Sensitivity of LATE Estimates to Violations of the Monotonicity Assumption

Claudia Noack

CDSE, University of Mannheim

Heidelberg, July 19, 2018

Local Average Treatment Effect

- Let D denote a binary treatment variable, Y_1 the potential outcome variable in the presence and Y_0 in the absence of treatment
- Average Treatment Effect: $\Delta = \mathbb{E}[Y_1 - Y_0]$
- But Δ is unobserved due to endogeneity of the treatment status
- Let Z denote the instrumental variable and Y_{dz} the potential outcome

Table: Groups with a binary instrumental variable Z and treatment status D

	Always Takers (AT)	Never Takers (NT)	Compliers (CO)	Defiers (DF)
Z=1	D=1	D=0	D=1	D=0
Z=0	D=1	D=0	D=0	D=1

Local Average Treatment Effect II

Assumption 1 (LATE)

1. *Random assignment:* $(Y_{11}, Y_{01}, Y_{10}, Y_{00}, D_0, D_1) \perp Z$
2. *Exclusion restriction:* $Y_d = Y_{d0} = Y_{d1}$ *for* $d \in \{0, 1\}$*;*
3. *Relevance:* $\mathbb{P}(D = 1|Z = 1) > \mathbb{P}(D = 1|Z = 0)$
4. *Monotonicity:* $\mathbb{P}(D_1 \geq D_0) = 1$

- Monotonicity rules out the existence of defiers
- Suppose above assumptions hold. The Wald estimand β^{IV}

$$\beta^{IV} = \frac{cov(Z, Y)}{cov(Z, D)} = \Delta_{CO}$$

where Δ_{CO} denotes the average treatment effect of compliers

Motivating Example

Angrist and Evans (1998)

- Analysis of the impact of having a third child on mother's wages (endogeneity problem)
- Instrument: Having two children of the same sex
  - Exclusion restriction, random assignment and relevance assumption seem to be plausible
  - Monotonicity: all parents weakly prefer children of different sexes
- Sex preferences might be correlated with potential wages

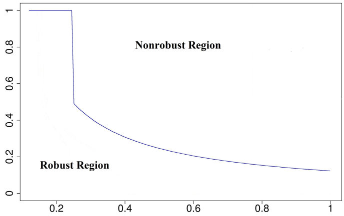
$\Rightarrow$ Validity of monotonicity and robustness of results might be questionable

Further Examples

Contribution

Sensitivity analysis of the monotonicity assumption

- Parameterization the degree of a violation of monotonicity into
  - Presence of defiers
  - Outcome heterogeneity of defiers and compliers
- Linking the degree of violation to implied treatment effects
- Identification of sensitivity parameters implying treatment effects being consistent with a certain empirical conclusion
- Approach: Identification of weakest sensitivity parameters using *breakdown frontiers*



Literature Review

Simple Case: Average Treatment Effect

Assumption 2

The IV satisfies random assignment, exclusion restriction and relevance.

From Angrist and Imbens (1995):

$$\beta^{IV} = \frac{\pi_{CO}}{\pi_{CO} - \pi_{DF}} \Delta_{CO} - \frac{\pi_{DF}}{\pi_{CO} - \pi_{DF}} \Delta_{DF}$$

Presence of Defiers

$$\lambda = \frac{\pi_{CO}}{\pi_{CO} - \pi_{DF}}$$

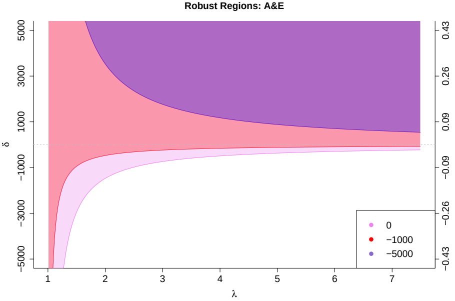
- $\lambda = 1$ absence of defiers; $\lambda \to \infty$ as many defiers as compliers

Outcome heterogeneity

$$\delta_a = \Delta_{CO} - \Delta_{DF}$$

Breakdown Frontier of Average Treatment Effect

$$BF_{ATE,CO}(\mu) = \left\{(\lambda, \delta_a) \in [1, \infty) \times \mathbb{R} : \delta_a = \frac{\mu - \beta^{IV}}{1 - \lambda}\right\}$$



Breakdown frontiers and robust regions for different claims of δ_a for $\lambda \in [1, 7.5]$

x-axis: λ, *Left y-axis:* δ_a *Right y-axis:* δ_a normalized by the sd of the outcome

More Interesting Case: Quantile Treatment Effect

Let $G_1(y) = \frac{cov(Z,\mathbf{1}[Y \leq y])}{cov(Z,D)}$. Based on Assumption 1 it holds that

$$G_d(y) = \frac{\pi_{CO}}{\pi_{CO} - \pi_{DF}} F_{Y_d^{CO}}(y) - \frac{\pi_{DF}}{\pi_{CO} - \pi_{DF}} F_{Y_d^{DF}}(y)$$

Presence of Defiers

$$\lambda = \frac{\pi_{CO}}{\pi_{CO} - \pi_{DF}}$$

Outcome Heterogeneity (Kolmogorov Smirnov Norm)

$$\sup_y \{ |F_{Y_d^{CO}}(y) - F_{Y_d^{DF}}(y)| \} \leq \delta_q \text{ with } \delta_q \in [0, 1]$$

Picture

Remark: One can further derive sharp upper and lower bounds of the prameter space. If parameter space is empty, Assumption 1 has to be violated.

Bounds

Bounding the Outcome Distributions

Theorem 1

Suppose Assumption 2 holds and the compliers outcome distributions is compatible with the sensitivity parameters. Denote the set of all CDFs on the support of the outcome variable Y by $\mathcal{F}_{\mathbb{Y}}$. Then, for $d \in \{0, 1\}$

$$\overline{F}_{Y_d^{co}}(y, \lambda, \delta_q) \leq F_{Y_d^{co}}(y) \leq \underline{F}_{Y_d^{co}}(y, \lambda, \delta_q).$$

Moreover, $\overline{F}_{Y_d^{co}}(y, \lambda, \delta_q)$ and $\underline{F}_{Y_d^{co}}(y, \lambda, \delta_q)$ are attainable as outcome CDFs. Thus, the bounds are sharp.

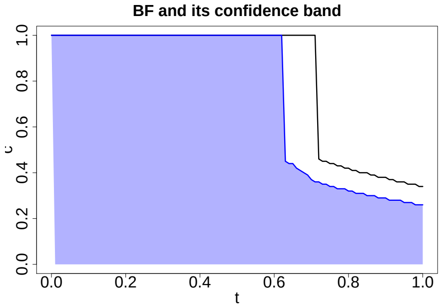
Bounds

Based on Theorem 1, one obtains quantile functions and quantile treatment effects $\underline{\Delta_{CO}}(\tau, \lambda, \delta_q)$. The breakdown frontier is given by:

$$BF_{QTE,CO}(\mu_\tau) = \{(\lambda, \delta_q) \in [0, 1]^2 : \delta_q = \arg \sup_{\delta_q^* \in [0,1]} \{\underline{\Delta_{CO}}(\tau, \lambda, \delta_q^*) \geq \mu_\tau\}\}$$

What is the Aim of Inference here?

$$\lim_{n \to \infty} \mathbb{P}(\widehat{RR}_{LB} \subseteq RR) \geq 1 - \alpha$$



Black line: Estimated breakdown frontier
Blue line: Confidence band at confidence level 0.9

Uniform Convergence

The breakdown frontier is a functional of some underlying parameters:

$$\phi : \ell^\infty(\mathbb{R}, \mathbb{R}^6) \to \ell^\infty([0,1]^2),\ \theta \mapsto \phi(\theta)$$

- $\theta = \left(F_{Y_{11}}, F_{Y_{10}}, F_{Y_{01}}, F_{Y_{00}}, G_1^+, G_0^+\right)$, where G_d^+ is a functional of outcome densities
- ϕ consists of among others mimimum, maximum, supremum, infimum operators

Theorem 2

Suppose Assumption 1 and additional regularity conditions are satisfied.

$$\sqrt{n}(\phi(\hat{\theta}) - \phi(\theta_0)) \xrightarrow{d} \phi'_{\theta_0}(\sqrt{n}(\hat{\theta} - \theta_0)) \equiv \mathcal{Z}_1,$$

where $\mathcal{Z}_1$ is a tight random element of $\ell^\infty([0,1]^2)$.

Outline of Proof

Step 1: Gaussianity of estimators of underlying parameters

- $G_d^+(y) = \int_{-\infty}^{y} \max\{0, f_1(z) - f_2(z)\}dz$, where $f_1(z)$ and $f_2(z)$ are densities
- $\sqrt{n}\left(\hat{G}_d^+(y) - G_d^+(y)\right) \to \mathbb{G}$, where $\mathbb{G}$ is a Gaussian process (similar to Anderson, Linton, and Whang, 2010)

Assumptions

Step 2: Hadamard Directional Differentiability

- ϕ is a composition of HDD functionals
- Chain rule for HDD functionals (Masten and Poirier, 2017)
- Delta-method for HDD functionals (Fang and Santos, 2016; Shapiro, 1991): If ϕ is HDD and $\sqrt{n}(\hat{\theta} - \theta_0) \to \mathbb{G}_0$ where $\mathbb{G}_0$ is a Gaussian process,

$$\sqrt{n}\left(\phi(\hat{\theta}) - \phi(\theta_0)\right) \xrightarrow{d} \phi'_{\theta_0}(\mathbb{G}_0) \equiv \mathcal{Z}$$

where $\mathcal{Z}$ is a tight random process with continuous path but unknown analytical distribution.

Illustration Definition

Bootstrap

- $\mathcal{Z}_1$ has unknown analytical distribution
- Analytical consistent estimates of ϕ'_{θ_0} difficult to obtain
- If there exists a consistent estimator of ϕ'_{θ_0}, bootstrap consistency follows from Fang and Santos (2016)
- A consistent estimator is proposed by Hong and Li (2016):

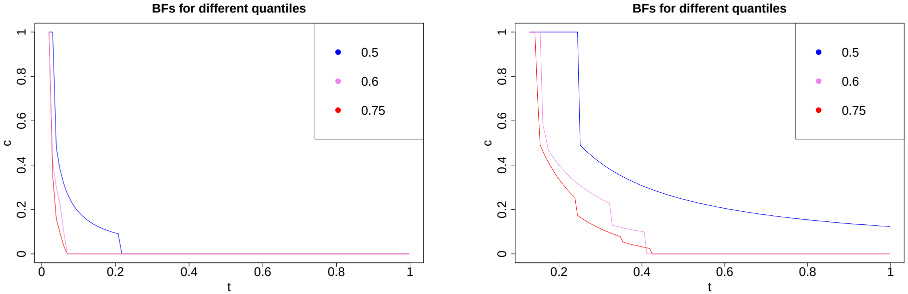
$$\tilde{\phi}'_{\theta_0}(\sqrt{n}(\hat{\theta}^* - \hat{\theta})) = \frac{\phi(\hat{\theta} + \varepsilon_n \sqrt{n}(\hat{\theta}^* - \hat{\theta})) - \phi(\hat{\theta})}{\varepsilon_n},$$

The tuning parameter ε_n is crucial.

$\Rightarrow$ Critical values can be bootstrapped to construct uniform confidence bands

Uniform Confidence Bands

Application to Angrist and Evans (1998)



Breakdown Frontier for $\overline{\Delta}_{CO}(\tau) \leq 0$
Left panel $\lambda \in [1, 6.06]$, *Right* panel $\lambda \in [1, 1.78]$

Thanks for your attention !

Anna Aizer and Joseph J. Doyle. Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, 130(2): 759–803, 2015.

Gordon Anderson, Oliver B. Linton, and Yoon-Jae Whang. Nonparametric Estimation of a Polarization Measure. *Working Paper: SSRN*, 2010.

Joshua Angrist and William Evans. Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, 88(3):450–77, 1998.

Joshua D. Angrist and Guido W. Imbens. Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association*, 90(430):431–442, 1995.

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434): 444–455, 1996.

David Autor, Andreas Ravndal Kostol, Magne Mogstad, and Bradley Setzler. Disability Benefits, Consumption Insurance, and Household Labor Supply. Working Paper 23466, National Bureau of Economic Research, June 2017.

Alexander Balke and Judea Pearl. Bounds on Treatment Effects from Studies with Imperfect Compliance. *Journal of the American Statistical Association*, 92:1171–1177, 1997.

Manudeep Bhuller, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. Incarceration, recidivism and employment. Working Paper 22648, National Bureau of Economic Research, 2016.

Sandra Black, Paul Devereux, and Kjell G Salvanes. Too Young to Leave the Nest? The Effects of School Starting Age. *The Review of Economics and Statistics*, 93(2):455–467, 2011.

Kristin Butcher and Anne Case. The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics*, 109(3):531–563, 1994.

David Card. The Causal Effect of Education on Earnings. *Handbook of Labor Economics*, 3: 1801 – 1863, 1999.

Gordon B. Dahl, Andreas Ravndal Kostol, and Magne Mogstad. Family Welfare Cultures. *The Quarterly Journal of Economics*, 129(4):1711–1752, 2014.

ClÃľment De Chaisemartin. Tolerating defiance? Local Average Treatment Effects without Monotonicity. *Forthcoming: Quantitative Economics*, 2017.

Will Dobbie, Jacob Goldin, and Crystal Yang. The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *AER*, 2018a.

Will Dobbie, Hans Grönqvist, Susan Niknami, Mårten Palme, and Mikael Priks. The Intergenerational Effects of Parental Incarceration. Working Paper 24186, National Bureau of Economic Research, 2018b.

Esther Duflo and Emmanuel Saez. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment*. *The Quarterly Journal of Economics*, 118(3):815–842, 2003.

Zheng Fang and Andres Santos. Inference on Directionally Differentiable Functions. *Working Paper: ARIX*, 2016.

Mario Fiorini and Katrien Stevens. Assessing the Monotonicity Assumption in IV and fuzzy RD designs. *Working Papers: University of Sydney, School of Economics:2014-13*, Feb 2016.

Eric French and Jae Song. The Effect of Disability Insurance Receipt on Labor Supply. *American Economic Journal: Economic Policy*, 6(2):291–337, 2014.

Donald P. Green and Daniel Wink. Using random judge assignments to estimate the effects of incarceration and probation on recidivism among drug offenders*. *Criminology*, 48(2): 357–387, 2010. ISSN 1745-9125. doi: 10.1111/j.1745-9125.2010.00189.x. URL http://dx.doi.org/10.1111/j.1745-9125.2010.00189.x.

Han Hong and Jessie Li. The Numerical Delta Method and Bootstrap. *Working Paper*, 2016.

Joel L. Horowitz and Charles F. Manski. Identification and Robustness with Contaminated and Corrupted Data. *Econometrica*, 63(2):281–302, 1995.

Martin Huber. Sensitivity checks for the local average treatment effect. *Economics Letters*, 123 (2):220 – 223, 2014.

Martin Huber and Giovanni Mellace. Testing instrument validity for LATE identification based on inequality moment constraints. *Review of Economics and Statistics*, 97(2):398–411, October 2011.

Martin Huber and Giovanni Mellace. Relaxing monotonicity in the identification of local average treatment effects. *Discussion Paper 2012-12, Department of Economics, University of St. Gallen*, 2012.

Martin Huber and Giovanni Mellace. Sharp IV bounds on average treatment effects under endogeneity and noncompliance. 2017.

Martin Killias, Marcelo Aebi, and Denis Ribeaud. Does community service rehabilitate better than short-term imprisonment?: Results of a controlled experiment. *The Howard Journal of Criminal Justice*, 39(1):40–57, 2000. ISSN 1468-2311. doi: 10.1111/1468-2311.00152. URL http://dx.doi.org/10.1111/1468-2311.00152.

Toru Kitagawa. A test for instrument validity. *Econometrica*, 83(5):2043–2063, 2015.

Tobias Klein. Heterogeneous Treatment Effects: Instrumental Variables without Monotonicity? (2738), 2007.

Patrick Kline and Andres Santos. Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure. *Quantitative Economics*, 4(2):231–267, 2013.

Jeffrey R. Kling. Incarceration length, employment, and earnings. *American Economic Review*, 96(3):863–876, June 2006. doi: 10.1257/aer.96.3.863. URL http://www.aeaweb.org/articles?id=10.1257/aer.96.3.863.

Cecilia Machado, Azeem M. Shaikan, and Edward J. Vytlacilh. Instrumental variables, monotonicity restrictions, and the sign of the average treatment effect. *Working Paper*, 2017.

Nicole Maestas, Kathleen Mullen, and Alexander Strand. Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. *American Economic Review*, 103(5):1797–1829, 2013.

Charles F. Manski and John V. Pepper. Monotone Instrumental Variables: With an Application to the Returns to Schooling. *Econometrica*, 68(4):997–1010, 2000.

Matthew A. Masten and Alexandre Poirier. Inference on Breakdown Frontiers. *Working Paper: SSRN*, May 2017.

Ismael Mourifie and Yuanyuan Wan. Testing Local Average Treatment Effect Assumptions. *The Review of Economics and Statistics*, 99(2):305–313, 2017.

Michael Mueller-Smith. The criminal and labor market impacts of incarceration. *Unpublished*, 2015.

Alexander Shapiro. Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30(1):169–186, Dec 1991.

Dylan S. Small and Zhiqiang Tan. A Stochastic Monotonicity Assumption for the Instrumental Variables Method. *Working Paper, Department of Statistics, Wharton School, University of Pennsylvania*, 2007.

JÃűrg Stoye. Essays on partial identification and satistical decisions. *Ph.D. thesis, Northewstern University*, 2005.

JÃűrg Stoye. Partial identification of spread parameters. *Quantitative Economics*, 2010.

Literature Review

Monotonicity Assumption

- TESTING MONOTONICITY ASSUMPTION: Kitagawa (2015), Huber and Mellace (2011), Autor, Kostol, Mogstad, and Setzler (2017), Mourifie and Wan (2017), Machado, Shaikan, and Vytlacilh (2017)
- BOUNDS ON THE TREATMENT EFFECTS: Angrist, Imbens, and Rubin (1996), Balke and Pearl (1997), Klein (2007), Manski and Pepper (2000), Huber and Mellace (2012), Small and Tan (2007), Huber and Mellace (2017)
- REINTERPRETATION OF WALD-ESTIMAND: De Chaisemartin (2017)
- SENSITIVITY ANALYSIS: Huber (2014), Fiorini and Stevens (2016)

Breakdown Frontier

- Horowitz and Manski (1995); Kline and Santos (2013); Masten and Poirier (2017); Stoye (2005, 2010)

Back

Empirical Papers Suffering from a Potential Violation of Monotonicity

Dobbie, Goldin, and Yang (2018a, AER), Dobbie, Grönqvist, Niknami, Palme, and Priks (2018b), Autor et al. (2017), Bhuller, Dahl, Løken, and Mogstad (2016) Aizer and Doyle (2015, QJE), Mueller-Smith (2015) Dahl, Kostol, and Mogstad (2014, QJE), French and Song (2014, AEJ: Public Policy), Maestas, Mullen, and Strand (2013, AER), Black, Devereux, and Salvanes (2011, REStat), Green and Wink (2010), Kling (2006, AER), Duflo and Saez (2003, QJE), Killias, Aebi, and Ribeaud (2000), Angrist and Evans (1998, AER), Butcher and Case (1994, QJE), . . .

Back

Bounding Parameter Space

Sensitivity Parameter λ

- $\pi_{CO} - \pi_{DF} = \mathbb{P}(D = 1|Z = 1) - \mathbb{P}(D = 1|Z = 0)$
- $\overline{\pi}_{CO} = \min\{\mathbb{P}(D = 1|Z = 1), \mathbb{P}(D = 0|Z = 0)\}$
- $\underline{\pi}_{CO} = \max\{\sup_{B\in\mathscr{B}}\{\mathbb{P}(Y \in B, D = 1|Z = 1) - \mathbb{P}(Y \in B, D = 1|Z = 0)\}, \sup_{B\in\mathscr{B}}\{\mathbb{P}(Y \in B, D = 0|Z = 0) - \mathbb{P}(Y \in B, D = 0|Z = 1)\}\}.$

Lemma 1

Let $\underline{\lambda} = \frac{\underline{\pi}_{CO}}{\pi_{CO}-\pi_{DF}}$ and $\overline{\lambda} = \frac{\overline{\pi}_{CO}}{\pi_{CO}-\pi_{DF}}$. If Assumption 1 holds, $\underline{\lambda} \leq \lambda \leq \overline{\lambda}$. These bounds are sharp.

Sensitivity Parameter δ_q

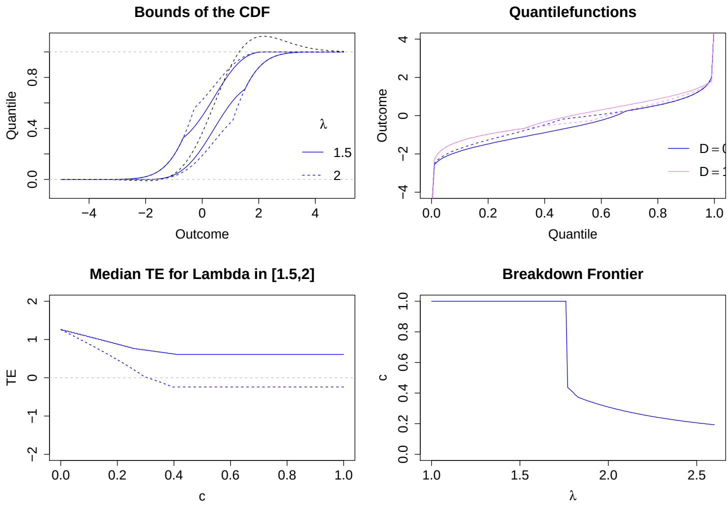
The domain of δ_q is bounded for a given value of λ by

$$\delta_q \geq \psi(\lambda)$$

$\psi(\lambda)$ is a known functional of the outcome distributions and densities

Back

Derivation of the Breakdown Frontier



If not otherwise stated: $c = 1$, $\tau = 0.5$, and $\delta_\tau \geq 0$

Illustration of the KS-Norm

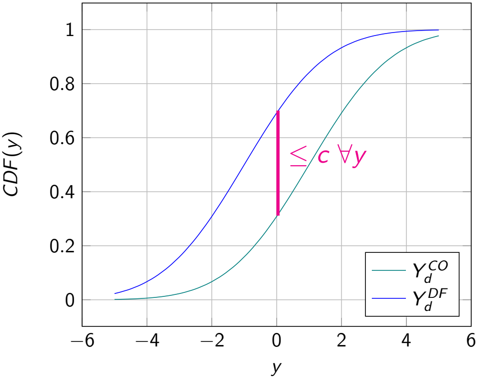


Figure: Visualization of the KS-Assumption

Back

Bounds of the Distributions - Definition

Let $G_d^{sup}(y) = \sup_{\tilde{y} \leq y} G_d(\tilde{y}),\ G_d^{inf}(y) = \inf_{\tilde{y} \geq y} G_d(\tilde{y}),\ G_d^{+}(y) = \int_{\underline{y}}^{y} \max\{0, \partial G_d(y) \backslash \partial y\}$

$$\overline{H}(y,t,c) = \frac{F_{Y_{dd}}(y)}{\pi_{CO}} - \max\{0, G_d^{sup}(y) + (1-\lambda)c, \frac{G_d^{sup}(y)}{\lambda}, \frac{F_{Y_{dd}}(y) - \pi_d}{\pi_{CO}}\}$$
$$+ \sup_{\tilde{y} \geq y} \left(\frac{G_d^{+}(\tilde{y})}{\lambda} - \inf_{\hat{y} \leq \tilde{y}} \left(\frac{G_d^{+}(\hat{y})}{\lambda} - \max\{0, G_d^{sup}(\hat{y}) + (1-\lambda)c, \frac{G_d^{sup}(\hat{y})}{\lambda}, \frac{F_{Y_{dd}}(\hat{y}) - \pi_d}{\pi_{CO}}\} \right) - \frac{F_{Y_{dd}}(\tilde{y})}{\pi_{CO}} \right)$$

$$\underline{H}(y,t,c) = \min\{1, G_d^{inf}(y) - (1-\lambda)c, \frac{G_d^{inf}(y) - (1-\lambda)}{\lambda}, \frac{F_{Y_{dd}}(y)}{\pi_{CO}}\} - \frac{G_d^{+}(y)}{\lambda}$$
$$- \inf_{\tilde{y} \geq y} \left(-\frac{G_d^{+}(\tilde{y})}{\lambda} + \frac{F_{Y_{dd}}(\tilde{y})}{\pi_{CO}} + \inf_{\hat{y} \leq \tilde{y}} \left(-\frac{F_{Y_{dd}}(\hat{y})}{\pi_{CO}} + \min\{1, G_d^{inf}(\hat{y}) - (1-\lambda)c, \frac{G_d^{inf}(\hat{y}) - (1-\lambda)}{\lambda}, \frac{F_{Y_{dd}}(\hat{y})}{\pi_{CO}}\} \right) \right)$$

Then

$$\overline{F}_{Y_d^{CO}}(y,t,c) = \max\{0, G_d^{sup}(y) + (1-\lambda)c, \frac{G_d^{sup}(y)}{\lambda}, \frac{F_{Y_{dd}}(y) - \pi_d}{\pi_{CO}}\} + \overline{H}(y,t,c)$$
$$\underline{F}_{Y_d^{CO}}(y,t,c) = \min\{1, G_d^{inf}(y) - (1-\lambda)c, \frac{G_d^{inf}(y) - (1-\lambda)}{\lambda}, \frac{F_{Y_{dd}}(y)}{\pi_{CO}}\} - \underline{H}(y,t,c).$$

back

Outcome Distributions of the Other Groups

$$\overline{F}_{Y_d^{DF}}(y,t,c) = \frac{1}{1-\lambda}\left(G_d(y) - \lambda \overline{F}_{Y_d^{CO}}(y,t,c)\right)$$

$$\overline{F}_{Y_1^{AT}}(y,t,c) = \frac{F_{Y_{11}}(y) - \pi_{CO}\underline{F}_{Y_1^{CO}}(y,t,c)}{\pi_{AT}}$$

$$\overline{F}_{Y_0^{NT}}(y,t,c) = \frac{F_{Y_{00}}(y) - \pi_{CO}\underline{F}_{Y_0^{CO}}(y,t,c)}{\pi_{NT}}$$

Lower bounds are obtained by replacing upper with lower bounds and vice versa.

back

Bounds of the Distributions - Graphical Illustration

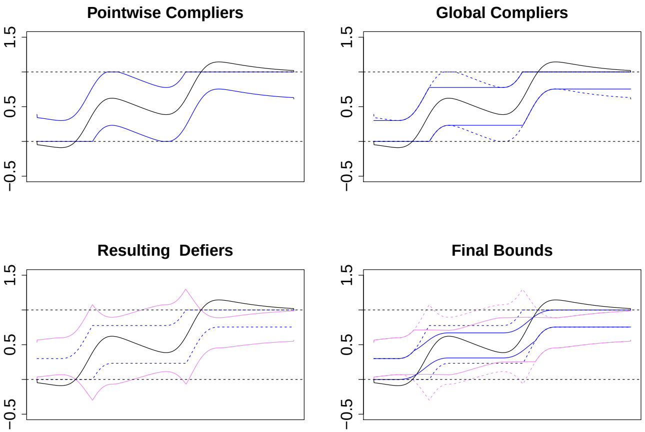


Figure: Derivation of the Bounds

Black solid line: Observed Function $G_d(y)$, *Blue lines:* Compliers, *Violet lines:* Defiers

Assumption 2 (Anderson et al., 2010)

1. $\{(Y_i, D_i, Z_i)\}_{i=1}^n$ are identically and independently distributed according to the distribution of (Y, D, Z).
2. The kernel is a 2nd order kernel function, being symmetric around zero, integrates to 1, two-times continuously differentiable, of bounded variation and zero-valued off $[-0.5, 0.5]$.
3. The bandwidth satisfies: (a) $nh^4 \to 0$, (b) $nh^2 \to \infty$, (c) $nh/\log(n) \to \infty$, (d) $nh^{(1+\gamma)/\gamma}/(\log(n)^{(1+2\gamma)/\gamma} \to \infty$, where $\gamma > 1$ is a constant satisfying (iv).
4. The densities $f_{YD|Z=z}(y_1|Z = z)$ and $f_{Y(1-D)|Z=z}(y_0|Z = z)$ are bounded, absolutely continuous and two times continuously differentiable with uniformly bounded derivatives for $z \in \{0, 1\}$.
5. $\mu(\{x \in [\underline{y}, y] : 0 < |f_1(x) - f_2(x)| < \varepsilon\}) = O_p(\varepsilon^\gamma)$ as $\varepsilon \to 0$ for some constant $\gamma > 1$ and uniformly over all y.
6. The outcome variable Y has compact support $\mathbb{Y}$.

Back

Hadamard Directional Differentiability

Definitions 3

Let $\mathbb{D}$ and $\mathbb{E}$ be Banach spaces with norms $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{E}}$. Let $\mathbb{D}_\phi \subseteq \mathbb{D}$ and $\mathbb{D}_0 \subseteq \mathbb{D}$. The map $\phi : \mathbb{D}_\phi \to \mathbb{E}$ is Hadamard directional differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_0$ if there is a continuous map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ such that

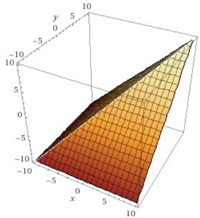
$$\lim_{n\to\infty} \| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h)\|_{\mathbb{E}} = 0$$

for all sequences $\{h_n\} \subset \mathbb{D}$ and $\{t_n\} \in \mathbb{R}_+$ such that $t_n \to 0$, $\|h_n - h\|_{\mathbb{D}} \to 0$, $h \in \mathbb{D}_0$ as $n \to \infty$ and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n.

If ϕ'_θ is a linear map, ϕ_θ is Hadamard differentiable.

Back

Hadamard Directional Differentiability (HDD)



Let $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ and $\phi(\theta) = \min\{\theta_1, \theta_2\}$. ϕ is not differentiable if $\theta_1 = \theta_2$.

ϕ is HDD, if its directional derivative at θ in direction $h = (h_1, h_2) \in \mathbb{R}^2$ exists and is continuous $\forall h_n \to h$ and $t_n \searrow 0$

$$\begin{aligned}\phi'_\theta(h) &= \lim_{n\to\infty} \frac{1}{t_n} \left(\phi(\theta + t_n h_n) - \phi(\theta)\right) \\ &= \mathbb{1}[\theta_1 < \theta_2]h_1 + \mathbb{1}[\theta_1 > \theta_2]h_2 + \mathbb{1}[\theta_1 = \theta_2]\min\{h_1, h_2\}\end{aligned}$$

Further Examples

- $\phi(\theta_1, \theta_2) = \theta_1 + \theta_2$
- $\phi(\theta_1, \theta_2) = \frac{\theta_1}{\theta_2}$ if $\theta_2 \neq 0$
- $\phi(\theta_1, \theta_2) = \max\{\theta_1, \theta_2\}$
- $\phi(\theta) = \inf_{\tau \in \mathcal{T}} \theta(\tau)$

Applications

- Chain rule (Masten and Poirier, 2017; Shapiro, 1991)
- Delta-method (Fang and Santos, 2016)

Back

Uniform Confidence Bands

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{t\in[0,1]} \sqrt{n} \left(\widehat{LB} - BP_{LQTE,CO}(t,\mu_\tau) \right) \leq 0 \right) \geq 1-\alpha.$$

Let us simplify the problem by only considering

$$\widehat{LB} = \widehat{BP}_{LQTE,CO}(\mu_t, t) - \hat{k}(t)$$

Given that $\hat{k}(t)$ satisfies the probability constraint, one can either assume

- *Constant band size*: $\hat{k}(t) = k$
- *Maximal size of robust region*: $\min \int_0^1 \hat{k}(t)dt$
- *Relative constant bandwidth*: $\min k$ such that $\hat{k}(t) = \frac{k}{\widehat{BF}(t,\mu_\tau)}$
- *Relative maximal robust region*: $\hat{k}(t) = \arg\min_{\tilde{k}(t)} \int_0^1 \frac{\tilde{k}(t)}{\widehat{BP}(t)} dt$
- *Maxmin*: $\hat{k}(t)$ such that $\hat{k}(t) = \arg\min_{\tilde{k}(t)} \min_t \widehat{BP}(t) - \tilde{k}(t)$

Back

Summary Statistics: Angrist and Evans (1998)

Data and Sample Restrictions

- Data set from 1980 restricted to women at the age of $20-36$ with at least two children
- Further restriction to women being white and having their first child at the age of $19-25$
- Outcome of interest: Wage

OLS and TSLS

Model	Estimate of D	Standard Deviation	Sample Size
OLS	-2892	47	211,983
TSLS	-1575	699	211,983
TSLS (A&E)	-1390	555	390,000