/4), C(3/4), 1-\frac{\sqrt{6}}{2}\right)$
F^{L}	$\mathrm{M}(\delta)=F^{\mathrm{L}}(\delta)\approx 0.49$	$M(\delta)=F^{\mathrm{L}}(\delta)=0$
$\mathcal{Y}_{\sup,\delta}$	Singleton, interior point	Singleton, interior point
W_{L},δ	$N(0,\sigma_{\mathrm{L}}^2)$	$\max\{N(0,\sigma_{\mathrm{L}}^2),0\}$
F^{U}	$m(\delta)\approx 0.06$, $F^{\mathrm{U}}(\delta)=1$	$1-m(\delta)=F^{\mathrm{U}}(\delta)\approx 0.9$
$\mathcal{Y}_{\inf,\delta}$	Two boundary points	Two interior points
$W_{U,\delta}$	$\Pr(F_n^{\mathrm{U}}(\delta)=1)\to 1$	$\inf_{y\in\mathcal{Y}_{\inf,\delta}}G(y,\delta)$
	(Case C3)	(Case C4)
(F_1, F_0, δ)	$\left(C(3/4), C(1/4), \frac{\sqrt{6}}{2}-1\right)$	$\left(C(3/4), C(1/4), -\frac{1}{8}\right)$
F^{L}	$M(\delta)=F^{\mathrm{L}}(\delta)\approx 0.1$	$M(\delta)\approx -0.06$, $F^{\mathrm{L}}(\delta)=0$
$\mathcal{Y}_{\sup,\delta}$	Two interior points	Two boundary points
$W_{L,\delta}$	$\sup_{y\in\mathcal{Y}_{\sup,\delta}}G(y,\delta)$	$\Pr(F_n^{\mathrm{L}}(\delta)=0)\to 1$
F^{U}	$1-m\ (\delta)=F^{\mathrm{U}}(\delta)=1$	$1-m(\delta)=F^{\mathrm{U}}(\delta)\approx 0.51$
$\mathcal{Y}_{\inf,\delta}$	Singleton, interior point	Singleton, interior point
$W_{U,\delta}$	$\min\{N(0,\sigma_{\mathrm{U}}^2),0\}$	$N(0,\sigma_{\mathrm{U}}^2)$

Table 2. DGPs (Case N1)–(Case N6).

	(Case N1)	(Case N2)	(Case N3)
(F_1, F_0, δ)	$(N(2,2), N(1,1), 1.3)$	$(N(2,2), N(1,1), 2.6)$	$(N(2,2), N(1,1), 4.5)$
F^{L}	$M(\delta) = F^{\mathrm{L}}(\delta) \approx 0.15$	$M(\delta) = F^{\mathrm{L}}(\delta) \approx 0.51$	$M(\delta) = F^{\mathrm{L}}(\delta) \approx 0.86$
$\mathcal{Y}_{\sup,\delta}$	Singleton	Singleton	Singleton
$W_{L,\delta}$	$N(0, \sigma_{\mathrm{L}}^2)$	$N(0, \sigma_{\mathrm{L}}^2)$	$N(0, \sigma_{\mathrm{L}}^2)$
F^{U}	$1-m(\delta) = F^{\mathrm{U}}(\delta) \approx 0.97$	$1-m(\delta) = F^{\mathrm{U}}(\delta) \approx 1$	$1-m(\delta) = F^{\mathrm{U}}(\delta) \approx 1$
$\mathcal{Y}_{\inf,\delta}$	Singleton	Singleton	Singleton
$W_{U,\delta}$	$N(0, \sigma_{\mathrm{U}}^2)$	$N(0, \sigma_{\mathrm{U}}^2)$	$N(0, \sigma_{\mathrm{U}}^2)$
	(Case N4)	(Case N5)	(Case N6)
(F_1, F_0, δ)	$(N(2,2), N(1,1), -2.4)$	$(N(2,2), N(1,1), -0.6)$	$(N(2,2), N(1,1), 0.7)$
F^{L}	$M(\delta) = F^{\mathrm{L}}(\delta) \approx 0$	$M(\delta) = F^{\mathrm{L}}(\delta) \approx 0$	$M(\delta) = F^{\mathrm{L}}(\delta) \approx 0.04$
$\mathcal{Y}_{\sup,\delta}$	Singleton	Singleton	Singleton
$W_{L,\delta}$	$N(0, \sigma_{\mathrm{L}}^2)$	$N(0, \sigma_{\mathrm{L}}^2)$	$N(0, \sigma_{\mathrm{L}}^2)$
F^{U}	$1-m(\delta) = F^{\mathrm{U}}(\delta) \approx 0.16$	$1-m(\delta) = F^{\mathrm{U}}(\delta) \approx 0.49$	$1-m(\delta) = F^{\mathrm{U}}(\delta) \approx 0.85$
$\mathcal{Y}_{\inf,\delta}$	Singleton	Singleton	Singleton
$W_{U,\delta}$	$N(0, \sigma_{\mathrm{U}}^2)$	$N(0, \sigma_{\mathrm{U}}^2)$	$N(0, \sigma_{\mathrm{U}}^2)$

performance of our bootstrap procedures for different values of $F^{\mathrm{L}}(\delta)$ and $F^{\mathrm{U}}(\delta)$. From (Case N1) to (Case N6), $F^{\mathrm{L}}(\delta)$ ranges from being very close to zero to about 0.86 and $F^{\mathrm{U}}(\delta)$ from 0.16 to almost 1.

We now present F_n^{L} and F_n^{U} for the normal marginals (DGPs (Case N1)–(Case N6)) and C (α) class of marginals (DGPs (Case C1)–(Case C4)). 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 eight 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).

Fig. 7(a) and (b) correspond to (Case C1)–(Case C4), while Fig. 7(c) corresponds to (Case N1)–(Case N6). 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 δ is less than 0.0005.

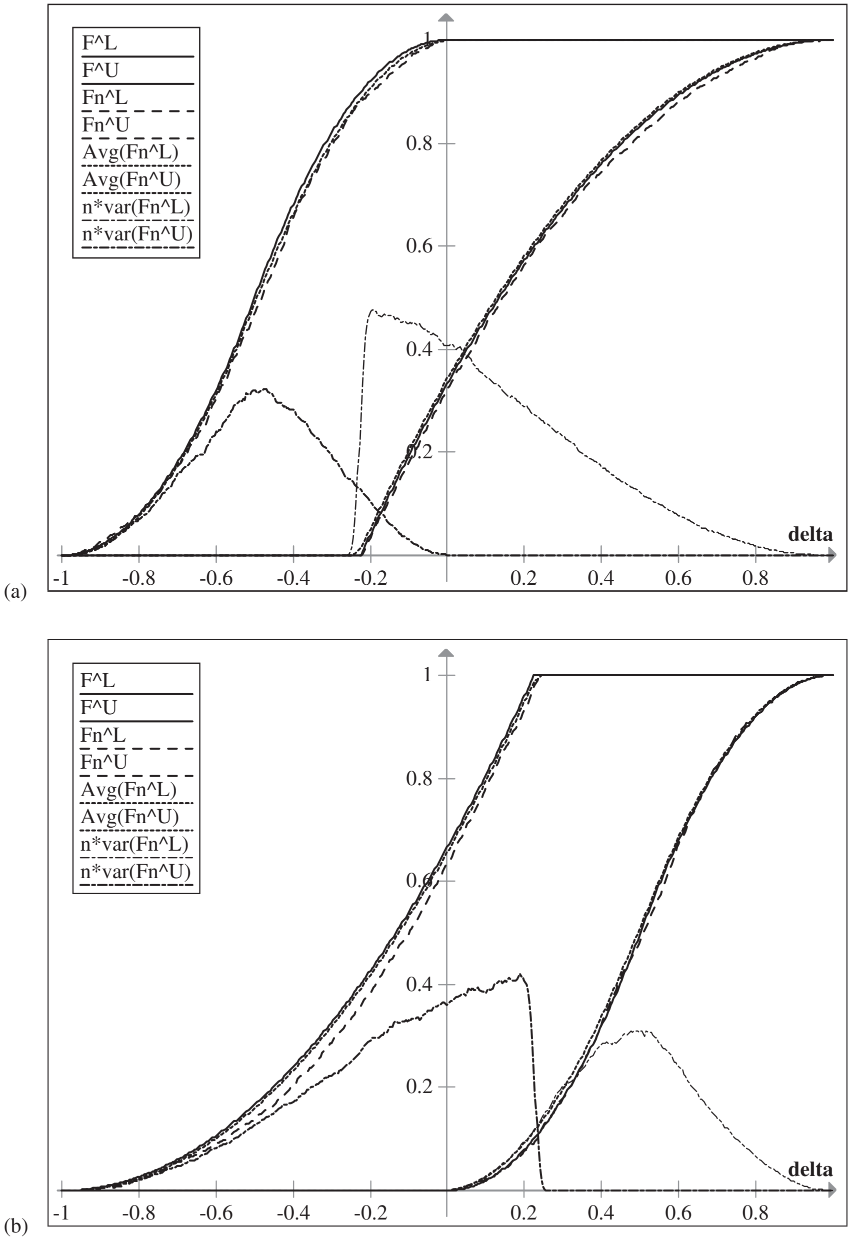


Fig. 7. (a) Estimates of the Distribution Bounds: (*C*(1/4), *C*(3/4)); (b) Estimates of the Distribution Bounds: (*C*(3/4), *C*(1/4)); and (c) Estimates of the Distribution Bounds: (*N*(2,2), *N*(1,1).

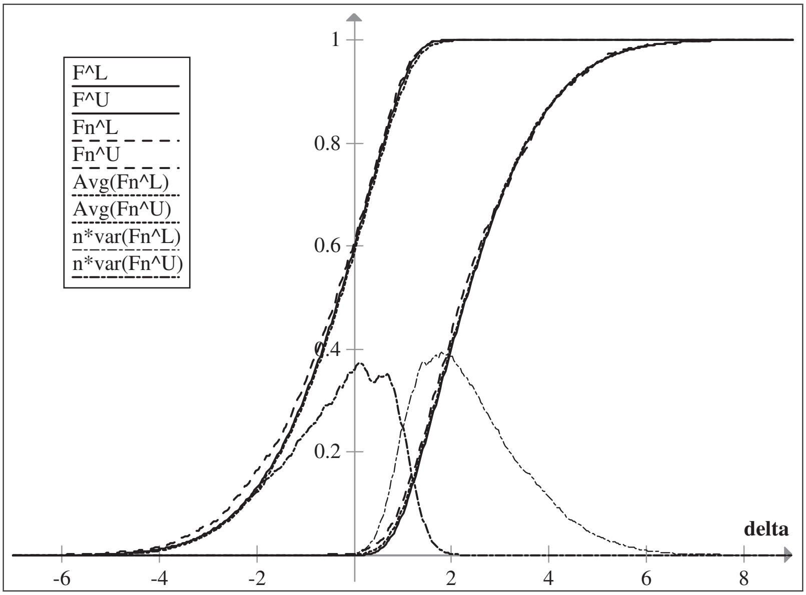
(c) 

Fig. 7. (*Continued*)

7.2. Simulation Results for Coverage Rates

In this and the next subsections, we present simulation results for the bootstrap CSs and the bootstrap bias-corrected estimators. For each DGP, we generated random samples of sizes $n_1 = n_0 = 300$ and 1,000, respectively. The number of replications we used is 2,500 and the number of bootstrap repetitions is $B = 1{,}999$ as suggested in Davidson and Mackinnon (2004, pp. 163–165). The shrinkage parameters are: $b_n = n_1^{-(1/3)}$ and $b'_n = 0.3 n_1^{-(0.95/2)}$, that is, $c = 1.0$, $a = 1/3$, $c' = 0.3$, and $a' = 0.05$ in the expressions in Section 5.1. We used the second procedure based on $W^*_{\mathrm{L},\delta}$ and $W^*_{\mathrm{U},\delta}$. We set $\alpha = 0.05$ throughout the simulations.

Table 3 presents the minimum values of coverage rates of the CSs defined in Theorem 3 (i) ($F_\Delta(\delta)$ columns) and the average values of $\underline{\hat{p}}$ with DGPs (Case C1)–(Case C4). AU :6

The CSs for DGPs (Case C2) and (Case C4) perform very well. As n grows, the coverage rates for DGPs (Case C2) and (Case C3) become closer to the nominal level $1-\alpha = 0.95$. Considering that (Case C2) and (Case C3) are cases where the estimator of one of the two bounds follows a normal

Table 3. Coverage Rates and avg($\underline{\hat{p}}$) for (Case C1)–(Case C4).

	(Case C1)		(Case C2)		(Case C3)		(Case C4)	
	$F_\Delta(\delta)$	avg($\underline{\hat{p}}$)	$F_\Delta(\delta)$	avg($\underline{\hat{p}}$)	$F_\Delta(\delta)$	avg($\underline{\hat{p}}$)	$F_\Delta(\delta)$	avg($\underline{\hat{p}}$)
$n = 300$	0.9320	0.9220	0.9360	0.9762	0.9356	0.9766	0.9312	0.9203
$n = 1{,}000$	0.9376	0.9228	0.9488	0.9780	0.9540	0.9786	0.9384	0.9213

Table 4. Coverage Rates and avg($\underline{\hat{p}}$) for (Case N1)–(Case N6).

	(Case N1)		(Case N2)		(Case N3)	
	$F_\Delta(\delta)$	avg($\underline{\hat{p}}$)	$F_\Delta(\delta)$	avg($\underline{\hat{p}}$)	$F_\Delta(\delta)$	avg($\underline{\hat{p}}$)
$n = 300$	0.9304	0.9628	0.9252	0.929	0.9332	0.9007
$n = 1{,}000$	0.9536	0.9626	0.9508	0.9479	0.9492	0.9050
	(Case N4)		(Case N5)		(Case N6)	
	$F_\Delta(\delta)$	avg($\underline{\hat{p}}$)	$F_\Delta(\delta)$	avg($\underline{\hat{p}}$)	$F_\Delta(\delta)$	avg($\underline{\hat{p}}$)
$n = 300$	0.950	0.9182	0.9176	0.9717	0.9444	0.9629
$n = 1{,}000$	0.9492	0.9293	0.950	0.9869	0.9492	0.9643

distribution asymptotically but the estimator of the other bound violates (A3) and (A4), our bootstrap procedure seems to perform very well. The minimum coverage rates for (Case C1) and (Case C4) in which the estimator of one of the two bounds degenerates asymptotically are about 0.93–0.94. They improve slowly as the sample size becomes larger. When $n = 1{,}000$, the coverage rates are still less than 0.94 but a little better than the coverage rates with $n = 300$. The average $\underline{\hat{p}}$ differs from DGP to DGP. (Case C1) and (Case C4), where $F_n^{\mathrm{L}}(\delta)$ or $F_n^{\mathrm{U}}(\delta)$ has a degenerate asymptotic distribution, have $\underline{\hat{p}}$ as low as about 0.92. (Case C2) and (Case C3) have $\underline{\hat{p}}$ about 0.98. In both cases, $\underline{\hat{p}}$ is far greater than $\alpha = 0.05$.

The coverage rates for DGPs (Case N1)–(Case N6) are in Table 4. Recall that in all of these cases, $\sqrt{n_1}(F_n^{\mathrm{L}}(\delta) - F^{\mathrm{L}}(\delta))$ and $\sqrt{n_1}(F_n^{\mathrm{U}}(\delta) - F^{\mathrm{U}}(\delta))$ have asymptotic normal distributions.

The coverage rates for $F_\Delta(\delta)$ increased from about 0.92–0.93 when $n = 300$ to almost 0.95 when $n = 1{,}000$. For (Case N4) and (Case N6), the coverage rates for $n = 300$ are already very good. As in DGPs (Case C1)–(Case C4), the average $\underline{\hat{p}}$ differs from DGP to DGP. Nonetheless, $\underline{\hat{p}}$ is greater than 0.05 for all cases.

7.3. Simulation Results for Bias-Corrected Estimators

In each replication, we computed the bootstrap biases and mean squared errors of F_n^{L} and F_n^{U} as well as $\widehat{F}_{n\mathrm{BC}}^{\mathrm{L}}$ and $\widehat{F}_{n\mathrm{BC}}^{\mathrm{U}}$, where we used the bootstrap bias-correction with the second bootstrap procedure discussed in Section 5.1. "Bias" and "$\sqrt{\mathrm{MSE}}$" in Table 5 represent the average bias and the square roots of the mean squared errors (MSE).

The direction of the bias without correction is as expected. The bias estimates are positive for F_n^{L} and negative for F_n^{U} for all DGPs except for the cases that $\sqrt{n_1}(F_n^{\mathrm{L}}(\delta) - F^{\mathrm{L}}(\delta))$ and $\sqrt{n_1}(F_n^{\mathrm{U}}(\delta) - F^{\mathrm{U}}(\delta))$ degenerate asymptotically (Case C1 for F_n^{L} and Case C4 for F_n^{U}). The bias-correction took

Table 5. Bias and MSE Reduction for (Case C1)–(Case C4).

		(Case C1)		(Case C2)	
		$F_n^{\mathrm{L}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$	$F_n^{\mathrm{L}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$
$n = 300$	Bias	0.0190	0.0003	0.0305	0.0142
	$\sqrt{\mathrm{MSE}}$	0.0382	0.0352	0.0429	0.0263
$n = 1{,}000$	Bias	0.0095	−0.0009	0.0152	0.0066
	$\sqrt{\mathrm{MSE}}$	0.0211	0.0197	0.0220	0.0130
		$F_n^{\mathrm{U}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$	$F_n^{\mathrm{U}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$
$n = 300$	Bias	0	0	−0.0292	−0.0064
	$\sqrt{\mathrm{MSE}}$	0	0	0.0361	0.0253
$n = 1{,}000$	Bias	0	0	−0.0150	−0.0031
	$\sqrt{\mathrm{MSE}}$	0	0	0.0187	0.0134
		(Case C3)		(Case C4)	
		$F_n^{\mathrm{L}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$	$F_n^{\mathrm{L}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$
$n = 300$	Bias	0.0292	0.0064	0	0
	$\sqrt{\mathrm{MSE}}$	0.0348	0.0247	0	0
$n = 1{,}000$	Bias	0.0144	0.0024	0	0
	$\sqrt{\mathrm{MSE}}$	0.0182	0.0131	0	0
		$F_n^{\mathrm{U}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$	$F_n^{\mathrm{U}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$
$n = 300$	Bias	−0.0306	−0.0141	−0.0192	−0.0004
	$\sqrt{\mathrm{MSE}}$	0.0430	0.0265	0.0382	0.0349
$n = 1{,}000$	Bias	−0.0159	−0.0070	−0.0099	0.0004
	$\sqrt{\mathrm{MSE}}$	0.0228	0.0136	0.0211	0.0194

effect with $n=300$ quite dramatically already. In (Case C1) for F_n^{L} and (Case C4) for F_n^{U}, where the asymptotic distributions of those estimators are normal, the magnitude of the bias reduces to roughly about 1/50–1/60 of the bias of F_n^{L} or F_n^{U}. For other DGPs, the magnitude of the bias-reduction is not as great but still the biases reduced by roughly about 1/1.5–1/4.5 of the bias of F_n^{L} or F_n^{U}. The relative magnitude of bias-reduction is similar in $n=1{,}000$ for (Case C2) or (Case C3). It is roughly about $1/2 \sim 1/5$ of the bias of F_n^{L} or F_n^{U}. The bias estimates of $\widehat{F}_{n\mathrm{BC}}^{\mathrm{L}}$ for (Case C1) and $\widehat{F}_{n\mathrm{BC}}^{\mathrm{U}}$ (Case C4) changed sign when $n=1{,}000$. The bootstrap bias-corrected estimators work quite well and we can see huge reduction in bias and changes of signs in (Case C1) for F_n^{L} and (Case C4) for F_n^{U} (where the normal asymptotics holds). We will see the sign change with the DGPs (Case N1)–(Case N6) as well. The bootstrap bias-corrected estimators also have smaller MSEs than F_n^{L} and F_n^{U} as shown in the table. The $\sqrt{\mathrm{MSE}}$ of $\widehat{F}_{n\mathrm{BC}}^{\mathrm{L}}$ and $\widehat{F}_{n\mathrm{BC}}^{\mathrm{U}}$ are roughly 2/3 of the $\sqrt{\mathrm{MSE}}$ of F_n^{L} and F_n^{U} for (Case C2) and (Case C3) but the reduction in $\sqrt{\mathrm{MSE}}$ is not as great in (Case C1) for F_n^{L} and (Case C4) for F_n^{U} as in other DGPs.

Table 6 show that results for (Case N1)–(Case N6) are similar. The sign change happened in all DGPs except for those in which $F^{\mathrm{L}}(\delta) \approx 0$ or $F^{\mathrm{U}}(\delta) \approx 1$. The relative magnitude of the bias in $\widehat{F}_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$ or $\widehat{F}_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$ to the bias in $F_n^{\mathrm{L}}(\delta)$ or $F_n^{\mathrm{U}}(\delta)$ ranges from 1/2 to 1/13. The reduction in $\sqrt{\mathrm{MSE}}$ is not sizable.

8. CONCLUSION

In this paper, we have provided a complete study on partial identification of and inference for the distribution of treatment effects for randomized experiments. For randomized experiments with a known value of a dependence measure between the potential outcomes such as Kendall's τ, we established tighter bounds on the distribution of treatment effects. Estimation of these bounds and inference for the distribution of treatment effects in this case can be done by following Sections 4 and 5 in this paper. When observable covariates are available such that the selection-on-observables assumption holds, Fan (2008) developed estimation and inference procedures for the distribution of treatment effects and Fan and Zhu (2009) established estimation and inference procedures for a general class of functionals of the joint distribution of potential outcomes

Table 6. Bias and MSE Reduction for (Case N1)–(Case N6).

		(Case N1)		(Case N2)		(Case N3)	
		$F_n^{\mathrm{L}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$	$F_n^{\mathrm{L}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$	$F_n^{\mathrm{L}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$
$n = 300$	Bias	0.0233	0.0023	0.0187	0.0011	0.0108	−0.0023
	$\sqrt{\mathrm{MSE}}$	0.0397	0.0354	0.0376	0.0343	0.0226	0.0214
$n = 1{,}000$	Bias	0.0106	−0.0008	0.0088	−0.0011	0.0049	−0.0024
	$\sqrt{\mathrm{MSE}}$	0.0207	0.0187	0.0205	0.0193	0.0121	0.0118
		$F_n^{\mathrm{U}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$	$F_n^{\mathrm{U}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$	$F_n^{\mathrm{U}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$
$n = 300$	Bias	−0.0182	0.0017	−0.0011	−0.0001	0	0
	$\sqrt{\mathrm{MSE}}$	0.0276	0.0207	0.0024	0.0005	0.0001	0
$n = 1{,}000$	Bias	−0.0087	0.0024	−0.0005	0.0	0.0	0.0
	$\sqrt{\mathrm{MSE}}$	0.0144	0.0120	0.0010	0.0001	0.0	0.0
		(Case N4)		(Case N5)		(Case N6)	
		$F_n^{\mathrm{L}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$	$F_n^{\mathrm{L}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$	$F_n^{\mathrm{L}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{L}}(\delta)$
$n = 300$	Bias	0.0	0.0	0.0013	0.0001	0.0192	−0.0009
	$\sqrt{\mathrm{MSE}}$	0.0002	0.0	0.0026	0.0005	0.0286	0.0210
$n = 1{,}000$	Bias	0.0	0.0	0.0005	0.0	0.0089	−0.0021
	$\sqrt{\mathrm{MSE}}$	0.0001	0.0	0.0005	0.0	0.0145	0.0118
		$F_n^{\mathrm{U}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$	$F_n^{\mathrm{U}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$	$F_n^{\mathrm{U}}(\delta)$	$F_{n\mathrm{BC}}^{\mathrm{U}}(\delta)$
$n = 300$	Bias	−0.0111	0.0024	−0.0195	−0.0017	−0.0229	−0.0019
	$\sqrt{\mathrm{MSE}}$	0.0228	0.0213	0.0381	0.0344	0.0385	0.0344
$n = 1{,}000$	Bias	−0.0055	0.0019	−0.0085	0.0014	−0.0104	0.0009
	$\sqrt{\mathrm{MSE}}$	0.0127	0.012	0.02	0.0187	0.0209	0.0189

including many commonly used inequality measures of the distribution of treatment effects.

This paper has focused on binary treatments. The results can be easily extended to multivalued treatments. For example, consider a randomized experiment on a treatment taking values in {0, 1, …, T}. Define the treatment effect between t and t' as $\Delta_{t',t} = Y_{t'} - Y_t$ for any $t, t' \in \{0, 1, \ldots, T\}$ and $t \neq t'$. Then by substituting Y_1 with $T_{t'}$ and Y_0 with Y_t, the results in this paper apply to $F_{\Delta t',t}$. The results in this paper can also be extended to continuous treatments, provided that the marginal distribution of the potential outcome corresponding to a given level of treatment intensity is identified.

UNCITED REFERENCES

Andrews (2000); Stoye (2008a)

NOTES

1. In the rest of this paper, we refer to ideal randomized experiments (data) as randomized experiments (data).

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

3. Frank et al. (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:

$$F^{\mathrm{L}}(\delta) = \Phi\left(\frac{-\sigma_1 s - \sigma_0 t}{\sigma_0^2 - \sigma_1^2}\right) + \Phi\left(\frac{\sigma_0 s - \sigma_1 t}{\sigma_0^2 - \sigma_1^2}\right) - 1$$

$$F^{\mathrm{U}}(\delta) = \Phi\left(\frac{-\sigma_1 s + \sigma_0 t}{\sigma_0^2 - \sigma_1^2}\right) + \Phi\left(\frac{\sigma_0 s + \sigma_1 t}{\sigma_0^2 - \sigma_1^2}\right)$$

4. 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. This would not affect the results to follow. For notational compactness, we assume that they are known.

ACKNOWLEDGMENTS

We thank the editors of the *Advances in Econometrics*, Vol. 24, T. Fomby, R. Carter Hill, Q. Li, and J. S. Racine, participants of the 7th annual Advances in Econometrics Conference, and two referees for helpful comments that improved both the exposition and content of this paper.

REFERENCES

Aakvik, A., Heckman, J., & Vytlacil, E. (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary among observationally identical persons: An application to Norwegian vocational rehabilitation programs. *Journal of Econometrics*, *125*, 15–51.

Abadie, A., Angrist, J., & Imbens, G. (2002). Instrumental variables estimation of quantile treatment effects. *Econometrica*, *70*, 91–117.

Abbring, J. H., & Heckman, J. (2007). Econometric evaluation of social programs, Part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. *The Handbook of Econometrics*, *6B*, 5145–5301.

Alsina, C. (1981). Some functional equations in the space of uniform distribution functions. *Equationes Mathematicae*, *22*, 153–164.

Andrews, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, *68*, 399–405.

Andrews, D. W. K., & Guggenberger, P. (2005a). *The limit of finite-sample size and a problem with subsampling*. Unpublished Manuscript, Cowles Foundation, Yale University, New Haven, CT.

Andrews, D. W. K., & Guggenberger, P. (2005b). *Hybrid and size-corrected subsampling methods*. Unpublished Manuscript, Cowles Foundation, Yale University, New Haven, CT.

Andrews, D. W. K., & Guggenberger, P. (2005c). *Applications of subsampling, hybrid, and size-correction methods*. Cowles Foundation Discussion Paper No. 1608, Yale University, New Haven, CT.

Andrews, D. W. K., & Guggenberger, P. (2007). *Validity of subsampling and 'plug-in asymptotic' inference for parameters defined by moment inequalities*. Unpublished Manuscript, Cowles Foundation, Yale University, New Haven, CT.

Andrews, D. W. K., & Soares, G. (2007). *Inference for parameters defined by moment inequalities using generalized moment selection*. Cowles Foundation Working Paper no. 1631, Yale University, New Haven, CT.

Biddle, J., Boden, L., & Reville, R. (2003). *A method for estimating the full distribution of a treatment effect, with application to the impact of workfare injury on subsequent earnings*. Mimeo. AU :7

Bitler, M., Gelbach, J., & Hoynes, H. W. (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, *96*, 988–1012.

Black, D. A., Smith, J. A., Berger, M. C., & Noel, B. J. (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.

Bugni, F.A. (2007). Bootstrap inference in partially identified models. Mimeo, Northwestern University.

Cambanis, S., Simons, G., & Stout, W. (1976). Inequalities for $\varepsilon k(X, Y)$ when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *36*, 285–294.

Canay, I. A. (2007). *EL inference for partially identified models: Large deviations optimality and bootstrap validity*. Manuscript, University of Wisconsin.

Carneiro, P., Hansen, K. T., & Heckman, J. (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.

Chernozhukov, V., & Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica*, *73*, 245–261.

Chernozhukov, V., Hong, H., & Tamer, E. (2007). Parameter set inference in a class of econometric models. *Econometrica*, *75*, 1243–1284.

Davidson, R., & Mackinnon, J. G. (2004). *Econometric theory and method*. Oxford University Press. AU :8

Dehejia, R. (1997). *A decision-theoretic approach to program evaluation*. Ph.D. Dissertation, Department of Economics, Harvard University.

Dehejia, R., & Wahba, S. (1999). Causal effects in non-experimental studies: Re-evaluating the evaluation of training programs. *Journal of the American Statistical Association*, *94*, 1053–1062.

Denuit, M., Genest, C., & Marceau, E. (1999). Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, *25*, 85–104.

Djebbari, H., & Smith, J. A. (2008). Heterogeneous impacts in PROGRESA. *Journal of Econometrics*, *145*, 64–80.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Annals of Statistics*, *2*, 267–277.

Embrechts, P., Hoeing, A., & Juri, A. (2003). Using copulae to bound the value-at-risk for functions of dependent risks. *Finance & Stochastics*, *7*(2), 145–167.

Fan, Y. (2008). *Confidence sets for distributions of treatment effects with covariates*. Vanderbilt University (work in progress). AU :9

Fan, Y., & Park, S. (2007a). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory* (forthcoming).

Fan, Y., & Park, S. (2007b). *Confidence sets for the quantile of treatment effects*. Manuscript, Vanderbilt University.

Fan, Y., & Park, S. (2007c). *Confidence sets for some partially identified parameters*. Manuscript, Vanderbilt University.

Fan, Y., & Wu, J. (2007). *Sharp bounds on the distribution of the treatment effect in switching regimes models*. Manuscript, Vanderbilt University.

Fan, Y., & Zhu, D. (2009). *Partial identification and confidence sets for parameters of the joint distribution of the potential outcomes*. Vanderbilt University (work in progress).

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, *75*, 259–276.

Firpo, S., & Ridder, G. (2008). *Bounds on functionals of the distribution of treatment effects*. Working Paper.

Frank, M. J., Nelsen, R. B., & Schweizer, B. (1987). Best-possible bounds on the distribution of a sum – a problem of Kolmogorov. *Probability Theory and Related Fields*, *74*, 199–211.

Galichon, A., & Henry, M. (2008). A test of non-identifying restrictions and confidence regions for partially identified parameters. *Journal of Econometrica* (forthcoming).

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, *66*, 315–331.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, *66*, 1017–1098.

Heckman, J., & Robb, R. (1985). Alternative methods for evaluating the impact of interventions. In: J. Heckman & B. Singer (Eds), *Longitudinal analysis of labor market data*. New York: Cambridge University Press.

Heckman, J., & Smith, J. (1993). Assessing the case for randomized evaluation of social programs. In: K. Jensen & P. K. Madsen (Eds), *Measuring labour market measures: Evaluating the effects of active labour market policies* (pp. 35–96). Copenhagen, Denmark: Danish Ministry of Labor.

Heckman, J., Smith, J., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for Heterogeneity in Programme Impacts. *Review of Economic Studies*, *64*, 487–535.

Heckman, J., & Vytlacil, E. (2007a). Econometric evaluation of social programs. Part I: Causal models, structural models and econometric policy evaluation. *The Handbook of Econometrics*, *6B*, 4779–4874.

Heckman, J., & Vytlacil, E. (2007b). Econometric evaluation of social programs. Part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *The Handbook of Econometrics*, *6B*, 4875–5143.

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, *71*, 1161–1189.

Horowitz, J. L., & Manski, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*, *95*, 77–84.

Imbens, G. W., & Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, *72*, 1845–1857.

Imbens, G. W., & Newey, W. (2005). *Identification and estimation of triangular simultaneous equations models without additivity*. Working Paper.

Imbens, G. W., & Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*, *64*, 555–574.

Joe, H. (1997). *Multivariate models and dependence concepts*. London, UK: Chapman & Hall/CRC.

Lalonde, R. (1995). The promise of public sector-sponsored training programs. *Journal of Economic Perspectives*, *9*, 149–168.

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.

Lee, L. F. (2002). *Correlation bounds for sample selection models with mixed continuous, discrete and count data variables*. Manuscript, The Ohio State University, Athens, OH.

Lee, M. J. (2005). *Micro-econometrics for policy, program, and treatment effects*. Oxford University Press.

Lehmann, E. L. (1974). *Nonparametrics: Statistical methods based on ranks*. San Francisco, CA: Holden-Day Inc.

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.

Manski, C. F. (1990). Non-parametric bounds on treatment effects. *American Economic Review, Papers and Proceedings*, *80*, 319–323.

Manski, C. F. (1997a). Monotone treatment effect. *Econometrica*, *65*, 1311–1334.

Manski, C. F. (1997b). The mixing problem in programme evaluation. *Review of Economic Studies*, *64*, 537–553.

Manski, C. F. (2003). *Partial identification of probability distributions*. New York: Springer-Verlag.

Moon, R., & Schorfheide, F. (2007). *A Bayesian look at partially-identified models*. Manuscript, University of Pennsylvania, Philadelphia, PA.

Nelsen, R. B. (1999). *An introduction to copulas*. New York: Springer.

Nelsen, R. B., Quesada-Molina, J. J., Rodriguez-Lallena, J. A., & Ubeda-Flores, M. (2001). Bounds on bivariate distribution functions with given margins and measures of association. *Communications in Statistics: Theory and Methods*, *30*, 1155–1162.

Nelsen, R. B., Quesada-Molina, J. J., Rodriguez-Lallena, J. A., & Ubeda-Flores, M. (2004). Best-possible bounds on sets of bivariate distribution functions. *Journal of Multivariate Analysis*, *90*, 348–358.

Nelsen, R. B., & Ubeda-Flores, M. (2004). A comparison of bounds on sets of joint distribution functions derived from various measures of association. *Communications in Statistics: Theory and Methods*, *33*, 2299–2305.

Romano, J., & Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference* (forthcoming).

Rosen, A. (2008). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *Journal of Econometrics*, *146*, 107–117.

Rosenbaum, P. R., & Rubin, D. B. (1983). 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.

Schweizer, B., & Sklar, A. (1983). *Probabilistic metric spaces*. New York: North-Holland.

Sklar, A. (1959). Functions de réartition à n dimensions et leures marges. *Publications de l'Institut de Statistique de L'Université de Paris*, *8*, 229–231.

Soares, G. (2006). Inference for partially identified models with inequality moment constraints. Working Paper, Yale University, New Haven, CT.

Stoye, J. (2008a). More on confidence intervals for partially identified parameters. *Econometrica* (forthcoming).

Stoye, J. (2008b). *Partial identification of spread parameters*. Working Paper, New York University, New York, NY.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *Annals of Probability*, *8*, 814–827.

Tesfatsion, L. (1976). Stochastic dominance and the maximization of expected utility. *Review of Economic Studies*, *43*, 301–315.

Williamson, R. C., & Downs, T. (1990). Probabilistic arithmetic I: Numerical Methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, *4*, 89–158.

APPENDIX A. PROOF OF EQ. (23)

Obviously, one can take $1-\underline{p} = \lim_{n_1\to\infty}\inf_{\theta_0\in[\theta_{\mathrm{L}},\theta_{\mathrm{U}}]}\Pr(\theta_0 \in \{\theta : T_n(\theta) \le 0)\}$. Now,

$$\begin{aligned}&\lim_{n_1\to\infty}\inf_{\theta_0\in[\theta_{\mathrm{L}},\theta_{\mathrm{U}}]}\Pr(\theta_0 \in \{\theta : T_n(\theta) \le 0)\\ &\quad = \inf \Pr[(W_{\mathrm{L},\delta} - h^{\mathrm{L}}(\theta_0))_+^2 + (W_{\mathrm{U},\delta} + h^{\mathrm{U}}(\theta_0))_-^2 = 0]\end{aligned}$$

We need to show that

$$\begin{aligned}&\inf \Pr[(W_{\mathrm{L},\delta} - h^{\mathrm{L}}(\theta_0))_+^2 + (W_{\mathrm{U},\delta} + h^{\mathrm{U}}(\theta_0))_-^2 = 0]\\ &\quad = \Pr\left[\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta) \le 0,\ \inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta) \ge 0\right]\end{aligned}$$

First, we consider the case with $W_{L,\delta} - h^{L}(\theta_0) \le 0$. We have:

$$\begin{aligned} & W_{L,\delta} - h^{L}(\theta_0) \le 0 \\ & \Leftrightarrow \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), -h_L(\delta)\right\} \le -\min\{h_L(\delta),0\} + h^{L}(\theta_0) \\ & \Leftrightarrow \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), -h_L(\delta)\right\} \le -h_L(\delta) + \lim_{n_1\to\infty}\sqrt{n_1}F_\Delta(\delta) \\ & \Leftrightarrow \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}M(\delta)\right\} \le \lim_{n_1\to\infty}\sqrt{n_1}[F_\Delta(\delta) - M(\delta)] \end{aligned}$$

since

$$\begin{aligned} h^{L}(\theta_0) &= -\lim_{n_1\to\infty}[\sqrt{n_1}F^{L}(\delta) - \sqrt{n_1}F_\Delta(\delta)] \\ &= -\lim_{n_1\to\infty}[\max\{\sqrt{n_1}M(\delta),0\} - \sqrt{n_1}F_\Delta(\delta)] \\ &= -\max\left\{\lim_{n_1\to\infty}\sqrt{n_1}M(\delta),0\right\} + \lim_{n_1\to\infty}\sqrt{n_1}F_\Delta(\delta) \end{aligned}$$

(i) If $F_\Delta(\delta) = F^{L}(\delta) = 0 > M(\delta)$, then

$$\begin{aligned} & \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}M(\delta)\right\} \le \lim_{n_1\to\infty}\sqrt{n_1}[F_\Delta(\delta) - M(\delta)] \\ & \Leftrightarrow \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), \infty\right\} \le \infty \\ & \Leftrightarrow \sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta) < \infty \end{aligned}$$

which holds trivially.

(ii) If $F_\Delta(\delta) = F^{L}(\delta) = 0 = M(\delta)$, then

$$\begin{aligned} & \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}M(\delta)\right\} \le \lim_{n_1\to\infty}\sqrt{n_1}[F_\Delta(\delta) - M(\delta)] \\ & \Leftrightarrow \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), 0\right\} \le 0 \\ & \Leftrightarrow \sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta) \le 0 \end{aligned}$$

(iii) If $F_\Delta(\delta) = F^{\mathrm{L}}(\delta) = M(\delta) > 0$, then

$$\max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}M(\delta)\right\} \le \lim_{n_1\to\infty}\sqrt{n_1}[F_\Delta(\delta) - M(\delta)]$$

$$\Leftrightarrow \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), -\infty\right\} \le 0$$

$$\Leftrightarrow \sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta) \le 0$$

(iv) If $F_\Delta(\delta) = F^{\mathrm{L}}(\delta) = 0 > M(\delta)$, then

$$\max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}M(\delta)\right\} \le \lim_{n_1\to\infty}\sqrt{n_1}[F_\Delta(\delta) - M(\delta)]$$

$$\Leftrightarrow \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), \infty\right\} \le \infty$$

$$\Leftrightarrow \sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta) < \infty$$

which holds trivially.

(v) If $F_\Delta(\delta) > F^{\mathrm{L}}(\delta) = 0 = M(\delta)$, then

$$\max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}M(\delta)\right\} \le \lim_{n_1\to\infty}\sqrt{n_1}[F_\Delta(\delta) - M(\delta)]$$

$$\Leftrightarrow \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), 0\right\} \le \infty$$

$$\Leftrightarrow \sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta) < \infty$$

which holds trivially.

(vi) If $F_\Delta(\delta) > F^{\mathrm{L}}(\delta) = M(\delta) > 0$, then

$$\max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}M(\delta)\right\} \le \lim_{n_1\to\infty}\sqrt{n_1}[F_\Delta(\delta) - M(\delta)]$$

$$\Leftrightarrow \max\left\{\sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta), \infty\right\} \le \infty$$

$$\Leftrightarrow \sup_{y\in\mathcal{Y}_{\sup,\delta}} G(y,\delta) < \infty$$

which holds trivially.

Summarizing (i)–(vi), we have

$$W_{L,\delta} - h^{L}(\theta_0) \le 0 \Leftrightarrow \sup_{y \in \mathcal{Y}_{\sup,\delta}} G(y,\delta) \le 0$$

if $F_{\Delta}(\delta) = F^{L}(\delta) = M(\delta) \ge 0$; otherwise it holds trivially.

Similarly to the $W_{L,\delta} - h^{L}(\theta_0) \ge 0$ case, we get

$$\begin{aligned}
&W_{U,\delta} + h^{U}(\theta_0) \ge 0\\
&\Leftrightarrow \min\left\{\inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta), -h_U(\delta)\right\} + \max\{h_U(\delta), 0\} + h^{U}(\theta_0) \ge 0\\
&\Leftrightarrow \min\left\{\inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta), -h_U(\delta)\right\} \ge -\max\{h_U(\delta), 0\} - \lim_{n\to\infty} \sqrt{n}[F^{U}(\delta) - F_{\Delta}(\delta)]\\
&\Leftrightarrow \min\left\{\inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta), -\lim_{n_1\to\infty} \sqrt{n_1}m(\delta)\right\} \ge -\lim_{n_1\to\infty}[1 + m(\delta) - F_{\Delta}(\delta)]
\end{aligned}$$

since

$$\begin{aligned}
h^{U}(\theta_0) &= \lim_{n_1\to\infty}[\sqrt{n_1}F^{U}(\delta) - \sqrt{n_1}F_{\Delta}(\delta)]\\
&= \lim_{n_1\to\infty} \sqrt{n_1}\min\{m(\delta), 0\} + \lim_{n_1\to\infty} \sqrt{n_1}(1 - F_{\Delta}(\delta))\\
&= \min\{h_U(\delta), 0\} + \lim_{n_1\to\infty} \sqrt{n_1}(1 - F_{\Delta}(\delta))
\end{aligned}$$

(i) If $1 + m(\delta) > 1 = F^{U}(\delta) = F_{\Delta}(\delta)$, then

$$\begin{aligned}
&\min\left\{\inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta), -\lim_{n_1\to\infty} \sqrt{n_1}m(\delta)\right\} \ge -\lim_{n_1\to\infty}[1 + m(\delta) - F_{\Delta}(\delta)]\\
&\Leftrightarrow \min\left\{\inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta), -\infty\right\} \ge -\infty\\
&\Leftrightarrow \inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta) \ge -\infty
\end{aligned}$$

which holds trivially.

(ii) If $1 + m(\delta) = 1 = F^{U}(\delta) = F_{\Delta}(\delta)$, then

$$\begin{aligned}
&\min\left\{\inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta), -\lim_{n_1\to\infty} \sqrt{n_1}m(\delta)\right\} \ge -\lim_{n_1\to\infty}[1 + m(\delta) - F_{\Delta}(\delta)]\\
&\Leftrightarrow \min\left\{\inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta), 0\right\} \ge 0\\
&\Leftrightarrow \inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta) \ge 0
\end{aligned}$$

(iii) If $1 > 1 + m(\delta) = F^{\mathrm{U}}(\delta) = F_{\Delta}(\delta)$, then

$$\begin{aligned}
&\min\left\{\inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}m(\delta)\right\} \geq -\lim_{n_1\to\infty}[1 + m(\delta) - F_{\Delta}(\delta)] \\
&\quad\Leftrightarrow \min\left\{\inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta), \infty\right\} \geq 0 \\
&\quad\Leftrightarrow \inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta) \geq 0
\end{aligned}$$

(iv) If $1 + m(\delta) > 1 = F^{\mathrm{U}}(\delta) > F_{\Delta}(\delta)$, then

$$\begin{aligned}
&\min\left\{\inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}m(\delta)\right\} \geq -\lim_{n_1\to\infty}[1 + m(\delta) - F_{\Delta}(\delta)] \\
&\quad\Leftrightarrow \min\left\{\inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta), -\infty\right\} \geq -\infty \\
&\quad\Leftrightarrow \inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta) \geq -\infty
\end{aligned}$$

which holds trivially.

(v) If $1 + m(\delta) = 1 = F^{\mathrm{U}}(\delta) > F_{\Delta}(\delta)$, then

$$\begin{aligned}
&\min\left\{\inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}m(\delta)\right\} \geq -\lim_{n_1\to\infty}[1 + m(\delta) - F_{\Delta}(\delta)] \\
&\quad\Leftrightarrow \min\left\{\inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta), 0\right\} \geq -\infty \\
&\quad\Leftrightarrow \inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta) \geq -\infty
\end{aligned}$$

which holds trivially.

(vi) If $1 > 1 + m(\delta) = F^{\mathrm{U}}(\delta) > F_{\Delta}(\delta)$, then

$$\begin{aligned}
&\min\left\{\inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta), -\lim_{n_1\to\infty}\sqrt{n_1}m(\delta)\right\} \geq -\lim_{n_1\to\infty}[1 + m(\delta) - F_{\Delta}(\delta)] \\
&\quad\Leftrightarrow \min\left\{\inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta), \infty\right\} \geq -\infty \\
&\quad\Leftrightarrow \inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta) \geq -\infty
\end{aligned}$$

which holds trivially. Summarizing (i)–(vi), we get

$$W_{\mathrm{U},\delta} + h^{\mathrm{U}}(\theta_0) \geq 0 \Leftrightarrow \inf_{y\in\mathcal{Y}_{\inf,\delta}} G(y,\delta) \geq 0$$

if $1 \geq 1 + m(\delta) = F^{\mathrm{U}}(\delta) = F_{\Delta}(\delta)$; otherwise it holds trivially.

Finally, we obtain:

$$
\begin{aligned}
&\inf \ \Pr[(W_{L,\delta} - h^{L})(\theta_0)_+^2 + (W_{U,\delta} + h^{U}(\theta_0))_-^2 = 0] \\
&\quad = \inf \ \Pr[W_{L,\delta} - h^{L}(\theta_0) \leq 0, W_{U,\delta} + h^{U}(\theta_0) \geq 0] \\
&\quad = \Pr\left[\sup_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta) \leq 0, \inf_{y \in \mathcal{Y}_{\inf,\delta}} G(y,\delta) \geq 0 \right]
\end{aligned}
$$

APPENDIX B. EXPRESSIONS FOR $y_{\sup,\delta}$, $y_{\inf,\delta}$, $m(\delta)$ AND $m(\delta)$ FOR SOME KNOWN MARGINAL DISTRIBUTIONS

Denuit et al. (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 $Y_1 \sim C(\alpha_1)$ and $Y_0 \sim C(\alpha_0)$. We now provide the functional form of $F_1(y) - F_0(y-\delta)$.

1. Suppose $\delta < 0$. Then $\mathcal{Y}_\delta = [0, 1+d]$.
   (a) If $a_0 + \delta \leq 0 < a_1 \leq 1 + \delta$, then

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

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

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

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

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

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

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

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

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

2. Suppose $\delta \ge 0$. Then $\mathcal{Y}_\delta = [\delta, 1]$.
   (a) If $\delta < a_0 + \delta \le a_1 < 1$, then
      (i) if $a_1 \ne a_0$ and $\delta \ne 0$, then

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

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

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

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

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

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

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

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

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

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

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

(Shifted) Exponential marginals. The marginal distributions are:

$$F_1(y) = 1 - \exp\left(-\frac{y-\theta_1}{\alpha_1}\right) \text{ for } y \in [\theta_1, \infty) \quad \text{and}$$

$$F_0(y) = 1 - \exp\left(-\frac{y-\theta_0}{\alpha_0}\right) \quad \text{for} \quad y \in [\theta_0, \infty), \quad \text{where } \alpha_1, \theta_1, \alpha_0, \theta_0 > 0$$

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 $\delta \leq \delta_c$,

$$F^{\mathrm{L}}(\delta) = \max\{M(\delta), 0\} = 0$$

$$\text{where } M(\delta) = \left(\left(\frac{\alpha_0}{\alpha_1}\right)^{\alpha_1/(\alpha_1-\alpha_0)} - \left(\frac{\alpha_0}{\alpha_1}\right)^{\alpha_0/(\alpha_1-\alpha_0)} \right) \exp\left(-\frac{\delta - (\theta_1 - \theta_0)}{\alpha_1 - \alpha_0}\right) < 0$$

$$\text{and } y_{\inf,\delta} = \frac{\alpha_0\alpha_1(\ln\, \alpha_1 - \ln\, \alpha_0) + \alpha_1\theta_0 - \alpha_0\theta_1 + \alpha_1\delta}{\alpha_1 - \alpha_0} \text{(an interior solution)}$$

$$F^{\mathrm{U}}(\delta) = 1 + \min\{m(\delta), 0\} = 1 + m(\delta)$$

$$\text{where } m(\delta) = \min\left\{ \exp\left(-\frac{\max\{\theta_1 - (\delta + \theta_0), 0\}}{\alpha_0}\right) - \exp\left(-\frac{\max\{\theta_0 + \delta - \theta_1, 0\}}{\alpha_1}\right), 0 \right\}$$

$$\text{and } y_{\sup,\delta} = \max\{\theta_1, \theta_0 + \delta\} \quad \text{or} \quad \infty \text{ (boundary solution)}$$

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

$$F^{\mathrm{L}}(\delta) = \max\{M(\delta), 0\} = M(\delta) > 0$$

$$\text{where } M(\delta) = 1 - \exp\left(-\frac{\delta + \theta_0 - \theta_1}{\alpha_1}\right) \text{and } y_{\inf,\delta} = \theta_0 + \delta$$

$$F^{\mathrm{U}}(\delta) = 1 + \min\{m(\delta), 0\} = 1$$

$$\text{since } m(\delta) = 0 \quad \text{and} \quad y_{\sup,\delta} = \infty$$

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

$$F^{\mathrm{L}}(\delta) = \max\{M(\delta), 0\} = M(\delta)$$

$$\text{where } M(\delta) = \begin{cases} 0 & \text{if } \delta \leq \theta_1 - \theta_0 \\ 1 - \exp\left(-\frac{\delta - (\theta_1 - \theta_0)}{\alpha}\right) > 0 & \text{if } \delta > \theta_1 - \theta_0 \end{cases}$$

$$\text{and } y_{\inf,\delta} = \begin{cases} \infty & \text{if } \delta < \theta_1 - \theta_0 \\ \text{any point in } \mathcal{R} & \text{if } \delta = \theta_1 - \theta_0 \\ \theta_0 + \delta & \text{if } \delta > \theta_1 - \theta_0 \end{cases}$$

$$F^{\mathrm{U}}(\delta) = 1 + \min\{m(\delta), 0\} = 1 + m(\delta)$$

$$\text{where } m(\delta) = \begin{cases} \exp\left(-\frac{\theta_1 - (\delta + \theta_0)}{\alpha}\right) - 1 < 0 & \text{if } \delta < \theta_1 - \theta_0 \\ 0 & \text{if } \delta \geq \theta_1 - \theta_0 \end{cases}$$

$$\text{and } y_{\sup,\delta} = \begin{cases} \theta_1 & \text{if } \delta < \theta_1 - \theta_0 \\ \text{any point in } \mathcal{R} & \text{if } \delta = \theta_1 - \delta_0 \\ \infty & \text{if } \delta > \theta_1 - \theta_0 \end{cases}$$

3. Suppose $\alpha_1 > \alpha_0$.
   (a) If $\delta < \delta_c$,

$$F^{L}(\delta) = \max\{M(\delta), 0\} = 0, \quad \text{since } M(\delta) = 0 \text{ and } y_{\inf,\delta} = \infty$$

$$F_{U}(\delta) = 1 + \min\{m(\delta), 0\} = 1 + m(\delta)$$

$$\text{where } m(\delta) = \exp\left(-\frac{\theta_1 - (\delta + \theta_0)}{\alpha_0}\right) - 1 < 0, \;\; y_{\sup,\delta} = \theta_1$$

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

$$F^{L}(\delta) = \max\{M(\delta), 0\} = M(\delta)$$

$$\text{where } M(\delta) = \max\left\{\exp\left(-\frac{\max\{\theta_1 - (\delta + \theta_0), 0\}}{\alpha_0}\right) - \exp\left(-\frac{\max\{\theta_0 + \delta - \theta_1, 0\}}{\alpha_1}\right), 0\right\}$$

$$\text{and } y_{\inf,\delta} = \max\{\theta_1, \theta_0 + \delta\} \quad \text{or} \quad \infty \text{ (boundary solution)}$$

$$F^{U} = 1 + \min\{m(\delta), 0\} = 1 + m(\delta)$$

$$\text{where } m(\delta) = \left(\left(\frac{\alpha_0}{\alpha_1}\right)^{\alpha_1/(\alpha_1-\alpha_0)} - \left(\frac{\alpha_0}{\alpha_1}\right)^{\alpha_0/(\alpha_1-\alpha_0)}\right)\exp\left(-\frac{\delta - (\theta_1 - \theta_0)}{\alpha_1 - \alpha_0}\right) < 0$$

$$\text{and } y_{\sup,\delta} = \frac{\alpha_0\alpha_1(\ln\,\alpha_1 - \ln\,\alpha_0) + \alpha_1\theta_0 - \alpha_0\theta_1 + \alpha_1\delta}{\alpha_1 - \alpha_0} \text{(an interior solution)}$$

(Shifted) Pareto marginals. The marginal distributions are:

$$F_1(y) = 1 - \left(\frac{\lambda_1}{\lambda_1 + y - \theta_1}\right)^{\alpha} \quad \text{for} \quad y \in [\theta_1, \infty) \quad \text{and}$$

$$F_0(y) = 1 - \left(\frac{\lambda_0}{\lambda_0 + y - \theta_0}\right)^{\alpha} \quad \text{for} \quad y \in [\theta_0, \infty), \quad \text{where } \alpha, \lambda_1, \theta_1, \lambda_0, \theta_0 > 0$$

Define

$$\delta_c = (\theta_1 - \theta_0) - (\max\{\lambda_1, \lambda_0\})^{\alpha/(\alpha+1)}(\lambda_1^{1/(\alpha+1)} - \lambda_0^{1/(\alpha+1)})$$

1. Suppose $\lambda_1 < \lambda_0$.
   (a) If $\delta \leq \delta_c$, then

$$F^{\mathrm{L}}(\delta) = \max\{M(\delta), 0\} = M(\delta)$$

$$\text{where } M(\delta) = (\lambda_0^{\alpha/(\alpha+1)} - \lambda_1^{\alpha/(\alpha+1)})\left(\frac{\lambda_1^{\alpha/(\alpha+1)} - \lambda_0^{\alpha/(\alpha+1)}}{\delta - \lambda_0 + \lambda_1 - \theta_1 + \theta_0}\right)^{\alpha} > 0$$

$$\text{and } y_{\inf,\delta} = \frac{(\delta + \theta_0 - \lambda_0)\lambda_1^{\alpha/(\alpha+1)} + (\lambda_1 - \theta_1)\lambda_0^{\alpha/(\alpha+1)}}{\lambda_1^{\alpha/(\alpha+1)} - \lambda_0^{\alpha/(\alpha+1)}} \text{(an interior solution)}$$

$$F^{\mathrm{U}}(\delta) = 1 + \min\{m(\delta), 0\} = 1 + m(\delta)$$

$$\text{where } m(\delta) = \min\left\{\left(\frac{\lambda_0}{\lambda_0 + \max\{\theta_1 - \delta - \theta_0, 0\}}\right)^{\alpha} - \left(\frac{\lambda_1}{\lambda_1 + \max\{\theta_0 + \delta - \theta_1, 0\}}\right)^{\alpha}, 0\right\}$$

$$\text{and } y_{\sup,\delta} = \max\{\theta_1, \theta_0 + \delta\} \quad \text{or} \quad \infty \text{ (boundary solution)}$$

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

$$F^{\mathrm{L}}(\delta) = \max\{M(\delta), 0\} = M(\delta)$$

$$\text{where } M(\delta) = 1 - \left(\frac{\lambda_1}{\lambda_1 + \theta_0 + \delta - \theta_1}\right)^{\alpha} \geq 0 \quad \text{and} \quad y_{\inf,\delta} = \theta_0 + \delta$$

$$F^{\mathrm{U}}(\delta) = 1 + \min\{m(\delta), 0\} = 1$$

$$\text{since } m(\delta) = 0 \quad \text{and} \quad y_{\sup,\delta} = \infty$$

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

$$F^{\mathrm{L}}(\delta) = \max\{M(\delta), 0\} = M(\delta)$$

$$\text{where } M(\delta) = \begin{cases} 0 & \text{if } \delta \leq \theta_1 - \theta_0 \\ 1 - \left(\dfrac{\lambda}{\lambda + \delta - (\theta_1 - \theta_0)}\right)^{\alpha} \geq 0 & \text{if } \delta > \theta_1 - \theta_0 \end{cases}$$

$$\text{and } y_{\inf,\delta} = \begin{cases} \infty & \text{if } \delta < \theta_1 - \theta_0 \\ \text{any point in } \mathcal{Y} & \text{if } \delta = \theta_1 - \theta_0 \\ \theta_0 + \delta & \text{if } \delta > \theta_1 - \theta_0 \end{cases}$$

$$F^{\mathrm{U}}(\delta) = 1 + \min\{m(\delta), 0\} = 1 + m(\delta)$$

$$\text{where } m(\delta) = \begin{cases} \left(\dfrac{\lambda}{\lambda - \delta + (\theta_1 - \theta_0)}\right)^{\alpha} - 1 & \text{if } \delta < \theta_1 - \theta_0 \\ 0 & \text{if } \delta \geq \theta_1 - \theta_0 \end{cases}$$

$$\text{and } y_{\sup,\delta} = \begin{cases} \theta_1 & \text{if } \delta < \theta_1 - \theta_0 \\ \text{any point in } \mathcal{Y} & \text{if } \delta = \theta_1 - \theta_0 \\ \infty & \text{if } > \theta_1 - \theta_0 \end{cases}$$

3. Suppose $\lambda_1 > \lambda_0$.
   (a) If $\delta < \delta_c$, then

$$F^{\mathrm{L}}(\delta) = \max\{M(\delta), 0\} = 0 \text{ since } M(\delta) = 0, \quad \text{and} \quad y_{\inf,\delta} = \infty$$

$$F^{\mathrm{U}}(\delta) = 1 + \min\{m(\delta), 0\} = 1 + m(\delta)$$

$$\text{where } m(\delta) = \left(\frac{\lambda_0}{\lambda_0 + \theta_1 - \delta - \theta_0}\right)^{\alpha} - 1 \leq 0 \text{ and } y_{\sup,\delta} = \theta_1$$

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

$$F^{\mathrm{L}}(\delta) = \max\{M(\delta), 0\} = M(\delta)$$

$$\text{where } M(\delta) = \max\left\{\left(\frac{\lambda_0}{\lambda_0 + \max\{\theta_1 - \delta - \theta_0, 0\}}\right)^{\alpha} - \left(\frac{\lambda_1}{\lambda_1 + \max\{\theta_0 + \delta - \theta_1, 0\}}\right)^{\alpha}, 0\right\}$$

$$\text{and } y_{\inf,\delta} = \max\{\theta_1, \theta_0 + \delta\} \text{ or } \infty \text{ (boundary solution)}$$

$$F^{\mathrm{U}}(\delta) = 1 + \min\{m(\delta), 0\} = 1 + m(\delta)$$

$$\text{where } m(\delta) = \left(\lambda_0^{\alpha/(\alpha+1)} - \lambda_1^{\alpha/(\alpha+1)}\right)\left(\frac{\lambda_1^{\alpha/(\alpha+1)} - \lambda_0^{\alpha/(\alpha+1)}}{\delta - \lambda_0 + \lambda_1 - \theta_1 + \theta_0}\right)^{\alpha} < 0$$

$$\text{and } y_{\sup,\delta} = \frac{(\delta + \theta_0 - \lambda_0)\lambda_1^{\alpha/(\alpha+1)} + (\lambda_1 - \theta_1)\lambda_0^{\alpha/(\alpha+1)}}{\lambda_1^{\alpha/(\alpha+1)} - \lambda_0^{\alpha/(\alpha+1)}} \text{ (an interior solution)}$$

AUTHOR QUERY FORM

	Book: AECO-V025 **Chapter: 1**	**Please e-mail or fax your responses and any corrections to:** **E-mail:** **Fax:**

Dear Author,

During the preparation of your manuscript for typesetting, some questions may have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list.

Disk use
Sometimes we are unable to process the electronic file of your article and/or artwork. If this is the case, we have proceeded by:

☐ Scanning (parts of) your article ☐ Rekeying (parts of) your article

☐ Scanning the artwork

Bibliography
If discrepancies were noted between the literature list and the text references, the following may apply:

☐ **The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.**

☐ ***UNCITED REFERENCES:*** **This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or delete it. Any reference not dealt with will be retained in this section.**

Queries and/or remarks

Location in Article	Query / remark	Response
AU:1	Please provide some brief text description in figure captions of Figures 3–6.	
AU:2	The references (Manski, 1997; Galichon and Henry, 2006; Stoye, 2008; Beresteanu and Molinari, 2006) have not been included in the reference list. Please check.	

AU:3	As per the style, Lemmas, Propositions, Theorems, Remarks, and Definitions have been renumbered throughout, please check and confirm..	
AU:4	Citation for Theorem 3.2 is provided, but there is no Therrem 3.2, please check and confirm.	
AU:5	In the following text "are results of the confidence sets", should "Theorem 3 (i)" be "Theorem 2 (i)." Please check and confirm.	
AU:6	In the following text "Table 3 presents the minimum values of coverage", should "Theorem 3 (i)" be "Theorem 2 (i)." Please check and confirm.	
AU:7	Please provide complete details (like, university name, working paper no.) for references (Biddle et al., 2003; Firpo & Ridder, 2008; Imbens & Newey, 2005; Soares, 2006; Stoye, 2008b).	
AU:8	Please provide location for references (Davidson and Mackinnon, 2004; Lee, 2005).	
AU:9	Please update references (Fan, 2008; Fan & Park, 2007a; Fan & Zhu, 2009; Galichon & Henry, 2008; Romano & Shaikh, 2008; Stoye, 2008a), if possible.	