Partial Identification of the Distribution of Treatment Effects in Switching Regimes Models and its Confidence Sets*

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

First version: July 2007
This version: November 2008

Abstract

In this paper, we establish sharp bounds on the joint distribution of potential outcomes and the distribution of treatment effects in parametric switching regimes models with **generalized hyperbolic errors** and in the semiparametric switching regimes models of Heckman (1990). Our results for parametric switching regimes models with generalized hyperbolic errors extend some existing results for Gaussian switching regimes models and our results for semiparametric switching regimes models supplement the point identification results of Heckman (1990). Compared with the corresponding sharp bounds when selection is random, we observe that self selection tightens the bounds on the joint distribution of the potential outcomes and the distribution of treatment effects. These bounds depend on the identified model parameters only and can be easily estimated once the identified model parameters are estimated. We demonstrate the feasibility of inference on the distribution of treatment effects by constructing an asymptotically uniformly valid and non-conservative confidence set in a semiparametric switching regimes model.

Keywords: Average treatment effect; treatment effect for the treated; copula; Fréchet-Hoeffding inequality; correlation bounds

JEL codes: C31, C35, C14

*This is a substantially revised version of an early paper circulated under the title: "Sharp Bounds on the Distribution of the Treatment Effect in Switching Regimes Models." We are grateful to Enrique Sentana and two anonymous referees for helpful comments on the early paper that significantly improved the current paper. We thank Bo Honore and seminar participants at U. Penn, NYU, Penn State University, University of Virginia, Oregon State University, University of Pittsburgh, and Far Eastern Econometric Society Meetings for helpful comments and discussions. Fan gratefully acknowledges financial support from the National Science Foundation.

1 Introduction

The class of switching regimes models (SRMs) or generalized sample selection models extends the Roy model of self-selection by allowing a more general decision rule for selecting into different states. The income maximizing Roy model of self-selection was developed to explain occupational choice and its consequences for the distribution of earnings when individuals differ in their endowments of occupation-specific skills. Heckman and Honore (1990) demonstrated that the identification of the joint distribution of potential outcomes is essential to the empirical content of the Roy model.

By allowing a more general decision/selection rule, SRMs enjoy a much wider scope of applications than the Roy model, but in any particular application, they are also limited in their ability to address a wide range of interesting economic/policy questions because of the non-identifiability[1] of the joint distribution of potential outcomes in SRMs. Even in the 'textbook' Gaussian SRM, the correlation coefficient between the potential outcomes or equivalently the joint distribution of the potential outcomes is not identifiable. In a study of a sectoral labor market using the Gaussian SRM, Vijverberg (1993) showed that a number of interesting economic questions including the share of 'productive' workers employed in a sector can not be answered without knowledge of the joint distribution of the two potential outcomes. When used to study treatment effect defined as the difference between the two potential outcomes, important distributional aspects of the treatment effect other than its mean are not identified in SRMs. This partly explains why the current literature has mainly focussed on various measures of average treatment effect including the average treatment effect (ATE), the treatment effect for the treated (TT), the local average treatment effect (LATE), and the marginal treatment effect (MTE). Heckman, Tobias, and Vytlacil (2003) derived expressions for these four average treatment effect parameters for a Gaussian copula SRM and a Student's t copula SRM with normal outcome errors and non-normal selection errors[2]. Heckman and Vytlacil (2005), among other things, showed that in a latent variable framework, ATE, TT, and LATE can be expressed in terms of MTE.

Recently two approaches have been proposed to deal with the non-identifiability problem of the joint distribution of potential outcomes in the 'textbook' Gaussian SRM and some of its extensions. By employing the positive semidefiniteness of the covariance matrix of the outcome errors and the selection error, Vijverberg (1993) showed that in the 'textbook' Gaussian SRM, although unidentified, useful bounds can be placed on the correlation coefficient between the potential outcomes, that is, it is partially identified. Koop and Poirier (1997), Poirier (1998), and Poirier and Tobias (2003) demonstrated via Bayesian approach that these bounds often provide informative information on the unidentified correlation coefficient. Since the joint distribution of the potential

[1] In this paper, we use identification and point identification interchangeably. We say a parameter is not identified if it is not point identified.

[2] They didn't use the concept of copulas, but their models can be interpreted this way.

outcomes in the 'textbook' Gaussian SRM depends on the unidentified correlation coefficient only (besides the identified marginal parameters), it is possible to place bounds on the joint distribution of the potential outcomes and on the distribution of the difference between the potential outcomes. In the second approach, restrictions are imposed on the dependence structure between the potential outcomes such that their joint distribution and the distribution of treatment effects are identified, see, e.g., Heckman, Smith, and Clements (1997), Biddle, Boden, and Reville (2003), Carneiro, Hansen, and Heckman (2003), Aakvik, Heckman, and Vytlacil (2005), Cunha and Heckman (2007), among others. Among other things, they demonstrated that knowledge of the joint distribution of potential outcomes and the distribution of the treatment effect allows a much richer analysis of policy effects than average treatment effects. Questions that can be addressed include the proportion of people participating in the program who benefit from it; the proportion of the total population that benefits from the program; and which groups in an initial position benefit or lose from the program.

This paper takes the first approach and makes several contributions to the current literature. First, we extend the partial identification results in Vijverberg (1993), Koop and Poirier (1997), Poirier (1998), Poirier and Tobias (2003), and Li, Poirier, and Tobias (2004) to a general class of SRMs in which the joint distribution of the outcome errors and the selection error is assumed to follow a trivariate generalized hyperbolic (GH) distribution referred to as GH-SRMS. The 'textbook' Gaussian and Student's t SRMs are members of GH-SRMs. In addition, GH-SRMs also allow the errors to follow asymmetric distributions. For GH-SRMs, we provide sharp bounds or partial identification results on the correlation coefficient of the potential outcomes, their joint distribution, and the distribution of treatment effects.

Second, we establish sharp bounds on the joint distribution of potential outcomes and the distribution of treatment effects in the general class of semiparametric SRMs in Heckman (1990) in which the joint distribution of the trio of errors is completely unspecified. Our results rely on and supplement the point identification results in Heckman (1990). The distribution bounds established in GH-SRMs rely on two special properties: (i) the only unidentified parameter in a GH-SRM is the correlation coefficient between the two potential outcomes and (ii) the joint distribution of the two potential outcomes in a GH-SRM is also GH. The fact that the joint distribution of the potential outcomes in a SRM is not identifiable and thus can never be verified empirically calls for a study of the robustness of the results for GH-SRMs to the implied joint distribution of the potential outcomes and the development of a general approach to bounding the joint distribution of the potential outcomes and the distribution of treatment effects that is robust to distributional assumptions on the outcome errors and the selection error.

Without specifying the joint distribution of the outcome errors and the selection error, the approach used to bound the distribution of the potential outcomes and the distribution of treatment effects in GH-SRMs breaks down. The new tool that we employ in this paper to establish bounds on the joint distribution of potential outcomes is the Fréchet-Hoeffding inequality on copulas. A straightforward application of this inequality allows us to bound the joint distribution of potential outcomes using the bivariate distributions of each outcome error and the selection error, where the latter distributions are known to be identified under general conditions, see Heckman (1990). To bound the distribution of treatment effects, we make use of existing results on sharp bounds on the distributions of functions of two random variables including the four simple arithmetic operations, see Williamson and Downs (1990). For a sum of two random variables, Makarov (1981), Rüschendorf (1982), and Frank, Nelsen, and Schweizer (1987) establish sharp bounds on its distribution, see also Nelsen (1999). These results have been used in Fan and Park (2006, 2008) to bound the distribution of treatment effects and the quantile function of treatment effects in the context of ideal social experiments where selection is random. Other applications of the Fréchet-Hoeffding inequality include Heckman, Smith, and Clements (1997) in which they bound the variance of treatment effects under the assumption of random selection; Manski (1997b) in which he established bounds on the mixture of two potential outcomes when the distribution of each outcome is known; and Fan (2005) in which she provided a systematic study on the estimation and inference on the correlation bounds.

Two interesting conclusions emerge from our results. First, in GH-SRMs with symmetric outcome errors, we find that the sharp bounds on the joint distribution of potential outcomes are robust to the implied joint distribution of the potential outcomes in the sense that these bounds remain valid for any distribution of the trio of errors as long as the implied bivariate distributions for each outcome error and the selection error are GH with symmetric outcome errors. In contrast, the sharp bounds on the treatment effect distribution in GH-SRMs are not robust to the implied joint distribution of the potential outcomes. We provide a detailed numerical comparison between sharp bounds on the treatment effect distribution relying on the trivariate Gaussian and Student's t distributions with those that do not specify the non-refutable joint distribution of potential outcomes. Our numerical results show that bounds relying on the trivariate Gaussian or Setudent's t assumption can be misleading. Second, we find that in general information on individual's selection decision can help improve sharp bounds on the joint distribution of the potential outcomes and the distribution of treatment effects. When the conditional distribution of one of the potential outcome errors given the selection error is degenerate at a finite value, our sharp bounds point identify the joint distribution of potential outcomes and the distribution of treatment effects.

The partial identification results established in this paper can be used to develop inference procedures for the joint distribution of potential outcomes and the distribution of treatment effects. There is a recent, but rapidly growing literature on inference for partially identified parameters, including Imbens and Manski (2004), Bugni (2007), Canay (2007), Chernozhukov, Hong, and Tamer

(2007), Fan and Park (2007), Romano and Shaikh (2006), Stoye (2007), Andrews and Guggenberger (2007), and Andrews and Soares (2007), among others. We refer the reader to Fan and Park (2007) for more references. A complete treatment of this important issue is beyond the scope of this paper. However, we demonstrate this feasibility by constructing an asymptotically uniformly valid and non-conservative confidence set (CS) for the distribution of treatment effects in a semiparametric SRM.

The rest of this paper is organized as follows. In Section 2, we introduce the class of GH-SRMs and discuss the identification/partial identification of the parameters in GH-SRMs. In particular, we extend existing work on correlation bounds in SRMs with trivariate normal or Student's t errors to our GH-SRMs. In Section 3, we establish sharp bounds on the joint distribution of potential outcomes and on the distribution of treatment effects for the whole population with a given observed covariate and the subpopulation with a given observed covariate receiving treatment in GH-SRMs. In Section 4, we establish sharp bounds on the joint distribution of potential outcomes and on the distribution of treatment effects for the whole population with a given observed covariate and the subpopulation with a given observed covariate receiving treatment in semiparametric SRMs in Heckman (1990). In Section 5 we provide a systematic comparison of the two sets of bounds when the two identified bivariate marginal distributions are GH. Section 6 presents an asymptotically uniformly valid and non-conservative CS for the distribution of treatment effects in a special class of semiparametric SRMs. The last section concludes. Technical proofs are relegated in the Appendix.

2 Generalized Hyperbolic Switching Regimes Models and Parameter Identification

Consider the following SRM:

$$
\begin{aligned}
Y_{1i} &= X_i'\beta_1 + U_{1i}, \\
Y_{0i} &= X_i'\beta_0 + U_{0i}, \\
D_i &= I_{\{W_i'\gamma+\epsilon_i>0\}},\ i = 1, \dots, n, \qquad (1)
\end{aligned}
$$

where X_i, W_i denote individual i's observed covariates and U_{1i}, U_{0i}, ϵ_i denote individual i's unobserved covariates. Here, D_i is the binary variable indicating participation of individual i in the program or treatment; it takes the value 1 if individual i participates in the program and takes the value zero if she chooses not to participate in the program, Y_{1i} is the outcome of individual i we observe if she participates in the program, and Y_{0i} is her outcome if she chooses not to participate in the program. For individual i, we always observe the covariates X_i, W_i, but observe Y_{1i} if $D_i = 1$ and Y_{0i} if $D_i = 0$. The vector of errors or unobserved covariates $(U_{1i}, U_{0i}, \epsilon_i)'$ is assumed to be independent of the vector of observed covariates $(X_i', W_i')'$. We also assume the existence of an exclusion restriction, i.e., there exists at least one element of W_i which is not contained in X_i.

Parametric SRMs supplement model (1) by distributional specifications for the vector of errors $(U_{1i}, U_{0i}, \epsilon_i)'$. Commonly used distributions include the trivariate normal and Student's t distributions. In this section, we introduce a general flexible class of parametric SRMs characterized by (1) with the trivariate error vector $(U_{1i}, U_{0i}, \epsilon_i)'$ following a generalized hyperbolic (GH) distribution. We refer to this class of parametric SRMs as GH-SRMs which nest both SRMs with the trivariate Gaussian and Student's t distributions as special cases. We discuss identification or partial identification of the parameters in GH-SRMs. For notational compactness, we omit the subscript i in the rest of Section 2 and Section 3.

2.1 GH-SRMs

We first present a brief review of d-simensional GH distributions. This part follows McNeil, Frey, and Embrechts (2005) and Mencia and Sentana (2005). Let V denote a d-dimensional random vector following a GH distribution with parameters $\lambda, \chi, \psi, \mu, \Sigma, \zeta$, where μ and ζ are $d \times 1$ vectors, Σ is a $d \times d$ matrix, and parameters λ, χ, ψ satisfy: $\chi > 0$, $\psi \geq 0$ if $\lambda < 0$; $\chi > 0, \psi > 0$, if $\lambda = 0$; and $\chi \geq 0, \psi > 0$, if $\lambda > 0$. To simplify exposition, we use $GH_d(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$ to denote this distribution and $GH_d(\cdot; \lambda, \chi, \psi, \mu, \Sigma, \zeta)$, $gh_d(\cdot; \lambda, \chi, \psi, \mu, \Sigma, \zeta)$ to denote respectively its distribution function and density function. If Σ is positive definite, then the probability density function of V has the following closed-form expression:

$$gh_d(v; \lambda, \chi, \psi, \mu, \Sigma, \zeta) = c \frac{K_{\lambda-(d/2)}\left(\sqrt{\left(\chi + (v-\mu)' \Sigma^{-1}(x-\mu)\right)\left(\psi + \zeta'\Sigma^{-1}\zeta\right)}\right) e^{(v-\mu)'\Sigma^{-1}\zeta}}{\left[\sqrt{\left(\chi + (v-\mu)' \Sigma^{-1}(v-\mu)\right)\left(\psi + \zeta'\Sigma^{-1}\zeta\right)}\right]^{(d/2)-\lambda}},$$

where $K_v(\bullet)$ is the modified Bessel function of the third kind and

$$c = \frac{\left(\sqrt{\chi\psi}\right)^{-\lambda} \psi^{\lambda}\left(\psi + \zeta'\Sigma^{-1}\zeta\right)^{(d/2)-\lambda}}{(2\pi)^{d/2} |\Sigma^{-1}|^{1/2} K_{\lambda}\left(\sqrt{\chi\psi}\right)}.$$

To understand the role of each parameter in $GH_d(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$, we recall the following definition of a GH distribution.

Definition 2.1 *We say $V \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$ with $\Sigma = AA'$ if V has the same distribution as $\mu + S\zeta + \sqrt{S}AZ$, where μ, ζ, A are respectively $d \times 1$, $d \times 1$, and $d \times k$ constant matrices, $Z \sim N_k(0, I_k)$, and S is independent of Z and follows the generalized inverse Gaussian (GIG) distribution, denoted as $N^{-1}(\lambda, \chi, \psi)$, with density function:*

$$f_S(s) = \frac{(\psi/\chi)^{\lambda/2}}{2K_{\lambda}\left(\sqrt{\chi\psi}\right)} s^{\lambda-1} \exp\left(-\frac{\chi s^{-1} + \psi s}{2}\right). \tag{2}$$

Based on Definition 2.1, we can put the parameters in $GH_d(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$ into two groups: those in the distribution of the mixing variable S denoted as (λ, χ, ψ) and the remaining parameters

(μ, Σ, ζ). When $\zeta = 0$, the distribution of V belongs to the class of symmetric GH distributions. A non-zero value of ζ introduces asymmetry into the distribution of V. It follows from the above definition that $V|S = s \sim N_d(\mu + s\zeta, s\Sigma)$. Thus, μ and Σ play the roles of the location vector and the dispersion matrix respectively. Among the parameters in the first group, ψ is a scale parameter and (λ, χ) allow for flexible tails in the distribution of V, see Jogensen (1982) and Mencia and Sentana (2005) for more discussion.

We are now ready to introduce the class of GH-SRMs. Let $V = (U_1, U_0, \epsilon)' \sim GH_3(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$. Given that the distributions $GH_3(\lambda, \chi/a, a\psi, \mu, a\Sigma, a\zeta)$ and $GH_3(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$ are identical for any $a > 0$, the parameters $\chi, \psi, \Sigma, \zeta$ are not separately identifiable without normalization. Following the convention in the literature on SRMs, we normalize the variance of the selection error ϵ to be 1. The parameters $\lambda, \chi, \psi, \mu, \Sigma, \zeta$ are restricted such that $E(V) = 0$ and $Var(\epsilon) = 1$. To ensure $E(V) = 0$, we let

$$\mu = -\zeta E(S), \tag{3}$$

where

$$E(S) = \left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}. \tag{4}$$

Let $\mu = (\mu_1, \mu_0, \mu_\epsilon)'$, $\zeta = (\zeta_1, \zeta_0, \zeta_\epsilon)'$, and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{10}\sigma_1\sigma_0 & \rho_{1\epsilon}\sigma_1\sigma_\epsilon \\ \rho_{10}\sigma_1\sigma_0 & \sigma_0^2 & \rho_{0\epsilon}\sigma_0\sigma_\epsilon \\ \rho_{1\epsilon}\sigma_1\sigma_\epsilon & \rho_{0\epsilon}\sigma_0\sigma_\epsilon & \sigma_\epsilon^2 \end{pmatrix}.$$

Note that in general Σ is not the variance-covariance matrix of the error vector V, as

$$Var(V) = E(S)\Sigma + Var(S)\zeta\zeta'.$$

From this, it follows that $Var(\epsilon) = E(S)\sigma_\epsilon^2 + \zeta_\epsilon^2 Var(S)$, where

$$Var(S) = \left(\frac{\chi}{\psi}\right)\left(\frac{K_{\lambda+2}\left(\sqrt{\chi\psi}\right)K_\lambda\left(\sqrt{\chi\psi}\right) - K_{\lambda+1}^2\left(\sqrt{\chi\psi}\right)}{K_\lambda^2\left(\sqrt{\chi\psi}\right)}\right). \tag{5}$$

Restricting $Var(\epsilon) = 1$ leads to

$$\sigma_\epsilon^2 = \left[1 - \zeta_\epsilon^2 Var(S)\right]/E(S). \tag{6}$$

The class of GH-SRMs is characterized by (1) with $(U_1, U_0, \epsilon)' \sim GH_3(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$, where μ satisfies (3) and Σ or σ_ϵ^2 satisfies (6). This class of models is very flexible. It includes the commonly used Gaussian and Student's t SRMs as special cases. Furthermore, it allows the errors $(U_1, U_0, \epsilon)'$ to have skewed and kurtotic distributions.

Example 2.1. (Normal) When $\lambda \to -\infty$ or $\lambda \to +\infty$ or $\chi \to +\infty$, $GH_3(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$ becomes trivariate normal and the corresponding GH-SRM reduces to the 'textbook' Gausssian SRM.

Example 2.2. (Student's t) When $\lambda = -v/2$, $\chi = v$, $\psi = 0$ and $\zeta \to 0$, $GH_3(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$ becomes trivariate Student's t with degree of freedom v and the GH-SRM reduces to Student's t SRM.

Example 2.3. (Skewed t) When $\lambda = -\frac{1}{2}v$, $\chi = v$, and $\psi = 0$, $GH_3(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$ is referred to as the asymmetric or skewed t distribution. It reduces to the standard trivariate Student's t distribution with degree of freedom v as $\zeta \to 0$.

2.2 Parameter Identification/Partial Identification

It follows from Proposition 3.13 in McNeil, Frey, and Embrechts (2005) that

$$\begin{aligned}
(U_1, \epsilon)' &\sim GH_2(\lambda, \chi, \psi, \mu_{1\epsilon}, \Sigma_{1\epsilon}, \zeta_{1\epsilon}), \\
(U_0, \epsilon)' &\sim GH_2(\lambda, \chi, \psi, \mu_{0\epsilon}, \Sigma_{0\epsilon}, \zeta_{0\epsilon}), \\
(U_1, U_0)' &\sim GH_2(\lambda, \chi, \psi, \mu_{10}, \Sigma_{10}, \zeta_{10}),
\end{aligned}$$

where for $i, j = 0, 1, \epsilon$,

$$\mu_{ij} = \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}, \ \zeta_{ij} = \begin{pmatrix} \zeta_i \\ \zeta_j \end{pmatrix}, \ \Sigma_{ij} = \begin{pmatrix} \sigma_i^2 & \sigma_i\sigma_j\rho_{ij} \\ \sigma_i\sigma_j\rho_{ij} & \sigma_j^2 \end{pmatrix}.$$

Since either Y_1 or Y_0 is observed for any given individual but never both, the joint distribution of U_1 and U_0 is in general not point identified in GH-SRMs. As a result, ρ_{10} is in general not point identified, as it only appears in the joint distribution of U_1, U_0, whereas the remaining parameters including $\rho_{1\epsilon}$ and $\rho_{0\epsilon}$ are point identified. However, since $\Sigma = Var\left(s^{-1/2}V|S = s\right)$ is positive semi-definite, we obtain: $\rho_{10} \in [\rho_L, \rho_U]$, where

$$\rho_L = \rho_{1\epsilon}\rho_{0\epsilon} - \sqrt{(1-\rho_{1\epsilon}^2)(1-\rho_{0\epsilon}^2)}, \ \rho_U = \rho_{1\epsilon}\rho_{0\epsilon} + \sqrt{(1-\rho_{1\epsilon}^2)(1-\rho_{0\epsilon}^2)}.$$

This result was first established in Vijverberg (1993) for Gaussian SRMs by using the fact that the variance covariance matrix of the error vector $(U_1, U_0, \epsilon)'$ must be positive semi-definite, see also Koop and Poirier (1997), Poirier (1998), and Poirier and Tobias (2003).

In Gaussian SRMs, Σ is the variance-covariance matrix of $(U_1, U_0, \epsilon)'$ and ρ_{10} is the correlation coefficient between U_1 and U_0. In general GH-SRMs, ρ_{10} is not the correlation coefficient between U_1 and U_0. Let $\overline{\rho}_{10}$ denote the correlation coefficient between U_1 and U_0. It is related to ρ_{10} through the following expression:

$$\overline{\rho}_{10} = \frac{E(S)\rho_{10}\sigma_1\sigma_0 + \zeta_1\zeta_0 Var(S)}{\sqrt{\left(E(S)\sigma_1^2 + \zeta_1^2 Var(S)\right)\left(E(S)\sigma_0^2 + \zeta_0^2 Var(S)\right)}}. \tag{7}$$

The bounds on ρ_{10} and (7) yield bounds on $\overline{\rho}_{10}$: $\overline{\rho}_L \leq \overline{\rho}_{10} \leq \overline{\rho}_U$, where

$$\begin{aligned}\bar{\rho}_L &= \frac{E(S)\,\rho_L\sigma_1\sigma_0 + \zeta_1\zeta_0 Var(S)}{\sqrt{\left(E(S)\,\sigma_1^2 + \zeta_1^2 Var(S)\right)\left(E(S)\,\sigma_0^2 + \zeta_0^2 Var(S)\right)}},\\ \bar{\rho}_U &= \frac{E(S)\,\rho_U\sigma_1\sigma_0 + \zeta_1\zeta_0 Var(S)}{\sqrt{\left(E(S)\,\sigma_1^2 + \zeta_1^2 Var(S)\right)\left(E(S)\,\sigma_0^2 + \zeta_0^2 Var(S)\right)}}.\end{aligned}$$

Since $\rho_{1\epsilon}, \rho_{0\epsilon}$ are point identified, the bounds ρ_L, ρ_U are point identified and thus $\bar{\rho}_L$, $\bar{\rho}_U$ are point identified. If $[\overline{\rho}_L, \overline{\rho}_U] \neq [-1, 1]$, then we say $\overline{\rho}_{10}$ is partially identified with identified interval (set) given by $[\overline{\rho}_L, \overline{\rho}_U]$. We point out that the bounds $\overline{\rho}_L$, $\overline{\rho}_U$ are sharp, as ρ_L and ρ_U are sharp for ρ_{10}. The following theorem shows that the identified interval for $\overline{\rho}_{10}$ may identify its sign and may even point identify $\overline{\rho}_{10}$.

THEOREM 2.1 *Let $\bar{\rho}_{1\epsilon} = Corr(U_1, \epsilon)$ and $\bar{\rho}_{0\epsilon} = Corr(U_0, \epsilon)$. (i) Suppose $Var(S) > 0$. Then $[\overline{\rho}_L, \overline{\rho}_U] = [-1, 1]$ if and only if $\bar{\rho}_{1\epsilon} = \bar{\rho}_{0\epsilon} = 0$ and $\zeta_1 = \zeta_0 = 0$. Suppose $Var(S) = 0$. Then $[\overline{\rho}_L, \overline{\rho}_U] = [-1, 1]$ if and only if $\bar{\rho}_{1\epsilon} = \bar{\rho}_{0\epsilon} = 0$; (ii) If $\overline{\rho}_{1\epsilon}^2 + \overline{\rho}_{0\epsilon}^2 > 1$ and $\overline{\rho}_{1\epsilon}$, $\overline{\rho}_{0\epsilon}$ have the same sign, then $\overline{\rho}_L > 0$; (iii) If $\overline{\rho}_{1\epsilon}^2 + \overline{\rho}_{0\epsilon}^2 > 1$ and $\overline{\rho}_{1\epsilon}$, $\overline{\rho}_{0\epsilon}$ have the opposite sign, then $\overline{\rho}_U < 0$; (iv) If $\overline{\rho}_{1\epsilon}^2 = 1$ or $\overline{\rho}_{0\epsilon}^2 = 1$, then $\bar{\rho}_L = \bar{\rho}_U$ implying that $\overline{\rho}_{10}$ is point identified.*

Theorem 2.1 (i) implies that when $(U_1, U_0)'$ follows a symmetric GH distribution, the bounds $\overline{\rho}_L, \overline{\rho}_U$ are informative or $\overline{\rho}_{10}$ is partially identified as long as at least one of the potential outcomes is correlated with the selection error. In addition, Theorem 2.1 (ii) and (iii) imply that it is possible to identify the sign of $\overline{\rho}_{10}$. The inequality: $\overline{\rho}_L \leq \overline{\rho}_{10} \leq \overline{\rho}_U$ characterizes the class of GH-SRMs consistent with the sample information; any GH-SRM with $\overline{\rho}_{10}$ violating it is inconsistent with the sample information.

3 Distribution Bounds in GH-SRMs

Let $\Delta = Y_1 - Y_0$ denote the individual treatment effect. Heckman, Tobias, and Vytlacil (2003) derived expressions for four treatment parameters of interest for a Gaussian copula model and a Student's t copula model with normal outcome errors and non-normal selection errors. They are respectively ATE, TT, LATE, and MTE. Let ATE denote the average treatment effect conditional on X, W: $ATE \equiv E(\Delta|X, W) = X'(\beta_1 - \beta_0)$. Note that

$$\Delta = ATE + (U_1 - U_0).$$

The individual treatment effect Δ may differ across individuals with the same observable covariates because of the unobserved heterogeneity $(U_1 - U_0)$. This motivates the study of the distribution of treatment effect Δ conditional on the observed covariates. Throughout the rest of this paper, we consider exclusively distributions conditional on the observed covariates without explicitly mentioning it.

The distribution of Δ depends on ρ_{10} or $\overline{\rho}_{10}$ and hence is not point identified in general. In this section, we establish partial identification results for the joint distribution of potential outcomes and the distribution of Δ for a randomly chosen individual from the whole population and from the population participating in the treatment (with the observed covariates x, w).

3.1 Sharp Bounds on the Joint Distribution of Potential Outcomes

Let F_{10}^Y denote the joint distribution of potential outcomes Y_1, Y_0 conditional on $X = x$, $W = w$. Let $\alpha_{10} = (\lambda, \chi, \psi, \mu_{10}, \Sigma_{10}, \zeta_{10})$ and α_{10}^- denote all the parameters in α_{10} except ρ_{10}.

In a GH-SRM, $F_{10}^Y(y_1, y_0) = GH_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \alpha_{10}\right)$,where

$$GH_2\left(u_1, u_0; \alpha_{10}\right) = \int_0^\infty f_S\left(s\right) \Phi_{\rho_{10}}\left(\frac{u_1 - \left(\zeta_1 s - \zeta_1 E\left(S\right)\right)}{\sigma_1\sqrt{s}}, \frac{u_0 - \left(\zeta_0 s - \zeta_0 E\left(S\right)\right)}{\sigma_0\sqrt{s}}\right) ds,$$

in which $E\left(S\right)$ and $f_S\left(\bullet\right)$ are given in (4) and (2) respectively, $\Phi_\rho(\cdot, \cdot)$ is the distribution function of a bivariate normal variable with zero means, unit variances, and correlation coefficient ρ. We now show that the bounds on ρ_{10} place bounds on the joint distribution $F_{10}^Y(y_1, y_0)$. Let C^{Gau} denote the Gaussian copula given by

$$C^{Gau}(u, v, \rho) = \Phi_\rho\left(\Phi^{-1}\left(u\right), \Phi^{-1}\left(v\right)\right), \ (u, v) \in [0, 1]^2.$$

Then we can write

$$\begin{aligned} & GH_2\left(u_1, u_0; \alpha_{10}\right) \\ = & \int_0^\infty f_S\left(s\right) C^{Gau}\left(\Phi\left(\frac{u_1 - \left(\zeta_1 s - \zeta_1 E\left(S\right)\right)}{\sigma_1\sqrt{s}}\right), \Phi\left(\frac{u_0 - \left(\zeta_0 s - \zeta_0 E\left(S\right)\right)}{\sigma_0\sqrt{s}}\right), \rho_{10}\right) ds. \end{aligned}$$

Since the Gaussian copula is increasing in concordance in ρ_{10} (see Joe (1997)), we obtain the following sharp bounds on the joint distribution of Y_1, Y_0:

$$\begin{aligned} & GH_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \left[\alpha_{10}^-, \rho_L\right]\right) \\ \leq & F_{10}^Y(y_1, y_0) \\ \leq & GH_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \left[\alpha_{10}^-, \rho_U\right]\right). \end{aligned} \tag{8}$$

For any fixed x and y_1, y_0, the bounds above are informative or $F_{10}^Y(y_1, y_0)$ as long as

$$GH_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \left[\alpha_{10}^-, \rho_L\right]\right) \neq 0 \text{ or } GH_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \left[\alpha_{10}^-, \rho_U\right]\right) \neq 1.$$

Moreover, if either $\overline{\rho}_{1\epsilon}^2 = 1$ or $\overline{\rho}_{0\epsilon}^2 = 1$, Theorem 2.1 (iv) implies that $\bar{\rho}_L = \bar{\rho}_U$ or equivalently $\rho_L = \rho_U$ and thus (8) point identifies $F_{10}^Y(y_1, y_0)$.

THEOREM 3.1 *Let $C_L(s,t) = \max(s+t-1,0)$ denote the Fréchet lower bound copula and $C_U(s,t) = \min(s,t)$ denote the Fréchet upper bound copula. Suppose $\zeta_1 = \zeta_0 = 0$ and $\bar{\rho}_{1\epsilon} = \bar{\rho}_{0\epsilon} = 0$. Then*

$$\begin{aligned} GH_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \left[\alpha_{10}^{-}, \rho_L\right]\right) &= C_L\left(GH_1\left(y_1 - x'\beta_1; \theta_1\right), GH_1\left(y_0 - x'\beta_0; \theta_0\right)\right), \\ GH_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \left[\alpha_{10}^{-}, \rho_U\right]\right) &= C_U\left(GH_1\left(y_1 - x'\beta_1; \theta_1\right), GH_1\left(y_0 - x'\beta_0; \theta_0\right)\right), \end{aligned}$$

where $\theta_1 = \left(\lambda, \chi, \psi, \mu_1, \sigma_1^2, \zeta_1\right)$ and $\theta_0 = \left(\lambda, \chi, \psi, \mu_0, \sigma_0^2, \zeta_0\right)$.

It is interesting to observe from (8) and Theorem 3.1 that in GH-SRMs, two sources of information contribute to the partial identification of $F_{10}^{Y}(y_1, y_0)$: (i) the partial identification of $\overline{\rho}_{10}$ or ρ_{10}; (ii) the point identification of the marginal distributions of the potential outcomes. When selection is random, i.e., the selection error is independent of the outcome errors, we have

$$\begin{aligned} & C_L\left(GH_1\left(y_1 - x'\beta_1; \theta_1\right), GH_1\left(y_0 - x'\beta_0; \theta_0\right)\right) \\ \leq\ & F_{10}^{Y}(y_1, y_0) \\ \leq\ & C_U\left(GH_1\left(y_1 - x'\beta_1; \theta_1\right), GH_1\left(y_0 - x'\beta_0; \theta_0\right)\right). \end{aligned} \tag{9}$$

This is a straightforward applicationof the Fréchet-Hoeffding inequality:

$$C_L(s,t) \leq C(s,t) \leq C_U(s,t), \text{ for all } (s,t) \in [0,1]^2, \tag{10}$$

where $C(\cdot,\cdot)$ is any copula function, as $Y_j \sim GH_1\left(\cdot - x'\beta_j; \theta_j\right)$, $j = 1, 0$. Theorem 3.1 shows that when the outcome errors follow symmetric GH distributions and are uncorrelated with the selection error, the bounds in (8) are the same as those in (9). In general, the bounds in (8) are sharper than those in (9). Thus, taking into account self-selection tightens the bounds on the joint distribution of potential outcomes.

Similar conclusions hold for the joint distribution of the potential outcomes for participants.To simplify the notation for sharp bounds on $F_{10}^{Y}(y_1, y_0|D=1)$, we let $F_{10|D=1}^{*}\left(u_1, u_0; \rho_{10|\epsilon}\right)$ denote the conditional distribution function of (U_1, U_0) given $D = 1$, where

$$\rho_{10|\epsilon} = \frac{\rho_{10} - \rho_{1\epsilon}\rho_{0\epsilon}}{\sqrt{\left(1-\rho_{1\epsilon}^2\right)\left(1-\rho_{0\epsilon}^2\right)}}.$$

It is easy to show that

$$\begin{aligned} & F_{10|D=1}^{*}\left(u_1, u_0; \rho_{10|\epsilon}\right) \\ =\ & \frac{\displaystyle\int_{-(w,x_c)'\gamma}^{+\infty} \int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}} \phi\left(\frac{\epsilon - \mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right) \Phi\left(\frac{u_1 - \bar{\mu}_1(s)}{\sqrt{(1-\rho_{1\epsilon}^2)\sigma_1^2 s}}, \frac{u_0 - \bar{\mu}_0(s)}{\sqrt{(1-\rho_{0\epsilon}^2)\sigma_0^2 s}}; \rho_{10|\epsilon}\right) ds dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w,x_c)'\gamma)}, \end{aligned}$$

where

$$\begin{aligned}\mu_j(s) &= \zeta_j(s - E(S)),\ j = 0, 1, \epsilon, \\ \bar{\mu}_j(s) &= \mu_j(s) + \rho_{j\epsilon}\frac{\sigma_j}{\sigma_\epsilon}\epsilon,\ j = 0, 1.\end{aligned}$$

When ρ_{10} reaches ρ_L (ρ_U), $\rho_{10|\epsilon}$ will reach -1 (1), so the sharp bounds on $F_{10}^Y(y_1, y_0|D = 1)$ are given by $F_{10}^L(y_1, y_0|D = 1)$ and $F_{10}^U(y_1, y_0|D = 1)$ respectively, where

$$\begin{aligned}F_{10}^Y(y_1, y_0|D &= 1) = F_{10|D=1}^*\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \rho_{10|\epsilon}\right), \\ F_{10}^L(y_1, y_0|D &= 1) = F_{10|D=1}^*\left(y_1 - x'\beta_1, y_0 - x'\beta_0; -1\right), \\ F_{10}^U(y_1, y_0|D &= 1) = F_{10|D=1}^*\left(y_1 - x'\beta_1, y_0 - x'\beta_0; 1\right).\end{aligned}$$

3.2 Sharp Bounds on the Distribution of Treatment Effects

In GH-SRMs, the individual treatment effect when $X = x$ is given by

$$\Delta = x'(\beta_1 - \beta_0) + U_1 - U_0.$$

Define $\gamma_1 = \sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon}$ and $\gamma_2^2 = \sigma_1^2 + \sigma_0^2 - 2\sigma_1\sigma_0\rho_{10}$. Then γ_2^2 satisfies $\sigma_L^2 \leq \gamma_2^2 \leq \sigma_U^2$, where

$$\sigma_U^2 = \sigma_1^2 + \sigma_0^2 - 2\rho_L\sigma_1\sigma_0,\ \sigma_L^2 = \sigma_1^2 + \sigma_0^2 - 2\rho_U\sigma_1\sigma_0.$$

Since $(U_1, U_0, \epsilon)' \sim GH_3(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$ and $(\Delta - x'(\beta_1 - \beta_0), U_0, \epsilon)' = B(U_1, U_0, \epsilon)'$, where

$$B = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

Proposition 3.13 in McNeil, Frey, and Embrechts (2005) implies that

$$\left(\Delta - x'(\beta_1 - \beta_0), U_0, \epsilon\right)' \sim GH_3\left(\lambda, \chi, \psi, B\mu, B\Sigma B', B\zeta\right),$$

where

$$\begin{aligned}B\mu &= \begin{pmatrix} \mu_1 - \mu_0 \\ \mu_0 \\ \mu_\epsilon \end{pmatrix},\ B\zeta = \begin{pmatrix} \zeta_1 - \zeta_0 \\ \zeta_0 \\ \zeta_\epsilon \end{pmatrix},\ \text{and} \\ B\Sigma B' &= \begin{pmatrix} \gamma_2^2 & \sigma_1\sigma_0\rho_{10} - \sigma_0^2 & \gamma_1\sigma_\epsilon \\ \sigma_1\sigma_0\rho_{10} - \sigma_0^2 & \sigma_0^2 & \sigma_0\sigma_\epsilon\rho_{0\epsilon} \\ \gamma_1\sigma_\epsilon & \sigma_0\sigma_\epsilon\rho_{0\epsilon} & \sigma_\epsilon^2 \end{pmatrix}.\end{aligned}$$

Applying Proposition 3.13 in McNeil, Frey, and Embrechts (2005) again, we get

$$\begin{aligned}\left(\Delta - x'(\beta_1 - \beta_0), \epsilon\right)' &\sim GH_2\left(\lambda, \chi, \psi, (B\mu)_{1\epsilon}, \left(B\Sigma B'\right)_{1\epsilon}, (B\zeta)_{1\epsilon}\right), \\ \Delta - x'(\beta_1 - \beta_0) &\sim GH_1\left(\lambda, \chi, \psi, \mu_1 - \mu_0, \gamma_2^2, \zeta_1 - \zeta_0\right), \\ \epsilon &\sim GH_1\left(\lambda, \chi, \psi, \mu_\epsilon, \sigma_\epsilon^2, \zeta_\epsilon\right),\end{aligned} \tag{11}$$

where

$$(B\mu)_{1\epsilon} = \begin{pmatrix} \mu_1 - \mu_0 \\ \mu_\epsilon \end{pmatrix}, \ (B\Sigma B')_{1\epsilon} = \begin{pmatrix} \gamma_2^2 & \gamma_1\sigma_\epsilon \\ \gamma_1\sigma_\epsilon & \sigma_\epsilon^2 \end{pmatrix}, \ (B\zeta)_{1\epsilon} = \begin{pmatrix} \zeta_1 - \zeta_0 \\ \zeta_\epsilon \end{pmatrix}.$$

Let $F_\Delta(\delta)$ and $F_\Delta(\delta|D=1)$ denote respectively the distribution of Δ conditional on $X = x$, $W = w$ and the distribution of Δ conditional on $X = x$, $W = w$, and $D = 1$. The Theorem below provides sharp bounds on $F_\Delta(\delta)$ and $F_\Delta(\delta|D=1)$.

THEOREM 3.2 *(i) It holds that* $F_\Delta^L(\delta) \leq F_\Delta(\delta) \leq F_\Delta^U(\delta)$, *where*

$$\begin{aligned} F_\Delta^U(\delta) &= \max_{\sigma_L^2 \leq \gamma_2^2 \leq \sigma_U^2} GH_1\left(\delta - x'(\beta_1 - \beta_0); \lambda, \chi, \psi, \mu_1 - \mu_0, \gamma_2^2, \zeta_1 - \zeta_0\right), \\ F_\Delta^L(\delta) &= \min_{\sigma_L^2 \leq \gamma_2^2 \leq \sigma_U^2} GH_1\left(\delta - x'(\beta_1 - \beta_0); \lambda, \chi, \psi, \mu_1 - \mu_0, \gamma_2^2, \zeta_1 - \zeta_0\right). \end{aligned}$$

If $\zeta_1 = \zeta_0$, *then*

$$\begin{aligned} F_\Delta^L(\delta) &= \begin{cases} GH_1\left(\delta - x'(\beta_1 - \beta_0); \lambda, \chi, \psi, \mu_1 - \mu_0, \sigma_U^2, 0\right) & \text{if } \delta \geq ATE \\ GH_1\left(\delta - x'(\beta_1 - \beta_0); \lambda, \chi, \psi, \mu_1 - \mu_0, \sigma_L^2, 0\right) & \text{if } \delta < ATE \end{cases}; \\ F_\Delta^U(\delta) &= \begin{cases} GH_1\left(\delta - x'(\beta_1 - \beta_0); \lambda, \chi, \psi, \mu_1 - \mu_0, \sigma_L^2, 0\right) & \text{if } \delta \geq ATE \\ GH_1\left(\delta - x'(\beta_1 - \beta_0); \lambda, \chi, \psi, \mu_1 - \mu_0, \sigma_U^2, 0\right) & \text{if } \delta < ATE \end{cases}. \end{aligned}$$

(ii) It holds that $F_\Delta^L(\delta|D=1) \leq F_\Delta(\delta|D=1) \leq F_\Delta^U(\delta|D=1)$, *where*

$$\begin{aligned} & F_\Delta^U(\delta|D=1) \\ = & \max_{\sigma_L^2 \leq \gamma_2^2 \leq \sigma_U^2} \int_{-\infty}^{\delta} \int_{-w'\gamma}^{\infty} \frac{gh_2\left(u - x'(\beta_1 - \beta_0), \epsilon; \lambda, \chi, \psi, (B\mu)_{1\epsilon}, (B\Sigma B')_{1\epsilon}, (B\zeta)_{1\epsilon}\right)}{1 - GH_1(-w'\gamma; \lambda, \chi, \psi, \mu_\epsilon, \sigma_\epsilon^2, \zeta_\epsilon)} du d\epsilon, \\ & F_\Delta^L(\delta|D=1) \\ = & \min_{\sigma_L^2 \leq \gamma_2^2 \leq \sigma_U^2} \int_{-\infty}^{\delta} \int_{-w'\gamma}^{\infty} \frac{gh_2\left(u - x'(\beta_1 - \beta_0), \epsilon; \lambda, \chi, \psi, (B\mu)_{1\epsilon}, (B\Sigma B')_{1\epsilon}, (B\zeta)_{1\epsilon}\right)}{1 - GH_1(-w'\gamma; \lambda, \chi, \psi, \mu_\epsilon, \sigma_\epsilon^2, \zeta_\epsilon)} du d\epsilon. \end{aligned}$$

For any fixed x and fixed δ, the bounds on $F_\Delta(\delta)$ are informative as long as $F_\Delta^L(\delta) \neq 0$ or $F_\Delta^U(\delta) \neq 1$. When either $\overline{\rho}_{1\epsilon}^2 = 1$ or $\overline{\rho}_{0\epsilon}^2 = 1$, we have $\rho_L = \rho_U$ and thus $\sigma_L^2 = \sigma_U^2$. Theorem 3.1 (i) implies that in this case, $F_\Delta(\cdot)$ is point identified. In general, $F_\Delta(\delta)$ is partially identified. However, consider the case that $\zeta_1 = \zeta_0$. In this case, the distribution of Δ is symmetric. Theorem 3.2 (i) implies that when $\delta = ATE$, $F_\Delta^L(\delta) = F_\Delta^U(\delta) = 0.5$. Hence $F_\Delta(ATE) = 0.5$, implying that the value of the distribution of Δ at the ATE is identified and that the median of the distribution of the outcome gain is the same as ATE.

We note that the same two sources of information contributing to the partial identification of the joint distribution also contribute to the partial identification of the distribution of Δ. Suppose $\zeta_1 = \zeta_0 = 0$ and $\bar{\rho}_{1\epsilon} = \bar{\rho}_{0\epsilon} = 0$. Theorem 2.1 (i) implies that $[\overline{\rho}_L, \overline{\rho}_U] = [-1, 1]$. In this case, $\sigma_L^2 = (\sigma_1 - \sigma_0)^2$ and $\sigma_U^2 = (\sigma_1 + \sigma_0)^2$. In general, $(\sigma_1 - \sigma_0)^2 \leq \sigma_L^2 \leq \sigma_U^2 \leq (\sigma_1 + \sigma_0)^2$. Theorem

3.2 (i) implies that taking into account self-selection tightens the bounds on $F_\Delta(\delta)$. Moreover, the following simple algebra demonstrates that the stronger the self-selection is, the tighter the bounds. For any δ, the width of the distribution bounds depends on σ_U and σ_L. Noting that

$$\sigma_U^2 - \sigma_L^2 = 4\sigma_1\sigma_0\sqrt{(1-\rho_{1\epsilon}^2)(1-\rho_{0\epsilon}^2)},$$

we conclude that the width of the distribution bounds on $F_\Delta(\delta)$ becomes narrower as the correlation between the selection error and the outcome errors become stronger.

Example 2.1. (Cont.) For SRMs with trivariate Gaussian errors, Theorem 3.2 (i) implies that $F_\Delta^L(\delta) \le F_\Delta(\delta) \le F_\Delta^U(\delta)$, where

$$\begin{aligned} F_\Delta^L(\delta) &= \begin{cases} \Phi\left(\frac{\delta - ATE}{\sigma_U}\right) & \text{if } \delta \ge ATE \\ \Phi\left(\frac{\delta - ATE}{\sigma_L}\right) & \text{if } \delta < ATE \end{cases}; \\ F_\Delta^U(\delta) &= \begin{cases} \Phi\left(\frac{\delta - ATE}{\sigma_L}\right) & \text{if } \delta \ge ATE \\ \Phi\left(\frac{\delta - ATE}{\sigma_U}\right) & \text{if } \delta < ATE \end{cases}. \end{aligned} \tag{12}$$

Example 2.2. (Cont.) Let $T_{[v]}(\cdot)$ denote the distribution function of the Student's t distribution with v degrees of freedom. For SRMs with trivariate Student's t errors, Theorem 3.2 (i) implies that $F_\Delta^L(\delta) \le F_\Delta(\delta) \le F_\Delta^U(\delta)$, where

$$\begin{aligned} F_\Delta^L(\delta) &= \begin{cases} T_{[v]}\left(\frac{\delta - ATE}{\sigma_U}\sqrt{\frac{v}{v-2}}\right) & \text{if } \delta \ge ATE \\ T_{[v]}\left(\frac{\delta - ATE}{\sigma_L}\sqrt{\frac{v}{v-2}}\right) & \text{if } \delta < ATE \end{cases}; \\ F_\Delta^U(\delta) &= \begin{cases} T_{[v]}\left(\frac{\delta - ATE}{\sigma_L}\sqrt{\frac{v}{v-2}}\right) & \text{if } \delta \ge ATE \\ T_{[v]}\left(\frac{\delta - ATE}{\sigma_U}\sqrt{\frac{v}{v-2}}\right) & \text{if } \delta < ATE \end{cases}. \end{aligned}$$

4 Distribution Bounds in Semiparametric SRMs

The bounds for GH-SRMs established in Section 3 depend crucially on the parametric distribution assumption, especially the implied joint GH distribution of the potential outcomes. Given that the assumption of joint GH distribution of the potential outcomes can never be verified empirically, it is important to investigate the robustness of these bounds to the corresponding distributional assumptions and to establish bounds that do not rely on them. This will be accomplished in the current section.

For generality, we adopt Heckman (1990)'s notation and consider the following semiparametric SRM:

$$\begin{aligned} Y_{1i} &= g_1(X_{1i}, X_{ci}) + U_{1i}, \\ Y_{0i} &= g_0(X_{0i}, X_{ci}) + U_{0i}, \\ D_i &= I_{\{(W_i, X_{ci})'\gamma + \epsilon_i > 0\}}, \ i = 1, \ldots, n, \end{aligned} \tag{13}$$

where both $g_1(x_1, x_c)$, $g_0(x_0, x_c)$ and the distribution of $(U_{1i}, U_{0i}, \epsilon_i)'$ are completely unknown. Heckman (1990) provided conditions under which the distributions of $(U_{1i}, \epsilon_i)'$ and $(U_{0i}, \epsilon_i)'$, $g_1(x_1, x_c)$, $g_0(x_0, x_c)$, and γ are point identified from the sample information alone. However, the joint distribution of $(U_{1i}, U_{0i})'$ is not (point) identified.

In this section, we provide sharp bounds on the joint distribution of U_{1i}, U_{0i} or Y_{1i}, Y_{0i} and the distribution of Δ_i. We assume independence of the errors $U_{1i}, U_{0i}, \epsilon_i$ and the regressors $X_{1i}, X_{0i}, X_{ci}, W_i$. We note that the covariance approach used in Section 3 is not applicable here, as the distribution of $(U_{1i}, U_{0i}, \epsilon_i)'$ is completely unknown. Instead we make use of the Fréchet-Hoeffding inequality in (10) and existing results on bounding the distribution of a difference of two random variables each having a given distribution function. Again, we omit the subscript i in the rest of Section 4 and Section 5.

4.1 Sharp Bounds on the Distribution of a Difference of Two Random Variables

Sharp bounds on distributions of functions of random variables Y_1 and Y_0 including the four simple arithmetic operations are presented in Williamson and Downs (1990). For a sum of two random variables, Makarov (1981), Rüschendorf (1982), and Frank, Nelsen, and Schweizer (1987) establish sharp bounds on its distribution, see also Nelsen (1999). Frank, Nelsen, and Schweizer (1987) demonstrate that their proof based on copulas can be extended to more general functions than the sum. In this subsection, we will present the relevant results for the difference between two random variables. Specifically, let $\Delta = Y_1 - Y_0$ and $F_\Delta(\cdot)$ denote the distribution function of Δ. The following lemma presents sharp bounds on $F_\Delta(\cdot)$ when only F_1 and F_0 are known.

Lemma 4.1 *Let* $F_{\min}(\delta) = \sup_{y_1} \max(F_1(y_1) - F_0(y_1 - \delta), 0)$ *and* $F_{\max}(\delta) = 1 + \inf_{y_1} \min(F_1(y_1) - F_0(y_1 - \delta), 0)$. *Then* $F_{\min}(\delta) \leq F_\Delta(\delta) \leq F_{\max}(\delta)$.

Viewed as an inequality among all possible distribution functions, the sharp bounds $F_{\min}(\delta)$ and $F_{\max}(\delta)$ 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 $F_{\min}(\delta) = F_\Delta(\delta) = F_{\max}(\delta)$. In fact, given any pair of distribution functions F_1 and F_0, the inequality: $F_{\min}(\delta) \leq F_\Delta(\delta) \leq F_{\max}(\delta)$ cannot be improved, that is, the bounds $F_{\min}(\delta)$ and $F_{\max}(\delta)$ 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. Unlike the sharp bounds on the correlation coefficient between Y_1, Y_0 or the joint distribution of Y_1, Y_0 which are reached at the Fréchet-Hoeffding lower and upper bounds for the distribution of Y_1, Y_0 when Y_1 and Y_0 are perfectly negatively dependent or perfectly positive dependent (see Fan (2005)), the 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. Frank, Nelsen, and Schweizer (1987) provided explicit expressions for copulas that reach the bounds on the distribution of Δ.

Explicit expressions for bounds on the distribution of a sum of two random variables are available for the case where the distributions of both random variables belong to the same family which includes the uniform, the normal, the Cauchy, and the exponential families, see Alsina (1981), Frank, Nelsen, and Schweizer (1987), and Denuit, Genest, and Marceau (1999). Below we provide expressions for $F_{\min}(\delta)$ and $F_{\max}(\delta)$ when both Y_1 and Y_0 are normal or Student's t.

Example 4.1. Let $Y_1 \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y_0 \sim N\left(\mu_0, \sigma_0^2\right)$. Fan and Park (2006) provide the following expressions for the bounds $F_{\min}(\delta)$ and $F_{\max}(\delta)$:

(i) If $\sigma_1 = \sigma_0 = \sigma$, then

$$F_{\min}(\delta) = \begin{cases} 0 & \text{if } \delta < \mu_1 - \mu_0, \\ 2\Phi\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right) - 1 & \text{if } \delta \geq \mu_1 - \mu_0, \end{cases} \tag{14}$$

$$F_{\max}(\delta) = \begin{cases} 2\Phi\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right) & \text{if } \delta < \mu_1 - \mu_0, \\ 1 & \text{if } \delta \geq \mu_1 - \mu_0. \end{cases} \tag{15}$$

(ii) If $\sigma_1 \neq \sigma_0$, then

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

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

where $s = \delta - (\mu_1 - \mu_0)$ and $t = \left(s^2 + 2\left(\sigma_1^2 - \sigma_0^2\right)\ln\left(\frac{\sigma_1}{\sigma_0}\right)\right)^{\frac{1}{2}}$.

Example 4.2. For $j = 0, 1$, we assume $\frac{Y_j-\mu_j}{\sigma_j}\sqrt{\frac{v_j}{v_j-2}} \sim t_{[v_j]}$, where $v_j > 2$, so that $E(Y_j) = \mu_j$, $Var(Y_j) = \sigma_j^2$ and $F_j(\delta) = T_{[v_j]}\left(\left(\frac{\delta-\mu_j}{\sigma_j}\right)\sqrt{\frac{v_j}{v_j-2}}\right)$.

By Lemma 4.1, $F_{\min}(\delta) = \max(F_1(x_1^*) - F_0(x_1^* - \delta), 0)$ and $F_{\max}(\delta) = 1 + \min(F_1(x_2^*) - F_0(x_2^* - \delta), 0)$, where x_1^* and x_2^* are the maximizer and minimizer of the function $[F_1(x) - F_0(x - \delta)]$ respectively, i.e., x_1^*, x_2^* satisfy the equation:

$$\frac{1}{\sigma_1}\sqrt{\frac{v_1}{v_1-2}}t_{[v_1]}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)\sqrt{\frac{v_1}{v_1-2}}\right) = \frac{1}{\sigma_0}\sqrt{\frac{v_0}{v_0-2}}t_{[v_0]}\left(\left(\frac{x-\mu_0-\delta}{\sigma_0}\right)\sqrt{\frac{v_0}{v_0-2}}\right).$$

In general, one must solve the above equation and hence evaluate $F_{\min}(\delta)$ and $F_{\max}(\delta)$ numerically. When $v_1 = v_0 \equiv v$ (say), we are able to get closed-form expressions for $F_{\min}(\delta)$ and $F_{\max}(\delta)$ as follows:

(i) If $\sigma_1 = \sigma_0 = \sigma$, then

$$F_{\min}(\delta) = \begin{cases} 0 & \text{if } \delta < \mu_1 - \mu_0, \\ 2T_{[v]}\left(\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right)\sqrt{\frac{v}{v-2}}\right) - 1 & \text{if } \delta \geq \mu_1 - \mu_0, \end{cases} \tag{16}$$

$$F_{\max}(\delta) = \begin{cases} 2T_{[v]}\left(\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right)\sqrt{\frac{v}{v-2}}\right) & \text{if } \delta < \mu_1 - \mu_0, \\ 1 & \text{if } \delta \geq \mu_1 - \mu_0. \end{cases} \tag{17}$$

(ii) If $\sigma_1 \neq \sigma_0$, then

$$\begin{aligned} & F_{\min}(\delta) \\ = \; & T_{[v]}\left(\left(\frac{\sigma_1^{2\kappa-1}s - \sigma_0^{\kappa}\sigma_1^{\kappa-1}t}{\sigma_1^{2\kappa} - \sigma_0^{2\kappa}}\right)\sqrt{\frac{v}{v-2}}\right) + T_{[v]}\left(\left(\frac{\sigma_1^{\kappa}\sigma_0^{\kappa-1}t - \sigma_0^{(2\kappa-1)}s}{\sigma_1^{2\kappa} - \sigma_0^{2\kappa}}\right)\sqrt{\frac{v}{v-2}}\right) - 1, \\ & F_{\max}(\delta) \\ = \; & T_{[v]}\left(\left(\frac{\sigma_1^{2\kappa-1}s + \sigma_0^{\kappa}\sigma_1^{\kappa-1}t}{\sigma_1^{2\kappa} - \sigma_0^{2\kappa}}\right)\sqrt{\frac{v}{v-2}}\right) - T_{[v]}\left(\left(\frac{\sigma_1^{\kappa}\sigma_0^{\kappa-1}t + \sigma_0^{(2\kappa-1)}s}{\sigma_1^{2\kappa} - \sigma_0^{2\kappa}}\right)\sqrt{\frac{v}{v-2}}\right) + 1, \end{aligned}$$

where $s = \delta - (\mu_1 - \mu_0)$, $\kappa = \frac{v}{v+1}$, and

$$t = \left(s^2 + \left(\sigma_1^{2\kappa} - \sigma_0^{2\kappa}\right)\left(\sigma_1^{2(1-\kappa)} - \sigma_0^{2(1-\kappa)}\right)(v-2)\right)^{\frac{1}{2}}.$$

It is easy to see that in both cases, the expressions for $F_{\min}(\delta)$ and $F_{\max}(\delta)$ reduce to those in Example 4.1 as $v \to +\infty$. For instance, consider the case where $\sigma_1 \neq \sigma_0$. As $v \to +\infty$, we have $\kappa \to 1$, $\sqrt{\frac{v}{v-2}} \to 1$, and $\left(\sigma_1^{2(1-\kappa)} - \sigma_0^{2(1-\kappa)}\right)(v-2) \to 2\log\left(\frac{\sigma_1}{\sigma_0}\right)$.

4.2 Semiparametric SRMs

Let $F_{1\epsilon}(u_1, \epsilon)$ and $F_{0\epsilon}(u_0, \epsilon)$ denote respectively the distribution functions of $(U_1, \epsilon)'$ and $(U_0, \epsilon)'$ in model (13). Since $F_{1\epsilon}(u_1, \epsilon)$ and $F_{0\epsilon}(u_0, \epsilon)$ are identified from the sample information, the joint distribution of U_1, U_0, ϵ belongs to the Fréchet class of trivariate distributions for which the (1,3) and (2,3) bivariate margins are given or fixed, denoted as $\mathcal{F}(F_{1\epsilon}, F_{0\epsilon})$. Joe (1997) showed that for any $F_{10\epsilon} \in \mathcal{F}(F_{1\epsilon}, F_{0\epsilon})$, it must satisfy

$$\int_{-\infty}^{\epsilon} C_L\left[F_{1|\epsilon}(u_1), F_{0|\epsilon}(u_0)\right] dF_\epsilon(\epsilon) \leq F_{10\epsilon}(u_1, u_0, \epsilon) \leq \int_{-\infty}^{\epsilon} C_U\left[F_{1|\epsilon}(u_1), F_{0|\epsilon}(u_0)\right] dF_\epsilon(\epsilon), \tag{18}$$

where $F_{j|\epsilon}(u_j)$ denote the conditional distribution of U_j given ϵ, $j = 1, 0$ and $F_\epsilon(\epsilon)$ the marginal distribution function of ϵ. Inequality (18) follows from the Fréchet-Hoeffding inequality and the expression: $F_{10\epsilon}(u_1, u_0, \epsilon) = \int_{-\infty}^{\epsilon} F_{10|\epsilon}(u_1, u_0) dF_\epsilon(\epsilon)$, where $F_{10|\epsilon}(u_1, u_0)$ is the conditional joint distribution of U_1, U_0 given ϵ.

THEOREM 4.2 *In a semiparametric SRM, the following inequalities hold.*

(i) *ATE: The joint distribution of potential outcomes satisfies*

$$F_{10}^{L}(y_1, y_0) \leq F_{10}^{Y}(y_1, y_0) \leq F_{10}^{U}(y_1, y_0), \tag{19}$$

where

$$\begin{aligned} F_{10}^{L}(y_1, y_0) &= \int_{-\infty}^{\infty} C_L \left[F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\right] dF_\epsilon(\epsilon), \\ F_{10}^{U}(y_1, y_0) &= \int_{-\infty}^{\infty} C_U \left[F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\right] dF_\epsilon(\epsilon). \end{aligned} \tag{20}$$

(ii) *TT: The joint distribution of potential outcomes for the treated satisfies*

$$F_{10}^{L}(y_1, y_0|D=1) \leq F_{10}^{Y}(y_1, y_0|D=1) \leq F_{10}^{U}(y_1, y_0|D=1),$$

where

$$F_{10}^{L}(y_1, y_0|D=1) = \frac{\int_{-(w,x_c)'\gamma}^{\infty} C_L \left(F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\right) dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w, x_c)'\gamma)},$$
$$F_{10}^{U}(y_1, y_0|D=1) = \frac{\int_{-(w,x_c)'\gamma}^{\infty} C_U \left(F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\right) dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w, x_c)'\gamma)}.$$

The result in (i) is presented in Lee (2002). It is an immediate consequence of (18) when $\epsilon = \infty$. To prove (ii), we note that

$$\begin{aligned} & F_{10}^{Y}(y_1, y_0|D=1) = P(U_{1i} \leq y_1 - g_1(x_1, x_c), U_{0i} \leq y_0 - g_0(x_0, x_c)|\epsilon_i > -(w, x_c)'\gamma) \\ = & \frac{P(U_{1i} \leq y_1 - g_1(x_1, x_c), U_{0i} \leq y_0 - g_0(x_0, x_c), \epsilon_i > -(w, x_c)'\gamma)}{P(\epsilon_i > -(w, x_c)'\gamma)} \\ = & \frac{\int_{-(w,x_c)'\gamma}^{\infty} F_{10|\epsilon}(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c)) dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w, x_c)'\gamma)}. \end{aligned} \tag{21}$$

Now since $F_{10|\epsilon}(y_1 - g_1(x_1, x_c), g_0(x_0, x_c))$ satisfies the Fréchet-Hoeffding inequality, we obtain the inequality in (ii).

The lower or upper bounds in Theorem 4.2 are reached when the two potential outcomes are conditionally (on ϵ) perfectly negatively or positively dependent on each other. One example is $\epsilon = U_1 - U_0$ in which U_1, U_0 are perfectly positively dependent conditional on ϵ and the upper bound is reached. These bounds take into account the self-selection process and are tighter than the bounds obtained under random selection. For instance, if selection is random, i.e., both U_1 and U_0 are independent of ϵ, then the bounds in Theorem 4.2 (i) become

$$F_{10}^{LI}(y_1, y_0) = C_L \left[F_1(y_1 - g_1(x_1, x_c)), F_0(y_0 - g_0(x_0, x_c))\right], \tag{22}$$

$$F_{10}^{UI}(y_1, y_0) = C_U \left[F_1(y_1 - g_1(x_1, x_c)), F_0(y_0 - g_0(x_0, x_c))\right]. \tag{23}$$

In general, $F_{10}^{LI}(y_1, y_0) \leq F_{10}^{L}(y_1, y_0)$ and $F_{10}^{UI}(y_1, y_0) \geq F_{10}^{U}(y_1, y_0)$ implying that the dependence between the outcome errors and the selection error improves on the bounds on $F_{10}^{Y}(y_1, y_0)$. When the distribution of either U_1 or U_0 conditional on ϵ is degenerate at a finite value, the lower and upper bounds in Theorem 4.2 (i) coincide and thus point identify $F_{10}(y_1, y_0)$ for any y_1, y_0.

We now consider sharp bounds on the distribution of $\Delta = Y_1 - Y_0$. Note that

$$ATE \equiv E(\Delta | X = x) = g_1(x_1, x_c) - g_0(x_0, x_c)$$

and $F_\Delta(\delta) = E\left[P(U_1 - U_0 \leq \{\delta - ATE\} | \epsilon)\right]$. Applying Lemma 4.1 to $P(U_1 - U_0 \leq \{\delta - ATE\} | \epsilon)$, we obtain the sharp bounds on the distribution of the treatment effect in Theorem 4.3 (i) below. Other bounds presented in Theorem 4.3 can be obtained in the same way.

THEOREM 4.3 *In a semiparametric SRM, the following inequalities hold.*

(i) *ATE:* $F_\Delta^L(\delta) \leq F_\Delta(\delta) \leq F_\Delta^U(\delta)$, *where*

$$\begin{aligned} F_\Delta^L(\delta) &= \int_{-\infty}^{+\infty} \left[\sup_u \max\left\{F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - ATE\}), 0\right\}\right] dF_\epsilon(\epsilon), \\ F_\Delta^U(\delta) &= \int_{-\infty}^{+\infty} \left[\inf_u \min\left\{1 - F_{0|\epsilon}(u - \{\delta - ATE\}) + F_{1|\epsilon}(u), 1\right\}\right] dF_\epsilon(\epsilon). \end{aligned}$$

(ii) *TT: The distribution of* Δ *for the treated satisfies*

$$F_\Delta^L(\delta | D = 1) \leq F_\Delta(\delta | D = 1) \leq F_\Delta^U(\delta | D = 1),$$

where

$$\begin{aligned} F_\Delta^L(\delta | D = 1) &= \frac{\int_{-(w,x_c)'\gamma}^{\infty} \left[\sup_u \max\left\{F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - ATE\}), 0\right\}\right] dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w, x_c)'\gamma)}, \\ F_\Delta^U(\delta | D = 1) &= \frac{\int_{-(w,x_c)'\gamma}^{\infty} \left[\inf_u \min\left\{1 - F_{0|\epsilon}(u - \{\delta - ATE\}) + F_{1|\epsilon}(u), 1\right\}\right] dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w, x_c)'\gamma)}. \end{aligned}$$

In contrast to sharp bounds on the joint distribution of potential outcomes, the sharp bounds on the distribution of the treatment effect are not reached at conditional perfect positive or negative dependence. Again, self-selection and the identified marginals of $F_{10}^{Y}(y_1, y_0)$ contribute to the partial identification of the distribution of Δ. When the distribution of either U_1 or U_0 conditional on ϵ is degenerate at a finite value, the lower and upper bounds in Theorem 4.3 (i) coincide and thus point identify $F_\Delta(\delta)$ for any δ.

When $\epsilon = U_1 - U_0$, the potential outcome errors are perfectly positively dependent conditional on ϵ. Let $F_{\Delta|\epsilon}^R$ and F_Δ^R denote respectively the conditional distribution of Δ on ϵ and the unconditional distribution of Δ in this case. Fan and Park (2006) shows that $F^R_{\Delta|\epsilon}$ second order stochastically dominates any outcome gain distribution conditional on ϵ, $F_{\Delta|\epsilon}$. Taking expectation with respect to ϵ, we obtain the following theorem.

THEOREM 4.4 *In a semiparametric SRM, F^R_{Δ} second order stochastically dominates any F_{Δ} consistent with the sample information.*

Unlike the average treatment parameters such as ATE and TT, the quantile of Δ is in general not identified. By inverting the distribution bounds in Theorem 4.3, we obtain sharp bounds on the quantile of the treatment effect[3] for the whole population and the subpopulation receiving treatment.

4.3 Some Applications of the Distribution Bounds

By using the distribution bounds established in the previous subsection, we can provide informative bounds on many interesting effects other than the average treatment effect. Some illustrative examples are discussed below, see Heckman, Smith, and Clements (1997) and Vijverberg (1993) for more examples.

1. The proportion of people participating in the program who benefit from it,

$$P(Y_1 > Y_0|D = 1) = P(\Delta > 0|D = 1) = 1 - F_{\Delta}(0|D = 1).$$

2. The proportion of the total population that benefits from the program,

$$P(Y_1 > Y_0|D = 1)P(D = 1) = \{1 - F_{\Delta}(0|D = 1)\}\, P(D = 1).$$

3. The share of 'productive' workers employed in sector 1,

$$P\left(D = 1|Y_1 > Y_0\right) = \frac{\{1 - F_{\Delta}(0|D = 1)\}\, P(D = 1)}{1 - F_{\Delta}(0)}.$$

4. The distribution of the potential outcome Y_1 of an individual with an above average Y_0,

$$P(Y_1 \leq y_1|U_0 > 0) = \frac{F_1\left(y_1 - g_1(x_1, x_c)\right) - F_{10}\left(y_1 - g_1(x_1, x_c), 0\right)}{1 - F_0(0)}.$$

[3] Recently, $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ has been used to study treatment effect heterogeneity and is referred to as the quantile treatment effect (QTE), see e.g., Heckman, Smith, and Clements (1997), Abadie, Angrist, and Imbens (2002), Chen, Hong, and Tarozzi (2004), Chernozhukov and Hansen (2005), Firpo (2007), 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.

5. The variance of the treatment effect for participants (Lee, 2002),

$$\sigma^2_{L,D=1} \leq Var(\Delta|D=1) \leq \sigma^2_{U,D=1},$$

where

$$\begin{aligned}
\sigma^2_{U,D=1} &= Var\left(Y_1|D=1\right) + Var\left(Y_0|D=1\right) \\
&\quad -2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(F^L_{10}(y_1,y_0|D=1) - F_1\left(y_1|D=1\right)F_0\left(y_0|D=1\right)\right)dy_1dy_0, \\
\sigma^2_{L,D=1} &= Var\left(Y_1|D=1\right) + Var\left(Y_0|D=1\right) \\
&\quad -2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(F^U_{10}(y_1,y_0|D=1) - F_1\left(y_1|D=1\right)F_0\left(y_0|D=1\right)\right)dy_1dy_0.
\end{aligned}$$

Similar techniques used in the previous subsection may help establish bounds on other parameters of interest. For example, the distribution of the potential outcome Y_1 of an individual with an above average Y_0 who selects into the program is given by

$$P(Y_1 \leq y_1|D=1, U_0 > 0) = \frac{P\left(Y_1 \leq y_1, \epsilon \geq -(w,x_c)'\gamma\right) - \int_{-(w,x_c)'\gamma}^{\infty} F_{10|\epsilon}(y_1 - g_1(x_1,x_c), 0)dF_\epsilon(\epsilon)}{P\left(\epsilon \geq -(w,x_c)'\gamma, U_0 > 0\right)},$$

where the probability in the denominator and the first probability in the numerator are identified from the sample information and the second term in the numerator can be bounded by applying the Fréchet-Hoeffding inequality to $F_{10|\epsilon}(y_1 - g_1(x_1,x_c),0)$.

5 A Comparison of the two sets of Bounds

The distribution bounds developed in Section 4 depend on the bivariate distributions of (U_1,ϵ) and (U_0,ϵ) which can be parametric or nonparametric. In this section, we first study these bounds when (U_j,ϵ), $j=1,0$, follow bivariate GH distributions and then compare them with those established in Section 3 for GH-SRMs. The difference between these two sets of bounds is that the former bounds are valid for any joint distribution of the errors U_1, U_0, ϵ, provided that the bivariate marginal distributions corresponding to $\{U_1,\epsilon\}$ and $\{U_0,\epsilon\}$ are bivariate GH distributions, while the bounds in Section 3 depend crucially on the joint GH distribution for the trio of errors $\{U_1,U_0,\epsilon\}$. Robustness of the former results to the joint distribution of U_1 and U_0 is a desirable property, as this joint distribution is not identifiable from the sample information alone and any distributional assumption imposed on it can never be verified empirically.

5.1 Bounds on F^Y_{10} in Semiparametric SRMs with Bivariate GH Distributions

For $j=0,1$, assume $(U_j,\epsilon)' \sim GH_2\left(\lambda,\chi,\psi,\mu_{j\epsilon},\Sigma_{j\epsilon},\zeta_{j\epsilon}\right)$, where

$$\mu_{j\epsilon} = \begin{pmatrix} \mu_j \\ \mu_\epsilon \end{pmatrix},\ \zeta_{j\epsilon} = \begin{pmatrix} \zeta_j \\ \zeta_\epsilon \end{pmatrix},\ \Sigma_{j\epsilon} = \begin{pmatrix} \sigma_j^2 & \sigma_j\sigma_\epsilon\rho_{j\epsilon} \\ \sigma_j\sigma_\epsilon\rho_{j\epsilon} & \sigma_\epsilon^2 \end{pmatrix}.$$

By making use of the fact that the conditional distribution of U_j given ϵ is GH for $j = 0, 1$**,** we show in the Appendix that the following theorem holds.

THEOREM 5.1 *In a semiparametric SRM with bivariate GH distributions for $\{U_j, \epsilon\}$ for $j = 0, 1$ with $\zeta_1 = \zeta_0 = 0$, we have:*

$$
\begin{aligned}
F_{10}^L(y_1, y_0) &= \int_{-\infty}^{\infty} \max\{F_{1|\epsilon}(y_1 - g_1(x_1, x_c)) + F_{0|\epsilon}(y_0 - g_0(x_0, x_c)) - 1, 0\} dF_\epsilon(\epsilon) \\
&= GH_2\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \left[\alpha_{10}^-, \rho_L\right]\right), \qquad (24)
\end{aligned}
$$

$$
\begin{aligned}
F_{10}^U(y_1, y_0) &= \int_{-\infty}^{\infty} \min\{F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\} dF_\epsilon(\epsilon) \\
&= GH_2\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \left[\alpha_{10}^-, \rho_U\right]\right). \qquad (25)
\end{aligned}
$$

We observe immediately that these bounds are the same as the bounds on the joint distribution of potential outcomes in GH-SRMs presented in Section 3. This is interesting, because it implies that the non-refutable GH assumption on the joint distribution of the potential outcomes in GH-SRMs with symmetric outcome errors does not improve on the bounds of this joint distribution. Heuristically, this is because the conditional copula for $\{U_1, U_0\}$ given ϵ implied by the trivariate GH distribution assumption in GH-SRMs is still a GH copula. Since the partial correlation between U_1 and U_0 ranges from -1 to 1 when $\zeta_1 = \zeta_0 = 0$, the conditional copula for $\{U_1, U_0\}$ given ϵ interpolates between the lower bound copula and the upper bound copula, resulting in the same bounds as if the conditional copula for $\{U_1, U_0\}$ is unrestricted at all.

5.2 Bounds on F_Δ in Semiparametric SRMs with Bivariate GH Distributions

Theorem 4.3 (i) provides bounds on F_Δ for any pair of bivariate distributions for $\{U_j, \epsilon\}$, $j = 0, 1$ including GH distributions. In this subsection, we use Examples 4.1 and 4.2 to simplify the expressions for bivariate normal marginals and bivariate Student's t marginals.

Suppose $\{U_j, \epsilon\}$ follows a bivariate normal distribution:

$$
\begin{pmatrix} U_j \\ \epsilon \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & \sigma_j\rho_{j\epsilon} \\ \sigma_j\rho_{j\epsilon} & 1 \end{pmatrix}\right].
$$

The distribution of U_j given ϵ follows a univariate normal distribution with mean $\sigma_j\rho_{j\epsilon}\epsilon$ and variance $\sigma_j^2(1-\rho_{j\epsilon}^2)$, $j = 1, 0$. Example 4.1 provides bounds on the distribution of Δ given ϵ, i.e., expressions for

$$
\sup_u \max\left\{F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - ATE\}), 0\right\}
$$

and

$$
\inf_u \min\left\{1 - F_{0|\epsilon}(u - \{\delta - ATE\}) + F_{1|\epsilon}(u), 1\right\}
$$

in Theorem 4.3 (i). Taking their expectations with respect to ϵ leads to the following bounds on $F_\Delta(\delta)$.

THEOREM 5.2 *In a semiparametric SRM with bivariate normal distributions for* $\{U_j, \epsilon\}$ *for* $j = 0, 1$, *we have:*

(i) *If* $\sigma_1\sqrt{1-\rho_{1\epsilon}^2} = \sigma_0\sqrt{1-\rho_{0\epsilon}^2}$ *and* $\rho_{j\epsilon}^2 \neq 1$, *then*

$$F_\Delta^L(\delta) = 2\int_A \Phi\left(\frac{\{\delta - ATE\} - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon}{2\sigma_1\sqrt{1-\rho_{1\epsilon}^2}}\right)\phi(\epsilon)d\epsilon - P(A),$$

$$F_\Delta^U(\delta) = 2\int_{A^C} \Phi\left(\frac{\{\delta - ATE\} - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon}{2\sigma_1\sqrt{1-\rho_{1\epsilon}^2}}\right)\phi(\epsilon)d\epsilon + P(A),$$

where $A = \{\epsilon : \{\delta - ATE\} \geq (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon\}$ *and* A^C *is the complement of* A. *When* $(\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon}) = 0$, A *is the whole real line if* $\delta \geq ATE$, *else* A *is an empty set;*

(ii) If $\sigma_1\sqrt{1-\rho_{1\epsilon}^2} \neq \sigma_0\sqrt{1-\rho_{0\epsilon}^2}$ and $\rho_{j\epsilon}^2 \neq 1$, then

$$\begin{aligned} F_\Delta^L(\delta) &= \int_{-\infty}^{+\infty} \Phi\left(\frac{\sigma_1\sqrt{(1-\rho_{1\epsilon}^2)}s - \sigma_0\sqrt{(1-\rho_{0\epsilon}^2)}t}{\sigma_1^2(1-\rho_{1\epsilon}^2) - \sigma_0^2(1-\rho_{0\epsilon}^2)}\right)\phi(\epsilon)d\epsilon \\ &\quad + \int_{-\infty}^{+\infty} \Phi\left(\frac{\sigma_1\sqrt{(1-\rho_{1\epsilon}^2)}t - \sigma_0\sqrt{(1-\rho_{0\epsilon}^2)}s}{\sigma_1^2(1-\rho_{1\epsilon}^2) - \sigma_0^2(1-\rho_{0\epsilon}^2)}\right)\phi(\epsilon)d\epsilon - 1, \end{aligned}$$

$$\begin{aligned} F_\Delta^U(\delta) &= \int_{-\infty}^{+\infty} \Phi\left(\frac{\sigma_1\sqrt{(1-\rho_{1\epsilon}^2)}s + \sigma_0\sqrt{(1-\rho_{0\epsilon}^2)}t}{\sigma_1^2(1-\rho_{1\epsilon}^2) - \sigma_0^2(1-\rho_{0\epsilon}^2)}\right)\phi(\epsilon)d\epsilon \\ &\quad - \int_{-\infty}^{+\infty} \Phi\left(\frac{\sigma_1\sqrt{(1-\rho_{1\epsilon}^2)}t + \sigma_0\sqrt{(1-\rho_{0\epsilon}^2)}s}{\sigma_1^2(1-\rho_{1\epsilon}^2) - \sigma_0^2(1-\rho_{0\epsilon}^2)}\right)\phi(\epsilon)d\epsilon + 1, \end{aligned}$$

where $s = \{\delta - ATE\} - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon$ and

$$t = \left(s^2 + 2\left[\sigma_1^2(1-\rho_{1\epsilon}^2) - \sigma_0^2(1-\rho_{0\epsilon}^2)\right]\ln\left(\frac{\sigma_1\sqrt{1-\rho_{1\epsilon}^2}}{\sigma_0\sqrt{1-\rho_{0\epsilon}^2}}\right)\right)^{\frac{1}{2}}.$$

In contrast to the sharp bounds on the joint distribution of the potential outcomes in Theorem 5.1, the bounds given above on the distribution of the outcome gain differ from the corresponding bounds in Gaussian SRMs and are in general wider, because they are valid for any trivariate distribution with bivariate normal marginals for (U_1, ϵ) and (U_0, ϵ), not necessarily the trivariate Normal distribution in Gaussian SRMs. On the one hand, imposing the trivariate normality assumption narrows the width of the bounds, but on the other hand, it may lead to misleading conclusions if the implied normality assumption for the joint distribution of potential outcomes is violated. To see the seriousness of this problem, remember in Gaussian SRMs, the value of the treatment effect distribution at its mean is always identified: $F_\Delta(ATE) = 0.5$. However, if the joint distribution of the potential outcomes is unknown, then $F_\Delta(ATE)$ is in general not identified and the bounds on $F_\Delta(ATE)$ depend on the parameters of the identified bivariate distributions.

In Figure 1, we plotted the two sets of bounds on $F_\Delta(\cdot)$ in Gaussian SRMs and semiparametric SRMs with bivariate normal marginals. We fixed $ATE = 0$, $\sigma_1^2 = 1$ and $\sigma_0^2 = 1$. For $\rho_{1\epsilon} = 0.5$, we chose a range of values for $\rho_{0\epsilon}$. We also plotted the bounds when $\rho_{1\epsilon} = \rho_{0\epsilon} = 0$. Solid curves are bounds in Theorem 5.2 assuming bivariate normality (BN) for (U_j, ϵ) only, while dashed curves are bounds in (6) assuming trivariate normality (TN) for (U_j, U_0, ϵ). Several general conclusions emerge from Figure 1. First, for any given set of parameter values, the bounds under bivariate normal marginals are always wider than the bounds under the trivariate normal assumption; Second, for given δ, the bounds in general become narrower as the dependence between U_0 and ϵ as measured by the magnitude of $\rho_{0\epsilon}$ increases except when $\delta = 0$ in Gaussian SRMs in which case the lower and upper bounds coincide and become 0.5. In the extreme cases where either $\rho_{1\epsilon}^2 = 1$ or $\rho_{0\epsilon}^2 = 1$, the two sets of bounds coincide and both identify the distribution of Δ. More specifically, we have

$$F_\Delta(\delta) = F_\Delta^L(\delta) = F_\Delta^U(\delta) = \Phi\left(\frac{\delta - ATE}{\sqrt{\sigma_1^2 + \sigma_0^2 - 2\rho_{0\epsilon}\sigma_1\sigma_0}}\right), \text{ if } \rho_{1\epsilon} = 1, \tag{26}$$

$$F_\Delta(\delta) = F_\Delta^L(\delta) = F_\Delta^U(\delta) = \Phi\left(\frac{\delta - ATE}{\sqrt{\sigma_1^2 + \sigma_0^2 + 2\rho_{0\epsilon}\sigma_1\sigma_0}}\right), \text{ if } \rho_{1\epsilon} = -1. \tag{27}$$

Third, the bounds corresponding to $(\rho_{1\epsilon}, \rho_{0\epsilon}) = (0, 0)$ are wider than the bounds when $(\rho_{1\epsilon}, \rho_{0\epsilon}) \neq (0, 0)$, because the former does not account for the information through self-selection.

To see how these bounds change with the variance parameters. In Figure 2, we plotted the bounds on $F_\Delta(\delta)$ against σ_0 at $\delta = 0, 1, 4$ when $\sigma_1^2 = 1$, $\rho_{1\epsilon} = 0.5$ and $\rho_{0\epsilon} = 0.5$. One interesting fact we observe is that the distribution bounds under both trivariate normality and bivariate normality become wider to some point and then narrower as σ_0 goes to ∞.

Suppose $\{U_j, \epsilon\}$ follows a bivariate Student's t distribution:

$$\left\{\sqrt{\frac{v}{v-2}}\frac{U_j}{\sigma_j}, \sqrt{\frac{v}{v-2}}\epsilon\right\} \sim t_{[v]}\left(\bullet, \bullet, \rho_{j\epsilon}\right), \ j = 1, 0.$$

To derive bounds on the distribution of Δ in this case, we make use of the fact that $U_j|\epsilon$ follows the univariate Student's t distribution with degrees of freedom $v + 1$, mean $\sigma_j\rho_{j\epsilon}\epsilon$, and variance $\sigma_j^2(1 - \rho_{j\epsilon}^2)\left(\frac{(v-2)+\epsilon^2}{v-1}\right)$, $j = 1, 0$. Example 4.2 provides bounds on the distribution of Δ given ϵ, i.e., expressions for

$$\sup_u \max\left\{F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - ATE\}), 0\right\}$$

and

$$\inf_u \min\left\{1 - F_{0|\epsilon}(u - \{\delta - ATE\}) + F_{1|\epsilon}(u), 1\right\}$$

in Theorem 4.3 (i). Taking their expectations with respect to ϵ leads to the bounds on $F_\Delta(\delta)$.

THEOREM 5.3 *In a semiparametric SRM with bivariate Student's t distributions for* $\{U_j, \epsilon\}$ *for* $j = 0, 1$*, we have:*

(i) *Suppose* $\sigma_1\sqrt{1-\rho_{1\epsilon}^2} = \sigma_0\sqrt{1-\rho_{0\epsilon}^2} \equiv \sigma$ *and* $\rho_{j\epsilon}^2 \neq 1$*. Let* $\overline{\sigma} = \sigma\sqrt{\left(\frac{(v-2)+\epsilon^2}{v-1}\right)}$*. Then*

$$F_\Delta^L(\delta) = 2\int_A T_{[v+1]}\left(\left(\frac{\delta - ATE - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon}{2\overline{\sigma}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon - P(A),$$

$$F_\Delta^U(\delta) = 2\int_{A^C} T_{[v+1]}\left(\left(\frac{\delta - ATE - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon}{2\overline{\sigma}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon + P(A),$$

where $A = \{\epsilon : \{\delta - ATE\} \geq (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon\}$ *and* A^C *is the complement of* A*.*

(ii) *Suppose* $\sigma_1\sqrt{1-\rho_{1\epsilon}^2} \neq \sigma_0\sqrt{1-\rho_{0\epsilon}^2}$ *and* $\rho_{j\epsilon}^2 \neq 1$*. Let* $\overline{\sigma}_1 = \sigma_1\sqrt{\left(1-\rho_{1\epsilon}^2\right)\left(\frac{(v-2)+\epsilon^2}{v-1}\right)}$ *and* $\overline{\sigma}_0 = \sigma_0\sqrt{\left(1-\rho_{0\epsilon}^2\right)\left(\frac{(v-2)+\epsilon^2}{v-1}\right)}$*. Then*

$$\begin{aligned} F_\Delta^L(\delta) &= \int_{-\infty}^{+\infty} T_{[v+1]}\left(\left(\frac{\overline{\sigma}_1^{2\kappa-1}s - \overline{\sigma}_0^{\kappa-1}\overline{\sigma}_1^{\kappa-1}t}{\overline{\sigma}_1^{2\kappa} - \overline{\sigma}_0^{2\kappa}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon \\ &+ \int_{-\infty}^{+\infty} T_{[v+1]}\left(\left(\frac{\overline{\sigma}_1^{\kappa}\overline{\sigma}_0^{\kappa-1}t - \overline{\sigma}_0^{2\kappa-1}s}{\overline{\sigma}_1^{2\kappa} - \overline{\sigma}_0^{2\kappa}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon - 1, \end{aligned}$$

$$\begin{aligned} F_\Delta^U(\delta) &= \int_{-\infty}^{+\infty} T_{[v+1]}\left(\left(\frac{\overline{\sigma}_1^{2\kappa-1}s + \overline{\sigma}_0^{\kappa}\overline{\sigma}_1^{\kappa-1}t}{\overline{\sigma}_1^{2\kappa} - \overline{\sigma}_0^{2\kappa}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon \\ &- \int_{-\infty}^{+\infty} T_{[v+1]}\left(\left(\frac{\overline{\sigma}_1^{\kappa}\overline{\sigma}_0^{\kappa-1}t + \overline{\sigma}_0^{2\kappa-1}s}{\overline{\sigma}_1^{2\kappa} - \overline{\sigma}_0^{2\kappa}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon + 1, \end{aligned}$$

where

$$s = \{(\delta - ATE) - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon\},\ \kappa = \frac{v+1}{v+2},$$

and

$$t = \left(s^2 + \left(\overline{\sigma}_1^{2\kappa} - \overline{\sigma}_0^{2\kappa}\right)\left(\left(\overline{\sigma}_1^{2(1-\kappa)} - \overline{\sigma}_0^{2(1-\kappa)}\right)(v-1)\right)\right)^{\frac{1}{2}}.$$

We evaluated these bounds for the same set of parameters used in the normal case for $v = 4$, see Figure 3. The same general qualitative conclusions hold as in the normal case. Comparing Figures 1 and 3, we observe that the degree of freedom parameter has little effect on the bounds at the ATE, but it has large effect on the bounds away from ATE. This is due to the fact that Student's t distribution has fatter tails than the normal distribution.

5.3 Bounds on $F_\Delta(\cdot|D=1)$ and the Propensity Score

In a SRM, the propensity score is given by

$$P(D=1|W,X_c) = P(\epsilon > -(W,X_c)'\gamma) = 1 - F_\epsilon(-(W,X_c)'\gamma).$$

Hence the smaller the value of $(W,X_c)'\gamma$, the less likely the individual with the value of $(W,X_c)'\gamma$ will participate in the program or the smaller the propensity score. Since there is a one-to-one relation between the propensity score and $(W,X_c)'\gamma$, we can group individuals in the population via their propensity score. For a given value of the propensity score, Theorem 4.3 (ii) provides sharp bounds on the distribution of Δ for participants with the given propensity score in semiparametric SRMs. Figure 4 depicts the distribution bounds for Δ for participants with $(W,X_c)'\gamma = -1.28$ or propensity score 0.1. Figures 4(a) and 4(b) are based on the normal assumption, while Figures 4(c) and 4(d) are based on the Student's t assumption with degree of freedom 4. We observe that the distribution bounds in Student's t case are generally wider than those in normal case. Moreover, plots with different values of the propensity score and/or the degree of freedom in the Student's t case reveal that the degree of skewness of each bound increases as the propensity score decreases and the bounds get tighter as the degree of freedom increases.

One important and potentially useful application of the distribution bounds established in Theorem 4.3 (ii) is to predict or bound the probability that an individual with a given propensity score will benefit from participating in the program in terms of Δ. Note that

$$F_\Delta^L(0|D=1) \leq P(\Delta \leq 0|D=1) \leq F_\Delta^U(0|D=1).$$

Hence $1-F_\Delta^L(0|D=1)$ is the maximum probability that an individual with a given propensity score will benefit from participating in the program and $1-F_\Delta^U(0|D=1)$ is the minimum probability that an individual with a given propensity score will benefit from participating in the program. To see how these probabilities change with respect to the propensity score, we plotted them against the propensity score in SRMs with bivariate normal distributions in Figures 5(b)-10(b). The expressions[4] for $F_\Delta^L(\delta|D=1)$ and $F_\Delta^U(\delta|D=1)$ are derived by using Theorem 4.3 (ii) and a similar argument to Theorem 5.2. Using these expressions, one can show[5] that the bounds $F_\Delta^L(\delta|D=1)$ and $F_\Delta^U(\delta|D=1)$ approach either 0 or 1 as the propensity score approaches zero. As a result, the bounds are informative for individuals with low propensity score and once they participate, with high probability, they either get hurt or benefit from the treatment.

In a SRM with bivariate normal distributions, TT is given by

$$TT = ATE + (\rho_{1\epsilon}\sigma_1 - \rho_{0\epsilon}\sigma_0)\,\lambda\left((W,X_c)'\gamma\right),$$

[4]They are tedious and hence not provided here, but they are available upon request.
[5]The proofs are elementary, but tedious. They are available upon request.

where $\lambda(\cdot)$ is the inverse mills ratio. For a given value of $(W, X_c)' \gamma$ or a given value of the propensity score, TT measures the average treatment effect for the subpopulation of participants with the given propensity score. It is composed of two terms: the first term is the average treatment effect for the population with covariates X_1, X_0, X_c, W and the second term is the effect due to selection on unobservables. Figures 5(a)-10(a) plotted TT and the second term in TT due to unobservables against the propensity score. Also plotted in each graph are the bounds on the median of the distribution $F_\Delta(\cdot|D = 1)$.

In Figures 5 and 6, ATE is zero. In Figure 5, $(\rho_{1\epsilon}, \rho_{0\epsilon}) = (0.5, -0.5)$ and TT is non-negative for all values of the propensity score. However, when the propensity score is greater than 0.54, there is a positive probability that an individual with the given propensity score will get hurt by participating in the program. This probability increases as the value of the propensity score increases. And for all values of the propensity score, there is always a positive probability that an individual with the given propensity score will benefit from participating in the program and this probability decreases as the value of the propensity score increases. Consequently, people with low propensity score would benefit from the program with high probability once they participate. In Figure 6, $(\rho_{1\epsilon}, \rho_{0\epsilon}) = (-0.5, 0.5)$ and TT is non-positive for all values of the propensity score. However, when the propensity score is less than 0.54, there is a positive probability that an individual with the given propensity score will benefit from participating in the program and this probability increases as the value of the propensity score increases. In addition, for all values of the propensity score, there is always a positive probability that an individual with the given propensity score will get hurt from participating in the program and this probability decreases as the value of the propensity score increases. The seemingly reversal roles of the two probabilities in Figures 5 and 6 are due to the reversal of the correlation values. Consider Figure 5 with $(\rho_{1\epsilon}, \rho_{0\epsilon}) = (0.5, -0.5)$. Heuristically, for small values of the propensity score, individuals participating in the program tend to have large selection errors ϵ. Given the positive correlation between Y_1 and ϵ, Y_1 would tend to be large for those participants. By the same token, the negative correlation between Y_0 and ϵ imply small Y_0. As a result, Δ tend to be large for participants with small propensity score. Figures 5 and 6 demonstrate clearly that average treatment effect parameters such as ATE and TT do not provide a complete picture of the effects of treatment when there is selection on unobserved variables, and the distribution bounds we established in this paper provide useful information that are missed by ATE and TT.

Figures 7 and 8 further support the conclusions we drew from Figures 5 and 6. They are similar to Figures 5 and 6 except that $ATE = -0.5$ in Figure 7 and $ATE = 0.5$ in Figure 8. In both figures, TT is positive for some values of the propensity score and negative for other values of the propensity score. The patterns of $\left[1 - F_\Delta^L(0|D = 1)\right]$ and $\left[1 - F_\Delta^U(0|D = 1)\right]$ as functions of the propensity score remain the same as in Figures 5 and 6. It is interesting to observe from Figures 7 and 8 that even when the ATE for the whole population is negative (-0.5) or positive (0.5), some subpopulations (those with the propensity score less than 0.73) will in general benefit or get hurt from the program if they join the program. The proportion of people in each subgroup who will benefit or get hurt from being in the program will also change with the level of ATE.

In Figures 9 and 10, we increased $\rho_{1\epsilon}$ to 0.95. Comparing these figures with Figures 5-8, we see clearly that the distribution bounds get tighter as $\rho_{1\epsilon}$ ($\rho_{0\epsilon}$) gets larger. When the magnitudes of $\rho_{1\epsilon}, \rho_{0\epsilon}$ are the same, the bounds are more informative when $\rho_{1\epsilon}$ and $\rho_{0\epsilon}$ have different signs than when they have the same sign.

Summarizing Figures 5-10, we conclude that the unobserved selection error has a large effect on those with low propensity score. That is, those who are less likely to participate in the program will most likely be affected by the program once they participate in the program. Whether they gain or lose from participating in the program once they participate depends on the sign of $(\rho_{1\epsilon}\sigma_1 - \rho_{0\epsilon}\sigma_0)$.

6 Confidence Sets for $F_\Delta(\delta)$ in Semiparametric SRMs

Given the sharp bounds established in this paper, statistical inference on the joint distribution of potential outcomes and the distribution of treatment effects falls in the currently active research area: inference for partially identified parameters. A complete treatment of this important issue for the general semiparametric SRMs in (13) is beyond the scope of this paper and left for future research. In this section, we demonstrate its feasibility by constructing an asymptotically uniformly valid and non-conservative confidence set for $F_\Delta(\delta)$ in a special class of semiparametric SRMs in which $g_1(x_1, x_c)$, $g_0(x_0, x_c)$, $F_{1\epsilon}$, and $F_{0\epsilon}$ are parametric. We emphasize here that this is a semiparametric SRM, as the joint distribution of U_1, U_0 is completely unspecified.

Theorem 4.2 (i) implies that $F_\Delta^L(\delta) \le F_\Delta(\delta) \le F_\Delta^U(\delta)$, where

$$\begin{aligned}
F_\Delta^L(\delta) &= \int_{-\infty}^{+\infty} \left\{ \sup_u \left[F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - [g_1(x_1,x_c) - g_0(x_0,x_c)]\}) \right] \right\}_+ dF_\epsilon(\epsilon), \\
F_\Delta^U(\delta) &= 1 + \int_{-\infty}^{+\infty} \left\{ \inf_u \left[F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - [g_1(x_1,x_c) - g_0(x_0,x_c)]\}) \right] \right\}_- dF_\epsilon(\epsilon),
\end{aligned}$$

where $(x)_+ = \max(x, 0)$ and $(x)_- = \min(x, 0)$. For notational compactness, we let $\theta^0 = F_\Delta(\delta)$, $\theta_L = F_\Delta^L(\delta)$, and $\theta_U = F_\Delta^U(\delta)$.

Let

$$\tau_\epsilon(u) = F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - [g_1(x_1,x_c) - g_0(x_0,x_c)]\})$$

and θ_τ denote the collection of unknown parameters in $\tau_\epsilon(u)$: $\tau_\epsilon(u) = \tau_\epsilon(u; \theta_\tau)$. Further, let $F_\epsilon(\epsilon) = F_\epsilon(\epsilon; \theta_\epsilon)$. Let $\vartheta_o \in \Theta$ be the vector of all the parameters in the model, i.e., $\vartheta_o = (\theta_\tau', \theta_\epsilon')'$.

Then

$$\theta_L = F_\Delta^L(\delta;\vartheta) = \int_{-\infty}^{+\infty} \left\{ \sup_u \left[\tau_\epsilon \left(u; \theta_\tau \right) \right] \right\}_+ dF_\epsilon(\epsilon;\theta_\epsilon),$$

$$\theta_U = F_\Delta^U(\delta;\vartheta) = 1 + \int_{-\infty}^{+\infty} \left\{ \inf_u \left[\tau_\epsilon \left(u; \theta_\tau \right) \right] \right\}_- dF_\epsilon(\epsilon;\theta_\epsilon).$$

Let $\widehat{\theta}_\tau$ and $\widehat{\theta}_\epsilon$ denote consistent estimators of θ_τ and θ_ϵ respectively and $\widehat{\vartheta} = \left(\widehat{\theta}'_\tau, \widehat{\theta}'_\epsilon\right)'$. Examples of $\widehat{\vartheta}$ include the maximum likelihood estimator and the two-step estimator. The plug-in estimators of θ_L and θ_U are given by

$$\widehat{\theta}_L = \widehat{F}_\Delta^L(\delta) = \int_{-\infty}^{+\infty} \left\{ \sup_u \left[\tau_\epsilon \left(u; \widehat{\theta}_\tau \right) \right] \right\}_+ dF_\epsilon(\epsilon;\widehat{\theta}_\epsilon),$$

$$\widehat{\theta}_U = \widehat{F}_\Delta^U(\delta) = 1 + \int_{-\infty}^{+\infty} \left\{ \inf_u \left[\tau_\epsilon \left(u; \widehat{\theta}_\tau \right) \right] \right\}_- dF_\epsilon(\epsilon;\widehat{\theta}_\epsilon).$$

For $\varepsilon > 0$, let $B_\Theta(\varepsilon) = \{\vartheta \in \Theta : \|\vartheta - \vartheta_o\| < \varepsilon\}$. Let $\mathcal{W}_i = (Y_i, X'_{1i}, X'_{0i}, X'_{ci}, W'_i, D_i)'$ with $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$ and $\mathcal{P}$ denote the collection of all the potential distributions of the random sample $\{\mathcal{W}_i\}_{i=1}^n$ such that Assumption 1 below holds.

Assumption 1. (i) For some $\varepsilon > 0$, there exists a function $\Gamma_P^j(\delta)[\vartheta - \vartheta_o]$ of $(\vartheta - \vartheta_o)$, $\vartheta \in B_\Theta(\varepsilon)$, such that for $j = L, U$,

$$\left| F_\Delta^j(\delta;\vartheta) - F_\Delta^j(\delta;\vartheta_o) - \Gamma_P^j(\delta)[\vartheta - \vartheta_o] \right| \leq M_\delta \|\vartheta - \vartheta_o\|^2 \tag{28}$$

with a constant M_δ that does not depend on P, and for each $\varepsilon > 0$,

$$\limsup_{n \geq 1} \sup_{P \in \mathcal{P}} P \left\{ \left| \sqrt{n} \Gamma_P^j(\delta) \left[\widehat{\vartheta} - \vartheta_o \right] - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\delta^j(\mathcal{W}_i) \right| > \varepsilon \right\} = 0, \tag{29}$$

where $\psi_\delta^j(\mathcal{W}_i)$ satisfies that there exists $\eta > 0$ such that $E\left[\psi_\delta^j(\mathcal{W}_i)\right] = 0$ and $\sup_{P \in \mathcal{P}} \left\|\psi_\delta^j(\mathcal{W}_i)\right\|_{P,2+\eta} < \infty$. (ii) $\left\|\widehat{\vartheta} - \vartheta_o\right\| = o_P\left(n^{-1/4}\right)$ uniformly in $P \in \mathcal{P}$.

The first condition in Assumption 1 (i) is concerned with differentiability of $F_\Delta^j(\delta;\vartheta)$, $j = L, U$, with respect to $\vartheta \in B_\Theta(\varepsilon)$. This holds under mild differentiability conditions on $\tau_\epsilon(u;\theta_\tau)$ with respect to θ_τ and $F_\epsilon(\epsilon;\theta_\epsilon)$ with respect to θ_ϵ. The second condition in Assumption 1 (i) imposes an asymptotic linear representation on $\Gamma_P^j(\delta)\left[\widehat{\vartheta} - \vartheta_o\right]$. This can be established using the asymptotic linear representation of $\left[\widehat{\vartheta} - \vartheta_o\right]$. Assumption 1 (ii) can be established by following the procedure of Theorem 3.2.5 of van der Vaart and Wellner (1996).

Under Assumption 1, one can show that unformly in $P \in \mathcal{P}$, we have

$$\sqrt{n} \begin{pmatrix} \widehat{\theta}_L - \theta_L \\ \widehat{\theta}_U - \theta_U \end{pmatrix} \Longrightarrow N(0, \Omega), \tag{30}$$

where Ω is the variance-covariance matrix of $\left(\psi_\delta^L(\mathcal{W}_i), \psi_\delta^U(\mathcal{W}_i)\right)'$. As a result, CSs developed in Andrews and Guggenberger (2007), Andrews and Soares (2007), Fan and Park (2007), and Stoye (2007), among others, are applicable to θ^0.

We briefly summarize the CS in Fan and Park (2007). Note that $\theta_L \le \theta^0 \le \theta_U$ is equivalent to $\theta^0 = \arg\min_{\theta\in[0,1]}\left\{(\theta_L-\theta)_+^2 + (\theta_U-\theta)_-^2\right\}$. We define the sample criterion function as

$$T_n(\theta^0) = n\left(\widehat{\theta}_L - \theta^0\right)_+^2 + n\left(\widehat{\theta}_U - \theta^0\right)_-^2. \tag{31}$$

Then a $(1-\alpha)$ level CS for θ^0 can be constructed as

$$CS_n = \{\theta \in [0,1] : T_n(\theta) \le c_{1-\alpha}(\theta)\} \tag{32}$$

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

Let $(Z_L, Z_U)' \sim N(0,\Omega)$. It follows from (30) that

$$T_n(\theta) \Longrightarrow \left(Z_L - h^L(\theta)\right)_+^2 + \left(Z_U + h^U(\theta)\right)_-^2$$

where

$$h^L(\theta) = -\lim_{n\to\infty}\sqrt{n}\left[\theta_L - \theta\right] \text{ and } h^U(\theta) = \lim_{n\to\infty}\sqrt{n}\left[\theta_U - \theta\right].$$

Let $\widehat{\Delta}^S = \widehat{\Delta} I\left\{\widehat{\Delta} > b_n\right\}$, where $\widehat{\Delta} \equiv \widehat{\theta}_U - \widehat{\theta}_L$ and b_n is a pre-specified sequence of positive numbers satisfying $b_n \to 0$ and $\sqrt{n}b_n \to \infty$. Then an asymptotically uniformly valid and non-conservative CS for θ_0 is given by CS_n with

$$c_{1-\alpha}(\theta) = \max\left\{cv_{1-\alpha}\left(0, \sqrt{n}\widehat{\Delta}^S, \widehat{\Omega}\right), cv_{1-\alpha}\left(\sqrt{n}\widehat{\Delta}^S, 0, \widehat{\Omega}\right)\right\},$$

where $cv_{1-\alpha}\left(h^L, h^U, \Omega\right)$ is the $1-\alpha$ quantile of the random variable $\left(Z_L - h^L\right)_+^2 + \left(Z_U + h^U\right)_-^2$ and $\widehat{\Omega}$ is a uniformly consistent estimator of Ω.

THEOREM 6.1 *Suppose Assumption 1 holds and* $0 < \alpha < 1/2$. *Then* CS_n *satisfies*

$$\lim_{n\to\infty}\inf_{\theta\in[0,1]}\inf_{P\in\mathcal{P}:\theta_0(P)=\theta}\Pr\left(\theta^0 \in CS_n\right) = 1-\alpha.$$

Remark 6.1. Note that $c_{1-\alpha}(\theta) = c_{1-\alpha}$ does not depend on θ. It follows from Fan and Park (2007) that

$$CS_n = \begin{cases} \left[\hat{\theta}_L - \sqrt{c_{1-\alpha}}\frac{1}{\sqrt{n}},\ \hat{\theta}_U + \sqrt{c_{1-\alpha}}\frac{1}{\sqrt{n}}\right] & \text{if } \sqrt{n}\widehat{\Delta} \ge -\sqrt{c_{1-\alpha}} \\ \left[\widehat{A}, \widehat{B}\right] & \text{if } -\sqrt{2c_{1-\alpha}} \le \sqrt{n}\widehat{\Delta} < -\sqrt{c_{1-\alpha}} \\ \varnothing & \text{if } \sqrt{n}\widehat{\Delta} < -\sqrt{2c_{1-\alpha}} \end{cases}$$

where

$$\widehat{A} \equiv \frac{\hat{\theta}_L + \hat{\theta}_U}{2} - \frac{1}{2\sqrt{n}}\sqrt{c_{1-\alpha} - \frac{n\widehat{\Delta}^2}{2}},\ \widehat{B} \equiv \frac{\hat{\theta}_L + \hat{\theta}_U}{2} + \frac{1}{2\sqrt{n}}\sqrt{c_{1-\alpha} - \frac{n\widehat{\Delta}^2}{2}}.$$

7 Conclusion

In this paper we have established sharp bounds on the joint distribution of potential outcomes and the distribution of treatment effects in GH-SRMs and in semiparametric SRMs of Heckman (1990). The means of the distributions of treatment effects that we consider correspond to the average treatment effect and the treatment effect for the treated. The results we obtain reveal the important role played by self selection, i.e., it helps tighten the bounds on these distributions.

As a first step, this paper has focused on partial identification. Estimation of the distribution bounds developed in this paper is straightforward in view of the identification results in Heckman (1990) and existing work on estimation of parametric/semiparametric sample selection models. Heckman (1990) provides a review of various nonparametric/semiparametric methods for estimating $g_1(x_1, x_c)$ and $g_0(x_0, x_c)$ without specifying the bivariate margins for (U_1, ϵ) and (U_0, ϵ), see also Ai (1997), Andrews and Schafgans (1998), Schafgans and Zinde-Walsh (2002), Das, Newey, and Vella (2003), Chen (2006), and Chen and Zhou (2006). Gallant and Nychka (1987) provide estimators of the unknown marginal distributions $F_{1\epsilon}$ and $F_{0\epsilon}$. It remains to establish a complete set of inference tools for the joint distribution of potential outcomes and the distribution of treatment effects.

Appendix: Technical Proofs

Proof of Theorem 2.1: (i) The 'if part' is obvious. Now we prove the 'only if' part.

First, we consider the case that $Var(S) > 0$. Let $a_j = \sqrt{E(S)\sigma_j^2}$, $b_j = sign(\zeta_j)\sqrt{\zeta_j^2 Var(S)}$, $j = 1, 0, \epsilon$. Then for any ρ, we have

$$\begin{aligned} & \left(a_1^2 + b_1^2\right)\left(a_0^2 + b_0^2\right) - \left(a_1 a_0 \rho + b_1 b_0\right)^2 \\ = & \left(1 - \rho^2\right)\left(a_1^2 a_0^2 + a_1^2 b_0^2\right) + \left(\rho a_1 b_0 - a_0 b_1\right)^2. \end{aligned} \tag{A.1}$$

Note that

$$\begin{aligned} \bar{\rho}_L &= \frac{E(S)\rho_L \sigma_1 \sigma_0 + \zeta_1 \zeta_0 Var(S)}{\sqrt{\left(E(S)\sigma_1^2 + \zeta_1^2 Var(S)\right)\left(E(S)\sigma_0^2 + \zeta_0^2 Var(S)\right)}} \\ &= \frac{a_1 a_0 \rho_L + b_1 b_0}{\sqrt{\left(a_1^2 + b_1^2\right)\left(a_0^2 + b_0^2\right)}}. \end{aligned}$$

If $\bar{\rho}_L = -1$, then the left hand and the right hand side expressions in (A.1) with ρ replaced by ρ_L equal to 0, which implies that $\rho_L^2 = 1$ and $\rho_L a_1 b_0 - a_0 b_1 = 0$. Similarly, since

$$\begin{aligned} \bar{\rho}_U &= \frac{E(S)\rho_U \sigma_1 \sigma_0 + \zeta_1 \zeta_0 Var(S)}{\sqrt{\left(E(S)\sigma_1^2 + \zeta_1^2 Var(S)\right)\left(E(S)\sigma_0^2 + \zeta_0^2 Var(S)\right)}} \\ &= \frac{a_1 a_0 \rho_U + b_1 b_0}{\sqrt{\left(a_1^2 + b_1^2\right)\left(a_0^2 + b_0^2\right)}}, \end{aligned}$$

$\bar{\rho}_U = 1$ implies that $\rho_U^2 = 1$ and $\rho_U a_1 b_0 - a_0 b_1 = 0$. Since $a_j > 0$ and $Var(S) > 0$, we must have $(\rho_L - \rho_U) b_0 = 0$ which implies that either $\rho_L = \rho_U$ or $b_0 = 0$. Since $\rho_L = \rho_U$ implies $\overline{\rho}_L = \overline{\rho}_U$ violating the condition, we must have $b_0 = 0$ implying $\zeta_0 = 0$**.** Now since $\rho_L \neq \rho_U$, it follows from $\rho_L^2 = \rho_U^2 = 1$ that $\rho_L = -1$ and $\rho_U = 1$. This in turn implies that $\rho_L = \bar{\rho}_L = -1$ and $\rho_U = \bar{\rho}_U = 1$. $\zeta_1 = 0$ follows from $a_1 b_0 + a_0 b_1 = 0$ and $a_1 b_0 - a_0 b_1 = 0$. From the expressions for ρ_L and ρ_U, it follows immediately that $\rho_{1\epsilon} = \rho_{0\epsilon} = 0$. Since

$$\bar{\rho}_{j\epsilon} = \frac{E(S)\rho_{j\epsilon}\sigma_1 \sigma_0 + \zeta_1 \zeta_0 Var(S)}{\sqrt{\left(E(S)\sigma_1^2 + \zeta_1^2 Var(S)\right)\left(E(S)\sigma_0^2 + \zeta_0^2 Var(S)\right)}},\ j = 1, 0,$$

with $\zeta_1 = \zeta_0 = 0$, we get $\bar{\rho}_{j\epsilon} = \rho_{j\epsilon} = 0$**,** $j = 1, 0$.

It remains to consider the case that $Var(S) = 0$. In this case, $\bar{\rho}_L = \rho_L$ and $\bar{\rho}_U = \rho_U$. It follows that $\rho_{1\epsilon} = \rho_{0\epsilon} = 0$. But when $Var(S) = 0$, $\bar{\rho}_{j\epsilon} = \rho_{j\epsilon} = 0$.

(ii) and (iii) Since the positive semi-definiteness of Σ implies the positive semi-definiteness of $E(S)\Sigma + Var(S)[\zeta_1, \zeta_0, \zeta_\epsilon]'[\zeta_1, \zeta_0, \zeta_\epsilon]$, which is the variance-covariance matrix of $(U_1, U_0, \epsilon)'$. The bounds $[\bar{\rho}_L, \bar{\rho}_U]$ implied by the positive semi-definiteness of Σ are in general tighter than the bounds implied by that of the variance-covariance matrix of $(U_1, U_0, \epsilon)'$. So we have:

$$\bar{\rho}_L \geq \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} - \sqrt{\left(1 - \bar{\rho}_{1\epsilon}^2\right)\left(1 - \bar{\rho}_{0\epsilon}^2\right)} \text{ and } \bar{\rho}_U \leq \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} - \sqrt{\left(1 - \bar{\rho}_{1\epsilon}^2\right)\left(1 - \bar{\rho}_{0\epsilon}^2\right)}.$$

When $\bar{\rho}_{1\epsilon}^2 + \bar{\rho}_{0\epsilon}^2 > 1$ and $\bar{\rho}_{1\epsilon}$, $\bar{\rho}_{0\epsilon}$ have the same sign, we have

$$\begin{aligned}\bar{\rho}_L &\geq \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} - \sqrt{\left(1-\bar{\rho}_{1\epsilon}^2\right)\left(1-\bar{\rho}_{0\epsilon}^2\right)} \\ &= \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} - \sqrt{1-\left(\bar{\rho}_{1\epsilon}^2+\bar{\rho}_{0\epsilon}^2\right)+\bar{\rho}_{1\epsilon}^2\bar{\rho}_{0\epsilon}^2} \\ &> \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} - \sqrt{\bar{\rho}_{1\epsilon}^2\bar{\rho}_{0\epsilon}^2} \\ &= 0.\end{aligned}$$

When $\bar{\rho}_{1\epsilon}^2 + \bar{\rho}_{0\epsilon}^2 > 1$ and $\bar{\rho}_{1\epsilon}$, $\bar{\rho}_{0\epsilon}$ have the opposite sign, we have

$$\begin{aligned}\bar{\rho}_U &\leq \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} + \sqrt{\left(1-\bar{\rho}_{1\epsilon}^2\right)\left(1-\bar{\rho}_{0\epsilon}^2\right)} \\ &= \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} + \sqrt{1-\left(\bar{\rho}_{1\epsilon}^2+\bar{\rho}_{0\epsilon}^2\right)+\bar{\rho}_{1\epsilon}^2\bar{\rho}_{0\epsilon}^2} \\ &< \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} + \sqrt{\bar{\rho}_{1\epsilon}^2\bar{\rho}_{0\epsilon}^2} \\ &= 0.\end{aligned}$$

(iv) Following the notation in **(i)**, we have

$$\begin{aligned}\bar{\rho}_{j\epsilon} &= \frac{E(S)\rho_{j\epsilon}\sigma_1\sigma_\epsilon + \zeta_j\zeta_\epsilon Var(S)}{\sqrt{\left(E(S)\sigma_j^2+\zeta_j^2 Var(S)\right)\left(E(S)\sigma_\epsilon^2+\zeta_\epsilon^2 Var(S)\right)}} \\ &= \frac{a_j a_\epsilon \rho_{j\epsilon} + b_j b_\epsilon}{\sqrt{\left(a_j^2+b_j^2\right)\left(a_\epsilon^2+b_\epsilon^2\right)}},\ j=0,1.\end{aligned}$$

Similar to **(i)**, we conclude that $\bar{\rho}_{j\epsilon}^2 = 1$ implies $\rho_{j\epsilon}^2 = 1$. If $\rho_{1\epsilon}^2 = 1$ or $\rho_{0\epsilon}^2 = 1$, we have $\rho_L = \rho_U$ which leads to $\bar{\rho}_L = \bar{\rho}_U$ and the result that $\bar{\rho}_{10}$ is point identified.

Proof of Theorem 3.1: When $\zeta_1 = \zeta_0 = 0$ and $\bar{\rho}_{1\epsilon} = \bar{\rho}_{0\epsilon} = 0$, Theorem 2.1 implies that $\rho_L = -1$ and $\rho_U = 1$. So we have:

$$\begin{aligned}&GH_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \left[\alpha_{10}^-, \rho_L\right]\right) \\ &= \int_0^\infty f_S(s)\, C^{Gau}\left(\Phi\left(\frac{y_1-x'\beta_1}{\sigma_1\sqrt{s}}\right), \Phi\left(\frac{y_0-x'\beta_0}{\sigma_0\sqrt{s}}\right), -1\right) ds \\ &= \int_0^\infty f_S(s) \max\left\{\Phi\left(\frac{y_1-x'\beta_1}{\sigma_1\sqrt{s}}\right) + \Phi\left(\frac{y_0-x'\beta_0}{\sigma_0\sqrt{s}}\right) - 1, 0\right\} ds,\end{aligned}$$

where the second equality follows from the fact that $C^{Gau}(\cdot,\cdot,-1) = C_L(\cdot,\cdot)$, see e.g., Joe (1997). When $\frac{y_1-x'\beta_1}{\sigma_1} \geq (\leq) - \frac{y_0-x'\beta_0}{\sigma_0}$, we have $\frac{y_1-x'\beta_1}{\sigma_1\sqrt{s}} \geq (\leq) - \frac{y_0-x'\beta_0}{\sigma_1\sqrt{s}}$ and

$$\Phi\left(\frac{y_1-x'\beta_1}{\sigma_1\sqrt{s}}\right) + \Phi\left(\frac{y_0-x'\beta_0}{\sigma_0\sqrt{s}}\right) - 1 = \Phi\left(\frac{y_1-x'\beta_1}{\sigma_1\sqrt{s}}\right) - \Phi\left(-\frac{y_0-x'\beta_0}{\sigma_0\sqrt{s}}\right) \geq (\leq)\, 0.$$

Thus, the sign of $\left[\Phi\left(\frac{y_1-x'\beta_1}{\sigma_1\sqrt{s}}\right)+\Phi\left(\frac{y_0-x'\beta_0}{\sigma_0\sqrt{s}}\right)-1\right]$ does not depend on s. We obtain:

$$
\begin{aligned}
& GH_2\left(y_1-x'\beta_1, y_0-x'\beta_0; \left[\alpha_{10}^{-}, \rho_L\right]\right) \\
= & \int_0^{\infty} f_S(s) \max\left\{\Phi\left(\frac{y_1-x'\beta_1}{\sigma_1\sqrt{s}}\right)+\Phi\left(\frac{y_0-x'\beta_0}{\sigma_0\sqrt{s}}\right)-1, 0\right\} ds \\
= & \max\left\{\int_0^{\infty} f_S(s)\left\{\Phi\left(\frac{y_1-x'\beta_1}{\sigma_1\sqrt{s}}\right)+\Phi\left(\frac{y_0-x'\beta_0}{\sigma_0\sqrt{s}}\right)-1\right\} ds, 0\right\} \\
= & \max\left\{GH\left(y_1-x'\beta_1;\theta_1\right)+GH\left(y_0-x'\beta_0;\theta_0\right)-1, 0\right\} \\
= & C_L\left(GH\left(y_1-x'\beta_1;\theta_1\right), GH\left(y_0-x'\beta_0;\theta_0\right)\right),
\end{aligned}
$$

where $\theta_1=\left(\lambda,\chi,\psi,\mu_1,\sigma_1^2,\zeta_1\right)$ and $\theta_0=\left(\lambda,\chi,\psi,\mu_0,\sigma_0^2,\zeta_0\right)$. Similarly, we can show that

$$
GH_2\left(y_1-x'\beta_1, y_0-x'\beta_0; \left[\alpha_{10}^{-}, \rho_U\right]\right) = C_U\left(GH\left(y_1-x'\beta_1;\theta_1\right), GH\left(y_0-x'\beta_0;\theta_0\right)\right).
$$

Proof of Theorem 3.2: It follows from (11) that

$$
F_{\Delta}(\delta) = GH_1\left(\delta - x'\left(\beta_1-\beta_0\right); \lambda,\chi,\psi,\mu_1-\mu_0,\gamma_2^2,\zeta_1-\zeta_0\right), \tag{A.2}
$$

$$
\begin{aligned}
& F_{\Delta}(\delta|D=1) \\
= & \int_{-\infty}^{\delta}\int_{-w'\gamma}^{\infty} \frac{gh_2\left(u-x'\left(\beta_1-\beta_0\right),\epsilon;\lambda,\chi,\psi,(B\mu)_{1\epsilon},(B\Sigma B')_{1\epsilon},(B\zeta)_{1\epsilon}\right)}{1-GH_1\left(-w'\gamma;\lambda,\chi,\psi,\mu_\epsilon,\sigma_\epsilon^2,\zeta_\epsilon\right)} du d\epsilon.
\end{aligned}
$$

Since the only unidentified parameter in $F_{\Delta}(\delta)$ and $F_{\Delta}(\delta|D=1)$ is γ_2^2, we get

$$
\begin{aligned}
F_{\Delta}^{L}(\delta) = & \min_{\sigma_L^2\leq\gamma_2^2\leq\sigma_U^2} GH_1\left(\delta-x'\left(\beta_1-\beta_0\right);\lambda,\chi,\psi,\mu_1-\mu_0,\gamma_2^2,\zeta_1-\zeta_0\right) \leq F_{\Delta}(\delta) \\
\leq & \max_{\sigma_L^2\leq\gamma_2^2\leq\sigma_U^2} GH_1\left(\delta-x'\left(\beta_1-\beta_0\right);\lambda,\chi,\psi,\mu_1-\mu_0,\gamma_2^2,\zeta_1-\zeta_0\right) = F_{\Delta}^{U}(\delta)
\end{aligned}
$$

and

$$
\begin{aligned}
& F_{\Delta}^{L}(\delta|D=1) \\
= & \min_{\sigma_L^2\leq\gamma_2^2\leq\sigma_U^2} \int_{-\infty}^{\delta}\int_{-w'\gamma}^{\infty} \frac{gh_2\left(u-x'\left(\beta_1-\beta_0\right),\epsilon;\lambda,\chi,\psi,(B\mu)_{1\epsilon},(B\Sigma B')_{1\epsilon},(B\zeta)_{1\epsilon}\right)}{1-GH_1\left(-w'\gamma;\lambda,\chi,\psi,\mu_\epsilon,\sigma_\epsilon^2,\zeta_\epsilon\right)} du d\epsilon \\
\leq & F_{\Delta}(\delta|D=1) \\
\leq & \max_{\sigma_L^2\leq\gamma_2^2\leq\sigma_U^2} \int_{-\infty}^{\delta}\int_{-w'\gamma}^{\infty} \frac{gh_2\left(u-x'\left(\beta_1-\beta_0\right),\epsilon;\lambda,\chi,\psi,(B\mu)_{1\epsilon},(B\Sigma B')_{1\epsilon},(B\zeta)_{1\epsilon}\right)}{1-GH_1\left(-w'\gamma;\lambda,\chi,\psi,\mu_\epsilon,\sigma_\epsilon^2,\zeta_\epsilon\right)} du d\epsilon \\
= & F_{\Delta}^{U}(\delta|D=1).
\end{aligned}
$$

If $\zeta_1=\zeta_0$, then

$$
F_{\Delta}(\delta) = GH_1\left(\delta-x'\left(\beta_1-\beta_0\right);\lambda,\chi,\psi,\mu_1-\mu_0,\gamma_2^2,0\right) = \int_0^{\infty} f_S(s)\,\Phi\left(\frac{\delta-x'\left(\beta_1-\beta_0\right)}{\gamma_2\sqrt{s}}\right) ds.
$$

Thus, $F_\Delta(\delta)$ is an increasing function of γ_2 when $\delta < ATE$ $(= x'(\beta_1 - \beta_0))$ and a decreasing function of γ_2 when $\delta \geq ATE$. We get the second result in (i).

Proof of Theorem 5.1: Since $(U_j, \epsilon) \sim GH_2(\lambda, \chi, \psi, \mu_{j\epsilon}, \Sigma_{j\epsilon}, \zeta_{j\epsilon})$, it follows that the conditional distribution function of U_j given ϵ is given by

$$F_{j|\epsilon}(u_j|\epsilon) = \frac{\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}}\phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right)\Phi\left(\frac{u_j-\bar{\mu}_j(s)}{\sqrt{(1-\rho_{j\epsilon}^2)\sigma_j^2 s}}\right)ds}{f_\epsilon(\epsilon)},\ j=0,1,$$

where

$$\begin{aligned}\mu_j(s) &= \zeta_j(s - E(S)),\ j = 0,1,\epsilon,\\ \bar{\mu}_j(s) &= \mu_j(s) + \rho_{j\epsilon}\frac{\sigma_j}{\sigma_\epsilon}\epsilon,\ j=0,1.\end{aligned}$$

Let

$$m_1(s) = \frac{y_1 - g_1(x_1,x_c) - \bar{\mu}_1(s)}{\sqrt{\left(1-\rho_{1\epsilon}^2\right)\sigma_1^2 s}} \text{ and } m_0(s) = \frac{y_0 - g_0(x_0,x_c) - \bar{\mu}_0(s)}{\sqrt{\left(1-\rho_{0\epsilon}^2\right)\sigma_0^2 s}}.$$

If $\sigma_0\sqrt{(1-\rho_{0\epsilon}^2)}\zeta_1 = -\sigma_1\sqrt{(1-\rho_{1\epsilon}^2)}\zeta_0$, then

$$\begin{aligned}&m_1(s) + m_0(s)\\ =\ &\frac{\sigma_0\sqrt{\left(1-\rho_{0\epsilon}^2\right)}\left(y_1 - g_1(x_1,x_c) - \rho_{1\epsilon}\frac{\sigma_1}{\sigma_\epsilon}\epsilon\right) + \sigma_1\sqrt{\left(1-\rho_{1\epsilon}^2\right)}\left(y_0 - g_0(x_0,x_c) - \rho_{0\epsilon}\frac{\sigma_0}{\sigma_\epsilon}\epsilon\right)}{\sqrt{\left(1-\rho_{1\epsilon}^2\right)\left(1-\rho_{0\epsilon}^2\right)\sigma_1^2\sigma_0^2 s}},\end{aligned}$$

which will not change sign with respect to s, so $\Phi(m_1(s)) + \Phi(m_0(s)) - 1$ $(= \Phi(m_1(s)) + \Phi(-m_0(s)))$ will not change sign with respect to s. Thus, we obtain

$$\begin{aligned}&C_L\left(F_{1|\epsilon}(y_1 - g_1(x_1,x_c)), F_{0|\epsilon}(y_0 - g_0(x_0,x_c))\right)\\ =\ &\max\left\{\frac{\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}}\phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right)\{\Phi(m_1(s)) + \Phi(m_0(s)) - 1\}ds}{f_\epsilon(\epsilon)}, 0\right\}\\ =\ &\frac{\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}}\phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right)\max\{\Phi(m_1(s)) + \Phi(m_0(s)) - 1, 0\}ds}{f_\epsilon(\epsilon)}\\ =\ &\frac{\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}}\phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right)\Phi(m_1(s), m_0(s), -1)ds}{f_\epsilon(\epsilon)}.\end{aligned} \tag{A.3}$$

Similarly, if $\sigma_0\sqrt{(1-\rho_{0\epsilon}^2)}\zeta_1 = \sigma_1\sqrt{(1-\rho_{1\epsilon}^2)}\zeta_0$, we have

$$\begin{aligned}&C_U\left(F_{1|\epsilon}(y_1 - g_1(x_1,x_c)), F_{0|\epsilon}(y_0 - g_0(x_0,x_c))\right)\\ =\ &\frac{\int \frac{f_S(s)}{(\sigma_\epsilon^2 s)^{1/2}}\phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right)\Phi(m_1(s), m_0(s), 1)ds}{f_\epsilon(\epsilon)}.\end{aligned} \tag{A.4}$$

Now suppose $Z = (Z_1, Z_0, Z_\epsilon)' \sim GH_3(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$. Then the density function of Z_ϵ is the same as that of ϵ and the conditional density function of Z_1, Z_0 given $Z_\epsilon = \epsilon$ is:

$$
\begin{aligned}
f_{Z_1,Z_0|Z_\epsilon=\epsilon}\left(z_1, z_0; \rho_{10|\epsilon}\right) &= \frac{\int \frac{f_S(s)}{(2\pi s)^{3/2}|\Sigma|^{1/2}} \exp\left\{-\left(z-\mu(s)\right)' \Sigma^{-1}\left(z-\mu(s)\right)/(2s)\right\} ds}{f_\epsilon(\epsilon)} \\
&= \frac{\int \frac{f_S(s)}{(s)^{3/2}\sigma_\epsilon|\Sigma_2|^{1/2}} \phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right) \phi\left(\frac{z_1-\bar{\mu}_1(s)}{\sqrt{(1-\rho_{1\epsilon}^2)\sigma_1^2 s}}, \frac{z_0-\bar{\mu}_0(s)}{\sqrt{(1-\rho_{0\epsilon}^2)\sigma_0^2 s}}; \rho_{10|\epsilon}\right) ds}{f_\epsilon(\epsilon)},
\end{aligned}
$$

where

$$
\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{10}\sigma_1\sigma_0 & \rho_{1\epsilon}\sigma_1\sigma_\epsilon \\ \rho_{10}\sigma_1\sigma_0 & \sigma_0^2 & \rho_{0\epsilon}\sigma_0\sigma_\epsilon \\ \rho_{1\epsilon}\sigma_1\sigma_\epsilon & \rho_{0\epsilon}\sigma_0\sigma_\epsilon & \sigma_\epsilon^2 \end{pmatrix}, \; \Sigma_2 = \begin{pmatrix} \left(1-\rho_{1\epsilon}^2\right)\sigma_1^2 & \left(\rho_{10}-\rho_{1\epsilon}\rho_{0\epsilon}\right)\sigma_1\sigma_0 \\ \left(\rho_{10}-\rho_{1\epsilon}\rho_{0\epsilon}\right)\sigma_1\sigma_0 & \left(1-\rho_{0\epsilon}^2\right)\sigma_0^2 \end{pmatrix},
$$

$$
\rho_{10|\epsilon} = \frac{\rho_{10}-\rho_{1\epsilon}\rho_{0\epsilon}}{\sqrt{\left(1-\rho_{1\epsilon}^2\right)\left(1-\rho_{0\epsilon}^2\right)}}, \; \mu(s) = \left(\mu_1(s), \mu_0(s), \mu_\epsilon(s)\right)'.
$$

Thus the conditional distribution function of Z_1, Z_0 given $Z_\epsilon = \epsilon$ is

$$
F_{Z_1,Z_0|Z_\epsilon=\epsilon}\left(z_1, z_0; \rho_{10|\epsilon}\right) = \frac{\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}} \phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right) \Phi\left(\frac{z_1-\bar{\mu}_1(s)}{\sqrt{(1-\rho_{1\epsilon}^2)\sigma_1^2 s}}, \frac{z_0-\bar{\mu}_0(s)}{\sqrt{(1-\rho_{0\epsilon}^2)\sigma_0^2 s}}; \rho_{10|\epsilon}\right) ds}{f_\epsilon(\epsilon)}. \tag{A.5}
$$

Comparing $(A.3)$ and $(A.4)$ with $(A.5)$, we obtain: if $\sigma_0\sqrt{\left(1-\rho_{0\epsilon}^2\right)}\zeta_1 = -\sigma_1\sqrt{\left(1-\rho_{1\epsilon}^2\right)}\zeta_0$, then

$$
\begin{aligned}
F_{10}^L(y_1, y_0) &= \int C_L\left(F_{1|\epsilon}\left(y_1 - g_1(x_1, x_c)\right), F_{0|\epsilon}\left(y_0 - g_0(x_0, x_c)\right)\right) dF_\epsilon(\epsilon) \\
&= \int F_{Z_1,Z_0|Z_\epsilon=\epsilon}\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); -1\right) dF_\epsilon(\epsilon) \\
&= F_{Z_1,Z_0}\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \rho_L\right) \\
&= GH_2\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \left[\alpha_{10}^-, \rho_L\right]\right),
\end{aligned}
$$

where the last equality follows, as $(Z_1, Z_0)' \sim GH_2\left(\left[\alpha_{10}^-, \rho_{10}\right]\right)$; If $\sigma_0\sqrt{\left(1-\rho_{0\epsilon}^2\right)}\zeta_1 = \sigma_1\sqrt{\left(1-\rho_{1\epsilon}^2\right)}\zeta_0$, then

$$
\begin{aligned}
F_{10}^U(y_1, y_0) &= \int C_U\left(F_{1|\epsilon}\left(y_1 - g_1(x_1, x_c)\right), F_{0|\epsilon}\left(y_0 - g_0(x_0, x_c)\right)\right) dF_\epsilon(\epsilon) \\
&= \int F_{Z_1,Z_0|Z_\epsilon=\epsilon}\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); 1\right) dF_\epsilon(\epsilon) \\
&= F_{Z_1,Z_0}\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \rho_U\right) \\
&= GH_2\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \left[\alpha_{10}^-, \rho_U\right]\right).
\end{aligned}
$$

So when $\zeta_1 = \zeta_0 = 0$, we have

$$
\begin{aligned}
F_{10}^L(y_1, y_0) &= GH_2\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \left[\alpha_{10}^-, \rho_L\right]\right), \\
F_{10}^U(y_1, y_0) &= GH_2\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \left[\alpha_{10}^-, \rho_U\right]\right).
\end{aligned}
$$

Proof of Theorem 6.1: The proof is similar to that of Fan and Park (2007) and thus omitted.

References

[1] Aakvik, A. , J. Heckman, and E. Vytlacil (2005), "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.

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

[3] Ai, C. (1997), "A Semiparametric Maximum Likelihood Estimator," *Econometrica* 65, 933-963.

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

[5] Andrews, D.W.K. and P. Guggenberger (2007), "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," Unpublished Manuscript, Cowles Foundation, Yale University.

[6] Andrews, D. W. K. and M. M. A. Schafgans (1998), "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies* 65, 497-517.

[7] Andrews, D.W.K. and G. Soares (2007), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," Manuscript, Yale University.

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

[9] Bugni, F. A. (2007), "Bootstrap Inference in Partially Identified Models," mimeo, Northwestern University.

[10] Canay, I.A. (2007), "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Manuscript, University of Wisconsin.

[11] Carneiro, P. , K. T. Hansen, and J. Heckman (2003), "Estimating Distributions of Treatment Effects With an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review* 44(2), 361-422.

[12] Chen, S. (2006), "Nonparametric Identification and Estimation of Truncated Regression Models," Working Paper, The Hong Kong University of Science and Technology.

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

[14] Chen, S. and Y. Zhou (2006), "Semiparametric and Nonparametric Estimation of Sample Selection Models Under Symmetry," Working Paper, The Hong Kong University of Science and Technology.

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

[16] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Parameter Set Inference in a Class of Econometric Models," *Econometrica* 75, 1243-1284.

[17] Cunha, F. and J. Heckman (2007), "A New Framework for the Analysis of Inequality," IZA DP No. 2565.

[18] Das, M., W. Newey, and F. Vella (2003), "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies* 70, 33-58.

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

[20] Fan, Y. (2005), "Sharp Correlation Bounds and Their Applications," Mimeo.

[21] Fan, Y. and S. Park (2006), "Sharp Bounds on the Distribution of the Treatment Effect and Their Statistical Inference," Manuscript, Vanderbilt University.

[22] Fan, Y. and S. Park (2007), "Confidence Sets for Some Partially Identified Parameters," Manuscript, Vanderbilt University.

[23] Fan, Y. and S. Park (2008), "Confidence Sets for the Quantile of Treatment Effects," Manuscript, Vanderbilt University.

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

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

[26] Gallant, R. and D. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica* 55, 363-390.

[27] Heckman, J. J. (1990), "Varieties of Selection Bias," *American Economic Review, Papers and Proceedings* 80, 313-318.

[28] Heckman, J. J. and Bo E. Honore (1990), "The Empirical Content of the Roy Model," *Econometrica* 58, 1121-1149.

[29] Heckman, J. J. , J. Smith, and N. Clements (1997), "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *Review of Economic Studies* 64, 487-535.

[30] Heckman, J., J. L. Tobias, and E. Vytlacil (2003), "Simple Estimators for Treatment Parameters in a Latent Variable Framework," *Review of Economics and Statistics* 85, 748-755.

[31] Heckman, J.J. and E. Vytlacil (2005), "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica* 73, 669-738.

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

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

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

[35] Jogensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*, Lecture Notes in Statistics 9. New York–Berlin: Springer-Verlag.

[36] Koop, G. and D. J. Poirier (1997), "Learning About the Across-Regime Correlation in Switching Regression Models," *Journal of Econometrics* 78, 217-227.

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

[38] Li, M., D. J. Poirier, and J. L. Tobias (2004), "Do Dropouts Suffer From Dropping Out? Estimation and Prediction of Outcome Gains in Generalized Selection Models," *Journal of Applied Econometrics* 19, 203-225.

[39] 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.

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

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

[42] McNeil, Frey, and Embrechts (2005), *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton University Press.

[43] Mencia, J. and Sentana, E. (2005), "Estimation and Testing of Dynamic Models with Generalized Hyperbolic Innovations," CEPR Discussion Papers, No. 5177.

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

[45] Poirier, D. J. (1998), "Revising Beliefs in Non-Identified Models," *Econometric Theory* 14, 483-509.

[46] Poirier, D. J. and J. L. Tobias (2003), "On the Predictive Distributions of Outcome Gains in the Presence of an Unidentified Parameter," *Journal of Business & Economic Statistics* 21, 258-268.

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

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

[49] Schafgans, M. M. A. and V. Zinde-Walsh (2002), "On Intercept Estimation in the Sample Selection Model," *Econometric Theory* 18, 40-50.

[50] Stoye, J. (2007), "More on Confidence Intervals for Partially Identified Parameters," Working paper, NYU.

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

[52] Vijverberg, W. P. M. (1993), "Measuring the Unidentified Parameter of the Extended Roy Model of Selectivity," *Journal of Econometrics* 57, 69-89.

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

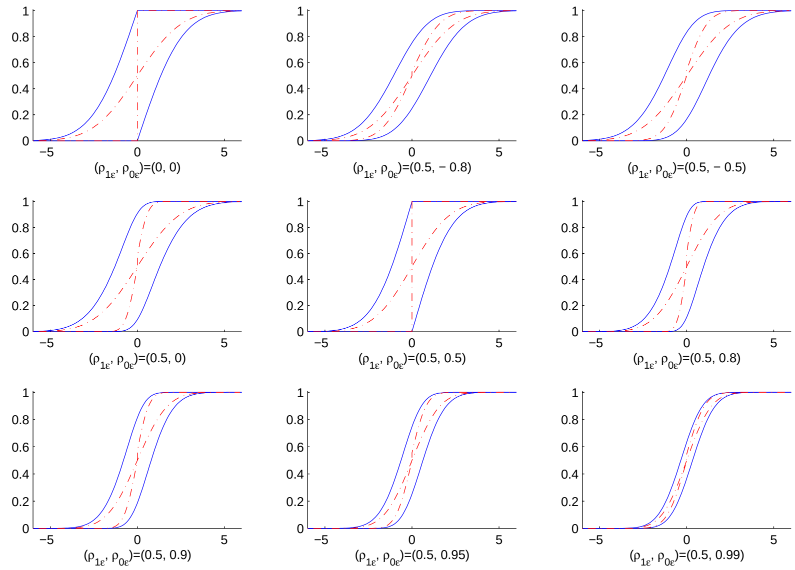

Figure 1: Sharp bounds on the distribution of the treatment effect, $\sigma_1 = \sigma_0 = 1$. Dashed curves are bounds under the trivariate normality assumption for $(U_{1i}, U_{0i}, \epsilon_i)$ and solid curves are bounds assuming bivariate normality for (U_{ji}, ϵ_i), $j = 1, 0$.

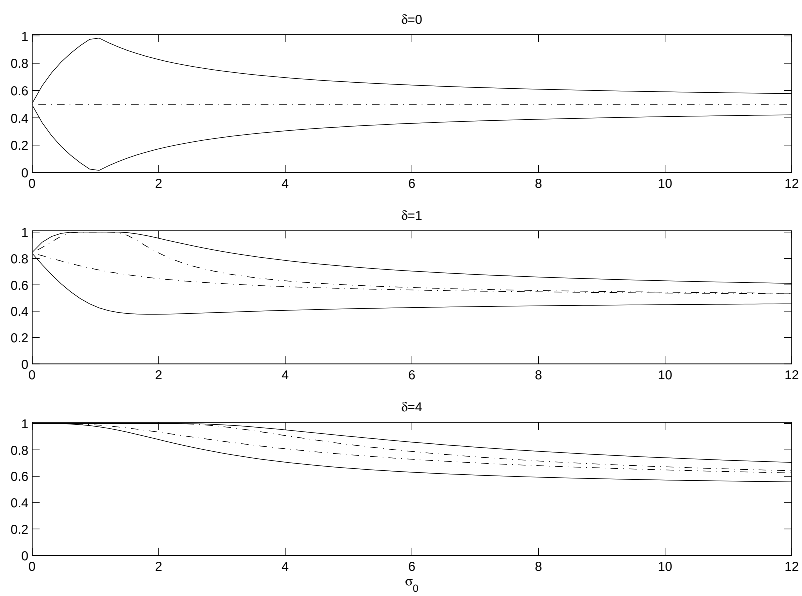

Figure 2: Sharp bounds on the distribution of the treatment effect at a given δ — $\sigma_1 = 1$, $\rho_{1\varepsilon} = 0.5$, and $\rho_{0\varepsilon} = 0.5$. Dashed curves are bounds under the trivariate normality assumption for $(U_{1i}, U_{0i}, \epsilon_i)$ and solid curves are bounds assuming bivariate normality for (U_{ji}, ϵ_i), $j = 1, 0$.

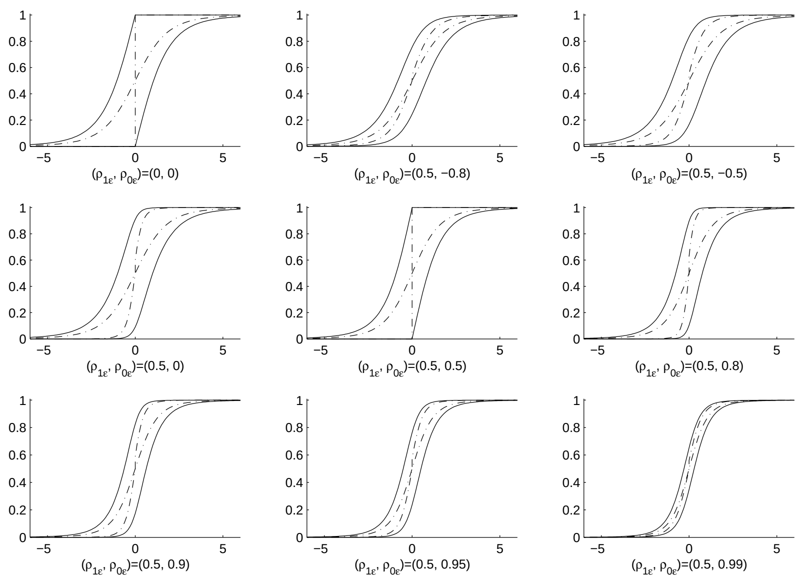

Figure 3: Sharp bounds on the distribution of the treatment effect, $\sigma_1 = \sigma_0 = 1$. Dashed curves are bounds assuming $(U_{1i}, U_{0i}, \epsilon_i)$ follows trivariate Student's t distribution with 4 degrees of freedom and solid curves are bounds assuming (U_{ji}, ϵ_i) follows bivariate Student's t distribution with 4 degrees of freedom.

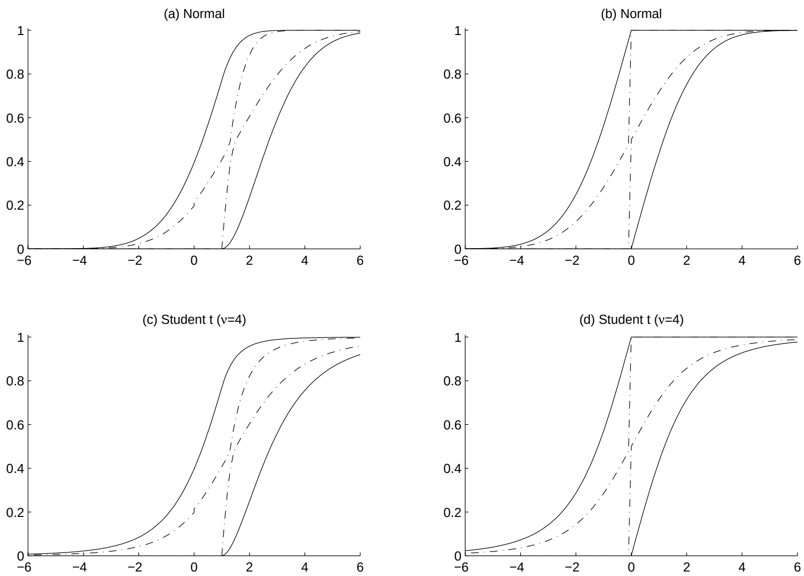

Figure 4: Sharp bounds on the distribution of the treatment effect for the treated — $ATE = 0$, $\sigma_1 = \sigma_0 = 1$, and the Propensity Score = 0.1. In (a) and (c), $\rho_{1\varepsilon} = 0.5$ and $\rho_{0\varepsilon} = -0.5$, while in (b) and (d), $\rho_{1\varepsilon} = \rho_{0\varepsilon} = 0.5$.

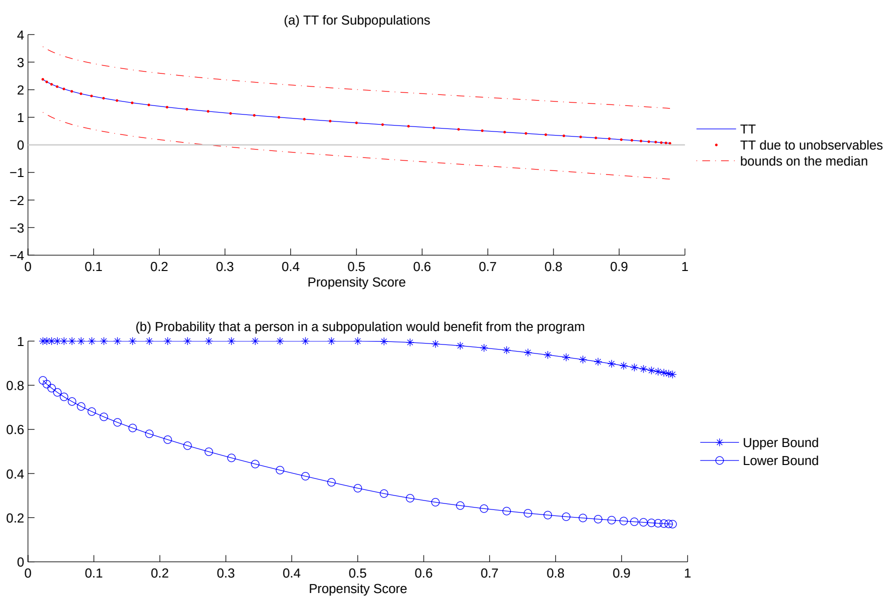

Figure 5: Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = 0.5$, $\rho_{0\varepsilon} = -0.5$, and $\sigma_1 = \sigma_0 = 1$.

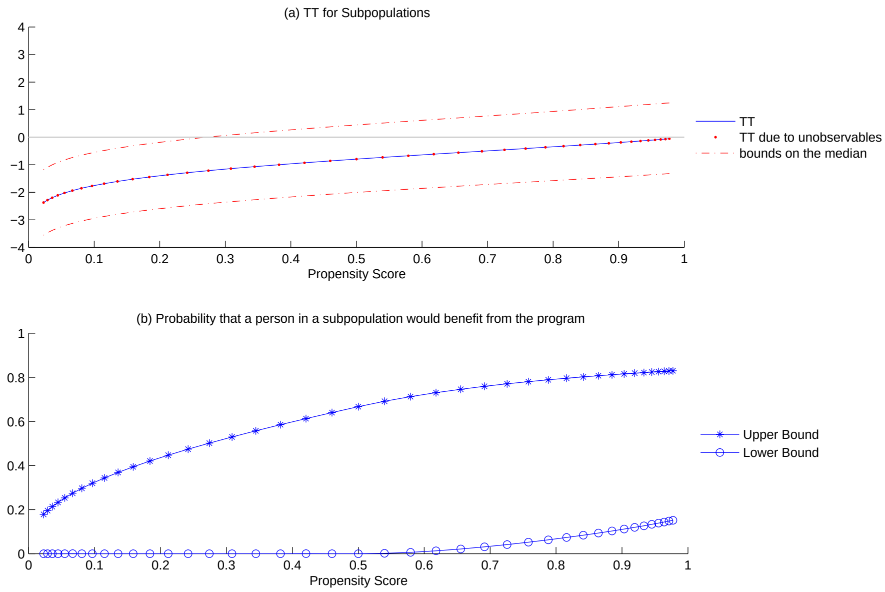

Figure 6: Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = -0.5$, $\rho_{0\varepsilon} = 0.5$, and $\sigma_1 = \sigma_0 = 1$.

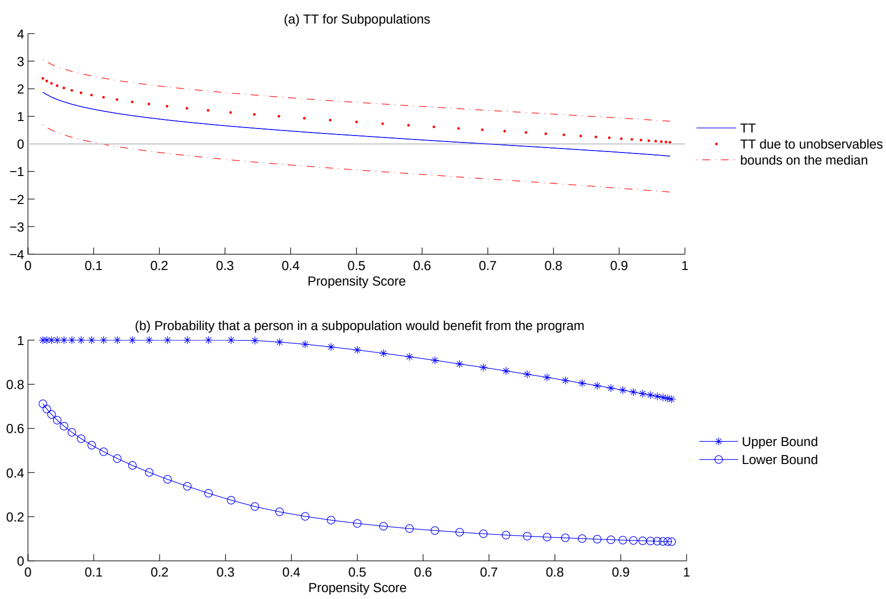

Figure 7: Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = -0.5$, $\rho_{1\varepsilon} = 0.5$, $\rho_{0\varepsilon} = -0.5$, and $\sigma_1 = \sigma_0 = 1$.

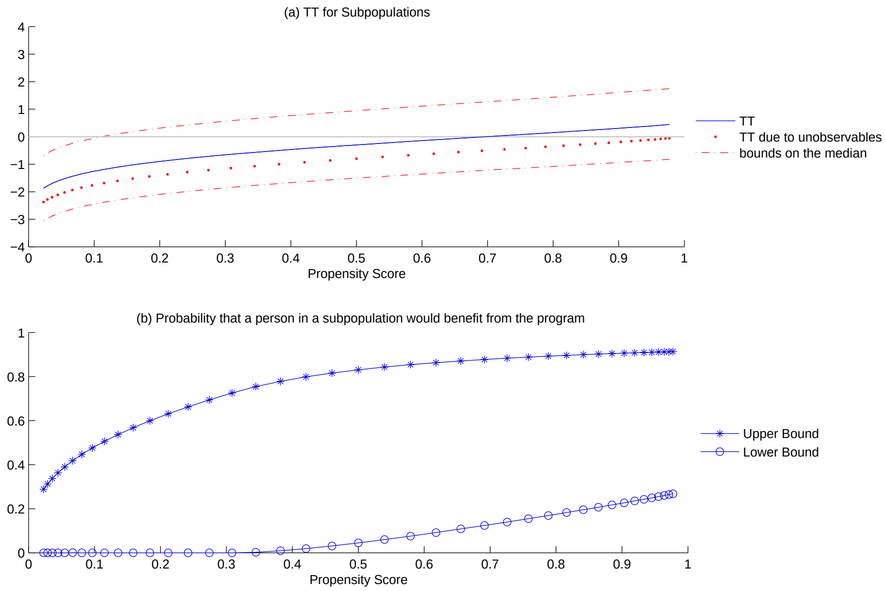

Figure 8: Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0.5$, $\rho_{1\varepsilon} = -0.5$, $\rho_{0\varepsilon} = 0.5$, and $\sigma_1 = \sigma_0 = 1$.

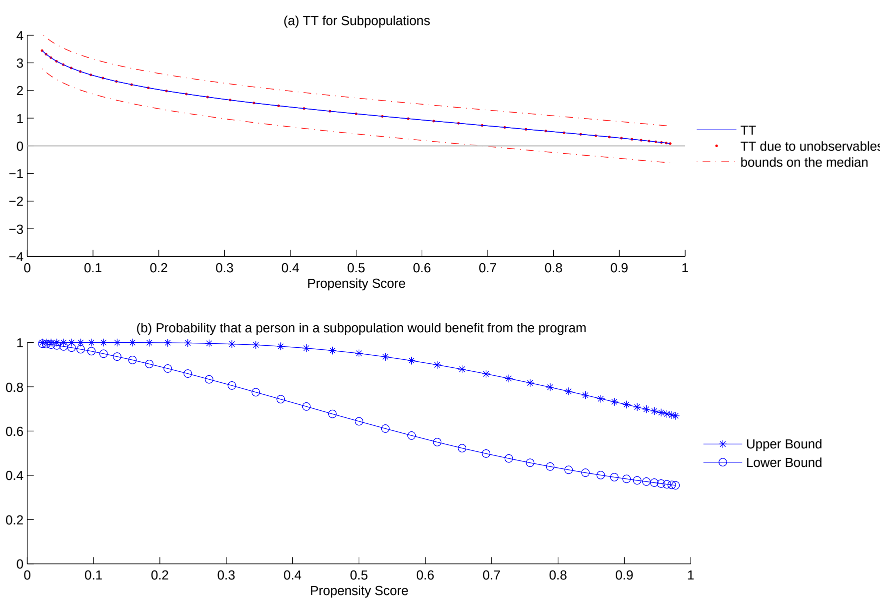

Figure 9: Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = 0.95$, $\rho_{0\varepsilon} = -0.5$, and $\sigma_1 = \sigma_0 = 1$.

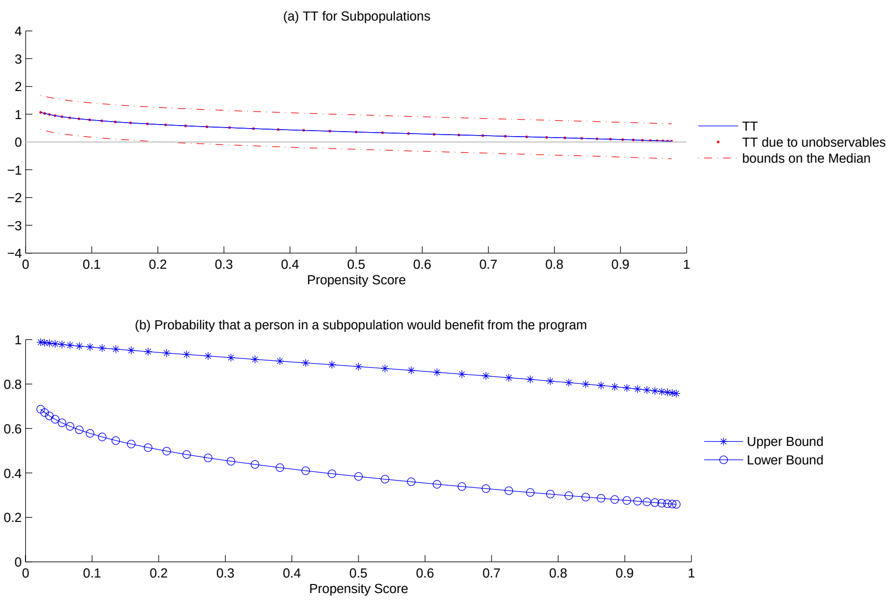

Figure 10: Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = 0.95$, $\rho_{0\varepsilon} = 0.5$, and $\sigma_1 = \sigma_0 = 1$.