Econometric Theory, **26**, 2010, 931–951.
doi:10.1017/S0266466609990168

SHARP BOUNDS ON THE DISTRIBUTION OF TREATMENT EFFECTS AND THEIR STATISTICAL INFERENCE

Yanqin Fan
Vanderbilt University

Sang Soo Park
University of North Carolina

In this paper, we propose nonparametric estimators of sharp bounds on the distribution of treatment effects of a binary treatment and establish their asymptotic distributions. We note the possible failure of the standard bootstrap with the same sample size and apply the fewer-than-*n* bootstrap to making inferences on these bounds. The finite sample performances of the confidence intervals for the bounds based on normal critical values, the standard bootstrap, and the fewer-than-*n* bootstrap are investigated via a simulation study. Finally we establish sharp bounds on the treatment effect distribution when covariates are available.

1. INTRODUCTION

Evaluating the effect of a treatment or a program is important in diverse disciplines, including social sciences and medical sciences. In medical sciences, randomized clinical trials are often used to evaluate the efficacy of a drug or a procedure in the treatment or prevention of disease. The central problem in the evaluation of a treatment is that any potential outcome that program participants would have received without the treatment is not observed. Because of this missing data problem, most work in the treatment effect literature has focused on the evaluation of various average treatment effects such as the mean of the treatment effects; see the recent book by Lee (2005) for discussion and references. However, empirical evidence strongly suggests that treatment effect heterogeneity prevails in many experiments, and various interesting effects of the treatment are missed

We thank Jinyong Hahn and two anonymous referees for their valuable suggestions that greatly improved the paper. We thank Jianqing Fan, Joel Horowitz, Chuck Manski, Per Mykland, Bryan Shepherd, Elie Tamer, and seminar participants in the Department of Statistics at the University of Chicago and in the Department of Economics at Northwestern University, UNC at Chapel Hill, and Princeton University for helpful discussions. We also thank Jeff Smith and Jörg Stoye for providing useful references. Y. Fan acknowledges financial support from the National Science Foundation. Address correspondence to Yanqin Fan, Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA; e-mail: yanqin.fan@vanderbilt.edu.

 0266-4666/10 $15.00

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 features of the treatment effects other than its mean may be of interest. Two approaches have been proposed in the literature to study the distribution of the treatment effects. The first is the bounding approach originated in Manski (1997a). Assuming monotone treatment response, Manski (1997a) developed sharp bounds on the distribution of the treatment effects. In the second approach, restrictions are imposed on the dependence structure between the potential outcomes such that their joint distribution and the distribution of the treatment effects are identified; see, e.g., Heckman et al. (1997), Biddle, Boden, and Reville (2003), Carneiro, Hansen, and Heckman (2003), and Aakvik, Heckman, and Vytlacil (2005), among others. Abbring and Heckman (2007) provides a detailed survey of recent analyses using the second approach.

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

Sharp bounds on the joint distribution of the potential outcomes with identified marginals are given by the Fréchet-Hoeffding lower and upper bound distributions; see Heckman and Smith (1993), Heckman et al. (1997), Manski (1997b), and Abbring and Heckman (2007) for their applications in program evaluation. For randomized experiments, Heckman et al. (1997) proposed nonparametric estimates of the Fréchet-Hoeffding distribution bounds and developed a test for the "common effect" model by testing the lower bound of the variance of the treatment effects. They also suggested an alternative test based on the difference between the quantile functions of the marginal distributions of the potential outcomes referred to as the quantile treatment effects (QTE); see Firpo (2007) or Section 2 for more references.

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

By making use of the expressions in Williamson and Downs (1990), we propose nonparametric estimators of sharp bounds on the distribution of the treatment effects for randomized experiments and establish their asymptotic properties. It turns out that the asymptotic distributions of these bounds may be discontinuous as functions of the values of the marginal distributions, providing additional examples for which the standard bootstrap with the same sample size may not be asymptotically valid. The failure of the standard bootstrap (bootstrap with the same sample size) in nonregular cases has been pointed out in Andrews (2000), Bickel, Götze, and van Zwet (1997), Beran (1997), and the references therein. Subsampling and fewer-than-n bootstrap have been proposed to rectify the failure of the standard bootstrap; see Andrews (2000), Bickel et al. (1997), Beran (1997), and Politis, Romano, and Wolf (1999) for discussion and references. In this paper, we apply the fewer-than-n bootstrap (Bickel et al. 1997; Bickel and Sakov, 2008) to construct confidence intervals for these sharp bounds. The finite sample performances of the confidence intervals based on the standard normal critical values, the standard bootstrap with the same sample size, and the fewer-than-n bootstrap are compared in a simulation study.

Given sharp bounds on the distribution of treatment effects, we obtain bounds on the class of D-parameters introduced in Manski (1997a). One example of a D-parameter is any quantile of the treatment effect distribution. In addition, we obtain sharp bounds on the class of D_2-parameters of the treatment effect distribution; see Stoye (2009) or Section 2 for the definition of a D_2-parameter. As pointed out in Stoye (2009), many inequality and risk measures are D_2-parameters. These results shed light on the relation and distinction between QTE and the quantile of the treatment effect distribution.

As an initial investigation of a unified approach to bounding or partially identifying the distribution of treatment effects, this paper has focused on randomized experiments. Numerous extensions of the methodologies developed in this paper are possible and worthwhile. Of immediate concern is the incorporation of covariates into the analysis. We extend sharp bounds in Williamson and Downs (1990) to take into account the presence of covariates under the selection-on-observables assumption commonly used in the treatment effect literature; see, e.g., Rosenbaum and Rubin (1983a, 1983b), Hahn (1998), Heckman, Ichimura, Smith, and Todd (1998), and Dehejia and Wahba (1999), among others. In general, taking into account observable covariates tightens the bounds.[1]

The rest of this paper is organized as follows: In Section 2, we review sharp bounds on the distribution of a difference of two random variables and provide bounds on parameters of the treatment effect distribution that respect either first- or second-order stochastic dominance.[2] In Section 3, we propose nonparametric estimators of the distribution bounds and establish their asymptotic properties. Results from a detailed simulation study are provided in Section 4. Section 5 provides sharp bounds on the treatment effect distribution when covariates are available. Section 6 concludes and discusses interesting extensions. Proofs are collected in the Appendix.

Throughout the paper, we use $\Longrightarrow$ 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

We consider a binary treatment with continuous outcomes and use Y_1 to denote the potential outcome from receiving treatment and Y_0 the outcome without treatment. Both Y_1 and Y_0 are one-dimensional. Let $F(y_1, y_0)$ denote the joint distribution of Y_1, Y_0 with marginals $F_1(\cdot)$ and $F_0(\cdot)$, respectively.

The characterization theorem of Sklar (1959) implies that there exists a copula[3] $C(u,v)$: $(u,v) \in [0,1]^2$ 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).

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

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

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

Heckman and Smith (1993), Heckman et al. (1997), and Manski (1997b) applied (1) in the context of program evaluation. Lee (2002) applied (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 treatment effect or outcome gain and $F_\Delta(\cdot)$ its distribution function. Given the marginals F_1 and F_0, sharp bounds on the distribution of Δ can be found in Williamson and Downs (1990).

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

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

$$F^L(\delta) = \max\left(\sup_y \{F_1(y) - F_0(y-\delta)\}, 0\right),$$
$$F^U(\delta) = 1 + \min\left(\inf_y \{F_1(y) - F_0(y-\delta)\}, 0\right). \quad \textbf{(2)}$$

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

Lemma 2.1 implies that the treatment effect distribution F_Δ first-order stochastically dominates F^U and is first-order stochastically dominated by F^L. Let $\succsim_{FSD}$ denote the first-order stochastic dominance relation. Then $F^L \succsim_{FSD} F_\Delta \succsim_{FSD} F^U$. We note that unlike sharp bounds on the joint distribution of Y_1, Y_0, sharp bounds on the distribution of 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

$$F_{\Delta'}(\delta) = \mathrm{E}1\{Y_1' - Y_0' \le \delta\} = \int_0^1 1\left\{F_1^{-1}(u) - F_0^{-1}(u) \le \delta\right\} du,$$

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

$$F_{\Delta''}(\delta) = \mathrm{E}1\{Y_1'' - Y_0'' \le \delta\} = \int_0^1 1\left\{F_1^{-1}(u) - F_0^{-1}(1-u) \le \delta\right\} du.$$

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

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

Theorem 1 in Stoye (2009) (see also Tesfatsion, 1976) shows that $F_{\Delta'} \succsim_{SSD} F_\Delta$ is equivalent to $\mathrm{E}\left[U(\Delta')\right] \le \mathrm{E}[U(\Delta)]$ or $\mathrm{E}\left[U(Y_1' - Y_0')\right] \le \mathrm{E}[U(Y_1 - Y_0)]$ for every convex real-valued function U. Corollary 2.3 in Tchen (1980) and Lemma 2.1 imply the conclusion of Lemma 2.2; see also Cambanis, Simons, and Stout (1976).

2.2. Bounds on *D*-Parameters

The sharp bounds on the treatment effect distribution imply 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 (2009) introduced another class of parameters that 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 (2009). He refers to the class of D-parameters as the class of D_1-parameters.

DEFINITION 2.1. *A population statistic θ is a D_1-parameter if it increases weakly with first-order stochastic dominance; that is, $F \succsim_{FSD} G$ implies $\theta(F) \ge \theta(G)$.*

Obviously, if θ is a D_1-parameter, then Lemma 2.1 implies $\theta(F^L) \ge \theta(F_\Delta) \ge \theta(F^U)$. In general, the bounds $\theta(F^L), \theta(F^U)$ on a D_1-parameter may not be sharp, as the bounds in Lemma 2.1 are pointwise 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 $\theta(F^L), \theta(F^U)$ are known to be sharp and can be expressed in terms of the quantile functions of the marginal distributions of the potential outcomes. Specifically, let $G^{-1}(u)$ denote the generalized inverse of a nondecreasing function G; that is, $G^{-1}(u) = \inf\{x | G(x) \ge u\}$. Then Lemma 2.1 implies, for $0 \le q \le 1$, $(F^U)^{-1}(q) \le F_\Delta^{-1}(q) \le (F^L)^{-1}(q)$, and the bounds are known to be sharp.

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

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

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

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 expressions for sharp bounds on the quantile function of the distribution of treatment effects in Lemma 2.3; see Williamson and Downs (1990). 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 2.4. *For* $q \in [0,1]$, *we have (i)* $F_{\Delta'}^{-1}(q) = \left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ *if* $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ *is an increasing function of* q*; (ii)* $F_{\Delta''}^{-1}(q) = \left[F_1^{-1}(q) - F_0^{-1}(1-q)\right]$.

The proof of Lemma 2.4 follows that of Proposition 3.1 in Embrechts et al. (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 $\varphi\left(F_Z^{-1}(q)\right)$, where F_Z is the distribution function of Z. For (i), we note that $F_{\Delta'}^{-1}(q)$ equals the quantile of $\left[F_1^{-1}(U) - F_0^{-1}(U)\right]$, where U is a uniform random variable on [0, 1]. Let $\varphi(U) = F_1^{-1}(U) - F_0^{-1}(U)$. Then $F_{\Delta'}^{-1}(q) = \varphi(q) = F_1^{-1}(q) - F_0^{-1}(q)$, provided that $\varphi(U)$ is an increasing function of U. For (ii), let $\varphi(U) = F_1^{-1}(U) - F_0^{-1}(1-U)$. Then $F_{\Delta''}^{-1}(q)$ equals the quantile of $\varphi(U)$. Since $\varphi(U)$ is always increasing in this case, we get $F_{\Delta''}^{-1}(q) = \varphi(q)$.

Note that the condition in (i) is a necessary condition; without this condition, $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ can fail to be a quantile function. Doksum (1974) and Lehmann (1974) used $\left[F_1^{-1}(F_0(y_0)) - y_0\right]$ to measure treatment effects. Recently, $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ has been used to study treatment effects heterogeneity and is referred to as the QTE; see, e.g., Heckman et al. (1997), Abadie et al. (2002), Chernozhukov and Hansen (2005), Firpo (2007), 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 2.4(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. In general, the quantile of the treatment effect distribution is different from QTE and is not identified, but it can be bounded; see Lemma 2.3.

DEFINITION 2.2. *A population statistic θ is a D_2-parameter if it increases weakly with second-order stochastic dominance, i.e., $F \succsim_{SSD} G$ implies $\theta(F) \geq \theta(G)$.*

If θ is a D_2-parameter, then Lemma 2.2 implies $\theta(F_{\Delta'}) \leq \theta(F_\Delta) \leq \theta(F_{\Delta''})$. Stoye (2009) 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 bounds on D_1-parameters of the treatment effect distribution implied by Lemma 2.1, bounds on D_2-parameters: $\theta(F_{\Delta'}), \theta(F_{\Delta''})$ are sharp and are reached when the potential outcomes are perfectly dependent on each other; see Cambanis et al. (1976). 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 a D_2-parameter 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 and proposed a test for the common effect model by testing the value of the lower bound of the variance of Δ. Stoye (2009) presents many other examples of D_2-parameters, including many well-known inequality and risk measures.

3. NONPARAMETRIC ESTIMATORS AND THEIR ASYMPTOTIC PROPERTIES

Suppose random samples $\{Y_{1i}\}_{i=1}^{n_1} \sim F_1$ and $\{Y_{0i}\}_{i=1}^{n_0} \sim F_0$ are available. Let $\mathcal{Y}_1$ and $\mathcal{Y}_0$ denote, respectively, the supports[4] of F_1 and F_0. Note that the bounds in Lemma 2.1 can be written as

$$F^L(\delta) = \sup_{y \in \mathcal{R}} \{F_1(y) - F_0(y-\delta)\}, \qquad F^U(\delta) = 1 + \inf_{y \in \mathcal{R}} \{F_1(y) - F_0(y-\delta)\}, \tag{3}$$

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 $\mathcal{Y}_1 = \mathcal{Y}_0 = \mathcal{R}$, (3) suggests the following plug-in estimators of $F^L(\delta)$ and $F^U(\delta)$:

$$F_n^L(\delta) = \sup_{y \in \mathcal{R}} \{F_{1n}(y) - F_{0n}(y-\delta)\}, \qquad F_n^U(\delta) = 1 + \inf_{y \in \mathcal{R}} \{F_{1n}(y) - F_{0n}(y-\delta)\}, \tag{4}$$

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

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

When either $\mathcal{Y}_1$ or $\mathcal{Y}_0$ is not the whole real line, we provide alternative expressions for $F^L(\delta)$ and $F^U(\delta)$ below, which turn out to be convenient for both computational purposes and for asymptotic analysis. Suppose $\mathcal{Y}_1 = [a, b]$ and $\mathcal{Y}_0 = [c, d]$ for $a, b, c, d \in \overline{\mathcal{R}} \equiv \mathcal{R} \cup \{-\infty, +\infty\}$, $a < b, c < d$ with $F_1(a) = F_0(c) = 0$ and $F_1(b) = F_0(d) = 1$. It is easy to see that

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

For any $\delta \in [a-d, b-c] \bigcap \mathcal{R}$, let $\mathcal{Y}_\delta = [a, b] \bigcap [c+\delta, d+\delta]$. Then

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

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

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

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

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

For any fixed δ, the consistency of $F_n^L(\delta)$ and $F_n^U(\delta)$ is obvious. By using $F_n^L(\delta)$ and $F_n^U(\delta)$, we can provide 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 will establish the asymptotic distributions of $\sqrt{n_1}\left(F_n^L(\delta) - F^L(\delta)\right)$ and $\sqrt{n_1}\left(F_n^U(\delta) - F^U(\delta)\right)$. Define

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

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

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

Then

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

We make the following assumptions.

Assumption 1. (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 $0 < \lambda < \infty$.

Assumption 2. The distribution functions F_1 and F_0 are twice differentiable with bounded density functions f_1 and f_0 on their supports.

Assumption 3. For a fixed $\delta \in [a-d, b-c] \bigcap \mathcal{R}$, the function $y \longmapsto F_1(y) - F_0(y-\delta)$ has a unique, well-separated interior maximum at $y_{\sup,\delta}$ on $\mathcal{Y}_\delta$.

Assumption 4. For a fixed $\delta \in [a-d, b-c] \bigcap \mathcal{R}$, the function $y \longmapsto F_1(y) - F_0(y-\delta)$ has a unique, well-separated interior minimum at $y_{\inf,\delta}$ on $\mathcal{Y}_\delta$.

We note that the uniqueness condition in Assumptions 3 and 4 can be restrictive and can be relaxed. We first establish the asymptotic distributions of $M_n(\delta)$ and $m_n(\delta)$.

PROPOSITION 3.1. *Suppose Assumptions 1 and 2 hold. For a given* δ, *let*

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

(i) If Assumption 3 also holds, then $\sqrt{n_1}[M_n(\delta) - M(\delta)] \Longrightarrow N(0, \sigma_L^2)$;
(ii) if Assumption 4 also holds, then $\sqrt{n_1}[m_n(\delta) - m(\delta)] \Longrightarrow N(0, \sigma_U^2)$.

Theorem 3.2 follows from Proposition 3.1.

THEOREM 3.2.

(i) Suppose Assumptions 1–3 hold. For any $\delta \in [a-d, b-c] \bigcap R$,

$$\sqrt{n_1}[F_n^L(\delta) - F^L(\delta)] \Longrightarrow \begin{cases} N(0, \sigma_L^2), & \textit{if } M(\delta) > 0; \\ \max\left\{N(0, \sigma_L^2), 0\right\} & \textit{if } M(\delta) = 0; \end{cases}$$

and $\Pr\left(F_n^L(\delta) = 0\right) \to 1$ *if* $M(\delta) < 0$.

(ii) Suppose Assumptions 1, 2, and 4 hold. For any $\delta \in [a-d, b-c] \bigcap R$,

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

$$\text{and} \quad \Pr\left(F_n^U(\delta) = 1\right) \rightarrow 1 \quad \text{if } m(\delta) > 0.$$

Theorem 3.2 shows that the asymptotic distribution of $F_n^L(\delta)$ ($F_n^U(\delta)$) depends on the value of $M(\delta)$ ($m(\delta)$). For example, if δ is such that $M(\delta) > 0$ ($m(\delta) < 0$), then $F_n^L(\delta)$ ($F_n^U(\delta)$) is asymptotically normally distributed, but if δ is such that $M(\delta) = 0$ ($m(\delta) = 0$), then the asymptotic distribution of $F_n^L(\delta)$ ($F_n^U(\delta)$) is truncated normal.

4. SIMULATION

In this section, we investigate the coverage rates of the asymptotic normal, the standard bootstrap, and the fewer-than-n bootstrap confidence intervals for $F^L(\delta)$ and $F^U(\delta)$ for δ values corresponding to $y_{\sup,\delta}$ ($y_{\inf,\delta}$) being an interior solution with $M(\delta) > 0$ and $M(\delta) = 0$ ($m(\delta) < 0$ and $m(\delta) = 0$). To implement the fewer-than-n bootstrap, we need to choose the subsample size. We use the procedure suggested in Bickel and Sakov (2008). Let m denote the subsample size and $\widehat{m}$ the value of m chosen by the procedure in Bickel and Sakov (2008) (see below for a detailed description of this procedure applied to our case). As shown by Bickel and Sakov (2008), $\widehat{m}$ has the desirable property that under general regularity conditions, when the standard bootstrap fails, $\widehat{m} \rightarrow \infty$ in probability and $\widehat{m}/n = o_p(1)$; and when the standard bootstrap works, $\widehat{m}/n = O_p(1)$. As a result, there is no loss in efficiency in using the fewer-than-n bootstrap with this adaptive rule of choosing the subsample size. On the other hand, subsampling requires a strictly smaller subsample size.

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

Step 1. Consider a sequence of m's of the form $m_j = \left[q^j n\right]$ for $j = 0, 1, 2, \ldots$, $0 < q < 1$, where $[\gamma]$ denotes the largest integer $\leq \gamma$.

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

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

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

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

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

The true marginal distributions and the values of δ used in the simulation are summarized in Table 1. In Example 1, $Y_1 \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y_0 \sim N\left(\mu_0, \sigma_0^2\right)$. When[5] $\sigma_1 \neq \sigma_0$, we get

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

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

where $s = \delta - (\mu_1 - \mu_0)$ and $t = \sqrt{s^2 + \left(\sigma_1^2 - \sigma_0^2\right)\ln\left(\frac{\sigma_1^2}{\sigma_0^2}\right)}$. For any δ, one can show that $M(\delta) > 0$ and $m(\delta) < 0$. Hence the standard bootstrap works for all δ's. The values of δ are chosen such that $F^L(\delta_1) \approx F^U(\delta_1) \approx 0.15$, $F^L(\delta_2) \approx F^U(\delta_2) \approx 0.5$, and $F^L(\delta_3) \approx F^U(\delta_3) \approx 0.85$ to see the effect of the relative position of δ on the coverage rates.

For $a \in (0,1)$, let $C(a)$ denote the distribution with distribution function given by:

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

In Example 2, for the lower bound, we choose $Y_1 \sim C\left(\frac{1}{4}\right)$ and $Y_0 \sim C\left(\frac{3}{4}\right)$. When $\delta = 1 - \frac{\sqrt{6}}{2}$, one can show that $M(\delta) = 0$, $y_{\sup,\delta} = 1 - \frac{\sqrt{6}}{4}$ is in the interior, and

TABLE 1. DGPs used in the simulation

	Estimators for	Marginal distributions F_1	F_0	δ δ_1	δ_2	δ_3
Example 1	$F^L(\delta)$	$N(2,2)$	$N(1,1)$	1.3	2.6	4.5
	$F^U(\delta)$	$N(2,2)$	$N(1,1)$	−2.4	−0.6	0.7
Example 2	$F^L(\delta)$	$C\left(\frac{1}{4}\right)$	$C\left(\frac{3}{4}\right)$	$\frac{1}{8}$	$1-\frac{\sqrt{6}}{2}$	.
	$F^U(\delta)$	$C\left(\frac{3}{4}\right)$	$C\left(\frac{1}{4}\right)$	$-\frac{1}{8}$	$\frac{\sqrt{6}}{2}-1$	.

$f_1'\left(y_{\sup,\delta}\right) - f_0'\left(y_{\sup,\delta} - \delta\right) = -16/3 < 0$. Theorem 3.2 implies that at $\delta = 1 - \sqrt{6}/2$, the asymptotic distribution of $F_n^L(\delta)$ is truncated normal. When $\delta = 1/8$, $y_{\sup,\delta} = 9/16$, $M(\delta) = 47/96 > 0$, $f_1'\left(y_{\sup,\delta}\right) - f_0'\left(y_{\sup,\delta} - \delta\right) = -16/3 < 0$. Theorem 3.2 implies that when $\delta = \frac{1}{8}$, the asymptotic distribution of $F_n^L(\delta)$ is normal. For the upper bound $F^U(\delta)$, we choose $Y_1 \sim C(3/4)$ and $Y_0 \sim C(1/4)$. Similarly to the lower bound case, we show that when $\delta = \sqrt{6}/2 - 1$, the asymptotic distribution of $F_n^U(\delta)$ is truncated normal, and when $\delta = -1/8$, the asymptotic distribution of $F_n^U(\delta)$ is normal.

For each data generating process (DGP) described in Table 1, we generate random samples of the same size n from F_1 and F_0 respectively. The sample sizes are $n = 1{,}000, 2{,}000, 4{,}000$, and the number of simulations is 1,000. To select the number of bootstrap repetitions B, we follow Davidson and Mackinnon (2004; pp. 163–165) by choosing B such that $\alpha(B+1)$ is an integer. Specifically, we use $B = 999$ for $\alpha = 0.05$. For Example 1, we construct confidence intervals for $F^L(\delta)$ and $F^U(\delta)$ for each δ by three methods. The first is the confidence interval based on the standard normal distribution. We denote the corresponding results by "Asymptotics" in Table 2 below. The second method uses the standard bootstrap confidence intervals and the results are denoted by "n-bootstrap" in Table 2. Finally, we use the fewer-than-n bootstrap confidence intervals. In the fewer-than-n bootstrap, we use $q = 0.95$. Here only one value for q is used, because the fewer-than-n bootstrap is only used for comparison purposes (the standard bootstrap works for this case). For Example 2, we use the standard normal distribution ("Asymptotics" in Table 3), the standard bootstrap ("n-bootstrap" in Table 3), and the fewer-than n bootstrap with two values for q: 0.75 and 0.95.

First, we discuss the coverage rates for normal distributions in Table 2. Clearly the coverage rates depend critically on the location of δ. For δ_2, all three methods lead to confidence intervals with very accurate coverage rates for both F^L

TABLE 2. Coverage rates: $(N(2,2), N(1,1))$

		$F^L(\delta)$			$F^U(\delta)$		
n	Method	δ_1	δ_2	δ_3	δ_1	δ_2	δ_3
1,000	Asymptotics	.929	.944	.937	.931	.949	.926
	n-bootstrap	.942	.954	.950	.950	.953	.939
$q = 0.95$	Fewer-than-n bootstrap	.948	.949	.948	.952	.951	.942
2,000	Asymptotics	.942	.944	.934	.943	.946	.927
	n-bootstrap	.949	.944	.946	.946	.952	.937
$q = 0.95$	Fewer-than-n bootstrap	.941	.944	.952	.949	.950	.939
4,000	Asymptotics	.935	.953	.936	.949	.949	.928
	n-bootstrap	.945	.957	.953	.951	.952	.936
$q = 0.95$	Fewer-than-n bootstrap	.944	.957	.952	.951	.952	.939

TABLE 3. Coverage rates: $(C(1/4), C(3/4))$ for F^L; $(C(3/4), C(1/4))$ for F^U

n	Method	$F^L(\delta)$		$F^U(\delta)$	
		δ_1	δ_2	δ_1	δ_2
1,000	Asymptotics	.933	.935	.947	.935
	n-bootstrap	.941	.961	.951	.958
	Fewer-than-n bootstrap ($q = 0.75$)	.943	.963	.951	.960
	Fewer-than-n bootstrap ($q = 0.95$)	.945	.963	.947	.962
2,000	Asymptotics	.952	.955	.940	.940
	n-bootstrap	.951	.970	.947	.959
	Fewer-than-n bootstrap ($q = 0.75$)	.944	.971	.946	.959
	Fewer-than-n bootstrap ($q = 0.95$)	.951	.969	.946	.959
4,000	Asymptotics	.948	.944	.952	.946
	n-bootstrap	.947	.963	.946	.963
	Fewer-than-n bootstrap ($q = 0.75$)	.949	.964	.947	.965
	Fewer-than-n bootstrap ($q = 0.95$)	.949	.962	.951	.961

and F^U. The coverage rates at δ_1 and δ_3 depend on the methods being used. Although in theory all three methods are asymptotically valid, in finite samples, confidence intervals based on normal critical values often substantially under-cover the true values at δ_1 and/or δ_3. For example, the coverage rates of confidence intervals based on normal critical values for $F^L(\delta)$ at $\delta = \delta_1$ and δ_3 are, respectively, .929 and .937 when $n = 1{,}000$ and .935 and .936 when $n = 4{,}000$. On the other hand, the standard bootstrap leads, respectively, to coverage rates of .942 and .950 when $n = 1{,}000$ and .945 and .953 when $n = 4{,}000$, supporting the asymptotic refinement of the standard bootstrap over asymptotic normality in this case. The fewer-than-n bootstrap delivers similar coverage rates to the standard bootstrap.

For Example 2, all three methods—the Asymptotics based on normal critical values, the n-bootstrap, and the fewer-than-n bootstrap with different values of q—perform similarly at δ_1, except that when $n = 1{,}000$, the Asymptotics under-covers for $F^L(\delta_1)$ with coverage rate .933. At δ_2, the n-bootstrap leads to coverage rates higher than .95 for almost all sample sizes, while the fewer-than-n bootstrap produces coverage rates that are slightly better than the n-bootstrap, but not by much. On the other hand, the Asymptotics provides coverage rates that are closer to .95 except when $n = 1{,}000$.

5. SHARP BOUNDS ON THE DISTRIBUTION OF TREATMENT EFFECTS WITH COVARIATES

In many applications, observations on a vector of covariates for individuals in the treatment and control groups are available. In this section, we extend our study on sharp bounds to take into account these covariates. For notational compactness, we let $n = n_1 + n_0$, so that there are n individuals altogether. For $i = 1, \ldots, n$, 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 program; see e.g., Rosenbaum and Rubin (1983a,1983b), Hahn (1998), Heckman et al. (1998), Dehejia and Wahba (1999), and Hirano, Imbens, and Ridder (2003), to name only a few.

Assumption 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 $X = x$.

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

In the following, we present sharp bounds on the distribution of Δ under Assumptions C1 and C2. For any fixed $x \in \mathcal{X}$, Lemma 2.1 provides sharp bounds on the conditional distribution of Δ given $X = x$: $F^L(\delta|x) \leq F_\Delta(\delta|x) \leq F^U(\delta|x)$, where

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

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

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. Assumptions C1 and C2 allow the identification of the conditional distributions $F_1(y|x)$ and $F_0(y|x)$ appearing in the sharp bounds on $F_\Delta(\delta|x)$. To see this, note that

$$\begin{aligned} F_1(y|x) &= P(Y_1 \leq y|X = x) = P(Y_1 \leq y|X = x, D = 1) \\ &= P(Y \leq y|X = x, D = 1), \end{aligned} \tag{5}$$

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

$$F_0(y|x) = P(Y \leq y|X = x, D = 0). \tag{6}$$

Given the random sample $\{Y_i, X_i, D_i\}_{i=1}^n$, nonparametric estimators of the bounds $F^L(\delta|x)$, $F^U(\delta|x)$ can be constructed easily from nonparametric estimators of $F_1(y_1|x)$ and $F_0(y_0|x)$. Sharp bounds on the unconditional distribution of Δ follow from those of the conditional distribution:

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

We note that if X is independent of (Y_1, Y_0), then the above bounds on $F_\Delta(\delta)$ reduce to those in Lemma 2.1. In general, X is not independent of (Y_1, Y_0), and the above bounds are tighter than those in Lemma 2.1.

6. CONCLUSION AND EXTENSIONS

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

Much work remains to be done. In terms of the sharp bounds, those in this paper do not make use of any prior information on the possible dependence between the potential outcomes. When such information is available, these bounds can be tightened. In a companion paper, we explore sharp bounds taking account of dependence information such as values of dependence measures of the potential outcomes. The focus on randomized experiments in this paper allows the identification of the marginal distributions. In cases where the marginal distributions themselves are not identifiable but bounds on them can be placed (see, e.g., Manski, 1994, 2003; Manski and Pepper, 2000; Shaikh and Vytlacil, 2005; Blundell, Gosling, Ichimura, and Meghir, 2007; Honore and Lleras-Muney, 2006), we can also place bounds on the treatment effect distribution.

In terms of statistical inference, this paper looked at inference on the sharp bounds themselves. The lower and upper bounds represent, respectively, the minimum and maximum probabilities that the treatment effects do not exceed a given value. Inference on them should be useful in its own right. Alternatively, as initiated in Horowitz and Manski (2000) and Imbens and Manski (2004), followed by Chernozhukov, Hong, and Tamer (2007) and Romano and Shaikh (2008), among others, one may construct confidence sets for the identified set or the true distribution instead of its bounds. The authors are currently investigating this issue by using the general approach developed in Andrews and Guggenberger (2009, 2010) for nonregular problems.

NOTES

1. Independent of this paper, Firpo and Ridder (2008) also studied sharp bounds on the distribution of treatment effects under the assumption of selection on observables and bounds on functionals of the distribution of treatment effects.

2. Horowitz and Manski (1995) first used the concept of "respect stochastic dominance." Manski (1997a) referred to parameters that respect first-order stochastic dominance as D-parameters.

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

4. In practice, the supports of F_1 and F_0 may be unknown, but they can be estimated by using the corresponding univariate order statistics in the usual way. This will not affect the results to follow. For notational compactness, we assume that they are known.

5. 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

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

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

REFERENCES

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

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

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. & P. Guggenberger (2009) Hybrid and size-corrected subsampling methods. *Econometrica* 77, forthcoming.

Andrews, D.W.K. & P. Guggenberger (2010) Asymptotic size and a problem with subsampling and with the *m* out of *n* Bootstrap. *Econometric Theory* 26, forthcoming.

Beran, R. (1997) Diagnosing bootstrap success. *Annals of the Institute of Statistical Mathematics* 49, 1–24.

Bickel, P.J., F. Götze, & W.R. Zwet (1997) Resampling fewer than *n* observations: Gains, losses, and remedies for losses. *Statistica Sinica* 7, 1–31.

Bickel, P. J., & A. Sakov (2008) On the choice of *m* in the *m* out of *n* bootstrap and its application to confidence bounds for extreme percentiles. *Statistica Sinica* 18, 967–985.

Biddle, J., L. Boden, & R. Reville (2003) A Method for Estimating the Full Distribution of a Treatment Effect, With Application to the Impact of Workfare Injury on Subsequent Earnings. Manuscript, Michiga State University.

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

Black, D.A., J.A. Smith, M.C. Berger, & B.J. Noel (2003) Is the threat of reemployment services more effective than the services themselves? Experimental evidence from the UI system. *American Economic Review* 93(3), 1313–1327.

Blundell, R., A. Gosling, H. Ichimura, & C. Meghir (2007) Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica* 75, 323–363.

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

Carneiro, P., K.T. Hansen, & J. Heckman (2003) Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* 44(2), 361–422.

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

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

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

Dehejia, R. (1997) A Decision-Theoretic Approach to Program Evaluation. Ph.D. dissertation, Harvard University.

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

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

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

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

Firpo, S. & G. Ridder (2008) Bounds on Functionals of the Distribution of Treatment Effects. Working paper, Sâo Paulo School of Economics, Fundacâo. Getulio Vargas (FGV).

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

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

Heckman, J., H. Ichimura, J. Smith, & P. Todd (1998) Characterizing Selection Bias Using Experimental Data. *Econometrica* 66, 1017–1098.

Heckman, J. & R. Robb (1985) Alternative methods for evaluating the impact of interventions. In J. Heckman & B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*. Cambridge University Press.

Heckman, J. & J. Smith (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. Danish Ministry of Labor.

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

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

Honore, B.E. & A. Lleras-Muney (2006) Bounds in competing risks models and the war on cancer. *Econometrica* 74, 1675–1698.

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

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

Imbens, G.W. & W. Newey (2005) Identification and Estimation of Triangular Simultaneous Equations Models without Additivity. Working paper, Department of Economics, University of California, Berkeley.

Imbens, G. W. & D. B. Rubin (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*. Chapman & Hall/CRC, London.

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, Ohio State University.

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*. 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. (1994) The selection problem. In C. Sims (ed.), *Advances in Econometrics, Sixth World Congress*, vol. 1. Cambridge University Press.

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*. Springer-Verlag.

Manski, C.F. & J. Pepper (2000) Monotone instrumental variables: With application to the returns to schooling. *Econometrica* 68, 997–1010.

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

Nelsen, R.B. (1999) *An Introduction to Copulas*. Springer.

Politis, D.N., J.P. Romano, & M. Wolf (1999) *Subsampling*. Springer-Verlag.

Romano, J. & A.M. Shaikh (2008) Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference* 138, 2786–2807.

Rosenbaum, P.R. & D.B. Rubin (1983a) Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, Series B* 45, 212–218.

Rosenbaum, P.R. & D.B. Rubin (1983b) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55.

Rüschendorf, L. (1982) Random variables with maximum sums. *Advances in Applied Probability* 14, 623–632.

Schweizer, B. & A. Sklar (1983) *Probabilistic Metric Spaces*. North-Holland.

Shaikh, A.M. & E. Vytlacil (2005) Threshold Crossing Models and Bounds on Treatment Effects: A Nonparametric Analysis. Working paper t0307, National Bureau of Economic Research.

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

Stoye, J. (2009) Partial Identification of Spread Parameters. Working paper, Department of Economics, New York University.

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.

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

van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer.

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

APPENDIX: Technical Proofs

Proof of Proposition 3.1. Since the proofs of (i) and (ii) are similar, we provide a proof for (i) only. Let $Q_n(y,\delta) = F_{1n}(y) - F_{0n}(y-\delta)$, $Q(y,\delta) = F_1(y) - F_0(y-\delta)$.

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

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

Proof of Step 1. By the classical Glivenko-Cantelli theorem, the sequences $\sup_y |F_{1n}(y) - F_1(y)|$ and $\sup_y |F_{0n}(y-\delta) - F_0(y-\delta)|$ converge in probability to zero. Consequently, the sequence $\sup_y |[F_{1n}(y) - F_{0n}(y-\delta)] - \left[F_1(y) - F_0(y-\delta)\right]|$ also converges in probability to zero. This and Assumption 3(i) imply that the sequence $\widehat{y}_{\sup,\delta}$ converges in probability to $y_{\sup,\delta}$; see, e.g., Theorem 5.7 in van der Vaart (1998).

Proof of Step 2. We use Theorem 3.2.5 in van der Vaart and Wellner (1996) to establish the rate of convergence for $\widehat{y}_{\sup,\delta}$. Given Assumption 2, the map $y \mapsto Q(y,\delta)$ is twice differentiable and has a maximum at $y_{\sup,\delta}$. By Assumption 3, the first condition of Theorem 3.2.5 in van der Vaart and Wellner (1996) is satisfied with $\alpha = 2$. To check the second condition of Theorem 3.2.5 in van der Vaart and Wellner (1996), we consider the centered process:

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

For any $\eta > 0$,

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

Note that the envelope function of the class of functions

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

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

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

Similarly, we can show that

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

Consequently,

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

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

Proof of Step 3. For a fixed δ, we get

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

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