

# Supplement to “Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity”

(*Quantitative Economics*, Vol. 8, No. 2, July 2017, 589–610)

QUANG VUONG

Department of Economics, New York University

HAIQING XU

Department of Economics, University of Texas at Austin

## APPENDIX B

### B.1 *Additional proofs*

**PROOF OF LEMMA 3.** For the if part, suppose  $\Delta_0(\cdot, x, z_1, z_2)$  is piecewise weakly monotone on  $\mathcal{Y}_{0x|X=x}$ . Then, by definition,  $\mathcal{Y}_{0x|X=x}$  can be partitioned into a sequence of non-overlapping intervals  $\{I_j : j = 1, \dots, J\}$ , where  $J \in \mathbb{N} \cup \{+\infty\}$ , such that  $\Delta_0(\cdot, x, z_1, z_2)$  is weakly monotone on every interval. By assumption,  $\phi_x(\cdot)$  is continuous and strictly increasing on  $\mathcal{Y}_{0x|X=x}$ . Then  $\mathcal{Y}_{1x|X=x}$  can be partitioned into a sequence of non-overlapping intervals  $\{\phi_x(I_j) : j = 1, \dots, J\}$ . Moreover, by (9), we have

$$\Delta_1(y, x, z_1, z_2) = \Delta_0(\phi_x^{-1}(y), x, z_1, z_2), \quad \forall y \in \mathcal{Y}_{1x|X=x}.$$

Clearly,  $\Delta_1(\cdot, x, z_1, z_2)$  is weakly monotone in every interval  $\phi_x(I_j)$ . The only if part can be shown similarly. The proof for the strict part is similar.  $\square$

**PROOF OF LEMMA 4.** We first show the if part. W.l.o.g., let  $\Delta_0(\cdot, x, z_1, z_2)$  be piecewise strictly monotone on  $\mathcal{Y}_{0x|X=x}$ . Thus  $\mathcal{Y}_{0x|X=x}$  can be partitioned into a sequence of non-overlapping intervals  $\{I_j : j = 1, \dots, J\}$  such that  $\Delta_0(\cdot, x, z_1, z_2)$  is strictly monotone on each  $I_j$ . Moreover, we can merge successive intervals such that if  $\Delta_0(\cdot, x, z_1, z_2)$  is strictly increasing (decreasing) on  $I_j$ , then it is strictly decreasing (increasing) on  $I_{j+1}$ . By the proof of Lemma 3,  $\mathcal{Y}_{1x|X=x}$  can be partitioned into the same number of non-overlapping intervals  $\{I'_j \equiv \phi_x(I_j) : j = 1, \dots, J\}$ , such that  $\Delta_1(\cdot, x, z_1, z_2)$  is strictly increasing (decreasing) on each interval  $I'_j$  whenever  $\Delta_0(\cdot, x, z_1, z_2)$  is strictly increasing (decreasing) on  $I_j$ .

---

Quang Vuong: [quvuong@nyu.edu](mailto:quvuong@nyu.edu)

Haiqing Xu: [h.xu@austin.utexas.edu](mailto:h.xu@austin.utexas.edu)

Because the intervals  $\{I_j; j = 1, \dots, J\}$  and  $\{I'_j; j = 1, \dots, J\}$  partition  $\mathcal{S}_{Y_{0x}|X=x}$  and  $\mathcal{S}_{Y_{1x}|X=x}$ , respectively, with increasing left endpoints, we can pair them into  $\{(I_j, I'_j); j = 1, \dots, J\}$ . Thus, for every  $y \in I_j$ , we solve  $\phi_x(y)$  by (9) as

$$\phi_x(y) = \Delta_{1j}^{-1}(\Delta_{0j}(y, x, z_1, z_2), x, z_1, z_2),$$

where  $\Delta_{0j}(\cdot, x, z_1, z_2)$  and  $\Delta_{1j}(\cdot, x, z_1, z_2)$  are the restrictions of  $\Delta_0(\cdot, x, z_1, z_2)$  and  $\Delta_1(\cdot, x, z_1, z_2)$  to  $I_j$  and  $I'_j$ , respectively. Thus,  $\phi_x$  is identified on each  $I_j$  and hence on  $\mathcal{S}_{Y_{0x}|X=x} = \bigcup_{j=1}^J I_j$ .

For the only if part, w.l.o.g., suppose  $\Delta_0(\cdot, x, z_1, z_2)$  is constant on a nondegenerate interval  $I \subseteq \mathcal{S}_{Y_{0x}|X=x}$ . By the proof of Lemma 3,  $\Delta_1(\cdot, x, z_1, z_2)$  is also constant on  $\phi_x(I)$ . It suffices to construct a continuous and strictly increasing function  $\tilde{\phi}_x \neq \phi_x$  such that (9) holds for  $\tilde{\phi}_x$ . Let  $\tilde{\phi}_x(y) = \phi_x(y)$  for all  $y \notin I$  and  $\tilde{\phi}_x(y) = \phi_x(g(y))$  for all  $y \in I$ , where  $g$  is an arbitrary continuous, strictly increasing mapping from  $I$  onto  $I$ . Clearly, there are plenty of choices for such a function  $g$ . Moreover,  $\tilde{\phi}_x \neq \phi_x$  if  $g(t)$  is not an identity mapping. By construction,  $\Delta_1(\tilde{\phi}_x(y), x, z_1, z_2) = \Delta_1(\phi_x(y), x, z_1, z_2)$  holds for all  $y \in I$  because  $\Delta_1(\cdot, x, z_1, z_2)$  is constant on  $\phi_x(I)$ . This equation also holds for all  $y \notin I$  by the definition of  $\tilde{\phi}_x$ . Then (9) holds for  $\tilde{\phi}_x$ .  $\square$

### B.2 Proof of the statement in footnote 14

Given  $x \in \mathcal{S}_X$ . The Chernozhukov and Hansen (2005) condition is that the Jacobian matrix

$$II'(y_0, y_1) \equiv \begin{pmatrix} f(y_0, D=0|X=x, Z=z_1) & f(y_1, D=1|X=x, Z=z_1) \\ f(y_0, D=0|X=x, Z=z_2) & f(y_1, D=1|X=x, Z=z_2) \end{pmatrix}$$

is continuous and full-rank in  $(y_0, y_1) \in \mathcal{L}$ . In particular,  $\Delta_d(\cdot, x, z_1, z_2)$  is continuously differentiable in  $y_d \in \mathcal{S}_{Y_{dx}|X=x}$  for  $d = 0, 1$ .

We first show by contradiction that  $\partial\Delta_0(\cdot, x, z_1, z_2)/\partial y_0 \neq 0$  and  $\partial\Delta_1(\phi_x(\cdot), x, z_1, z_2)/\partial y_1 \neq 0$  on  $\mathcal{S}_{Y_{0x}|X=x}$ . Suppose not. Then there exists  $y_0^* \in \mathcal{S}_{Y_{0x}|X=x}$  such that  $\partial\Delta_0(y_0^*, x, z_1, z_2)/\partial y_0 = 0$  and  $\partial\Delta_1(\phi_x(y_0^*), x, z_1, z_2)/\partial y_1 = 0$  for some  $y_0^* \in \mathcal{S}_{Y_{0x}|X=x}$ . Because  $\partial\Delta_d(\cdot, x, z_1, z_2)/\partial y_d = (-1)^d[f(\cdot, D=d|X=x, Z=z_1) - f(\cdot, D=d|X=x, Z=z_2)]$ , it follows that  $II'(y_0^*, y_1^*)$  is not full rank, where  $y_1^* \equiv \phi_x(y_0^*) \in \mathcal{S}_{Y_{1x}|X=x}$ .

Because  $\mathcal{S}_{Y_{1x}|X=x} = \phi_x(\mathcal{S}_{Y_{0x}|X=x})$ , we have  $\partial\Delta_d(\cdot, x, z_1, z_2)/\partial y_d \neq 0$  on  $\mathcal{S}_{Y_{dx}|X=x}$  for  $d = 0, 1$ . Because of continuity,  $\partial\Delta_d(\cdot, x, z_1, z_2)/\partial y_d > 0$  or  $< 0$  on the projection of the rectangular  $\mathcal{L}$ , that is,  $\Delta_d(\cdot, x, z_1, z_2)$  is strictly monotone on the interval projection of  $\mathcal{L}$  onto  $\mathcal{S}_{Y_{dx}|X=x}$  for  $d = 0, 1$ .  $\square$

### B.3 Partial identification when Assumption D fails

In Section 2, Assumption D is a support condition that relates to the effectiveness of the instrumental variable. Lemma 1 highlights that, in addition to being a valid instrument, the instrumental variable also needs to be fully effective to extend point identification of the counterfactual mapping from the complier group to the whole population.<sup>19</sup> Other-

<sup>19</sup>Parameterization of  $\phi_x$  or assuming additive separability of  $\varepsilon$  in  $h$  can also achieve point identification of  $\phi_x$  on the full support of  $Y_{0x}$  given  $X = x$ . See also Angrist and Fernandez-Val (2013).

wise, there exists a nontrivial subgroup in the population for which exogenous variations of instrumental variables are not, in general, sufficient to point identify their treatment effects. When Assumption D fails, we call the instrumental variable  $Z$  partially effective. In this case, the counterfactual mapping on the support of the complier group provides bounds for the ITE and the structural function.

For a generic random variable  $W$ , let  $Q_W^*(\tau) = \sup\{w : \mathbb{P}(W \leq w) \leq \tau\}$  for any  $\tau \in [0, 1]$ . Because  $\{w : \mathbb{P}(W \leq w) \leq \tau\} \cup \{w : \mathbb{P}(W \leq w) \geq \tau\} = \mathbb{R}$ , we have  $Q_W(\tau) \leq Q_W^*(\tau)$  with equality if  $Q_W(\tau)$  is an inner point in the support of  $W$ . Moreover, we have  $Q_W^*(1) = +\infty$  and  $Q_W(0) = -\infty$ .

**COROLLARY 1.** *Suppose that Assumptions A to C hold. For every  $x \in \mathcal{S}_X$ , let  $z_1, z_2 \in \mathcal{S}_{Z|X=x}$  be such that  $p(x, z_1) < p(x, z_2)$ . Then the counterfactual mapping  $\phi_x(\cdot)$  is partially identified by*

$$\phi_{x,\ell}(y) \equiv Q_{Y_{1x}|C_x}(F_{Y_{0x}|C_x}(y)) \leq \phi_x(y) \leq Q_{Y_{1x}|C_x}^*(F_{Y_{0x}|C_x}(y)) \equiv \phi_{x,u}(y) \quad (11)$$

for all  $y \in \mathcal{S}_{Y|D=0, X=x}$ . This partially identifies the ITE by (4).<sup>20</sup> In addition, suppose that Assumption E holds. Then the structural function  $h(d, x, \tau)$  is partially identified by

$$h_\ell(d, x, \tau) \leq h(d, x, \tau) \leq h_u(d, x, \tau),$$

where  $h_s(1, x, \tau)$  is the  $\tau$ th quantile of the distribution  $\mathbb{P}[YD + \phi_{x,s}(Y)(1-D) \leq t | X = x]$ ;  $h_s(0, x, \tau)$  is the  $\tau$ th quantile of the distribution  $\mathbb{P}[\varphi_{x,s}(Y)D + Y(1-D) \leq t | X = x]$  with  $\varphi_{x,s}(\cdot)$  defined in footnote 20 for  $s = \ell, u$ .

Corollary 1 establishes interval identification for the ITE and the structural function  $h$  as well. Note that the upper and lower bounds of  $\phi_x(y)$  collapse to each other when  $y \in \mathcal{S}_{Y_{0x}|C_x}^\circ$ , the interior of  $\mathcal{S}_{Y_{0x}|C_x}$ . That is,  $\phi_x(\cdot)$  is point identified on  $\mathcal{S}_{Y_{0x}|C_x}^\circ$ . A similar property applies to the structural function  $h(d, x, \tau)$  when  $\tau \in \mathcal{S}_{\varepsilon|C_x}^\circ$ . Our bounds are constructive, but not necessarily sharp because of the exogeneity in Assumption E, which imposes restrictions beyond monotonicity

**PROOF OF COROLLARY 1.** By the proof of Lemma 1, we have

$$F_{Y_{0x}|C_x}(h(0, x, \tau)) = F_{Y_{1x}|C_x}(h(1, x, \tau))$$

for any  $\tau \in \mathcal{S}_{\varepsilon|C_x}^\circ$ . Since  $F_{Y_{1x}|C_x}$  is continuous and weakly increasing at  $h(1, x, \tau)$ , we have

$$Q_{Y_{1x}|C_x}(F_{Y_{0x}|C_x}(h(0, x, \tau))) \leq h(1, x, \tau) \leq Q_{Y_{1x}|C_x}^*(F_{Y_{0x}|C_x}(h(0, x, \tau))).$$

Let  $y = h(0, x, \tau) \in \mathcal{S}_{Y_{0x}|C_x}^\circ$ . Then  $h(1, x, \tau) = \phi_x(y)$  and the above equation becomes

$$\phi_{x,\ell}(y) \leq \phi_x(y) \leq \phi_{x,u}(y),$$

<sup>20</sup>Note that (11) implies that for all  $y \in \mathcal{S}_{Y|D=1, X=x}$ ,

$$\varphi_{x,\ell}(y) \equiv Q_{Y_{0x}|C_x}(F_{Y_{1x}|C_x}(y)) \leq \phi_x^{-1}(y) \leq Q_{Y_{0x}|C_x}^*(F_{Y_{1x}|C_x}(y)) \equiv \varphi_{x,u}(y).$$

It can be shown that  $\varphi_{x,\ell}(y) = \inf\{y_0 : \phi_{x,u}(y_0) \geq y\}$  and  $\varphi_{x,u}(y) = \sup\{y_0 : \phi_{x,\ell}(y_0) \leq y\}$ .

which gives the desired result by continuous extension to  $\mathcal{S}_{Y_{0x}|C_x} = \mathcal{S}_{Y|D=0, X=x}$ . By a similar argument, we obtain the result in footnote 20.

Moreover, by (6),  $\mathbb{P}[YD + \phi_{x,\ell}(Y)(1-D) \leq \cdot | X = x]$  and  $\mathbb{P}[YD + \phi_{x,u}(Y)(1-D) \leq \cdot | X = x]$  are the upper and lower bounds of  $\mathbb{P}(h(1, x, \varepsilon) \leq \cdot | X = x)$ , respectively. Therefore, their  $\tau$ th quantiles serve as the lower and upper bounds of  $h(1, x, \tau)$ , respectively. A similar argument holds for  $h(0, x, \tau)$ .  $\square$

#### B.4 Partial identification when Assumptions C' (ii) and D' fail

Motivated by Shaikh and Vytlacil (2011), Assumption C' (ii) can be relaxed, leading to partial identification. For generic random variables  $W$  and  $S$ , let  $K_{W|S}$  be the Kolmogorov (conditional) c.d.f. defined by  $K_{W|S}(w|s) = \mathbb{P}(W < w | S = s)$ . By definition, we have  $K_{W|S}(w|s) \leq F_{W|S}(w|s)$  with equality if  $w$  is not a mass point of  $W$  given  $S = s$ .

**COROLLARY 2.** *Suppose that Assumptions A', B, D', and E' hold. Fix  $x \in \mathcal{S}_X$ . Suppose also that Assumption C' (i) holds with  $\tilde{x}$ . Then  $\phi_{x,\tilde{x}}(\cdot)$  is partially identified by*

$$\begin{aligned} \underline{\phi}_{x,\tilde{x}}(\cdot) &\equiv Q_{Y_{1x}|m(x,z_1) < \eta \leq m(x,z_2)}(K_{Y_{0\tilde{x}}|m(\tilde{x},\tilde{z}_1) < \eta \leq m(\tilde{x},\tilde{z}_2)}(\cdot)) \leq \phi_{x,\tilde{x}}(\cdot) \\ &\leq Q_{Y_{1x}|m(x,z_1) < \eta \leq m(x,z_2)}(F_{Y_{0\tilde{x}}|m(\tilde{x},\tilde{z}_1) < \eta \leq m(\tilde{x},\tilde{z}_2)}(\cdot)) \equiv \overline{\phi}_{x,\tilde{x}}(\cdot), \quad \forall y \in \mathcal{S}_{Y|D=0, X=\tilde{x}}. \end{aligned}$$

Moreover, for any  $\tau \in (0, 1)$ , the lower and upper bounds of  $h(1, x, \tau)$  are identified as the  $\tau$ th quantile of the distributions  $\mathbb{P}(Y \leq \cdot; D = 1 | X = x, Z = z) + \mathbb{P}[\underline{\phi}_{x,\tilde{x}}(Y) \leq \cdot; D = 0 | X = \tilde{x}, Z = \tilde{z}]$  and  $\mathbb{P}(Y \leq \cdot; D = 1 | X = x, Z = z) + \mathbb{P}[\overline{\phi}_{x,\tilde{x}}(Y) \leq \cdot; D = 0 | X = \tilde{x}, Z = \tilde{z}]$ , respectively, where  $z \in \mathcal{S}_{Z|X=x}$  and  $\tilde{z} \in \mathcal{S}_{Z|X=\tilde{x}}$  satisfy  $p(x, z) = p(\tilde{x}, \tilde{z})$ .<sup>21</sup>

In Corollary 2, the lower and upper bounds collapse to each other if, in addition, Assumption C' (ii) holds for  $\tilde{x}$ .

Similarly to Corollary 1, we can also drop the support condition D' to construct bounds under Assumption C' (i).

**COROLLARY 3.** *Suppose that Assumptions A', B, and E' hold. Fix  $x \in \mathcal{S}_X$ . Suppose also that Assumption C' (i) holds with  $\tilde{x}$ . Then  $\phi_{x,\tilde{x}}(\cdot)$  is partially identified by*

$$\begin{aligned} \phi_{x,\tilde{x}}^L(\cdot) &\equiv Q_{Y_{1x}|m(x,z_1) < \eta \leq m(x,z_2)}(K_{Y_{0\tilde{x}}|m(\tilde{x},\tilde{z}_1) < \eta \leq m(\tilde{x},\tilde{z}_2)}(\cdot)) \leq \phi_{x,\tilde{x}}(\cdot) \\ &\leq Q_{Y_{1x}|m(x,z_1) < \eta \leq m(x,z_2)}^*(F_{Y_{0\tilde{x}}|m(\tilde{x},\tilde{z}_1) < \eta \leq m(\tilde{x},\tilde{z}_2)}(\cdot)) \equiv \phi_{x,\tilde{x}}^H(\cdot), \quad \forall y \in \mathcal{S}_{Y|D=0, X=\tilde{x}}. \end{aligned}$$

Moreover, for any  $\tau \in (0, 1)$ , the lower and upper bounds of  $h(1, x, \tau)$  are identified as the  $\tau$ th quantiles of the distributions  $\mathbb{P}(Y \leq \cdot; D = 1 | X = x, Z = z) + \mathbb{P}[\phi_{x,\tilde{x}}^L(Y) \leq \cdot; D = 0 | X = \tilde{x}, Z = \tilde{z}]$  and  $\mathbb{P}(Y \leq \cdot; D = 1 | X = x, Z = z) + \mathbb{P}[\phi_{x,\tilde{x}}^H(Y) \leq \cdot; D = 0 | X = \tilde{x}, Z = \tilde{z}]$ , respectively, where  $z \in \mathcal{S}_{Z|X=x}$  and  $\tilde{z} \in \mathcal{S}_{Z|X=\tilde{x}}$  satisfy  $p(x, z) = p(\tilde{x}, \tilde{z})$ .

<sup>21</sup>This corollary was first presented at the Bilkent Workshop in Econometric Theory and Applications, 2013.

When there are multiple  $\tilde{x}$  satisfying Assumption C'(i), we can use them jointly to tighten our bounds in Corollaries 2 and 3 by taking the sup and the inf over such  $\tilde{x}$  following Shaikh and Vytlacil (2011). For the same reason as in Appendix B.3, our bounds are constructive, but not necessarily sharp because of Assumption E'.

**PROOF OF COROLLARY 2.** Fix  $x, \tilde{x} \in \mathcal{X}$  satisfying Assumption C'(i). Let further  $z_1, z_2 \in \mathcal{Z}|_{X=x}$  and  $\tilde{z}_1, \tilde{z}_2 \in \mathcal{Z}|_{X=\tilde{x}}$  such that  $p(x, z_1) = p(\tilde{x}, \tilde{z}_1) < p(x, z_2) = p(\tilde{x}, \tilde{z}_2)$ . By the proof of Corollary 1, we have

$$\mathcal{Q}_{Y_{1x}|m(x,z_1) < \eta \leq m(x,z_2)}(F_{Y_{0\tilde{x}}|m(\tilde{x},\tilde{z}_1) < \eta \leq m(\tilde{x},\tilde{z}_2)}(h(0, \tilde{x}, \tau))) = h(1, x, \psi(0, \tilde{x}, \tau)) \geq h(1, x, \tau),$$

where the last inequality comes from the fact that  $\psi(0, \tilde{x}, \tau) \geq \tau$ . Therefore,

$$\begin{aligned} \mathcal{Q}_{Y_{1x}|m(x,z_1) < \eta \leq m(x,z_2)}(F_{Y_{0\tilde{x}}|m(\tilde{x},\tilde{z}_1) < \eta \leq m(\tilde{x},\tilde{z}_2)}(y)) &\geq h(1, x, h^{-1}(0, \tilde{x}, y)) \\ &= \phi_{x,\tilde{x}}(y), \quad y \in \mathcal{Y}_{0\tilde{x}}. \end{aligned}$$

For the lower bound, let  $\varphi(0, \tilde{x}, \tau) = \inf\{e : h(0, \tilde{x}, e) = h(0, \tilde{x}, \tau)\}$ . By Assumption A', we have

$$K_{Y_{0\tilde{x}}|m(\tilde{x},\tilde{z}_1) < \eta \leq m(\tilde{x},\tilde{z}_2)}(h(0, \tilde{x}, \tau)) = F_{\varepsilon|m(\tilde{x},\tilde{z}_1) < \eta \leq m(\tilde{x},\tilde{z}_2)}(\varphi(0, \tilde{x}, \tau)).$$

Therefore,

$$\mathcal{Q}_{Y_{1x}|m(x,z_1) < \eta \leq m(x,z_2)}(K_{Y_{0\tilde{x}}|m(\tilde{x},\tilde{z}_1) < \eta \leq m(\tilde{x},\tilde{z}_2)}(h(0, \tilde{x}, \tau))) = h(1, x, \varphi(0, \tilde{x}, \tau)) \leq h(1, x, \tau),$$

which provides a lower bound for  $h(1, x, h^{-1}(0, \tilde{x}, \cdot))$ .

Given the interval identification of  $\phi_{x,\tilde{x}}(\cdot)$ , the bounds of  $h(1, x, \cdot)$  follow a similar argument as in the proof in Corollary 1.  $\square$

The proof of Corollary 3 simply follows the arguments in Corollaries 1 and 2.

### B.5 Characterization of model restrictions

In empirical applications, an important question is whether to adopt a model with or without the selection equation (2). The selection equation provides a simple and constructive identification result, but introduces additional restrictions on the data. We now characterize all such restrictions. These are useful for developing model selection and model specification tests.

Formally, we denote these two models by

$$\mathcal{M}_0 \equiv \{[h, F_{\varepsilon D|XZ}] : \text{Assumption A holds}\},$$

$$\mathcal{M}_1 \equiv \{[h, m, F_{\varepsilon \eta|XZ}] : \text{Assumptions A and B hold}\}.$$

To simplify, hereafter we assume  $\mathcal{XZ} = \mathcal{X} \times \{z_1, z_2\}$  as well as  $p(x, z_1) < p(x, z_2)$  for all  $x \in \mathcal{X}$ ; see Assumption C. On the other hand, we do not impose Assumptions D to F which identify these two models. We say that a conditional distribution  $F_{YD|XZ}$  of observables is rationalized by model  $\mathcal{M}$  if and only if there exists a structure in  $\mathcal{M}$  that generates  $F_{YD|XZ}$ . The next lemma provides a necessary and sufficient condition for  $\mathcal{M}_0$  to be rationalized by the data.

LEMMA 5. A conditional distribution  $F_{YD|XZ}$  can be rationalized by  $\mathcal{M}_0$  if and only if the following statements hold:

- (i) The function  $F_{Y|DXZ}(\cdot|d, x, z)$  is a continuous conditional c.d.f.
- (ii) For every  $x \in \mathcal{S}_X$ , there exists a continuous and strictly increasing mapping  $g_x : \mathbb{R} \rightarrow \mathbb{R}$  such that  $g_x$  maps  $\mathcal{S}_{(1-D)Y+Dg_x^{-1}(Y)|X=x}$  onto  $\mathcal{S}_{DY+(1-D)g_x(Y)|X=x}$ ; the support  $\mathcal{S}_{(1-D)Y+Dg_x^{-1}(Y)|X=x}$  is a nondegenerate subset of  $\mathbb{R}$ . Moreover, we have

$$\Delta_0(\cdot, x, z_1, z_2) = \Delta_1(g_x(\cdot), x, z_1, z_2) \quad \text{on } \mathcal{S}_{(1-D)Y+Dg_x^{-1}(Y)|X=x}. \quad (12)$$

In Lemma 5, the key restriction is the existence of a solution to (12), which may not be unique.

We now turn to  $\mathcal{M}_1$ .

THEOREM 5. A conditional distribution  $F_{YD|XZ}$  rationalized by  $\mathcal{M}_0$  can also be rationalized by  $\mathcal{M}_1$  if and only if for any  $x \in \mathcal{S}_X$ ,  $\Delta_d(\cdot, x, z_1, z_2)$  for  $d \in \{0, 1\}$  are continuous and strictly increasing on  $\mathcal{S}_{(1-D)Y+Dg_x^{-1}(Y)|X=x}^o$  and  $\mathcal{S}_{DY+(1-D)g_x(Y)|X=x}^o$ , respectively.

Note that, by definition,  $\Delta_d(\cdot, x, z_1, z_2)$  for  $d = 0, 1$  are flat elsewhere. Specifically, from their definitions, they vanish below their supports and are equal to  $\mathbb{P}(D = 1|X = x, Z = z_2) - \mathbb{P}(D = 1|X = x, Z = z_1)$  above their supports.

PROOF OF LEMMA 5. For the only if part, note that Assumptions A(ii) and A(iii) imply  $F_{\varepsilon|X,Z}(\cdot|x, z)$  is continuous. Moreover, because

$$F_{\varepsilon|X,Z}(\cdot|x, z) = \sum_{d=0,1} F_{\varepsilon|DXZ}(\cdot|d, x, z) \times \mathbb{P}(D = d|X = x, Z = z), \quad (13)$$

where  $F_{\varepsilon|DXZ}(\cdot|d, x, z)$  are monotone functions and  $\mathbb{P}(D = d|X = x, Z = z)$  are non-negative, then  $F_{\varepsilon|DXZ}(\cdot|d, x, z)$  is also continuous. By Assumption A(i), condition (i) holds. For (ii), let  $g_x(\cdot) = \phi_x(\cdot)$  on  $\mathcal{S}_{Y_{0x}|X=x}$ . Thus, conditional on  $X = x$ , we have  $(1 - D)Y + Dg_x^{-1}(Y) = h(0, x, \varepsilon)$  and  $DY + (1 - D)g_x(Y) = h(1, x, \varepsilon)$ . The desired result follows from Section 3.

For the if part, we construct a structure  $S = [h, F_{\varepsilon D|XZ}]$  that rationalizes the given distribution  $F_{YD|XZ}$ . Fix an arbitrary  $x$ . Let  $h(0, x, e) = e$  and  $h(1, x, e) = g_x(e)$  for  $e \in \mathbb{R}$ . We define the distribution  $F_{\varepsilon D|XZ}$  by the conditional distribution of  $(g_x^{-1}(Y) \cdot D + Y \cdot (1 - D), D)$  given  $X$  and  $Z$ . It is straightforward to see that the constructed structure  $S$  satisfies Assumption A(i). Moreover, Assumption A(iii) is satisfied by integrating out  $z$  in (13). We now verify Assumption A(ii), that is,  $Z \perp \varepsilon|X = x$ .

By (12), we have

$$\begin{aligned} & \mathbb{P}[Y \leq g_x(y); D = 1|X = x, Z = z_1] + \mathbb{P}[Y \leq y; D = 0|X = x, Z = z_1] \\ &= \mathbb{P}[Y \leq g_x(y); D = 1|X = x, Z = z_2] + \mathbb{P}[Y \leq y; D = 0|X = x, Z = z_2] \\ &= \mathbb{P}[Y \leq g_x(y); D = 1|X = x] + \mathbb{P}[Y \leq y; D = 0|X = x], \end{aligned}$$

where the second equality is because  $Z$  is binary. Hence, for any  $\tau \in [0, 1]$  and  $z = z_1, z_2$ ,

$$\begin{aligned} \mathbb{P}_S(\varepsilon \leq \tau | X = x, Z = z) \\ = \mathbb{P}(Y \leq \tau; D = 0 | X = x, Z = z) + \mathbb{P}(Y \leq g_x(\tau); D = 1 | X = x, Z = z), \end{aligned}$$

which is invariant with  $z \in \{z_1, z_2\}$  by the previous equality. Therefore,  $Z \perp \varepsilon | X$ .

Now it suffices to show that the constructed structure  $S = [h, F_{\varepsilon D | X Z}]$  generates the given distribution  $F_{Y D | X Z}$ . This is true because for any  $(y, d, x, z) \in \mathcal{S}_{Y D X Z}$ , we have

$$\begin{aligned} \mathbb{P}_S(Y \leq y; D = 0 | X = x, Z = z) &= \mathbb{P}_S(\varepsilon \leq y; D = 0 | X = x, Z = z) \\ &= \mathbb{P}(Y \leq y; D = 0 | X = x, Z = z). \end{aligned}$$

The last step comes from the construction of  $F_{\varepsilon D | X Z}$ . Similarly,

$$\begin{aligned} \mathbb{P}_S(Y \leq y; D = 1 | X = x, Z = z) &= \mathbb{P}_S(g_x(\varepsilon) \leq y; D = 1 | X = x, Z = z) \\ &= \mathbb{P}(Y \leq y; D = 1 | X = x, Z = z). \quad \square \end{aligned}$$

**PROOF OF THEOREM 5.** For the only if part, let  $g_x(\cdot) = \phi_x(\cdot)$  on  $\mathcal{S}_{Y_{0x} | X=x}$ . Thus, conditional on  $X = x$ , we have  $(1 - D)Y + Dg_x^{-1}(Y) = h(0, x, \varepsilon)$  and  $DY + (1 - D)g_x(Y) = h(1, x, \varepsilon)$ . Hence,  $\mathcal{S}_{(1-D)Y + Dg_x^{-1}(Y) | X=x}^o = \mathcal{S}_{Y_{0x} | X=x}^o$  and  $\mathcal{S}_{DY + (1-D)g_x(Y) | X=x}^o = \mathcal{S}_{Y_{1x} | X=x}^o$ . From the proof of Lemma 1 it follows that  $\Delta_d(\cdot, x, z_1, z_2)$  is continuous and strictly increasing on  $\mathcal{S}_{Y_{dx} | X=x}^o$ .

We now show the if part. Suppose the strict monotonicity of  $\Delta_d(\cdot, x, z_1, z_2)$  holds for a structure  $S_0 = [h, F_{\varepsilon D | X Z}]$  satisfying Assumption A. Fix  $X = x$ . For notational simplicity, we suppress  $x$  and hence the conditioning on  $X = x$  in the following analysis. It suffices to construct an observationally equivalent structure  $S_1 \in \mathcal{M}_1$ . First, let  $\eta \sim U(0, 1)$  and  $m(z) = p(z)$  so that  $\mathbb{P}_{S_1}(D = 1 | X = x, Z = z) = p(x, z)$  when (2) holds. Next, let  $h(0, e) = e$  and  $h(1, e) = g(e)$  for all  $e \in \mathbb{R}$ . To construct  $S_1$ , it suffices to define  $F_{\varepsilon | \eta}$ . For any  $e \in \mathbb{R}$ , let

$$\mathbb{P}_{S_1}(\varepsilon \leq e | \eta) = \begin{cases} \frac{\mathbb{P}(Y \leq h(1, e), D = 1 | Z = z_1)}{m(z_1)}, & \text{if } \eta \leq m(z_1), \\ \frac{\Delta_1(h(1, e), z_1, z_2)}{m(z_2) - m(z_1)}, & \text{if } m(z_1) < \eta \leq m(z_2), \\ \frac{\mathbb{P}(Y \leq h(0, e), D = 0 | Z = z_2)}{1 - m(z_2)}, & \text{if } \eta > m(z_2). \end{cases}$$

By construction and Lemma 5(i),  $\mathbb{P}_{S_1}(\varepsilon \leq e | \eta)$  is continuous and (weakly) increasing in  $e \in \mathbb{R}$  since  $h(d, \cdot)$ , and  $\Delta_d(\cdot, z_1, z_2)$  are continuous and strictly increasing functions. Given  $\eta \sim U[0, 1]$ ,  $(\varepsilon, \eta)$  are jointly continuously distributed. Moreover, let  $F_{\varepsilon \eta | Z} = F_{\varepsilon \eta}$ . Hence,  $F_{\varepsilon \eta}$  is a proper conditional c.d.f. that satisfies Assumptions A(ii), A(iii), B(i), and B(ii). Using the constructed  $m, h$ , and  $F_{\varepsilon, \eta}$ , it is straightforward that  $S_1 \in \mathcal{M}_1$  is observationally equivalent to  $S_0$ .  $\square$

## REFERENCES

- Angrist, I. D. and I. Fernandez-Val (2013), “ExtrapoLATE-ing: External validity and overidentification in the late framework.” In *Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress*, Vol. 51, 401. Cambridge University Press. [2]
- Chernozhukov, V. and C. Hansen (2005), “An IV model of quantile treatment effects.” *Econometrica*, 73 (1), 245–261. [2]
- Shaikh, A. M. and E. J. Vytlacil (2011), “Partial identification in triangular systems of equations with binary dependent variables.” *Econometrica*, 79 (3), 949–955. [4, 5]

---

Co-editor Rosa L. Matzkin handled this manuscript.

Manuscript received 19 May, 2015; final version accepted 9 October, 2016; available online 25 October, 2016.