Econometrica Supplementary Material

SUPPLEMENT TO "IDENTIFICATION OF NONSEPARABLE MODELS USING INSTRUMENTS WITH SMALL SUPPORT"

(*Econometrica*, Vol. 83, No. 3, May 2015, 1185–1197)

BY ALEXANDER TORGOVITSKY

This supplement contains (i) sufficient conditions for point-identification when X is a vector and Z is continuously distributed; (ii) sufficient conditions for point-identification when X is a vector and Z is binary; and (iii) examples that illustrate the restrictiveness of the identifying assumptions.

S1. SUFFICIENT CONDITIONS FOR POINT-IDENTIFICATION WHEN Z IS CONTINUOUS

THIS SECTION CONTAINS A PROOF AND DISCUSSION OF THE FOLLOWING RESULT, which establishes point-identification when X is a vector and Z is continuously distributed. First, I introduce some notation required for the case where X is a vector. Define $\mathbf{F}_{X|Z}(x|z) \equiv (F_{X_1|Z}(x_1|z), \ldots, F_{X_{d_x}|Z}(x_{d_x}|z))$ as the vector of conditional (marginal) ranks of $X \in \mathbb{R}^{d_x}$. Then, just as in the scalar case, Assumptions FS and C imply that if

$$\mathbf{F}_{X|Z}\big(x^a|z^a\big) = \overline{v} = \mathbf{F}_{X|Z}\big(x^b|z^b\big) \quad \text{for} \quad x^a \in \mathcal{X}^\circ_{z^a} \text{ and } x^b \in \mathcal{X}^\circ_{z^b},$$

then the distributions of $\varepsilon|X = x^a, Z = z^a$ and $\varepsilon|X = x^b, Z = z^b$ are the same and equal to that of $\varepsilon|V = \overline{v}$. Hence, using the same arguments as in the main text, one obtains

$$I^g(x, e) = D^g\big(\mathbf{F}_{X|Z}(x|z), e\big) \quad \text{for} \quad x \in \mathcal{X}^\circ_z, e \in \mathcal{E}^\circ_{x,z}, \tag{S1}$$

and any $g \in \mathcal{G}^*$, where $D^g(v, e)$ is defined as before but with $v \in \mathbb{R}^{d_x}$. The analysis now proceeds from (S1). The aim, as before, is to show that $I^g(x, e)$ is not a function of x.

THEOREM S1: *Suppose that $\mathcal{X}$ is convex, Z is a continuously distributed d_z vector, and every element of $\mathcal{G}$ is everywhere continuously differentiable. Let $H(x, z)$ be the $d_z \times d_x$ matrix with (j, k) element $\nabla_{z_j} F_{X_k|Z}(x_k|z)$. Then g^* is point-identified on* $\operatorname{supp}(X, \varepsilon)$ *if for every x in a dense subset $\mathcal{X}_d$ of $\mathcal{X}$ and every $y \in \mathcal{Y}^\circ_x$, there exists a $\overline{z}$ with $x \in \mathcal{X}^\circ_{\overline{z}}$ and $y \in \mathcal{Y}^\circ_{x,\overline{z}}$ for which $\nabla_x F_{X|Z}(x|\overline{z})$ exists and $H(x, \overline{z})$ exists and has rank d_x.*

PROOF: Fix an $x \in \mathcal{X}_d$ and $e \in \mathcal{E}^\circ_x$ so that $y \equiv g^*(x, e) \in \mathcal{Y}^\circ_x$ and take $\overline{z}$ in the statement of the theorem. Then $e \in \mathcal{E}^\circ_{x,\overline{z}}$ because $y \in \mathcal{Y}^\circ_{x,\overline{z}}$, and so (S1) holds in a neighborhood of $(x, \overline{z}, e)$. Differentiating (S1) with respect to z_j gives

$$H_j(x, \overline{z}) \nabla_v D^g\big(\mathbf{F}_{X|Z}(x|\overline{z}), e\big)' = \nabla_{z_j} I^g(x, e) = 0, \tag{S2}$$

 DOI: 10.3982/ECTA9984

where $H_j(x, \overline{z})$ is the jth row of $H(x, \overline{z})$ and $\nabla_v D^g(v, e)'$ is a d_x column vector.[1] Stacking (S2) across j gives

$$H(x, \overline{z})\nabla_v D^g\big(\mathbf{F}_{X|Z}(x|\overline{z}), e\big)' = 0_{d_z}.$$

Because $H(x, \overline{z})$ has full rank, this implies $\nabla_v D^g(\mathbf{F}_{X|Z}(x|\overline{z}), e)' = 0_{d_x}$. Now differentiate (S1) with respect to x_k at $(x, \overline{z}, e)$. This yields

$$f_{X_k|Z}(x_k|\overline{z})\nabla_{v_k} D^g\big(\mathbf{F}_{X|Z}(x|\overline{z}), e\big) = 0 = \nabla_{x_k} I^g(x, e)$$

for each k. Hence $\nabla_x I^g(x, e) = 0_{d_x}$ for all $x \in \mathcal{X}$, because $\nabla_x I^g(\cdot, e)$ is continuous and $\mathcal{X}_d$ is a dense subset of $\mathcal{X}$.[2] Since $\mathcal{X}$ is convex, this implies that $I^g(x, e)$ is constant in x, that is, $I^g(x, e) = J^g(e)$ for some function J^g, every $x \in \mathcal{X}$, and every $e \in \mathcal{E}_x^\circ$ (hence $\mathcal{E}_x$, by continuity); see, for example, Theorem 9.19 of Rudin (1976).[3] As discussed in the main text, this implies that g^* is point-identified on $\text{supp}(X, \varepsilon)$. *Q.E.D.*

The conditions on the dependence between X and Z in Theorem S1 are weak. When X is scalar, they are similar to the local relevance condition used by Chesher (2003) to point-identify $\nabla_x g^*(Q_{X|Z}(v|z), Q_{\varepsilon|\eta}(t|Q_\eta(v)))$ under local restrictions.[4] In contrast to Chesher's result, here relevance is assumed to hold globally (for a dense subset of $\mathcal{X}$) to point-identify $g^*(x, e)$ at any prespecified x and e. For vector X, the conditions are a nonlinear generalization of the classical relevance (or rank) condition in the linear model.[5] The usual order condition ($d_z \geq d_x$) is necessary for H to have full rank.

S2. SUFFICIENT CONDITIONS FOR POINT-IDENTIFICATION WHEN X IS A VECTOR AND Z IS BINARY

This section contains a proof and discussion of the following result, which establishes point-identification when X is a vector and Z is binary.

[1] Assumptions C and G imply that $D^g(\cdot, e)$ is differentiable at $\mathbf{F}_{X|Z}(x|\overline{z})$ when g and g^* are everywhere continuously differentiable.

[2] Assumption G implies that $I^g(\cdot, e)$ is everywhere continuously differentiable if g and g^* are.

[3] The assumption that $\mathcal{X}$ is convex could be relaxed to the assumption that it is a closed region Olmsted (1961, p. 280).

[4] Chesher also allows η to enter g^* directly, so actually he point-identifies

$$\nabla_x g^*\big(Q_{X|Z}(v|z), Q_{\varepsilon|\eta}\big(t|Q_\eta(v)\big), Q_\eta(v)\big).$$

A concise treatment of Chesher's result can be found in Section 3.1 of Chesher (2007).

[5] To see this, suppose that $X = \Gamma Z + \eta$ for a $d_x \times d_z$ matrix, Γ. By FS.E, $F_{X_k|Z}(x_k|z) = F_{\eta_k}(x_k - \Gamma_k z)$, where Γ_k is the kth row of Γ. Differentiating yields $H(x, z) = \Gamma' B(x, z)$ for a $d_x \times d_x$ diagonal matrix $B(x, z)$ with entries $-f_{\eta_k}(x_k - \Gamma_k z)$. The diagonal matrix $B(x, z)$ has rank d_x for $x \in \mathcal{X}_z^\circ$, so $H(x, z)$ has rank d_x only if Γ has rank d_x.

THEOREM S2: *Suppose that* $\mathcal{X} = \mathcal{X}_z = \mathcal{X}_{z,1} \times \cdots \times \mathcal{X}_{z,d_x}$ *for* $z \in \mathcal{Z} = \{0, 1\}$, *where* $\mathcal{X}_{z,k} = [\underline{x}_k, \overline{x}_k]$ *is a compact interval for each* k *and* z. *Also assume that* $\mathcal{E}_{x,z}$ *does not depend on* x *or* z. *Suppose that the sets* $\mathcal{X}_k^\dagger \equiv \{x_k \in \mathcal{X}_k : F_{X_k|Z}(x_k|0) = F_{X_k|Z}(x_k|1)\}$ *are finite for each* k. *Then* g^* *is point-identified on* $\operatorname{supp}(X, \varepsilon)$.

Theorem S2 uses the strong assumption that (X, Z) has rectangular support and so may be of less practical interest. However, Theorem S2 is still interesting from a theoretical standpoint, because it demonstrates a case in which the traditional order condition ($d_z \geq d_x$) is not needed for point-identification. The main assumption in Theorem S2 is that the sets $\mathcal{X}_k^\dagger$ are finite, as in Theorem 2. (The nonemptyness assumed in Theorem 2 is already implied here by the conditions on the joint support of (X, Z).)

PROOF OF THEOREM S2: To ease notation, let $\mathcal{X}_k \equiv \mathcal{X}_{z,k}$, which by assumption does not depend on z. For each $k = 1, \ldots, d_x$, let $\{\xi_k^l\}_{l=0}^{L_k}$ denote the elements of the finite set $\mathcal{X}_k^\dagger$ arranged to be increasing. Since $\underline{x}_k, \overline{x}_k \in \mathcal{X}_k^\dagger$ and $\mathcal{X}_k$ is an interval, it follows that $L_k \geq 1$, $\xi_k^0 = \underline{x}_k$, $\xi_k^{L_k} = \overline{x}_k$, and $\mathcal{X}_k = \bigcup_{l=0}^{L_k-1} \mathcal{X}_k^l$, with $\mathcal{X}_k^l \equiv [\xi_k^l, \xi_k^{l+1}]$. Given that $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_{d_x}$, this construction divides $\mathcal{X}$ into nonempty cells each contained in $\mathcal{X}$. Let $\mathcal{J}$ denote the collection of all d_x-tuples of integers drawn from $\bigotimes_{k=1}^{d_x}\{0, \ldots, L_k - 1\}$ and $j = (j_1, \ldots, j_{d_x})$ an element of $\mathcal{J}$. Then $\mathcal{X} = \bigcup_{j \in \mathcal{J}} \mathcal{X}^j$, where $\mathcal{X}^j \equiv \bigotimes_{k=1}^{d_x} \mathcal{X}_k^{j_k} \subseteq \mathcal{X}$ is an individual cell of $\mathcal{X}$. Also, define $\mathcal{X}_k^\circ \equiv (\underline{x}_k, \overline{x}_k)$, $\mathcal{X}_k^{l,\circ} \equiv (\xi_k^l, \xi_k^{l+1})$, and $\mathcal{X}^{j,\circ} \equiv \bigotimes_{k=1}^{d_x} \mathcal{X}_k^{j_k,\circ}$.

For each k, define

$$\pi_k(x_k) \equiv Q_{X_k|Z}\big(F_{X_k|Z}(x_k|1)|0\big)$$

and let $\pi(x) \equiv (\pi_1(x_1), \ldots, \pi_{d_x}(x_{d_x}))$. Fix a $j \in \mathcal{J}$ and let x^0 be an arbitrary element of $\mathcal{X}^{j,\circ}$. Consider the recursive sequence $x^{n+1} = \pi(x^n)$ for $n \geq 0$. The kth component of this sequence is contained in $\mathcal{X}_k^{j_k,\circ} \equiv (\xi_k^{j_k}, \xi_k^{j_k+1})$. Otherwise, there would exist an N such that $x_k^N \in \mathcal{X}_k^{j_k,\circ}$ but $x_k^{N+1} \notin \mathcal{X}_k^{j_k,\circ}$, say, because $x_k^{N+1} \leq \xi_k^{j_k}$. Then

$$F_{X_k|Z}\big(\xi_k^{j_k}|1\big) = F_{X_k|Z}\big(\xi_k^{j_k}|0\big) \geq F_{X_k|Z}\big(x_k^{N+1}|0\big) = F_{X_k|Z}\big(x_k^N|1\big),$$

which contradicts $x_k^N > \xi_{j_k}^k$ under Assumption C. A symmetric argument provides a contradiction if $x_k^{N+1} \geq \xi_k^{j_k+1}$. Hence $x_k^n \in \mathcal{X}_k^{j_k,\circ}$ for each k and so $x^n \in \mathcal{X}^{j,\circ} \subseteq \mathcal{X}^\circ$ for each n.

Let

$$\sigma_k(x_k) \equiv \operatorname{sgn}\big[F_{X_k|Z}(x_k|1) - F_{X_k|Z}(x_k|0)\big].$$

Then for each k, $\sigma_k(x_k)$ is either 1 or -1 for all $x_k \in \mathcal{X}_k^{j_k,\circ}$ because $F_{X_k|Z}(\cdot|1) - F_{X_k|Z}(\cdot|0)$ is continuous by Assumption C and, by construction, nonzero on $\mathcal{X}_k^{j_k,\circ}$. If $\sigma_k(x_k^0) = -1$, then x_k^n is decreasing in n because

$$\pi_k(x_k) = Q_{X_k|Z}\big(F_{X_k|Z}(x_k|1)|0\big) < Q_{X_k|Z}\big(F_{X_k|Z}(x_k|0)|0\big) = x_k$$

for any $x_k \in \mathcal{X}_k^{j_k,\circ}$ and $x_k^n \in \mathcal{X}_k^{j_k,\circ}$ for all n. If $\sigma_k(x_k^0) = 1$, then x_k^n is increasing by a symmetric argument. In either case, x_k^n converges because $\mathcal{X}_k$ is compact. Furthermore, by Assumption C,

$$\begin{aligned} F_{X_k|Z}\big(\lim x_k^n|0\big) &= \lim F_{X_k|Z}\big(\pi_k\big(x_k^n\big)|0\big) \\ &= \lim F_{X_k|Z}\big(x_k^n|1\big) = F_{X_k|Z}\big(\lim x_k^n|1\big), \end{aligned}$$

so $\lim x_k^n \in \mathcal{X}_k^\dagger$ for each k. Since $x_k^n \in \mathcal{X}_k^{j_k,\circ} \equiv (\xi_k^{j_k}, \xi_k^{j_k+1})$ for all n, this implies that $\lim x_k^n = \xi_k^{j_k}$ if $\sigma_k(x_k^0) = -1$ and $\lim x_k^n = \xi_k^{j_k+1}$ if $\sigma_k(x_k^0) = 1$. Notice that this limiting point may depend on j, but it does not depend on $x^0 \in \mathcal{X}^{j,\circ}$, since σ_k is constant on $\mathcal{X}_k^{j_k,\circ}$.

The definition and properties of x^n together with (S1) imply that

$$I^g\big(x^{n+1}, e\big) = D^g\big(\mathbf{F}_{X|Z}\big(\pi\big(x^n\big)|0\big), e\big) = D^g\big(\mathbf{F}_{X|Z}\big(x^n|1\big), e\big) = I^g\big(x^n, e\big)$$

for all n and any $e \in \mathcal{E}^\circ$, since $\mathcal{E}_{x,z}^\circ = \mathcal{E}^\circ$ for all x, z by assumption. As noted in the main text, I^g is continuous everywhere due to Assumption G, so $I^g(x^0, e) = I^g(\lim x^n, e) \equiv J^g(e)$. Since x^0 was an arbitrary element of $\mathcal{X}^{j,\circ}$, and $\lim x^n$ depends only on j and not on $x^0 \in \mathcal{X}^{j,\circ}$, $I^g(x, e) = I^g(\lim x^n, e) \equiv J^g(e)$ for all $x \in \mathcal{X}^{j,\circ}$. By the continuity of I^g, it follows that $I^g(x, e) = J^g(e)$ for all $x \in \mathcal{X}^j \equiv \operatorname{cl} \mathcal{X}^{j,\circ}$ and $e \in \mathcal{E}$.

The preceding argument was for an arbitrary cell $j \in \mathcal{J}$. The conclusion that $I^g(x, e)$ does not vary in x over $\mathcal{X}^j$ also applies to any cell $\mathcal{X}^{j'}$ that is adjacent to $\mathcal{X}^j$. Since $\mathcal{X}^{j'}$ and $\mathcal{X}^j$ share a boundary, their intersection is nonempty. Thus, $I^g(x, e) = J^g(e)$ for all $x \in \mathcal{X}^j \cup \mathcal{X}^{j'}$ as well. Repeating this argument a finite number of times shows that $I^g(x, e) = J^g(e)$ for all $x \in \mathcal{X} = \bigcup_{j \in \mathcal{J}} \mathcal{X}^j$ and $e \in \mathcal{E}$. As discussed in the main text, this implies that g^* is point-identified on $\operatorname{supp}(X, \varepsilon)$. *Q.E.D.*

S3. EXAMPLE: THE CAUSAL EFFECT OF CLASS SIZE ON SCHOOLING OUTCOMES

In this section, I illustrate the restrictiveness of the identifying assumptions in the context of determining the causal effect of class size X on a measure of schooling outcomes Y. In this example, W stands for observable covariates including school characteristics and socioeconomic variables. The unobservable ε aggregates the many other factors involved in determining schooling outcomes Y, including parental involvement, unobserved family background characteristics, and preferences for class size that lead to sorting. The interpretation of G.S in terms of rank invariance was given in Section 1. By itself, G.S has no observable content.[6] It would fail, for example, if the marginal effect of class size on outcomes is controlled by a different unobservable than the one explaining differences in performance for a fixed class size (say 20), for example, if $Y = \varepsilon + \widetilde{\varepsilon}(X - 20)$ for distinct unobservables ε and $\widetilde{\varepsilon}$. Assumption FS depends on the particular instrument used. Consider two instruments from the literature.

EXAMPLE 1: Hoxby (2000) constructs an instrument Z from exogenous fluctuations in the number of enrolled students caused by changes in the timing of births around the calendar year.[7] Let $\overline{s}(W, \widetilde{\eta})$ represent the number of students who would be enrolled if the timing of births were nonvarying, where $\widetilde{\eta}$ is some (possibly multidimensional) random vector that may be arbitrarily dependent with ε. Let Z represent proportional exogenous fluctuations in enrollment (i.e., Z takes values around 1 and $Z > 0$) so that the actual number of enrolled students is $s(W, Z, \widetilde{\eta}) = Z\overline{s}(W, \widetilde{\eta})$. Also, suppose that $\overline{c}(W, \widetilde{\eta})$ denotes the number of classes that the school would maintain in a baseline year and let the actual number of classes be given by $c(W, Z, \widetilde{\eta}) = d(W, Z)\overline{c}(W, \widetilde{\eta})$, where $d > 0$ and $d(W, 1) = 1$. Assuming that classes are split into equal sizes,

$$\begin{aligned} X &= \frac{s(W, Z, \widetilde{\eta})}{c(W, Z, \widetilde{\eta})} = \frac{Z}{d(W, Z)} \frac{\overline{s}(W, \widetilde{\eta})}{\overline{c}(W, \widetilde{\eta})} \\ &\equiv h_1(W, Z)h_2(W, \widetilde{\eta}) \equiv h_1(W, Z)\eta, \end{aligned}$$

where $h_1(W, Z) \equiv Z/d(W, Z) > 0$ and $\eta \equiv h_2(W, \widetilde{\eta})$ so that FS.S is satisfied. Hoxby argues that Z is conditionally exogenous, so if one agrees with her argument, then $(\widetilde{\eta}, \varepsilon) \perp\!\!\!\perp Z|W$, which implies also $(\eta, \varepsilon) \perp\!\!\!\perp Z|W$ and, hence, that FS.E holds. Theorem S1 establishes point-identification of g^* if Z has some marginal effect on class sizes (conditional on W) for all class sizes under consideration. This is a weak assumption that can be checked in the data. Assumption FS.S may fail if d also depends on $\widetilde{\eta}$, so that the way a school district changes class size in response to birth fluctuations is influenced by factors affecting the number of students.

EXAMPLE 2: Feinstein and Symons (1999) use geographic indicator variables for Z. Variation in these indicators corresponds to different local authorities (a unit of local government in England), which have different policies on class size. The authors cite work on the determinants of migration to argue that geographic location at the local authority level is exogenous to schooling

[6]Although, as shown in Theorem 1, the model is testable in its entirety.

[7]The exact construction of Hoxby's instrument is more subtle, but it is not important for the current discussion.

outcomes after conditioning on measures of social class, parents' education, and parental interest. If one agrees with their argument, then FS.E holds when W is a set of controls containing these variables. Consider a school in local authority A that has relatively small class size compared to other schools in A with similar socioeconomic factors W. Assumption FS.S requires that were this school to be counterfactually located in local authority B, then it would also have a small class size relative to other W-comparable schools in B. Theorem 2 establishes point-identification of g^* if variation in local authority is correlated with the realized, absolute level of class size in a regular way.

REFERENCES

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441. [2]

——— (2007): "Identification of Non-Additive Structural Functions," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Vol. 3, ed. by R. Blundell, W. Newey, and T. Persson. New York: Cambridge University Press. [2]

FEINSTEIN, L., AND J. SYMONS (1999): "Attainment in Secondary School," *Oxford Economic Papers*, 51, 300–321. [5]

HOXBY, C. M. (2000): "The Effects of Class Size on Student Achievement: New Evidence From Population Variation," *Quarterly Journal of Economics*, 115, 1239–1285. [5]

OLMSTED, J. M. H. (1961): *Advanced Calculus*. New York: Appleton-Century-Crofts. [2]

RUDIN, W. (1976): *Principles of Mathematical Analysis*. New York: McGraw-Hill. [2]

Dept. of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, U.S.A.; a-torgovitsky@northwestern.edu.

Manuscript received April, 2011; final revision received November, 2014.