Supplement to "Partial identification of the distribution of treatment effects with an application to the Knowledge is Power Program (KIPP)"

(*Quantitative Economics*, Vol. 12, No. 1, January 2021, 143–171)

Brigham R. Frandsen
Department of Economics, Brigham Young University

Lars J. Lefgren
Department of Economics, Brigham Young University

Appendix A: Proofs

The following result is crucial to Theorem 2.

Lemma. *Let X and Y be random variables with marginal distributions F_X and F_Y, where $\mathcal{Y}$ is the support of Y. Suppose continuously distributed random variable X is weakly stochastically increasing in Y. Then*

$$\underline{F}_{X|Y}(x|y) \leq \Pr(X \leq x|Y = y) \leq \bar{F}_{X|Y}(x|y),$$

where

$$\underline{F}_{X|Y}(x|y) = \begin{cases} 0, & x < F_X^{-1}\big(F_Y(y)\big), \\ \dfrac{F_X(x) - F_Y(y)}{1 - F_Y(y)}, & x \geq F_X^{-1}\big(F_Y(y)\big), \end{cases}$$

and

$$\bar{F}_{X|Y}(x|y) = \begin{cases} \dfrac{F_X(x)}{F_Y(y)}, & x \leq F_X^{-1}\big(F_Y(y)\big), \\ 1, & x \geq F_X^{-1}\big(F_Y(y)\big). \end{cases}$$

Furthermore, the bounds are sharp.

Proof. Take the lower bound first. Assume $x \geq F_X^{-1}(F_Y(y))$ since the bound is trivially satisfied otherwise. Define $g_x^*(\cdot) := \Pr(X \leq x|Y = \cdot)$. Stochastic increasingness puts constraints on the class of functions, $\mathcal{G}_x$, to which $g_x^*(\cdot)$ belongs:

$$\mathcal{G}_x = \left\{ g_x : \begin{array}{ll} (1) & g_x\big(y''\big) \leq g_x\big(y'\big),\ y' < y'' \\ (2) & \int_{\mathcal{Y}} g_x(s)\, dF_Y(s) = F_X(x) \\ (3) & 0 \leq g_x(y) \leq 1,\ y \in \mathcal{Y} \end{array} \right\}.$$

Brigham R. Frandsen: frandsen@byu.edu
Lars J. Lefgren: lars_lefgren@byu.edu

© 2021 The Authors. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at http://qeconomics.org. https://doi.org/10.3982/QE1273

The first constraint imposes stochastic increasingness. The second requires that the conditional distribution integrate to the marginal. The third requires that the function be bounded between zero and one. The lower bound, $\underline{F}_{X|Y}(x|y)$, is therefore the minimum of $g_x(y)$ where the minimization is over the function class $\mathcal{G}_x$:

$$\underline{F}_{X|Y}(x|y) = \min_{g_x(\cdot)\in\mathcal{G}_x} g_x(y).$$

Constraint (1) in the definition of $\mathcal{G}_x$ means $g_x(y) \geq g_x(y'')$ for all $y'' \geq y$, which at the lower bound will clearly bind, implying $g_x(y'')$ will equal some constant, K_x, for all $y'' \geq y$. Constraint (3) will clearly bind with $g_x(y') = 1$ for all $y' < y$. The remaining constraint, (2), then determines the result:

$$\begin{aligned}
F_X(x) &= \int_{\mathcal{Y}} g_x(s)\, dF_Y(s) \\
&= F_Y(y) + \int_{[y,\infty)\cap\mathcal{Y}} K_x\, dF_Y(s) \\
&= F_Y(y) + K_x \int_{[y,\infty)\cap\mathcal{Y}} dF_Y(s) \\
&= F_Y(y) + K_x\big(1 - F_Y(y)\big) \\
&\Leftrightarrow K_x = \frac{F_X(x) - F_Y(y)}{1 - F_Y(y)}.
\end{aligned}$$

Now take the upper bound. Assume $x \leq F_X^{-1}(F_Y(y))$ since the bound is trivially satisfied otherwise. The upper bound for $\Pr(X \leq x|Y = y)$, $\bar{F}_{X|Y}(x|y)$, is the maximum of $g_x(y)$ over $g_x(\cdot) \in \mathcal{G}_x$:

$$\max_{g_x(\cdot)\in\mathcal{G}_x} g_x(y).$$

Constraint (1) in the definition of $\mathcal{G}_x$ means $g_x(y) \leq g_x(y')$ for all $y' \leq y$, which at the upper bound will clearly bind, implying $g_x(y')$ will equal some constant, C_x, for all $y' \leq y$. Constraint (3) will clearly bind with $g_x(y'') = 0$ for all $y'' > y$. The remaining constraint then determines the result:

$$\begin{aligned}
F_X(x) &= \int_{\mathcal{Y}} g_x(s)\, dF_Y(s) \\
&= \int_{(-\infty,y]\cap\mathcal{Y}} C_x\, dF_Y(s) \\
&= C_x \int_{(-\infty,y]\cap\mathcal{Y}} dF_Y(s) \\
&= C_x F_Y(y) \\
&\Leftrightarrow C_x = \frac{F_X(x)}{F_Y(y)}.
\end{aligned}$$

To see sharpness, let Y be marginally uniformly distributed on the unit interval. The lower bound is obtained when conditional on $Y = s$ for all $s \in [0, y)$, $X = Y$ almost surely

and conditional on $Y = s'$ for all $s' \in [y, 1]$ X is distributed uniformly on $[y, 1]$. Note that this satisfies (mutual) stochastic increasingness. Similarly, the upper bound is obtained when conditional on $Y = s$ for all $s \in (y, 1]$, $X = Y$ almost surely and conditional on $Y = s'$ for all $s' \in [0, y]$ X is distributed uniformly on $[0, y]$. Note that this also satisfies (mutual) stochastic increasingness. □

Proof of Theorem 2. Note that by definition

$$\Pr\big(\Delta \le t|Y(0)\big) = \Pr\big(Y(1) \le Y(0) + t|Y(0)\big).$$

Since $Y(1)$ is stochastically increasing in $Y(0)$, Lemma A applies to this case, taking $x = Y(0) + t$; $y = Y(0)$; $F_X = F_1$; $F_Y = F_0$. Making these substitutions in the lemma's result gives the result in the theorem. The argument for the lower bound is similar. Bounds on the conditional distribution of Δ given $Y(1)$ can be obtained by imposing that $Y(0)$ is stochastically increasing in $Y(1)$ in an analogous manner. Note that by definition

$$\Pr\big(\Delta \le t|Y(1)\big) = \Pr\big(Y(0) \ge Y(1) - t|Y(1)\big) = 1 - \Pr\big(Y(0) \le Y(1) - t|Y(1)\big)$$

since $Y(0)$ is assumed to be continuously distributed. We can thus simply interchange $Y(1)$ and $Y(0)$ in (1) and take the complement to obtain an upper bound:

$$F^U_{\Delta|Y(1)}\big(t|Y(1)\big) := \begin{cases} 1, & Y(1) + t < \tilde{Y}(0), \\ \dfrac{1 - F_0\big(Y(1) + t\big)}{1 - F_1\big(Y(1)\big)}, & Y(1) + t \ge \tilde{Y}(0), \end{cases} \tag{A1}$$

and similarly with (2) to obtain a lower bound:

$$F^L_{\Delta|Y(1)}\big(t|Y(1)\big) := \begin{cases} \dfrac{F_1\big(Y(1)\big) - F_0\big(Y(1) + t\big)}{F_1\big(Y(1)\big)}, & Y(1) + t \le \tilde{Y}(0), \\ 0, & Y(1) + t \ge \tilde{Y}(0), \end{cases} \tag{A2}$$

where $\tilde{Y}(0) := F_0^{-1}(F_1(Y(1)))$. □

Proof of Theorem 3. First, define $\mathbb{D}$ to be the set of continuous, bounded mappings from $\bar{\mathcal{Y}} \times \mathcal{D}$ to $\mathbb{R}^4$ and $\mathbb{E}$ the set of bounded continuous mappings from $\bar{\mathcal{Y}} \times \mathcal{D}$ to $\mathbb{R}^2$. The spaces $\bar{\mathcal{Y}}$ and $\mathcal{D}$ are defined in the statement of the theorem. One element of $\mathbb{D}$, evaluated at $v = (y, t)'$, is

$$\theta(v) := \begin{pmatrix} E\big[1(Y_i \le y)\big] \\ E\big[1(Y_i \le y)D_i\big] \\ E\big[1(Y_i \le y + t)D_i\big] \\ E[D_i] \end{pmatrix}.$$

Note that the estimand,

$$H(v) := \begin{pmatrix} \dfrac{F_1(y + t) - F_0(y)}{1 - F_0(y)} \\ \dfrac{F_1(y + t)}{F_0(y)} \end{pmatrix},$$

can be written as a function of θ:

$$H = \begin{pmatrix} \dfrac{E\big[1(Y \le y+t)D_i\big]\big(1-E[D_i]\big) - E[D_i]\big(E\big[1(Y \le y)\big] - E\big[1(Y \le y)D_i\big]\big)}{E[D_i]\big(1-E[D_i] - E\big[1(Y \le y)\big] + E\big[1(Y \le y)D_i\big]\big)} \\ \dfrac{\big(1-E[D_i]\big)E\big[1(Y \le y+t)D_i\big]}{E[D_i]\big(E\big[1(Y \le y)\big] - E\big[1(Y \le y)D_i\big]\big)} \end{pmatrix}$$
$$= \phi(\theta),$$

where $\phi : \mathbb{D} \to \mathbb{E}$ is a Hadamard differentiable map with Hadamard derivative evaluated at $\theta(v)$ of

$$\phi'_\theta(h) = J(v)\gamma(v)h,$$

where Jacobians J and γ are defined in the statement of the theorem. Define T_n as

$$T_n(v) = \begin{pmatrix} n^{-1}\sum_{i=1}^{n} 1(Y_i \le y) \\ n^{-1}\sum_{i=1}^{n} 1(Y_i \le y)D_i \\ n^{-1}\sum_{i=1}^{n} 1(Y_i \le y+t)D_i \\ n^{-1}\sum_{i=1}^{n} D_i \end{pmatrix},$$

and note that $\hat{H} = \phi(T_n)$. Given i.i.d. data, by the Donsker theorem, $\sqrt{n}(T_n - \theta)$ converges uniformly to a Gaussian process with zero mean function and covariance function $\Sigma(v, \tilde{v})$ given in the statement of the theorem. Then by Theorem 20.8 in van der Vaart (1998), we have that $\sqrt{n}(\hat{H} - H)$ converges to a Gaussian process with zero mean function and covariance function

$$J(v)\gamma(v)\Sigma(v, \tilde{v})\gamma(\tilde{v})'J(\tilde{v})'. \qquad \square$$

Proof of Theorem 4. Upper and lower bounds (5) and (6) on the overall distribution of treatment effects can be written as a function of the following vector of intermediate objects:

$$\theta_0 = (H, S, p),$$

where H is defined in the text,

$$p = \Pr(D_i = 1),$$

and $S : \mathcal{M} \to \mathbb{R}^2$ is the following map from a Donsker class $\mathcal{M}$ to $\mathbb{R}^2$:

$$S(m) = E\big[m(Y_i, D_i)\big].$$

Note that

$$m_0(y,d) := \begin{pmatrix} \dfrac{\max\{0, H_1(y)\}(1-d)}{1-p} \\ \dfrac{\min\{1, H_2(y)\}d}{p} \end{pmatrix}$$

is a member of a Donsker class as is

$$\hat{m}(y,d) = \begin{pmatrix} \dfrac{\max\{0, \hat{H}_1(y)\}(1-d)}{1-\hat{p}} \\ \dfrac{\min\{1, \hat{H}_2(y)\}d}{\hat{p}} \end{pmatrix}$$

with probability approaching one. Defining $\hat{S}(m) := \frac{1}{n}\sum_{i=1}^{n} m(Y_i, D_i)$, the sample counterpart to θ_0 is

$$\hat{\theta} = \begin{pmatrix} \hat{H} \\ \hat{S}(m) \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} D_i \end{pmatrix}, \quad m \in \mathcal{M}.$$

Since $\mathcal{M}$ is Donsker, and given the result in Theorem 3, we have that $\sqrt{n}(\hat{\theta} - \theta_0)$ converges in distribution to

$$\mathbb{G}_0 = \begin{pmatrix} \mathbb{G}_H \\ \mathbb{G}_S \\ \mathbb{G}_p \end{pmatrix} \in \ell^\infty(\bar{\mathcal{Y}} \times \mathcal{D})^2 \times \ell^\infty(\mathcal{M}) \times \mathbb{R}.$$

The treatment effect distribution bounds (5) and (6) can be written as the following function of θ_0:

$$\phi(\theta_0) = S\left(\begin{bmatrix} \dfrac{\big(H_1(Y_i) \vee 0\big)(1-D_i)}{1-p} \\ \dfrac{\big(H_2(Y_i) \wedge 1\big)D_i}{p} \end{bmatrix}\right)$$

with sample counterpart

$$\phi(\hat{\theta}) = \hat{S}\left(\begin{bmatrix} \dfrac{\big(\hat{H}_1(Y_i) \vee 0\big)(1-D_i)}{1-p} \\ \dfrac{\big(\hat{H}_2(Y_i) \wedge 1\big)D_i}{p} \end{bmatrix}\right).$$

Noting that the condition in the theorem's statement that $\Pr(F_1(Y_i + t) = F_0(Y_i)) = 0$ means $\Pr(H_1(Y_i) = 0) = 0$ and $\Pr(H_2(Y_i) = 1) = 0$, the map ϕ is Hadamard differentiable

with linear and continuous Hadamard derivative

$$\phi'_{\theta_0}(h) = h_S\big(m_0(Y_i, D_i)\big) + E\left[\begin{array}{c} \dfrac{h_1(Y_i)(1-D_i)1\{H^L > 0\}}{1-p} \\ \dfrac{h_2(Y_i)D_i 1\{H^U < 1\}}{p} \end{array}\right] + h_p E\left[\begin{array}{c} \dfrac{\big(H_1(Y_i) \vee 0\big)(1-D_i)}{(1-p)^2} \\ -\dfrac{\big(H_2(Y_i) \wedge 1\big)D_i}{p^2} \end{array}\right],$$

for direction $h = (h_1, h_2, h_S, h_p)'$. Theorem 2.1 in Fang and Santos (2018) applied to this setting then implies

$$\sqrt{n}\begin{pmatrix} \hat{F}^L_\Delta - F^L_\Delta \\ \hat{F}^U_\Delta - F^U_\Delta \end{pmatrix} \to \phi'_{\theta_0}(\mathbb{G}_0). \qquad \square$$

Proof of Theorem 6. The proof follows the proof of Theorem 2, but with all probabilities conditioned on $D(1) > D(0)$. $\square$

Theorem. *Let $\hat{Y}(0)$ and $\hat{Y}(1)$ be linear projections of potential outcomes on X with corresponding coefficients of determination R_0^2 and R_1^2. Then* $\mathrm{Cov}(Y(1), Y(0)) \geq 0$ *implies*

$$\mathrm{Corr}\big(\hat{Y}(1), \hat{Y}(0)\big) \geq -\sqrt{\big(1-R_0^2\big)\big(1-R_1^2\big)/\big(R_0^2 R_1^2\big)}.$$

Proof. Define $\varepsilon(1) = Y(1) - \hat{Y}(1)$ and $\varepsilon(0) = Y(0) - \hat{Y}(0)$. Note that by construction $\mathrm{Cov}(\hat{Y}(1), \varepsilon(0)) = \mathrm{Cov}(\hat{Y}(0), \varepsilon(1)) = 0$. Also, note that $\mathrm{Var}(\varepsilon(1)) = (1-R_1^2)\,\mathrm{Var}(Y(1))$ and $\mathrm{Var}(\varepsilon(0)) = (1-R_0^2)\,\mathrm{Var}(Y(0))$. Since $\varepsilon(0)$ is orthogonal to $\hat{Y}(1)$ and $\varepsilon(1)$ is orthogonal to $\hat{Y}(0)$, the covariance between potential outcomes can be written:

$$\mathrm{Cov}\big(Y(0), Y(1)\big) = \mathrm{Cov}\big(\hat{Y}(0), \hat{Y}(1)\big) + \mathrm{Cov}\big(\varepsilon(0), \varepsilon(1)\big). \tag{A3}$$

The Cauchy–Schwarz inequality implies

$$\mathrm{Cov}\big(\varepsilon(0), \varepsilon(1)\big) \leq \sqrt{\big(1-R_0^2\big)\,\mathrm{Var}\big(Y(0)\big)\big(1-R_1^2\big)\,\mathrm{Var}\big(Y(1)\big)}.$$

Inserting this into (A3) yields an upper bound on the covariance between potential outcomes:

$$\mathrm{Cov}\big(Y(0), Y(1)\big) \leq \mathrm{Cov}\big(\hat{Y}(0), \hat{Y}(1)\big) + \sqrt{\big(1-R_0^2\big)\,\mathrm{Var}\big(Y(0)\big)\big(1-R_1^2\big)\,\mathrm{Var}\big(Y(1)\big)}.$$

This upper bound is nonnegative when

$$\mathrm{Cov}\big(\hat{Y}(0), \hat{Y}(1)\big) \geq -\sqrt{\big(1-R_0^2\big)\,\mathrm{Var}\big(Y(0)\big)\big(1-R_1^2\big)\,\mathrm{Var}\big(Y(1)\big)},$$

or, equivalently,

$$\mathrm{Corr}\big(\hat{Y}(0), \hat{Y}(1)\big) \geq -\sqrt{\frac{\big(1-R_0^2\big)\big(1-R_1^2\big)}{R_0^2 R_1^2}}. \qquad \square$$

Appendix B: Incorporating covariates

For expositional simplicity, the bounds above were developed without additional covariates and assuming exogenous treatment assignment. In practice, the bounds may be substantially tightened by incorporating additional covariates X. In this section, we show how the framework can be extended to this cases.

When additional covariates are available that predict the outcome Y, the bounds may be tightened by adopting the following conditional version of the stochastic increasingness assumption.

Definition 1. $Y(0)$ and $Y(1)$ are **mutually stochastically increasing** conditional on X if $\Pr(Y(1) \leq t|Y(0) = y, X)$ and $\Pr(Y(0) \leq t|Y(1) = y, X)$ are each nonincreasing in y almost everywhere.

The following bounds on the treatment effect cdf conditional on $Y(0)$ and X are conditional versions of (1) and (2) and follow from the conditional mutual stochastic increasingness condition above:

$$F^L_{\Delta|Y(0),X}\big(t|Y(0),X\big) := \begin{cases} 0, & Y(0)+t < \tilde{Y}(1|X), \\ \dfrac{F_{1|X}\big(Y(0)+t|X\big) - F_{0|X}\big(Y(0)|X\big)}{1 - F_{0|X}\big(Y(0)|X\big)}, & Y(0)+t \geq \tilde{Y}(1|X), \end{cases}$$

$$F^U_{\Delta|Y(0),X}\big(t|Y(0),X\big) := \begin{cases} \dfrac{F_{1|X}\big(Y(0)+t|X\big)}{F_{0|X}\big(Y(0)|X\big)}, & Y(0)+t \leq \tilde{Y}(1|X), \\ 1, & Y(0)+t \geq \tilde{Y}(1|X), \end{cases}$$

where $\tilde{Y}(1|X) := F^{-1}_{1|X}(F_{0|X}(Y(0)|X))$; expressions for the treatment effect cdf conditional on $Y(1)$ and X are similar.

Bounds on the distribution of treatment effects conditional on $Y(0)$ (or $Y(1)$) alone—which are frequently of greater interest than the distribution conditional on $Y(0)$ and X—can be obtained by integrating the conditional bounds over the conditional distribution of X given $Y(0)$:

$$F^L_{\Delta|Y(d)}\big(t|Y(d)\big) = E\big[F^L_{\Delta|Y(d),X}\big(t|Y(d),X\big)|Y(d)\big], \tag{A4}$$

$$F^U_{\Delta|Y(d)}\big(t|Y(d)\big) = E\big[F^U_{\Delta|Y(d),X}\big(t|Y(d),X\big)|Y(d)\big], \tag{A5}$$

for $d = 0$ or 1. As before, bounds on the average treatment effect conditional on $Y(d)$ can be formed by integrating over the cdf bounds:

$$\Delta^L\big(Y(d)\big) = \int t\, dF^U_{\Delta|Y(d)}\big(t|Y(d)\big),$$

$$\Delta^U\big(Y(d)\big) = \int t\, dF^L_{\Delta|Y(d)}\big(t|Y(d)\big).$$

A simple method for computing bounds on the marginal cdf of the treatment effect is to average over the bounds on the conditional cdf as before, though these bounds may not

be sharp:

$$\tilde{F}^L_\Delta(t) = \max_{d\in\{0,1\}} E\big[F^L_{\Delta|Y(d),X}\big(t|Y(d),X\big)\big], \tag{A6}$$

$$\tilde{F}^U_\Delta(t) = \min_{d\in\{0,1\}} E\big[F^U_{\Delta|Y(d),X}\big(t|Y(d),X\big)\big]. \tag{A7}$$

Sharp bounds may be computed by directly searching over the space of bivariate copulae that satisfy mutual stochastic increasingness condition to find bounds conditional on X

$$F^L_{\Delta|X}(t|X) = \inf_{H(\cdot,\cdot)\in\mathcal{C}^{SI}} \int\int 1\big(F^{-1}_{1|X}(v|X) - F^{-1}_{0|X}(u|X) \le t\big)H(u,v)\,du\,dv,$$

$$F^U_{\Delta|X}(t|X) = \sup_{H(\cdot,\cdot)\in\mathcal{C}^{SI}} \int\int 1\big(F^{-1}_{1|X}(v|X) - F^{-1}_{0|X}(u|X) \le t\big)H(u,v)\,du\,dv,$$

which can then be averaged to produce bounds on the marginal cdf:

$$F^L_\Delta(t) = E\big[F^L_{\Delta|X}(t|X)\big], \tag{A8}$$

$$F^U_\Delta(t) = E\big[F^U_{\Delta|X}(t|X)\big]. \tag{A9}$$

The bounds (A8) and (A9) are sharp, but in practice are prohibitively costly to calculate as they require an infinite-dimensional optimization at each covariate value X. In the application, we estimate the integrated conditional bounds (A6) and (A7), which are slightly less tight but computationally feasible.

The conditional cdf bounds (A4) and (A5) can be consistently estimated by plugging in consistent estimators for the conditional cdfs $F_{1|X}$ and $F_{0|X}$. For the case where D_i is exogenous, the bounds can be constructed via the following steps for each untreated observation j:

1. Nonparametrically regress an indicator $1(Y_i \le Y_j)$ on X_i in the untreated subsample and construct predicted value $\hat{F}_{0|X}(Y_j|X_j)$
2. Nonparametrically regress an indicator $1(Y_i \le Y_j(0)+t)$ on X_i in the treated subsample and construct predicted value $\hat{F}_{1|X}(Y_j(0)+t|X_j)$
3. Form estimates of the bounds

$$\hat{F}^L_{\Delta|0,X}\big(t|Y_j(0),X_j\big) := \max\left\{0, \frac{\hat{F}_{1|X}\big(Y_j(0)+t|X_j\big) - \hat{F}_{0|X}\big(Y_j(0)|X_j\big)}{1-\hat{F}_{0|X}\big(Y_j(0)|X_j\big)}\right\}, \tag{A10}$$

$$\hat{F}^U_{\Delta|0,X}\big(t|Y_j(0),X_j\big) := \min\left\{1, \frac{\hat{F}_{1|X}\big(Y_j(0)+t|X_j\big)}{\hat{F}_{0|X}\big(Y_j(0)|X_j\big)}\right\}. \tag{A11}$$

The bounds (A4) and (A5) on the conditional distribution of treatment effects given $Y(0)$ can be constructed by nonparametrically regressing the estimates (A10) and (A11) on Y_i in the untreated sample. Bounds (3) and (4) on the conditional expectation of treatment effects given $Y(0)$ can be computed by numerically integrating the estimates

for (A4) and (A5) on a discrete grid. Analogous steps can be followed for the distribution of treatment effects given $Y(1)$.

The bound estimates are themselves functions of estimators for potential outcome conditional cdfs, $\hat{F}_{0|X}$ and $\hat{F}_{1|X}$. Several methods exist for estimating conditional cdfs; which is most suitable will depend on the specific empirical setting. For example, when treatment is exogenous and X has continuous components, the semiparametric distribution regression approach of Chernozhukov, Fernández-Val, and Melly (2013) may be most appropriate. When X is discrete, standard least squares regressions where treatment is fully interacted with X may be used.

Appendix C: Algorithm for approximating overall bounds

Let $\{C[i,j]\}_{i=1,\ldots,k,j=1,\ldots,k}$ be the elements of a $k \times k$ matrix which discretely approximates a bivariate copula function. By definition, each marginal distribution is uniform, which implies the following constraints:

$$\left\{\sum_{s=1}^{k} C[s,j] = 1\right\}_{j=1}^{k},$$

$$\left\{\sum_{s=1}^{k} C[i,s] = 1\right\}_{i=1}^{k-1}.$$

Stochastic increasingness means each conditional cdf is decreasing in the conditioning dimension, which implies the following set of constraints:

$$\left\{\left\{\sum_{s=1}^{i} C[s,j] \geq \sum_{s=1}^{i} C[s,j+1]\right\}_{j=1}^{k-1}\right\}_{i=1}^{k-1},$$

and

$$\left\{\left\{\sum_{s=1}^{j} C[i,s] \geq \sum_{s=1}^{j} C[i+1,s]\right\}_{i=1}^{k-1}\right\}_{j=1}^{k-1}.$$

Let the set of discrete copulae satisfying the above $k^2 + (k-1)^2$ constraints be denoted $\mathcal{C}_k$. Given estimates of the separate conditional distributions of $Y(0)$ and $Y(1)$ given X (obtained possibly via the methods described in Section 3.2) the lower bound on $F_{\Delta|X}$ can be approximated by solving the following linear program:

$$\min_{\{C[i,j]\}} \sum_{j=1}^{k}\sum_{i=1}^{k} 1\big(F_{1|X}^{-1}\big(r(i)\big) - F_{0|X}^{-1}\big(r(j)\big) \leq t\big) C[i,j]$$

$$\text{subject to } C[\cdot,\cdot] \in \mathcal{C}_k,$$

where

$$r(i) = \frac{i}{n} - \frac{1}{2n}.$$

The upper bound can be approximated by replacing the min with a max. Since the objective function and all constraints are linear, the program can be solved using efficient dual-simplex linear programming routines. Unconditional bounds can then be obtained by integrating the conditional bounds over X. In practice, the algorithm works well for $k \approx 100$ and discrete X with a moderate number of cells.

Appendix D: Simulations

This section illustrates the bounds on the distribution of treatment effects derived above using numerical simulations. The simulations adopt the following data generating process. Untreated potential outcomes are generated as $Y_i(0) = \beta X_i + \varepsilon_i$. The treated potential outcome is $Y_i(1) = Y_i(0) + \delta$.[1] The treatment indicator D_i is assigned independently of X_i and ε_i by random lottery whereby half the sample receives $D_i = 1$ and half receive $D_i = 0$. The observed and unobserved variables are generated according to

$$\begin{pmatrix} X_i \\ \varepsilon_i \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix}\right).$$

In the simulated model, σ_X^2 is set to one and β is set to $\sqrt{R^2}$ so that the variance of $Y_i(0)$ remains equal to one, and the R^2 between $Y_i(0)$ and X_i is $R^2 = 1 - \sigma_\varepsilon^2$. The simulations vary σ_ε^2 from 0.01 to 1, corresponding to an R^2 between $Y_i(0)$ and X from 0.99 to zero. The simulations also vary the treatment effect size δ from -1 to 1.

The first set of simulations illustrates how the bounds on the average treatment effect conditional on $Y_i(0)$ vary across the values of $Y_i(0)$. These simulations set the R^2 between $Y_i(0)$ and X_i to 0.7, corresponding to $\sigma_\varepsilon^2 = 0.3$ and $\beta = \sqrt{0.7} \approx 0.84$ and set the treatment effect size to $\delta = 1$. Figure A1 plots the bounds (3) and (4) as a function of $Y(0)$. The bounds always include the true treatment effect $\delta = 1$, and are tightest in the middle of the $Y(0)$ distribution, and widen in the tails. Notice that although in the simulated model the treatment effect is positive across the entire distribution of $Y(0)$, the bounds reach into negative territory for very high values of $Y(0)$, since the stochastic increasingness assumption allows for mean reversion; individuals with high values of $Y(0)$ have a larger probability of drawing a value of $Y(1)$ lower than $Y(0)$.

The second set of simulations shows how these bounds on the average treatment effect conditional on $Y_i(0)$ depend on the informativeness of the covariate X. These simulations set the treatment effect $\delta = 1$ and vary the R^2 between $Y_i(0)$ and X_i from zero to 0.99. Figure A2 plots the bounds (3) and (4) at $Y(0) = 0$ (i.e., at the median) as a function of the R^2. They show that the bounds tighten dramatically as the covariate X more strongly predicts outcomes.

The next set of simulations illustrates how bounds on the fraction of individuals harmed by treatment (i.e., the treatment effect cdf evaluated at zero) conditional on $Y(0)$ depends on the size of the treatment effect δ. As above, these simulations set the

[1] The simulations adopt a constant treatment effect so that the benchmark "truth" is the clearest: the fraction of individuals hurt will be one when $\delta < 0$ and zero when $\delta \geq 0$, and the average treatment effect conditional on $Y(0)$ will be δ for all $Y(0)$.

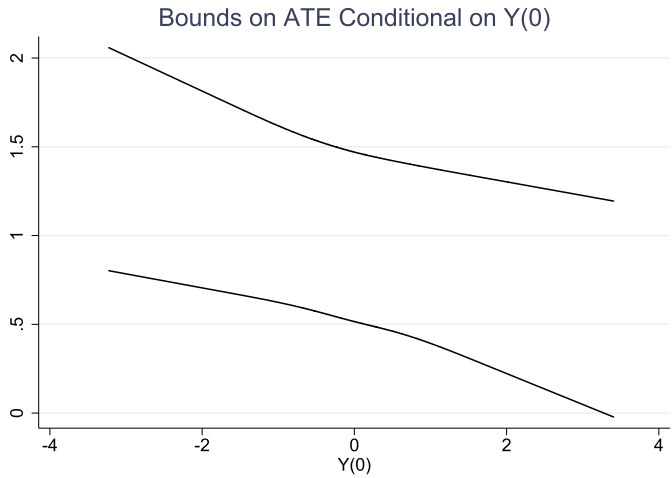

FIGURE A1. Simulated bounds on the average treatment effect conditional on untreated potential outcome. The true treatment effect is one for all values of $Y(0)$.

R^2 between $Y_i(0)$ and X_i to 0.7. Figure A3 plots the bounds (A4) and (A5) evaluated at zero as a function of δ for $Y(0) = 0$. Since the simulated model has constant treatment effects, the true fraction is one on the left side of the graph (where the treatment effect is negative) and zero on the right side. When the treatment effect is sufficiently large in magnitude, the bounds are quite tight. When the treatment effect is zero or slightly positive, the bounds are completely uninformative, spanning zero and one.

The next set of simulations shows how the bounds on the fraction of individuals hurt conditional on $Y_i(0)$ depend on the informativeness of the covariate X. These simulations set the treatment effect δ equal to one, and vary the R^2 between $Y_i(0)$ and X_i from zero to 0.99. Figure A4 plots the bounds (A4) and (A5) evaluated at zero as a function

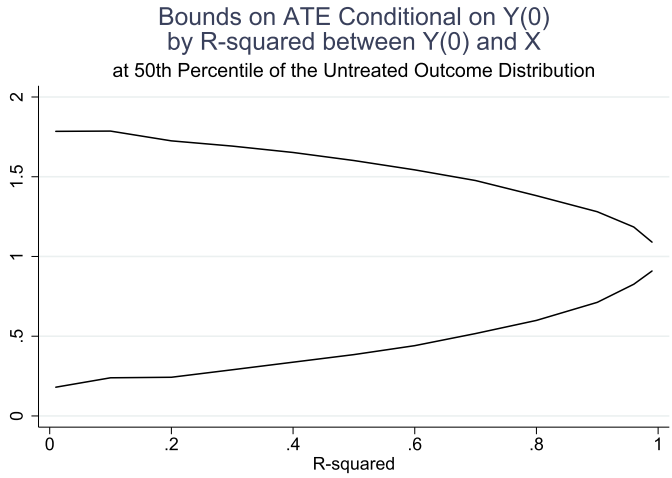

FIGURE A2. Simulated bound on the average treatment effect conditional on $Y(0) = 0$ as a function of the R^2 between $Y(0)$ and X. The true treatment effect is one.

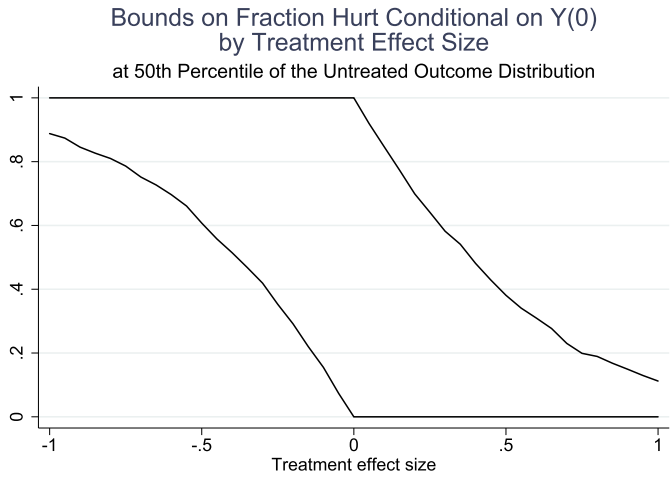

FIGURE A3. Simulated bound on the fraction hurt by treatment conditional on $Y(0) = 0$ as a function of the treatment effect. The true fraction is one when the treatment effect is negative (left side of the plot) and zero when the treatment effect is positive.

of R^2 for $Y(0) = 0$. Since the (constant) treatment effect in this simulation is positive, the true fraction is zero. On the far left, where the covariate has no predictive power, the bounds are quite wide, the upper bound reaching 0.3, but the bounds tighten dramatically as R^2 increases.

The next set of simulations shows how the bounds on the overall fraction of individuals hurt by treatment vary with the treatment effect size δ. Again, R^2 is set to 0.7 for these simulations. Figure A5 plots the Williamson–Downs bounds (which make no restrictions), our stochastic increasingness bounds calculated by integrating the conditional bounds, and the stochastic increasingness bounds calculated by searching over

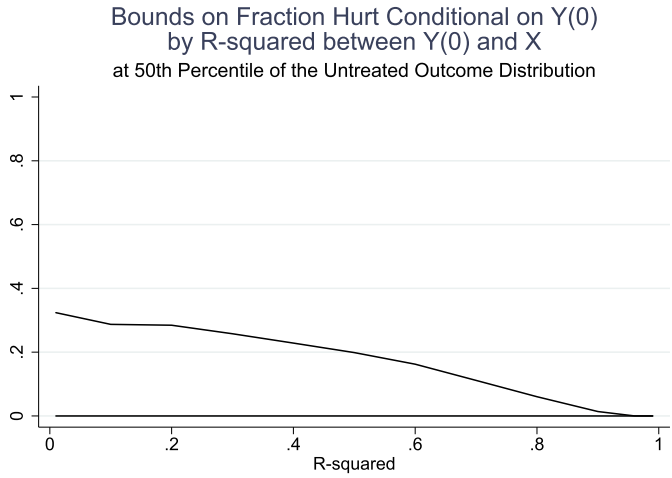

FIGURE A4. Simulated bound on the fraction hurt by treatment conditional on $Y(0) = 0$ as a function of the R^2 between $Y(0)$ and X. The true fraction in the simulation is zero.

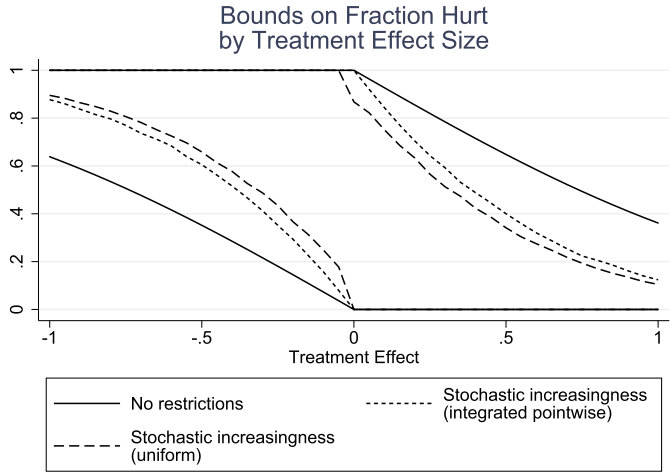

FIGURE A5. Simulated bounds on the fraction hurt by treatment as a function of the treatment effect. The true fraction is one when the treatment effect is negative (left side of the plot) and zero when the treatment effect is positive. The solid bounds impose no restrictions. The short–dashed bounds impose stochastic increasingness by integrating over the conditional (pointwise) bounds. The long-dashed bounds impose stochastic increasingness uniformly by searching over the space of copulae that satisfy stochastic increasingness.

the space of copula functions, each evaluated at zero, for a range of treatment effects sizes δ. All bounds are tightest when the treatment effect is large in magnitude. The stochastic increasingness bounds are much tighter than the classical Williamson–Downs bounds.

The next set of simulations shows how the bounds on the overall fraction of individuals hurt by treatment vary with the predictive power of the covariate X. Again, the treatment effect δ is set to one, and R^2 varies from zero to 0.99. Figure A6 plots the Williamson–Downs bounds (which make no restrictions), our stochastic increasingness bounds calculated by integrating the conditional bounds, and the stochastic increasingness bounds calculated by searching over the space of copula functions, each evaluated at zero, for a range of values of R^2. Since the treatment effect is positive, the true fraction is zero. On the left side of the plot, where the covariate has little explanatory power, the bounds we propose are quite wide, spanning zero to 0.35. However, even these are much tighter than the bounds that impose no restrictions, which span zero to over 0.6. As the R^2 between $Y(0)$ and X increases, the bounds tighten substantially.

The next set of simulations demonstrates how the bounds on the treatment effect cdf vary by treatment effect heterogeneity. To accomplish this, we modify the simulation setup by replace the constant treatment effect δ by a variable treatment effect δ_i, so that outcomes are generated according to

$$Y_i(0) = \beta X_i + \varepsilon_i,$$

$$Y_i(1) = Y_i(0) + \delta_i.$$

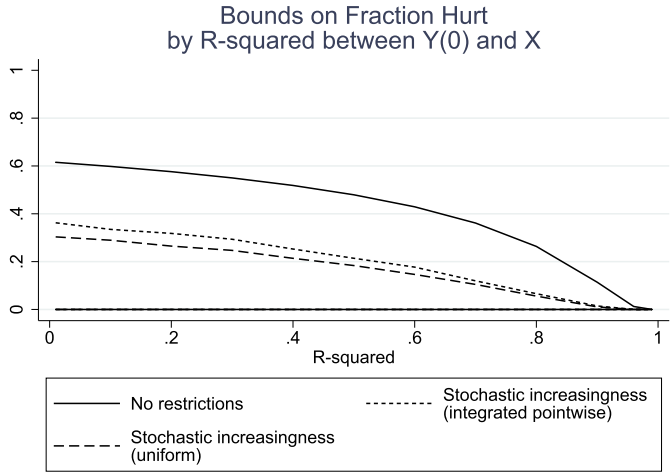

FIGURE A6. Simulated bound on the fraction hurt by treatment as a function of the R^2 between $Y(0)$ and X. The true fraction is zero. The solid bounds impose no restrictions. The short–dashed bounds impose stochastic increasingness by integrating over the conditional (pointwise) bounds. The long-dashed bounds impose stochastic increasingness uniformly by searching over the space of copulae that satisfy stochastic increasingness.

In order to hold fixed the marginal distributions of outcomes as the treatment effect distribution varies, we generate the treatment effect as

$$\delta_i = \alpha_0 + \alpha_1 \varepsilon_i + \eta_i,$$

where η_i normally distributed with variance $\sigma_\eta^2 = \sigma_\delta^2 - \alpha_1^2 \sigma_\varepsilon^2$, and $\alpha_1 = -0.5\sigma_\delta^2/\sigma_\varepsilon^2$. We generate ε_i and X_i as before. For these simulations, we set the R^2 between $Y_i(0)$ and X_i at 0.7, and the average treatment effect at $\alpha_0 = 0.5$. We vary the treatment effect standard deviation σ_δ between zero and one. Figure A7 plots the upper and lower bounds on the fraction hurt (i.e., treatment effect cdf evaluated at zero) along with the true fraction hurt for each value of σ_δ. Since the bounds depend only the marginal distributions of potential outcomes, which do not change in our setup as the treatment effect distribution varies, the bounds are flat, although they always contain the truth.

The final set of simulations investigates the finite-sample performance of inference procedures based on the asymptotic results in Theorem 4. These simulations are based on the following data-generating process:

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix} \right),$$

so that the true treatment effect distribution is $\delta_i := (Y_i(1) - Y_i(0)) \sim N(0, \sigma_1^2 + \sigma_0^2)$. The lower bound on the treatment effect cdf implied by stochastic increasingness in this

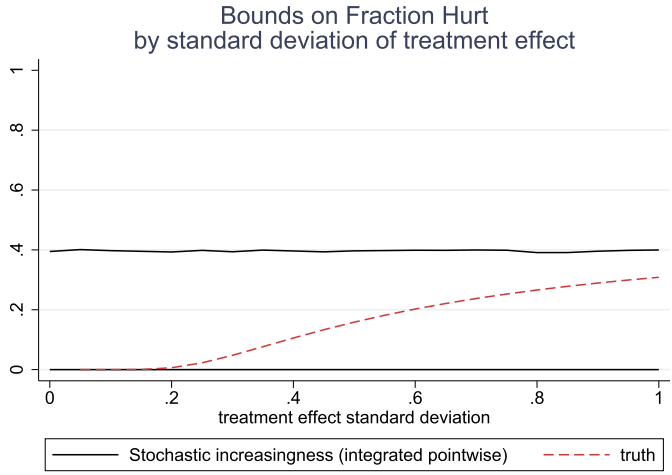

FIGURE A7. Simulated bounds on the fraction hurt by treatment as a function of the standard deviation of the treatment effect distribution. The treatment effect is normally distributed with a mean of 0.5. $Y(0)$ is normally distributed with mean zero and standard deviation one. The R-squared between $Y(0)$ and the covariate is 0.7.

example is

$$F_{\Delta}^{L}(t) = E\left[\max\left(0, \frac{\Phi\left(\frac{Y_i(0)+t}{\sigma_1}\right) - \Phi\left(\frac{Y_i(0)}{\sigma_0}\right)}{1-\Phi\left(\frac{Y_i(0)}{\sigma_0}\right)}\right)\right],$$

where Φ denotes the standard normal cdf. In each simulated dataset, we compute the lower bound estimate based on (9), estimate its sample variance using the bootstrap, apply Theorem 4 to construct nominal 95% confidence intervals for the lower bound, and check whether the confidence interval covers the true lower bound. We then report the coverage rate over 1000 simulated datasets. Figure A8 plots the coverage rate for a nominal 95% confidence interval at sample sizes ranging from $n = 100$ to $n = 1000$ for the cdf lower bound evaluated at $t = 0.5$, setting $\sigma_0 = 1$ and $\sigma_1 = 2$. In this case, the true lower bound is about $F_{\Delta}^{L}(0.5) \approx 0.151$. The figure shows the asymptotic approximation is excellent in finite samples, even for sample sizes as low as 100, which is smaller than our empirical example. The simulated coverage rate is very near the nominal rate over the whole range of sample sizes considered. We also investigate the performance of the inference procedure over the support of the treatment effect distribution. Figure A9 plots the coverage rate over a range of values from $t = -1$ to $t = 1$, fixing the sample size at $n = 500$. Here also we see the actual coverage rate is very close to the nominal rate over the range of treatment effect values considered.

APPENDIX E: BOUND ESTIMATION DETAILS

The empirical application in Section 4 reports estimates of bounds computed using the Stata command `tedistbounds` command, available from the authors on request. The

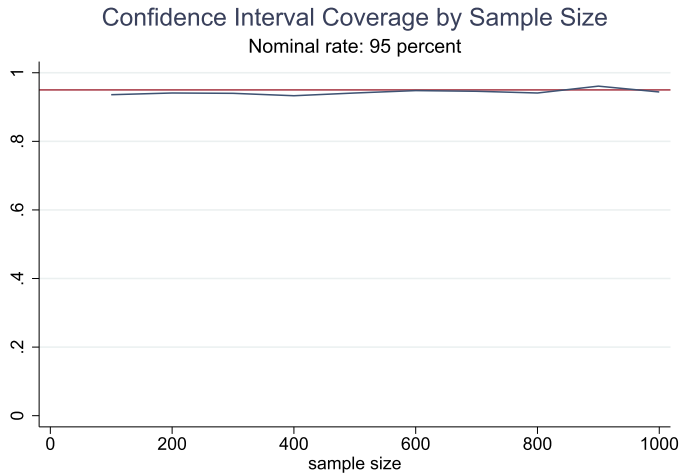

FIGURE A8. Simulated coverage rate of a nominal 95% confidence interval for the lower bound of the treatment effect cdf evaluated at $t = 0.5$ using normally distributed potential outcomes as described in the Appendix. Based on 1000 iterations.

Stata command implements estimation of the bounds in the context of an endogenous treatment with and instrumental variable, handling exogenous treatments as a special case where the instrument is identical to the treatment. The estimation proceeds as follows. First, we construct Abadie's (2003) κ weights:

$$\kappa = 1 - \frac{D(1-Z)}{1-\tau} - \frac{(1-D)Z}{\tau},$$

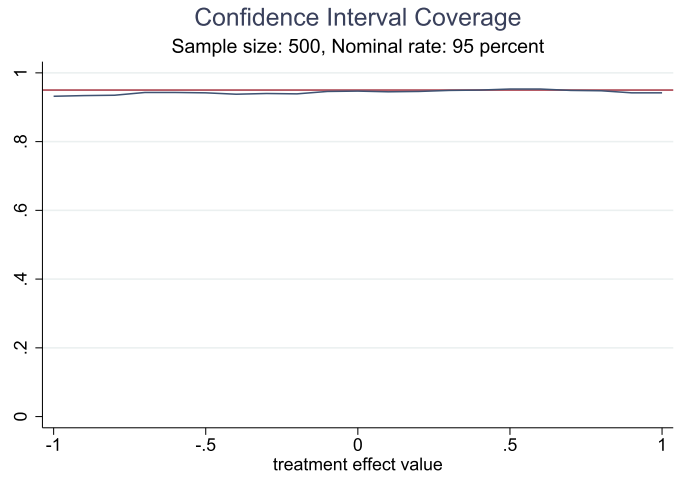

FIGURE A9. Simulated coverage rate of a nominal 95% confidence interval for the lower bound of the treatment effect cdf using normally distributed potential outcomes as described in the Appendix and a simulated sample size of $n = 500$. Based on 1000 iterations.

where $\tau = E[Z]$. Next, for a chosen pair of values (y, t) we calculate the bound on the conditional cdf of the compliers' treatment effect distribution evaluated at t, conditional on $Y(0) = y$ by the following procedure. We first estimate $F_{0|C,X}(y|X)$ via a κ-weighted least squares regression of $1(Y \leq y)$ on treatment, covariates X, and interactions of treatment with covariates. The fitted values from this regression where treatment is zero comprise the estimates $\hat{F}_{0|C,X}(y|X)$. Next, we estimate $F_{1|C,X}(y+t|X)$ via a κ-weighted least squares regression of $1(Y \leq y+t)$ on treatment, covariates, and their interactions. The fitted values from this regression, but evaluated with treatment set to one, comprise estimates $\hat{F}_{1|C,X}(y+t|X)$. We then combine the estimates obtained this way using formulae (A10) and (A11) to obtain the estimated bounds $\hat{F}_{\Delta|0,X,C}(t|y,X)$. To obtain the bound on the unconditional distribution of compliers' treatment effects, repeat this procedure with y set to each observed value of the outcome among the $D = 0$ subsample, and take a κ-weighted average:

$$\hat{F}_{\Delta|C}(t) = \frac{\sum_{i:D_i=0} \kappa_i \hat{F}_{\Delta|0,X,C}(t|Y_i, X_i)}{\sum_{i:D_i=0} \kappa_i}.$$

References

Abadie, A. (2003), "Semiparametric instrumental variable estimation of treatment response models." *Journal of Econometrics*, 113, 231–263. [16]

Chernozhukov, V., Fernández-Val, and B. Melly (2013), "Inference on counterfactual distributions." *Econometrica*, 81 (6), 2205–2268. [9]

Fang, Z. and A. Santos (2018), "Inference on directionally differentiable functions." *The Review of Economic Studies*, 86 (1), 377–412. ISSN 0034-6527. Available at https://doi.org/10.1093/restud/rdy049. [6]

van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge University Press, New York, NY. [4]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 14 January, 2019; final version accepted 14 July, 2020; available online 7 August, 2020.