Sharp Analytical Instrumental Variable Bounds: Limitations & Methods

Sina Akbari

April 16, 2026

Statistical Laboratory
University of Cambridge
EuroCIM 2026, Oxford

Joint work with:

Arefe Boushehrian, Mohammad Reza Badri, Negar Kiyavash

EPFL

Homage

In Memory of Lives Lost–Iran Protests, 7–8 January 2026.

Outline

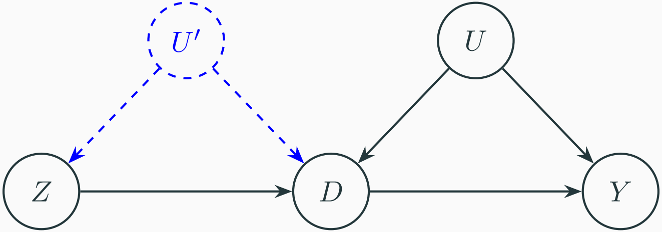
Overview

Setup (I)

Binary Treatment D

Discrete Outcome Y taking n values $\gamma_0, \ldots, \gamma_{n-1}$

Categorical Instrument Z taking values in $[\ell] = \{0, \ldots, \ell - 1\}$



Assumptions

Assumption 1 (Individual-level exclusion)

$Y^{(d,z)} = Y^{(d,z')}$ *almost surely for all* $z, z' \in [\ell]$ *and every* $d \in \{0,1\}$.

Assumption 2 (Random assignment)

The variables $D^{(z)}$ *for* $z \in [\ell]$ *exist, and,*

$$Z \perp\!\!\!\perp \left(Y^{(0)}, Y^{(1)}, D^{(0)}, D^{(1)}, \ldots, D^{(\ell-1)}\right).$$

Assumption 3 (Consistency)

$$Y = (1-D)Y^{(0)} + DY^{(1)}, \quad D = \sum_{z\in[\ell]} \mathbb{1}(Z=z)D^{(z)}.$$

Alternative Assumptions

Due to Song et al. [2024] (keynote by T.R. yesterday), alternatively:

Assumption (Individual-level exclusion)

$Y^{(d,z)} = Y^{(d,z')}$ *almost surely for all* $z, z' \in [\ell]$ *and every* $d \in \{0, 1\}$.

Assumption (Joint independence)

$$Z \perp\!\!\!\perp (Y^{(0)}, Y^{(1)}).$$

Assumption (Consistency)

$$Y = (1 - D)Y^{(0)} + DY^{(1)}.$$

- Full data law $\mathcal{Q}^f(Y^{(0)}, Y^{(1)}, D^{(0)}, \ldots, D^{(\ell-1)}, Z)$
- Marginal law $\mathcal{Q}(Y^{(0)}, Y^{(1)}, D^{(0)}, \ldots, D^{(\ell-1)})$, parameterized as

$$q_{ij,\mathbf{d}} := \mathcal{Q}(Y^{(0)} = \gamma_i, Y^{(1)} = \gamma_j, (D^{(0)}, \ldots, D^{(\ell-1)}) = \mathbf{d}),$$

- Observed law $\mathcal{P}(Y, D, Z)$, parameterized as

$$p_{id,z} := \mathcal{P}(Y = \gamma_i, D = d \mid Z = z).$$

Characterization

Proposition 1

Under Assumptions 1, 2 and 3, given the observed law $\mathcal{P}$, the marginal law $Q(\cdot)$ is sharply characterized as follows:

$$\begin{cases} p_{y0,z} = \sum_{j\in[n]} \sum_{\substack{\mathbf{d}\in\{0,1\}^\ell,\\ \mathbf{d}_z=0}} q_{yj,\mathbf{d}} & \forall y \in [n],\ z \in [\ell], \\ p_{y1,z} = \sum_{i\in[n]} \sum_{\substack{\mathbf{d}\in\{0,1\}^\ell,\\ \mathbf{d}_z=1}} q_{iy,\mathbf{d}} & \forall y \in [n],\ z \in [\ell], \\ q_{ij,\mathbf{d}} \geq 0 & \forall i,j \in [n],\ \mathbf{d} \in \{0,1\}^\ell, \end{cases} \tag{1}$$

where $\mathbf{d}_z$ is the z-th element of vector $\mathbf{d}$.

Matrix Representation

Equation (1) can be expressed in matrix form:

$$M^\top \mathbf{q} = \mathbf{p}, \qquad \mathbf{q} \geq 0,$$

where M is a binary matrix such that

$$M_{(ij,\mathbf{d}),(yd,z)} = 1$$

if $q_{ij,\mathbf{d}}$ appears in the equation corresponding to $p_{yd,z}$ in Equation (1), and is 0 otherwise.

Sharp ATE Bounds

ATE

$$\mathit{ATE}(\mathcal{Q}^f) := \mathbb{E}_{\mathcal{Q}^f}[Y^{(1)} - Y^{(0)}] = \sum_{i,j\in[n]} \sum_{\mathbf{d}\in\{0,1\}^\ell} (\gamma_j - \gamma_i)\, q_{ij,\mathbf{d}} = \mathbf{c}^\top \mathbf{q},$$

where the coefficient vector $\mathbf{c}$ is defined component-wise as

$$c_{ij,\mathbf{d}} = \gamma_j - \gamma_i,$$

LP Formulation

Sharp bounds on ATE:

$$L(\mathcal{P}) \leq ATE(\mathcal{Q}^f) \leq U(\mathcal{P}),$$

where

$$L(\mathcal{P}) = \min_{\mathbf{q}} \quad \mathbf{c}^\top \mathbf{q} \quad \text{s.t.} \quad M^\top \mathbf{q} = \mathbf{p}, \quad \mathbf{q} \geq 0 \tag{2}$$

and

$$U(\mathcal{P}) = \max_{\mathbf{q}} \quad \mathbf{c}^\top \mathbf{q} \quad \text{s.t.} \quad M^\top \mathbf{q} = \mathbf{p}, \quad \mathbf{q} \geq 0. \tag{3}$$

Strategy: Run 2 separate LPs to get sharp bounds.

What if we wanted closed-form symbolic bounds?

Dual LP

The dual LPs associated with Equations (2) and (3) are

$$\max_{\mathbf{v}\in\mathbb{R}^{2\ell n}} \quad \mathbf{v}^\top \mathbf{p} \quad \text{s.t.} \quad M\mathbf{v} \leq \mathbf{c}, \tag{4}$$

and

$$\min_{\mathbf{v}\in\mathbb{R}^{2\ell n}} \quad \mathbf{v}^\top \mathbf{p} \quad \text{s.t.} \quad M\mathbf{v} \geq \mathbf{c}, \tag{5}$$

respectively.

Strategy: Enumerate the extreme points of $M\mathbf{v} \leq \mathbf{c}$ and $M\mathbf{v} \geq \mathbf{c}$, evaluate the objective in the extreme points. First proposed by Balke and Pearl [1997].

Promising, But Is It Feasible?

n	# Extreme Points	Runtime
2	8	0.04s
3	52	0.15s
4	260	0.52s
5	1156	3.52s
6	4868	246s
7	???	> 1 day!

- ‘Computationally intensive and lacking intuitiveness’ [Evans, 2012]
- ‘now feasible in many more settings due to increased computational power’ [Sachs et al., 2026]
- Infeasible even for moderate support sizes [common wisdom by now?]

Are Existing Methods Inefficient?

Yes, and No.

Informal results:

- The number of terms in the ATE bounds grows exponentially.
- No shortcut methods! Any bounding technique claiming (or conjecturing) to be 'sharp' has to produce at least exponentially many terms. ***Pay the price for sharpness.***
- The computational complexity of off-the-shelf methods for extreme point enumeration grows SUPER-exponentially.

LP Structure

- Bad news: the polyhedral structure is general enough to result in exponentially many extreme points.
- Good news: the problem is structured enough to reveal the extreme points without necessitating numerical methods.

Proposition 2

For any observed data law $\mathcal{P}$ we have $U(\mathcal{P}) \equiv -L(\bar{\mathcal{P}})$, where the law $\bar{\mathcal{P}}$ is constructed based on $\mathcal{P}$ as follows:

$$\bar{\mathcal{P}}(Y = y, D = d, Z = z) = \mathcal{P}(Y = y, D = 1 - d, Z = z).$$

A 50% reduction already!

Further Structure (I)

Feasible set: $M^\top \mathbf{v} \leq \mathbf{c}$.

Definition 1 (Active Constraint Matrix)

For a point $\mathbf{v}$, denoted by $M_{\mathbf{v}}$, it is the submatrix of M and $\mathbf{c}_{\mathbf{v}}$ the corresponding subvector of $\mathbf{c}$ associated with active (tight) constraints at $\mathbf{v}$, i.e., $M_{\mathbf{v}}\mathbf{v} = \mathbf{c}_{\mathbf{v}}$.

Lemma 2

The rank of matrix M is $4n - 1$.

Lemma 3

Let $\mathbf{v}$ be a feasible point satisfying $M\mathbf{v} \leq \mathbf{c}$. Then $\mathbf{v}$ is a vertex if and only if

$$\mathit{rank}(M_{\mathbf{v}}) = 4n - 1.$$

Further Structure (II)

Definition 4 (Signature of a Point)

Denoted by $\mathbf{b}^{\mathbf{v}} \in \{0,1\}^{4n}$ for a point $\mathbf{v}$, it is the column-wise logical OR of the rows of $M_{\mathbf{v}}(**01 \cup **10, ***)$. That is, for each index (y, d, z), the entry $b^{\mathbf{v}}_{ydz}$ equals 1 if there exists at least one entry equal to 1 in the column $M_{\mathbf{v}}(**01 \cup **10, ydz)$, and equals 0 otherwise.

Proposition 3

The active constraint matrix $M_{\mathbf{v}}$ uniquely determines $\mathbf{b}^{\mathbf{v}}$, and conversely, $\mathbf{b}^{\mathbf{v}}$ uniquely determines $M_{\mathbf{v}}$.

Remark 1 (Informal)

$\mathbf{v}$ and $M_{\mathbf{v}}$ uniquely determine each other. Hence $\mathbf{v}$ uniquely determines $\mathbf{b}^{\mathbf{v}}$, and conversely $\mathbf{b}^{\mathbf{v}}$ uniquely determines $\mathbf{v}$.

Definition 5 (Admissible signature)

$S = S_1 \cup S_2 \cup S_3$, where $\mathbf{b} \in$

$$S_1 \text{ iff } \begin{cases} \exists\, t \in [n-1] \text{ such that } b_{i00} = b_{i01} = 1 \text{ for all } i \geq t \text{ and } b_{i00} \neq b_{i01} \text{ for all } i < t. \\ \forall\, i \in [n], b_{i10} \neq b_{i11}. \\ \exists\, i, j \in [n] \text{ such that } b_{i10} = b_{j11} = 1. \end{cases}$$

$$S_2 \text{ iff } \begin{cases} b_{(n-1)00} = b_{(n-1)01} = b_{010} = b_{011} = 1. \\ \forall\, 0 \leq i < n-1,\, b_{i00} \neq b_{i01}. \\ \forall\, 0 < j \leq n-1, b_{j10} \neq b_{j11}. \end{cases}$$

$$S_3 \text{ iff } \begin{cases} \exists\, t \in [n-1] \text{ s.t. } b_{i10} = b_{i11} = 1 \text{ for all } i \leq t+1 \text{ and } b_{i10} \neq b_{i11} \text{ for all } i > t+1. \\ \forall\, i \in [n], b_{i00} \neq b_{i01}. \\ \exists\, i, j \in [n] \text{ such that } b_{i00} = b_{j01} = 1. \end{cases}$$

Theorem 6

Every extreme point has an admissible signature.

Vertex Map

Definition 7 (Vertex map)

Given an admissible signature $\mathbf{b} \in S$, we define the vector $\mathbf{u}(\mathbf{b}) \in \mathbb{R}^{n \times 2 \times 2}$ as follows:

$$u_{i00} = -\gamma_i - \alpha \text{ if } b_{i00} = 1, \text{ and } u_{i00} = \gamma_0 \text{ otherwise;}$$

$$u_{i10} = \gamma_i \text{ if } b_{i10} = 1, \text{ and } u_{i10} = -\gamma_{n-1} - \alpha \text{ otherwise;}$$

$$u_{i01} = -\gamma_i \text{ if } b_{i01} = 1, \text{ and } u_{i01} = \gamma_0 + \alpha \text{ otherwise;}$$

$$u_{i11} = \gamma_i + \alpha \text{ if } b_{i11} = 1, \text{ and } u_{i11} = -\gamma_{n-1} \text{ otherwise,}$$

where

$$\alpha = \begin{cases} -\gamma_0 - \gamma_t & \text{if } \mathbf{b} \in S_1, \\ -\gamma_0 - \gamma_{n-1} & \text{if } \mathbf{b} \in S_2, \\ -\gamma_t - \gamma_{n-1} & \text{if } \mathbf{b} \in S_3. \end{cases}$$

Sharp ATE Bounds In A Nutshell

Theorem 8

The vertex map of every admissible signature is an extreme point.

Theorem 9

Under Assumptions 1, 2, and 3, the ATE admits the following valid and sharp bounds:

$$\max_{\mathbf{v}\in\mathcal{V}} \mathbf{v}^\top \mathbf{p} \;\leq\; ATE \;\leq\; -\max_{\mathbf{v}\in\mathcal{V}} \mathbf{v}^\top \bar{\mathbf{p}}, \tag{6}$$

where $\mathcal{V} = \{\mathbf{u}(\mathbf{b}) : \mathbf{b} \in S\}$.

Remarks

- $|S| = |\mathcal{V}| = 5 \times 4^{n-1} - 2^{n+2} + 4$, which is the number of terms in the upper/lower bound.
- Constructing the set S, and therefore the bounds takes time $O(|S|)$. This is optimal.

Multi-valued Instrument

Theorem 10 (Exponential lower bound)

Under Assumptions 1, 2 and 3, if a set of linear functionals of $\mathcal{P}$ is a sharp ATE bound, then it must contain at least

$$\ell\Big((\ell-1)^{n-1}-(\ell-1)\Big)$$

terms, where ℓ is the number of values the instrument can take.

Testable Implications / Falsification

Sharp Testable Implications (I)

$$\min_{\mathbf{q}} \quad 0 \quad \text{s.t.} \quad M^\top \mathbf{q} = \mathbf{p}, \quad \mathbf{q} \geq 0.$$

The IV model can be falsified if and only if the latter is not feasible. Equivalently,

$$\{\max_{\mathbf{r}\in\mathbb{R}^{2\ell n}} \quad \mathbf{p}^\top \mathbf{r} \quad \text{subject to} \quad M\mathbf{r} \leq 0\} \quad \leq \quad 0.$$

Sharp Testable Implications (II)

Define $\mathcal{K} := \{\mathbf{r} \in \mathbb{R}^{2\ell n} : M\mathbf{r} \leq 0\}$. The implications of the IV model boil down to

$$\mathbf{p}^\top \mathbf{r} \leq 0 \quad \text{for all } \mathbf{r} \in \mathcal{K}. \tag{7}$$

Since $\mathcal{K}$ is a polyhedral cone, it can be expressed as a cone combination of finitely many, e.g. ω, extreme rays (see, e.g., [Rockafellar, 1970]):

$$\mathcal{K} = \text{cone}(r_1, \ldots, r_\omega)$$

Equation (7) is therefore equivalent to the finite family of inequalities

$$\mathbf{p}^\top r_i \leq 0 \qquad \forall i \in \{1, \ldots, \omega\}.$$

Theorem 11

Let $\mathcal{P}$ be the observed probability distribution. The sharp testable implications of the IV model under Assumptions 1, 2, and 3 are the following inequalities:

$$\begin{cases} \sum_{k\in[n]} -p_{k0,1} + \sum_{k\in T} (p_{k1,0} - p_{k1,1}) \leq 0 & \forall T \subseteq [n-1], T \neq \varnothing \\ \sum_{k\in[n]} -p_{k1,0} + \sum_{k\in T} (p_{k0,1} - p_{k0,0}) \leq 0 & \forall T \subset [n], T \neq \varnothing \\ \sum_{k\in[n]} -p_{k0,0} + \sum_{k\in T} (p_{k1,1} - p_{k1,0}) \leq 0 & \forall T \subseteq [n-1], T \neq \varnothing \end{cases} \quad (8)$$

Corollary 12

The number of sharp IV inequalities in a binary instrument setting is $2^{n+1} - 4$.

Multi-valued Instrument

Theorem 13

Suppose Assumptions 1, 2 and 3 hold. For any $y' \in [n-1]$*,* $j' \in [\ell]$*, and non-constant* $(j_0, \ldots, j_{n-1}) \in ([\ell] \setminus \{j'\})^n$*, the following inequality holds:*

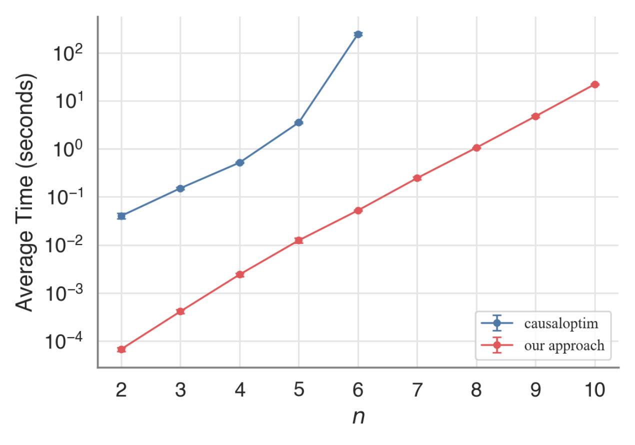
$$p_{y'1,j'} \leq \sum_{j \in \{j_0,\ldots,j_{n-1}\}} p_{y'1,j} + \sum_{y=0}^{n-1} p_{y0,j_y}. \tag{9}$$

Moreover, none of these inequalities are implied by the others (no redundancy), and their total number is

$$(n-1)\,\ell\,\big((\ell-1)^n - (\ell-1)\big).$$

n	Terms in ATE Bounds	IV Inequalities
2	8	4
3	52	12
4	260	28
5	1156	60
6	4868	124
7	19972	252
8	80900	508
9	325636	1020

Comparison in Action



causaloptim [Sachs et al., 2020]

Future Directions

Some of it coming soon!

- Multi-treatment
- Any linear functional of $\mathcal{Q}$?
- Complete set of bounds for $\ell > 2$

References

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92:1171–1176, 1997. URL https://api.semanticscholar.org/CorpusID:18365761.

Robin J Evans. Graphical methods for inequality constraints in marginalized dags. In *2012 IEEE International Workshop on Machine Learning for Signal Processing*, pages 1–6. IEEE, 2012.

Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970. ISBN 9781400873173. doi: **doi:10.1515/9781400873173**. URL https://doi.org/10.1515/9781400873173.

Michael C. Sachs, Gustav Jonzon, Arvid Sjölander, and Erin E. Gabriel. A general method for deriving tight symbolic bounds on causal effects. *Journal of Computational and Graphical Statistics*, 32: 567 – 576, 2020. URL https://api.semanticscholar.org/CorpusID:244955008.

Michael C Sachs, Erin E Gabriel, Robin J Evans, and Arvid Sjölander. Deriving complete constraints in hidden variable models. *arXiv preprint arXiv:2601.11242*, 2026.

Yilin Song, F Richard Guo, KC Chan, and Thomas S Richardson. The categorical instrumental variable model: Characterization, partial identification, and statistical inference. *arXiv preprint*

Wrap-up

Take away

- If you want sharp bounds/ sharp IV inequalities, you need to pay an exponential cost. NO SHORTCUTS.
- So far, people have been paying a super-exponential cost. That excess cost is unnecessary.
- If you don't have enough compute resources, if you are not comfortable with programming, LPs, etc., we have done it once and for all.