ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 79, No. 3 (May, 2011), 949–955

PARTIAL IDENTIFICATION IN TRIANGULAR SYSTEMS OF EQUATIONS WITH BINARY DEPENDENT VARIABLES

AZEEM M. SHAIKH
University of Chicago, Chicago, IL 60637, U.S.A.

EDWARD J. VYTLACIL
Yale University, New Haven, CT 06520-8281, U.S.A.

Econometrica, Vol. 79, No. 3 (May, 2011), 949–955

NOTES AND COMMENTS

PARTIAL IDENTIFICATION IN TRIANGULAR SYSTEMS OF EQUATIONS WITH BINARY DEPENDENT VARIABLES

BY AZEEM M. SHAIKH AND EDWARD J. VYTLACIL[1]

This paper studies the special case of the triangular system of equations in Vytlacil and Yildiz (2007), where both dependent variables are binary but without imposing the restrictive support condition required by Vytlacil and Yildiz (2007) for identification of the average structural function (ASF) and the average treatment effect (ATE). Under weak regularity conditions, we derive upper and lower bounds on the ASF and the ATE. We show further that the bounds on the ASF and ATE are sharp under some further regularity conditions and an additional restriction on the support of the covariates and the instrument.

KEYWORDS: Partial identification, simultaneous equation model, binary dependent variable, endogeneity, threshold crossing model, weak separability, average structural function, average treatment effect.

1. INTRODUCTION

THIS PAPER STUDIES the special case of the triangular system of equations in Vytlacil and Yildiz (2007), where both dependent variables are binary. Under the weak separability assumptions imposed by Vytlacil and Yildiz (2007), such a model may, without loss of generality, be written as[2]

$$Y = I\{\nu_1(D, X) \geq \varepsilon_1\}, \qquad (1)$$
$$D = I\{\nu_2(Z) \geq \varepsilon_2\}.$$

Here, Y denotes the observed binary outcome of interest, D denotes the observed binary endogenous regressor, X and Z are observed random vectors, and ε_1 and ε_2 are unobserved random variables. We additionally assume some mild regularity on the distribution of $(\varepsilon_1, \varepsilon_2)$ and that X and Z are exogenous in the sense that $(X, Z) \perp\!\!\!\perp (\varepsilon_1, \varepsilon_2)$. Under these assumptions, we derive upper and lower bounds on the average structural function (ASF) and the average treatment effect (ATE), which may be expressed, respectively, as

$$G_1(d, x) = \Pr\{Y_d = 1 | X = x\},$$
$$\Delta G_1(x) = \Pr\{Y_1 = 1 | X = x\} - \Pr\{Y_0 = 1 | X = x\},$$

[1]An earlier version of this paper titled "Threshold Crossing Models and Bounds on Treatment Effects: A Nonparametric Analysis" appeared in May 2005 as NBER Technical Working Paper 307. 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 Vytlacil was in residence at Hitotsubashi University. This research was supported by NSF SES-05-51089 and DMS-08-20310.

[2]This can be shown by appropriately adapting arguments in Vytlacil (2002).

 DOI: 10.3982/ECTA9082

where $Y_d = I\{\nu_1(d, X) \geq \varepsilon_1\}$ and (d, x) denotes a potential realization of (D, X). Vytlacil and Yildiz (2007) established identification of the ASF and the ATE when the support of the distribution of X conditional on $\Pr\{D = 1|Z\}$ is sufficiently rich. This support condition would be expected to fail near the boundaries of the support of X. In particular, it would be expected to fail when X is a discrete random variable. In this paper, we do not impose any such support restriction. Under further assumptions, we show 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. In subsequent work, Chiburis (2010) showed that our bounds may not be sharp when these additional assumptions are not satisfied.

2. IDENTIFICATION ANALYSIS

Formally, we will make use of the following assumptions in our analysis:

ASSUMPTION 2.1: $(X, Z) \perp\!\!\!\perp (\varepsilon_1, \varepsilon_2)$.

ASSUMPTION 2.2: *The distribution of* $(\varepsilon_1, \varepsilon_2)$ *has strictly positive density with respect to* (*w.r.t.*) *Lebesgue measure on* $\mathbf{R}^2$.

ASSUMPTION 2.3: *The support of the distribution of* (X, Z), $\mathrm{supp}(X, Z)$, *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*.

Our analysis below is similar to Chesher (2005), but his analysis requires a rank condition that can only hold in trivial cases when D is binary. Jun, Pinkse, and Xu (2009) relaxed this rank condition so that it may hold nontrivially when D is binary, but they impose an additional assumption on the dependence between ε_1 and ε_2.

Note that it follows from Assumptions 2.1 and 2.2 that we may, without loss of generality, normalize $\varepsilon_2 \sim U(0, 1)$ 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 $(\varepsilon_1, \varepsilon_2)$ 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.

Consider first identification of $G_1(1, x)$. By equation (1) and Assumption 2.1, we have that $\Pr\{Y_1 = 1|X\} = \Pr\{Y_1 = 1|X, P(Z)\}$ and $\Pr\{D =$

$1|X, P(Z)\} = P(Z)$. Since the events $\{D = 1, Y = 1\}$ and $\{D = 1, Y_1 = 1\}$ are the same,

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

The terms $P(Z)$ and $\Pr\{D = 1, Y = 1|X, P(Z)\}$ are identified, but the term $\Pr\{Y_1 = 1|X, P(Z), D = 0\}$ is not identified. Since Y is binary, this unidentified term is bounded from above and below by 1 and 0, so

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

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 $P(Z)$. Parallel reasoning provides bounds on $\Pr\{Y_0 = 1|X = x\}$.

The next lemma uses equation (1) 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. 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 1 or 0 that are identified.

LEMMA 2.1: *Suppose Y and D are determined by* (1) *and that Assumptions* 2.1 *and* 2.2 *hold*. *Let*

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

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 $p > p'$*.

For the proof, see the Supplemental Material (Shaikh and Vytlacil (2011)).

Before proceeding with the statement of the main theorem, we illustrate the use of Lemma 2.1 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

$$H(x, x') = E[h(x, x', P, P')|P > P'], \tag{2}$$

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, that is, $p > p'$ 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 2.1 that $\nu_1(1, x') \geq \nu_1(0, x)$. Therefore,

$$\begin{aligned}
&\Pr\{Y_0 = 1|D = 1, X = x, P = p\} \\
&\quad = \Pr\{\varepsilon_1 \leq \nu_1(0, X)|D = 1, X = x, P = p\} \\
&\quad \leq \Pr\{\varepsilon_1 \leq \nu_1(1, X)|D = 1, X = x', P = p\} \\
&\quad = \Pr\{Y = 1|D = 1, X = x', P = p\},
\end{aligned}$$

where the first and third equalities follow from equation (1), 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 0 and 1.

We now state our main theorem. 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, the supremum over the empty set is 0, and the infimum over the empty set is 1.

THEOREM 2.1: *Suppose Y and D are determined by* (1). *Let* $\mathbf{X}_{0+}(x) = \{x' : H(x, x') \geq 0\}$, $\mathbf{X}_{0-}(x) = \{x' : H(x, x') \leq 0\}$, $\mathbf{X}_{1+}(x) = \{x' : H(x', x) \geq 0\}$, *and* $\mathbf{X}_{1-}(x) = \{x' : H(x', x) \leq 0\}$, *where $H(x, x')$ is defined in* (2) *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 statements*:

(i) *If Assumptions 2.1 and 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*

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

$$L_1(x) = \sup_p \Big\{ \Pr\{D = 1, Y = 1 | X = x, P = p\} + \sup_{x' \in \mathbf{X}_{1+}(x)} \Pr\{D = 0, Y = 1 | X = x', P = p\} \Big\},$$

$$U_0(x) = \inf_p \Big\{ \Pr\{D = 0, Y = 1 | X = x, P = p\} + p \inf_{x' \in \mathbf{X}_{0+}(x)} \Pr\{Y = 1 | D = 1, X = x', P = p\} \Big\},$$

$$U_1(x) = \inf_p \Big\{ \Pr\{D = 1, Y = 1 | X = x, P = p\} + (1 - p) \inf_{x' \in \mathbf{X}_{1-}(x)} \Pr\{Y = 1 | D = 0, X = x', P = p\} \Big\}.$$

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

$$L_0(x) = \Pr\{D = 0, Y = 1 | X = x, P = \underline{p}\} + \sup_{x' \in \mathbf{X}_{0-}(x)} \Pr\{D = 1, Y = 1 | X = x', P = \underline{p}\},$$

$$L_1(x) = \Pr\{D = 1, Y = 1 | X = x, P = \overline{p}\} + \sup_{x' \in \mathbf{X}_{1+}(x)} \Pr\{D = 0, Y = 1 | X = x', P = \overline{p}\},$$

$$U_0(x) = \Pr\{D = 0, Y = 1 | X = x, P = \underline{p}\} + \underline{p} \inf_{x' \in \mathbf{X}_{0+}(x)} \Pr\{Y = 1 | D = 1, X = x', P = \underline{p}\},$$

$$U_1(x) = \Pr\{D = 1, Y = 1 | X = x, P = \overline{p}\} + (1 - \overline{p}) \inf_{x' \in \mathbf{X}_{1-}(x)} \Pr\{Y = 1 | D = 0, X = x', P = \overline{p}\},$$

where $\underline{p} = \inf\{p : p \in \operatorname{supp}(P)\}$ *and* $\overline{p} = \sup\{p : p \in \operatorname{supp}(P)\}$.

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

The proof is given in the Supplemental Material.

As a corollary, we have immediately that the sign of $\Delta G_1(x)$ is identified whenever $h(x, x, p, p')$ is well defined for some $p > p'$. This will be the case whenever Assumption 2.5 holds.

COROLLARY 2.1: *Suppose that Y and D satisfy* (1) *and that Assumptions* 2.1, 2.2, *and* 2.5 *hold. Then the sign of $\Delta G_1(x)$ is identified.*

REMARK 2.1: The bounds of Theorem 2.1 reduce to those in Manski (1989) if Assumption 2.5 does not hold. The bounds are 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 are different. See the analysis in Bhattacharya, Shaikh, and Vytlacil (2008) for details.

REMARK 2.2: It is interesting to ask when the upper and lower bounds will equal one another for the ASF or the ATE, that is, when the bounds imply that the ASF or the ATE is identified. Suppose that $\text{supp}(P, X) = \text{supp}(P) \times \text{supp}(X)$ and that the sets $\mathbf{X}_{d+}(x)$ and $\mathbf{X}_{d-}(x)$ for $d \in \{0, 1\}$ are nonempty. Consider $G_1(0, x)$. The analysis for $G_1(1, x)$ and $\Delta G_1(x)$ is similar. The width of the bounds on $G_1(0, x)$ is equal to

$$\begin{aligned}(3)\qquad &\inf_{x' \in \mathbf{X}_{0+}(x)} \Pr\{D = 1, Y = 1 | X = x', P = \underline{p}\} \\ &\quad - \sup_{x' \in \mathbf{X}_{0-}(x)} \Pr\{D = 1, Y = 1 | X = x', P = \underline{p}\}.\end{aligned}$$

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 (3) is less than or equal to

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

Since (3) is greater than or equal to 0 by construction, it follows that $G_1(0, x)$ is identified whenever there exists x^* such that $H(x, x^*) = 0$. Using Lemma 2.1, we may state this condition equivalently as the existence of a x^* such that $\nu_1(1, x) = \nu_1(0, x^*)$.

REMARK 2.3: It is worth noting that there are several testable implications of equation (1) and Assumptions 2.1 and 2.2. A straightforward implication is that $\Pr\{D = 1 | X, Z\}$ does not depend on X, and, as noted earlier, Lemma 2.1 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.

REFERENCES

BHATTACHARYA, J., A. SHAIKH, AND E. VYTLACIL (2008): "Treatment Effect Bounds Under Monotonicity Conditions: An Application to Swan–Ganz Catheterization," *American Economic Review, Papers and Proceedings*, 98, 351–356. [954]

CHESHER, A. (2005): "Nonparametric Identification Under Discrete Variation," *Econometrica*, 73, 1525–1550. [950]

CHIBURIS, R. (2010): "Semiparametric Bounds on Treatment Effects," *Journal of Econometrics*, 159, 267–275. [950]

JUN, S. J., J. PINKSE, AND H. XU (2009): "Tighter Bounds in Triangular Systems," Working Paper, Penn State University. [950]

MANSKI, C. (1989): "Anatomy of the Selection Problem," *Journal of Human Resources*, 24, 343–360. [954]

MANSKI, C., AND J. PEPPER (2000): "Monotone Instrumental Variables With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. [954]

SHAIKH, A. M., AND E. J. VYTLACIL (2011): "Supplement to 'Partial Identification in Triangular Systems of Equations With Binary Dependent Variables': Appendix," *Econometrica Supplemental Material*, 79, http://www.econometricsociety.org/ecta/Supmat/9082_proofs.pdf. [951]

VYTLACIL, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341. [949]

VYTLACIL, E., AND N. YILDIZ (2007): "Dummy Endogenous Variables in Weakly Separable Models," *Econometrica*, 75, 757–779. [949,950]

Dept. of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, U.S.A.; amshaikh@uchicago.edu

and

Dept. of Economics, Yale University, New Haven, CT 06520-8281, U.S.A.; edward.vytlacil@yale.edu.

Manuscript received February, 2010; final revision received October, 2010.