Econometrica Supplementary Material

SUPPLEMENT TO "IDENTIFYING EFFECTS OF MULTIVALUED TREATMENTS"

(*Econometrica*, Vol. 86, No. 6, November 2018, 1939–1963)

SOKBAE LEE
Department of Economics, Columbia University and Institute for Fiscal Studies

BERNARD SALANIÉ
Department of Economics, Columbia University

Appendix A gives an identification result for the zero-index case, which was not dealt with in the text. It also provides a characterization of Heckman and Pinto's unordered monotonicity property as a subcase of our more general framework. Appendix B collects proofs of some of the results in the main text. Appendix C fills in the details of the entry game introduced in Section 2, and Appendix D compares our results with those of Heckman, Urzua, and Vytlacil (2008) in more detail. Finally, Appendix E discusses a more general form of threshold conditions than the "rectangular" threshold conditions in Assumption 2.1.

APPENDIX A: ADDITIONAL RESULTS

A.1. *Identification With a Zero Index*

THEOREM 3.1 REQUIRED THAT the index of treatment k be nonzero (Assumption 3.1). It therefore does not apply to Example 3, for instance. Recall that in that example,

$$D_0 = \mathcal{D}_0(\boldsymbol{S}) = 1 - S_1 - S_2 - S_3 + S_1 S_2 + S_1 S_3 + S_2 S_3$$

and treatment 0 has degree $m^0 = 2 < J^0 = 3$.

Note, however, that steps 1 and 2 of the proof of Theorem 3.1 apply to zero-index treatments as well; the relevant polynomial of Heaviside functions has leading term

$$H(q_1 - v_1)H(q_2 - v_2) + H(q_1 - v_1)H(q_3 - v_3) + H(q_2 - v_2)H(q_3 - v_3),$$

and we can take the derivative in (q_1, q_2), for instance, to obtain an equation that replaces (6.4):

$$\frac{\partial^2}{\partial q_1 \partial q_2} B_0(\boldsymbol{q}) = \int b_0(q_1, q_2, v_3)\, dv_3.$$

Applying this to $B_0(\boldsymbol{q}) = \Pr[D = 0 | \boldsymbol{Q}(\boldsymbol{Z}) = \boldsymbol{q}]$ and $b_0(\boldsymbol{v}) = f_{\boldsymbol{V}}(\boldsymbol{v})$, and then to $B_0(\boldsymbol{q}) = E[Y D_0 | \boldsymbol{Q}(\boldsymbol{Z}) = \boldsymbol{q}]$ and $b_0(\boldsymbol{v}) = E[G(Y_0) | \boldsymbol{V} = \boldsymbol{v}] f_{\boldsymbol{V}}(\boldsymbol{v})$, identifies

$$\int f_{V_1,V_2,V_3}(q_1, q_2, v_3)\, dv_3 = f_{V_1,V_2}(v_1, v_2)$$

Sokbae Lee: sl3841@columbia.edu
Bernard Salanié: bsalanie@columbia.edu

© 2018 The Econometric Society https://doi.org/10.3982/ECTA14269

and

$$\int E\big[G(Y_0)|V_1=q_1,V_2=q_2,V_3=v_3\big]f_{V_1,V_2,V_3}(q_1,q_2,v_3)\,dv_3$$
$$=E\big[G(Y_0)|V_1=q_1,V_2=q_2\big]f_{V_1,V_2}(v_1,v_2).$$

Dividing through identifies a local counterfactual outcome:

$$E\big[G(Y_0)|V_1=q_1,V_2=q_2\big].$$

Under Assumption 3.5, this also identifies $EG(Y_0)$. Moreover, we can apply the same logic to the pairs (q_1,q_3) and (q_2,q_3) to get further information on the treatment effects.

This argument applies more generally. It allows us to state the following theorem:

THEOREM A.1—Identification With a Zero Index: *Let Assumptions* 2.1, 2.2, *and* 3.2 *hold. Fix a value* $\boldsymbol{q}$ *in* $\tilde{\mathcal{Q}}$, *so that Assumptions* 3.3 *and* 3.4 *also hold at* $\boldsymbol{q}$. *Let m be the degree of treatment k. Take l to be any subset of* $\boldsymbol{J}$ *that corresponds to a leading term in the expansion of the indicator function of* $\{D=k\}$. *Denote* $\widetilde{T}$ *the differential operator*

$$\widetilde{T}=\frac{\partial^m}{\prod\limits_{i=1,\dots,m}\partial_{l_i}}.$$

Then, for $\boldsymbol{q}=(\boldsymbol{q}^l,\boldsymbol{q}^{J-l})$,

$$f_{V^l}\big(\boldsymbol{q}^l\big)=\frac{1}{c_l^k}\widetilde{T}\Pr[D=k|\boldsymbol{Q}(\boldsymbol{Z})=\boldsymbol{q}],$$
$$E\big[G(Y_k)|V^l=\boldsymbol{q}^l\big]=\frac{\widetilde{T}E\big[G(Y)D_k|\boldsymbol{Q}(\boldsymbol{Z})=\boldsymbol{q}\big]}{\widetilde{T}\Pr\big[D=k|\boldsymbol{Q}(\boldsymbol{Z})=\boldsymbol{q}\big]}.$$

PROOF OF THEOREM A.1: The proof of Theorem A.1 is basically the same as that of Theorem 3.1. Steps 1 and 2 of the proof of Theorem 3.1 do not rely on any assumption about indices. They show that if we define

$$W_l(\boldsymbol{q})=\int\prod_{j\in l}H(q_j-v_j)b_k(\boldsymbol{v})\,d\boldsymbol{v},$$

where the set $l\subset\boldsymbol{J}$, then its cross-derivative with respect to $(\boldsymbol{q}^l)$ is

$$\int b_k\big(\boldsymbol{q}^l,\boldsymbol{v}_{-l}\big)\,d\boldsymbol{v}_{-l},$$

where $\boldsymbol{v}_{-l}$ collects all components of $\boldsymbol{v}$ whose indices are not in l.

Now let m be the degree of treatment k. In the sum (6.3), take any term l such that $|l|=m$. Recall that $\widetilde{T}$ denotes the differential operator

$$\widetilde{T}=\frac{\partial^m}{\prod\limits_{i=1,\dots,m}\partial_{l_i}}.$$

By the formula above, applying $\widetilde{T}$ to term l gives

$$c_l \int b_k\big(\boldsymbol{q}^l, \boldsymbol{v}_{-l}\big)\, d\boldsymbol{v}_{-l}.$$

Moreover, applying $\widetilde{T}$ to any other term l' obviously gives zero if term l' has degree less than m. Now take any other term l' of degree m. As $\widetilde{T}$ takes at least one derivative along a direction that is not in l', that term must also contribute zero.

This proves that

$$\widetilde{T}B_k(\boldsymbol{q}) = c_l^k \int b_k\big(\boldsymbol{q}^l, \boldsymbol{v}_{-l}\big)\, d\boldsymbol{v}_{-l};$$

note that it also implies that $\widetilde{T}B_k(\boldsymbol{q})$ only depends on $\boldsymbol{q}^l$.

Applying this first to $b_k(\boldsymbol{v}) = f_V(\boldsymbol{v})$ and $B_k(\boldsymbol{q}) = \Pr(D = k|\boldsymbol{Q}(\boldsymbol{Z}) = \boldsymbol{q})$, then to $b_k(\boldsymbol{v}) = E[G(Y_k)|\boldsymbol{V} = \boldsymbol{v}]f_V(\boldsymbol{v})$ and $B_k(\boldsymbol{q}) = E[G(Y)D_k|\boldsymbol{Q}(\boldsymbol{Z}) = \boldsymbol{q}]$ exactly as in the proof of Theorem 3.1, we get

$$\int f_V\big(\boldsymbol{q}^l, \boldsymbol{v}_{-l}\big)\, d\boldsymbol{v}_{-l} = \frac{1}{c_l^k}\widetilde{T}\Pr(D = k|\boldsymbol{Q}(\boldsymbol{Z}) = \boldsymbol{q}),$$

$$\int E\big[G(Y_k)|\boldsymbol{V} = \big(\boldsymbol{q}^l, \boldsymbol{v}_{-l}\big)\big]f_V\big(\boldsymbol{q}^l, \boldsymbol{v}_{-l}\big)\, d\boldsymbol{v}_{-l} = \frac{1}{c_l^k}\widetilde{T}E\big(G(Y)D_k|\boldsymbol{Q}(\boldsymbol{Z}) = \boldsymbol{q}\big).$$

Since the left-hand sides are simply $f_{V^l}(\boldsymbol{v}^l)$ and $E[G(Y_k)|V^l = \boldsymbol{q}^l]f_{V^l}(\boldsymbol{v}^l)$, the conclusion of the theorem follows immediately. *Q.E.D.*

Theorem A.1 is a generalization of Theorem 3.1 (just take $m = J$). It calls for three remarks. First, we could weaken its hypotheses somewhat. We could, for instance, replace $(0,1)^J$ with $(0,1)^m$ in the statement of Assumption 3.5.

Second, when $m < J$, the treatment effects are over-identified. This is obvious from the equalities in Theorem A.1, in which the right-hand side depends on $\boldsymbol{q}$ but the left-hand side only depends on $\boldsymbol{q}^l$.

Finally, considering several treatment values can identify even more, since $\boldsymbol{V}$ is assumed to be the same across k. Theorem 3.1 would then imply that if there is any treatment value k with a nonzero index, then the joint density f_V is identified from that treatment value.

A.2. *Further Analysis of Unordered Monotonicity*

Our formalism allows us to derive a new characterization of the unordered monotonicity property defined by Heckman and Pinto (2018). Take any treatment value k. In our model, a change in instruments $\boldsymbol{Z}$ acts on the treatment assigned to an observation with unobserved characteristics $\boldsymbol{V}$ through the indicator functions $S_j = \mathbb{1}(V_j < Q_j(\boldsymbol{Z}))$, which depend on the thresholds $\boldsymbol{Q}(\boldsymbol{Z})$.

Unordered monotonicity requires that there exist changes in thresholds $\Delta\boldsymbol{Q}$ such that, for $\boldsymbol{Q}' = \boldsymbol{Q} + \Delta\boldsymbol{Q}$,

$$\Pr\big\{d_k(\boldsymbol{V}, \boldsymbol{Q}) = 0 \text{ and } d_k\big(\boldsymbol{V}, \boldsymbol{Q}'\big) = 1\big\} \times \Pr\big\{d_k(\boldsymbol{V}, \boldsymbol{Q}) = 1 \text{ and } d_k\big(\boldsymbol{V}, \boldsymbol{Q}'\big) = 0\big\} = 0,$$

where the probabilities are computed over the joint distribution of $\boldsymbol{V}$.

In our framework, several thresholds are typically relevant for each treatment value. This makes the analysis of unordered monotonicity complex in general. To understand why, we start from the expression (2.2) of D_k as a polynomial of $\boldsymbol{S} = (S_1, \ldots, S_J)$ for $S_j(\boldsymbol{V}, \boldsymbol{Q}) = \mathbb{1}(V_j < Q_j)$. For any change in thresholds $\Delta\boldsymbol{Q}$ that induces changes in the indicators $\Delta\boldsymbol{S}$, Taylor's theorem yields

$$\Delta D_k = \sum_{m=1}^{J} \sum_{\alpha_1+\cdots+\alpha_J=m} \frac{1}{\alpha_1!\alpha_2!\cdots\alpha_J!} \frac{\partial^m \mathcal{D}_k(\boldsymbol{S})}{\partial S_1^{\alpha_1} \partial S_2^{\alpha_2} \cdots \partial S_J^{\alpha_J}} \prod_{l=1}^{J} \Delta S_l^{\alpha_l}, \tag{A.1}$$

where α_j is a nonnegative integer for $j = 1, \ldots, J$. Note that this is an exact expansion since $\mathcal{D}_k$ is a polynomial. Moreover, note that, given a change in one threshold ΔQ_j, only S_j changes and

$$\Delta S_j = \mathbb{1}(0 < V_j - Q_j < \Delta Q_j) - \mathbb{1}(\Delta Q_j < V_j - Q_j < 0). \tag{A.2}$$

(We do not need to distinguish between the weak and strict inequalities since the distribution of V_j is absolutely continuous with respect to the Lebesgue measure.)

The changes ΔS_j can only take the values 0 or ± 1. In general, higher order terms in expansion (A.1) may be nonzero. However, if the changes in thresholds $\Delta\boldsymbol{Q}$ are small, then we can neglect the higher order terms since the values of $\boldsymbol{V}$ for which several ΔS_j are nonzero occur with very small probability. To make this more precise, we use the following definition:

DEFINITION A.1—Two-Way Flows: A change in thresholds $\Delta\boldsymbol{Q}$ generates two-way flows for treatment value k if and only if

$$\lim_{\varepsilon\to 0}\left(\frac{\Pr\big(D_k(0) = 0 \text{ and } D_k(\varepsilon) = 1\big)}{\varepsilon} \times \frac{\Pr\big(D_k(0) = 1 \text{ and } D_k(\varepsilon) = 0\big)}{\varepsilon}\right) > 0$$

for $D_k(\varepsilon) \equiv d_k(\boldsymbol{V}, \boldsymbol{Q} + \varepsilon\Delta\boldsymbol{Q})$.

We now provide new characterizations of unordered monotonicity.

THEOREM A.2—Characterizing Unordered Monotonicity in the Small: *Fix a value $\boldsymbol{Q}$ of the thresholds. Denote*

$$\nabla\mathcal{D}_k(\boldsymbol{S}) = \frac{\partial \mathcal{D}_k}{\partial \boldsymbol{S}}(\boldsymbol{S}).$$

Assume that $J \geq 2$ and that there exist two values $j_1 \neq j_2$ such that $\nabla_{j_1}\mathcal{D}_k$ and $\nabla_{j_2}\mathcal{D}_k$ are not identically zero. Then:

1. *If each component of $\nabla\mathcal{D}_k(\boldsymbol{S})$ has a constant sign when $\boldsymbol{S}$ varies over $\{0, 1\}^J$, then some changes in thresholds do not generate two-way flows, and some others do.*
2. *If the sign of any component $\nabla_j\mathcal{D}_k(\boldsymbol{S})$ changes when S_j switches between 0 and 1, then any change in thresholds generates two-way flows.*

(*In these two statements, we take 0 to have the same sign as both -1 and $+1$.*)

PROOF OF THEOREM A.2: Take $\varepsilon > 0$ small. Remember that, given a change in thresholds $\varepsilon\Delta Q_j$,

$$\Delta S_j = \mathbb{1}(0 < V_j - Q_j < \varepsilon\Delta Q_j) - \mathbb{1}(\varepsilon\Delta Q_j < V_j - Q_j < 0),$$

which is zero or has the sign of ΔQ_j.

Under our assumptions on the distribution of $\boldsymbol{V}$, the probability that $\Delta S_j \neq 0$ is of order ε; the probability that $\Delta S_j \Delta S_l \neq 0$ is of order ε^2, etc. Given Definition A.1, we only need to work on the first-order terms in expansion (A.1) since the other terms generate vanishingly small corrections. That is, we use

$$\begin{aligned}\Delta D_k &\simeq \sum_{j=1}^{J} \nabla_j \mathcal{D}_k(\boldsymbol{S}) \times \Delta S_j \\ &= \sum_{j=1}^{J} \nabla_j \mathcal{D}_k(\boldsymbol{S}) \times \big(\mathbb{1}(0 < V_j - Q_j < \varepsilon \Delta Q_j) - \mathbb{1}(\varepsilon \Delta Q_j < V_j - Q_j < 0)\big).\end{aligned} \tag{A.3}$$

• *Proof of part 1:* To prove part 1 of the theorem, assume that each derivative $\nabla_j \mathcal{D}_k$ has a constant sign, independent of $\boldsymbol{S} \in \{0, 1\}^J$.

Then it is easy to find changes $\boldsymbol{\Delta Q}$ that only generate one-way flows. First, take each ΔQ_j to have the sign of $\nabla_j \mathcal{D}_k$.

Since each ΔS_j has the sign of the corresponding ΔQ_j, each product term in the sum (A.3) is nonnegative, and so is the change in D_k. Obviously, changing the signs of all ΔQ_j's would generate one-way flows in the opposite direction.

It is equally easy to find changes in instruments that generate two-way flows. Take the indices j_1 and j_2 referred to in the statement of the theorem. Take $\Delta Q_m = 0$ for $m \neq j_1, j_2$. Then expansion (A.3) becomes

$$\Delta D_k \simeq \nabla_{j_1} \mathcal{D}_k(\boldsymbol{S}) \times \Delta S_{j_1} + \nabla_{j_2} \mathcal{D}_k(\boldsymbol{S}) \times \Delta S_{j_2}.$$

Choose some $\Delta Q_{j_1}, \Delta Q_{j_2} \neq 0$ such that

$$\nabla_{j_1} \mathcal{D}_k(\boldsymbol{S}) \times \Delta Q_{j_1} \quad \text{and} \quad \nabla_{j_2} \mathcal{D}_k(\boldsymbol{S}) \times \Delta Q_{j_2}$$

have opposite signs (which do not vary with $\boldsymbol{S}$ by assumption).

Take $|V_{j_1} - Q_{j_1}|$ small and $|V_{j_2} - Q_{j_2}|$ not small, so that ΔS_{j_1} has the sign of ΔQ_{j_1} and $\Delta S_{j_2} = 0$; then ΔD_k has the sign of $\nabla_{j_1} \mathcal{D}_k(\boldsymbol{S}) \times \Delta Q_{j_1}$. Permuting j_1 and j_2 generates the opposite sign; therefore, such a change in thresholds generates two-way flows.

• *Proof of part 2:* To prove part 2 of the theorem, take j such that $\nabla_j \mathcal{D}_k$ changes sign when the sign of $V_j - Q_j$ changes (so that S_j switches between 0 and 1). Let $\Delta Q_m = 0$ for all $m \neq j$, so that

$$\Delta D_k \simeq \nabla_j \mathcal{D}_k(\boldsymbol{S}) \times \Delta S_j.$$

By the assumption in part 2, the sign of ΔD_k is the sign of ΔS_j for some values of $\boldsymbol{V}$ and the opposite sign for other values. Take any change in the threshold ΔQ_j. Since ΔS_j is zero or has the sign of ΔQ_j, ΔD_k must take opposite values as $\boldsymbol{V}$ varies. *Q.E.D.*

To illustrate the theorem, first consider the double hurdle model, for which $\nabla \mathcal{D}_1(\boldsymbol{S}) = (S_2, S_1) \geq 0$. This case is covered by part 1 of Theorem A.2. Changes such that ΔQ_1 and ΔQ_2 have the same sign do not generate two-way flows, but changes that generate $\Delta Q_1 \Delta Q_2 < 0$ do.

Now turn to the model of Example 1, where $\nabla \mathcal{D}_2(\boldsymbol{S}) = (1 - 2S_2, 1 - 2S_1)$. This corresponds to part 2 of the theorem, since the sign of $(1 - 2s)$ depends on $s = 0, 1$. Using the expansion (A.3) gives, with $j_1 = 1$, $j_2 = 2$,

$$\Delta D_2 \simeq (1 - 2S_2) \times \Delta S_1 + (1 - 2S_1) \times \Delta S_2.$$

Depending on the values of $\boldsymbol{V}$ and therefore of S_1 and S_2, this can be

$$\Delta S_1 + \Delta S_2, \qquad \Delta S_1 - \Delta S_2, \qquad \Delta S_2 - \Delta S_1, \quad \text{or} \quad -\Delta S_1 - \Delta S_2.$$

To get one-way flows only, we would need to change thresholds to induce $\Delta S_1, \Delta S_2 = \pm 1$ such that the four numbers above have the same sign. But that is clearly impossible. Hence *any* change in instruments creates two-way flows.

APPENDIX B: ADDITIONAL PROOFS

B.1. *Proof of Corollary 3.2*

First, consider the average treatment effect. Under Assumption 3.5, we have that

$$EG(Y_k) = \int E\big(G(Y_k)|\boldsymbol{V} = \boldsymbol{v}\big) f_{\boldsymbol{V}}(\boldsymbol{v})\, d\boldsymbol{v},$$

which implies (3.2) immediately.

Now consider $E[G(Y_k) - G(Y_\ell)|D = m]$. Note that

$$\begin{aligned}
&E\big[G(Y_k) - G(Y_\ell)|D = m, \boldsymbol{Q}(\boldsymbol{Z}) = \boldsymbol{q}\big] \\
&\quad = E\big[G(Y_k) - G(Y_\ell)|d_m(\boldsymbol{V}, \boldsymbol{q}) = 1\big] \\
&\quad = \frac{\displaystyle\int \mathbb{1}\big(d_m(\boldsymbol{v}, \boldsymbol{q}) = 1\big) E\big[G(Y_k) - G(Y_\ell)|\boldsymbol{V} = \boldsymbol{v}\big] f_{\boldsymbol{V}}(\boldsymbol{v})\, d\boldsymbol{v}}{\displaystyle\int \mathbb{1}\big(d_m(\boldsymbol{v}, \boldsymbol{q}) = 1\big) f_{\boldsymbol{V}}(\boldsymbol{v})\, d\boldsymbol{v}}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&E\big[G(Y_k) - G(Y_\ell)|D = m\big] \\
&\quad = EE\big[G(Y_k) - G(Y_\ell)|D = m, \boldsymbol{Q}(\boldsymbol{Z})\big]\big] \\
&\quad = \int \frac{\displaystyle\int \mathbb{1}\big(d_m(\boldsymbol{v}, \boldsymbol{q}) = 1\big) E\big[G(Y_k) - G(Y_\ell)|\boldsymbol{V} = \boldsymbol{v}\big] f_{\boldsymbol{V}}(\boldsymbol{v})\, d\boldsymbol{v}}{\displaystyle\int \mathbb{1}\big(d_m(\boldsymbol{v}, \boldsymbol{q}) = 1\big) f_{\boldsymbol{V}}(\boldsymbol{v})\, d\boldsymbol{v}}\, dF_{\boldsymbol{Q}(\boldsymbol{Z})|D}(\boldsymbol{q}|m).
\end{aligned}$$

By Bayes's rule, we have that

$$dF_{\boldsymbol{Q}(\boldsymbol{Z})|D}(\boldsymbol{q}|m) = \frac{\Pr[D = m|\boldsymbol{Q}(\boldsymbol{Z}) = \boldsymbol{q}]}{\Pr(D = m)}\, dF_{\boldsymbol{Q}(\boldsymbol{Z})}(\boldsymbol{q}).$$

Since

$$\Pr[D = m|\boldsymbol{Q}(\boldsymbol{Z}) = \boldsymbol{q}] = \int \mathbb{1}\big(d_m(\boldsymbol{v}, \boldsymbol{q}) = 1\big) f_{\boldsymbol{V}}(\boldsymbol{v})\, d\boldsymbol{v},$$

we have that

$$
\begin{aligned}
&E\big[G(Y_k)-G(Y_\ell)|D=m\big] \\
&\quad=\int\frac{\int \mathbb{1}\big(d_m(\boldsymbol{v},\boldsymbol{q})=1\big)E\big[G(Y_k)-G(Y_\ell)|\boldsymbol{V}=\boldsymbol{v}\big]f_V(\boldsymbol{v})\,d\boldsymbol{v}}{\Pr(D=m)}\,dF_{Q(\mathbf{Z})}(\boldsymbol{q}) \\
&\quad=\frac{\int \Pr\big(d_m(\boldsymbol{v},\boldsymbol{Q}(\boldsymbol{Z}))=1\big)E\big[G(Y_k)-G(Y_\ell)|\boldsymbol{V}=\boldsymbol{v}\big]f_V(\boldsymbol{v})\,d\boldsymbol{v}}{\Pr(D=m)} \\
&\quad=\int \Delta^{(k,\ell)}_{\mathrm{MTE}}(\boldsymbol{v})\omega^m_{\mathrm{ATT}}(\boldsymbol{v})\,d\boldsymbol{v}.
\end{aligned}
$$

We now move to the identification of the policy-relevant treatment effects. Recall that in the proof of Theorem 3.1 (see equation (6.1)), we have that

$$
\begin{aligned}
&E\big[G(Y)D_k|\boldsymbol{Q}(\boldsymbol{Z})=\boldsymbol{q}\big] \\
&\quad=\int \mathbb{1}\big(d_k(\boldsymbol{v},\boldsymbol{q})=1\big)E\big[G(Y_k)|\boldsymbol{V}=\boldsymbol{v}\big]f_V(\boldsymbol{v})\,d\boldsymbol{v}.
\end{aligned}
$$

Since $G(Y)=\sum_{k\in\mathcal{K}}G(Y)D_k$, we then have that

$$
\begin{aligned}
E\big[G(Y)\big]&=\sum_{k\in\mathcal{K}}E\big[E\big[G(Y)D_k|\boldsymbol{Q}(\boldsymbol{Z})=\boldsymbol{q}\big]\big] \\
&=\sum_{k\in\mathcal{K}}\int \Pr\big[d_k(\boldsymbol{v},\boldsymbol{Q}(\boldsymbol{Z}))=1\big]E\big[G(Y_k)|\boldsymbol{V}=\boldsymbol{v}\big]f_V(\boldsymbol{v})\,d\boldsymbol{v}.
\end{aligned}
$$

Similarly, we have that

$$
\begin{aligned}
E[D]&=\sum_{k\in\mathcal{K}}kE\big[E[D_k|\boldsymbol{Q}(\boldsymbol{Z})=\boldsymbol{q}]\big] \\
&=\sum_{k\in\mathcal{K}}k\int \Pr\big[d_k(\boldsymbol{v},\boldsymbol{Q}(\boldsymbol{Z}))=1\big]f_V(\boldsymbol{v})\,d\boldsymbol{v}
\end{aligned}
$$

and that

$$
\begin{aligned}
E[D_k=1]&=E\big[E[D_k|\boldsymbol{Q}(\boldsymbol{Z})=\boldsymbol{q}]\big] \\
&=\int \Pr\big[d_k(\boldsymbol{v},\boldsymbol{Q}(\boldsymbol{Z}))=1\big]f_V(\boldsymbol{v})\,d\boldsymbol{v}.
\end{aligned}
$$

The desired results follow immediately since the new policy only changes $\boldsymbol{Q}$ to $\boldsymbol{Q}^*$, while everything else remains the same.

B.2. *Proof of Theorem 4.1*

It follows from (2.1) in the main paper that

$$
Q_1(\boldsymbol{Z})+Q_2(\boldsymbol{Z})=2P_0(\boldsymbol{Z})+P_2(\boldsymbol{Z}). \tag{B.1}
$$

The right-hand side of (B.1) is identified directly from the data. Suppose that $\tilde{Q}_1(\mathbf{Z})$ and $\tilde{Q}_2(\mathbf{Z})$ also satisfy $\tilde{Q}_1(\mathbf{Z}) + \tilde{Q}_2(\mathbf{Z}) = 2P_0(\mathbf{Z}) + P_2(\mathbf{Z})$, as well as Assumption 4.1. Then, writing $\Delta_j(\mathbf{Z}) = Q_j(\mathbf{Z}) - \tilde{Q}_j(\mathbf{Z})$ $(j = 1, 2)$ gives $\Delta_1(\mathbf{Z}) = -\Delta_2(\mathbf{Z})$. But by Assumption 4.1, Δ_1 does not depend on Z_2, and Δ_2 does not depend on Z_1. Therefore, we must have $\tilde{Q}_1(Z_1) = Q_1(Z_1) + C$ and $\tilde{Q}_2(Z_2) = Q_2(Z_2) - C$, where C is a constant. This proves that Q_1 and Q_2 are identified up to an additive constant.

Further, take any $(z_1^0, z_2^0) \in \mathcal{Z}$. If we take $Q_2(z_2) = P(z_1^0, z_2) - C_1^0$ for some constant C_1^0, then by (B.1),

$$Q_1(z_1) = P(z_1, z_2) - P\big(z_1^0, z_2\big) + C_1^0. \tag{B.2}$$

Since the right-hand side of (B.2) should not depend on z_2, we set

$$\begin{aligned} Q_1(z_1) &= P\big(z_1, z_2^0\big) - P\big(z_1^0, z_2^0\big) + C_1^0, \\ Q_2(z_2) &= P\big(z_1^0, z_2\big) - C_1^0. \end{aligned}$$

To describe the possible range of C_1^0, note that we require that

$$\begin{aligned} &\Pr(D = 0) = \Pr\big[Q_1(Z_1) > 0 \text{ and } Q_2(Z_2) > 0\big] > 0, \\ &\Pr(D = 1) = \Pr\big[Q_1(Z_1) < 1 \text{ and } Q_2(Z_2) < 1\big] > 0, \\ &\Pr(D = 2) = \Pr\big[Q_1(Z_1) > 0 \text{ and } Q_2(Z_2) < 1\big] + \Pr\big[Q_1(Z_1) < 1 \text{ and } Q_2(Z_2) > 0\big] > 0. \end{aligned}$$

That is, C_1^0 must satisfy the following restrictions:

$$\begin{aligned} \Pr\big[P\big(z_1^0, z_2^0\big) - P\big(Z_1, z_2^0\big) < C_1^0 < P\big(z_1^0, Z_2\big)\big] &> 0, \\ \Pr\big[P\big(z_1^0, Z_2\big) - 1 < C_1^0 < 1 + P\big(z_1^0, z_2^0\big) - P\big(Z_1, z_2^0\big)\big] &> 0, \\ \Pr\big[\max\big\{P\big(z_1^0, z_2^0\big) - P\big(Z_1, z_2^0\big), P\big(z_1^0, Z_2\big) - 1\big\} < C_1^0\big] & \\ + \Pr\big[C_1^0 < \min\big\{1 + P\big(z_1^0, z_2^0\big) - P\big(Z_1, z_2^0\big), P\big(z_1^0, Z_2\big)\big\}\big] &> 0. \end{aligned}$$

B.3. *Proof of Theorem 4.2*

Recall that we denote $H(z_1, z_2) = \Pr(D = 1 | Z_1 = z_1, Z_2 = z_2)$ the propensity score. Under our exclusion restrictions, $H(z_1, z_2) = F_{V_1,V_2}(G_1(z_1), G_2(z_2))$.

Let $f_V(v_1, v_2)$ denote the density of $V = (V_1, V_2)$. By construction,

$$H(z_1, z_2) = F_V\big(G_1(z_1), G_2(z_2)\big) = \int_0^{G_1(z_1)} \int_0^{G_2(z_2)} f_V(v_1, v_2)\, dv_1\, dv_2. \tag{B.3}$$

Differentiating both sides of (B.3) with respect to z_1 gives

$$\frac{\partial H}{\partial z_1}(z_1, z_2) = G_1'(z_1) \int_0^{G_2(z_2)} f_V\big(G_1(z_1), v_2\big)\, dv_2. \tag{B.4}$$

Now letting $z_2 \to b_2$ on both sides of (B.4) yields

$$\lim_{z_2 \to b_2} \frac{\partial H}{\partial z_1}(z_1, z_2) = G_1'(z_1) \bigg[\lim_{z_2 \to b_2} \int_0^{G_2(z_2)} f_V\big(G_1(z_1), v_2\big)\, dv_2\bigg]. \tag{B.5}$$

The expression inside the brackets on the right-hand side of (B.5) is 1 since $\lim_{z_2\to b_2} G_2(z_2) = 1$ and the marginal distribution of V_2 is $U[0,1]$. Therefore, we identify G_1 by

$$G_1(z_1) = \int_{a_1}^{z_1} \lim_{t_2\to b_2} \frac{\partial H}{\partial z_1}(t_1, t_2)\, dt_1. \tag{B.6}$$

Analogously, we identify G_2 by

$$G_2(z_2) = \int_{a_2}^{z_2} \lim_{t_1\to b_1} \frac{\partial H}{\partial z_2}(t_1, t_2)\, dt_2. \tag{B.7}$$

Returning to (B.3), since G_1 and G_2 are strictly increasing, we identify F_V by

$$F_V(v_1, v_2) = H\big(G_1^{-1}(v_1), G_2^{-1}(v_2)\big).$$

B.4. *Proof of Theorem 4.3*

B.4.1. *Proof of Part 1*

Given our differentiability assumptions, we can take derivatives of the formula

$$\phi\big(H(z_1, z_2)\big) = \phi\big(G_1(z_1)\big) + \phi\big(G_2(z_2)\big) \tag{B.8}$$

over $\mathcal{N}$. Using

$$\frac{\partial^2(\phi\circ H)}{\partial z_1 \partial z_2}(z_1, z_2) = 0,$$

we obtain

$$\phi''(h)\frac{\partial H}{\partial z_1}(z_1, z_2)\frac{\partial H}{\partial z_2}(z_1, z_2) + \phi'(h)\frac{\partial^2 H}{\partial z_1 \partial z_2}(z_1, z_2) = 0$$

with $h = H(z_1, z_2)$.

Take any smooth curve contained in $\mathcal{N}$ and parameterize it as $h \to (z_1(h), z_2(h))$ with $h = H(z_1(h), z_2(h))$; then we have a differential equation

$$\phi''(h)\frac{\partial H}{\partial z_1}\big(z_1(h), z_2(h)\big)\frac{\partial H}{\partial z_2}\big(z_1(h), z_2(h)\big) + \phi'(h)\frac{\partial^2 H}{\partial z_1 \partial z_2}\big(z_1(h), z_2(h)\big) = 0. \tag{B.9}$$

Using (B.8), the partial derivatives H_1 and H_2 cannot take the value zero on $\mathcal{N}$ since G_1' and G_2' are never zero. Therefore, we can rewrite (B.9) as

$$\frac{\phi''}{\phi'}(h) = -\frac{H_{12}}{H_1 H_2}\big(z_1(h), z_2(h)\big)$$

over $\mathcal{N}$.

We note that this equation incorporates a sign constraint and over-identifying restrictions. For ϕ to be strictly decreasing and convex, we require $H_{12}/(H_1 H_2) \geq 0$. Moreover, on any admissible curve, the ratio $H_{12}/(H_1 H_2)$ must be the same function of h, which we denote $R(h)$.

B.4.2. *Proof of Part 2*

From now on, we denote $(\underline{h}, \bar{h}) \subset (0,1)$ the image of $\mathcal{N}$ by H.
We use the fact that $\partial \log(-\phi'(h))/\partial h = \phi''(h)/\phi'(h)$ to obtain

$$\log\bigl(-\phi'(h)\bigr) = \int_h^{\bar{h}} R(t)\,dt + \log\bigl(-\phi'(\bar{h})\bigr),$$

so that

$$\phi'(h) = \phi'(\bar{h}) \exp\biggl(\int_h^{\bar{h}} R(t)\,dt\biggr).$$

Denoting

$$\mathbb{T}(h) := \int_h^{\bar{h}} dk \exp\biggl(\int_k^{\bar{h}} R(t)\,dt\biggr)$$

gives us $\phi(h) = \phi(\bar{h}) - \phi'(\bar{h})\mathbb{T}(h)$. Note that, by construction, $\mathbb{T}$ is a decreasing function and $\mathbb{T}(\bar{h}) = 0$. Moreover, $\phi'(\bar{h})$ cannot be zero since ϕ would be constant.

B.4.3. *Proof of Part 3*

If ϕ solves (B.8), then clearly so does $\alpha\phi$ for any $\alpha > 0$; we normalize $\phi'(\bar{h}) = -1$. Hence, from now on, $\phi(h) = \phi(\bar{h}) - \mathbb{T}(h)$. The constant $\phi(\bar{h})$ must be nonnegative since ϕ cannot take negative values. Moreover, since ϕ is convex, $\phi'(\bar{h}) = -1$, and $\phi(1) = 0$, we must have $\phi(\bar{h}) \leq 1 - \bar{h}$. If, moreover, $\bar{h} = \sup_{z\in\mathcal{N}} \Pr(D = 1|\mathbf{Z} = z) = 1$, then $\phi(\bar{h}) = \phi(1) = 0$; this defines directly $\phi(h) = -\mathbb{T}(h)$ over $(\underline{h}, 1)$.

B.4.4. *Proof of Part 4*

Since the model is well-specified, there is a solution G_1, G_2 (the thresholds of the true DGP). In addition, since any other admissible $(\tilde{G}_1, \tilde{G}_2)$ must satisfy

$$\phi\bigl(\tilde{G}_1(z_1)\bigr) + \phi\bigl(\tilde{G}_2(z_2)\bigr) = \phi\bigl(H(z_1, z_2)\bigr) = \phi\bigl(G_1(z_1)\bigr) + \phi\bigl(G_2(z_2)\bigr)$$

on $\mathcal{N}$, it must be that

$$\phi\bigl(\tilde{G}_1(z_1)\bigr) = \phi\bigl(G_1(z_1)\bigr) - k,$$
$$\phi\bigl(\tilde{G}_2(z_2)\bigr) = \phi\bigl(G_2(z_2)\bigr) + k,$$

for some constant k. Any such constant must be such that $\phi(G_1(z_1)) - k$ and $\phi(G_2(z_2)) + k$ are both nonnegative for all z_1 and z_2 in the projections of $\mathcal{N}$. That is,

$$-\inf \phi\bigl(G_2(z_2)\bigr) \leq k \leq \inf \phi\bigl(G_1(z_1)\bigr).$$

If, moreover, $\sup_{z\in\mathcal{N}} \Pr(D = 1|\mathbf{Z} = z) = 1$, then $\bar{h} = 1$. Take a sequence (z_n) such that $H(z_n)$ converges to $\bar{h} = 1$. Then $\phi(H(z_n))$ converges to zero, so that both $\phi(G_1(z_{1n}))$ and $\phi(G_2(z_{2n}))$ must converge to zero. The double inequality above implies that $k = 0$, and G_1 and G_2 are point-identified on the projections of $\mathcal{N}$.

APPENDIX C: THE ENTRY GAME

Let us return to Example 2, in which two firms $j = 1, 2$ are considering entry into a new market. Firm j has profit π_j^m if it becomes a monopoly, and $\pi_j^d < \pi_j^m$ if both firms enter. We saw that if $\pi_j^m > 0 > \pi_j^d$ for both firms, then there are two symmetric equilibria, with only one firm operating. Now assume that we observe not only the number of entrants as in Example 2, but also their identity. With profits given by $\pi_j^m = V_j - Q_j(\boldsymbol{Z})$ and $\pi_j^d = \bar{V}_j - \bar{Q}_j(\boldsymbol{Z})$, if only firm 1 entered then we know that $\pi_1^m > 0$ and $\pi_2^d < 0$, so that

$$V_1 > Q_1(\boldsymbol{Z}) \quad \text{and} \quad \bar{V}_2 < \bar{Q}_2(\boldsymbol{Z}).$$

That still leaves two possible cases:

1. $\pi_2^m < 0$, and the unique equilibrium has only firm 1 entering the market;
2. $\pi_2^m > 0$, and there is another, symmetric equilibrium with only firm 2 entering.

Now let us postulate an equilibrium selection rule that has a threshold structure: when both π_1^m and π_m^2 are positive, firm 1 is selected to be the unique entrant if and only if $U < q(\boldsymbol{Z})$. Then the necessary and sufficient set of conditions for the entry of firm 1 only is

$$V_1 > Q_1(\boldsymbol{Z}) \quad \text{and} \quad \big(V_2 < Q_2(\boldsymbol{Z}) \text{ or } \big(\bar{V}_2 < \bar{Q}_2(\boldsymbol{Z}) \text{ and } U < q(\boldsymbol{Z})\big)\big).$$

This is again a special case of the general framework we analyze in this paper.

APPENDIX D: DETAILED DISCUSSION OF HECKMAN, URZUA, AND VYTLACIL (2008)

Heckman, Urzua, and Vytlacil (2008) considered a multinomial discrete choice model for treatment. They posited

$$D = k \quad \iff \quad R_k(\boldsymbol{Z}) - U_k > R_l(\boldsymbol{Z}) - U_l \quad \text{for } l = 0, \dots, K-1 \text{ such that } l \neq k,$$

where the U's are continuously distributed and independent of $\boldsymbol{Z}$.

Define

$$\boldsymbol{R}(\boldsymbol{Z}) = (R_k(\boldsymbol{Z}) - R_l(\boldsymbol{Z}))_{l \neq k} \quad \text{and} \quad \boldsymbol{U} = (U_k - U_l)_{l \neq k}.$$

Then $D_k = \mathbb{1}(\boldsymbol{R}(\boldsymbol{Z}) > \boldsymbol{U})$; and defining $Q_l(\boldsymbol{Z}) = \Pr[\boldsymbol{U}_l < \boldsymbol{R}_l(\boldsymbol{Z}) | \boldsymbol{Z}]$ allows us to write the treatment model as

$$D = k \quad \text{iff} \quad \boldsymbol{V} < \boldsymbol{Q}(\boldsymbol{Z}), \tag{D.1}$$

where each V_l is distributed as $U[0, 1]$.

The applications they considered are GED certification (with three treatments: permanent high school dropout, GED, high school degree) and randomized trials with imperfect compliance (e.g., no training, classroom training, and job search assistance).

They then studied the identification of marginal and local average treatment effects under assumptions that are similar to ours: continuous instruments that generate enough dimensions of variation in the thresholds. They assumed that $\boldsymbol{V}$ is continuously distributed with full support; that $(\boldsymbol{U}, \boldsymbol{V}) \perp\!\!\!\perp Z$; and that all treatments have positive probabilities. More importantly, they made either

- assumption (a): for each treatment j, there is a component of $\boldsymbol{Z}$ that drives some variation in R_j conditional on the other components, and in R_j only;

• assumption (b): for each treatment j, there is a component of $\boldsymbol{Z}$ that drives continuous variation in R_j conditional on the other components, and no variation in the other components of R.

For any subset of treatments $\mathcal{J} \subset \mathcal{K}$, they defined $Y_{\mathcal{J}}$ to be the outcome when the agent chooses the best treatment from $\mathcal{J}$. They also defined $\Delta_{\mathcal{J},\mathcal{L}} = Y_{\mathcal{J}} - Y_{\mathcal{L}}$, and in particular, the MTE

$$E\big(\Delta_{\mathcal{J},\mathcal{L}}|\boldsymbol{Z}, R_{\mathcal{J}}(\boldsymbol{Z}) = R_{\mathcal{L}}(\boldsymbol{Z})\big).$$

They showed that

• if we take $\mathcal{J} = \{j\}$ and $\mathcal{L} = \mathcal{K} - \{j\}$, then the LATE is identified under (a) and the MTE is identified under (b);

• if we take any $\mathcal{J}$ and $\mathcal{L} = \mathcal{K} - \mathcal{J}$, then the results are similar but the MTEs and LATEs are defined by conditioning on the values of the Q's rather than on the Z's.

They did not invoke any large support assumptions to obtain identification results mentioned just above.

However, if we take $\mathcal{J} = \{j\}$ and $\mathcal{L} = \{l\}$, then their corresponding identification results (see Theorem 3 of Heckman, Urzua, and Vytlacil (2008)) require a large support condition. To see their logic, suppose that $K = 3$ and that one of the R_j's is sufficiently negative that the probability of choosing one of the choices is arbitrarily small. This case effectively reduces to the binary treatment case; their LIV estimand, which is the limit of a sequence of Wald estimands, identifies the MTE.

We do not rely on this type of identification-at-infinity strategy since we identify the MTE via multidimensional cross-derivatives. Note that our identification results are conditional on the assumption that $\boldsymbol{Q}$ is already identified. A more stringent assumption on the support of $\boldsymbol{Z}$ might be necessary to identify $\boldsymbol{Q}$, as demonstrated in Matzkin (1993, 2007). In this sense, our assumptions are not necessarily weaker than those of Heckman, Urzua, and Vytlacil (2008). We view our identification results and theirs as complementing each other.

APPENDIX E: NON-RECTANGULAR THRESHOLD CONDITIONS

The threshold conditions we postulated in Assumption 2.1 have the "rectangular" form $V_j < Q_j(\boldsymbol{Z})$. Suppose that the threshold conditions $j = 1, \dots, J$ have the more general form

$$\boldsymbol{\alpha}_j \cdot \boldsymbol{U} \leq R_j(\boldsymbol{Z}),$$

where the $\boldsymbol{\alpha}_j$ are possibly unknown parameter vectors in $\mathbb{R}^L$ and $\boldsymbol{U} = (U_1, \dots, U_L)$ is independent of $\boldsymbol{Z}$. For notational simplicity, assume that each (scalar) random variable $u_j \equiv \boldsymbol{\alpha}_j \cdot \boldsymbol{U}$ has positive density everywhere; denote H_j its cdf. Then, each threshold condition can be written equivalently as

$$V_j \equiv H_j(u_j) < H_j\big(R_j(\boldsymbol{Z})\big) \equiv Q_j(\boldsymbol{Z}).$$

By construction, each V_j is distributed uniformly over $[0, 1]$. Moreover, since each threshold Q_j is an increasing function of the corresponding R_j only, any exclusion restriction assumed on either form applies equally to the other, so that we can hope to identify the thresholds Q_j under suitable assumptions. If they are indeed identified, then we can apply Theorem 3.1 to recover the joint density of $\boldsymbol{V} = (V_1, \dots, V_J)$ and the MTE conditional on $\boldsymbol{v}$.

The random variables $\boldsymbol{V}$ and the thresholds $\boldsymbol{Q}$ are only auxiliary objects, and the analyst is likely to be more interested in the $\boldsymbol{U}$ and $\boldsymbol{R}$. If the cdf H_j were known, then we could write $R_j = H_j^{-1}(Q_j)$ and by the change-of-variables formula,

$$f_{\boldsymbol{u}}(u_1, \dots, u_J) = f_V\big(H_1^{-1}(u_1), \dots, H_J^{-1}(u_J)\big) \times \prod_{j=1}^{J} H_j'(u_j).$$

In turn, knowing the joint distribution of $\boldsymbol{u}$ directly gives the density of $\boldsymbol{U}$ if $L = J$ and the matrix $\boldsymbol{\alpha}$ whose rows are the vectors $\boldsymbol{\alpha}_j'$ is invertible:

$$f_U(\boldsymbol{U}) = f_{\boldsymbol{u}}(\boldsymbol{\alpha U}) \times |\boldsymbol{\alpha}|.$$

If, more realistically, the H_j and $\boldsymbol{\alpha}_j$ are unknown, we may still use other restrictions. As an illustration, take a recursive system, where the matrix $\boldsymbol{\alpha}$ is lower-triangular with diagonal terms equal to 1. Then, since $U_2 = u_2 - \alpha_{21} u_1 = H_2^{-1}(V_2) - \alpha_{21} H_1^{-1}(V_1)$, the independence of U_1 and U_2, for instance, would translate into the independence of V_1 and of the variable

$$W_2 \equiv H_2^{-1}(V_2) - \alpha_{21} H_1^{-1}(V_1).$$

Now $V_2 = H_2(W_2 + \alpha_{21} U_1)$, so this in turn implies that the (identified) distribution of V_2 conditional of V_1 must satisfy

$$F_{V_2|V_1}\big(H_2\big(w_2 + \alpha_{21} H_1^{-1}(v_1)\big)|v_1\big) = F_{W_2}(w_2) = H_2(w_2)$$

for all w_2 and v_1. But as the right-hand side does not depend on v_1, this imposes restrictions that only hold for some choices of H_1, H_2, and α_{21}. If we only know H_2, then

$$w_2 + \alpha_{21} H_1^{-1}(v_1) = F_{V_2|V_1}^{-1}\big(H_2(w_2)|v_1\big)$$

over-identifies the product $\alpha_{21} H_1^{-1}(v_1)$; and if we also know H_1, then it over-identifies α_{21}. These results extend directly to higher dimensional systems.

REFERENCES

HECKMAN, J., AND R. PINTO (2018): "Unordered Monotonicity," *Econometrica*, 86 (1), 1–35.[3]

HECKMAN, J. J., S. URZUA, AND E. VYTLACIL (2008): "Instrumental Variables in Models With Multiple Outcomes: The General Unordered Case," *Annales d'économie et de statistique*, 91/92, 151–174.[1,11, 12]

MATZKIN, R. L. (1993): "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58 (1), 137–168.[12]

——— (2007): "Heterogeneous Choice," in *Advances in Economics and Econometrics: Theory and Applications*, Vol. 2, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge University Press, 75–110, Chapter 4. [12]

Co-editor Elie Tamer handled this manuscript.

Manuscript received 25 March, 2016; final version accepted 29 May, 2018; available online 22 June, 2018.