

Partial Identification in Triangular Systems of Equations with Binary Dependent Variables *

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

Edward J. Vytlačil
Department of Economics
Yale University
edward.vytlacil@yale.edu

July 15, 2010

Abstract

This paper studies models for binary outcome variables that contain a binary endogenous regressor. More specifically, we consider a nonparametric, triangular system of equations with binary dependent variables. The main assumption we impose is a weak separability condition on each equation, or, equivalently, a threshold crossing model on each equation. In this setting, we construct upper and lower bounds on the Average Structural Function (ASF) and the Average Treatment Effect (ATE) under weak regularity conditions. The resulting bounds are narrower the greater the strength of the instrument and the greater the degree to which the exogenous covariates that enter the outcome equation can compensate for variation in the endogenous regressor. We show further that the bounds on the ASF and ATE are sharp under an additional restriction on the support of the covariates and the instrument.

JEL Codes: C14, C35

KEYWORDS: Partial Identification, Simultaneous Equation Model, Binary Dependent Variable, Endogeneity, Threshold Crossing Model, Weak Separability, Average Structural Function, Average Treatment Effect

ACKNOWLEDGEMENTS: We would like to thank Hide Ichimura, Jim Heckman, Whitney Newey, and Jim Powell for very helpful comments on this paper. This research was conducted in part while Edward Vytlačil was in residence at Hitotsubashi University. This research was supported by NSF SES-05-51089 and DMS-0820310.

*An earlier version of this paper titled “Threshold Crossing Models and Bounds on Treatment Effects: A Non-parametric Analysis” appeared in May 2005 as NBER Technical Working Paper 307.

1 Introduction

This paper studies models for binary outcome variables that contain a binary endogenous regressor. More specifically, we consider a nonparametric, triangular system of equations with binary dependent variables. The main assumption we impose is a weak separability condition on each equation, or, equivalently, a threshold crossing model on each equation. This structure nests the bivariate probit model with structural shift of Heckman (1978) as a special case. In this setting, we consider the problem of partially identifying the Average Structural Function (ASF) and the Average Treatment Effect (ATE), thereby extending the identification results of Vytlacil and Yildiz (2007).

In order to define this structure precisely, let D denote the binary endogenous regressor and let Y denote the outcome of interest. For example, D might denote receipt of job training and Y later employment, or D might denote receipt of a medical intervention and Y later mortality. See Bhattacharya et al. (2009) for an application of the methodology developed in this paper to the evaluation of the impact of Swan-Ganz catheterization on patient mortality. Consider the following triangular system of equations:

$$\begin{aligned} Y &= g_1(D, X, \epsilon_1) \\ D &= g_2(Z, \epsilon_2) . \end{aligned} \tag{1}$$

Here, X and Z are observed random vectors that may share elements in common, and ϵ_1 and ϵ_2 are unobserved random variables. Following Blundell and Powell (2004), our object of interest is the Average Structural Function (ASF)

$$G_1(d, x) = \int g_1(d, x, \epsilon_1) dF_{\epsilon_1} ,$$

where (d, x) denotes a potential realization of the random vector (D, X) . The ASF averages against the unconditional distribution of ϵ_1 , not the distribution of ϵ_1 conditional on the possibly endogenous regressor D , and thus gives the expected outcome of Y if D were determined exogenously. We also consider

$$\Delta G_1(x) = G_1(1, x) - G_1(0, x) ,$$

which is often referred to as the Average Treatment Effect (ATE) in the treatment effect literature.

The main assumption we impose is that g_1 and g_2 both satisfy weak separability of the observed regressors from the unobserved error term. As will be further discussed in Section 2, for a binary dependent variable, such an assumption is equivalent to assuming that the function is weakly increasing in the error term, as in Chesher (2005), assuming the monotonicity restriction considered by Imbens and Angrist (1994), or assuming that the model can be represented as a threshold crossing model with an additively separable latent error, as in Heckman and Vytlacil (2005). For ease of

analysis, we will work with the threshold crossing representation of the model, i.e.,

$$\begin{aligned} Y &= I\{\nu_1(D, X) \geq \epsilon_1\} \\ D &= I\{\nu_2(Z) \geq \epsilon_2\} . \end{aligned} \tag{2}$$

If one assumes that ν_1 and ν_2 are linear functions and that (ϵ_1, ϵ_2) has a bivariate normal distribution, then the above model reduces to the classical bivariate probit with structural shift considered in Heckman (1978). We will not impose any such parametric functional form or parametric distributional assumptions in this paper.

In addition to the weak separability assumption described above, we will require some mild regularity of the distribution of (ϵ_1, ϵ_2) . We will also assume that X and Z are exogenous in the sense that $(X, Z) \perp\!\!\!\perp (\epsilon_1, \epsilon_2)$. Note that D may still be endogenous in the Y equation due to possible dependence between ϵ_1 and ϵ_2 . For example, those who receive the job training might have the worst human capital, or those who receive the medical intervention might have the worst latent health. The resulting bounds on the ASF or ATE are substantially narrower than alternative bounds that do not impose our weak separability restrictions. Under certain restrictions on the distribution of (X, Z) and the functions ν_1 and ν_2 in (2), we show further that the bounds we derive on the ASF and ATE are sharp in the sense that for any value lying between the upper and lower bounds, there will exist a distribution of unobservable variables satisfying all of the assumptions of our analysis that is consistent with both the distribution of the observed data and the proposed value of the ASF or the ATE.

Identification of the ASF and ATE with this structure was previously considered by Vytlacil and Yildiz (2007). They show that when the support of the distribution of X conditional on $\Pr\{D = 1|Z\}$ is sufficiently rich it is possible to point identify the ASF and the ATE. Their support condition will fail if, for example, X is a discrete random variable, and would be expected to fail near the boundaries of the support of X if X has bounded support. In this paper, we investigate what can be inferred about the ASF or the ATE without imposing this support restriction. To this end, we first use a modified instrumental variable-like procedure to determine what variation in X over-compensates or under-compensates for *ceteris paribus* variation in D , and then use this information to construct bounds on the ASF or the ATE. The resulting bounds are smaller the greater the variation there is in X conditional on $\Pr\{D = 1|Z\}$, and collapse to point identification under the Vytlacil and Yildiz (2007) condition of sufficient variation in X conditional on $\Pr\{D = 1|Z\}$.

As mentioned earlier, our weak separability restriction on the functions g_1 and g_2 is equivalent to imposing that the functions are weakly increasing in the error terms ϵ_1 and ϵ_2 , respectively. We do not impose the stronger requirement that either function is strictly increasing in its error term, as to do so would imply under our regularity conditions on the distribution of (ϵ_1, ϵ_2) that Y or D must be continuous. For this reason, we cannot follow the control variate-approach used, e.g., in Altonji

and Matzkin (2005), Blundell and Powell (2004), Chesher (2003), and Imbens and Newey (2010), which would require g_2 to be strictly increasing in ϵ_2 . Similarly, we cannot follow the quantile instrumental variable-approach used in Chernozhukov and Hansen (2005) and Chernozhukov et al. (2007), which would require g_1 to be strictly increasing in ϵ_1 .

Our analysis is similar to Chesher (2005), who only assumes that g_1 and g_2 are weakly increasing in ϵ_1 and ϵ_2 , respectively. In his analysis, the object of interest is g_1 itself, while in this paper we focus more modestly on the ASF and the ATE. More importantly, his analysis requires a rank condition that cannot hold except in trivial cases when D is binary. When D is binary, the rank condition under which he constructs bounds for $g_1(0, x, \tau)$ is that there exists some value z_0 such that $\Pr\{D = 1|Z = z_0\} \leq \tau \leq 0$ and for $g_1(1, x, \tau)$ that there exists some value z_0 such that $1 \leq \tau \leq \Pr\{D = 1|Z = z_0\}$. These conditions cannot hold for any value of τ except $\tau = 0$ or $\tau = 1$, in which case the ASF is identified following arguments in Heckman and Vytlačil (2001). See Jun et al. (2009) for extensions of his analysis and Chesher (2007) for related analysis that considers partial identification of g_1 without imposing any restrictions on g_2 .

The analysis of this paper has recently been extended in subsequent work by Chiburis (2009). While we show that our bounds are sharp whenever the support of (X, Z) may be written as the product of the support of X and the support of Z , Chiburis (2009) shows that our bounds may not be sharp without this restriction. On the other hand, he presents numerical evidence that suggests that our bounds will often be close to the sharp bounds even when this restriction fails. Moreover, our bounds are much simpler to describe than the sharp bounds derived in Chiburis (2009). Chiburis (2009) also considers restrictions beyond what we impose, such as linear latent index restrictions and parametric distributional assumptions.

The remainder of the paper is organized as follows. In Section 2, we formally define our assumptions and analyze the connection between our assumptions and the assumptions considered in the previous literature. Our main results are contained in Section 3. We conclude with a numerical example in Section 4.

2 Model and Assumptions

In addition to assuming that Y and D are determined by (2), we will make use of the following assumptions in our analysis:

Assumption 2.1 $(X, Z) \perp\!\!\!\perp (\epsilon_1, \epsilon_2)$.

Assumption 2.2 The distribution of (ϵ_1, ϵ_2) has strictly positive density w.r.t. Lebesgue measure on \mathbf{R}^2 .

Assumption 2.3 The support of the distribution of (X, Z) , $\text{supp}(X, Z)$, is compact.

Assumption 2.4 The functions $\nu_1(\cdot)$, and $\nu_2(\cdot)$ are continuous.

Assumption 2.5 The distribution of $\nu_2(Z)|X$ is nondegenerate.

In the derivation of our bounds, we will exploit the assumption that Y and D are determined by (2) and Assumptions 2.1 - 2.2. Formally, our analysis will not require Assumption 2.5, but when it fails our bounds will reduce to those of Manski (1989), who imposes no structure on the equations determining Y and D . In this sense, though formally our results will not require a variable in Z that is not in X , they will be nontrivial only when there is a variable in Z that is not contained in X . When this is the case, any regressor in X that is not in Z will provide an additional source of identifying power in our analysis. We will make use of Assumptions 2.3 and 2.4 only when arguing that the bounds are sharp.

As discussed in Vytlačil (2006), the existence of a threshold crossing representation with an additive latent error as in (2) is equivalent to several other nonparametric monotonicity conditions considered in the literature. In fact, by combining results from the previous literature, we have the following lemma:

Lemma 2.1 *For $f : \mathcal{W} \times \mathcal{E} \mapsto \{0, 1\}$, where $\mathcal{W} \subseteq \mathbf{R}^{K_W}$, $\mathcal{E} \subseteq \mathbf{R}^{K_E}$, the following statements are equivalent:*

- (i) *For any $w, \tilde{w} \in \mathcal{W}$, $f(w, e^*) > f(\tilde{w}, e^*)$ for some $e^* \in \mathcal{E} \Rightarrow f(w, e) \geq f(\tilde{w}, e)$ for all $e \in \mathcal{E}$.*
- (ii) *There exists a function $\nu : \mathcal{E} \mapsto \mathbf{R}$ with range $\mathcal{R}(\nu)$ and a function $g : \mathcal{W} \times \mathcal{R}(\nu) \mapsto \mathbf{R}$ with $g(w, t)$ weakly increasing in t such that $f(w, e) = g(w, \nu(e))$ for all $(w, e) \in \mathcal{W} \times \mathcal{E}$.*
- (iii) *There exists a function $\nu : \mathcal{W} \mapsto \mathbf{R}$ with range $\mathcal{R}(\nu)$ and a function $g : \mathcal{R}(\nu) \times \mathcal{E} \mapsto \mathbf{R}$ with $g(t, e)$ weakly increasing in t such that $f(w, e) = g(\nu(w), e)$ for all $(w, e) \in \mathcal{W} \times \mathcal{E}$.*
- (iv) *There exists a function $\nu : \mathcal{W} \mapsto \mathbf{R}$ and a function $\lambda : \mathcal{E} \mapsto \mathbf{R}$ such that $f(w, e) = I\{\nu(w) \geq \lambda(e)\}$ for all $(w, e) \in \mathcal{W} \times \mathcal{E}$.*

PROOF: The equivalence between (i) and (iii) follows from Theorem C.1 of Vytlačil and Yildiz (2007). The equivalences between (i) and (iv) and between (ii) and (iv) follow from straightforward modifications to the proof of Theorem 1 of Vytlačil (2002). ■

Restriction (i) in Lemma 2.1 is imposed on the model for D by Imbens and Angrist (1994). Imbens and Angrist (1994) refer to this restriction as “monotonicity,” whereas Heckman and Vytlačil (2005) refer to it as a uniformity condition. This restriction on D without the corresponding

restriction on Y does not narrow the resulting sharp bounds on the ASF or ATE compared to not imposing this structure on D . See Balke and Pearl (1997) for the case of binary Z and Heckman and Vytlacil (2001) for the case of general Z . In comparison, in this paper, we impose the structure symmetrically on both Y and D and show that the resulting bounds on the ASF and ATE are much narrower than those obtained when not imposing this structure.

Restriction (ii) in Lemma 2.1 imposes that e is weakly separable from w , while restriction (iii) imposes that w is weakly separable from e . In general, e weakly separable from w is not equivalent to w weakly separable from e , though these restrictions are equivalent for binary valued functions. Chesher (2005) imposes restriction (ii) on the structural equations for both Y and D and studies partial identification of the equation determining Y instead of the ASF or the ATE, as we do in this paper. More importantly, as discussed in Section 1, when D is binary, the rank condition required in his analysis will not hold except in trivial cases. Vytlacil and Yildiz (2007) impose restriction (iii) on the structural equations for both Y and D and show that the ASF and the ATE are identified whenever the support of X conditional on $\Pr\{D = 1|Z\}$ is sufficiently rich. In this paper, we study partial identification of the ASF and the ATE without such a restriction on the support of (X, Z) .

Restriction (iv) in Lemma 2.1 is a threshold crossing model with an additively separable latent error. It is imposed on the equation determining D by Heckman and Vytlacil (2005), though they do not impose the corresponding structure on Y . Therefore, as discussed above, the resulting bounds on the ASF and the ATE in their setting are the same as those obtained when not imposing this structure on D .

Finally, note that these restrictions are not equivalent to the monotone treatment response (MTR) restrictions of Manski and Pepper (2000). MTR requires that the equation determining Y is weakly monotonic in D , with the direction of the monotonicity known *a priori* and the direction not depending on $(X, \epsilon_1, \epsilon_2)$. In other words, MTR requires that D is weakly separable from (X, ϵ_1) and that the direction of the monotonicity in D is known *a priori*. See Bhattacharya et al. (2008) for further discussion.

3 Identification Analysis

We begin by noting that it follows from Assumptions 2.1 - 2.2 that we may without loss of generality impose the normalization that $\epsilon_2 \sim \text{Uniform}[0, 1]$ and $\nu_2(Z) = P(Z) = \Pr\{D = 1|Z\}$. We may sometimes write P in place of $P(Z)$. After such a normalization, Assumption 2.2 becomes the requirement that the distribution of (ϵ_1, ϵ_2) has a strictly positive density w.r.t. Lebesgue measure on $\mathbf{R} \times [0, 1]$. Furthermore, note that Assumptions 2.1- 2.4 imply that P is bounded away from 0

and 1. We will henceforth work with the normalized model.

We will use the following notation from the treatment effect literature:

$$Y_d \equiv g_1(d, X, \epsilon_1) = I\{\nu_1(d, X) \geq \epsilon_1\} .$$

With this notation, and using that $X \perp\!\!\!\perp \epsilon_1$ by Assumption 2.1, the ASF and the ATE can be expressed as

$$\begin{aligned} G_1(d, x) &= E[Y_d|X = x] = \Pr\{Y_d = 1|X = x\} , \\ \Delta G_1(x) &= E[Y_1 - Y_0|X = x] = \Pr\{Y_1 = 1|X = x\} - \Pr\{Y_0 = 1|X = x\} . \end{aligned}$$

For the identification analysis, we assume that the distribution of the observed data, (Y, X, D, Z) , is known. Consider first identification of $\Pr\{Y_1 = 1|X = x\} = G_1(1, x)$. By equation (2) and Assumption 2.1, we have that

$$\Pr\{Y_1 = 1|X\} = \Pr\{Y_1 = 1|X, P(Z)\}$$

and

$$\Pr\{D = 1|X, P(Z)\} = P(Z) .$$

By the definition of Y_1 , we have that the event $\{D = 1, Y = 1\}$ is the same event as $\{D = 1, Y_1 = 1\}$. Thus,

$$\begin{aligned} \Pr\{Y_1 = 1|X, P(Z)\} &= \Pr\{D = 1, Y_1 = 1|X, P(Z)\} + \Pr\{D = 0, Y_1 = 1|X, P(Z)\} \\ &= \Pr\{D = 1, Y = 1|X, P(Z)\} + (1 - P(Z)) \Pr\{Y_1 = 1|X, P(Z), D = 0\} . \end{aligned}$$

The terms $P(Z)$ and $\Pr\{D = 1, Y = 1|X, P(Z)\}$ are identified from the distribution of (Y, D, X, Z) , but the term $\Pr\{Y_1 = 1|X, P(Z), D = 0\}$ is not identified from the distribution of (Y, D, X, Z) . Since Y is binary, this unidentified term is bounded from above and below by one and zero, which immediately implies bounds on $\Pr\{Y_1 = 1|X\}$,

$$\Pr\{D = 1, Y_1 = 1|X, P(Z)\} \leq \Pr\{Y_1 = 1|X\} \leq \Pr\{D = 1, Y_1 = 1|X, P(Z)\} + (1 - P(Z)) .$$

Since $\Pr\{Y_1 = 1|X\}$ does not depend on $P(Z)$, we can take the supremum of the lower bounds and the infimum of the upper bounds over values of $P(Z)$. Furthermore, using the threshold crossing structure of the equation determining D , Heckman and Vytlacil (2001) show that the supremum of the lower bounds and the infimum of the upper bounds are both achieved at the largest value in the support of $P(Z)$ conditional on X . Parallel reasoning provides bounds on $\Pr\{Y_0 = 1|X = x\}$. Denote by $\text{supp}(P|X)$ the support of $P(Z)$ conditional on X . We have the following lemma, adapted from Heckman and Vytlacil (2001):

Lemma 3.1 (Heckman and Vytlacil (2001)) Suppose that Y and D are determined by (2) and that Assumptions 2.1 - 2.2 hold. Then, whenever all conditional probabilities are well defined, we have the following:

$$\begin{aligned} \sup_{p \in \text{supp}(P|X)} \Pr\{D = 1, Y = 1|X, P(Z) = p\} \\ \leq \Pr\{Y_1 = 1|X\} \leq \inf_{p \in \text{supp}(P|X)} \{\Pr\{D = 1, Y = 1|X, P(Z) = p\} + (1 - p)\} \end{aligned}$$

$$\begin{aligned} \sup_{p \in \text{supp}(P|X)} \Pr\{D = 0, Y = 1|X, P(Z) = p\} \\ \leq \Pr\{Y_0 = 1|X\} \leq \inf_{p \in \text{supp}(P|X)} \{\Pr\{D = 0, Y = 1|X, P(Z) = p\} + p\} , \end{aligned}$$

and these expressions can be simplified to

$$\Pr\{D = 1, Y = 1|X, P(Z) = \bar{p}_X\} \leq \Pr\{Y_1 = 1|X\} \leq \Pr\{D = 1, Y = 1|X, P(Z) = \bar{p}_X\} + (1 - \bar{p}_X)$$

$$\Pr\{D = 0, Y = 1|X, P(Z) = \underline{p}_X\} \leq \Pr\{Y_0 = 1|X\} \leq \Pr\{D = 0, Y_1 = 1|X, P(Z) = \underline{p}_X\} + \underline{p}_X ,$$

where

$$\begin{aligned} \underline{p}_X &= \inf\{p : p \in \text{supp}(P|X)\} \\ \bar{p}_X &= \sup\{p : p \in \text{supp}(P|X)\} . \end{aligned}$$

PROOF: Follows from Theorem 2 of Heckman and Vytlacil (2001). ■

Heckman and Vytlacil (2001) impose our weak separability restriction only on the equation determining D . They show that the bounds of Lemma 3.1 cannot be improved upon without imposing more structure. By imposing the same weak separability restriction on the Y equation as on the D equation, we can considerably reduce the width of the bounds. The next lemma is central to our ability to improve upon Heckman and Vytlacil (2001). It uses the weak separability restrictions on the equations determining Y and D together with the other assumptions of our analysis to determine the sign of $\nu_1(1, x') - \nu_1(0, x)$ from a modified instrumental variables-like term that is identified from the distribution of the observed data. Depending on the sign of $\nu_1(1, x') - \nu_1(0, x)$, we will then be able to bound $\Pr\{Y_1 = 1|D = 0, X = x, P = p\}$ and $\Pr\{Y_0 = 1|D = 1, X = x, P = p\}$ from above or below by terms other than one or zero that are identified from the distribution of the observed data. In other words, we first determine whether variation in X over-compensates or under-compensates for *ceteris paribus* variation in D and use this information in the construction of our bounds.

Lemma 3.2 *Suppose that Y and D are determined by (2) and that Assumptions 2.1 - 2.2 hold. Let*

$$h(x, x', p, p') = (\Pr\{D = 1, Y = 1|X = x', P = p\} - \Pr\{D = 1, Y = 1|X = x', P = p'\}) \\ - (\Pr\{D = 0, Y = 1|X = x, P = p'\} - \Pr\{D = 0, Y = 1|X = x, P = p\}) .$$

Then, whenever all conditional probabilities are well defined, we have for $p > p'$ that $h(x, x', p, p')$ and $\nu_1(1, x') - \nu_1(0, x)$ share the same sign. In particular, the sign of $h(x, x', p, p')$ does not depend on p or p' provided $p > p'$.

PROOF: See Appendix. ■

Before proceeding with the statement of the main theorem, we illustrate the use of Lemma 3.2 in characterizing the possible values for $\Pr\{Y_1 = 1|D = 0, X = x, P = p\}$ and $\Pr\{Y_0 = 1|D = 1, X = x, P = p\}$. Denote by P' a random variable distributed independently of P with the same distribution as P . Define

$$H(x, x') = E[h(x, x', P, P')|P > P'] , \quad (3)$$

where $h(x, x', p, p') = 0$ whenever it is not well defined. Suppose there exists $p > p'$ for which $h(x, x', p, p')$ is well defined, i.e., $p > p'$ with both p and p' in $\text{supp}(P|X = x) \cap \text{supp}(P|X = x')$. Recall that the sign of $h(x, x', p, p')$ does not depend on p or p' provided $p > p'$. If $H(x, x') \geq 0$, then it follows from Lemma 3.2 that $\nu_1(1, x') \geq \nu_1(0, x)$. Therefore,

$$\begin{aligned} \Pr\{Y_0 = 1|D = 1, X = x, P = p\} &= \Pr\{\epsilon_1 \leq \nu_1(0, X)|D = 1, X = x, P = p\} \\ &\leq \Pr\{\epsilon_1 \leq \nu_1(1, X)|D = 1, X = x', P = p\} \\ &= \Pr\{Y = 1|D = 1, X = x', P = p\} , \end{aligned}$$

where the first and third equalities follow from equation (2), and the inequality follows from the fact that $\nu_1(1, x') \geq \nu_1(0, x)$ and Assumption 2.2. If, on the other hand, $H(x, x') \leq 0$, then we can argue along similar lines to bound $\Pr\{Y_0 = 1|D = 1, X = x, P = p\}$ from below by $\Pr\{Y = 1|D = 1, X = x', P = p\}$. We can thus bound the unidentified terms $\Pr\{Y_0 = 1|D = 1, X = x, P = p\}$ and $\Pr\{Y_1 = 1|D = 0, X = x, P = p\}$ by lower and upper bounds that differ from zero and one.

We now state our main theorem, which relies critically on Lemma 3.2. In the statement of the theorem, it is understood that all supremums and infimums are only taken over regions where all conditional probabilities are well defined, and we adopt the convention that the supremum over the empty set is zero and the infimum over the empty set is one.

Theorem 3.1 Suppose that Y and D are determined by (2). Let

$$\begin{aligned}\mathbf{X}_{0+}(x) &= \{x' : H(x, x') \geq 0\} \\ \mathbf{X}_{0-}(x) &= \{x' : H(x, x') \leq 0\} \\ \mathbf{X}_{1+}(x) &= \{x' : H(x', x) \geq 0\} \\ \mathbf{X}_{1-}(x) &= \{x' : H(x', x) \leq 0\} ,\end{aligned}$$

where $H(x, x')$ is defined in (3) if $h(x, x', p, p')$ is well defined for some $p > p'$, and with each set understood to be empty if $h(x, x', p, p')$ is not well defined for any $p > p'$. Then, we have the following:

(i) If Assumptions 2.1 - 2.2 hold, then $G_1(d, x) \in [L_d(x), U_d(x)]$ for $d \in \{0, 1\}$, and $\Delta G_1(x) \in [L_\Delta(x), U_\Delta(x)]$, where $L_\Delta(x) = L_1(x) - U_0(x)$, $U_\Delta(x) = U_1(x) - L_0(x)$, and

$$\begin{aligned}L_0(x) &= \sup_p \{\Pr\{D = 0, Y = 1 | X = x, P = p\} + \sup_{x' \in \mathbf{X}_{0-}(x)} \Pr\{D = 1, Y = 1 | X = x', P = p\}\} , \\ L_1(x) &= \sup_p \{\Pr\{D = 1, Y = 1 | X = x, P = p\} + \sup_{x' \in \mathbf{X}_{1+}(x)} \Pr\{D = 0, Y = 1 | X = x', P = p\}\} , \\ U_0(x) &= \inf_p \{\Pr\{D = 0, Y = 1 | X = x, P = p\} + p \inf_{x' \in \mathbf{X}_{0+}(x)} \Pr\{Y = 1 | D = 1, X = x', P = p\}\} , \\ U_1(x) &= \inf_p \{\Pr\{D = 1, Y = 1 | X = x, P = p\} + (1-p) \inf_{x' \in \mathbf{X}_{1-}(x)} \Pr\{Y = 1 | D = 0, X = x', P = p\}\} .\end{aligned}$$

(ii) If Assumptions 2.1 - 2.2 hold and $\text{supp}(P, X) = \text{supp}(P) \times \text{supp}(X)$, then the above expressions for $L_d(x)$ and $U_d(x)$ for $d \in \{0, 1\}$ simplify as follows:

$$\begin{aligned}L_0(x) &= \Pr\{D = 0, Y = 1 | X = x, P = \underline{p}\} + \sup_{x' \in \mathbf{X}_{0-}(x)} \Pr\{D = 1, Y = 1 | X = x', P = \underline{p}\} , \\ L_1(x) &= \Pr\{D = 1, Y = 1 | X = x, P = \bar{p}\} + \sup_{x' \in \mathbf{X}_{1+}(x)} \Pr\{D = 0, Y = 1 | X = x', P = \bar{p}\} , \\ U_0(x) &= \Pr\{D = 0, Y = 1 | X = x, P = \underline{p}\} + \underline{p} \inf_{x' \in \mathbf{X}_{0+}(x)} \Pr\{Y = 1 | D = 1, X = x', P = \underline{p}\} , \\ U_1(x) &= \Pr\{D = 1, Y = 1 | X = x, P = \bar{p}\} + (1-\bar{p}) \inf_{x' \in \mathbf{X}_{1-}(x)} \Pr\{Y = 1 | D = 0, X = x', P = \bar{p}\} ,\end{aligned}$$

where

$$\begin{aligned}\underline{p} &= \inf\{p : p \in \text{supp}(P)\} \\ \bar{p} &= \sup\{p : p \in \text{supp}(P)\} .\end{aligned}$$

(iii) If Assumptions 2.1 - 2.4 hold and $\text{supp}(P, X) = \text{supp}(P) \times \text{supp}(X)$, then the above bounds are sharp.

PROOF: See Appendix. ■

As a corollary to Theorem 3.1, we have immediately that the sign of $\Delta G_1(x)$ is always identified from the distribution of the observed data whenever $h(x, x, p, p')$ is well defined for some $p > p'$. This will be the case whenever Assumption 2.5 holds. This result is critical to the analysis of Bhattacharya et al. (2009). See also Abrevaya et al. (2010), who in related analysis show how to identify and test for the direction of the effect of an endogenous regressor in a semiparametric model.

Corollary 3.1 *Suppose that Y and D satisfy (2) and that Assumptions 2.1 - 2.2 and 2.5 hold. Then, the sign of $\Delta G_1(x)$ is identified.*

Remark 3.1 The bounds of Theorem 3.1 reduce to those in Manski (1989) if there is no variable in Z that is excluded from X , i.e., if Assumption 2.5 does not hold. The bounds will be smaller the more variation there is in X conditional on $P(Z)$. In the extreme case where X is degenerate conditional on $P(Z)$, the bounds reduce to the same form as the Manski and Pepper (2000) bounds under monotone treatment response even though the assumptions of the bounds are different. See the analysis in Bhattacharya et al. (2008) for details. ■

Remark 3.2 Part (iii) of Theorem 3.1 immediately implies that the bounds will be sharp if there are no X regressors in the Y equation, i.e., if $Y = \nu_1(D, \epsilon_1)$. ■

Remark 3.3 It is interesting to ask when the upper and lower bounds will equal one another for the ASF or the ATE, i.e., when the bounds imply that the ASF or the ATE is point-identified. Suppose that $\text{supp}(P, X) = \text{supp}(P) \times \text{supp}(X)$, and further suppose that the sets $\mathbf{X}_{d+}(x)$ and $\mathbf{X}_{d-}(x)$ for $d \in \{0, 1\}$ are nonempty. First consider $G_1(0, x) = E[Y_0|X = x]$. The analysis for $G_1(1, x)$ is similar. The width of the bounds on $G_1(0, x)$ is given by $U_0(x) - L_0(x)$ and is equal to

$$\inf_{x' \in \mathbf{X}_{0+}(x)} \Pr\{D = 1, Y = 1|X = x', P = \underline{p}\} - \sup_{x' \in \mathbf{X}_{0-}(x)} \Pr\{D = 1, Y = 1|X = x', P = \underline{p}\} . \quad (4)$$

Suppose there exists x^* such that $H(x, x^*) = 0$. It follows that $x^* \in \mathbf{X}_{0+}(x) \cap \mathbf{X}_{0-}(x)$ and (4) is less than or equal to

$$\Pr\{D = 1, Y = 1|X = x^*, P = \underline{p}\} - \sup_{x' \in \mathbf{X}_{0-}(x)} \Pr\{D = 1, Y = 1|X = x', P = \underline{p}\} ,$$

which in turn is less than or equal to zero. Since (4) is greater than or equal to zero by construction, it follows that $G_1(0, x)$ is identified whenever there exists x^* such that $H(x, x^*) = 0$. Using Lemma 3.2, we may state this condition equivalently as the existence of a x^* such that $\nu_1(1, x) = \nu_1(0, x^*)$.

Similarly, it is possible to show that the width of the bounds on the ATE is equal to

$$\begin{aligned} & \inf_{x' \in \mathbf{X}_{1-}(x)} \Pr\{D = 0, Y = 1 | X = x', P = \bar{p}\} - \sup_{x' \in \mathbf{X}_{1+}(x)} \Pr\{D = 0, Y = 1 | X = x', P = \bar{p}\} \\ & + \inf_{x' \in \mathbf{X}_{0+}(x)} \Pr\{D = 1, Y = 1 | X = x', P = \underline{p}\} - \sup_{x' \in \mathbf{X}_{0-}(x)} \Pr\{D = 1, Y = 1 | X = x', P = \bar{p}\} . \end{aligned}$$

Using the same reasoning as above, we see that the average treatment effect is identified whenever there exists x^* and x^{**} such that $H(x^*, x) = 0$ and $H(x, x^{**}) = 0$. Equivalently, the average treatment effect is identified whenever there exists x^* and x^{**} such that $\nu_1(1, x) = \nu_1(0, x^*)$ and $\nu_1(1, x^{**}) = \nu_1(0, x)$. ■

Remark 3.4 For the special case of our model in which there are no exogenous covariates and Z is binary, Bhattacharya et al. (2009) adapt the results of Romano and Shaikh (2008) and Romano and Shaikh (2010) for inference on the ATE. More generally, the bounds described in Theorem 3.1 fall into the framework studied by Chernozhukov et al. (2009), and the methods described there can be adapted for inference on the ASF or ATE. ■

Remark 3.5 It is worth noting that there are several testable implications of equation (2) and Assumptions 2.1 - 2.2. A straightforward implication is that $\Pr\{D = 1 | X, Z\}$ does not depend on X , and, as noted earlier, Lemma 3.2 implies that $h(x, x', p, p')$ does not depend on p or p' provided $p > p'$ whenever all conditional probabilities are well defined. It is also possible to show that for $d \in \{0, 1\}$, there exists a real-valued function $Q_d(\cdot)$ such that

$$\Pr\{Y = 1, D = d | X, Z\} = \Pr\{Y = 1, D = d | Q_d(X), P(Z)\} .$$

Moreover, $\Pr\{Y = 1, D = 1 | Q_1(X) = q, P = p\}$ is strictly increasing in both q and p , while $\Pr\{Y = 1, D = 0 | Q_0(X) = q, P = p\}$ is strictly increasing in q and strictly decreasing in p . ■

4 Numerical Illustration

We now provide a numerical illustration of the bounds on the ATE from Theorem 3.1. We first consider an example without any X variation to be exploited when forming the bounds, and then consider an example with such variation. Variation in X greatly reduces the width of the bounds.

Consider the following special case of our model without X regressors:

$$\begin{aligned} Y &= I\{\alpha D - \epsilon_1 \geq 0\} \\ D &= I\{\delta Z - \epsilon_2 \geq 0\} , \end{aligned} \tag{5}$$

with $(\epsilon_1, \epsilon_2) \sim N(0, I)$ and $\text{supp}(Z) = \{-1, 1\}$. In Figure 1, we hold constant the strength of the effect of the treatment on the outcome while varying the strength of the instrument from very weak to very strong. In particular, we set $\alpha = 1/4$ and plot the ATE together with the bounds of Theorem 3.1 for $\delta \in (0, 2]$. The width of the bounds is decreasing in δ , and the width asymptotically approaches zero as δ goes to infinity, i.e., as the strength of the instrument becomes arbitrarily large. This phenomenon is an example of “identification at infinity,” as in Heckman (1990). In Figure

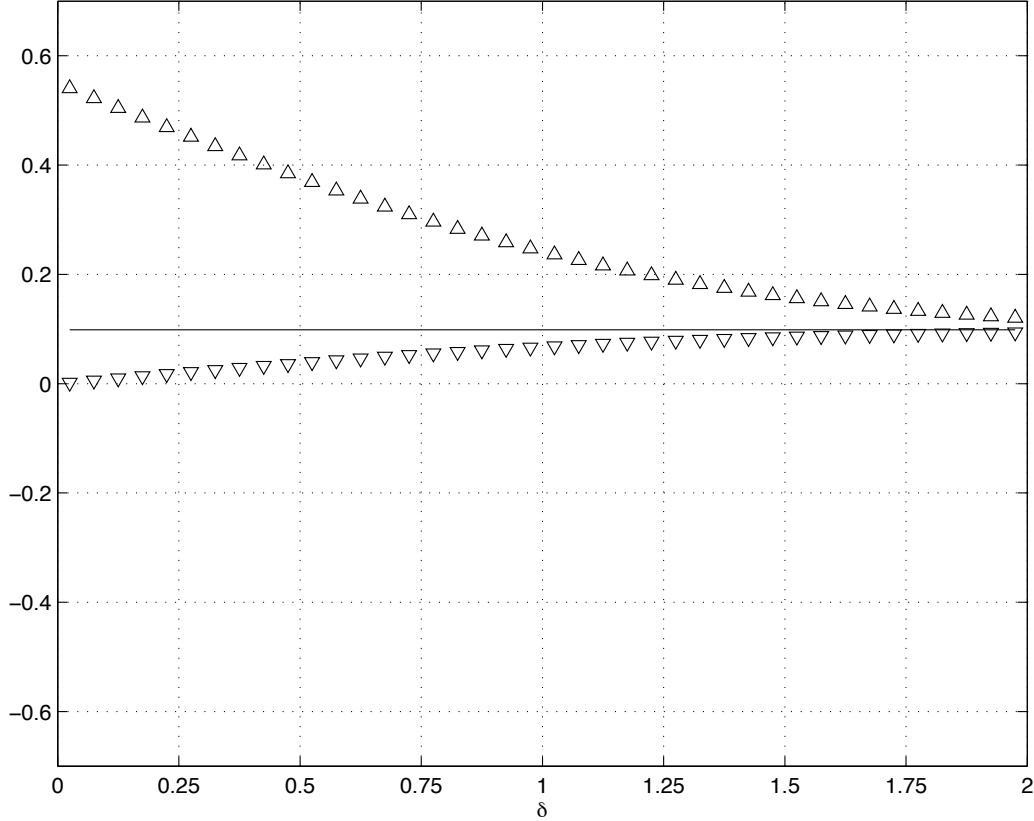


Figure 1: Bounds of Theorem 3.1 on the ATE for the model (5) with $\alpha = 1/4$. The ATE is indicated by the solid line and the upper and lower bounds are indicated, respectively, by the upwards and downwards pointing triangles.

2, we hold constant the strength of the instrument while varying the strength of the effect of the treatment. In particular, we set $\delta = 1/4$ and plot the ATE together with the bounds of Theorem 3.1 for $\alpha \in [-2, 2]$. The width of the bounds decrease as α approaches zero with a discontinuity at the point $\alpha = 0$, where the ATE is identified and equal to zero.

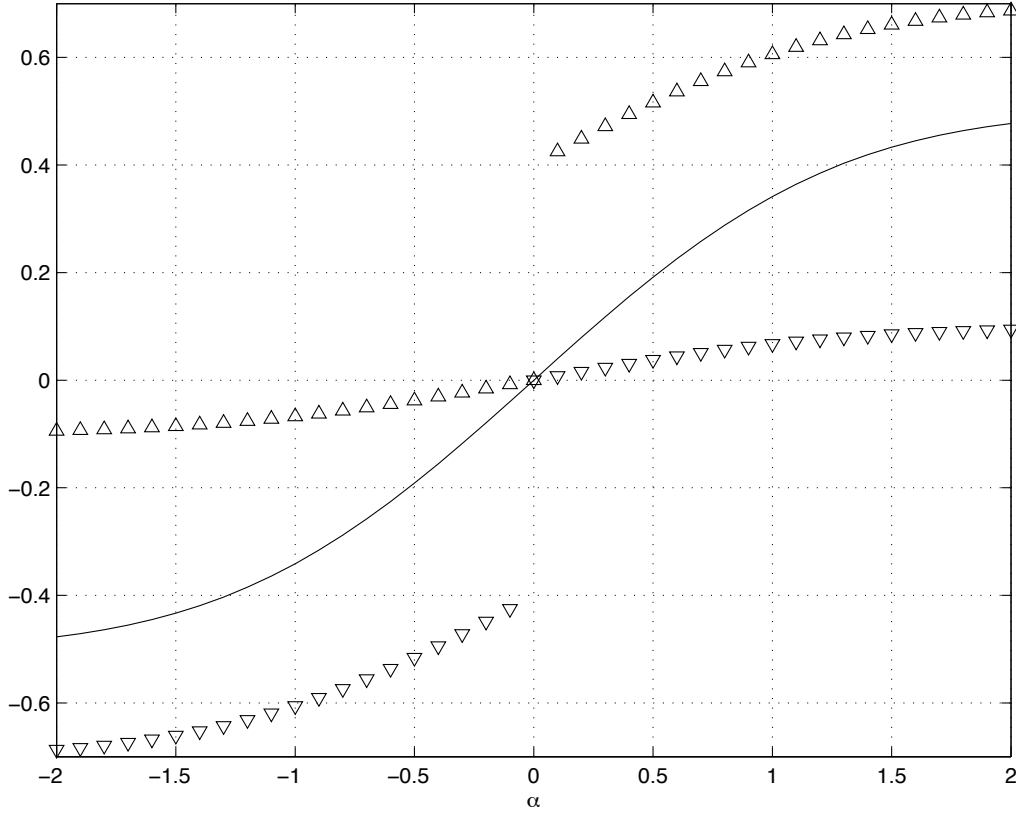


Figure 2: Bounds of Theorem 3.1 on the ATE for the model (5) with $\delta = 1/4$. The ATE is indicated by the solid line and the upper and lower bounds are indicated, respectively, by the upwards and downwards pointing triangles.

Now consider the following special case of our model with X covariates:

$$\begin{aligned} Y &= I\{\beta X + \alpha D - \epsilon_1 \geq 0\} \\ D &= I\{\delta Z - \epsilon_2 \geq 0\} , \end{aligned} \tag{6}$$

with $(\epsilon_1, \epsilon_2) \sim N(0, I)$ and $\text{supp}(X, Z) = \{-2, -1, 0, 1, 2\} \times \{-1, 1\}$. Note that this model conditional on $X = 0$ reduces to (5). In Figure 3, we hold constant both the strength of the effect of the treatment and the strength of the instrument while varying the effect of the exogenous regressor X . In particular, we set $\delta = 1/4$, $\alpha = 1/4$, and plot $\Delta G_1(0)$ together with the bounds of Theorem 3.1 for $\beta \in [1/8, 1/4]$. By comparing the bounds in Figure 3 with the bounds in Figures 1 and 2, we see that the bounds that exploit variation in X dramatically improve upon the bounds that do not exploit variation in X . Note that the bounds provide point identification when $\beta = 1/8$ or $1/4$, which corresponds to the case where there is variation in X that exactly compensates for variation in D .

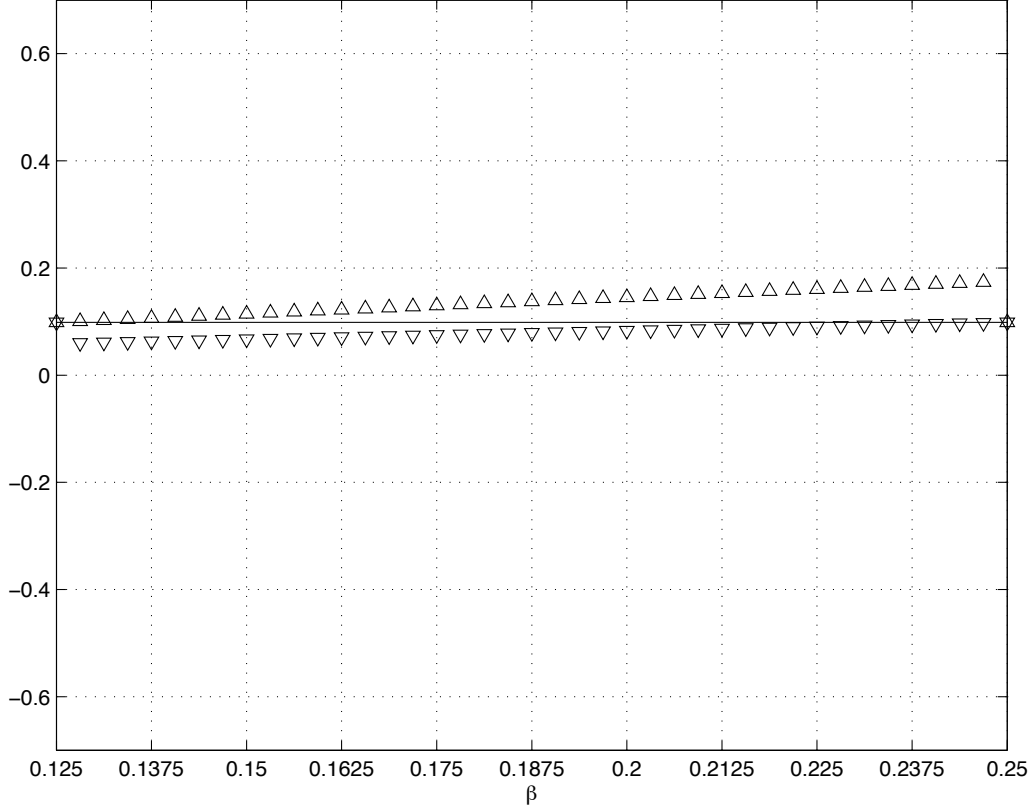


Figure 3: Bounds of Theorem 3.1 on $\Delta G_1(0)$ for the model (6) with $\alpha = 1/4$ and $\delta = 1/4$. The solid line represents $\Delta G_1(0)$ and the upwards and downwards pointing triangles represent, respectively, the upper and lower bounds.

5 Appendix

PROOF OF LEMMA 3.2: First recall the simplifications following from Assumptions 2.1 - 2.2 noted at the beginning of Section 3. Next, note from equation (2) and Assumption 2.1 that

$$\Pr\{D = 1, Y = 1|X = x', P = p\} = \Pr\{\epsilon_2 \leq p, \epsilon_1 \leq \nu_1(1, x')\}$$

and

$$\Pr\{D = 1, Y = 1|X = x', P = p'\} = \Pr\{\epsilon_2 \leq p', \epsilon_1 \leq \nu_1(1, x')\}.$$

Thus, for $p > p'$,

$$\Pr\{D = 1, Y = 1|X = x', P = p\} - \Pr\{D = 1, Y = 1|X = x', P = p'\}$$

is equal to

$$\Pr\{p' < \epsilon_2 \leq p, \epsilon_1 \leq \nu_1(1, x')\}.$$

It follows similarly that

$$\Pr\{D = 0, Y = 1|X = x, P = p'\} = \Pr\{\epsilon_2 > p', \epsilon_1 \leq \nu_1(0, x)\}$$

and

$$\Pr\{D = 0, Y = 1|X = x, P = p\} = \Pr\{\epsilon_2 > p, \epsilon_1 \leq \nu_1(0, x)\} .$$

Therefore,

$$\Pr\{D = 0, Y = 1|X = x, P = p'\} - \Pr\{D = 0, Y = 1|X = x, P = p\}$$

is equal to

$$\Pr\{p' < \epsilon_2 \leq p, \epsilon_1 \leq \nu_1(0, x)\} .$$

Hence,

$$h(x, x', p, p') = \begin{cases} \Pr\{p' < \epsilon_2 \leq p, \nu_1(0, x) < \epsilon_1 \leq \nu_1(1, x')\} & \text{if } \nu_1(1, x') > \nu_1(0, x) \\ 0 & \text{if } \nu_1(1, x') = \nu_1(0, x) \\ -\Pr\{p' < \epsilon_2 \leq p, \nu_1(1, x') < \epsilon_1 \leq \nu_1(0, x)\} & \text{if } \nu_1(1, x') < \nu_1(0, x) . \end{cases}$$

The desired conclusion now follows immediately from Assumption 2.2. ■

PROOF OF THEOREM 3.1: Consider part (i) of the theorem. We derive bounds on $G_1(0, x) = \Pr\{Y_0 = 1|X = x\}$; the bounds on $G_1(1, x)$ and on $\Delta G_1(x)$ follow from parallel arguments.

Note that

$$\Pr\{Y_0 = 1|X = x, P = p\} = \Pr\{D = 0, Y_0 = 1|X = x, P = p\} + \Pr\{D = 1, Y_0 = 1|X = x, P = p\} .$$

By Lemma 3.2, equation (2) and Assumption 2.1,

$$\Pr\{D = 1, Y_0 = 1|X = x, P = p\} \leq \Pr\{D = 1, Y = 1|X = x', P = p\}$$

for all $x' \in \mathbf{X}_{0+}(x)$ and

$$\Pr\{D = 1, Y_0 = 1|X = x, P = p\} \geq \Pr\{D = 1, Y = 1|X = x', P = p\}$$

for all $x' \in \mathbf{X}_{0-}(x)$. Thus, $\Pr\{Y_0 = 1|X = x, P = p\}$ is bounded from below by

$$\Pr\{D = 0, Y = 1|X = x, P = p\} + \sup_{x' \in \mathbf{X}_{0-}(x)} \Pr\{D = 1, Y = 1|X = x', P = p\}$$

and from above by

$$\Pr\{D = 0, Y = 1|X = x, P = p\} + p \inf_{x' \in \mathbf{X}_{0+}(x)} \Pr\{Y = 1|D = 1, X = x', P = p\} ,$$

where all supremums and infimums are only taken over regions where all conditional probabilities are well defined, and with the convention that the supremum over the empty set is zero and the infimum over the empty set is one. The stated result now follows by noting that equation (2) and Assumption 2.1 imply that $\Pr\{Y_0 = 1|X = x\} = \Pr\{Y_0 = 1|X = x, P = p\}$.

Consider part (ii) of the theorem. We prove the result for the term $L_0(x)$; the result for the other terms follows from parallel arguments.

Suppose $\text{supp}(P)$ is not a singleton, for otherwise there is nothing to prove. Since $\text{supp}(X, P) = \text{supp}(X) \times \text{supp}(P)$, $h(x, x', p, p')$ is well defined for some $p < p'$ with $(p, p') \in \text{supp}(P)^2$ and any $(x, x') \in \text{supp}(X)^2$. Hence, by Lemma 3.2, we have that

$$\mathbf{X}_{0-}(x) = \{x' : \nu_1(1, x') \leq \nu_1(0, x)\} . \quad (7)$$

It follows from Assumptions 2.3 - 2.4 that $\mathbf{X}_{0-}(x)$ is compact. Hence, by Assumption 2.4, there exists $x_0^l(x) \in \mathbf{X}_{0-}(x)$ such that

$$\nu_1(1, x_0^l(x)) = \sup_{x' \in \mathbf{X}_{0-}(x)} \nu_1(1, x') .$$

From equation (2), we therefore have for any $p \in \text{supp}(P)$ that

$$\sup_{x' \in \mathbf{X}_{0-}(x)} \Pr\{D = 1, Y = 1 | X = x', P = p\} = \Pr\{D = 1, Y = 1 | X = x_0^l(x), P = p\} ,$$

from which it follows that

$$L_0(x) = \sup_p \{\Pr\{D = 0, Y = 1 | X = x, P = p\} + \Pr\{D = 1, Y = 1 | X = x_0^l(x), P = p\}\} .$$

To complete the argument, note for any $p > p'$ that

$$\begin{aligned} & (\Pr\{D = 0, Y = 1 | X = x, P = p\} + \Pr\{D = 1, Y = 1 | X = x_0^l(x), P = p\}) \\ & - (\Pr\{D = 0, Y = 1 | X = x, P = p'\} + \Pr\{D = 1, Y = 1 | X = x_0^l(x), P = p'\}) \\ & = \Pr\{\epsilon_1 \leq \nu_1(1, x_0^l(x)), p' < \epsilon_2 \leq p\} - \Pr\{\epsilon_1 \leq \nu_1(0, x), p' < \epsilon_2 \leq p\} \leq 0 , \end{aligned}$$

where the final inequality follows from the fact that $x_0^l(x) \in \mathbf{X}_{0-}(x)$ and (7).

Finally, consider part (iii) of the theorem. Before proceeding, we introduce some notation. Let $(\epsilon_1^*, \epsilon_2^*)$ denote a random vector with $(\epsilon_1^*, \epsilon_2^*) \perp\!\!\!\perp (X, Z)$ and with $(\epsilon_1^*, \epsilon_2^*)$ having density $f_{1,2}^*$ with respect to Lebesgue measure on \mathbf{R}^2 . Let f_2^* denote the corresponding marginal density of ϵ_2^* and let $f_{1|2}^*$ denote the corresponding density of ϵ_1^* conditional on ϵ_2^* . Let $f_{1,2}$, $f_{1|2}$ and f_2 denote the corresponding density functions for (ϵ_1, ϵ_2) . We will also make use of $F_{1,2}$, the c.d.f. for (ϵ_1, ϵ_2) , and $F_{1,-2}$, the c.d.f. for $(\epsilon_1, -\epsilon_2)$.

In order to show that our bounds on $G_1(0, x)$, $G_1(1, x)$ and $G_1(1, x) - G_1(0, x)$ are sharp, it suffices to show that for any $x \in \text{supp}(X)$ and $(s_0, s_1) \in [L_0(x), U_0(x)] \times [L_1(x), U_1(x)]$ there exists a density function $f_{1,2}^*$ such that:

(A) $f_{1,2}^*$ is strictly positive on \mathbf{R}^2 .

(B) the proposed model is consistent with the observed data, i.e.,

- (i) $\Pr\{D = 1 | X = \tilde{x}, P = p\} = \Pr\{\epsilon_2^* \leq p\}$
- (ii) $\Pr\{Y = 1 | D = 1, X = \tilde{x}, P = p\} = \Pr\{\epsilon_1^* \leq \nu_1(1, \tilde{x}) \mid \epsilon_2^* \leq p\}$
- (iii) $\Pr\{Y = 1 | D = 0, X = \tilde{x}, P = p\} = \Pr\{\epsilon_1^* \leq \nu_1(0, \tilde{x}) \mid \epsilon_2^* > p\}$

for all $(\tilde{x}, p) \in \text{supp}(X, P)$.

(C) the proposed model is consistent with the specified values of $G_1(0, x)$ and $G_1(1, x)$, i.e.,

- (i) $\Pr\{\epsilon_1^* \leq \nu_1(0, x)\} = s_0$
- (ii) $\Pr\{\epsilon_1^* \leq \nu_1(1, x)\} = s_1$.

Let $x \in \text{supp}(X)$ and $(s_0, s_1) \in [L_0(x), U_0(x)] \times [L_1(x), U_1(x)]$ be given. We prove the result for the case where $\mathbf{X}_{d-}(x) \neq \emptyset$, $\mathbf{X}_{d+}(x) \neq \emptyset$, and $\mathbf{X}_{d-}(x) \cap \mathbf{X}_{d+}(x) = \emptyset$ for $d \in \{0, 1\}$; the result in the other cases follows from analogous arguments. Note that by arguing as in Remark 3.3 this implies in particular that $L_d(x) < U_d(x)$ for $d \in \{0, 1\}$. For brevity, we also only consider $(s_0, s_1) \in (L_0(x), U_0(x)) \times (L_1(x), U_1(x))$; the case where s_d equals $L_d(x)$ or $U_d(x)$ for some $d \in \{0, 1\}$ follows from a straightforward modification of the argument below.

Recall that $h(x, x', p, p')$ is well defined for some $p < p'$ with $(p, p') \in \text{supp}(P)^2$ and any $(x, x') \in \text{supp}(X)^2$ because $\text{supp}(X, P) = \text{supp}(X) \times \text{supp}(P)$. Arguing as in the proof of part (ii) of the theorem, we have that

$$\begin{aligned} L_0(x) &= \Pr\{D = 0, Y = 1 | X = x, P = \underline{p}\} + \Pr\{D = 1, Y = 1 | X = x_0^l(x), P = \underline{p}\} \\ U_0(x) &= \Pr\{D = 0, Y = 1 | X = x, P = \underline{p}\} + \Pr\{D = 1, Y = 1 | X = x_0^u(x), P = \underline{p}\} \\ L_1(x) &= \Pr\{D = 1, Y = 1 | X = x, P = \bar{p}\} + \Pr\{D = 0, Y = 1 | X = x_1^l(x), P = \bar{p}\} \\ U_1(x) &= \Pr\{D = 1, Y = 1 | X = x, P = \bar{p}\} + \Pr\{D = 0, Y = 1 | X = x_1^u(x), P = \bar{p}\}, \end{aligned} \quad (8)$$

where $x_d^l(x)$ and $x_d^u(x)$ for $d \in \{0, 1\}$ denote evaluation points such that

$$\begin{aligned} \Pr\{D = 1, Y = 1 | X = x_0^l(x), P = \underline{p}\} &= \sup_{x' \in \mathbf{X}_{0-}(x)} \Pr\{D = 1, Y = 1 | X = x', P = \underline{p}\} \\ \Pr\{D = 1, Y = 1 | X = x_0^u(x), P = \underline{p}\} &= \inf_{x' \in \mathbf{X}_{0+}(x)} \Pr\{D = 1, Y = 1 | X = x', P = \underline{p}\} \\ \Pr\{D = 0, Y = 1 | X = x_1^l(x), P = \bar{p}\} &= \sup_{x' \in \mathbf{X}_{1+}(x)} \Pr\{D = 0, Y = 1 | X = x', P = \bar{p}\} \\ \Pr\{D = 0, Y = 1 | X = x_1^u(x), P = \bar{p}\} &= \inf_{x' \in \mathbf{X}_{1-}(x)} \Pr\{D = 0, Y = 1 | X = x', P = \bar{p}\}. \end{aligned}$$

Let

$$\begin{aligned} s_0^* &= s_0 - \Pr\{D = 0, Y = 1 | X = x, P = \underline{p}\} \\ s_1^* &= s_1 - \Pr\{D = 1, Y = 1 | X = x, P = \bar{p}\}. \end{aligned}$$

Using equation (8) and the fact that $s_d \in (L_d(x), U_d(x))$ for $d \in \{0, 1\}$, we have that

$$\begin{aligned} s_0^* &\in (F_{1,2}(\nu_1(1, x_0^l(x)), \underline{p}), F_{1,2}(\nu_1(1, x_0^u(x)), \underline{p})) \\ s_1^* &\in (F_{1,-2}(\nu_1(0, x_1^l(x)), -\bar{p}), F_{1,-2}(\nu_1(0, x_1^u(x)), -\bar{p})). \end{aligned} \quad (9)$$

These intervals are non-empty because $L_d(x) < U_d(x)$ for $d \in \{0, 1\}$. It follows by Lemma 3.2 that

$$\nu_1(d, x_{1-d}^l(x)) < \nu_1(1-d, x) < \nu_1(d, x_{1-d}^u(x)) \quad (10)$$

for $d \in \{0, 1\}$, where the strict inequalities follow from our assumption that $\mathbf{X}_{d-}(x) \cap \mathbf{X}_{d+}(x) = \emptyset$ for $d \in \{0, 1\}$. Furthermore, by the construction of $x_d^l(x)$ and $x_d^u(x)$ for $d \in \{0, 1\}$, it must be the case for $d \in \{0, 1\}$ and $\tilde{x} \in \text{supp}(X)$ that

$$\nu_1(d, \tilde{x}) \notin (\nu_1(d, x_{1-d}^l(x)), \nu_1(d, x_{1-d}^u(x))). \quad (11)$$

We now construct the proposed density $f_{1,2}^*$ as follows. Let $f_{1,2}^*(t_1, t_2) = f_{1|2}^*(t_1|t_2)f_2^*(t_2)$, where $f_2^*(t_2) =$

$f_2(t_2) = I\{0 \leq t_2 \leq 1\}$ and

$$f_{1|2}^*(t_1|t_2) = \begin{cases} a(t_2)f_{1|2}(t_1|t_2) & \text{if } \nu_1(1, x_0^l(x)) < t_1 < \nu_1(0, x) \text{ and } t_2 < \underline{p} \\ b(t_2)f_{1|2}(t_1|t_2) & \text{if } \nu_1(0, x) \leq t_1 < \nu_1(1, x_0^u(x)) \text{ and } t_2 < \underline{p} \\ c(t_2)f_{1|2}(t_1|t_2) & \text{if } \nu_1(0, x_1^l(x)) \leq t_1 < \nu_1(1, x) \text{ and } t_2 > \bar{p} \\ d(t_2)f_{1|2}(t_1|t_2) & \text{if } \nu_1(1, x) \leq t_1 < \nu_1(0, x_1^u(x)) \text{ and } t_2 > \bar{p} \\ f_{1|2}(t_1|t_2) & \text{otherwise ,} \end{cases}$$

with

$$\begin{aligned} a(t_2) &= \frac{\Pr\{\nu_1(1, x_0^l(x)) < \epsilon_1 < \nu_1(1, x_0^u(x)) | \epsilon_2 = t_2\}}{\Pr\{\nu_1(1, x_0^l(x)) < \epsilon_1 < \nu_1(0, x) | \epsilon_2 = t_2\}} \frac{s_0^* - F_{1,2}(\nu_1(1, x_0^l(x)), \underline{p})}{F_{1,2}(\nu_1(1, x_0^u(x)), \underline{p}) - F_{1,2}(\nu_1(1, x_0^l(x)), \underline{p})} \\ b(t_2) &= \frac{\Pr\{\nu_1(1, x_0^l(x)) < \epsilon_1 < \nu_1(1, x_0^u(x)) | \epsilon_2 = t_2\} - a(t_2) \Pr\{\nu_1(1, x_0^l(x)) < \epsilon_1 < \nu_1(0, x) | \epsilon_2 = t_2\}}{\Pr\{\nu_1(0, x) < \epsilon_1 < \nu_1(1, x_0^u(x)) | \epsilon_2 = t_2\}} \\ c(t_2) &= \frac{\Pr\{\nu_1(0, x_1^l(x)) < \epsilon_1 < \nu_1(0, x_1^u(x)) | \epsilon_2 = t_2\}}{\Pr\{\nu_1(0, x_1^l(x)) < \epsilon_1 < \nu_1(1, x) | \epsilon_2 = t_2\}} \frac{s_1^* - F_{1,-2}(\nu_1(0, x_1^l(x)), -\bar{p})}{F_{1,-2}(\nu_1(0, x_1^u(x)), -\bar{p}) - F_{1,-2}(\nu_1(0, x_1^l(x)), -\bar{p})} \\ d(t_2) &= \frac{\Pr\{\nu_1(0, x_1^l(x)) < \epsilon_1 < \nu_1(0, x_1^u(x)) | \epsilon_2 = t_2\} - c(t_2) \Pr\{\nu_1(0, x_1^l(x)) < \epsilon_1 < \nu_1(1, x) | \epsilon_2 = t_2\}}{\Pr\{\nu_1(1, x) < \epsilon_1 < \nu_1(0, x_0^u(x)) | \epsilon_2 = t_2\}} . \end{aligned}$$

These quantities are well defined because of the fact that the intervals in (9) are non-empty, (10) and Assumption 2.2.

We now argue that $f_{1,2}^*$ satisfies claim (A), i.e., that it is a strictly positive density on \mathbf{R}^2 . For this purpose, it suffices to show that $f_{1|2}^*$ integrates to one and is strictly positive on \mathbf{R} . First consider whether $f_{1|2}^*$ integrates to one. For $t_2 \in [\underline{p}, \bar{p}]$, $f_{1|2}^*(\cdot|t_2) = f_{1|2}(\cdot|t_2)$ and so the result follows immediately. For $t_2 < \underline{p}$,

$$\begin{aligned} & \int_{-\infty}^{\infty} f_{1|2}^*(t_1|t_2) dt_1 \\ &= \int_{-\infty}^{\nu_1(1, x_0^l(x))} f_{1|2}(t_1|t_2) dt_1 + a(t_2) \int_{\nu_1(1, x_0^l(x))}^{\nu_1(0, x)} f_{1|2}(t_1|t_2) dt_1 + b(t_2) \int_{\nu_1(0, x)}^{\nu_1(1, x_0^u(x))} f_{1|2}(t_1|t_2) dt_1 \\ & \quad + \int_{\nu_1(1, x_0^u(x))}^{\infty} f_{1|2}(t_1|t_2) dt_1 \\ &= \Pr\{\epsilon_1 \leq \nu_1(1, x_0^l(x)) | \epsilon_2 = t_2\} + \Pr\{\nu_1(1, x_0^l(x)) < \epsilon_1 < \nu_1(1, x_0^u(x)) | \epsilon_2 = t_2\} \\ & \quad + \Pr\{\epsilon_1 \geq \nu_1(1, x_0^u(x)) | \epsilon_2 = t_2\} \\ &= 1. \end{aligned}$$

A similar argument shows that $\int f_{1|2}^*(t_1|t_2) dt_1 = 1$ for $t_2 > \bar{p}$.

Since $f_{1|2}$ is strictly positive on \mathbf{R} , in order to establish that $f_{1|2}^*$ is strictly positive on \mathbf{R} it suffices to show that $a(t_2), b(t_2), c(t_2)$ and $d(t_2)$ are all strictly positive. Consider $a(t_2)$ and $b(t_2)$; the proof for $c(t_2)$ and $d(t_2)$ follows from similar arguments. From (9), we have that $s_0^* > F_{1,2}(\nu_1(1, x_0^l(x)), \underline{p})$, which together with (10) and Assumption 2.2 implies that $a(t_2) > 0$. Similarly, from (9), we have that $s_0^* < F_{1,2}(\nu_1(1, x_0^u(x)), \underline{p})$, which implies that

$$\frac{s_0^* - F_{1,2}(\nu_1(1, x_0^l(x)), \underline{p})}{F_{1,2}(\nu_1(1, x_0^u(x)), \underline{p}) - F_{1,2}(\nu_1(1, x_0^l(x)), \underline{p})} < 1 .$$

It therefore follows from (10) and Assumption 2.2 that

$$\begin{aligned} & \Pr\{\nu_1(1, x_0^l(x)) < \epsilon_1 < \nu_1(1, x_0^u(x)) | \epsilon_2 = t_2\} - a(t_2) \Pr\{\nu_1(1, x_0^l(x)) < \epsilon_1 < \nu_1(0, x) | \epsilon_2 = t_2\} \\ &= \Pr\{\nu_1(1, x_0^l(x)) < \epsilon_1 < \nu_1(1, x_0^u(x)) | \epsilon_2 = t_2\} \left(1 - \frac{s_0^* - F_{1,2}(\nu_1(1, x_0^l(x)), \underline{p})}{F_{1,2}(\nu_1(1, x_0^u(x)), \underline{p}) - F_{1,2}(\nu_1(1, x_0^l(x)), \underline{p})} \right) \\ &> 0, \end{aligned}$$

so $b(t_2) > 0$.

We now argue that $f_{1,2}^*$ satisfies claim (B). Since $f_2^* = f_2$, we have immediately that $\Pr\{\epsilon_2^* \leq p\} = \Pr\{D = 1 | X = \tilde{x}, P = p\}$ for all $(\tilde{x}, p) \in \text{supp}(X, P)$. Consider $\Pr\{\epsilon_1^* \leq \nu_1(1, \tilde{x}) | \epsilon_2^* \leq p\}$. From (11), we have that $\nu_1(1, \tilde{x}) \leq \nu_1(1, x_0^l(x))$ or $\nu_1(1, \tilde{x}) \geq \nu_1(1, x_0^u(x))$ for any $\tilde{x} \in \text{supp}(X)$. For $(\tilde{x}, p) \in \text{supp}(X, P)$ such that $\nu_1(1, \tilde{x}) \leq \nu_1(1, x_0^l(x))$, we have

$$\begin{aligned} \Pr\{\epsilon_1^* \leq \nu_1(1, \tilde{x}) | \epsilon_2^* \leq p\} &= \frac{1}{p} \int_0^p \int_{-\infty}^{\nu_1(1, \tilde{x})} f_{1,2}^*(t_1, t_2) dt_1 dt_2 \\ &= \frac{1}{p} \int_0^p \int_{-\infty}^{\nu_1(1, \tilde{x})} f_{1,2}(t_1, t_2) dt_1 dt_2 \\ &= \Pr\{\epsilon_1 \leq \nu_1(1, \tilde{x}) | \epsilon_2 \leq p\} = \Pr\{Y = 1 | D = 1, X = \tilde{x}, P = p\}. \end{aligned}$$

For $(\tilde{x}, p) \in \text{supp}(X, P)$ such that $\nu_1(1, \tilde{x}) \geq \nu_1(1, x_0^u(x))$, we have

$$\begin{aligned} & \Pr\{\epsilon_1^* \leq \nu_1(1, \tilde{x}) | \epsilon_2^* \leq p\} \\ &= \frac{1}{p} \int_0^p \int_{-\infty}^{\nu_1(1, \tilde{x})} f_{1,2}^*(t_1, t_2) dt_1 dt_2 \\ &= \frac{1}{p} \left\{ \int_{\underline{p}}^p \int_{-\infty}^{\nu_1(1, \tilde{x})} f_{1,2}(t_1, t_2) dt_1 dt_2 + \int_0^{\underline{p}} \left[\int_{-\infty}^{\nu_1(1, x_0^l(x))} f_{1|2}(t_1 | t_2) dt_1 + a(t_2) \int_{\nu_1(1, x_0^l(x))}^{\nu_1(0, x)} f_{1|2}(t_1 | t_2) dt_1 \right. \right. \\ &\quad \left. \left. + b(t_2) \int_{\nu_1(0, x)}^{\nu_1(1, x_0^u(x))} f_{1|2}(t_1 | t_2) dt_1 + \int_{\nu_1(1, x_0^u(x))}^{\nu_1(1, \tilde{x})} f_{1|2}(t_1 | t_2) dt_1 \right] dt_2 \right\} \\ &= \frac{1}{p} \left\{ \Pr\{\epsilon_1 \leq \nu_1(1, \tilde{x}), \underline{p} < \epsilon_2 \leq p\} + \Pr\{\epsilon_1 \leq \nu_1(1, \tilde{x}), \epsilon_2 \leq \underline{p}\} \right\} \\ &= \Pr\{\epsilon_1 \leq \nu_1(1, \tilde{x}) | \epsilon_2 \leq p\} = \Pr\{Y = 1 | D = 1, X = \tilde{x}, P = p\}. \end{aligned}$$

The proof that $\Pr\{\epsilon_1^* \leq \nu_1(0, \tilde{x}) | \epsilon_2^* > p\} = \Pr\{Y = 1 | D = 0, X = \tilde{x}, P = p\}$ for all $(\tilde{x}, p) \in \text{supp}(X, P)$ follows from an analogous argument.

Finally, we argue that $f_{1,2}^*$ satisfies claim (C). Consider $\Pr\{\epsilon_1^* \leq \nu_1(0, x)\}$. From (11), we have that

$\nu_1(1, x) \leq \nu_1(1, x_0^l(x))$ or $\nu_1(1, x) \geq \nu_1(1, x_0^u(x))$. In the former case, we have that

$$\begin{aligned}
& \Pr\{\epsilon_1^* \leq \nu_1(0, x)\} \\
&= \int_0^1 \int_{-\infty}^{\nu_1(0, x)} f_{1,2}^*(t_1, t_2) dt_1 dt_2 \\
&= \left\{ \int_0^{\underline{p}} \left(\int_{-\infty}^{\nu_1(1, x_0^l(x))} f_{1,2}^*(t_1, t_2) dt_1 + \int_{\nu_1(1, x_0^l(x))}^{\nu_1(0, x)} f_{1,2}^*(t_1, t_2) dt_1 \right) dt_2 + \int_{\underline{p}}^1 \int_{-\infty}^{\nu_1(0, x)} f_{1,2}^*(t_1, t_2) dt_1 dt_2 \right\} \\
&= \left\{ \int_0^{\underline{p}} \left(\int_{-\infty}^{\nu_1(1, x_0^l(x))} f_{1,2}(t_1, t_2) dt_1 + a(t_2) \int_{\nu_1(1, x_0^l(x))}^{\nu_1(0, x)} f_{1,2}(t_1, t_2) dt_1 \right) dt_2 \right. \\
&\quad \left. + \int_{\underline{p}}^1 \int_{-\infty}^{\nu_1(0, x)} f_{1,2}(t_1, t_2) dt_1 dt_2 \right\} \\
&= s_0^* + \Pr\{D = 0, Y = 1 | X = x, P = \underline{p}\} = s_0 .
\end{aligned}$$

In the latter case, it suffices to show that

$$\int_{\underline{p}}^1 \int_{-\infty}^{\nu_1(0, x)} f_{1,2}^*(t_1, t_2) dt_1 dt_2 = \int_{\underline{p}}^1 \int_{-\infty}^{\nu_1(0, x)} f_{1,2}(t_1, t_2) dt_1 dt_2 .$$

For this purpose, it suffices to show that

$$\int_{\bar{p}}^1 \int_{\nu_1(0, x_1^l(x))}^{\nu_1(0, x_1^u(x))} f_{1,2}^*(t_1, t_2) dt_1 dt_2 = \int_{\bar{p}}^1 \int_{\nu_1(0, x_1^l(x))}^{\nu_1(0, x_1^u(x))} f_{1,2}(t_1, t_2) dt_1 dt_2 ,$$

since outside of this region of integration $f_{1,2}^* = f_{1,2}$. Note that

$$\begin{aligned}
& \int_{\bar{p}}^1 \int_{\nu_1(0, x_1^l(x))}^{\nu_1(0, x_1^u(x))} f_{1,2}^*(t_1, t_2) dt_1 dt_2 \\
&= \int_{\bar{p}}^1 c(t_2) \int_{\nu_1(0, x_1^l(x))}^{\nu_1(0, x)} f_{1|2}(t_1 | t_2) dt_1 dt_2 + \int_{\bar{p}}^1 d(t_2) \int_{\nu_1(0, x)}^{\nu_1(0, x_1^u(x))} f_{1|2}(t_1 | t_2) dt_1 dt_2 \\
&= \int_{\bar{p}}^1 c(t_2) \Pr\{\nu_1(0, x_1^l(x)) < \epsilon_1 < \nu_1(0, x) | t_2\} dt_2 + \int_{\bar{p}}^1 d(t_2) \Pr\{\nu_1(0, x) < \epsilon_1 < \nu_1(0, x_1^u(x)) | t_2\} dt_2 \\
&= \int_{\bar{p}}^1 \Pr\{\nu_1(0, x_1^l(x)) < \epsilon_1 < \nu_1(1, x_1^u(x)) | t_2\} dt_2 \\
&= \int_{\underline{p}}^1 \int_{\nu_1(0, x_1^l(x))}^{\nu_1(0, x_1^u(x))} f_{1,2}(t_1, t_2) dt_1 dt_2 ,
\end{aligned}$$

as desired. The proof that $\Pr\{\epsilon_1^* \leq \nu_1(1, x)\} = s_1$ follows from an analogous argument. ■

References

- ABREVAYA, J., HAUSMAN, J. and KHAN, S. (2010). Testing for causal effects in a generalized model with endogenous regressors. *Econometrica*. Forthcoming.
- ALTONJI, J. G. and MATZKIN, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, **73** 1053–1102.

- BALKE, A. and PEARL, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, **92** 1171–1176.
- BHATTACHARYA, J., SHAIKH, A. and VYTLACIL, E. (2008). Treatment effect bounds under monotonicity conditions: An application to Swan-Ganz catheterization. *American Economic Review, Papers and Proceedings*, **98** 351–6.
- BHATTACHARYA, J., SHAIKH, A. M. and VYTLACIL, E. J. (2009). Treatment effect bounds: An application to Swan-Ganz catheterization. Working paper.
- BLUNDELL, R. and POWELL, J. (2004). Endogeneity in semiparametric binary response models. *Review of Economic Studies*, **71** 655–79.
- CHERNOZHUKOV, V. and HANSEN, C. (2005). An IV model of quantile treatment effects. *Econometrica*, **73** 245–261.
- CHERNOZHUKOV, V., IMBENS, G. and NEWEY, W. (2007). Instrumental variable identification and estimation of nonseparable models via quantile conditions. *Journal of Econometrics*, **142** 379–398.
- CHERNOZHUKOV, V., LEE, S. and ROSEN, A. (2009). Intersection bounds: Estimation and inference. Working paper.
- CHESHER, A. (2003). Identification in nonseparable models. *Econometrica*, **71** 1405–1441.
- CHESHER, A. (2005). Nonparametric identification under discrete variation. *Econometrica*, **73** 1525–1550.
- CHESHER, A. (2007). Endogeneity and discrete outcomes. Working Paper.
- CHIBURIS, R. (2009). Semiparametric bounds on treatment effects. Working paper.
- HECKMAN, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, **46** 931–959.
- HECKMAN, J. (1990). Varieties of selection bias. *American Economic Review*, **80** 313–318.
- HECKMAN, J. and VYTLACIL, E. (2001). Instrumental variables, selection models, and tight bounds on the average treatment effect. In *Econometric Evaluations of Active Labor Market Policies in Europe* (M. Lechner and F. Pfeiffer, eds.). Physica-Verlag, Heidelberg; New York.
- HECKMAN, J. J. and VYTLACIL, E. J. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, **73** 669–738.

- IMBENS, G. and ANGRIST, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** 467–475.
- IMBENS, G. W. and NEWKEY, W. K. (2010). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*. Forthcoming.
- JUN, S. J., PINKSE, J. and XU, H. (2009). Tighter bounds in triangular systems. Working paper.
- MANSKI, C. (1989). Anatomy of the selection problem. *Journal of Human Resources*, **24** 343–360.
- MANSKI, C. and PEPPER, J. (2000). Monotone instrumental variables with an application to the returns to schooling. *Econometrica*, **68** 997–1010.
- ROMANO, J. and SHAIKH, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference*, **138** 2786–2807.
- ROMANO, J. and SHAIKH, A. M. (2010). Inference for the identified set in partially identified econometric models. *Econometrica*, **78** 169–211.
- VYTLACIL, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, **70** 331–341.
- VYTLACIL, E. (2006). A note on additive separability and latent index models of binary choice: Representation results. *Oxford Bulletin of Economics and Statistics*, **68** 515–518.
- VYTLACIL, E. and YILDIZ, N. (2007). Dummy endogenous variables in weakly separable models. *Econometrica*, **75** 757–779.