Linear programming approach to partially identified econometric models

Andrei Voronin

UCLA

December 16, 2024

The Question

Consider a linear program (LP):

$$B(\theta) \equiv \min_{Mx \geq c} p'x, \text{ where } \theta = (p, M, c) \in \mathbb{R}^d \times \mathbb{R}^{q \times d} \times \mathbb{R}^q$$

The value $\theta_0(\mathbb{P})$ is an identified feature of probability measure $\mathbb{P}$.

We are interested in $B(\mathbb{P}) = B(\theta_0(\mathbb{P}))$.

Key structure:

1. $B(\mathbb{P})$ is a measure-dependent linear program
2. All parameters p, M, c are to be estimated

Examples of LP estimation

Conditions in the AICM class result in LPs:

- Blundell et al. (2007), Gundersen et al. (2012), Siddique (2013), De Haan (2017), Cygan-Rehm et al. (2017), among others.

Example 1 (MIV in Manski and Pepper (2000))

$\mathbb{E}[Y(t)|Z=z]$ *is non-decreasing in* $z \in \mathcal{Z}$ *for each* $t \in \mathcal{T}$*.*

Example 2 (Roy model in Laffférs (2019))

For each $t \in \mathcal{T}$*, the individual's choice is, on average, optimal*
$\mathbb{E}[Y(t)|T=t, Z=z] = \max_{d \in \mathcal{T}} \mathbb{E}[Y(d)|T=t, Z=z]$*.*

LP often appears outside of AICM class:

- Mogstad et al. (2018), Syrgkanis et al. (2021), Andrews et al. (2023) among others, see Kline and Tamer (2023) for a review.

LP discontinuity

$$B(b) = \min_{x,y} x \quad \text{s.t.}: y \geq (1+b)x,\ y \leq x,\ x \in [-1;1],$$

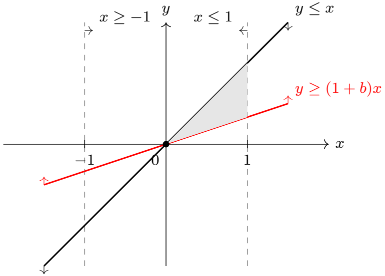


Figure: $b < 0,\ B(b) = 0$

LP discontinuity

$$B(b) = \min_{x,y} x \quad \text{s.t.}: y \geq (1+b)x,\ y \leq x,\ x \in [-1;1],$$

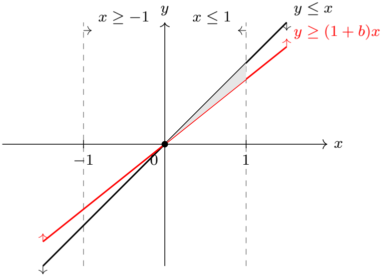


Figure: $b < 0,\ B(b) = 0$

LP discontinuity

$$B(b) = \min_{x,y} x \quad \text{s.t.}: y \geq (1+b)x,\ y \leq x,\ x \in [-1;1],$$

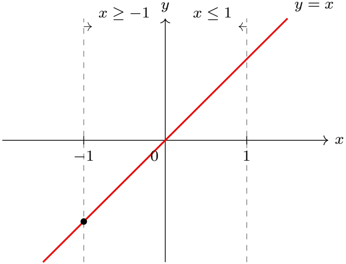


Figure: $b = 0,\ B(b) = -1$

Key point: $B(\cdot)$ is discontinuous, $B(b) = -\mathbb{1}\{b \geq 0\}$.

LP discontinuity

Suppose we estimate b as $b_n = n^{-1} \sum_{i=1}^{n} U_i$ with $U_i \sim U[-1+2b; 1]$ i.i.d.:

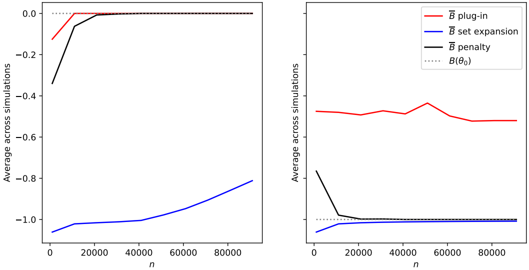


Figure: Comparison of estimators for two measures with $b = -0.02$ and $b = 0$, left to right. Average values over 400 simulations.

Aside: At $b = 0$ if intercept is noisy $B(b_n)$ does not exist w.p. $1/2$ $\forall n \in \mathbb{N}$

Contributions

Estimation

- Develop the first generally $\sqrt{n}$-consistent estimator of $B(\mathbb{P})$
- Develop exact, computationally efficient inference

Contributions

Estimation

- Develop the first generally $\sqrt{n}$-consistent estimator of $B(\mathbb{P})$
- Develop exact, computationally efficient inference
- Show that $\nexists$ uniformly consistent estimator absent further conditions
- Formulate the weakest condition ($\mathcal{P}^\delta$) for its existence
- Show uniformity of our estimator over $\mathcal{P}^\delta$

Contributions

Estimation

- Develop the first generally $\sqrt{n}$-consistent estimator of $B(\mathbb{P})$
- Develop exact, computationally efficient inference
- Show that $\nexists$ uniformly consistent estimator absent further conditions
- Formulate the weakest condition ($\mathcal{P}^\delta$) for its existence
- Show uniformity of our estimator over $\mathcal{P}^\delta$

Identification via LP (not in this talk)

- Provide a general identification result for 'AICM': LP sharp bounds

Contributions

Estimation

- Develop the first generally $\sqrt{n}$-consistent estimator of $B(\mathbb{P})$
- Develop exact, computationally efficient inference
- Show that $\nexists$ uniformly consistent estimator absent further conditions
- Formulate the weakest condition ($\mathcal{P}^{\delta}$) for its existence
- Show uniformity of our estimator over $\mathcal{P}^{\delta}$

Identification via LP (not in this talk)

- Provide a general identification result for 'AICM': LP sharp bounds

Application (not in this talk)

- Introduce a new condition (cMIV) that tightens classical bounds
- Develop a test for cMIV
- Apply results to estimating returns to education in Colombia
- cMIV yields a lower bound of 5.91% for the return to college education, classical conditions do not produce an informative bound

Problematic scenarios

Define $\Theta_I(\theta) \equiv \underbrace{\{x \in \mathbb{R}^d | Mx \geq c\}}_{\text{Identified set}}$ and $\mathcal{A}(\theta) \equiv \arg\min_{\Theta_I(\theta)} p'x$

Definition 1

Slater's condition (SC) asserts that $Relint(\Theta_I(\theta_0)) \neq \emptyset$.

Intuition: SC rules out point-identification along a perturbed direction

Problematic scenarios

Define $\Theta_I(\theta) \equiv \underbrace{\{x \in \mathbb{R}^d | Mx \geq c\}}_{\text{Identified set}}$ and $\mathcal{A}(\theta) \equiv \underset{\Theta_I(\theta)}{\arg\min}\, p'x$

Definition 1

Slater's condition (SC) asserts that $Relint(\Theta_I(\theta_0)) \neq \emptyset$.

Intuition: SC rules out point-identification along a perturbed direction

Definition 2

Linear independence constraint qualification (LICQ) asserts that the submatrix of binding constraints at any $x \in \mathcal{A}(\theta_0)$ is full-rank.

Intuition: LICQ rules out over-identification at an optimum

Problematic scenarios

Define $\Theta_I(\theta) \equiv \underbrace{\{x \in \mathbb{R}^d | Mx \geq c\}}_{\text{Identified set}}$ and $\mathcal{A}(\theta) \equiv \underset{\Theta_I(\theta)}{\arg\min}\, p'x$

Definition 1

Slater's condition (SC) asserts that $Relint(\Theta_I(\theta_0)) \neq \emptyset$.

Intuition: SC rules out point-identification along a perturbed direction

Definition 2

Linear independence constraint qualification (LICQ) asserts that the submatrix of binding constraints at any $x \in \mathcal{A}(\theta_0)$ is full-rank.

Intuition: LICQ rules out over-identification at an optimum

Definition 3

The notion of flat faces refers to the situation where $|\mathcal{A}(\theta_0)| \neq 1$.

Pointwise assumptions

Assumption (A0: Pointwise setup)

Suppose that at the fixed true parameter θ_0:

1. $\underbrace{\Theta_I(\theta_0) \neq \emptyset}_{\text{The model cannot be rejected}}$ *and* $\Theta_I(\theta_0) \subseteq \mathcal{X}$ *for a known compact* $\mathcal{X}$
2. *There is a $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ for θ_0*

Key: we do not assume SC, LICQ or no-flat-faces - unlike previous work.

Penalty-function estimator

Fix a $w \in \mathbb{R}^q_{++}$ and introduce the following:

$$L(x;\theta,w) \equiv p'x + \underbrace{w'(c - Mx)^+}_{\text{Penalty term}}$$

$$\tilde{B}(\theta;w) \equiv \min_{x\in\mathcal{X}} L(x;\theta,w), \quad \tilde{\mathcal{A}}(\theta;w) \equiv \arg\min_{x\in\mathcal{X}} L(x;\theta,w)$$

Lemma 1

If $\exists\lambda^*$ - KKT vector in the true LP such that $w > \lambda^*$, then:

1. *optimal values coincide:* $B(\theta_0) = \tilde{B}(\theta_0;w)$
2. *solutions coincide:* $\mathcal{A}(\theta_0) = \tilde{\mathcal{A}}(\theta_0;w)$

Proof

- In general, $\tilde{B}(\theta_0;w) \le B(\theta_0)$

Consistency of penalty-function estimator

Theorem 1

For any $w_n \to \infty$ *w.p.* 1 *as. and* $\frac{w_n}{\sqrt{n}} \xrightarrow{p} 0$, *we have:*

$$|\tilde{B}(\hat{\theta}_n; w_n \iota) - B(\theta_0)| = O_p\left(\frac{w_n}{\sqrt{n}}\right)$$

Comments:

- At a fixed measure eventually $w_n > \max_j \lambda_j^*$ for some λ^*
- Intuitively, $\frac{w_n}{\sqrt{n}}$ rate from $w_n \iota'(\hat{c}_n - \hat{M}_n x)^+ = O_p(\frac{w_n}{\sqrt{n}})$ for $x \in \Theta_I(\theta_0)$.
- We can do better by dropping that term.

$\sqrt{n}$-consistency of the debiased estimator

Theorem 2

Suppose $\mathcal{A}(\theta_0) \subseteq Int(\mathcal{X})$. For any $w_n \to \infty$ with $\frac{w_n}{\sqrt{n}} \xrightarrow{p} 0$:

$$\sup_{\tilde{\mathcal{A}}(\hat{\theta}_n; w_n)} |p'x - B(\theta_0)| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

Intuition:

1. The (biased) estimator selects a correct 'vertex' w.p. approaching 1
2. Once we get the 'vertex', can drop the penalty

A $\sqrt{n}-$consistent debiased estimator:

$$\hat{B}(\hat{\theta}_n; w_n) \equiv \sup_{\tilde{\mathcal{A}}(\hat{\theta}_n; w_n)} p'x$$

Sample splitting for asymptotic normality

Split the data $\mathcal{D}_n = \mathcal{D}_n^{(1)} \cup \mathcal{D}_n^{(2)}$ randomly, in proportion $\gamma \in (0;1)$

1. On $\mathcal{D}_n^1$, estimate $\hat{\theta}_n^{(1)}$, and:

$$\hat{x} \in \underset{\tilde{\mathcal{A}}(\hat{\theta}_n^{(1)};w_n)}{\arg\max}\ p'x, \quad \hat{A} \equiv \{j \in [q] : \hat{M}^{(1)\prime}{}_j \hat{x} = 0\}$$

$$\hat{v} \in \underset{v \in \mathbb{R}^{|\hat{A}|}:||v|| \leq \overline{v}}{\arg\min} ||p - \hat{M}^{(1)\prime}{}_{\hat{A}} v||^2$$

2. On $\mathcal{D}_n^{(2)}$, simply compute $\hat{\theta}_n^{(2)} = (\hat{M}_n^{(2)}, \hat{c}_n^{(2)})$

Exact inference

Theorem 3

Suppose $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$*, and we have an estimator* $\widehat{\Sigma}_n \xrightarrow{p} \Sigma < \infty$*. Under a non-degeneracy condition, for any* $w_n \to \infty$ *with* $w_n = o_p(\sqrt{n})$*, for any* $\alpha > 0$*:*

$$\mathbb{P}\left[\frac{\sqrt{n_2}}{\sigma(\hat{A}, \hat{v}, \hat{x}, \hat{\Sigma}_n)}\left(\hat{v}'({\hat{c}^{(2)}}_{\hat{A}} - {\hat{M}^{(2)}}_{\hat{A}}\hat{x}) + p'\hat{x} - B(\theta_0)\right) \leq z_{1-\alpha}\right] \to 1 - \alpha,$$

Comments:

- Closed-form for $\sigma(\cdot) \to$ no resampling needed
- If explicit Σ_n is not available, can bootstrap it from $\hat{\theta}_n$

Uniform asymptotic theory

Lemma 2

Suppose the estimand $V : (\mathcal{P}, ||\cdot||_{TV}) \to (\mathbb{R}, |\cdot|)$ is discontinuous at $\mathbb{P}_0 \in \mathcal{P}$. Then, there exists no uniformly consistent estimator $\hat{V}_n = \hat{V}_n(X)$, which is a sequence of measurable functions of the data $X \sim \mathbb{P}^n$. Moreover, if $\delta > 0$ is the jump at $\mathbb{P}_0$, then:

$$\inf_{\hat{V}_n} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[||V(\mathbb{P}) - \hat{V}_n(X(\mathbb{P}^n))||] \geq \frac{\delta}{2}, \quad \forall n \in \mathbb{N},$$

where infinum is taken over all measurable functions of the data.

Proof

- The Lemma is proven via Le Cam's binary method.

Negative result

Assumption (U0: Uniform setup)

The functional $\theta_0(\cdot)$ and the set of measures $\mathcal{P}$ are such that:

1. $\theta_0 : (\mathcal{P}, ||\cdot||_{TV}) \to (\mathbb{R}^S, ||\cdot||_2)$ *is a continuous functional*
2. $\theta_0(\mathcal{P}) = \{y \in \mathbb{R}^S$ ***s.t.*** $\Theta_I(y) \neq \emptyset, \Theta_I(y) \subseteq \mathcal{X}\}$

- We have seen that $B(\theta)$ is discontinuous
- So, under U0, $B \circ \theta_0$ is discontinuous

Theorem 4

Under U0, there exists no uniformly consistent estimator of $B(\mathbb{P})$.

- Is there a weak condition, under which it exists?

The δ-condition

Theorem 5

Under A0, $\exists x^* \in \mathcal{A}(\theta_0)$, the associated KKT vector λ^* and a subset of binding inequalities $J^* \subseteq \{1, \ldots, q\}$ with $|J^*| = rk(M_{J^*}) = d$, such that:

$$x^* = M_{J^*}^{-1} c_{J^*}$$

$$\lambda^*_{J^*} = M_{J^*}^{-1\prime} p$$

$$\lambda^*_i = 0, \; i \notin J^*$$

Assumption (U1: δ-condition)

For some $\delta > 0$, the collection $\mathcal{P}^\delta$ and the functional $\theta_0(\cdot)$ satisfy $\forall \mathbb{P} \in \mathcal{P}^\delta$:

$$\max_{J^*} \sigma_d(M_{J^*}(\theta_0(\mathbb{P}))) > \delta,$$

where J^* are defined above.

Geometry of δ-condition

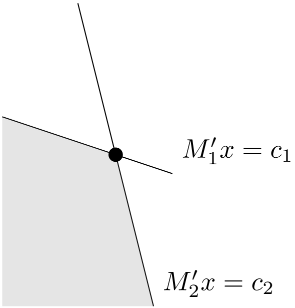


Figure: Optimal vertex $J = \{1, 2\}$

LICQ holds, $\delta-$condition holds with $\delta = \sigma_2(M_{\{1,2\}}) \gg 0$

Geometry of δ-condition

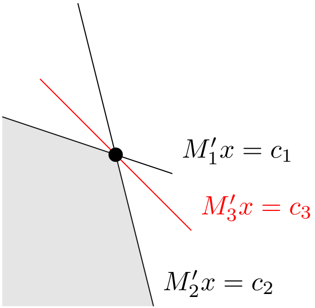
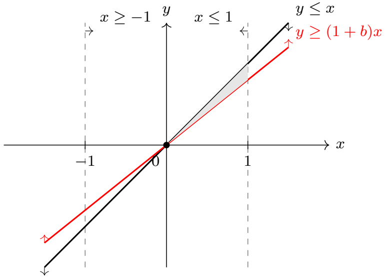


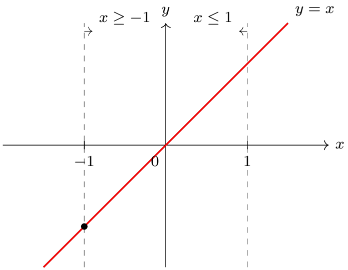
Figure: Optimal vertex $J = \{1, 2, 3\}$

LICQ fails, $\delta-$condition holds with $\delta = \sigma_2(M_{\{1,2\}}) \gg 0$

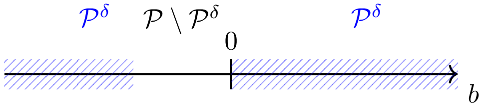
δ-condition in the baseline example



(a) $b \approx 0^- \Rightarrow \delta \approx -\frac{b}{2}$



(b) $b \geq 0 \Rightarrow \delta \gg 0$



(c) Set of b satisfying a δ-condition

Properties of the δ-condition

The usual uniform conditions are:

$$\mathcal{P}^{Slater;\varepsilon} \equiv \{\mathbb{P} \in \mathcal{P} | \text{Volume}(\Theta_I(\theta(\mathbb{P}))) > \varepsilon\}$$
$$\mathcal{P}^{LICQ;\varepsilon} \equiv \{\mathbb{P} \in \mathcal{P} | \mathcal{M}(v) \in \mathbb{R}^{d\times d}, \sigma_d(\mathcal{M}(v)) > \varepsilon \; \forall v \in \mathcal{V}(\mathbb{P})\},$$

$\mathcal{V}-$ all vertices of Θ_I, $\mathcal{M}(\cdot)-$ matrix of binding constraints

1. $\lim_{n\to\infty} \mathcal{P}^{Slater;1/n} \cup \mathcal{P}^{LICQ;1/n} \subset \mathcal{P} = \lim_{n\to\infty} \mathcal{P}^{1/n}$, the inclusion is strict
2. $\mathcal{P}^{LICQ;\varepsilon} \subset \mathcal{P}^{\delta}$ for any $\delta \leq \varepsilon$, the inclusion is strict
3. If M is normalized, $\forall \varepsilon > 0$, $\exists\, \delta$ s.t. $\mathcal{P}^{Slater;\varepsilon} \subset \mathcal{P}^{\delta}$, the inclusion is strict

$\tilde{B}_n$ is uniformly consistent over $\mathcal{P}^\delta$

Theorem 6

Suppose: i) $\exists\delta > 0$*:* $\mathcal{P}^* \subseteq \mathcal{P}^\delta$*, ii)* $\hat{\theta}_n(\cdot) \to \theta_0(\cdot)$ *at rate* $\sqrt{n}$ *uniformly. Setting* $w_n = ||\hat{p}_n||\delta^{-1} + \zeta$ *for any globally fixed* $\zeta > 0$ *yields,* $\forall\varepsilon > 0$ *and* $r_n \ll \sqrt{n}$*:*

$$\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathcal{P}^*} \mathbb{P}[\sup_{m\geq n} r_m |\tilde{B}(\hat{\theta}_m, w_m) - B(\theta_0(\mathbb{P}))| \geq \varepsilon] = 0. \tag{1}$$

Moreover, (1) *holds at rate* $\frac{\sqrt{n}}{w_n}$ *for any* $w_n \to \infty$ *with* $\frac{w_n}{\sqrt{n}} \to 0$*.*

Uniform consistency

Put differently, for any $w_n \to \infty$ with $\frac{w_n}{\sqrt{n}} \to 0$, for $\tilde{B}_n$ there is:

$\frac{\sqrt{n}}{w_n}$ uniform consistency under U1: $\sup_{\delta>0} \lim_{n\to\infty} \sup_{\mathbb{P}\in\mathcal{P}^\delta} \mathbb{P}[\dots] = 0$

No uniform consistency under U0: $\lim_{n\to\infty} \underbrace{\sup_{\delta>0} \sup_{\mathbb{P}\in\mathcal{P}^\delta}}_{\sup_{\mathbb{P}\in\mathcal{P}}} \mathbb{P}[\dots] \neq 0$

Comments:

- The debiased estimator converges at least $\frac{\sqrt{n}}{w_n}$ **uniformly** over $\mathcal{P}^\delta$ (*)
- $\hat{B}_n$ actual uniform rate appears to be $\sqrt{n}$, unless SC, LICQ, NFF all fail

Simulations

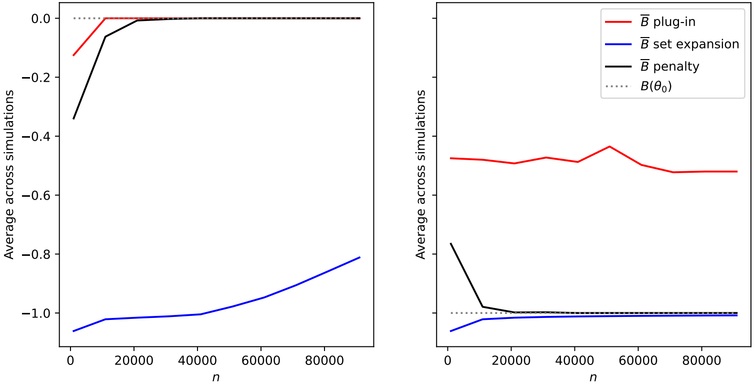
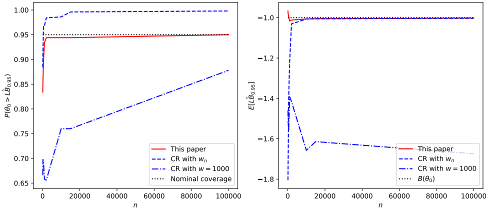


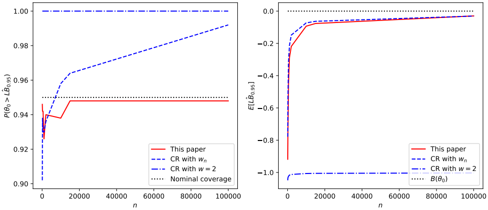
Figure: **Left:** $b = -0.02$ $(\alpha = 0.12)$ & SC holds; **Right**: $b = 0$ $(\alpha = 0.75)$ & SC fails.
Parameters: $N_{sim} = 400$, $w_n = \delta_{0.15}^{-1} \frac{\ln \ln n}{\ln \ln 100}$, $\sqrt{\kappa_n} = \ln \ln n$

$$\min_{x,y} x \quad \text{s.t.} : y \geq (1+b_n)x + \kappa_n,\ y \leq (1+\zeta_n)x + \zeta_n,\ x \in [-1-\kappa_n; 1+\kappa_n]$$

$$b_n = b + \overline{U^b}, \kappa_n = \overline{U^\kappa}, \zeta_n = \overline{U^\zeta} \text{ with } U_i^t \sim U[-0.5; 0.5] \text{ i.i.d. across } i, t$$



(a) $b = 0$



(b) $b = -0.1$ (angle 3°)

Q&A

Thank you for your attention!
avoronin@ucla.edu

I am grateful to Andres Santos, Denis Chetverikov, Bulat Gafarov, Zhipeng Liao, Jinyong Hahn, Rosa Matzkin, Tim Armstrong, Shuyang Sheng, Kirill Ponomarev, Manu Navjeevan and Daniel Ober-Reynolds as well as all the participants of the 2024 California Econometrics Conference for the valuable discussions and criticisms.
All mistakes are mine.

References I

Andrews, I., Roth, J., & Pakes, A. (2023). Inference for Linear Conditional Moment Inequalities. *The Review of Economic Studies*, *90*(6), 2763–2791. https://doi.org/10.1093/restud/rdad004

Bertsekas, D. P. (1975). Necessary and sufficient conditions for a penalty method to be exact. *Mathematical programming*, *9*(1), 87–99.

Blundell, R., Gosling, A., Ichimura, H., & Meghir, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, *75*(2), 323–363.

Chetverikov, D. (2019). Testing regression monotonicity in econometric models. *Econometric Theory*, *35*(4), 729–776. https://doi.org/10.1017/S0266466618000282

Cygan-Rehm, K., Kuehnle, D., & Oberfichtner, M. (2017). Bounding the causal effect of unemployment on mental health: Nonparametric evidence from four countries. *Health Economics*, *26*(12), 1844–1861. https://doi.org/https://doi.org/10.1002/hec.3510

De Haan, M. (2017). The effect of additional funds for low-ability pupils: A non-parametric bounds analysis. *The Economic Journal*, *127*(599), 177–198.

References II

Gundersen, C., Kreider, B., & Pepper, J. (2012).The impact of the national school lunch program on child health: A nonparametric bounds analysis [Annals Issue on "Identification and Decisions", in Honor of Chuck Manski's 60th Birthday]. *Journal of Econometrics*, *166*(1), 79–91. https://doi.org/https://doi.org/10.1016/j.jeconom.2011.06.007

Kline, B., & Tamer, E. (2023).Recent developments in partial identification. *Annual Review of Economics*, *15*(Volume 15, 2023), 125–150. https://doi.org/https://doi.org/10.1146/annurev-economics-051520-021124

Lafférs, L. (2019).Bounding average treatment effects using linear programming. *Empirical economics*, *57*, 727–767.

Manski, C. F., & Pepper, J. V. (2000).Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, *68*(4), 997–1010. Retrieved November 10, 2023, from http://www.jstor.org/stable/2999533

References III

Mogstad, M., Santos, A., & Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, *86*(5), 1589–1619.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton University Press. https://doi.org/doi:10.1515/9781400873173

Siddique, Z. (2013). Partially identified treatment effects under imperfect compliance: The case of domestic violence. *Journal of the American Statistical Association*, *108*(502), 504–513. https://doi.org/10.1080/01621459.2013.779836

Syrgkanis, V., Tamer, E., & Ziani, J. (2021). Inference on auctions with weak assumptions on information. *arXiv preprint arXiv:1710.03830*.

Tao, T., & Vu, V. (2010). Random matrices: The distribution of the smallest singular values. *Geometric And Functional Analysis*, *20*, 260–297.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint* (Vol. 48). Cambridge university press.

Identification result in this paper

Outcome: $Y \in \mathcal{Y} \subseteq \mathbb{R}$, treatment: $T \in \mathcal{T} \subseteq \mathbb{R}$, covariates: $Z \in \mathcal{Z} \subseteq \mathbb{R}^{d_Z}$

$\mathcal{T} = \mathcal{O} \sqcup \mathcal{U}$: if $T \in \mathcal{U}$, Y - unobserved. For the talk, $\mathcal{T} = \mathcal{O}$.

$$Y = \sum_{t \in \mathcal{T}} \mathbb{1}\{T = t\} Y(t)$$

Potential outcomes $\mathbb{Y} \equiv (Y(t))_{t \in \mathcal{T}} \in \mathbb{R}^{N_T}$

$\rightarrow$ conditional moments $m(P) \equiv (\mathbb{E}_P[\mathbb{Y}|T = d, Z = z])_{d \in \mathcal{T}, z \in \mathcal{Z}}$

Target: $\beta^*(\mathbb{P}) = \mu^*(\mathbb{P})' m(\mathbb{P})$ for identified μ^* (e.g. ATE)

Identification result in this paper

For identified matrices: $A^*, \tilde{A}$, vectors: $b^*, \tilde{b}$, the model is:

$$\mathcal{P}^* \equiv \{P \in \mathcal{P} | A^*(P)m(P) + b^*(P) \geq 0,\ \tilde{A}(P)\mathbb{Y} + \tilde{b}(P) \geq 0 \text{ } P\text{-a.s.}\}$$

Split $m(\cdot)$ into identified $\overline{x}$ and counterfactual moments x:

$$\overline{x} \equiv (\mathbb{E}[Y(t)|T = t, Z = z])_{z,t}, \quad x \equiv (\mathbb{E}[Y(t)|T = d, Z = z])_{z,\ t \neq d}$$

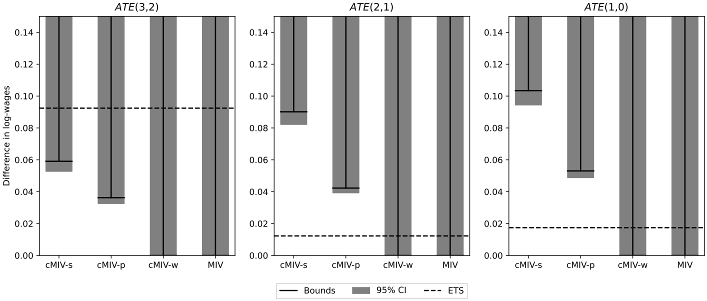
- $A^*, \tilde{A}$ and $F_{T,Z} \rightarrow$ identified M
- $b^*, \tilde{b}$ and $F_{T,Z} \rightarrow$ identified c
- μ^* and $F_{T,Z} \rightarrow$ identified p, $\overline{p}$

For any M^*, b^* and relevant $\tilde{M}, \tilde{b}$, sharp identified set for β^* is:

$$\mathcal{B}^* = \{\beta \in \mathbb{R} | \inf_{Mx \geq c} p'x \leq \beta - \overline{p}'\overline{x} \leq \sup_{Mx \geq c} p'x\}$$

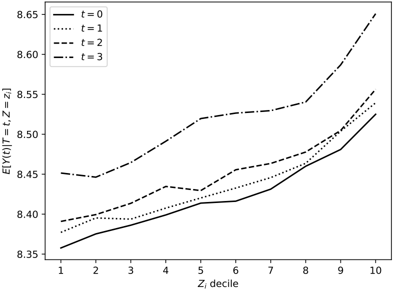
Returns to education in Colombia

- **Data**: 664633 observations from Colombian labor force
- **Variables**: Saber test results (Z), average wages (Y), schooling (S)
- Split Z into deciles
- Education levels T: primary, secondary, high school and university



- **AICM**: (c)MIV + bounded outcomes + MTR ($Y(t') \geq Y(t)$ if $t' > t$)
- **Result**: university education $\rightarrow$ average wage $\uparrow$ by $\geq 5.91\%$

Testing cMIV



(a) Estimated conditional moments

t	R_t^{st}	$R_{t;0.1}^{crit}$	p-value	n_t
0	0.98	2.33	0.34	274295
1	-1.17	2.17	0.95	143299
2	-1.51	2.30	1.00	216336
3	1.86	2.38	0.08	30703

(b) Results of the monotonicity test. Columns: 2. estimated Chetverikov (2019) test-statistic; 3. 10% critical values, corresponding to 2.6% individual critical value; 3. p-value against the individual null. The overall p-value is 29%.

Selecting a reasonable δ

Impossible to estimate, but can select a reasonable ”conservative” δ

Theorem 7 (Tao and Vu (2010))

Let Ξ_d be a sequence of $d \times d$ matrices with $[\Xi_d]_{ij} \sim \xi_{ij}$, independently across i, j where ξ_{ij} are such that $\mathbb{E}[\xi] = 0$, $Var(\xi) = 1$ and $\mathbb{E}[|\xi|^{C_0}] < \infty$ for some sufficiently large C_0, then:

$$\sqrt{d}\sigma_d(\Xi_d) \xrightarrow{d} \Pi \tag{2}$$

- The distribution of ξ_{ij} is any: possibly discrete, not identical.
- Normalize the matrix: $||\hat{M}_{.j}|| = 1$ for each row, or $\hat{M} \to \hat{M}/\hat{\sigma}(\hat{M})$
- Pick $\delta = \frac{\left(\sqrt{1-2\ln(1-\alpha)}-1\right)^2}{\sqrt{d}}$ - the $\alpha-$quantile of Π (we use $\alpha = 0.2$)
- Set $w_n = ||\hat{p}_n||\delta^{-1}\frac{\kappa_n}{\kappa_{100}}$ for some $\kappa_n \to \infty$, $\kappa_n = o(\sqrt{n})$.

Proof of Lemma 2

Proof.

Let $\delta > 0$ be a jump at $\mathbb{P}_0$. Construct a sequence $\{\mathbb{P}_n\} \subset \mathcal{P}$ such that for some $0 < \vartheta < 1$:

$$||\mathbb{P}_0 - \mathbb{P}_n||_{TV} < \vartheta n^{-1} \tag{3}$$

While $||V(\mathbb{P}_0) - V(\mathbb{P}_n)|| > \delta$. Recall that:

$$||\mathbb{P}_0^n - \mathbb{P}_n^n||_{TV} \leq n||\mathbb{P}_0 - \mathbb{P}_n||_{TV} \tag{4}$$

It follows that:

$$||\mathbb{P}_0^n - \mathbb{P}_n^n||_{TV} \leq \vartheta \tag{5}$$

Using the binary Le Cam's method[1], one obtains $\forall n$:

$$\inf_{\hat{V}_n} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[||V(\mathbb{P}) - \hat{V}_n(X(\mathbb{P}^n))||] \geq \frac{\delta(1-\vartheta)}{2} \tag{6}$$

Recalling that $0 < \vartheta < 1$ and δ were chosen arbitrarily and taking supremum over δ as well as sending $\vartheta \to 0$ yields the result. $\square$

Proof of Lemma 1.i

If w in the linear penalty function is component-wise larger than the KKT vector λ at a local minimum of the original problem, then this local minimum is also a local minimum of the penalized unconstrained problem (see Bertsekas (1975)). The claim then follows from the fact that any local minimum of a convex program is also global.

Proof of Lemma 1.ii

Suppose that $(\overline{\lambda}, w)$ are the KKT vector and the penalty vector that satisfy Assumption A0 and $\overline{x}$ is the associated optimum of the initial LP and $\overline{B} \equiv p'\overline{x}$. Note that one direction of ii) is trivial, since any $\tilde{x}$ that is optimal in the initial problem yields the same value in the penalized problem.

For another direction, suppose x^* is a local (global) minimum of the penalized problem. If x^* is feasible, it is also an optimum of the initial problem. Suppose it is not feasible. By the assumption on $(w, \overline{\lambda})$:

$$p'x^* + w'(c - M'x^*)^+ > p'x^* + \overline{\lambda}'(c - M'x^*) \tag{7}$$

The definition of a KKT vector in Rockafellar (1970) also requires that:

$$\overline{B} = \underset{x \in \mathbb{R}^{N(S-1)}}{inf} \; p'x + \overline{\lambda}'(c - M'x) \leq p'x^* + \overline{\lambda}'(c - M'x^*) \tag{8}$$

Therefore,

$$\overline{B} = p'x^* + w'(c - M'x^*) > p'x^* + \overline{\lambda}'(c - M'x^*) \geq \overline{B} \tag{9}$$

Which yields a contradiction, so there can be no such x^*. Thus, the sets of optimal solutions coincide. Return

Three forms of cMIV

Consider $Z \in \mathbb{R}$ and bounded outcomes $Y(t) \in [K_0, K_1]$ a.s.

Assumption (cMIV-s)

Suppose that for any $t \in \mathcal{T}$, $A \subseteq \mathcal{T}$ and $z, z' \in \mathcal{Z}$ s.t. $z' > z$ we have:

$$\mathbb{E}[Y(t)|T \in A, Z = z'] \geq \mathbb{E}[Y(t)|T \in A, Z = z] \tag{10}$$

Assumption (cMIV-w)

Suppose MIV holds and for any $t \in \mathcal{T}$ and $z, z' \in \mathcal{Z}$ s.t. $z' > z$ we have:

$$\Big\{\mathbb{E}[Y(t)|T \neq t, Z = z'] \geq \mathbb{E}[Y(t)|T \neq t, Z = z] \tag{11}$$

Assumption (cMIV-p)

Suppose MIV holds and for any $t \in \mathcal{T}, d \in \mathcal{T} \setminus \{t\}$ and $z, z' \in \mathcal{Z}$ s.t. $z' > z$ we have:

$$\mathbb{E}[Y(t)|T = d, Z = z] - \text{ monotone} \tag{12}$$

cMIV bounds are tighter than MIV

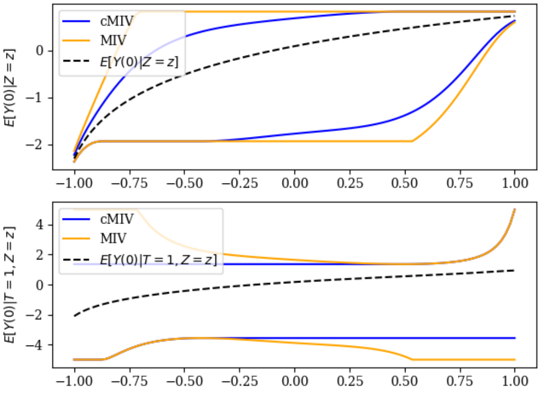


Figure: Sharp bounds for a DGP satisfying cMIV

Sharp bounds:

cMIV in words

Let $Y(t)$ be the individual's wage, $T \in \{0, 1\}$ - college degree, and Z - ability (e.g. IQ).

MIV assumption implies that:

- 'Smarter' individuals can do better both with and without a college degree on average: $\mathbb{E}[Y(t)|Z = z]$ - monotone

cMIV additionally assumes:

- Among those who have a college degree, a 'smarter' individual could have done relatively better than their counterpart if both did not have it: $\mathbb{E}[Y(0)|Z = z, T = 1]$ - monotone
- Among those who do not have a college degree, a 'smarter' individual could have done relatively better than their counterpart if both had it: $\mathbb{E}[Y(1)|Z = z, T = 0]$ - monotone

Example: education selection (1)

Suppose there is an innate '**effort**' level η s.t. $\eta \perp\!\!\!\perp Z$. Roy model:

$$Y(t) = \beta_0(t) + \beta_1(t)Z + \beta_2(t)\eta + \varepsilon(t) \quad (13)$$

$$T = \mathbb{1}\{\mathbb{E}[Y(1) - Y(0)|Z, \eta] + \nu \geq 0\} \quad (14)$$

where $\varepsilon(t) \perp\!\!\!\perp (Z, T, \eta)$ and $\nu \perp\!\!\!\perp (Z, \eta, \varepsilon(\cdot))$.

Let $\delta_z \equiv \beta_1(1) - \beta_1(0)$ and $\delta_\eta \equiv \beta_2(1) - \beta_2(0)$ - the differential effects of Z, η.

MIV:

$$\beta_1(t) \geq 0,\ t = 0, 1 \quad (15)$$

cMIV: MIV and

$$\underbrace{\beta_1(0)z}_{\text{direct effect}} + \underbrace{\beta_2(0)\mathbb{E}[\eta|\delta_z z + \delta_\eta \eta + \tilde{\nu} \geq 0]}_{\text{selection given } T=1} - \text{increasing} \quad (16)$$

$$\underbrace{\beta_1(1)z}_{\text{direct effect}} + \underbrace{\beta_2(1)\mathbb{E}[\eta|\delta_z z + \delta_\eta \eta + \tilde{\nu} \leq 0]}_{\text{selection given } T=0} - \text{increasing} \quad (17)$$

Example: education selection (2)

cMIV: (15) and

$$\underbrace{\beta_1(0)z}_{\text{direct effect}} + \underbrace{\beta_2(0)\mathbb{E}[\eta|\delta_z z + \delta_\eta \eta + \tilde{\nu} \geq 0]}_{\text{selection given } T=1} - \text{increasing} \tag{18}$$

$$\underbrace{\beta_1(1)z}_{\text{direct effect}} + \underbrace{\beta_2(1)\mathbb{E}[\eta|\delta_z z + \delta_\eta \eta + \tilde{\nu} \leq 0]}_{\text{selection given } T=0} - \text{increasing} \tag{19}$$

Suppose $\beta_1(t), \beta_2(t) \geq 0,\ t = 0, 1$

- δ_Z and δ_η have different signs $\rightarrow$ cMIV implied by MIV
- δ_Z and δ_η have the same sign $\rightarrow$ cMIV requires $\beta_1(t)$ to be larger

Takeaway:

- Z has to affect the potential outcomes **directly** and strongly enough
- In the presence of unobserved heterogeneity η with $sgn(\delta_\eta) = sgn(\delta_Z)$, Z's direct effect relative to its effect on **selection** must be greater than that for η

In other words, Z should be **relatively weak** and **strongly monotone**

Example: education selection (3)

1. Suppose education $\rightarrow$ jobs where ability Z gives a comparative advantage $\delta_Z > 0$, no education $\rightarrow$ jobs that are more effort-intensive $\delta_\eta < 0$.

   Positive conditional association b/w Z, η:

   - Given $T = 0$, $Z \underset{+}{\sim} \eta$ as else a higher Z-person would select into $T = 1$
   - Given $T = 1$, $Z \underset{+}{\sim} \eta$ as else a higher η-person would select into $T = 0$

2. Suppose to get a degree one needs to be either hardworking or of high ability:

$$T = \mathbb{1}\{\eta + Z \geq 0\}$$

   Negative conditional association b/w Z, η:

   - Given $T = 0$, $Z \underset{-}{\sim} \eta$ as a higher Z person would have gotten a degree, if not for lower effort
   - Given $T = 1$, $Z \underset{-}{\sim} \eta$ as at higher Z one does not need to be as hardworking to get a degree

Testing cMIV-p

$Y(t) = f(t, Z, T, \eta, \xi)$ where η is an unobs. r. vector, noise $\xi \perp\!\!\!\perp (T, Z, \eta)$

Homogeneity of $f(\cdot)$ + MIV $\rightarrow$ cMIV-p is **testable**:

Proposition 1

Suppose that a): i) $Y(t) = g(t, \xi) + h(t)\psi(Z, \eta)$, $h(t) \neq 0$ and ii) MIV, strictly for zome z, z'; or b): i) $Y(t) = g(t, \xi, T) + h(t)\psi(Z, \eta)$, ii) $\frac{h(t)}{h(d)} > 0 \; \forall t, d \in \mathcal{T}$ and iii) MIV. Then Assumption cMIV-p holds iff $\mathbb{E}[Y(t)|T = t, Z = z]$ are all monotone.

- MP (2009) discusses **HLR**: $Y(t) = \beta t + \eta$ under MIV $\implies$ a.i) or b.i)

Using regression monotonicity (Chetverikov, 2019), will test:

$$\mathcal{H}_0 : \mathbb{E}[Y(t)|T = t, Z = z] - \text{monotone in } z$$
$$\mathcal{H}_a : o/w$$

- If $\mathcal{H}_0$ is not rejected and we believe in homogeneity - can assume cMIV
- Applied work has inspected this monotonicity w/o theoretical justification

DGP for Figure 1

$$
\begin{aligned}
Y(t) &= c + \alpha t + \beta\eta + g(Z) \\
T &= \mathbb{1}\{\varepsilon + f(Z) \geq 0\} \\
\eta &= \min\{u, \max\{\varepsilon, l\}\} \\
\varepsilon &\sim \mathcal{N}(0,1)
\end{aligned}
$$

With:

$$
\begin{aligned}
t &= 0 \\
[l, u] &= [-4, 2] \\
z &\sim U[-1, 1] \\
f(z) &= -2z^4 \\
g(z) &= \ln(z + 1.1) \\
\beta &= 0.1
\end{aligned}
$$

Return

cMIV-p, cMIV-s sharp bounds

Suppose $\mathcal{Z} = \{z_1, z_2, \ldots, z_N\} \subset \mathbb{R}$, s.t. $z_i < z_j$ for $i < j$ and let $S \equiv N_T(N_T - 1)$ and $x^j \equiv (\mathbb{E}[Y(t)|T = d, Z = z_j])'_{d \neq t}$. Using Theorem 1:
Under cMIV-s and cMIV-p, sharp bounds on $\mathbb{E}[Y(t)]$ have the form:

$$\min_{Mx \geq c} \left\{ \sum_{j=1}^{N} P[Z = z_j] \cdot p^{j\prime} x^j \right\} + \sum_{j=1}^{N} P[T = t, Z = z_j] \mathbb{E}[Y(t)|T = t, Z = z_j]$$

$$\leq \mathbb{E}[Y(t)] \leq$$

$$\max_{Mx \geq c} \left\{ \sum_{j=1}^{N} P[Z = z_j] \cdot p^{j\prime} x^j \right\} + \sum_{j=1}^{N} P[T = t, Z = z_j] \mathbb{E}[Y(t)|T = t, Z = z_j]$$

Where

$$M \equiv \begin{bmatrix} -I_S & \ldots & 0 & 0 \\ G_N & -G_{N-1} & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & G_2 & -G_1 \\ 0 & \ldots & 0 & I_S \end{bmatrix}, \quad c \equiv \begin{pmatrix} -K_1 \cdot \iota_S \\ -\Delta c_N \\ \vdots \\ -\Delta c_2 \\ K_0 \cdot \iota_S \end{pmatrix}, \quad x = \begin{pmatrix} x^N \\ \vdots \\ x^1 \end{pmatrix}$$

G_j, c_j - cMIV-p G_j, c_j - cMIV-s   Return

G_j, c_j for cMIV-s

Let $\mathcal{F} \equiv 2^{\mathcal{T}} \setminus \{\{t\}, \emptyset\}$. Let $Q \equiv |\mathcal{F}| = 2^{N_T} - 2$. Fix the ordering of subsets of $\mathcal{F}$, so that $\mathcal{F} = \{A^1, A^2, \dots A^Q\}$.

$$\mathbb{E}[Y(t)|T \in A^k, Z = z_j] \geq \mathbb{E}[Y(t)|T \in A^k, Z = z_{j-1}],\ k = 1, \dots, Q,\ j = 2, \dots N_Z$$
$$\mathbb{E}[Y(t)|T = d, Z = z_N] \leq K_1, d \in \mathcal{T} \setminus \{t\}$$
$$\mathbb{E}[Y(t)|T = d, Z = z_1] \geq K_0, d \in \mathcal{T} \setminus \{t\}$$

The whole set of information given by cMIV-s can be represented as follows:

$$G_j x^j - G_{j-1} x^{j-1} \geq -\Delta c_j, j = 2, \dots, N_Z$$
$$x^N \leq K_1 \iota$$
$$x^1 \geq K_0 \iota$$

Where:

$$G_j \equiv \left(\mathbb{1}\left\{d \in A^k\right\} \frac{P[T = d|Z = z_j]}{P[T \in A^k|Z = z_j]} \right)_{k \in \overline{1,Q}, d \neq t} \in \mathbb{R}^{Q \times (N_T - 1)}$$

$$c_j \equiv \left(\mathbb{1}\left\{t \in A^k\right\} \frac{P[T = t|Z = z_j]}{P[T \in A^k|Z = z_j]} \mathbb{E}[Y(t)|T = t, Z = z_j] \right)_{k \in \overline{1,Q}} \in \mathbb{R}^{Q}$$

Return

G_j, c_j for cMIV-p

cMIV-p implies:

$$\mathbb{E}[Y(t)|Z = z_j] \geq \mathbb{E}[Y(t)|Z = z_{j-1}],\ j = 2, \ldots N_Z$$
$$\mathbb{E}[Y(t)|T = d, Z = z_j] \geq \mathbb{E}[Y(t)|T = d, Z = z_{j-1}],\ d \in \mathcal{T} \setminus \{t\},\ j = 2, \ldots N_Z$$
$$\mathbb{E}[Y(t)|T = d, Z = z_N] \leq K_1, d \in \mathcal{T} \setminus \{t\}$$
$$\mathbb{E}[Y(t)|T = d, Z = z_1] \geq K_0, d \in \mathcal{T} \setminus \{t\}$$

The whole set of information given by cMIV-s can be represented as follows:

$$\begin{aligned} &G_j x^j - G_{j-1} x^{j-1} \geq -\Delta c_j, j = 2, \ldots, N_Z \\ &x^N \leq K_1 \iota \\ &x^1 \geq K_0 \iota \end{aligned}$$

Recall that $p^j \equiv (P[T = d|Z = z_j])_{d \neq t}$ and we have:

$$G_j \equiv \begin{pmatrix} p^{j\prime} \\ I_{N_T - 1} \end{pmatrix} \in \mathbb{R}^{N_T \times (N_T - 1)}$$
$$c_j \equiv \begin{pmatrix} P[T = t|Z = z_j]\mathbb{E}[Y(t)|T = t, Z = z_j] \\ 0_{N_T - 1} \end{pmatrix} \in \mathbb{R}^{N_{T-1}}$$

Analytical sharp bounds under cMIV-w

Denote ℓ_j, ℓ_j^{-t} - s.l.b. for $\mathbb{E}[Y(t)|Z = z_j]$ and $\mathbb{E}[Y(t)|T \neq t, Z = z_j]$, then:
If i) cMIV-w holds or ii) $T \in \{0, 1\}$ and cMIV-s holds, then $\ell_1^{-t} = K_0$,
$\ell_1 = P[T = t|Z = z_1]\mathbb{E}[Y(t)|T = t, Z = z_1] + P[T \neq t|Z = z_1]K_0$ and for $j \geq 2$:

$$\Delta\ell_j = \left(\Delta P[T \neq t|Z = z_j]\ell_{j-1}^{-t} + \delta_j\right)^+ \quad (20)$$

$$\Delta\ell_j^{-t} = \frac{1}{P[T \neq t|Z = z_j]}\left(\Delta P[T \neq t|Z = z_j]\ell_{j-1}^{-t} + \delta_j\right)^- \quad (21)$$

Where:

$$\delta_j \equiv \Delta\left\{P[T = t|Z = z_j]\mathbb{E}[Y(t)|T = t, Z = z_j]\right\} \quad (22)$$

Sharp upper bounds u_i, u_i^{-t} are obtained analogously. Moreover,

$$\sum_{i=1}^{N} P[Z = z_i]\ell_i(t) \leq \mathbb{E}[Y(t)] \leq \sum_{i=1}^{N} P[Z = z_i]u_i(t) \quad (23)$$

In the absence of additional information, these bounds are sharp. Return

Simultaneous equations

$$q^k(p) = \alpha^k(p) + \beta^k(p)Z + \gamma^k(p)\eta + \kappa^k(p)\varepsilon^k,\ k \in \{s, d\}$$
$$P \in \{p \in \mathbb{R} | \mathbb{E}[q^s(p)|Z, \eta] = \mathbb{E}[q^d(p)|Z, \eta]\},$$

where η is unobserved with $\mathbb{E}[\eta|Z = z] = 0$, and $\mathbb{E}[\varepsilon^k] = 0$, $\varepsilon^k \perp\!\!\!\perp (\eta, Z, \varepsilon^{-k})$.

- All functions are continuous, support is full (for illustrative purposes)
- Define $\delta_z(p) \equiv \beta^s(p) - \beta^d(p)$ and $\delta_\eta(p)$, with $\delta_p(p) \equiv \alpha^s(p) - \alpha^d(p)$
- The model is *complete* and *coherent* iff:
  1. $\delta_p(p)$ is strictly increasing;
  2. $\delta_\eta(p)$ and $\delta_Z(p)$ are constant
- For concreteness, $\beta^s(p), \gamma^s(p) > 0$, and we want to estimate $\mathbb{E}[q^s(p)]$

$$(MIV) : \beta^s(p) \geq 0,\ \forall p$$

$$(cMIV) : (MIV) + \quad \left| \frac{\beta^s(p) - \beta^d(p)}{\beta^s(p)} \right| \leq \left| \frac{\gamma^s(p) - \gamma^d(p)}{\gamma^s(p)} \right| \vee sgn(\delta_\eta) \neq sgn(\delta_z)$$

Same idea: cMIV requires the instrument to be relatively weak and strongly monotone. Return