Supplement to "Partial identification by extending subdistributions"

(*Quantitative Economics*, Vol. 10, No. 1, January 2019, 105–144)

Alexander Torgovitsky
Department of Economics, University of Chicago

This supplement contains a discussion of applying PIES to a bivariate ordered response model and a two-sector Roy model.

S.1. Extension to a bivariate ordered response model

In this section, I briefly extend the discussion in Section 4 to demonstrate the key derivations for applying Theorem 1 to a bivariate ordered response model. The model is given by

$$Y_1 = \sum_{j=1}^{J} y_{1j} \mathbb{1}\big[g_{1(j-1)}(Y_2, X) < U_1 \leq g_{1j}(Y_2, X)\big] \tag{S-1}$$

$$\text{and} \quad Y_2 = \sum_{k=1}^{K} y_{2k} \mathbb{1}\big[g_{2(k-1)}(X) < U_2 \leq g_{2k}(X)\big], \tag{S-2}$$

where $Y = (Y_1, Y_2)$ are random variables with supports $\mathcal{Y}_1 \equiv \{y_{11}, \dots, y_{1J}\}$ and $\mathcal{Y}_2 \equiv \{y_{21}, \dots, y_{2K}\}$ ordered to be increasing,

$$g \equiv (g_1, g_2) \equiv (g_{10}, g_{11}, \dots, g_{1J}, g_{20}, g_{21}, \dots, g_{2K})$$

is an unknown vector of functions, X is a random vector with support $\mathcal{X}$, and $U \equiv (U_1, U_2)$ is a bivariate latent variable. As before, let $\mathcal{F}$ denote the set of all proper bivariate conditional distribution functions $F : \overline{\mathbb{R}}^2 \times \mathcal{X} \to [0, 1]$, and let $\mathcal{F}^\dagger$ denote the admissible subset of $\mathcal{F}$. The parameter θ is the function g, with admissible set $\mathcal{G}^\dagger$, and any $g \in \mathcal{G}^\dagger$ satisfying $g_{10} = g_{20} = -\infty$ and $g_{1J} = g_{2K} = +\infty$. The model reduces to the bivariate binary response model in Section 4 of the main text by letting $J = K = 2$, $y_{11} = y_{21} = 0$, and $y_{12} = y_{22} = 1$.

The observational equivalence function $\omega_{y|x}$ defined in (11) of the main text is given by

$$\begin{aligned}
&\omega_{(y_{1j}, y_{2k})|x}(g, F) \\
&\quad \equiv \mathbb{P}_S[Y_1 \leq y_{1j}, Y_2 \leq y_{2k} | X = x]
\end{aligned}$$

Alexander Torgovitsky: torgovitsky@uchicago.edu
This research was supported in part by National Science Foundation Grant SES-1530538.

Available at http://qeconomics.org. https://doi.org/10.3982/QE634

$$= \sum_{k'=1}^{k} \mathbb{P}_F\big[U_1 \le g_{1j}(y_{2k'}, x), Y_2 = y_{2k'} | X = x\big]$$

$$= \sum_{k'=1}^{k} \mathbb{P}_F\big[U_1 \le g_{1j}(y_{2k'}, x), U_2 \in \big(g_{2(k'-1)}(x), g_{2k'}(x)\big] | X = x\big]$$

$$= \sum_{k'=1}^{k} F\big(g_{1j}(y_{2k'}, x), g_{2k'}(x)|x\big) - F\big(g_{1j}(y_{2k'}, x), g_{2(k'-1)}(x)|x\big) \tag{S-3}$$

for all $j = 1, \dots, J$, $k = 1, \dots, K$, and $x \in \mathcal{X}$. From (S-3), one can see that to satisfy (U2.ω) in Theorem 1, $\mathcal{U}_x(g)$ must be chosen so that it contains the set

$$\big\{\{g_{1j}(y_{2k}, x)\}_{k=1}^{K}\big\}_{j=0}^{J} \times \{g_{2k}(x)\}_{k=0}^{K} \tag{S-4}$$

for any fixed $g \in \mathcal{G}^\dagger$. Single equation ordered response models are nested by taking $K = 1$, since in this case the equation for Y_2 becomes trivially satisfied given that every $g \in \mathcal{G}^\dagger$ satisfies $g_{20}(x) = -\infty$ and $g_{2K}(x) = +\infty$ for all x. In this case, (S-3) reduces to

$$\omega_{(y_{1j}, y_{2k})|x}(g, F) = F\big(g_{1j}(y_{2k'}, x), +\infty|x\big),$$

which just depends on the marginal distribution of U_1, as one would expect.

S.2. A two-sector Roy model

Consider the binary treatment potential outcomes model

$$W = \mathbb{1}[T = 1]W_1 + \mathbb{1}[T = 2]W_2, \tag{S-5}$$

where $T \in \{0, 1\}$ is a binary treatment and (W_1, W_2) are latent potential outcomes corresponding to different states of this treatment. The researcher observes (W, T, X) where X is a vector of covariates with respect to which certain exclusion and/or independence conditions might be maintained. To make the relationship between X and (W_1, W_2) explicit, consider the latent variable formulation

$$W_t = g_t(X, U_t) \quad \text{for } t = 1, 2, \tag{S-6}$$

where U_t, $t = 1, 2$ are latent random variables and g_t, $t = 1, 2$ are unknown functions. The functions g_t can be parameterized, or a completely agnostic approach can be taken by setting $g_t(X, U_t) = U_t$, in which case U_t is simply a relabeling of the potential outcome W_t. In addition to (S-6), analysis of this problem frequently maintains a weakly separable selection equation

$$T = 1 + \mathbb{1}\big[U_3 \le g_3(X)\big], \tag{S-7}$$

where U_3 is a latent variable and g_3 is an unknown function (Vytlacil (2002), Heckman and Vytlacil (2005)). In the framework of Section 3.3, (S-5)–(S-7) comprise a two-equation model

$$
\begin{aligned}
W &= \mathbb{1}[T=1]g_1(X, U_1) + \mathbb{1}[T=2]g_2(X, U_2), \\
T &= 1 + \mathbb{1}\big[U_3 \leq g_3(X)\big],
\end{aligned}
$$

with a two-dimensional random vector $Y \equiv (W, T)$, the usual vector of covariates, X, and an $L = 3$-dimensional vector of unobservables (U_1, U_2, U_3). Let $\mathcal{F}$ denote the set of all proper trivariate conditional distribution functions $F : \overline{\mathbb{R}}^3 \to [0, 1]$, with $\mathcal{F}^\dagger$ the admissible subset of $\mathcal{F}$. The parameter θ in this context is the triple of functions $g = (g_1, g_2, g_3)$ with admissible set $\mathcal{G}$.

Suppose that $\mathcal{G}$ only contains triples g such that g_1 and g_2 are weakly increasing and left-continuous in their latent components. Denote the generalized inverse of any such g_1 and g_2 in these components by $g_1^{-1}(x, \cdot)$ and $g_2^{-1}(x, \cdot)$.[1] Then the mapping $\omega_{y|x}$ defined in (11) in the main text is given by

$$
\begin{aligned}
\omega_{(w,d)|x}(g, F) &\equiv \mathbb{P}_S[W \leq w, T \leq t | X = x] \\
&= \begin{cases}
F\big(g_1^{-1}(x, w), +\infty, g_3(x)|x\big), & \text{if } t = 1, \\
F\big(g_1^{-1}(x, w), +\infty, g_3(x)|x\big) & \\
\quad + F\big(+\infty, g_2^{-1}(x, w), +\infty|x\big) & \\
\quad - F\big(+\infty, g_2^{-1}(x, w), g_3(x)|x\big), & \text{if } t = 2.
\end{cases}
\end{aligned}
$$

Denote the support of W by $\mathcal{W}$, and let $\overline{\mathcal{W}}$ be a subset of $\mathcal{W}$ that is chosen by the researcher. In order to satisfy (U2.ω) in Theorem 1, $\mathcal{U}_x(g)$ must be chosen to contain the set

$$
\big\{g_1^{-1}(x, w), \pm\infty\big\}_{w \in \overline{\mathcal{W}}} \times \big\{g_2^{-1}(x, w), \pm\infty\big\}_{w \in \overline{\mathcal{W}}} \times \big\{g_3(x)\big\}.
$$

If $\overline{\mathcal{W}}$ is a strict subset of $\mathcal{W}$, then the characterization provided by Theorem 1 will not be sharp, but can be made arbitrarily sharp by making $\overline{\mathcal{W}}$ arbitrarily large.

A commonly maintained identifying assumption in the two-sector Roy model is that X is independent of the latent variables $U \equiv (U_1, U_2, U_3)$; see, for example, Assumption 1 in Eisenhauer, Heckman, and Vytlacil (2015). This can be imposed via condition A1 of Assumption A. Using Assumption A2, one could refine this strategy by requiring X to only be independent of certain components of U. In both cases, these assumptions could be imposed so that only part of the components of X are used in these independence statements, perhaps conditional on other components. Much additional flexibility is possible, including weakening independence to location restrictions of the sort discussed in Sections 2 and 4.

For the selection equation, (S-7), it is common to maintain a nonparametric view of the function g_3, in which case the marginal distribution of U_3 can be normalized,

[1] That is, $g_t^{-1}(x, w) \equiv \sup\{u : g_t(x, u) \leq w\}$ for $t = 1, 2$.

with a typical choice being the uniform distribution on the $[0, 1]$ interval. This type of normalization can be accommodated in PIES through Assumption A3; see Example 1. If W is continuously distributed, a common restriction to impose on g_1 and g_2 is additive separability in their respective latent variables. If W is a binary or more generally ordered discrete outcome, one might adopt a specification for g_1 and g_2 that is similar to those in Sections 4 and S.1.

Mourifié, Henry, and Méango (2015) derived analytic expressions for sharp identified sets of certain parameters in nonparametric two-sector Roy models. The PIES approach provides a general method to compute these sets that is also applicable to the types of semiparametric Roy models commonly used in empirical work. However, the PIES procedure is computational, and does not provide analytic expressions for identified sets.

References

Eisenhauer, P., J. J. Heckman, and E. Vytlacil (2015), "The generalized Roy model and the cost–benefit analysis of social programs." *Journal of Political Economy*, 123, 413–443. [3]

Heckman, J. J. and E. Vytlacil (2005), "Structural equations, treatment effects, and econometric policy evaluation." *Econometrica*, 73, 669–738. [3]

Mourifié, I., M. Henry, and R. Méango (2015), "Sharp bounds for the Roy model." SSRN Working Paper 2043117. [4]

Vytlacil, E. (2002), "Independence, monotonicity, and latent index models: An equivalence result." *Econometrica*, 70, 331–341. [3]

Co-editor Rosa L. Matzkin handled this manuscript.

Manuscript received 22 October, 2015; final version accepted 6 April, 2018; available online 1 May, 2018.