

University of Toronto  
Department of Economics



Working Paper 498

**SHARP BOUNDS ON TREATMENT EFFECTS IN A BINARY  
TRIANGULAR SYSTEM**

By Ismael MOURIFIÃ

October 01, 2013

# SHARP BOUNDS ON TREATMENT EFFECTS IN A BINARY TRIANGULAR SYSTEM

ISMAEL MOURIFIÉ

**ABSTRACT.** This paper considers the evaluation of the average treatment effect ( $ATE$ ) in a triangular system with binary dependent variables. I impose a threshold crossing model on both endogenous regressor and the outcome. No parametric functional form or distributional assumptions are imposed. Shaikh and Vytlacil (2011, SV) proposed bounds on  $ATE$  which are sharp only under a restrictive condition on the support of the covariates and the instruments, which rules out a wide range of models and many relevant applications. In some cases, when SV's support condition fails, their bounds retrieve the same empirical content as the model with unrestricted endogenous regressor. In this setting, I provide a methodology which allows to construct sharp bounds on the  $ATE$  by efficiently using variation on covariates without imposing support restrictions.

**Keywords:** partial identification, threshold crossing model, triangular system, average treatment effect, endogeneity, program social evaluation.

**JEL subject classification:** C14, C31, C35

---

*Date:* June 22, 2013. I am deeply grateful to Marc Henry for his constant guidance, inspiration and encouragement. I am also grateful to Louis-Philippe Béland, Ivan Canay, Sung Jae Jun, Joris Pinkse, Christoph Rothe, Alexander Torgovitsky, Bernard Salanié for helpful discussions and comments from participants to Second CIREQ-CeMMAP Conference on Incomplete models and from seminar audiences in Cambridge, Chicago, Columbia, Exeter Business School, PennState, University of Toronto and Warwick economics departments. Parts of this paper were written while I was visiting Penn State, I thank my hosts for their hospitality and support. Correspondence address: Department of Economics, University of Toronto, 150 St. George Street, Toronto ON M5S 3G7, Canada, ismael.yacoub@gmail.com.

## INTRODUCTION

This paper considers the evaluation of the average treatment effect ( $ATE$ ) of a binary endogenous regressor on a binary outcome when I impose a threshold crossing model on both the endogenous regressor and the outcome. This model encompasses many important applications in different areas of economics including labor economics as in Battistin and Rettore (2002), education, as in Canton and Bloom (2004), Beffy, Fougère and Maurel (2010), health economics as in Bhattacharya, Shaikh, and Vytlačil (2008), Carpenter and Dobkin (2009), political economy as in Lee (2008) among many others.

The joint threshold crossing model was recently investigated by Shaikh and Vytlačil (2011), but their proposed bounds are sharp only under a critical restriction imposed on the support of the covariates and the instruments. The support condition required is very strong and fails when we have deterministic treatment or partially deterministic treatment. Even without deterministic treatment, SV's support condition is likely to fail for a wide range of models. Basically, the SV critical support condition is more likely to hold in the rare case when there is no common exogenous covariates between the outcome and the treatment equation. Indeed, SV takes advantage of the threshold crossing condition imposed on the endogenous regressor, to refine known bounds on the  $ATE$  in the model with unrestricted endogenous regressor. However, when the support condition fails they do not take full advantage of the threshold crossing condition imposed on the endogenous regressor. In some cases, their bounds do not have any empirical content beyond the model with unrestricted endogenous regressor. I show throughout this paper, how it is possible to fully exploit the second threshold crossing restriction imposed on the endogenous regressor without imposing any support restrictions. More specifically, I show under the joint threshold crossing model, that the sign of the marginal average effect may be identified, and observable bounds of the marginal average effect can be derived. I take advantage of that to construct sharp bounds on the  $ATE$ .

Therefore, this paper complements SV's work by providing a methodology which allows to construct sharp bounds on the  $ATE$  by efficiently using variation on covariates and which does not need to impose any support restrictions. The proposed methodology requires only mild regularity conditions on the distribution of unobservable variables and an usual exogeneity assumption between the covariates (except the binary endogenous regressor) and the unobservable variables. Inference for the proposed bounds can be easily carried out using existing inferential methods, specifically Chernozhukov, Lee, and Rosen (2013) or Andrews and Shi (2013). The proof of sharpness of the

proposed bounds is based on copula theory and a characterization theorem proposed by Chiburis (2010). A similar objective is pursued by Chiburis (2010), but his approach relies on an algorithm to determine existence of a copula, which is computationally infeasible in many cases of interest. I provide a methodology to reduce the computational burden of the Chiburis (2010) technique. However, the method proposed in this paper remains much simpler to apply. In addition, this method can be easily extended to a triangular system with nonbinary-valued discrete endogenous regressors and continuous outcome.

This joint threshold crossing model is a special case of nonparametric triangular systems. Imbens and Newey (2009) and Kasy (2011) develop nonparametric identifications results in triangular systems by using the “control function” approach. Their results hold when dependent and endogenous variables are continuous, but do not extend to the present context. This model is also a particular case of Chesher (2005), but his analysis requires a strong rank condition and an additional assumption on the joint distribution of the unobservable variables. His rank condition rules out the case where the endogenous regressor is binary. Jun, Pinkse and Xu (2010, JPX) relax this rank condition but still maintain an additional assumption on the joint distribution of the unobservable variables. Moreover, identification of the *ATE* in this model was previously considered by Vytlačil and Yildiz (2007, VY). They showed that under a strong support condition it is possible to point identify the *ATE*. Jun, Pinkse and Xu (2011) weaken the VY support condition by using the identification method proposed in JPX, but still maintain an additional assumption on the joint distribution of the unobservable variables. Here, I do not impose such restrictions, which might be very restrictive.

The rest of the paper is organized as follows. The next section revisits single threshold crossing models, when no structural form is assumed for the binary endogenous regressor. The following section considers joint threshold crossing models, explains why SV’s bounds fails to be sharp without their support condition and proposes a methodology to sharpen their bounds in the case, where the latter fails to hold. The third and fourth sections present an numerical illustration and discuss the inference procedure. The last section concludes and proofs are collected in the appendix.

## 1. THRESHOLD CROSSING MODEL WITH UNRESTRICTED BINARY TREATMENT

I adopt in this section the framework of the potential outcomes model  $Y = Y_1D + Y_0(1 - D)$ , where  $Y$  is an observed outcome,  $D$  denotes the observed binary endogenous regressor and  $Y_1, Y_0$

are unobserved potential outcomes. Potential outcomes are as follows:

$$Y_d = 1\{F(d, x, u) > 0\}, d = 1, 0 \quad (1.1)$$

where  $1\{.\}$  denotes the indicator function and  $F$  is an unknown function of a vector of exogenous regressors  $X$ , and unobserved random variable  $u$ . The formal assumptions I use in this section may be expressed as follows:

**Assumption 1.** *The functions  $F(d, x, u)$ ,  $d=1,0$ , both have weakly separable errors. As shown in Vytlacil (2002) and Vytlacil and Yildiz (2007), potential outcomes can then be written  $Y_d = 1\{\nu(d, x) > u\}$  without loss of generality.*

**Assumption 2.**  *$(X, Z)$  and  $u$  are statistically independent, where  $Z$  is an available instrument.*

**Assumption 3.** *The distribution of  $u$  has positive density w.r.t Lebesgue measure on  $\mathbb{R}$ .*

According to equation (1.1) we have  $\mathbb{E}[Y_d|X, Z] = \mathbb{E}[Y_d|X]$ . It follows from assumptions 1 and 3 that we may impose, without loss of generality, the normalization that  $u$  is uniformly distributed on  $[0,1]$  ( $u \sim \text{Uniform}[0,1]$ ). This normalization is very convenient, since it implies  $\mathbb{E}[Y_d | X = x] = P(Y_d = 1 | X = x)$  and bounds on treatment effects parameters can be derived from bounds on the structural parameters  $\nu(1, x)$  and  $\nu(0, x)$ . Then we may define the average structural function (*ASF*) and the average treatment effect (*ATE*), respectively, as follows:

$$\begin{aligned} \nu(d, x) &= P(Y_d = 1 | X = x) \\ \Delta\nu(x) &= P(Y_1 = 1 | X = x) - P(Y_0 = 1 | X = x). \end{aligned}$$

In all this section, I shall use the notation  $P(i, j|x, z) = P(Y = i, D = j|X = x, Z = z)$ . Now, I will provide a proposition which recalls known result on sharp bounds on the *ASF* when the binary endogenous regressor is unrestricted.

**Proposition 1** (Chiburis (2010)). *Suppose  $Y, D$  determined by model (1.1). Let  $\text{Dom}(X)$  and  $\text{Dom}(Z)$  denote the respective domains of the random variables  $X$  and  $Z$ . Under assumptions 1, 2 and 3, the following bounds are sharp for the average structural function (*ASF*) in the model (1.1): For each  $x$  in  $\text{Dom}(X)$ ,*

- if  $\nu(0, x) \leq \nu(1, x)$

$$\begin{aligned} \sup_z \{P(1, 0|x, z)\} &\leq \nu(0, x) \leq \inf_z \{P(Y = 1 | x, z)\} \\ \sup_z \{P(Y = 1|x, z)\} &\leq \nu(1, x) \leq \inf_z \{P(1, 1 | x, z) + P(D = 0 | x, z)\}, \end{aligned}$$

- if  $\nu(0, x) \geq \nu(1, x)$

$$\begin{aligned} \sup_z \{P(Y = 1|x, z)\} &\leq \nu(0, x) \leq \inf_z \{P(1, 0|x, z) + P(D = 1|x, z)\} \\ \sup_z \{P(1, 1|x, z)\} &\leq \nu(1, x) \leq \inf_z \{P(Y = 1|x, z)\}, \end{aligned}$$

where the supremum and the infimum are taken over the domain of the random variable  $Z$  ( $Dom(Z)$ ).

One proof of this result is given by Chiburis (2010), but I give an alternative proof which considers this model as a particular case of discrete outcome models with multiple equilibria. (See Appendix A). The proof I propose is convenient since it allows to derive sharp bounds for model with specific sector heterogeneity i.e  $Y_d = 1\{F(d, x, u_d) > 0\}$ , as it has been done in subsequent paper, see Henry and Mourifié (2012). It, also, allows to derive sharp bounds in a case of non-binary discrete endogenous regressor. In a case where the bounds cross for one of the two cases the sign of the *ATE* is identified. For instance,  $\sup_z \{P(1, 0|x, z)\} > \inf_z \{P(Y = 1 | x, z)\}$  but  $\sup_z \{P(Y = 1|x, z)\} \leq \inf_z \{P(1, 0|x, z) + P(D = 1|x, z)\}$  and  $\sup_z \{P(1, 1|x, z)\} \leq \inf_z \{P(Y = 1|x, z)\}$  then  $\Delta\nu(x) < 0$ . If the bounds cross in both cases, the joint assumption of weakly separable errors and the presence of a valid instrument  $Z$  is rejected. When no instrument is available, the previous sharp bounds of proposition 1 become:

- if  $\nu(0, x) \leq \nu(1, x)$

$$\begin{aligned} P(1, 0|x) &\leq \nu(0, x) \leq P(Y = 1|x) \\ P(Y = 1 | x) &\leq \nu(1, x) \leq P(1, 1|x) + P(D = 0|x), \end{aligned}$$

- if  $\nu(0, x) \geq \nu(1, x)$

$$\begin{aligned} P(Y = 1|x) &\leq \nu(0, x) \leq P(1, 0|x) + P(D = 1|x) \\ P(1, 1|x) &\leq \nu(1, x) \leq P(Y = 1|x), \end{aligned}$$

which are the same as Manski and Pepper's (2000) bounds on the *ASF*. In such a context, it is immediately apparent that the sign of the *ATE* is not identified and the weakly separable errors assumption cannot be falsified.

## 2. JOINT THRESHOLD CROSSING MODEL

In this section, I put more structure on the previous model by assuming that the binary endogenous regressor has the following structure  $D = 1\{G(x, z, v) > 0\}$ , where  $G$  is an unknown function weakly separable in  $v$ . Then I propose the following model:

$$\begin{aligned} Y_d &= 1\{F(d, x, u) > 0\}, d = 0, 1 \\ D &= 1\{G(x, z, v) > 0\}. \end{aligned} \tag{2.1}$$

This model can be summarized without loss of generality ( see Vytlačil (2002) and Vytlačil and Yildiz (2007)) as follows:

$$\begin{aligned} Y_d &= 1\{\nu(d, x) > u\}, d = 0, 1 \\ D &= 1\{p(x, z) > v\}, \end{aligned} \tag{2.2}$$

where  $Y = Y_1D + Y_0(1 - D)$  denotes the observed binary outcome of interest,  $D$  denotes the observed binary endogenous regressor,  $(X, Z)$  is a vector of exogenous regressors,  $(u, v)$  are unobserved random variables. The formal assumptions I use in this section may be expressed as follows:

**Assumption 4.**  $(X, Z)$  and  $(u, v)$  are statistically independent.

**Assumption 5.** The distribution of  $(u, v)$  has positive density w.r.t Lebesgue measure on  $\mathbb{R}^2$ .

It follows from assumptions 4 and 5, that we may impose, without loss of generality, the normalization that  $u, v \sim U[0, 1]$ ,  $\nu(d, x) = \mathbb{E}[Y_d | X = x] = P(Y_d = 1 | X = x)$ ,  $P(X, Z) = P(D = 1 | X, Z)$  and  $\nu(d, x) = \mathbb{E}[Y_d | X = x, P(X, Z) = p]$ . Then, the *ASF* is  $\nu(d, x) = P(Y_d = 1 | X = x)$  and the *ATE* is  $\Delta\nu(x) = P(Y_1 = 1 | X = x) - P(Y_0 = 1 | X = x)$ . In all this section, I shall use the notation  $P(i, j | x, p) = P(Y = i, D = j | X = x, P(X, Z) = p)$ . Similar analysis has been carried out, previously, by SV. Their work provides bounds on the *ASF*, which are based on observable quantities, and which exploit covariates variation. SV used the joint threshold crossing equations determining  $Y$  and  $D$  and additional assumptions to determine the sign of  $[\nu(1, x') - \nu(0, x)]$  from the distribution of observed data, and then take advantage of this information to construct bounds on *ASF* which exploit variation on covariates. Denote by  $Supp(P(X, Z) | X)$  the support of  $P(X, Z)$

conditional on  $X$  and write  $P(X, Z) = P$  in the rest of the paper. Before going into details, let's provide a simple intuition of the main idea of this paper. We have

$$\nu(0, x) = P(u \leq \nu(0, x), v \geq p(x, z)) + P(u \leq \nu(0, x), v \leq p(x, z)),$$

where  $P(u \leq \nu(0, x), v \geq p(x, z)) = P(1, 0|x, p)$ , but the second term  $P(u \leq \nu(0, x), v \leq p(x, z)) = P(Y_0 = 1, D = 1|X = x)$  is the unobserved counterfactual. SV proposed to bound this counterfactual by exploiting variation on covariates. Indeed, SV's idea suggests that, we may bound the unobserved counterfactual for untreated individuals ( $D = 0$ ) with characteristic  $x$  by using information on treated individuals ( $D = 1$ ) with different characteristics  $x'$  whenever they have exactly the same probability to be treated. In fact, if we have a treated individual with characteristic  $x'$  belonging to the following set  $\Delta(x) = \{x' : \nu(0, x) \leq \nu(1, x')\} \cap \{x' : \exists z' \in \text{Dom}(Z), p(x, z) = p(x', z')\}$ , the proposed bounds of SV for the unobserved counterfactual can be summarized as follows

$$P(u \leq \nu(0, x), v \leq p(x, z)) \leq \begin{cases} P(u \leq \nu(1, x'), v \leq p(x', z')) & \text{if } x' \in \Delta(x) \\ p(x, z) & \text{if } \Delta(x) = \emptyset \end{cases}$$

where  $P(u \leq \nu(1, x'), v \leq p(x', z')) = P(1, 1|x', p')$ . Their idea is quite interesting, but not sufficient to provide the sharp bounds. My argument relies on the fact that, under the threshold crossing model assumption imposed on the treatment ( $D$ ), we may bound the unobserved counterfactual  $\mathbb{P}(Y_0 = 1, D = 1|x, z)$  by using information on treated individuals with different characteristics  $x'$  even if they have different probabilities to be treated. In fact, if we have a treated individual with characteristic  $x'$  belonging to the following subset  $\tilde{\Delta}(x) = \{x' : \nu(0, x) \leq \nu(1, x')\} \cap \{x' : \exists z' \in \text{Dom}(Z), p(x, z) \leq p(x', z')\}$ , the unobserved counterfactual may be bounded as follows

$$P(u \leq \nu(0, x), v \leq p(x, z)) \leq \begin{cases} P(1, 1|x', p') & \text{if } x' \in \tilde{\Delta}(x) \\ p(x, z) & \text{if } \tilde{\Delta}(x) = \emptyset. \end{cases}$$

Then, it is immediately apparent that, the subset  $\Delta(x)$  is necessary but not sufficient to construct the sharp bound for  $\nu(0, x)$ . Instead of using  $\Delta(x)$ , I propose to visit  $\tilde{\Delta}(x)$ . Since  $\Delta(x) \subseteq \tilde{\Delta}(x)$ , it is easy to see that we may get an improvement over SV's bounds by using  $\tilde{\Delta}(x)$  instead of  $\Delta(x)$ , especially when  $\Delta(x)$  is empty or  $\Delta(x) = \{x\}$ . When  $\text{Supp}(P|X = x) = \text{Supp}(P|X = x')$  we have  $\tilde{\Delta}(x) = \Delta(x)$ , this fact explains why the SV bounds would be sharp when  $\text{Supp}(P|X = x) = \text{Supp}(P|X = x')$ . This first idea is not sufficient to fully characterize all the empirical content of the model. In this first idea I show that, to bound the *ASF* for an untreated individual ( $D = 0$ ) with characteristic  $x$ , we may use information on a treated individual ( $D = 1$ ) with different characteristic  $x'$ . My second idea relies on the fact that, under the threshold crossing model assumption imposed on the treatment ( $D$ ), to bound the *ASF* for an untreated individual ( $D = 0$ ) with characteristic  $x$ , we



may, also, use information on other untreated individual ( $D = 0$ ) with different characteristic  $x'$ . In fact, if we have two untreated individuals with characteristics  $x$  and  $x'$  such that  $\nu(0, x) < \nu(0, x')$ , we have the following

$$\begin{aligned} P(\nu(0, x) \leq u \leq \nu(0, x')) &= P(\nu(0, x) \leq u \leq \nu(0, x'), v \geq p(x, z)) \\ &\quad + P(\nu(0, x) \leq u \leq \nu(0, x'), v \leq p(x, z)) \\ &\geq P(\nu(0, x) \leq u \leq \nu(0, x'), v \geq p(x, z)) \\ &\geq P(u \leq \nu(0, x'), v \geq p(x, z)) - P(u \leq \nu(0, x), v \geq p(x, z)), \end{aligned}$$

then for all  $p(x, z) \leq p(x', z')$  we have

$$\begin{aligned} \nu(0, x') - \nu(0, x) &\geq P(u \leq \nu(0, x'), v \geq p(x', z')) - P(u \leq \nu(0, x), v \geq p(x, z)) \\ &\geq P(1, 0|x', p') - P(1, 0|x, p). \end{aligned}$$

Thus for all  $x'$  belonging to  $\{x' : \nu(0, x) \leq \nu(0, x')\} \cap \{x' : \exists z' \in \text{Dom}(Z), p(x, z) \leq p(x', z')\}$ , we shall have

$$\nu(0, x') - \nu(0, x) \geq \max[P(1, 0|x', p') - P(1, 0|x, p), 0]. \quad (2.3)$$

We can easily see that SV's bounds do not recover this feature of the model. For instance, in a case where  $\text{Supp}(P|X = x) \cap \text{Supp}(P|X = x') = \emptyset$ , SV's bounds become the bounds derived in proposition 1, and then, if  $\nu(1, x') > \nu(0, x') > \nu(0, x) > \nu(1, x)$  we have  $\nu(0, x') - \nu(0, x) \geq P(1, 0|x', p') - P(1, 0|x, p) - P(D = 1|x, p)$  which is wider than the bound proposed in (2.3). To the best of our knowledge, this paper is the first to construct bounds on the *ATE* which respect such feature of the model. As we can remark throughout the above discussion the signs of the following quantities  $[\nu(1, x') - \nu(0, x)]$ ,  $[\nu(d, x') - \nu(d, x)]$ , for  $d = 1, 0$  are very important in our analysis.

Model (3.1) has the additional nice feature that it allows identification of the sign of the following marginal average effect

$$\mathbb{E}[Y_1|X = x'] - \mathbb{E}[Y_0|X = x] = [\nu(1, x') - \nu(0, x)] \quad (2.4)$$

and

$$\mathbb{E}[Y_d|X = x'] - \mathbb{E}[Y_d|X = x] = [\nu(d, x') - \nu(d, x)], \quad d=1,0 \quad (2.5)$$

under very mild assumptions. SV showed that  $[\nu(1, x') - \nu(0, x)]$  share the same sign as the following observable function  $h(x, x', p, p') = (P(1, 1|x', p) - P(1, 1|x', p')) - (P(1, 0|x, p') - P(1, 0|x, p))$ , when  $P$  is not degenerate, and both  $p$  and  $p'$  belong to  $\text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$  such that

$p' < p$ . As we can remark, the SV idea cannot identify the sign of  $[\nu(1, x') - \nu(0, x)]$  when  $\text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$  is empty or a singleton. However, the sign of  $[\nu(1, x') - \nu(0, x)]$  would still be identified. Indeed, if there are  $p'_1 < p'_2 \in \text{Supp}(P | X = x')$  and  $p_1 < p_2 \in \text{Supp}(P | X = x)$  such that  $[p'_1, p'_2] \subseteq [p_1, p_2]$  the sign of  $[\nu(1, x') - \nu(0, x)]$  would be identified using the following observable function  $[P(1, 1|x', p'_2) - P(1, 1|x', p'_1)] - [P(1, 0|x, p_1) - P(1, 0|x, p_2)]$ . In the Lemma 1 in Appendix A, I show how the sign of  $[\nu(1, x') - \nu(0, x)]$  may be identified under weaker assumptions. Moreover, I show that the sign of  $[\nu(d, x') - \nu(d, x)]$   $d = 1, 0$  may also be identified under very mild assumptions.

This interesting feature of the model, will help to reduce considerably the computation burden of the proposed bounds, as we will see later.

Now, let recall the SV bounds. SV take advantage of the knowledge of the sign of  $[\nu(1, x') - \nu(0, x)]$ , to construct an upper bound for  $\nu(0, x)$ , when  $\text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$  is not empty. Therefore, if  $p \in \text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$ , and  $h(x, x', p, p') \geq 0$  we have

$$\begin{aligned} \nu(0, x) &= P(u \leq \nu(0, x), v \geq p) + P(u \leq \nu(0, x), v \leq p) \\ &\leq P(u \leq \nu(0, x), v \geq p) + P(u \leq \nu(1, x'), v \leq p), \end{aligned}$$

for all  $x' \in \mathbf{X}(x) = \{x' : h(x, x', p, p') \geq 0\}$ . Hence,

$$\nu(0, x) \leq \inf_p \{P(1, 1|x, p) + p \inf_{x' \in \mathbf{X}(x)} P(1, 1|x', p')\}.$$

Similarly, we can derive the lower bound for  $\nu(0, x)$ , and also the lower and upper bounds for  $\nu(1, x)$ . Hereafter, I adopt the convention that the supremum over empty set is zero and the infimum over the empty set is one.

**Remark 1.** *SV take advantage of the knowledge of the sign of  $[\nu(1, x') - \nu(0, x)]$ , to construct an upper bound for  $\nu(0, x)$  only whenever there exist  $p' < p$  belong to  $\text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$ . Lemma 1 showed that we may take advantage of the knowledge of the sign of  $[\nu(1, x') - \nu(0, x)]$ , to construct an upper bound for  $\nu(0, x)$  even if  $\text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$  is empty or a singleton.*

SV showed that these bounds are sharp under further assumptions, which can be expressed as follows:

**Assumption 6.** *The functions  $\nu(0, \cdot)$ ,  $\nu(1, \cdot)$  and  $p(\cdot)$  are continuous.*

**Assumption 7.** *The support of the distribution of  $(X, Z)$ ,  $Supp(X, Z)$ , is compact.*

**Assumption 8.** *(critical support condition)  $Supp(X, P) = Supp(X) \times Supp(P)$ .*

Their result remains restrictive due to the strong restriction imposed by the “critical support condition”.

**2.1. Plausibility of the “critical support condition”.** Assumption 8 implies that  $Supp(P | X = x) = Supp(P | X = x')$  for all  $(x, x') \in Dom(X) \times Dom(X)$ , in others terms for all  $(x, x') \in Dom(X) \times Dom(X)$  and  $z \in Dom(Z)$ , there exist  $z' \in Z$  such that  $p(x, z) = p(x', z')$ . This type of “perfect matching restriction” may be difficult to achieve in many applications.

**2.1.1. partially deterministic treatment.** In the treatment effect setting, there are many applications where we have additional information about the treatment assignment: it’s known that the treatment assignment mechanism depends (at least in part) on the value of observed variables. Then, the treatment may be deterministic for some characteristics  $(x', z')$  such that the treated probability is degenerate in some points i.e  $(p(x', z') = 1 \text{ or } p(x', z') = 0)$ . One well suited example for this type of deterministic treatment is the financial aid selection rule. Van der Klaauw (2002) studied financial aid selection rule and showed that USA’s colleges use generally SRT (Standard Reasoning Test) and GPA (Grade Point Average) to rank students into a small number of categories and then decide on a particular selection rule. Let

$$G_l = \begin{cases} 1 & \text{if } 0 \leq GPA \leq C_1, \\ 2 & \text{if } C_1 \leq GPA \leq C_2, \\ \vdots & \\ \vdots & \\ L & \text{if } C_{L-1} \leq GPA \leq C_L, \end{cases}$$

denote the financial aid group. Denote by  $Z$  the vector of others characteristics such as special awards, recommendation letters or/and extracurricular activities. The selection rule is usually the following: the higher ranked category is directly selected i.e  $(p(C_{L-1} \leq GPA \leq C_L, z) = 1 \text{ for all } z \in Dom(Z))$ , the lower one is directly excluded i.e  $(p(0 \leq GPA \leq C_1, z) = 0 \text{ for all } z \in Dom(Z))$  and for the others, we look at others characteristics such as special awards, recommendation letters or/and extracurricular activities. i.e  $(p(C_1 \leq GPA \leq C_{L-1}, z) \in (0, 1))$ . In this generic example, there does not exist  $(z, z') \in Dom(Z) \times Dom(Z)$  such that  $P(G_1, z) = P(G_l, z')$  for  $1 < l \leq L$ , then SV’s “critical support condition” fails. We have showed in the latter example, that SV’s “critical support condition” fails when we have deterministic treatment or partially deterministic

treatment. Furthermore, our numerical illustration shows that even without deterministic treatment SV's "critical support condition" is likely to fail.

**2.1.2. Semiparametric bounds.** Depending on the economic model you have, it is possible to assume some functional forms or parametric forms for either  $\nu(d, x)$  or  $p(x, z)$  or both of them. For instance, we may have  $\nu(d, x) = F(x'\beta + d\alpha)$  or  $p(x, z) = \Phi(x'\gamma + z'\eta)$ . Manski (1988) discussed the identification of the single index model where  $F$  or  $\Phi$  is unknown and showed that this model fits a wide range of model. Here, I will assume a linear index model for  $D$ , with  $\Phi$  an unknown, strictly increasing function. Then, we have the following semiparametric model:

$$\begin{aligned} Y &= 1\{\nu(D, x) > u\} \\ D &= 1\{\Phi(x\gamma + z\eta) > v\}, \end{aligned} \tag{2.6}$$

with  $Dom(X) \subset \mathbb{R}$ ,  $Dom(Z) = \{0, 1\}$  and  $\Phi(\cdot) \in [0, 1]$ . We can easily see that the SV support condition fails to hold for this specification which fits a wide range of models. Indeed, since  $\Phi(\cdot)$  is strictly increasing,  $Supp(P | X = 0) = \{\Phi(0), \Phi(\eta)\} \neq \{\Phi(\gamma), \Phi(\gamma + \eta)\} = Supp(P | X = 1)$  for all  $(\gamma, \eta) \neq (0, 0)$ . Moreover, since in this case  $Supp(P | X = x) \cap Supp(P | X = x')$  is empty, SV's bounds fail to improve bounds that we found in proposition 1. It means that their bounds do not take advantage of the linear index structure imposed on  $D$ . In a case where,  $Dom(Z)$  is discrete non-binary the SV critical support condition still fails to hold. However, when the instrument  $Z$  has a continuous and a large support such that  $\lim_{z \rightarrow +\infty} \Phi(x\gamma + z\eta) = 1$ , and  $\lim_{z \rightarrow -\infty} \Phi(x\gamma + z\eta) = 0$  for all  $x \in Dom(X)$  and  $\eta > 0$ , the SV critical support condition would hold. But, in such a context, the partial identification approach entertained by SV and this present paper is less relevant since we have identification at infinity of the object of interest  $\nu(d, x)$ , for all  $x \in Dom(X)$ .

Basically, the SV critical support condition is more likely to hold only when  $p(x, z)$  does not depend on  $x$ , which is only true in the rare case of a complete dichotomy between variables in the outcome equation and variables in the treatment equation.

**2.2. Failure of sharpness of SV's bounds without the critical support condition.** SV's bounds take advantage of the observability of the sign of  $[\nu(1, x') - \nu(0, x)]$  when  $Supp(P | X = x) \cap Supp(P | X = x')$  is not empty and not reduced to a singleton. Whenever  $Supp(P | X = x) \cap Supp(P | X = x')$  is empty or reduced to a singleton, SV's bounds do not take advantage of the additional weak separability restriction that we impose on the equation determining  $D$ . I will now show, how it is possible to sharpen bounds on the *ASF* and the *ATE*, without imposing the

“critical support condition” (assumption 8). Before formalizing our idea, I will define some subsets summarized in Table 1.

TABLE 1. Collection of sets

$$\begin{aligned} \mathbf{P}^+(x', p) &= \{p(x', z') = p' \in \text{Supp}(P|X = x') : p \leq p'\} \\ \mathbf{P}^-(x', p) &= \{p(x', z') = p' \in \text{Supp}(P|X = x') : p \geq p'\} \\ \Omega_{\mathbf{d}_1 \mathbf{d}_2}^+(x) &= \{x' : \nu(d_1, x) \leq \nu(d_2, x')\} \\ \Omega_{\mathbf{d}_1 \mathbf{d}_2}^-(x) &= \{x' : \nu(d_1, x) \geq \nu(d_2, x')\} \end{aligned}$$

**2.3. Sharpening the bounds.** I will show in a first step, that it is still possible to improve proposition 1’s bounds when  $p(x, z) \notin \text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$ , by taking advantage simultaneously of the sign of  $[\nu(1, x') - \nu(0, x)]$  and the range  $\mathbf{P}^+(x', p)$ . In a second step I will show how it is possible to narrow SV’s bounds by also using the sign of  $[\nu(0, x') - \nu(0, x)]$  and the range  $\mathbf{P}^+(x', p)$ .

**First step:**

When  $p(x, z) \notin \text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$ , we cannot identify  $P(u \leq \nu(1, x'), v \leq p(x, z))$  from the data. Then SV proposed in this case to bound  $P(u \leq \nu(1, x'), v \leq p(x, z))$  from above by  $P(v \leq p(x, z)) = p(x, z)$ . But whenever it is possible to find  $p(x', z')$  in  $\mathbf{P}^+(x', p)$ , I propose to bound  $P(u \leq \nu(1, x'), v \leq p(x, z))$  from above by  $P(u \leq \nu(1, x'), v \leq p(x', z')) = P(1, 1|x', p')$ , which may be lower than  $P(v \leq p(x, z)) = p(x, z)$  in some cases. The upper bound for  $\nu(0, x)$  that we can build with this strategy is lower than the one proposed by SV. Indeed, for all  $x' \in \Omega_{\mathbf{01}}^+(x)$  and  $p(x', z') \in \mathbf{P}^+(x', p)$  we have:

$$\begin{aligned} \nu(0, x) &= P(u \leq \nu(0, x), v \geq p(x, z)) + P(u \leq \nu(0, x), v \leq p(x, z)) \\ &\leq P(u \leq \nu(0, x), v \geq p(x, z)) + P(u \leq \nu(1, x'), v \leq p(x, z)) \\ &\leq P(u \leq \nu(0, x), v \geq p(x, z)) + \min[P(u \leq \nu(1, x'), v \leq p(x', z')), p(x, z)] \\ &\leq P(1, 0|x, p) + \min[\inf_{\Omega_{\mathbf{01}}^+(x)} \inf_{\mathbf{P}^+(x', p)} P(1, 1|x', p'), p]. \end{aligned}$$

The upper bound for  $\nu(0, x)$  we just built is lower than SV’s bounds, but may not be sharp, since it is also possible to take advantage of the knowledge of the sign of  $[\nu(0, x') - \nu(0, x)]$  and the range of  $\mathbf{P}^+(x, z)$ .

**Second step:**

Let's assume that there exists  $(x', z^*)$  such that  $p(x', z^*) = 0$ , then

$$\begin{aligned}
 \nu(0, x') &= P(u \leq \nu(0, x')) \\
 &= P(u \leq \nu(0, x'), v \geq p(x', z^*)) + P(u \leq \nu(0, x'), v \leq p(x', z^*)) \\
 &= P(u \leq \nu(0, x'), v \geq p(x', z^*)) \\
 &= P(1, 0|x', p^*).
 \end{aligned}$$

Thus  $\nu(0, x')$  is point identified. As it was pointed out by Vytlacil and Yildiz (2007), we also have point identification if there exists  $(x'', z'')$  such that  $\nu(0, x') = \nu(1, x'')$  and  $p(x', z') = p(x'', z'')$ . Indeed, we have:

$$\begin{aligned}
 \nu(0, x') &= P(u \leq \nu(0, x'), v \geq p') + P(u \leq \nu(0, x'), v \leq p') \\
 &= P(u \leq \nu(0, x'), v \geq p') + P(u \leq \nu(1, x''), v \leq p'') \\
 &= P(1, 0|x', p') + P(1, 1|x'', p'').
 \end{aligned}$$

Moreover, we may have identification under weaker assumptions as showed in Jun, Pinkse and Xu (2011). Previously we bounded  $P(u \leq \nu(0, x), v \leq p(x, z))$  by  $P(u \leq \nu(1, x'), v \leq p(x', z'))$  because the first term is non identifiable from the data while the second term may be identifiable. If  $\nu(0, x')$  is point identified we can identify  $P(u \leq \nu(0, x'), v \leq p(x', z'))$  from the data. Indeed, since  $P(u \leq \nu(0, x'), v \leq p(x', z')) = \nu(0, x') - P(u \leq \nu(0, x'), v \geq p(x', z'))$ , we have  $P(u \leq \nu(0, x'), v \leq p(x', z')) = P(1, 0|x', p^*) - P(1, 0|x', p')$  or  $P(u \leq \nu(0, x), v \leq p(x, z)) = P(1, 1|x'', p'')$  depending on the source of identification. Then, depending on the sign of  $[\nu(0, x') - \nu(0, x)]$  we will be able to bound  $\nu(0, x)$  from above by terms other than  $P(1, 0|x, p) + \min[\inf_{\Omega_{01}^+(x)} \inf_{\mathbf{P}^+(x', p)} P(1, 1|x', p'), p]$ . Indeed, for all  $x' \in \Omega_{00}^+(x)$  and  $p' \in \mathbf{P}^+(x', p)$ , such that  $\nu(0, x')$  is point identified we have the following

$$\nu(0, x) \leq P(u \leq \nu(0, x), v \geq p(x, z)) + P(u \leq \nu(0, x'), v \leq p(x', z')).$$

This means that we may take advantage from the identification of a given *ASF* (i.e  $\nu(0, x')$ ) to refine bounds for others *ASF* (i.e  $\nu(0, x)$ ). Moreover, our bounds should respect the necessary condition derived in (2.3). Therefore, I proposed the following strategy to take into account those features of

the model. For all  $x' \in \Omega_{00}^+(x)$  and  $p' \in \mathbf{P}^+(x', p)$  we have:

$$\begin{aligned} \nu(0, x') - \nu(0, x) &\geq \max[P(1, 0|x', p') - P(1, 0|x, p), 0] \\ &\geq \sup_p \left\{ \sup_{\mathbf{P}^+(x', p)} \max[P(1, 0|x', p') - P(1, 0|x, p), 0] \right\}. \end{aligned}$$

Similarly, we can derive the lower bound for  $\nu(0, x)$ , and also the lower and upper bounds for  $\nu(1, x)$ . I have just shown that  $\nu(0, x)$  and  $\nu(1, x)$  should respect some necessary conditions. The following theorem proves that these necessary conditions are sufficient to fully characterize the empirical content of the model. The proof is quite involved and it relies on copula theory and a characterization theorem in Chiburis (2010).

**Theorem 1.** *Suppose  $Y, D$  determined by model (3.1). Under assumptions 4 and 5, the characterization of the identified set for  $\nu(0, \cdot), \nu(1, \cdot)$  is the following*

$$\begin{aligned} &\sup_p \left\{ P(1, 0|x, p) + \sup_{\Omega_{01}^-(x)} \sup_{\mathbf{P}^-(x', p)} P(1, 1|x', p') \right\} \\ &\leq \nu(0, x) \\ &\leq \inf_p \left\{ P(1, 0|x, p) + \min \left( \inf_{\Omega_{01}^+(x)} \inf_{\mathbf{P}^+(x', p)} P(1, 1|x', p'), p \right) \right\}, \end{aligned} \tag{2.7}$$

$$\begin{aligned} &\sup_p \left\{ P(1, 1|x, p) + \sup_{\Omega_{10}^-(x)} \sup_{\mathbf{P}^+(x', p)} P(1, 0|x', p') \right\} \\ &\leq \nu(1, x) \\ &\leq \inf_p \left\{ P(1, 1|x, p) + \min \left( \inf_{\Omega_{10}^+(x)} \inf_{\mathbf{P}^-(x', p)} P(1, 0|x', p'), (1-p) \right) \right\} \end{aligned} \tag{2.8}$$

for all  $x' \in \Omega_{00}^+(x)$

$$\nu(0, x') - \nu(0, x) \geq \sup_p \left\{ \sup_{\mathbf{P}^+(x', p)} \max[P(1, 0|x', p') - P(1, 0|x, p), 0] \right\} \tag{2.9}$$

and for all  $x' \in \Omega_{11}^+(x)$

$$\nu(1, x') - \nu(1, x) \geq \sup_p \left\{ \sup_{\mathbf{P}^-(x', p)} \max[P(1, 1|x', p') - P(1, 1|x, p), 0] \right\}. \tag{2.10}$$

We can remark in this characterization of the identified set that there exists a dependence between the sharp bounds for  $\nu(d, x)$  and  $\nu(d, x')$ . An equivalent characterization of the identified set which provides more intuition on this dependence may be derived. Let's assume that we know the sharp

bounds  $[SL_0(x'), SU_0(x')]$  for  $\nu(0, x')$ , (2.9) may be rewritten equivalently as follows: for all  $x' \in \Omega_{00}^+(x)$  and  $p' \in \mathbf{P}^+(x', p)$

$$\begin{aligned} \nu(0, x) - P(u \leq \nu(0, x), v \geq p) &\leq \nu(0, x') - P(u \leq \nu(0, x'), v \geq p') \\ &\leq SU_0(x') - P(u \leq \nu(0, x'), v \geq p'). \end{aligned}$$

Hence,

$$\begin{aligned} \nu(0, x) &\leq P(u \leq \nu(0, x), v \geq p) + SU_0(x') - P(u \leq \nu(0, x'), v \geq p') \\ &\leq P(1, 0|x, p) + (SU_0(x') - P(1, 0|x', p')). \end{aligned}$$

Therefore,

$$\nu(0, x) \leq P(1, 0|x, p) + \inf_{\Omega_{00}^+(x)} \inf_{\mathbf{P}^+(x', p)} (SU_0(x') - P(1, 0|x', p')). \quad (2.11)$$

By combining the upper bounds for  $\nu(0, x)$  derived in (2.7) and (2.11). We may propose the following upper bound for  $\nu(0, x)$ .

$$\begin{aligned} \nu(0, x) \leq \inf_p \left\{ P(1, 0|x, p) + \min \left[ \min_{\Omega_{01}^+(x)} \inf_{\mathbf{P}^+(x', p)} P(1, 1|x', p'), p \right], \right. \\ \left. \inf_{\Omega_{00}^+(x)} \inf_{\mathbf{P}^+(x', p)} (SU_0(x') - P(1, 0|x', p')) \right\}. \end{aligned}$$

Similarly, we can derive the lower bound for  $\nu(0, x)$ , and also the lower and upper bounds for  $\nu(1, x)$ . Then, we have the following equivalent characterization of the identified set for  $\nu(0, \cdot)$  and  $\nu(1, \cdot)$ . Hereafter, I shall denote by  $SL_d(x)$  and  $SU_d(x)$  respectively the sharp lower bound and sharp upper bound for  $\nu(d, x)$   $d=0,1$ .

**Corollary 1.** *Suppose  $Y, D$  determined by model (3.1). Under assumptions 4 and 5, the characterization of the identified set for  $\nu(0, \cdot), \nu(1, \cdot)$  is the following*

$$\begin{aligned} &\sup_p \left\{ P(1, 0|x, p) + \max \left[ \sup_{\Omega_{01}^-(x)} \sup_{\mathbf{P}^-(x', p)} P(1, 1|x', p'), \sup_{\Omega_{00}^-(x)} \sup_{\mathbf{P}^-(x', p)} (SL_0(x') - P(1, 0|x', p')) \right] \right\} \\ &\leq \nu(0, x) \\ &\leq \inf_p \left\{ P(1, 0|x, p) + \min \left[ \min_{\Omega_{01}^+(x)} \inf_{\mathbf{P}^+(x', p)} P(1, 1|x', p'), p \right], \right. \\ &\quad \left. \inf_{\Omega_{00}^+(x)} \inf_{\mathbf{P}^+(x', p)} (SU_0(x') - P(1, 0|x', p')) \right\} \end{aligned} \quad (2.12)$$



and

$$\begin{aligned}
& \sup_p \left\{ P(1, 1|x, p) + \max \left[ \sup_{\Omega_{10}^-(x)} \sup_{\mathbf{P}^+(x', p)} P(1, 0 | x', p'), \sup_{\Omega_{11}^-(x)} \sup_{\mathbf{P}^+(x', p)} (SL_1(x') - P(1, 1|x', p')) \right] \right\} \\
& \leq \nu(1, x) \\
& \leq \inf_p \left\{ P(1, 1|x, p) + \min \left[ \min \left( \inf_{\Omega_{10}^+(x)} \inf_{\mathbf{P}^-(x', p)} P(1, 0|x', p'), (1-p) \right), \right. \right. \\
& \quad \left. \left. \inf_{\Omega_{11}^+(x)} \inf_{\mathbf{P}^-(x', p)} (SU_1(x') - P(1, 1|x', p')) \right] \right\}.
\end{aligned} \tag{2.13}$$

The intuition which allows us to derive the sharp bounds for  $\nu(0, x)$  by using variation in covariates in corollary 1 is the following: the width of the bounds on  $\nu(0, x')$  depends on  $\text{Supp}(P|X = x')$ . Then, when  $\text{supp}(P|X = x')$  is large the bounds for  $\nu(0, x')$  become narrower. Using the relation that our model imposes between  $\nu(0, x')$  and  $\nu(0, x)$ , we may refine the bounds for  $\nu(0, x)$  using narrower bounds for  $\nu(0, x')$ . This fact explains why the sharp bounds for  $\nu(0, x)$  depend on  $SL_0(x')$  and  $SU_0(x')$ , which are the sharp bounds for  $\nu(0, x')$ . This dependence vanishes for the lower bound  $SL_0(x)$  when  $x' \notin \Omega_{00}^-(x)$  and for the upper bound when  $x' \notin \Omega_{00}^+(x)$ . Therefore, when  $\Omega_{00}^+(x) = \emptyset$ , the upper bound for  $\nu(0, x)$  becomes  $SU_0(x) = \inf_p \left\{ P(1, 0|x, p) + \min(\inf_{\Omega_{01}^+(x)} \inf_{\mathbf{P}^+(x', p)} P(1, 1|x', p'), p) \right\}$  which no longer depends on others *ASF* sharp bounds.

**2.4. Computation of the bounds.** For the construction of our bounds we need to know the ordering of the elements of the set  $S_1 = \{\nu(d, x) : x \in \text{Dom}(X) \text{ and } d \in \{0, 1\}\}$ , in order to compute the collections of subsets defined in Table 1. Without restrictions on the true ordering on  $S_1$ , one may go over all possible orderings of  $S_1$ , and keep only orderings for which the bounds derived in Theorem 1 do not cross. Unfortunately, this method may be very costly, even under parametric restrictions for  $\nu$ . Indeed, to derive sharp bounds on *ATE* for the model (3.1), Chiburis (2010) proposes to visit all possible orderings of  $S_1$ . In his empirical example, he assumes parametric restrictions for  $\nu$  to reduce the number of orderings to be checked, but still fails to determine the bounds on the *ATE* in some cases due to computational intractability, even for a very limited  $\text{Dom}(X)$ . Even though our proposed bounds are easier to derive than bounds in Chiburis (2010), it is valuable to find a methodology to reduce the number of orderings to be checked. The properties of the function  $h(x, x', p, p')$  and the functions  $\tilde{h}_d(x, x', p, p')$ ,  $h(d, x, x', p_1, p_2, p'_1, p'_2)$ ,  $\tilde{h}_d(x, x', p_1, p_2, p'_1, p'_2)$  defined in Lemma 1 help to identify a partial ordering on  $S_1$ , which can dramatically reduce the number of orderings to be checked, particularly when  $\text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$  is large. This is illustrated in this following example:

**Example 1.** Denoting  $P(Y = 1, D = d | X = x, Z = z) = p(1, d | x, z)$ , consider  $X = \{0, 1\}$ ,  $Z = \{0, 1\}$ , and the following observables:

$$\begin{aligned} p(1, 1 | 0, 0) &= \beta_1 & p(1, 0 | 0, 0) &= \gamma_1 \\ p(1, 1 | 0, 1) &= \beta_2 & p(1, 0 | 0, 1) &= \gamma_2 \\ p(1, 1 | 1, 0) &= \beta_3 & p(1, 0 | 1, 0) &= \gamma_3 \\ p(1, 1 | 1, 1) &= \beta_4 & p(1, 0 | 1, 1) &= \gamma_4, \end{aligned}$$

with  $0 < \beta_i, \gamma_i < 1$  for  $i = 1 \dots 4$ ; and  $p(x, z)$  such that

- (1)  $\beta_1 - \beta_2 > \gamma_2 - \gamma_1 > \beta_3 - \beta_4 > \gamma_4 - \gamma_3$
- (2)  $p(0, 1) = p(1, 1) < p(0, 0) < p(1, 0)$ .

Those conditions are sufficient to determine the sign of some marginal average effects. We can describe the set of all possible orderings on  $S_1$  as follows:

$$\nu(d_1, d'_1) < \nu(d_2, d'_2) < \nu(d_3, d'_3) < \nu(d_4, d'_4),$$

where  $\{d_i, d'_i\} \neq \{d_j, d'_j\}$  for  $i \neq j$  and  $d_i, d'_i \in \{0, 1\}$ , hence there are 24 possible orderings.

We have  $\text{Supp}(P | X = 0) = \{p(0, 1), p(0, 0)\}$  and  $\text{Supp}(P | X = 1) = \{p(1, 1), p(1, 0)\}$ . Thus,  $\text{Supp}(P | X = 0) \neq \text{Supp}(P | X = 1)$ . The approach of SV can allow to identify only the sign of  $[\nu(1, 1) - \nu(0, 1)]$  and  $[\nu(1, 0) - \nu(0, 0)]$  from the observable function  $h(x, x', p, p')$ .

$$\begin{aligned} \text{sign}[\nu(1, 1) - \nu(0, 1)] &= \text{sign}[h(1, 1, p(1, 1), p(1, 0))] = + \\ \text{sign}[\nu(1, 0) - \nu(0, 0)] &= \text{sign}[h(0, 0, p(0, 0), p(0, 1))] = +. \end{aligned}$$

Then, we have the following partial ordering  $\nu(0, 1) \leq \nu(1, 1)$  and  $\nu(0, 0) \leq \nu(1, 0)$  on  $S_1$ . Among the 24 possible orderings only 6 are compatible with restrictions imposed by this partial ordering. We can see that even in a worst case when  $\text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$  is empty or a singleton, the number of orderings to be checked drops from 24 to 6 (i.e when  $\text{Supp}(P | X = x)$  is not a singleton we can always identify the sign of  $\text{sign}[\nu(1, x) - \nu(0, x)]$ ). Moreover, we can see that the sign of  $[\nu(0, 1) - \nu(1, 0)]$  may be identified using the function  $h(0, 0, 1, p(1, 1), p(1, 0), p(0, 1), p(0, 0))$  defined in Lemma 1. Indeed,  $\text{sign}[\nu(0, 1) - \nu(1, 0)] = +$ . Then, we have only one ordering compatible with the data in this generic case:  $\nu(0, 0) < \nu(1, 0) < \nu(0, 1) < \nu(1, 1)$ .

Since the number of orderings to be checked is reduced, it is now valuable to propose an efficient procedure to construct our bounds. The following iterative procedure can be used. The idea is

to use the last information we obtained on previous sharp bounds, to sharpen remaining *ASF* bounds. For instance, first, construct  $SU_0(x)$  such that  $\Omega_{00}^+(x) = \emptyset$ , then construct  $SU_0(x')$  such that  $x = \Omega_{00}^+(x')$ , then  $SU_0(x'')$  such that  $\{x, x'\} = \Omega_{00}^+(x'')$  and then iterate the strategy. The same iterative procedure holds to derive the lower sharp bounds for every element of the set  $\{\nu(0, x) : x \in \text{Dom}(X)\}$ , and also the lower and upper sharp bounds for every element of the set  $\{\nu(1, x) : x \in \text{Dom}(X)\}$ . For instance, in the latter example, one plausible ordering to be checked is the following:

$$\nu(0, 0) < \nu(1, 0) < \nu(0, 1) < \nu(1, 1).$$

Our procedure proposes to construct the lower bounds in the following ordering:

$$SL_0(0), \quad SL_1(0), \quad SL_0(1), \quad SL_1(1),$$

and the upper bounds in the following ordering:

$$SU_1(1), \quad SU_0(1), \quad SU_1(0), \quad SU_0(0).$$

**Remark 2.** *The methodology I used to derive the sharp bounds on ATE could not be used in the model with specific sector heterogeneity i.e  $Y_d = 1\{F(d, x, u_d) > 0\}$  where  $u_1 \neq u_0$ . However, this methodology directly holds if I impose the following rank similarity assumption of Chernozhukov and Hansen (2005):  $u_d|v \sim u|v$ . This assumption is weaker than the assumption I used throughout this paper i.e  $u_1 = u_0$ .*

### 3. NUMERICAL ILLUSTRATION

Now, I provide a numerical illustration of the bounds on *ATE* using Theorem 1. In addition to the bounds proposed in Theorem 1, I will compute the SV bounds. Consider the following special case of the model:

$$\begin{aligned} Y &= 1\{\alpha D + x\beta > \epsilon_1\} \\ D &= 1\{x\gamma + z\eta > \epsilon_2\}, \end{aligned} \tag{3.1}$$

with  $\text{Dom}(X) = [-2, 2]$  and  $\text{Dom}(Z) = \{0, 1\}$  and  $(\epsilon_1, \epsilon_2) \sim N(0, \Sigma)$  where

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

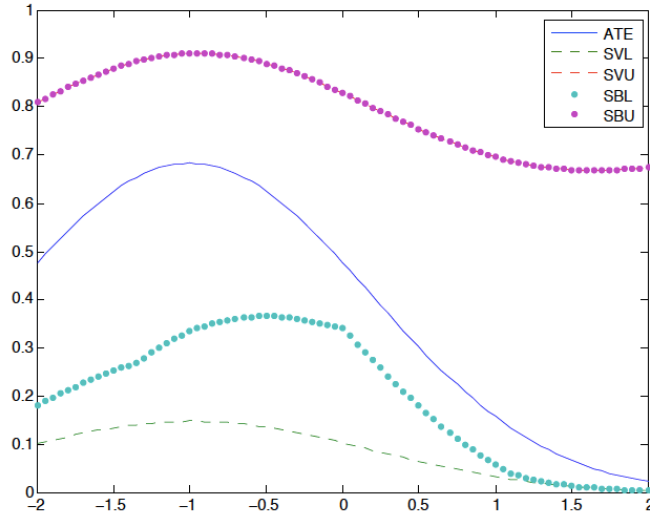
We can easily see that the SV support condition fails to hold for  $\gamma \neq 0$ . Indeed,  $\text{Supp}(P \mid X = x) \cap \text{Supp}(P \mid X = x')$  is either empty or reduces to a singleton, for all  $x \in [-2, 2]$ . I will construct the bounds by fixing  $\alpha = 2$ ,  $\rho = \frac{1}{2}$  while varying other parameters. I consider all the ordering induced by the parametric form  $\alpha'D + x\beta'$ . In fact, every couple of parameter  $(\alpha', \beta')$  induces one

ordering. Before constructing the bounds I will apply Lemma 1 to reduce the number of ordering to be checked. For example, the function  $h(x, x, p, p')$  allows to identify the sign of  $[\nu(1, x) - \nu(0, x)]$  for all  $x$ , which implies that  $\alpha' > 0$ . Now, we may visit only the ordering induced by  $(\alpha' > 0, \beta')$ . In this numerical illustration, I consider all the ordering induced by  $\alpha' \in (0, 5)$  and  $\beta' \in [\beta - 2.5, \beta + 2.5]$ . It is possible to consider a larger space, but within the simulation we note that most of the orderings induced by  $(\alpha', \beta')$  are rejected whenever those orderings deviate slightly from the true ordering induced by the true parameters  $(\alpha, \beta)$ .

All the figures show the  $ATE(x)$  while  $x$  varying from  $[-2, 2]$  for different values of the parameters. Here, I will describe four different facts:

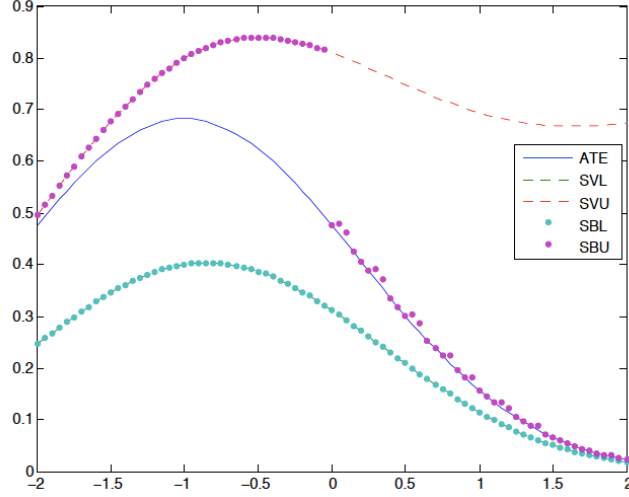
- (1) The figure 1 represents the case where  $(\beta = 1, \gamma = \frac{1}{5}, \eta = \frac{1}{2})$ . We can see that our lower

FIGURE 1. Sharp bounds on ATE when  $(\beta = 1, \gamma = \frac{1}{5}, \eta = \frac{1}{2})$ .



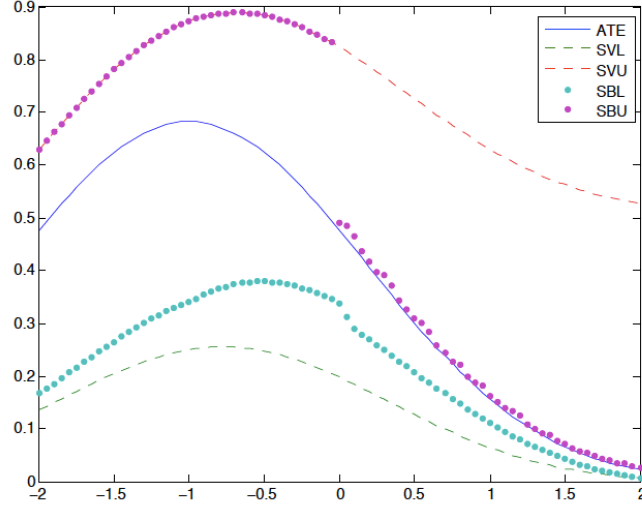
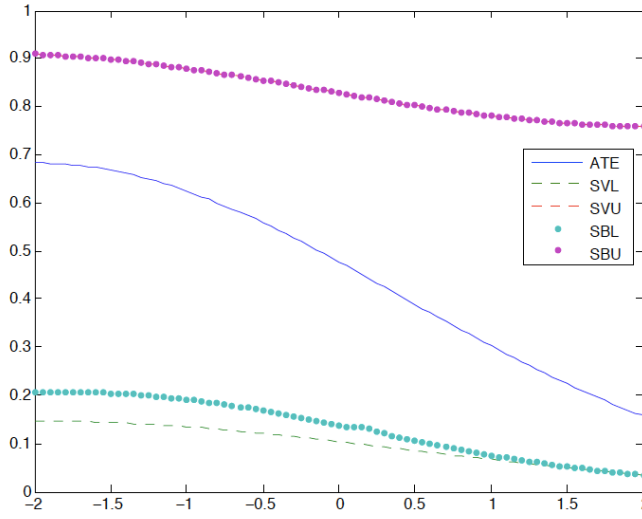
bound improves significantly on the SV lower bound, while the upper bound is exactly the same. Indeed, since  $\nu(0, x') \leq \nu(1, x)$  for many values of  $(x, x')$  our bounds refine the lower bound of  $\nu(1, x)$ , similarly for the upper bound of  $\nu(0, x')$ . However, the bounds do not refine the SV upper bound for two main reasons. There are only few values of  $(x, x')$  such that  $\nu(1, x) \leq \nu(0, x')$  and in the case where it holds it is likely to have  $p(x', z') \geq p(x, z)$ , while we need to have  $p(x', z') \leq p(x, z)$  if we want to refine  $\nu(1, x)$  using  $\nu(0, x')$ .

- (2) In figure 2, I increase the strength of the effect of the instrument  $(\beta = 1, \gamma = \frac{1}{5}, \eta = 4)$ . I note two important facts: First, I am now able to refine the upper bound of  $\nu(1, x)$  using  $\nu(0, x')$

FIGURE 2. Sharp bounds on ATE when  $(\beta = 1, \gamma = \frac{1}{5}, \eta = 4)$ .

since it is likely that  $p(x', 0) \leq p(x, 1)$ . The discontinuity denotes the point where both following conditions start to hold simultaneously :  $\nu(1, x) \leq \nu(0, x')$  and  $p(x', 0) \leq p(x, 1)$ . Second, when the strength of the instrument increases we tend to have identification. This phenomenon is an example of identification at infinity as in Heckman (1990). We can see that the SV bounds do not respect this feature when their support condition fails.

- (3) In figure 3, I decrease the strength of the covariate in the selection equation, to see how the bounds behave when the SV support condition almost holds ( $\beta = 1, \gamma = \frac{1}{100}, \eta = 1$ ). We see that our bounds improve on the SV bounds. Indeed, when we are getting closer to the “perfect matching restriction” our bounds become better in a continuous way, but the SV bounds are not sensitive to that and jump directly to the tightest bounds when  $\gamma = 0$ .
- (4) In figure 4, I reduce the strength of the covariate in the outcome equation, ( $\beta = 1, \gamma = \frac{1}{5}, \eta = \frac{1}{2}$ ), I note that both types of bounds are very wide, and I get only a small improvement over the SV bounds. Indeed, the strength of this present analysis is based on the variation of the covariates. Then, when  $Dom(X)$  is small, our improvement over the SV might be small. This fact explains why Chiburis (2010) found only a small improvement over the SV bounds within his empirical example where the domain of  $X$  was  $\{0, 1\}^2$ .

FIGURE 3. Sharp bounds on ATE when  $(\beta = 1, \gamma = \frac{1}{100}, \eta = 1)$ .FIGURE 4. Sharp bounds on ATE when  $(\beta = 1, \gamma = \frac{1}{5}, \eta = \frac{1}{2})$ .

#### 4. INFERENCE

Although this paper focuses on identification, I will briefly discuss inference in finite samples for our bounds. Even if the bounds derived in Theorem 1 seem involved, the bounds are indeed simple enough to lend themselves to existing inferential methods, specifically Chernozhukov, Lee,

and Rosen (2013). In fact, we have the following:

$$\begin{aligned}
\nu(0, x) &\leq \inf_p \left\{ P(1, 0|x, p) + \min \left( \inf_{\Omega_{01}^+(x)} \inf_{\mathbf{P}^+(x', p)} P(1, 1|x', p'), p \right) \right\} \\
&= \inf_p \left\{ P(1, 0|x, p) + \min \left( \inf_{x' \in \text{Dom}(X)} \inf_{p' \in \text{Supp}(P|X=x')} [P(1, 1|x', p') + 1\{p' > p\} \right. \right. \\
&\quad \left. \left. + 1\{\nu(0, x) > \nu(1, x')\}], p \right) \right\} \\
&= \inf_p \inf_{x' \in \text{Dom}(X)} \inf_{p' \in \text{Supp}(P|X=x')} \left\{ P(1, 0|x, p) + [P(1, 1|x', p') + 1\{p' > p\} \right. \\
&\quad \left. + 1\{\nu(0, x) > \nu(1, x')\}], p \right\}
\end{aligned}$$

The first equality holds since if  $x' \notin \Omega_{01}^+(x)$  or  $p' \notin \mathbf{P}^+(x', p)$  we have

$$\min \left( \inf_{x' \in \text{Dom}(X)} \inf_{p' \in \text{Supp}(P|X=x')} [P(1, 1|x', p') + 1\{p' > p\} + 1\{\nu(0, x) > \nu(1, x')\}], p \right) = p.$$

This reformulation allows to write our bounds as intersection bounds in the sense of Chernozhukov, Lee, and Rosen (2013). Then, their inference procedure can be directly used to construct the confidence set of our bounds.

## 5. CONCLUSION

I have considered the special case of joint threshold crossing model, where no parametric form or distributional assumptions are imposed. I provided sharp bounds on the Average Treatment effect (*ATE*) when I imposed only mild regularity conditions on the distribution of unobservable variables. I presented a methodology which allows to construct sharp bounds on the *ATE* by efficiently using variation on covariates without imposing any support restrictions. A numerical illustration showed that my proposed bounds significantly improve on the Shaikh and Vytlacil (2011) bounds, which where until now, the tightest feasible bounds proposed in the literature for this model. There are several natural extensions of this work. First, this methodology may be easily used to provide sharp bounds for others functionals of treatment, not just the average. Second, this methodology efficiently exploits variation on covariates to sharpen bounds and it may be extended to narrow cross-sectional bounds using time variation in panel data. Finally, it has been extended to provide sharp bounds on the *ATE* in the generalized discrete Roy model.

## APPENDIX A. PROOF OF RESULTS IN THE MAIN TEXT

**Lemma 1.** *Suppose  $\text{Supp}(P|X = x)$  is not a singleton for all  $x \in \text{Dom}(X)$ .*

- (1) *(Lemma 1, SV) Let  $p$  and  $p'$  belong to  $\text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$  such that  $p' < p$ , then*

$$\text{sign}([\nu(1, x') - \nu(0, x)]) = \text{sign}(h(x, x', p, p'))$$

$$\text{where } h(x, x', p, p') = (P(1, 1|x', p) - P(1, 1|x', p')) - (P(1, 0|x, p') - P(1, 0|x, p)).$$

- (2) *Let  $p'_1 < p'_2 \in \text{Supp}(P | X = x')$  and  $p_1 < p_2 \in \text{Supp}(P | X = x)$  such that  $[p'_1, p'_2] \subseteq [p_1, p_2]$  then,*

$$h(1, x, x', p_1, p_2, p'_1, p'_2) \geq 0 \Rightarrow [\nu(1, x') - \nu(0, x)] \geq 0$$

$$\text{where } h(1, x, x', p_1, p_2, p'_1, p'_2) = (P(1, 1|x', p'_2) - P(1, 1|x', p'_1)) - (P(1, 0|x, p_1) - P(1, 0|x, p_2)).$$

- (3) *Let  $p'_1 < p'_2 \in \text{Supp}(P | X = x')$  and  $p_1 < p_2 \in \text{Supp}(P | X = x)$  such that  $[p'_1, p'_2] \subseteq [p_1, p_2]$  then,*

$$h(0, x, x', p_1, p_2, p'_1, p'_2) \geq 0 \Rightarrow [\nu(0, x') - \nu(1, x)] \geq 0$$

$$\text{where } h(0, x, x', p_1, p_2, p'_1, p'_2) = (P(1, 0|x', p'_1) - P(1, 0|x', p'_2)) - (P(1, 1|x, p_2) - P(1, 1|x, p_1)).$$

- (4) *Let  $p$  and  $p'$  belong to  $\text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$  such that  $p' < p$ , then*

$$\text{sign}([\nu(1, x') - \nu(1, x)]) = \text{sign}(\tilde{h}_1(x, x', p, p'))$$

$$\text{where } \tilde{h}_1(x, x', p, p') = (P(1, 1|x', p) - P(1, 1|x', p')) - (P(1, 1|x, p) - P(1, 1|x, p')).$$

- (5) *Let  $p$  and  $p'$  belong to  $\text{Supp}(P | X = x) \cap \text{Supp}(P | X = x')$  such that  $p' < p$ , then*

$$\text{sign}([\nu(0, x') - \nu(0, x)]) = \text{sign}(\tilde{h}_0(x, x', p, p'))$$

$$\text{where } \tilde{h}_0(x, x', p, p') = (P(1, 0|x', p') - P(1, 0|x', p)) - (P(1, 0|x, p') - P(1, 1|x, p)).$$

- (6) *Let  $p'_1 < p'_2 \in \text{Supp}(P | X = x')$  and  $p_1 < p_2 \in \text{Supp}(P | X = x)$  such that  $[p'_1, p'_2] \subseteq [p_1, p_2]$  then,*

$$\tilde{h}_1(x, x', p_1, p_2, p'_1, p'_2) \geq 0 \Rightarrow [\nu(1, x') - \nu(1, x)] \geq 0$$

$$\text{where } \tilde{h}_1(x, x', p_1, p_2, p'_1, p'_2) = (P(1, 1|x', p'_2) - P(1, 1|x', p'_1)) - (P(1, 1|x, p_2) - P(1, 1|x, p_1)).$$

- (7) *Let  $p'_1 < p'_2 \in \text{Supp}(P | X = x')$  and  $p_1 < p_2 \in \text{Supp}(P | X = x)$  such that  $[p'_1, p'_2] \subseteq [p_1, p_2]$  then,*

$$\tilde{h}_0(x, x', p_1, p_2, p'_1, p'_2) \geq 0 \Rightarrow [\nu(0, x') - \nu(0, x)] \geq 0$$

$$\text{where } \tilde{h}_1(x, x', p_1, p_2, p'_1, p'_2) = (P(1, 0|x', p'_1) - P(1, 0|x', p'_2)) - (P(1, 0|x, p_1) - P(1, 1|x, p_2)).$$



*Proof of lemma 1.* I will prove cases (2), (4) and (6). The cases (3), (5), and (7) can be similarly proved.

- Case (2) Let  $p'_1 < p'_2 \in \text{Supp}(P \mid X = x')$  and  $p_1 < p_2 \in \text{Supp}(P \mid X = x)$  such that  $[p'_1, p'_2] \subseteq [p_1, p_2]$ .

$$\begin{aligned} h(1, x, x', p_1, p_2, p'_1, p'_2) &= (P(1, 1|x', p'_2) - P(1, 0|x', p'_1)) - (P(1, 0|x, p_1) - P(1, 1|x, p_2)) \\ &= P(u \leq \nu(1, x'), p'_1 < v < p'_2) - P(u \leq \nu(0, x), p_1 < v < p_2) \\ &\leq P(u \leq \nu(1, x'), p_1 < v < p_2) - P(u \leq \nu(0, x), p_1 < v < p_2). \end{aligned}$$

The last inequality holds since  $[p'_1, p'_2] \subseteq [p_1, p_2]$ . Therefore, if  $h(1, x, x', p_1, p_2, p'_1, p'_2) \geq 0$  then  $P(u \leq \nu(1, x'), p_1 < v < p_2) - P(u \leq \nu(0, x), p_1 < v < p_2) \geq 0$ , which implies that  $[\nu(1, x') - \nu(0, x)] \geq 0$  since we have the following:

$$\begin{aligned} &\left[ P(u \leq \nu(1, x'), p_1 < v < p_2) \right. \\ &\left. - P(u \leq \nu(0, x), p_1 < v < p_2) \right] = \begin{cases} P(\nu(0, x) \leq u \leq \nu(1, x'), p_1 < v < p_2) & \text{if } \nu(1, x') > \nu(0, x) \\ 0 & \text{if } \nu(1, x') = \nu(0, x) \\ -P(\nu(1, x') \leq u \leq \nu(0, x), p_1 < v < p_2) & \text{if } \nu(1, x') < \nu(0, x). \end{cases} \end{aligned}$$

- Case (4) Let  $p$  and  $p'$  belong to  $\text{Supp}(P \mid X = x) \cap \text{Supp}(P \mid X = x')$  such that  $p' < p$ .

$$\begin{aligned} \tilde{h}_d(x, x', p, p') &= (P(1, d|x', p) - P(1, d|x', p')) - (P(1, d|x, p) - P(1, d|x, p')) \\ &= P(u \leq \nu(d, x'), p' < v < p) - P(u \leq \nu(d, x), p' < v < p), \end{aligned}$$

then

$$\tilde{h}_d(x, x', p, p') = \begin{cases} P(\nu(d, x) \leq u \leq \nu(d, x'), p' < v < p) & \text{if } \nu(d, x') > \nu(d, x) \\ 0 & \text{if } \nu(d, x') = \nu(d, x) \\ -P(\nu(d, x') \leq u \leq \nu(d, x), p' < v < p) & \text{if } \nu(d, x') < \nu(d, x). \end{cases}$$

- Case (6) Let  $p'_1 < p'_2 \in \text{Supp}(P \mid X = x')$  and  $p_1 < p_2 \in \text{Supp}(P \mid X = x)$  such that  $[p'_1, p'_2] \subseteq [p_1, p_2]$ .

$$\begin{aligned} \tilde{h}_1(x, x', p_1, p_2, p'_1, p'_2) &= (P(1, 1|x', p'_2) - P(1, 0|x', p'_1)) - (P(1, 1|x, p_2) - P(1, 1|x, p_1)) \\ &= P(u \leq \nu(1, x'), p'_1 < v < p'_2) - P(u \leq \nu(1, x), p_1 < v < p_2) \\ &\leq P(u \leq \nu(1, x'), p_1 < v < p_2) - P(u \leq \nu(1, x), p_1 < v < p_2). \end{aligned}$$

The last inequality holds since  $[p'_1, p'_2] \subseteq [p_1, p_2]$ . Therefore, if  $\tilde{h}_1(x, x', p_1, p_2, p'_1, p'_2) \geq 0$  then  $P(u \leq \nu(1, x'), p_1 < v < p_2) - P(u \leq \nu(1, x), p_1 < v < p_2) \geq 0$ , which implies that

$[\nu(1, x') - \nu(1, x)] \geq 0$  since we have the following:

$$\begin{aligned} & \left[ P(u \leq \nu(1, x'), p_1 < v < p_2) \right. \\ & \left. - P(u \leq \nu(1, x), p_1 < v < p_2) \right] = \begin{cases} P(\nu(1, x) \leq u \leq \nu(1, x'), p_1 < v < p_2) & \text{if } \nu(1, x') > \nu(1, x) \\ 0 & \text{if } \nu(1, x') = \nu(1, x) \\ -P(\nu(1, x') \leq u \leq \nu(1, x), p_1 < v < p_2) & \text{if } \nu(1, x') < \nu(1, x). \end{cases} \end{aligned}$$

This completes our proof.  $\square$

*Proof of proposition 1.* Under assumption 1, if  $\nu(0, X) < \nu(1, X)$  the model (1.1) can be written in the form of a multi-valued mapping  $G_\nu(\cdot, X)$  from unobservable  $u$  to observable  $(Y, D, X, Z)$  in the following way:

$$\begin{aligned} G_\nu : u & \longmapsto (y, d, x, z) \\ [0, \nu(0, x)] & \longmapsto \{(1, 0, x, z); (1, 1, x, z)\} \\ [\nu(0, x), \nu(1, x)] & \longmapsto \{(0, 0, x, z); (1, 1, x, z)\} \\ [\nu(1, x), 1] & \longmapsto \{(0, 1, x, z); (0, 0, x, z)\} \end{aligned}$$

When  $\nu(0, X) > \nu(1, X)$  we can derived an analogous multi-valued mapping:

$$\begin{aligned} G_\nu : u & \longmapsto (y, d, x, z) \\ [0, \nu(1, x)] & \longmapsto \{(1, 0, x, z); (1, 1, x, z)\} \\ [\nu(1, x), \nu(0, x)] & \longmapsto \{(1, 0, x, z); (0, 1, x, z)\} \\ [\nu(0, x), 1] & \longmapsto \{(0, 1, x, z); (0, 0, x, z)\} \end{aligned}$$

Let call  $P$  the distribution of observable variables  $(Y, D) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\} = (\mathcal{Y}, \mathcal{D})$  which can be estimated from the data and  $U$  the uniform distribution on  $[0, 1]$ , of the unobservable variable  $u$ . The model relating outcomes variables  $(Y, D)$  and unobservable variable  $u$  is given by:

$$P(\{(Y, D) \in G_\nu(u \mid X, Z)\}) = 1 \text{ X,Z-as for some } \nu$$

such as  $\nu$  is generally non-unique, which prompts the following definition.

**Definition 1.**  $\nu(d, x) = P(Y_d = 1 \mid X = x)$  belongs to the identified set if and only if  $P(\{(Y, D) \in G_\nu(u \mid X, Z)\}) = 1 \text{ X, Z a.s}$

According to Theorem 1 of Galichon and Henry (2011), the identified set is equal to the set  $\nu$  such that the following inequalities hold:

$$\mathbb{P}(A \mid X, Z) \leq U(\{u \mid G_\nu(u \mid X, Z) \cap A \neq \emptyset\}), \forall A \in 2^{\mathcal{Y}, \mathcal{D}} \quad (\text{A.1})$$

We find exactly 6 non redundant inequalities which can be rewritten as follows:

$$\begin{aligned} P(Y = 1, D = 0 \mid X, Z) &\leq \nu(0, X) \\ P(Y = 0, D = 0 \mid X, Z) &\leq 1 - \nu(0, X) \\ P(Y = 1, D = 1 \mid X, Z) &\leq \nu(1, X) \\ P(Y = 0, D = 1 \mid X, Z) &\leq 1 - \nu(1, X) \\ P(Y = 0 \mid X, Z) &\leq \max(1 - \nu(0, X), 1 - \nu(1, X)) \\ P(Y = 1 \mid X, Z) &\leq \max(\nu(0, X), \nu(1, X)). \end{aligned}$$

□

This completes our proof.

*Proof of theorem 1.*

**Definition 2.** (Nelsen 2006) A two-dimensional subcopula (or briefly subcopula) is a function  $C$  with the following properties:

- (1)  $\text{Domain}(C) = D_1 \times D_2$  where  $D_1$  and  $D_2$  are subsets of  $[0, 1]$  containing 0 and 1.
- (2)  $C(u_1, v_1) - C(u_1, v_2) - C(u_2, v_1) + C(u_2, v_2) \geq 0$ , for all  $u_1, u_2 \in D_1$  and  $v_1, v_2 \in D_2$  such that  $u_1 \geq u_2$  and  $v_1 \geq v_2$ .
- (3)  $C(u, 1) = u$  and  $C(1, v) = v$  for all  $u \in D_1$  and for all  $v$  in  $D_2$ .

**Claim 1.** Under assumptions 4 and 5,  $\nu(d, x) : \{0, 1\} \times \text{Dom}(X) \rightarrow [0, 1]$  is in the identified set if and only if there exists a subcopula  $C$  whose domain is  $S_1 \cup \{0, 1\} \times S_2 \cup \{0, 1\}$  such that:

- (1)  $C(u, 0) = C(0, v) = 0$ , for all  $u \in S_1 \cup \{0, 1\}$  and for all  $v$  in  $S_2 \cup \{0, 1\}$ .
- (2)  $C(\nu(1, x), p) = P(1, 1 \mid x, p)$  and  $C(\nu(0, x), p) = \nu(0, x) - P(1, 0 \mid x, p)$  for all  $(x, p) \in \text{Dom}(X, \text{supp}(P \mid X))$ .

where  $S_2 = \{p(x, z) : (x, z) \in \text{Dom}(x, z)\}$ .

The proof of this claim is given in Chiburis (2010).

In the main text, we prove that the characterization of the identified set derived in the theorem (1) is equivalent to the characterization derived in the corollary (1). Then to prove the theorem (1) I will prove that the bounds proposed in the corollary (1) are sharp. For the sake of simplicity, I shall use in this section the following notation,  $L_0(x, p) = \sup_{\Omega_{01}^-(x)} \sup_{\mathbf{P}^-(x', p)} P(1, 1|x', p')$  and  $M_0(x, p) = \sup_{\Omega_{00}^-(x)} \sup_{\mathbf{P}^-(x', p)} (SL_0(x') - P(1, 0|x', p'))$ . Throughout the main text, we showed that  $\nu(0, x)$  lies inside the following interval  $[SL_0(x), SU_0(x)]$ . To show that these bounds are sharp, it is sufficient to construct a subcopula which respects conditions cited in claim 1 when  $\nu(0, x)$  equals  $SL_0(x)$  or  $SU_0(x)$ . Then, assume that,

$$\begin{aligned} \nu(0, x) &= SL_0(x) \\ &= \sup_p \{P(1, 0|x, p) + \max[L_0(x, p), M_0(x, p)]\}, \end{aligned}$$

where the supremum is taken over  $Supp(P | X)$ . I will now show that the following function is a subcopula on domain  $S_1 \cup \{0, 1\} \times S_2 \cup \{0, 1\}$ :

$$\begin{aligned} C(\nu(1, x), p) &= P(u \leq \nu(1, x), v \leq p) \\ C(\nu(0, x), p) &= -P(u \leq \nu(0, x), v \geq p) + \sup_p \{P(1, 0|x, p) \\ &\quad + \max[L_0(x, p), M_0(x, p)]\}. \end{aligned}$$

By construction, our function verifies properties (1) and (3) of definition 2, it remains to verify property (2). When  $Supp(P \times X) = Supp(P) \times Supp(X)$ , property (2) imposes restrictions on  $\nu(0, x)$  for all  $p, p' \in Supp(P) = Supp(P | X) = Supp(P | X')$ . It's no longer the case when we have  $Supp(P \times X) \neq Supp(P) \times Supp(X)$  because of additional data observability constraints. Indeed,  $C(\nu(1, x), p(x', z')) = P(u \leq \nu(1, x), v \leq p(x', z'))$  cannot be identified from the data when  $p(x', z') \notin Supp(P | X)$ . So, property (2) doesn't always impose additional testable constraints. To clarify this point, consider the two following situations:

- (1)  $Supp(P | X) \cap Supp(P | X') = \emptyset$ ,  $u_1 = \nu(0, x)$ ,  $u_2 = \nu(1, x')$ ,  $v_1 = p(x', z')$  and  $v_2 = p(x, z)$ .

Then property 2 doesn't impose additional restrictions on  $\nu(0, x)$  since we cannot identify

$$C(\nu(1, x), p(x', z')) = P(u \leq \nu(1, x), v \leq p(x', z')).$$

- (2)  $Supp(P | X) \cap Supp(P | X') = \emptyset$ ,  $u_1 = \nu(0, x)$ ,  $u_2 = \nu(1, x')$ ,  $v_1 = p(x, z)$  and  $v_2 = p(x', z')$ .

The only constraint from property 2 is:  $C(\nu(1, x), p(x, z)) \geq C(\nu(1, x'), p(x', z'))$ . I now prove in 2 steps, that our proposed function verifies property (2). Before going over these steps, we need a technical result:

**Claim 2.** For all  $(u_1, u_2, v_1, v_2) \in [0, 1]^4$  such that  $u_1 \geq u_2$  and  $v_1 \geq v_2$ , we have:

$$\begin{aligned} P(u_2 \leq u \leq u_1, v_2 \leq v \leq v_1) &= (P(u \leq u_1, v \leq v_1) + P(u \leq u_2, v \geq v_1)) \\ &\quad - (P(u \leq u_1, v \leq v_2) + P(u \leq u_2, v \geq v_2)) \\ &\geq 0. \end{aligned}$$

**First step :** Let  $p \in \text{Supp}(P \mid X) \cap \text{Supp}(P \mid X') \neq \emptyset$

(1) Let  $(x, x')$  satisfy  $\nu(0, x) \geq \nu(1, x')$ .

$$\begin{aligned} C(\nu(0, x), p) - C(\nu(1, x'), p) &= -(P(u \leq \nu(0, x), v \geq p) + P(u \leq \nu(1, x'), v \leq p)) \\ &\quad + \sup_{p \in \text{Supp}(P \mid X)} \{P(u \leq \nu(0, x), v \geq p) + \max[L_0(x, p), M_0(x, p)]\} \\ &\geq -(P(u \leq \nu(0, x), v \geq p) + P(u \leq \nu(1, x'), v \leq p)) \\ &\quad + \sup_{p \in \text{Supp}(P \mid X)} \{P(u \leq \nu(0, x), v \geq p) + L_0(x, p)\} \\ &\geq -(P(u \leq \nu(0, x), v \geq p) + P(u \leq \nu(1, x'), v \leq p)) \\ &\quad + \sup_{p \in \text{Supp}(P \mid X)} \{P(u \leq \nu(0, x), v \geq p) \\ &\quad + \sup_{\Omega_{01}^-(x)} \sup_{P^-(x^*, p)} P(u \leq \nu(1, x^*), v \leq p^*)\} \\ &\geq 0. \end{aligned}$$

The last inequality holds because  $p \in \text{Supp}(P \mid X)$ ,  $x' \in \Omega_{01}^-(x)$  and  $p \in P^-(x', p)$ . Also,  $C(\nu(0, x), p) - C(\nu(1, x'), p)$  is increasing in  $p$  by the first equality. Indeed, by claim 2  $(P(u \leq \nu(1, x'), v \leq p) + P(u \leq \nu(0, x), v \geq p))$  is decreasing in  $p$  since  $\nu(0, x) \geq \nu(1, x')$ . Then, for all  $p' < p \in \text{Supp}(P \mid X) \cap \text{Supp}(P \mid X')$  we have  $C(\nu(0, x), p) - C(\nu(1, x'), p) \geq C(\nu(0, x), p') - C(\nu(1, x'), p')$ . Thus,  $C(\nu(0, x), p) - C(\nu(1, x'), p) - C(\nu(0, x), p') + C(\nu(1, x'), p') \geq 0$ . So, the property 2 is verified.

(2) Let  $(x, x')$  satisfy  $\nu(0, x) \leq \nu(1, x')$ .

$$\begin{aligned}
C(\nu(0, x), p) - C(\nu(1, x'), p) &= -(P(u \leq \nu(0, x), v \geq p) + P(u \leq \nu(1, x'), v \leq p)) \\
&\quad + \sup_{p \in \text{Supp}(P|X)} \{P(u \leq \nu(0, x), v \geq p) + \max[L_0(x, p), M_0(x, p)]\} \\
&\leq -(P(u \leq \nu(0, x), v \geq p) + P(u \leq \nu(0, x), v \leq p)) \\
&\quad + \sup_{p \in \text{Supp}(P|X)} \{P(u \leq \nu(0, x), v \geq p) + \max[L_0(x, p), M_0(x, p)]\} \\
&\leq -P(u \leq \nu(0, x)) \\
&\quad + \sup_{p \in \text{Supp}(P|X)} \{P(u \leq \nu(0, x), v \geq p) + \max[L_0(x, p), M_0(x, p)]\} \\
&\leq -P(u \leq \nu(0, x)) + SL_0(x) \\
&\leq 0.
\end{aligned}$$

The first inequality holds because  $\nu(0, x) \leq \nu(1, x')$ . Also,  $C(\nu(0, x), p) - C(\nu(1, x'), p)$  is increasing in  $p$  by the first equality. Indeed, by claim 2 ( $P(u \leq \nu(1, x'), v \leq p) + P(u \leq \nu(0, x), v \geq p)$ ) is increasing in  $p$  since  $\nu(0, x) \leq \nu(1, x')$ .

(3) Let  $(x, x')$  satisfy  $\nu(0, x) \geq \nu(0, x')$

$$\begin{aligned}
C(\nu(0, x), p) - C(\nu(0, x'), p) &= -(P(u \leq \nu(0, x), v \geq p) - P(u \leq \nu(0, x'), v \geq p)) \\
&\quad + \sup_{p \in \text{Supp}(P|X)} \{P(u \leq \nu(0, x), v \geq p) + \max[L_0(x, p), M_0(x, p)]\} \\
&\quad - \sup_{p' \in \text{Supp}(P|X')} \{P(u \leq \nu(0, x'), v \geq p') + \max[L_0(x', p'), M_0(x', p')]\} \\
&\geq -(P(u \leq \nu(0, x), v \geq p) - P(u \leq \nu(0, x'), v \geq p)) \\
&\quad + \sup_{p \in \text{Supp}(P|X)} \{P(u \leq \nu(0, x), v \geq p) + M_0(x, p)\} - SL_0(x') \\
&\geq -(P(u \leq \nu(0, x), v \geq p) - P(u \leq \nu(0, x'), v \geq p)) \\
&\quad + \sup_{p \in \text{Supp}(P|X)} \{P(u \leq \nu(0, x), v \geq p) \\
&\quad + \sup_{\Omega_{00}^-(x)} \sup_{P^-(x^*, p)} (SL_0(x^*) - P(u \leq \nu(0, x^*), v \geq p^*))\} \\
&\quad - SL_0(x') \\
&\geq -(P(u \leq \nu(0, x), v \geq p) - P(u \leq \nu(0, x'), v \geq p)) \\
&\quad + \sup_{p \in \text{Supp}(P|X)} \{P(u \leq \nu(0, x), v \geq p) - P(u \leq \nu(0, x'), v \geq p)\} \\
&\geq 0.
\end{aligned}$$

The fourth inequality holds because  $x' \in \Omega_{00}^-(x)$  and  $p \in \mathbf{P}^-(x', p)$ . Also,  $C(\nu(0, x), p) - C(\nu(0, x'), p)$  is increasing in  $p$  by the first equality.

(4) Let  $(x, x')$  satisfy  $\nu(0, x) \leq \nu(0, x')$

by interchanging  $x$  by  $x'$  in point (3) we obviously get that  $C(\nu(0, x), p) - C(\nu(0, x'), p)$  is decreasing in  $p$  and greater than 0.

**Second step :**  $Supp(P \mid X) \cap Supp(P \mid X') = \emptyset$

(1)  $\nu(0, x) \geq \nu(1, x')$  and  $p(x, z) \geq p(x', z')$

$$\begin{aligned} C(\nu(0, x), p) - C(\nu(1, x'), p') &\geq C(\nu(0, x), p) - C(\nu(1, x'), p) \\ &\geq 0. \end{aligned}$$

The last inequality holds by point (1) of first step.

(2)  $\nu(0, x) \leq \nu(1, x')$  and  $p(x, z) \leq p(x', z')$

$$\begin{aligned} C(\nu(0, x), p) - C(\nu(1, x'), p') &\leq C(\nu(0, x), p) - C(\nu(1, x'), p) \\ &\leq 0. \end{aligned}$$

The last inequality holds by point (2) of first step.

(3)  $\nu(0, x) \geq \nu(0, x')$  and  $p(x, z) \geq p(x', z')$

$$\begin{aligned} C(\nu(0, x), p) - C(\nu(0, x'), p) &\geq C(\nu(0, x), p) - C(\nu(0, x'), p) \\ &\geq 0. \end{aligned}$$

The last inequality holds by point (3) of first step.

(4)  $\nu(0, x) \leq \nu(0, x')$

$$\begin{aligned} C(\nu(0, x), p) - C(\nu(0, x'), p) &\geq C(\nu(0, x), p) - C(\nu(0, x'), p) \\ &\geq 0. \end{aligned}$$

The last inequality holds by point (4) of first step.

Then the property (2) holds. We can proceed in the same way for  $\nu(0, x) = SU_0(x)$ . This completes our proof.  $\square$

## REFERENCES

- ANDREWS, D., AND X. SHI (2013): “Nonparametric inference based on conditional moment inequalities,” unpublished manuscript.
- BATTISTIN, E., AND E. RETTORE (2002): “Testing for Programme Effects in a Regression Discontinuity Design with Imperfect Compliance,” *Journal of the Royal Statistical Society*, 165(1), 39–57.
- BEFFY, M., D. FOUGÈRE, AND A. MAUREL (2010): “The Effect of Part-Time Work on Post-Secondary Educational Attainment: New Evidence from French Data,” unpublished manuscript.
- BHATTACHARYA, J., A. SHAIKH, AND E. VYTLACIL (2008): “Treatment effect bounds under monotonicity conditions: An application to Swan-Ganz catheterization,” *American Economic Review*, 98, 351–356.
- BLUNDELL, R., A. GOSLING, H. ICHIMURA, AND C. MEGHIR (2007): “Changes in the distribution of male and female wages accounting for employment composition using bounds,” *Econometrica*, 75, 323–363.
- CANTON, E., AND A. BLOM (2004): “Can Student Loans Improve Accessibility to Higher Education and Student Performance? An Impact Study of the Case of SOFES, Mexico,” unpublished manuscript.
- CARPENTER, C., AND C. DOBKIN (2009): “The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age,” *American Economic Journal: Applied Economics*, 1, 164–182.
- CHERNOZHUKOV, V., AND C. HANSEN (2005): “An iv model of quantile treatment effects,” *Econometrica*, 73(1), 245–261.
- CHERNOZHUKOV, V., S. LEE, AND A. ROSEN (2013): “Intersection bounds: Estimation and Inference,” forthcoming in *Econometrica*.
- CHESHER, A. (2003): “Identification in nonseparable models,” *Econometrica*, 71, 1405–1441.
- (2005): “Nonparametric identification under discrete variation,” *Econometrica*, 73, 1525–1550.
- CHIBURIS, R. C. (2010): “Semiparametric Bounds on Treatment effect,” *Journal of Econometrics*, 159, 267–275.
- FAN, Y., AND S. S. PARK (2010): “Sharp bounds on the distribution of treatment effects and their statistical inference,” *Econometric Theory*, 26, 931–951.



- FAN, Y., AND J. S. WU (2010): “Partial identification of the distribution treatment effects in switching regime models and its confidence sets,” *the Review of Economic Studies*, 77, 1002–1041.
- GALICHON, A., AND M. HENRY (2011): “Set identification in models with multiple equilibria,” *Review of Economic Studies*, 78(4), 1264–1298.
- HECKMAN, J. (1990): “Varieties of selection bias,” *American Economic Review*, 80, 313–318.
- HENRY, M., AND I. MOURIFIÉ (2012): “Sharp bounds in the binary rox model,” unpublished manuscript.
- IMBENS, G., AND T. LEMIEUX (2008): “Regression discontinuity designs: A guide to practice,” *Journal of Econometrics*, 142, 615–635.
- IMBENS, G., AND W. NEWEY (2009): “Identification and estimation of triangular simultaneous equations models without additivity,” *Econometrica*, 77, 1481–1512.
- JUN, S., J. PINKSE, AND H. XU (2010): “Tighter bounds in triangular systems,” *Journal of Econometrics*, 161(2), 122–128.
- (2011): “Discrete Endogenous Variables in Weakly Separable Models,” unpublished manuscript.
- KASY, M. (2011): “Identification in triangular systems using control functions,” forthcoming in *Econometric Theory*.
- KLAAUW, W. V. D. (2002): “Estimating the effect of financial aid offers on college enrollement: a regression-discontinuity approach,” *International Economic Review*, 43, 1249–1287.
- LEE, D. (2008): “Randomized experiments from non-random selection in U.S. House elections,” *Journal of Econometrics*, 142, 675–697.
- LEE, D., AND T. LEMIEUX (2009): “Regression discontinuity designs in economics,” unpublished manuscript.
- MANSKI, C. (1988): “Identification of Binary Response Models,” *Journal of the American Statistical Association*, 83(403), 729–738.
- (1989): “Anatomy of the selection problem,” *Journal of Human resources*, 24, 243–360.
- (1990): “Nonparametric bounds on treatment effects,” *American Economic Review*, 80, 319–323.
- MANSKI, C., AND J. PEPPER (2000): “Monotone instrumental variables with an application to the returns to schooling,” *Econometrica*, 68, 997–1010.
- NELSEN, R. B. (2006): *An introduction to copulas*. Springer, New york.

- SHAIKH, A., AND E. VYTLACIL (2011): “Partial identification in triangular systems of equations with binary dependent variables,” *Econometrica*, 79, 949–955.
- VYTLACIL, E., AND N. YILDIZ (2007): “Dummy Endogenous Variables in Weakly Separable Models,” *Econometrica*, 75, 757–779.