Fundamental Limits and Optimal Methods for Sharp Analytical Causal Bounds in Instrumental Variable Models

Arefe Boushehrian[*1] **Mohammad Reza Badri**[*1] **Sina Akbari**[2] **Negar Kiyavash**[1]

[1]College of Management of Technology, EPFL, Switzerland
[2]Statistical Laboratory, University of Cambridge, UK

Abstract

Bounding causal effects analytically, rather than numerically, is appealing for its interpretability and conceptual clarity. Existing sharp methods rely on optimization-based approaches such as the Balke–Pearl framework, whose computational complexity grows rapidly. An alternative line of work derives bounds heuristically using probability laws and generic inequalities, and some recent papers have claimed or conjectured that this approach can yield sharp analytical bounds with substantially lower complexity. In this paper, we show that this perceived advantage is illusory. In particular, in a discrete instrumental variable setting, we show that any sharp analytical bound for the average treatment effect must be expressible as a maximum (minimum) over a collection of linear terms whose cardinality grows exponentially in the number of values taken by the outcome. In parallel, we show that the number of instrumental variable inequalities itself also grows exponentially. Consequently, bounds and inequalities expressed using only polynomially many such terms cannot be sharp. As a constructive complement, the paper is accompanied by codes implemented in python and R to derive sharp analytical bounds and sharp inequalities with optimal computational complexity, matching the lower bounds proven in this paper. These codes are available online.

1 INTRODUCTION

Bounding causal effects under partial identification has a long history in statistics, econometrics, and causal inference. When point identification fails, sharp bounds provide the tightest possible characterization of causal estimands consistent with observed data and maintained assumptions. Seminal contributions by Manski [1990] and Manski [2003] formalized partial identification as a coherent framework, while subsequent work on instrumental variable (IV) models clarified how structural assumptions can result in narrower identification regions [Balke, 1995, Balke and Pearl, 1997, Robins, 1989, Pearl, 1995].

Among the available tools for bounding causal estimands, *analytical bounds*, closed-form expressions written directly as functionals of observed probabilities, are especially appealing due to their interpretability and transparency. However, the computational complexity of deriving sharp analytical bounds remains poorly understood.

Two broad approaches have been developed to derive analytical bounds in IV models. The first formulates the bounding problem as an optimization over the space of full data distributions consistent with the observed data and the structural assumptions. This approach originates in Balke [1995], and has been extended in multiple directions [Cheng and Small, 2006, Richardson and Robins, 2014, Sachs et al., 2023, Duarte et al., 2024]. These methods are guaranteed to yield sharp bounds, but their computational cost grows rapidly with the cardinalities of the observed variables, making them difficult to apply in more complex discrete settings.

The second approach derives inequalities algebraically from the observed data law using logical implications of the causal model, the laws of probability, and generic inequalities, e.g. Fréchet inequalities [Fréchet, 1935, Manski, 1990, Kédagni and Mourifié, 2020, Finkelstein and Shpitser, 2020]. While these bounds are not guaranteed to be sharp, they are often easier to interpret and compute. Because these approaches do not explicitly enumerate vertices of the full-data polytope, it has been conjectured, implicitly or explicitly [Bonet, 2001, Kédagni and Mourifié, 2020, Shu et al., 2025] that sharp analytical bounds and complete testable implications might be obtained through these approaches without incurring the limiting computational complexity of the optimization-based approaches.

---
*Authors contributed equally.

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:689–744.

This paper shows that such a computational separation is impossible. Focusing on the discrete instrumental variable (IV) model, we establish fundamental lower bounds on the complexity of *both* problems of deriving (i) sharp partial identification bounds for the average treatment effect (ATE) and (ii) sharp testable implications of the IV model. Here, *sharp* implies valid and tight: the bounds cannot be improved, and the implications are necessary and sufficient for compatibility with the IV model.

First, for partial identification, we prove that when the outcome variable Y takes n distinct values, any sharp upper (or lower) bound for the ATE must be representable as the minimum (or maximum) over a collection of linear expressions whose cardinality grows exponentially in n. Consequently, no polynomial-size family of linear analytical expressions can yield sharp bounds in general.

Second, we provide a complete and sharp characterization of the testable implications of the discrete IV model. We show that the set of observed data distributions compatible with the model is defined by a family of linear inequalities corresponding to the extreme rays of a dual cone. Crucially, we prove that the number of necessary and sufficient IV inequalities also grows exponentially in the support size of the outcome. Thus, any complete system of testable implications must be exponentially large in the worst case. This extends and sharpens earlier results on instrumental inequalities [Pearl, 1995, Bonet, 2001, Kédagni and Mourifié, 2020] by providing necessity, sufficiency, and explicit lower bounds on their number in multi-valued settings.

Both results show that the exponential growth observed in linear programming approaches, therefore, reflects an intrinsic combinatorial feature of the IV model rather than an artifact of a particular computational technique.

As a constructive complement to our theoretical results, we develop a software package that computes sharp upper and lower bounds on the ATE as well as sharp IV inequalities (testable implications) with *optimal output-sensitive computational complexity*, that is, time linear in the number of sharp terms (or inequalities). Rather than enumerating extreme points of the dual linear program, our implementation exploits the explicit structural characterizations derived in this paper to generate sharp bounds and IV inequalities directly. This avoids the combinatorial overhead inherent in off-the-shelf linear programming or vertex-enumeration routines. Simulation studies demonstrate substantial computational gains. While existing packages based on generic optimization or polyhedral enumeration become practically intractable even for moderate outcome support sizes, our method computes the full collection of sharp bounds and necessary and sufficient IV inequalities in milliseconds, even when the outcome support is large.

RELATED WORK

The identification of causal effects in IV models dates back to classical econometric analyses of simultaneous equations and compliance behavior. Under additional structural assumptions such as monotonicity, point identification of effects is possible [Imbens and Angrist, 1994, Angrist and Imbens, 1991, Chamberlain, 1986, Manski and Pepper, 2000]. Without such strengthening assumptions, causal effects are generally only partially identified.

Sharp bounds for IV models were first derived via linear programming in Balke [1995]. Extensions to multi-valued instruments appeared in Cheng and Small [2006] and Richardson and Robins [2014], building on earlier nonparametric bounds of Robins [1989] and Manski [1990]. Symbolic and algorithmic characterizations of causal bounds for a class of graphs more general than the standard IV model were developed in Sachs et al. [2023]. Our results complement this work by proving that the observed computational complexity is unavoidable for sharp bounds in the discrete IV setting, rather than an artifact of their approach. In addition, we characterize the solutions to these linear programming formulations of the problem, circumventing the need for the compute-intensive task of numerically enumerating the extreme points.

On the testability side, Pearl [1995] derived the original instrumental inequalities for the binary case, and Bonet [2001] strengthened these results. Kédagni and Mourifié [2020] provided necessary and sufficient generalized instrumental inequalities for the binary treatment and outcome setting under joint independence. Most recently, Song et al. [2024] characterized the restrictions on the joint distribution of potential outcomes in a categorical IV model.

Another related approach to deriving the testable implications is based on Artstein's inequalities [Artstein, 1983]. This perspective characterizes compatibility with a latent-variable model through the existence of a random selection from a set-valued map, and has been used to derive sharp instrumental variable inequalities [Beresteanu et al., 2012]. More recently, Luo et al. [2025] studied how to select non-redundant Artstein inequalities for sharp identification in models with set-valued predictions, recovering the restrictions characterized by Song et al. [2024] in the IV model.

These approaches are closely related to ours in that they characterize compatibility restrictions implied by IV assumptions. However, existing numerical and linear-programming-based approaches typically solve an instance-dependent optimization or feasibility problem for a given observed distribution, and therefore must be re-run when the observed data law changes. In contrast, our results provide explicit closed-form symbolic solutions for both sharp ATE bounds and falsification constraints in the discrete IV model. This provides two important advantages: (i) once derived, these solutions can be evaluated on any observed distribution without additional numerical optimization, and (ii) because of the explicit characterization, the bounds or the inequalities can be generated sequentially, avoiding the memory bottleneck in moderate to large support sizes.

Deriving symbolic bounds is also crucial for the task of inference, as shown by recent work [Ben-Michael, 2025]. Since existing symbolic approaches often derive such bounds by enumerating feasible solutions or vertices, which can become computationally intractable as the problem size grows, our closed-form characterizations can help tractable inference while avoiding unnecessary generic enumeration.

Finally, continuous IV models have been studied under structural or semiparametric assumptions [Imbens and Angrist, 1994, Kitagawa, 2021, Beresteanu et al., 2012]. However, without additional restrictions, continuous IV models impose no testable constraints on the observed distribution [Bonet, 2001, Gunsilius, 2021]. Our focus is therefore on the discrete setting, where rich geometric structure emerges.

Contributions. We make three main contributions:

1. **Sharp analytical ATE bounds.** We prove that any sharp representation of the ATE bounds in the discrete IV model must involve exponentially many linear terms in the outcome support size, showing that the computational burden is unavoidable when deriving sharp bounds. We provide an explicit and complete characterization of these bounds in the binary instrument setting.
2. **Sharp testable implications.** We establish exponential (in the outcome support size) lower bounds on the number of sharp testable implications (IV inequalities) in the discrete IV model. We provide these sharp inequalities explicitly in the binary instrument setting.
3. **Computationally optimal code.** We provide codes for computing the sharp bounds and inequalities with optimal time complexity in the binary instrument setting, which enables tractable computation of bounds and inequalities in discrete IV models.

2 PROBLEM SETUP

For $n \in \mathbb{N}$, we denote the set $\{0, \cdots, n-1\}$ by $[n]$. Throughout the paper, vectors are shown by bold letters (e.g., $\mathbf{p}$), random variables by capital letters (e.g., Y), and probability measures by calligraphic letters (e.g., $\mathcal{P}$).

We consider the instrumental variable model with categorical variables. We let D denote a dichotomous treatment variable. We assume without loss of generality that $D \in \{0, 1\}$. We denote the observed outcome variable and the instrument by Y and Z, respectively. Following Neyman-Rubin potential outcome model, let $Y^{(d,z)}$ denote the potential outcome of Y, if the treatment and the instrument (possibly contrary to the fact) were set to $D = d$ and $Z = z$, respectively. We assume these variables exist throughout. The variables $D^{(z)}$ are defined similarly, as the potential outcome of D, had (possibly contrary to the fact) the instrumental variable been set to $Z = z$. The existence of these variables, however, is only essential to some of our results.

We assume the outcome variable Y takes values in a finite real-valued set $\{\gamma_0, \gamma_1, \cdots, \gamma_{n-1}\}$, where $\gamma_0 < \gamma_1 < \cdots < \gamma_{n-1}$ without loss of generality. We further assume that the instrument Z takes values in $[\ell]$.

We require an exclusion restriction in our setting.

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

The individual-level exclusion restriction posits that there is no direct effect of Z on the outcome Y other than through the treatment of interest D. We maintain this assumption throughout the manuscript and consequently simplify the notation and use $Y^{(d)}$ to denote potential outcomes.

In addition to the exclusion restriction, we will work under an IV independence assumption. We consider the following two versions.

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 (Joint independence)**.**

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

Assumption 2 is stronger than Assumption 3, in the sense that the former implies the latter. In the remainder of this paper, we work under Assumption 2 for ease of exposition. However, we will show that most of our results are valid under Assumption 3. Finally, we make the standard consistency assumption:

Assumption 4 (Consistency)**.**

$$Y = (1 - D)Y^{(0)} + DY^{(1)}, \tag{4a}$$

$$D = \textstyle\sum_{z \in [\ell]} \mathbb{1}(Z = z) D^{(z)}, \tag{4b}$$

where $\mathbb{1}(\cdot)$ *is the indicator function.*

Let $\mathcal{Q}^f$ and $\mathcal{P}$ represent the full data and the observed data laws, respectively. For example, when $\ell = 2$ (binary instrument), $\mathcal{Q}^f = \mathcal{L}(Y^{(0)}, Y^{(1)}, D^{(0)}, D^{(1)}, Z)$, and $\mathcal{P} = \mathcal{L}(Y, D, Z)$. We denote by $\mathcal{Q}$ the marginal law of $(Y^{(0)}, Y^{(1)}, D^{(0)}, \ldots, D^{(\ell-1)})$, after marginalizing out Z, induced by $\mathcal{Q}^f$. Moreover, for every

$z \in [\ell]$, we define the stochastic kernel $\mathcal{P}_z$ as the regular conditional distribution[1] of (Y, D) given $Z = z$ under $\mathcal{P}$. That is, $\mathcal{P}_z := \mathcal{P}(Y, D \mid Z = z)$. Finally, we use $q_{ij,\mathbf{d}}$ and $p_{id,z}$ as shorthand for $\mathcal{Q}\left(Y^{(0)} = \gamma_i, Y^{(1)} = \gamma_j, (D^{(0)}, \ldots, D^{(\ell-1)}) = \mathbf{d}\right)$ and $\mathcal{P}_z(Y = \gamma_i, D = d)$, respectively.

In this work, we consider the following two problems.

Sharp Partial Identification Bounds for the Average Treatment Effect. We consider the problem of deriving sharp bounds for the average treatment effect, defined as

$$\begin{aligned}\operatorname{ATE}(\mathcal{Q}^f) &:= \mathbb{E}_{\mathcal{Q}^f}[Y^{(1)} - Y^{(0)}] \\ &= \textstyle\sum_{i,j\in[n]} \sum_{\mathbf{d}\in\{0,1\}^\ell} (\gamma_j - \gamma_i)\, q_{ij,\mathbf{d}}.\end{aligned} \tag{1}$$

Specifically, the objective is to derive functionals $U(\cdot)$ and $L(\cdot)$ of the observed data law $\mathcal{P}$ such that the following inequalities are valid and sharp

$$L(\mathcal{P}) \leq \operatorname{ATE}(\mathcal{Q}^f) \leq U(\mathcal{P}), \tag{2}$$

meaning that (i) they contain the true value of the estimand for every full data law that is compatible with the maintained assumptions and the observed distribution (valid); and, (ii) they are the tightest possible among all valid bounds (sharp). Specifically, for every observed law $\mathcal{P}$ conforming to the IV model, there exist full data distributions $\mathcal{Q}^f_L$ and $\mathcal{Q}^f_U$ satisfying the maintained assumptions and having the marginal $\mathcal{P}$ that attain these bounds, i.e.

$$\operatorname{ATE}(\mathcal{Q}^f_L) = L(\mathcal{P}), \qquad \operatorname{ATE}(\mathcal{Q}^f_U) = U(\mathcal{P}).$$

In general, one has to derive the functional $L(\cdot)$ and $U(\cdot)$ separately. However, in the following proposition we show that using the symmetry of the ATE functional, $U(\mathcal{P})$ can be expressed as $L(\cdot)$ evaluated at a different point (and vise versa).

Proposition 1. *For any observed data law $\mathcal{P}$ we have $U(\mathcal{P}) \equiv -L(\bar{\mathcal{P}})$, where the law $\bar{\mathcal{P}}$ can be constructed based on $\mathcal{P}$ as follows:*

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

Proof. Let $\bar{\mathcal{Q}}^f$ be defined as

$$\begin{aligned}&\bar{\mathcal{Q}}^f(Y^{(0)} = \gamma_i, Y^{(1)} = \gamma_j, (D^{(0)}, \ldots, D^{(\ell)}) = \mathbf{d}, Z = z) = \\ &\quad \mathcal{Q}^f(Y^{(0)} = \gamma_j, Y^{(1)} = \gamma_i, (D^{(0)}, \ldots, D^{(\ell)}) = \mathbf{d}, Z = z),\end{aligned}$$

which is the probability measure resulting from relabeling $Y^{(0)}$ and $Y^{(1)}$ by swapping them. By definition,

$$\begin{aligned}\operatorname{ATE}(\bar{\mathcal{Q}}^f) &:= \mathbb{E}_{\bar{\mathcal{Q}}^f}[Y^{(1)} - Y^{(0)}] \\ &= \mathbb{E}_{\mathcal{Q}^f}[Y^{(0)} - Y^{(1)}] = -\operatorname{ATE}(\mathcal{Q}^f).\end{aligned} \tag{3}$$

Let $\bar{\mathcal{P}}$ be the observed data law induced by $\bar{\mathcal{Q}}^f$ under consistency. Equation (3) implies that if $L(\bar{\mathcal{P}})$ is a (sharp) lower bound for $\operatorname{ATE}(\bar{\mathcal{Q}}^f)$, then $-L(\bar{\mathcal{P}})$ is a (sharp) upper bound for $\operatorname{ATE}(\mathcal{Q}^f)$; that is, $U(\mathcal{P}) \equiv -L(\bar{\mathcal{P}})$. The law $\bar{\mathcal{P}}$ can be easily constructed based on $\mathcal{P}$ as follows.

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

$\square$

Therefore, it is sufficient to characterize the lower bound functional $L(\cdot)$ for our purposes. We call $\bar{\mathcal{P}}$ the conjugate observed law, and similar to $\mathcal{P}$, we define conditionals $\bar{\mathcal{P}}_z$ and use $\bar{p}_{ij,\mathbf{d}}$ as shorthand for its conditionals.

Sharp Testable Implications. It is known that the instrumental variable model has testable implications (see, e.g. Pearl [1995], Kédagni and Mourifié [2020]). In particular, the instrumental variable model imposes restrictions on the observed data law $\mathcal{P}$, and if these restrictions are violated, the IV model is falsified. Here, we consider the problem of deriving the set of *valid* and *sharp* testable implications, in the sense that (i) if the true data-generating process follows the instrumental variable model, then $\mathcal{P}$ satisfies these implications, and (ii) if $\mathcal{P}$ satisfies this set of testable implications, then there exists a full data law $\mathcal{Q}^f$ that follows the IV model and induces (marginalizes to) $\mathcal{P}$. In other words, the IV model cannot be falsified.

3 LINEAR PROGRAMMING FORMULATION AND THE DUAL

In this section, we formalize the set of necessary and sufficient (linear) constraints that the full data law must satisfy under Assumptions 1, 2 and 4. Additionally, we review a unified linear programming approach to solving both classes of problems considered in this work.

Proposition 2. *Under Assumptions 1, 2 and 4, given the observed law $\mathcal{P}$, the marginal law of potential outcome variables $Q(\cdot)$ is sharply characterized as follows:*

$$\begin{cases} p_{y0,z} = \sum\limits_{j\in[n]} \sum\limits_{\substack{\mathbf{d}\in\{0,1\}^\ell,\\ d_z=0}} q_{yj,\mathbf{d}} & \forall y \in [n],\ z \in [\ell], \\ p_{y1,z} = \sum\limits_{i\in[n]} \sum\limits_{\substack{\mathbf{d}\in\{0,1\}^\ell,\\ d_z=1}} q_{iy,\mathbf{d}} & \forall y \in [n],\ z \in [\ell], \\ q_{ij,\mathbf{d}} \geq 0 & \forall i,j \in [n],\ \mathbf{d} \in \{0,1\}^\ell, \end{cases} \tag{4}$$

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

The proof of Proposition 2 is given in Appendix A.1. Proposition 2 states that for any IV model, these equations hold (necessity), and for every pair $\mathcal{P}, \mathcal{Q}$ that satisfy these equations, there exists a full data law $\mathcal{Q}^{f*}$ conforming to the IV model that marginalizes to $\mathcal{Q}$ and $\mathcal{P}$ (sufficiency).

[1] We assume without loss of generality that $\mathcal{P}(Z) > 0$. Otherwise, one can discard the values z with zero probability and rearrange.

Let $\mathbf{p}, \bar{\mathbf{p}} \in \mathbb{R}^{2\ell n}$ and $\mathbf{q} \in \mathbb{R}^{2^\ell n^2}$ be vector representations of $p_{yd,z}$ and $\bar{p}_{yd,z}$ for all y, d, z and $q_{ij,\mathbf{d}}$ for all $i, j, \mathbf{d}$, respectively. Equation (4) can be expressed in matrix form:

$$M^\top \mathbf{q} = \mathbf{p}, \qquad \mathbf{q} \geq 0, \tag{5}$$

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 (4), and is 0 otherwise. In Appendix, we provide an example where $n = \ell = 2$ and the corresponding matrix M is given (see Section E.1).

3.1 LINEAR PROGRAMMING AND THE DUAL PROBLEM

Based on Proposition 2 and its matrix representation (Equation 5), the set of all full data laws consistent with the observed distribution $\mathbf{p}$ is given by the convex polytope

$$\Gamma(\mathbf{p}) := \left\{ \mathbf{q} \in \mathbb{R}^{2^\ell n^2} : M^\top \mathbf{q} = \mathbf{p},\ \mathbf{q} \geq 0 \right\}. \tag{6}$$

Below, we review how the problems we consider in this paper are cast into linear programming (LP) problems using the latter observation.

Partial Identification for ATE. The ATE is a linear functional of the full data law. In particular,

$$\mathrm{ATE}(\mathcal{Q}^f) = \mathbf{c}^\top \mathbf{q},$$

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

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

for all $i, j \in [n]$ and all $\mathbf{d} \in \{0, 1\}^\ell$ (see Equation 1).

Since the bounds are obtained as global optima, the sharp identification region for the ATE is given by solutions to the linear programs

$$L(\mathcal{P}) = \min_{\mathbf{q}} \quad \mathbf{c}^\top \mathbf{q} \quad \text{s.t.} \quad M^\top \mathbf{q} = \mathbf{p}, \quad \mathbf{q} \geq 0 \tag{7}$$

and

$$\begin{aligned} U(\mathcal{P}) &= \max_{\mathbf{q}} \quad \mathbf{c}^\top \mathbf{q} \quad \text{s.t.} \quad M^\top \mathbf{q} = \mathbf{p}, \quad \mathbf{q} \geq 0 \\ &= -L(\bar{\mathcal{P}}) \\ &= -\min_{\mathbf{q}} \quad \mathbf{c}^\top \mathbf{q} \quad \text{s.t.} \quad M^\top \mathbf{q} = \bar{\mathbf{p}}, \quad \mathbf{q} \geq 0, \end{aligned} \tag{8}$$

where U, L are defined in Equation (2). Since $\Gamma(\mathbf{p})$ (the feasibility region) is a convex polytope, the resulting interval

$$[L(\mathcal{P}), U(\mathcal{P})] = [L(\mathcal{P}), -L(\bar{\mathcal{P}})]$$

is the sharp identification region for the ATE.

Testable Implications of IV Model. To test whether a given observed probability vector $\mathbf{p}$ is compatible with the IV model, we seek necessary and sufficient conditions on $\mathbf{p}$ such that existence of a vector $\mathbf{q}$ satisfying the linear system in Equation (5) is guaranteed. In other words, we want to determine whether the set $\Gamma(\mathbf{p})$ is nonempty. This can be cast as deciding the feasibility of the following linear program:

$$\min_{\mathbf{q}} \quad 0 \quad \text{s.t.} \quad M^\top \mathbf{q} = \mathbf{p}, \quad \mathbf{q} \geq 0. \tag{9}$$

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

Dual Formulation. The dual formulation provides an alternative characterization of both sharp identification regions and testable implications. By the strong duality theorem of linear programming, whenever the feasible set $\Gamma(\mathbf{p})$ is nonempty, the optimal values of the primal and dual problems coincide (see, e.g., [Boyd and Vandenberghe, 2004, Section 5.2.3]). The dual has two desirable advantages: (i) solving the dual can be computationally more efficient as it is defined over parameters corresponding to the observed data with dimension $2\ell n$, whereas the primal is defined over parameters corresponding to the full data law with dimension $2^\ell n^2$; (ii) more importantly, the dual admits a form where the constraints do not depend on the observed vector $\mathbf{p}$. As we shall see in Section 3.2, the latter is essential for deriving analytical (closed-form) solutions to the LPs.

The dual LPs associated with Equations (7) and (8) are

$$\max_{\mathbf{v} \in \mathbb{R}^{2\ell n}} \quad \mathbf{v}^\top \mathbf{p} \quad \text{s.t.} \quad M\mathbf{v} \leq \mathbf{c}, \tag{10}$$

and

$$-\max_{\mathbf{v} \in \mathbb{R}^{2\ell n}} \quad \mathbf{v}^\top \bar{\mathbf{p}} \quad \text{s.t.} \quad M\mathbf{v} \leq \mathbf{c}, \tag{11}$$

respectively, where $\mathbf{v}$ is the vector of dual variables corresponding to the equality constraints of primal LPs. Every dual-feasible vector $\mathbf{v}$ in Equation (10) or (11) generates a valid bound which holds uniformly over all full data laws consistent with the observed distribution. The optimal dual solution yields the tightest such bound.

In the testable implications setting, a fundamental result from linear programming (a theorem of the alternative; see, e.g., [Boyd and Vandenberghe, 2004, Section 5.8]) implies that the primal problem of Equation (9) is feasible if and only if

$$\{\max_{\mathbf{r} \in \mathbb{R}^{2\ell n}} \quad \mathbf{p}^\top \mathbf{r} \quad \text{subject to} \quad M\mathbf{r} \leq 0\} \quad \leq \quad 0. \tag{12}$$

In other words, each vector $\mathbf{r}$ satisfying $M\mathbf{r} \leq 0$ imposes the linear inequality $\mathbf{p}^\top \mathbf{r} \leq 0$ that must be satisfied by the observed distribution $\mathbf{p}$. These inequalities are necessary and sufficient for the existence of a full data law $\mathbf{q} \geq 0$ satisfying $M^\top \mathbf{q} = \mathbf{p}$.

3.2 DUAL SOLUTION VIA EXTREME POINTS AND RAYS

In this section, we discuss how to derive closed-form solutions of the dual LPs. We denote the feasible regions of Equations (10) and (11) by

$$\mathcal{H} := \{\mathbf{v} \in \mathbb{R}^{2\ell n} : M\mathbf{v} \leq \mathbf{c}\},$$

and the corresponding feasible region of Equation (12) by

$$\mathcal{K} := \{\mathbf{r} \in \mathbb{R}^{2\ell n} : M\mathbf{r} \leq 0\}.$$

Note that these sets do not depend on the observed data law; they remain invariant across every instance of the IV model, and so are their extreme points. Closed-form solutions can therefore be obtained through finding the extreme points and rays of these feasible regions. We provide the necessary formalism below.

Both $\mathcal{H}$ and $\mathcal{K}$ are polyhedra. However, since M might not have full column rank, these polyhedra contain affine directions given by the kernel of M, defined as

$$\ker(M) := \{x : Mx = 0\}.$$

Consequently, feasible points are not isolated in $\mathbb{R}^{2\ell n}$: if $\mathbf{v}$ is feasible, then so is $\mathbf{v} + \mathbf{s}$ for any $\mathbf{s} \in \ker(M)$. To eliminate this intrinsic degeneracy, we work modulo $\ker(M)$.

Definition 1 (M-equivalent and M-distinct)**.** *Two vectors $\mathbf{v}_1, \mathbf{v}_2$ are M-equivalent if*

$$\mathbf{v}_1 - \mathbf{v}_2 \in \ker(M),$$

or equivalently, $M\mathbf{v}_1 = M\mathbf{v}_2$. Two vectors are called M-distinct if they are not M-equivalent.

Working in the quotient space $\mathbb{R}^{2\ell n} \setminus \ker(M)$ removes affine directions and restores a proper notion of extremality.

Definition 2 (Vertex)**.** *A vector $\mathbf{v} \in \mathcal{H}$ is a vertex of $\mathcal{H}$ if for any representation*

$$\mathbf{v} = \lambda\mathbf{v}_1 + (1-\lambda)\mathbf{v}_2, \qquad \lambda \in (0,1), \quad \mathbf{v}_1, \mathbf{v}_2 \in \mathcal{H},$$

the vectors $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2$ are M-equivalent.

A vertex is a point in $\mathcal{H}$ that cannot be written as a nontrivial convex combination of two M-distinct feasible points. The dual optimum will be attained at a vertex of $\mathcal{H}$ (see Appendix A.2 for a more detailed discussion). Therefore, sharp identification bounds can be obtained by evaluating $\mathbf{p}^\top\mathbf{v}$ and $\bar{\mathbf{p}}^\top\mathbf{v}$ over the (finite) set of vertices.

The set $\mathcal{K}$ is a polyhedral cone. The feasibility of the primal problem is equivalent to

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

Hence, the extreme rays of $\mathcal{K}$ determine all necessary and sufficient IV inequalities.

Definition 3 (Extreme ray)**.** *A nonzero vector $\mathbf{r} \in \mathcal{K}$ is an extreme ray of $\mathcal{K}$ if for any decomposition*

$$\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2, \quad \mathbf{r}_1, \mathbf{r}_2 \in \mathcal{K},$$

there exist $\alpha, \beta \geq 0$ such that

$$M\mathbf{r}_1 = \alpha M\mathbf{r}, \quad M\mathbf{r}_2 = \beta M\mathbf{r}$$

or equivalently,

$$\mathbf{r}_1 = \alpha\mathbf{r} + \mathbf{s}_1, \qquad \mathbf{r}_2 = \beta\mathbf{r} + \mathbf{s}_2$$

for some $\mathbf{s}_1, \mathbf{s}_2 \in \ker(M)$.

Each extreme ray $\mathbf{r}$ generates the affine set

$$\mathcal{E}_\mathbf{r} = \{\alpha\mathbf{r} + \mathbf{s} : \alpha > 0,\ \mathbf{s} \in \ker(M)\}.$$

Since $\mathcal{K}$ is a polyhedral cone, it admits only finitely many M-distinct extreme rays [Rockafellar, 1970] (see Appendix A.3 for a more detailed discussion). Consequently, $\mathcal{K}$ can be written as the conic hull of finitely many extreme rays. It follows that the infinite family of inequalities $\mathbf{p}^\top\mathbf{r} \leq 0$ for all $\mathbf{r} \in \mathcal{K}$ reduces to a finite family indexed by the extreme rays.

The above discussion yields a unifying principle:

- Sharp identification bounds are obtained by enumerating the vertices of $\mathcal{H}$.
- Sharp testable implications are obtained by enumerating the extreme rays of $\mathcal{K}$.

We shall explicitly characterize the vertices of $\mathcal{H}$ and the extreme rays of $\mathcal{K}$ in Sections 4 and 5, respectively, and present closed-form sharp bounds and IV inequalities accordingly.

4 PARTIAL IDENTIFICATION OF ATE

We first present the sharp ATE bounds for the case where $\ell = 2$ (binary instrument) in Section 4.1. Later in Section 4.2, we provide an exponential lower bound on the number of terms in the bounds for an arbitrary value of ℓ (multi-valued instrument).

4.1 BINARY INSTRUMENT

To compute the vertices of $\mathcal{H}$ explicitly in this case, we use the special structure of $\mathcal{H}$. An extensive list of the useful properties of $\mathcal{H}$ along with their proof is available in Appendix B. Using these properties, for every feasible point $\mathbf{v} \in \mathcal{H}$, we build a corresponding binary vector $\mathbf{b}$, which we call the signature of that feasible point. See Definition 10 in Appendix B.3 for the details of constructing a signature vector. We next define a set S of such signature vectors that, as we shall see, correspond to the M-distinct vertices of $\mathcal{H}$.

Definition 4 (Admissible signatures)**.** *The set of admissible signatures, denoted by S, is defined as $S = S_1 \cup S_2 \cup S_3$, where each S_i is a set of binary vectors $\mathbf{b} \in \mathbb{R}^{n\times 2\times 2}$ s.t.*

1. $\mathbf{b} \in S_1$ *iff*
   - $\exists\, t \in [n-1]$ *such that* $b_{i00} = b_{i01} = 1$ *for all* $i \geq t$ *and* $b_{i00} \neq b_{i01}$ *for all* $i < t$.
   - *For all* $i \in [n]$, $b_{i10} \neq b_{i11}$.
   - *There exist* $i, j \in [n]$ *such that* $b_{i10} = b_{j11} = 1$.
2. $\mathbf{b} \in S_2$ *iff*
   - $b_{(n-1)00} = b_{(n-1)01} = b_{010} = b_{011} = 1$.
   - *For all* $0 \leq i < n-1$, $b_{i00} \neq b_{i01}$.
   - *For all* $0 < j \leq n-1$, $b_{j10} \neq b_{j11}$.
3. $\mathbf{b} \in S_3$ *iff*
   - $\exists\, t \in [n-1]$ *such that* $b_{i10} = b_{i11} = 1$ *for all* $i \leq t+1$ *and* $b_{i10} \neq b_{i11}$ *for all* $i > t+1$.
   - *For all* $i \in [n]$, $b_{i00} \neq b_{i01}$.
   - *There exist* $i, j \in [n]$ *such that* $b_{i00} = b_{j01} = 1$.

In Appendix B, we first show that given a set of M-distinct vertices of $\mathcal{H}$, every vertex induces a unique admissible signature $\mathbf{b} \in S$. Conversely, given an admissible signature, we can reconstruct the corresponding vertex, as outlined below.

Definition 5 (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:*

- $u_{i00} = -\gamma_i - \alpha$ *if* $b_{i00} = 1$, *and* $u_{i00} = \gamma_0$ *otherwise;*
- $u_{i10} = \gamma_i$ *if* $b_{i10} = 1$, *and* $u_{i10} = -\gamma_{n-1} - \alpha$ *otherwise;*
- $u_{i01} = -\gamma_i$ *if* $b_{i01} = 1$, *and* $u_{i01} = \gamma_0 + \alpha$ *otherwise;*
- $u_{i11} = \gamma_i + \alpha$ *if* $b_{i11} = 1$, *and* $u_{i11} = -\gamma_{n-1}$ *otherwise,*

where

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

This following result establishes a one-to-one correspondence between the set of admissible signatures S and the M-distinct vertices of $\mathcal{H}$.

Theorem 1. *The set $\mathcal{V} = \{\mathbf{u}(\mathbf{b}) : \mathbf{b} \in S\}$ is a set of M-distinct vertices of $\mathcal{H}$, and any other vertex of $\mathcal{H}$ is M-equivalent to a member of $\mathcal{V}$.*

Corollary 1. *The number of M-distinct vertices of $\mathcal{H}$ is exactly*

$$5 \times 4^{n-1} - 2^{n+2} + 4,$$

which is the number of admissible signatures.

Sharp ATE bounds can be expressed in terms of the M-distinct vertices of $\mathcal{H}$ as follows.

Theorem 2. *Under Assumptions 1, 2, and 4, the ATE admits the following valid and sharp bounds:*

$$\max_{\mathbf{v}\in\mathcal{V}} \mathbf{v}^\top \mathbf{p} \ \leq\ \text{ATE} \ \leq\ -\max_{\mathbf{v}\in\mathcal{V}} \mathbf{v}^\top \bar{\mathbf{p}}. \tag{13}$$

We presented the sharp bounds on ATE under random assignment (Assumption 2). However, our next result indicates that the ATE admits the same set of sharp bounds under the weaker assumption of joint independence.

Theorem 3. *The ATE bounds of Equation* (13) *are valid and sharp under Assumptions 1, 3 and 4a.*

Notice that Theorem 3 does not require the variables $D^{(z)}$ to be defined. For $n = 2$, the corresponding S, $\mathcal{V}$, and ATE bounds are provided in Appendix E.3.

Remark 1. *The set of admissible signatures S can be generated in time $O(|S|)$ by conditioning on the threshold t in the definitions of S_1 and S_3 and then iterating over the corresponding binary vector choices. Moreover, each admissible signature can be mapped to its associated vertex in $O(n)$ time; see remark 8. Hence, all vertices in $\mathcal{V}$ can be generated in time $O(n|\mathcal{V}|)$, which is optimal in output size, as it is linear in the total length of the generated vertices.*

4.2 MULTI-VALUED INSTRUMENT

To show that the number of linear terms in the ATE bounds grows exponentially in the multi-valued instrument case ($\ell > 2$), we construct a family of M-distinct vertices whose cardinality grows as $\Omega(\ell^n)$.

Theorem 4. *For any $a \in [\ell]$, and any $\mathbf{s} = (s_0, \ldots, s_{n-1}) \in ([\ell] \setminus \{a\})^n$, define $\mathbf{w}(a, \mathbf{s}) \in \mathbb{R}^{n\times 2\times \ell}$ as*

$$\begin{aligned} w_{y1,j} &= 0, && \forall y,\ j \neq a, \\ w_{y1,a} &= \gamma_y - \gamma_{n-1}, && \forall y, \\ w_{y0,j} &= 0, && \forall y,\ j \notin \{a, s_y\}, \\ w_{y0,s_y} &= \gamma_{n-1} - \gamma_y, && \forall y, \\ w_{y0,a} &= \gamma_0 - \gamma_{n-1}, && \forall y. \end{aligned}$$

Then $\{\mathbf{w}(a,\mathbf{s}) : a \in [\ell],\ \mathbf{s} \in ([\ell] \setminus \{a\})^n$ non-constant$\}$ is a set of M-distinct vertices of $\mathcal{H}$ (the feasible region of Equation 10). The number of these M-distinct vertices equals

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

Corollary 2 (Exponential lower bound)**.** *Under Assumptions 1, 2 and 4 (or alternatively, 1, 3 and 4a), 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.

5 TESTABLE IMPLICATIONS

In a similar fashion to the previous section, we first present the sharp implications expressed as IV inequalities in a binary instrument case ($\ell = 2$) in Section 5.1. Later in Section 5.2, we provide an exponential lower bound on the number of sharp inequalities for an arbitrary value of ℓ.

5.1 BINARY INSTRUMENT

As mentioned in Section 3, the implications of the IV model boil down to

$$\mathbf{p}^\top \mathbf{r} \leq 0 \quad \text{for all } \mathbf{r} \in \mathcal{K}. \tag{14}$$

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, Section 19, Part IV]):

$$\mathcal{K} = \text{cone}(r_1, \dots, r_\omega). \tag{15}$$

Equation (14) is therefore equivalent to the finite family of inequalities

$$\mathbf{p}^\top r_i \leq 0 \qquad \forall i \in \{1, \dots, \omega\}, \tag{16}$$

where $\{r_1, \dots, r_\omega\}$ are M-distinct extreme rays of $\mathcal{K}$. That is, the inequalities of Equation (16) are sufficient to test the IV model. In Appendix C, we characterize these extreme rays (see Proposition 7). We further show that each inequality of Equation (16) is necessary, that is, for each $i \in \{1, \dots, \omega\}$, there exists a vector $\mathbf{p}$ such that $\mathbf{p}$ satisfies all inequalities but the one corresponding to r_i (see Section C.3). Therefore, to obtain an explicit set of necessary and sufficient (valid and sharp) IV inequalities, it suffices to characterize the M-distinct extreme rays r_i of $\mathcal{K}$. The following result presents these sharp implications.

Theorem 5. *Let $\mathcal{P}$ be the observed probability distribution, and $p_{yd,z}$ defined as in Section 2. The sharp testable implications of the IV model under Assumptions 1, 2, and 4 are the following inequalities:*

$$\begin{cases} \displaystyle\sum_{k \in [n]} -p_{k0,1} + \sum_{k \in T} (p_{k1,0} - p_{k1,1}) \leq 0, \\ \qquad\qquad \forall T \subseteq [n-1],\ T \neq \varnothing, \\ \displaystyle\sum_{k \in [n]} -p_{k1,0} + \sum_{k \in T} (p_{k0,1} - p_{k0,0}) \leq 0, \\ \qquad\qquad \forall T \subset [n],\ T \neq \varnothing, \\ \displaystyle\sum_{k \in [n]} -p_{k0,0} + \sum_{k \in T} (p_{k1,1} - p_{k1,0}) \leq 0, \\ \qquad\qquad \forall T \subseteq [n-1],\ T \neq \varnothing. \end{cases} \tag{17}$$

Corollary 3. *The number of sharp IV inequalities in a binary instrument setting is $2^{n+1} - 4$.*

Remark 2. *For a binary instrument, Bonet [2001] derived 2^{n+1} IV inequalities, showing that they are necessary but not sufficient. In fact, some of those inequalities were shown to be redundant within the same paper. More recently, Kédagni and Mourifié [2020] obtained $2^{n^2} + 2n(4 + 2^n) + 2^{n+1}$ inequalities and proved their necessity, noting that this set may not be sufficient (and clearly contains redundancies). Theorem 5 provides a non-redundant, necessary and sufficient set of size $2^{n+1} - 4$.*

For $n = 2$, the corresponding extreme rays and IV inequalities are provided in Appendix E.4.

5.2 MULTI-VALUED INSTRUMENT

We now show that the number of IV inequalities grows exponentially when the instrument Z takes $\ell > 2$ values.

To this end, it suffices to find a set of M-distinct extreme rays of $\mathcal{K}$ with exponentially increasing cardinality, and construct the corresponding IV inequalities. The following result presents such a set of inequalities.

Theorem 6. *Suppose Assumptions 1, 2 and 4 hold. For any $y' \in [n-1]$, $j' \in [\ell]$, and non-constant $(j_0, \dots, j_{n-1}) \in ([\ell] \setminus \{j'\})^n$, the following inequality holds:*

$$p_{y'1,j'} \leq \sum_{j \in \{j_0, \dots, j_{n-1}\}} p_{y'1,j} + \sum_{y=0}^{n-1} p_{y0,j_y}. \tag{18}$$

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

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

Corollary 4 (Exponential lower bound)**.** *Any sharp set of linear IV (in $\mathcal{P}$) inequalities must contain at least $(n-1)\,\ell\,\big((\ell-1)^n - (\ell-1)\big)$ inequalities.*

6 SIMULATIONS

As shown in Theorem 2 and Theorem 5, both the number of linear terms in the ATE bounds and the number of testable implications grows exponentially in the outcome support n. These numbers are provided in Table 1 for $2 \leq n \leq 10$. These results establish an exponential-time lower bound on the time complexity of deriving closed-form bounds and testable implications. Our approach of explicitly characterizing the vertices (extreme rays) of the dual feasible set allows for the derivation of the bounds (inequalities) in time linear in the number of terms appearing in the bound (inequalities). We can therefore use these characterizations to derive the ATE bounds as well as testable implications efficiently, i.e., with time complexity matching the lower bound.

In order to assess whether this approach is advantageous to the existing ones, we compared the running time of deriving

n	Terms in ATE Bound	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

Table 1: Number of linear terms in the ATE bounds, and the number of testable implications. n represents the size of the outcome support.

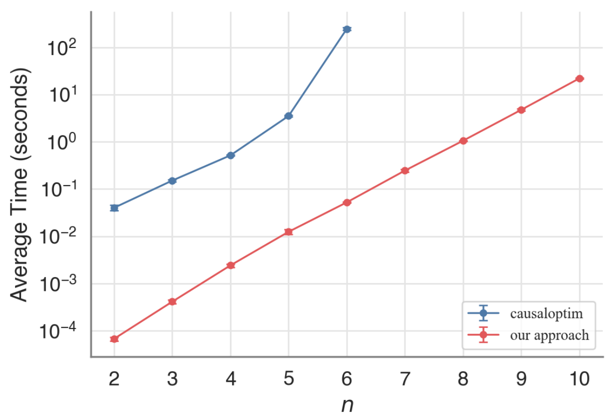

Figure 1: Average running time of our approach vs causaloptim [Sachs et al., 2023] for enumerating the vertices of the dual LP and producing the ATE bounds.

sharp closed-form ATE bounds using our approach vs the approach of Sachs et al. [2023], who enumerate the vertices of the feasible set using off-the-shelf methods. We used the causaloptim library in R [Sachs et al., 2023] for this comparison. The code implementing our approach is available online[2].

Figure 1 illustrates the running times across different values of n (the size of the outcome support). As expected, our approach requires time exponential in n (linear trend in the log-scale plot). However, as shown in Figure 1, the approach of Sachs et al. [2023] requires time *super-exponential* in n to enumerate the vertices, albeit the number of vertices is only exponential. With support sizes as small as $n = 6$, we can already witness more than 1000 times faster running times through our approach. Importantly, both methods yield the exact same sharp bounds, while our approach runs several orders of magnitude faster, taking milliseconds compared with over 24 hours for causaloptim when $|Y| = 7$.

The super-exponential scaling of causaloptim stems from its reliance on general-purpose polyhedral enumeration. Such enumeration procedures do not scale linearly with the number of resulting vertices or extreme rays; instead, they search over a much larger implicit candidate space. In contrast, our approach exploits the structural patterns of the discrete IV model identified in this paper, avoiding this generic enumeration bottleneck.

We note that confidence intervals are included in Figure 1; however, they are visually negligible due to the deterministic nature of both algorithms. The minor observed variation in running times is driven only by external factors such as operating-system scheduling and background processes.

[2]https://github.com/ArefeBoushehrian/Analytical-Causal-Bounds-in-Instrumental-Variable-Models

7 CONCLUSION

We studied the problem of deriving sharp analytical closed-form ATE bounds and testable implications of the discrete IV model. We showed that the number of linear expressions in any set of sharp bounds for the ATE, as well as the number of sharp testable implications in such models grow exponentially in the outcome support. This rules out any approach trying to derive sharp linear bounds (inequalities) in polynomial time. On the positive side, we explicitly characterized the sharp ATE bounds and the sharp IV inequalities in the binary instrument setting. As shown in our simulations, this approach makes deriving bounds and inequalities tractable for moderate sizes of outcome support, which were untractable with previously existing approaches. While this paper focuses on the fundamental IV setting, it provides groundwork for extending sharp characterizations to broader graph classes and estimands; we leave these extensions to future work.

Acknowledgements

The work of AB, MB and NK was supported by the Swiss National Science Foundation (SNSF) under the NCCR Automation grant agreement agreement 51NF40_225155 and project grant 200021_204355. SA was supported by an SNSF Postdoc.Mobility fellowship under grant P500PT_230240.

References

Joshua Angrist and Guido Imbens. Sources of identifying information in evaluation models, 1991.

Zvi Artstein. Distributions of random sets and random selections. *Israel Journal of Mathematics*, 46(4):313–324, 1983. doi: 10.1007/BF02762891.

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal* *of the American Statistical Association*, 92(439):1171–1176, 1997. doi: 10.1080/01621459.1997.10474074.

Alexander Abraham Balke. *Probabilistic Counterfactuals: Semantics, Computation, and Applications*. PhD thesis, University of California, Los Angeles, 1995.

Eli Ben-Michael. Partial identification via conditional linear programs: Estimation and policy learning, 2025.

Arie Beresteanu, Ilya Molchanov, and Francesca Molinari. Partial identification using random set theory. *Journal of Econometrics*, 166(1):17–32, 2012.

Blai Bonet. Instrumentality tests revisited. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 48–55, 2001.

Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Gary Chamberlain. Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics*, 32(2):189–218, 1986. doi: 10.1016/0304-4076(86)90038-2.

Jing Cheng and Dylan S Small. Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(5):815–836, 2006.

Guilherme Duarte, Noam Finkelstein, Dean Knox, Jonathan Mummolo, and Ilya Shpitser. An automated approach to causal inference in discrete settings. *Journal of the American Statistical Association*, 119(547):1778–1793, 2024.

Noam Finkelstein and Ilya Shpitser. Deriving bounds and inequality constraints using logical relations among counterfactuals. In *Conference on uncertainty in artificial intelligence*, pages 1348–1357. PMLR, 2020.

Maurice Fréchet. Généralisation du théoreme des probabilités totales. *Fundamenta mathematicae*, 25(1):379–387, 1935.

Florian F Gunsilius. Nontestability of instrument validity under continuous treatments. *Biometrika*, 108(4):989–995, 2021.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994. doi: 10.2307/2951620.

Toru Kitagawa. The identification region of the potential outcome distributions under instrument independence. *Journal of Econometrics*, 225(2):231–253, 2021. doi: 10.1016/j.jeconom.2021.03.006.

Désiré Kédagni and Ismael Mourifié. Generalized instrumental inequalities: testing the instrumental variable independence assumption. *Biometrika*, 107(3):661–675, 02 2020. ISSN 0006-3444. doi: 10.1093/biomet/asaa003. URL `https://doi.org/10.1093/biomet/asaa003`.

Ye Luo, Kirill Ponomarev, and Hai Wang. Selecting inequalities for sharp identification in models with set-valued predictions. Working paper, 2025.

Charles F Manski. Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323, 1990.

Charles F Manski. *Partial identification of probability distributions*. Springer, 2003.

Charles F. Manski and John V. Pepper. Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, 68(4):997–1010, 2000. doi: 10.1111/1468-0262.00144.

Judea Pearl. On the testability of causal models with latent and instrumental variables. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, page 435–443, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1558603859.

Thomas S. Richardson and James M. Robins. ACE bounds; SEMs with equilibrium conditions. *Statistical Science*, 29(3), August 2014. ISSN 0883-4237. doi: 10.1214/14-sts485. URL `http://dx.doi.org/10.1214/14-STS485`.

James M. Robins. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology: A Focus on AIDS*, pages 113–159. U.S. Public Health Service, 1989.

Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970. ISBN 9781400873173. doi: 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(2):567–576, 2023. doi: 10.1080/10618600.2022.2071905.

Xin Shu, Shuai Wang, and Ang Li. Identification of probabilities of causation: A complete characterization. *arXiv preprint arXiv:2505.15274*, 2025.

Yilin Song, F Richard Guo, KC Chan, and Thomas S Richardson. The categorical instrumental variable model: Characterization, partial identification, and statistical inference. *arXiv preprint arXiv:2405.09510*, 2024.

Fundamental Limits and Optimal Methods for Sharp Analytical Causal Bounds in Instrumental Variable Models (Supplementary Material)

Arefe Boushehrian[*1] **Mohammad Reza Badri**[*1] **Sina Akbari**[2] **Negar Kiyavash**[1]

[1]College of Management of Technology, EPFL, Switzerland
[2]Statistical Laboratory, University of Cambridge, UK

This supplementary material is organized as follows:

- The proof of Proposition 2 along with some preliminaries for the rest of our results is provided in Appendix A.
- The proofs of Theorems 1 and 2 are provided in Section B.
- The proof of Theorem 5 is provided in Section C.
- The proofs of Theorems 4 and 6 are provided in Section D.

A PRELIMINARIES

A.1 PROOF OF PROPOSITION 2

Proposition 2. *Under Assumptions 1, 2 and 4, given the observed law $\mathcal{P}$, the marginal law of potential outcome variables $Q(\cdot)$ is sharply characterized as follows:*

$$\begin{cases} p_{y0,z} = \sum\limits_{j\in[n]} \sum\limits_{\substack{\mathbf{d}\in\{0,1\}^\ell,\\ d_z=0}} q_{yj,\mathbf{d}} & \forall y \in [n],\ z \in [\ell], \\ p_{y1,z} = \sum\limits_{i\in[n]} \sum\limits_{\substack{\mathbf{d}\in\{0,1\}^\ell,\\ d_z=1}} q_{iy,\mathbf{d}} & \forall y \in [n],\ z \in [\ell], \\ q_{ij,\mathbf{d}} \geq 0 & \forall i,j \in [n],\ \mathbf{d} \in \{0,1\}^\ell, \end{cases} \tag{4}$$

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

Proof. ($\Rightarrow$) Consider the first equation for a fixed y and z. The right-hand-side summation is equal to $\mathcal{Q}(Y^{(0)} = y, D^{(z)} = 0)$. Under Assumption 2, the latter is also equal to $\mathcal{Q}(Y^{(0)} = y, D^{(z)} = 0 \mid Z = z)$, which, under consistency (Assumption 4) equals $\mathcal{P}(Y = y, D = 0 \mid Z = z) = p_{y0,z}$. The second set of equations hold similarly. The the inequalities $q_{ij,\mathbf{d}} \geq 0$ hold because $\mathcal{Q}$ is a probability measure.

($\Leftarrow$) By Theorem 1 of Song et al. [2024], it suffices to show that the linear relations Equation (4) imply the family of inequalities

$$\mathcal{Q}\Big(Y^{(0)} \in \mathcal{V}^{(0)}, Y^{(1)} \in \mathcal{V}^{(1)}\Big) \leq \sum_{d=0}^{1} \mathcal{P}\Big(Y \in \mathcal{V}^{(d)}, D = d \mid Z = z\Big), \qquad \forall z \in [\ell], \tag{19}$$

for every choice of nonempty sets $\mathcal{V}^{(0)}, \mathcal{V}^{(1)} \subseteq \{\gamma_0, \ldots, \gamma_{n-1}\}$.

Fix an arbitrary $z \in [\ell]$. We show that Equation (19) follows solely from Equation (4) for this fixed z.

Left-hand side: Expanding the definition of $\mathcal{Q}$ yields

$$\mathcal{Q}\Big(Y^{(0)} \in \mathcal{V}^{(0)}, Y^{(1)} \in \mathcal{V}^{(1)}\Big) = \sum_{\mathbf{d}\in\{0,1\}^\ell} \sum_{y_0\in\mathcal{V}^{(0)}} \sum_{y_1\in\mathcal{V}^{(1)}} q_{y_0y_1,\mathbf{d}}.$$

Right-hand side: Using (4),

$$\sum_{d=0}^{1} \sum_{y\in\mathcal{V}^{(d)}} p_{yd,z} = \sum_{y\in\mathcal{V}^{(0)}} \sum_{y_1} \sum_{\substack{\mathbf{d}\in\{0,1\}^\ell,\\ d_z=0}} q_{yy_1,\mathbf{d}} + \sum_{y\in\mathcal{V}^{(1)}} \sum_{y_0} \sum_{\substack{\mathbf{d}\in\{0,1\}^\ell,\\ d_z=1}} q_{y_0y,\mathbf{d}},$$

Since all entries of $\mathbf{q}$ are nonnegative, Hence,

$$\sum_{d=0}^{1} \sum_{y\in\mathcal{V}^{(d)}} p_{yd,z} \;\geq\; \sum_{\mathbf{d}\in\{0,1\}^\ell} \sum_{y_0\in\mathcal{V}^{(0)}} \sum_{y_1\in\mathcal{V}^{(1)}} q_{y_0y_1,\mathbf{d}}.$$

The right-hand side equals the left-hand side of (19), proving the inequality. □

A.2 VERTICES OF $\mathcal{H}$

We define the M-equivalence class of a vertex $\mathbf{v}$ as

$$[\mathbf{v}] = \{\, \mathbf{v} + \mathbf{s} \mid \mathbf{s} \in \ker(M) \,\}.$$

The Definition 2 captures extremality after quotienting out the intrinsic affine directions induced by $\ker(M)$.

Lemma 1 (Quotient characterization)**.** *Let $\pi : \mathbb{R}^{2\ell n} \to \mathbb{R}^{2\ell n}/\ker(M)$ denote the canonical projection, i.e., the map that identifies any two vectors $x, y \in \mathbb{R}^{2\ell n}$ such that $x - y \in \ker(M)$. Then $\mathbf{v}$ is a vertex of $\mathcal{H}$ in the sense of Definition 2 if and only if $\pi(\mathbf{v})$ is an extreme point of the projected polyhedron $\pi(\mathcal{H})$.*

Proof. ($\Rightarrow$) Assume $\mathbf{v}$ is a vertex and suppose that

$$\pi(\mathbf{v}) = \lambda\pi(\mathbf{v}_1) + (1-\lambda)\pi(\mathbf{v}_2)$$

for some $\lambda \in (0,1)$ and $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{H}$. By linearity of π, this implies

$$\pi\Big(\mathbf{v} - \lambda\mathbf{v}_1 - (1-\lambda)\mathbf{v}_2\Big) = \mathbf{0},$$

hence $\mathbf{v} = \lambda\mathbf{v}_1 + (1-\lambda)\mathbf{v}_2 + \mathbf{s}$ for some $\mathbf{s} \in \ker(M)$. Since $M\mathbf{s} = 0$, we also have $M(\mathbf{v}_1+\mathbf{s}) = M\mathbf{v}_1, M(\mathbf{v}_2+\mathbf{s}) = M\mathbf{v}_2$ and $\mathbf{v}_1 + \mathbf{s}, \mathbf{v}_2 + \mathbf{s} \in \mathcal{H}$ (because feasibility depends only on $M\cdot$). Thus we can rewrite

$$\mathbf{v} = \lambda(\mathbf{v}_1 + \mathbf{s}) + (1-\lambda)(\mathbf{v}_2 + \mathbf{s})$$

with both points in $\mathcal{H}$. By Definition 2, it follows that $M(\mathbf{v}_1 + \mathbf{s}) = M\mathbf{v}$ and $M(\mathbf{v}_2 + \mathbf{s}) = M\mathbf{v}$, i.e. $\pi(\mathbf{v}_1) = \pi(\mathbf{v})$ and $\pi(\mathbf{v}_2) = \pi(\mathbf{v})$. Therefore $\pi(\mathbf{v})$ is an extreme point of $\pi(\mathcal{H})$.

($\Leftarrow$) Conversely, assume that $\pi(\mathbf{v})$ is an extreme point of $\pi(\mathcal{H})$ and write

$$\mathbf{v} = \lambda\mathbf{v}_1 + (1-\lambda)\mathbf{v}_2, \qquad \lambda \in (0,1), \quad \mathbf{v}_1, \mathbf{v}_2 \in \mathcal{H}.$$

Applying π yields

$$\pi(\mathbf{v}) = \lambda\pi(\mathbf{v}_1) + (1-\lambda)\pi(\mathbf{v}_2).$$

By extremality of $\pi(\mathbf{v})$, we get $\pi(\mathbf{v}_1) = \pi(\mathbf{v}_2) = \pi(\mathbf{v})$, which is equivalent to $M\mathbf{v}_1 = M\mathbf{v}_2 = M\mathbf{v}$. This is exactly the condition in Definition 2. □

Lemma 2. *For any observed law $\mathcal{P}$ conforming to the IV and $\mathbf{s} \in \ker(M)$,*

$$\mathbf{p}^\top \mathbf{s} = 0.$$

Proof. Since $M^\top \mathbf{q} = \mathbf{p}$,

$$\mathbf{p}^\top \mathbf{s} = (M^\top \mathbf{q})^\top \mathbf{s} = \mathbf{q}^\top M\mathbf{s} \overset{(a)}{=} \mathbf{q}^\top 0 = 0,$$

where in (a) we used the fact that $\mathbf{s} \in \ker(M)$, we have $M\mathbf{s} = 0$. □

Consequently, for any linear objective $\mathbf{p}^\top \mathbf{v}$, it suffices to consider *one* representative per vertex class $[\mathbf{v}]$.

By the fundamental theorem of linear programming, whenever the primal is feasible and bounded, the dual optimum over $\mathcal{H}$ is attained at an extreme point of $\pi(\mathcal{H})$, equivalently at a vertex M-equivalence class of $\mathbf{v}$ of $\mathcal{H}$ in the sense of Definition 2. Therefore, sharp identification bounds can be obtained by evaluating $\mathbf{p}^\top \mathbf{v}$ over a finite set containing one representative from each vertex class.

A.3 EXTREME RAYS OF $\mathcal{K}$

The set $\mathcal{K} := \{\mathbf{r} \in \mathbb{R}^{2\ell n} : M\mathbf{r} \leq \mathbf{0}\}$ is a polyhedral cone. Moreover, by standard LP duality, feasibility of the primal system is equivalent to

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

Hence, sharp testable implications (i.e., a necessary and sufficient finite family of IV inequalities) can be obtained by characterizing the extreme structure of $\mathcal{K}$.

M-Equivalence under Scaling and Kernel Shifts. Because $\ker(M) \neq \{\mathbf{0}\}$ (see Remark 4), the cone $\mathcal{K}$ is generally not pointed and contains a lineality space. In particular, $\ker(M) \subseteq \mathcal{K}$ and $\ker(M) \subseteq -\mathcal{K}$. Accordingly, we work with rays modulo positive scaling and kernel shifts.

Definition 6 (Extreme ray class)**.** *For nonzero $\mathbf{r}, \mathbf{r}' \in \mathcal{K}$, we write $\mathbf{r} \sim \mathbf{r}'$ if there exists $\lambda > 0$ such that*

$$\lambda \mathbf{r} - \mathbf{r}' \in \ker(M).$$

We call the M-equivalence class

$$\mathcal{E}_{\mathbf{r}} := \{\lambda \mathbf{r} + \mathbf{s} : \lambda > 0,\ \mathbf{s} \in \ker(M)\}$$

the ray class *generated by* $\mathbf{r}$.

Definition 7 (M-Distinct extreme rays)**.** *Two extreme rays $\mathbf{r}_1, \mathbf{r}_2 \notin \ker(M)$ are called M-distinct if*

$$\mathcal{E}_{\mathbf{r}_1} \neq \mathcal{E}_{\mathbf{r}_2}.$$

Lemma 3. *For any two M-distinct extreme rays $\mathbf{r}_1$ and $\mathbf{r}_2$, we have*

$$\mathcal{E}_{\mathbf{r}_1} \cap \mathcal{E}_{\mathbf{r}_2} = \varnothing.$$

Proof. Suppose, by contradiction, that there exists $\mathbf{r}' \in \mathcal{E}_{\mathbf{r}_1} \cap \mathcal{E}_{\mathbf{r}_2}$. Then there exist $\alpha_1, \alpha_2 > 0$ and $\mathbf{s}_1, \mathbf{s}_2 \in \ker(M)$ such that

$$\mathbf{r}' = \alpha_1 \mathbf{r}_1 + \mathbf{s}_1 \qquad \text{and} \qquad \mathbf{r}' = \alpha_2 \mathbf{r}_2 + \mathbf{s}_2.$$

Rearranging yields

$$\mathbf{r}_1 = \frac{\alpha_2}{\alpha_1} \mathbf{r}_2 + \frac{\alpha_2}{\alpha_1}(\mathbf{s}_2 - \mathbf{s}_1),$$

which implies $\mathbf{r}_1 \in \mathcal{E}_{\mathbf{r}_2}$. By symmetry, $\mathbf{r}_2 \in \mathcal{E}_{\mathbf{r}_1}$, hence $\mathcal{E}_{\mathbf{r}_1} = \mathcal{E}_{\mathbf{r}_2}$, a contradiction. □

The following lemma formalizes the fact that extremality is a property of the ray class and coincides with extremality in the quotient cone.

Lemma 4 (Quotient characterization of extreme rays)**.** *Let $\pi : \mathbb{R}^{2\ell n} \to \mathbb{R}^{2\ell n} \setminus \ker(M)$ be the canonical projection and let $\widehat{\mathcal{K}} := \pi(\mathcal{K})$. Then $\mathbf{r} \in \mathcal{K} \setminus \ker(M)$ is an extreme ray in the sense of Definition 3 if and only if $\pi(\mathbf{r})$ generates an extreme ray of the pointed polyhedral cone $\widehat{\mathcal{K}}$. Moreover, if $\mathbf{r} \sim \mathbf{r}'$, then $\pi(\mathbf{r})$ and $\pi(\mathbf{r}')$ generate the same ray in $\widehat{\mathcal{K}}$.*

Proof. First, note that $\pi(\mathbf{r}) = \pi(\mathbf{r}')$ holds if and only if $\mathbf{r} - \mathbf{r}' \in \ker(M)$, hence $\mathbf{r} \sim \mathbf{r}'$ implies that $\pi(\mathbf{r})$ and $\pi(\mathbf{r}')$ lie on the same ray in $\widehat{\mathcal{K}}$. Now suppose that $\mathbf{r}$ is not extreme in the sense of Definition 3. Then there exist $\mathbf{r}_1, \mathbf{r}_2 \in \mathcal{K}$ such that $\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2$ and neither summand belongs to the ray class $\mathcal{E}_{\mathbf{r}}$, i.e., $\mathbf{r}_t \notin \{\alpha\mathbf{r} + \mathbf{s} : \alpha > 0,\ \mathbf{s} \in \ker(M)\}$. Applying π yields $\pi(\mathbf{r}) = \pi(\mathbf{r}_1) + \pi(\mathbf{r}_2)$ with $\pi(\mathbf{r}_1), \pi(\mathbf{r}_2) \in \widehat{\mathcal{K}}$, and neither $\pi(\mathbf{r}_t)$ lies on the ray generated by $\pi(\mathbf{r})$. Hence $\pi(\mathbf{r})$ does not generate an extreme ray of $\widehat{\mathcal{K}}$.

Conversely, if $\pi(\mathbf{r})$ does not generate an extreme ray of $\widehat{\mathcal{K}}$, then we can write $\pi(\mathbf{r}) = \widehat{r}_1 + \widehat{r}_2$ with $\widehat{r}_1, \widehat{r}_2 \in \widehat{\mathcal{K}}$ such that neither lies on the ray of $\pi(\mathbf{r})$. Choose preimages $\mathbf{r}_1, \mathbf{r}_2 \in \mathcal{K}$ with $\pi(\mathbf{r}_t) = \widehat{r}_t$. Then $\pi(\mathbf{r} - \mathbf{r}_1 - \mathbf{r}_2) = 0$, so $\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2 + \mathbf{s}$ for some $\mathbf{s} \in \ker(M)$. Since $\mathbf{s} \in \mathcal{K}$ and $\mathcal{K}$ is a cone, we may absorb $\mathbf{s}$ into, say, $\mathbf{r}_2$ and obtain a decomposition $\mathbf{r} = \tilde{\mathbf{r}}_1 + \tilde{\mathbf{r}}_2$ with $\tilde{\mathbf{r}}_1, \tilde{\mathbf{r}}_2 \in \mathcal{K}$ whose projections do not lie on the ray of $\pi(\mathbf{r})$; equivalently, neither summand belongs to the ray class $\mathcal{E}_{\mathbf{r}}$. Thus $\mathbf{r}$ is not extreme in the sense of Definition 3. □

Corollary 5 (Finite generation and reduction to finitely many inequalities)**.** *The quotient cone $\widehat{\mathcal{K}} = \pi(\mathcal{K})$ is a pointed polyhedral cone and therefore admits only finitely many extreme rays. Consequently, there exist $\mathbf{r}_1, \ldots, \mathbf{r}_\omega \in \mathcal{K}$ such that*

$$\mathcal{K} = \ker(M) + \textit{cone}(\mathbf{r}_1, \ldots, \mathbf{r}_\omega),$$

and every element of $\mathcal{K}$ can be written as $\mathbf{s} + \sum_{i=1}^{\omega} \lambda_i \mathbf{r}_i$ for some $\mathbf{s} \in \ker(M)$ and $\lambda_i \geq 0$.

In particular, if $\mathbf{p}^\top \mathbf{s} = 0$ for all $\mathbf{s} \in \ker(M)$, then the family of inequalities

$$\mathbf{p}^\top \mathbf{r} \leq 0 \quad \forall \mathbf{r} \in \mathcal{K}$$

is equivalent to the finite family

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

Proof. Since $\mathcal{K}$ is polyhedral, $\widehat{\mathcal{K}} = \pi(\mathcal{K})$ is also polyhedral. By construction it is pointed, hence it has finitely many extreme rays; see, e.g., [Rockafellar, 1970, Section 19]. Choosing one representative $\mathbf{r}_i$ in each extreme-ray class and lifting back to $\mathcal{K}$ yields $\mathcal{K} = \ker(M) + \text{cone}(\mathbf{r}_1, \ldots, \mathbf{r}_\omega)$.

Finally, if $\mathbf{r} = \mathbf{s} + \sum_i \lambda_i \mathbf{r}_i$, then

$$\mathbf{p}^\top \mathbf{r} = \mathbf{p}^\top \mathbf{s} + \sum_i \lambda_i \mathbf{p}^\top \mathbf{r}_i = \sum_i \lambda_i \mathbf{p}^\top \mathbf{r}_i,$$

so $\mathbf{p}^\top \mathbf{r} \leq 0$ for all $\mathbf{r} \in \mathcal{K}$ holds if and only if $\mathbf{p}^\top \mathbf{r}_i \leq 0$ for all i. □

Lineality Space. Since $\ker(M)$ is a linear subspace, there exists finite nonzero vectors $\mathbf{s}_1, \cdots \mathbf{s}_t$ such that

$$\ker(M) = \text{span}(\mathbf{s}_1, \cdots, \mathbf{s}_t) = \text{cone}(\mathbf{s}_1, \cdots, \mathbf{s}_t) + \text{cone}(-\mathbf{s}_1, \cdots, -\mathbf{s}_t).$$

Consequently,

$$\mathcal{K} = \text{cone}(\mathbf{s}_1, \cdots, \mathbf{s}_t, -\mathbf{s}_1, \cdots, -\mathbf{s}_t, \mathbf{r}_1, \ldots, \mathbf{r}_\omega),$$

where $\mathbf{r}_1, \ldots, \mathbf{r}_\omega$ are representatives of extreme-ray classes in $\mathcal{K} \setminus \ker(M)$. Moreover, note that $\mathbf{s}_1, \cdots, \mathbf{s}_t$ and $-\mathbf{s}_1, \cdots, -\mathbf{s}_t$ satisfy Definition 3, thus, they are extreme-rays in $\mathcal{K} \cap \ker(M)$.

Therefore, if $\mathbf{p}^\top \mathbf{s}_i \leq 0$ and $-\mathbf{p}^\top \mathbf{s}_i \leq 0$ are satisfied for all $i \leq t$, then we have $\mathbf{p}^\top \mathbf{s} = 0$ for all $\mathbf{s} \in \ker(M)$. As a result, based on Corollary 5,

$$\mathbf{p}^\top \mathbf{r} \leq 0 \quad \forall \mathbf{r} \in \mathcal{K}$$

is equivalent to the finite family

$$\begin{aligned} \mathbf{p}^\top \mathbf{r}_i &\leq 0 \qquad \forall i \in \{1, \ldots, \omega\} \\ \mathbf{p}^\top \mathbf{s}_i &\leq 0 \qquad \forall i \in \{1, \ldots, t\} \\ -\mathbf{p}^\top \mathbf{s}_i &\leq 0 \qquad \forall i \in \{1, \ldots, t\}. \end{aligned}$$

Consequently, sharp testable implications can be obtained by enumerating one representative from each extreme-ray class along with $\mathbf{s}_1, \cdots, \mathbf{s}_t$ and $-\mathbf{s}_1, \cdots, -\mathbf{s}_t$ from the extreme ray class $\mathcal{E}_0$ which denote the $\ker(M)$.

B VERTEX ENUMERATION

In this section, we provide the proofs and notations for Section 4. First, in Section B.1, we introduce the basic definitions and notation required for the problem. In Section B.2, we present fundamental lemmas that characterize the vertices and the structure of the dual LP matrix. In Section B.3, we define the signature binary number associated with a vertex $\mathbf{v}$ and prove lemmas describing its structure. Next, in Section B.4, we establish the existence of a bijection between these signatures and the vertices. Finally, in Section B.5, we prove the main results, Theorem 1 and Theorem 2, using the previously established lemmas.

B.1 DEFINITIONS AND NOTATIONS

In this subsection, we introduce the notation, definitions, and preliminary results required for the subsequent proofs.

Definition 8 (Active Constraint Matrix). *For a point $\mathbf{v}$, let $M_{\mathbf{v}}$ denote 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., $M_{\mathbf{v}}\mathbf{v} = \mathbf{c}_{\mathbf{v}}$. We call $M_{\mathbf{v}}$ the* active constraint matrix *at $\mathbf{v}$.*

Definition 9. *Let $M(\alpha, \beta)$ denote the submatrix of M obtained by restricting to the rows indexed by α and the columns indexed by β. Here, a row index α is of the form $y_0y_1d_0d_1$, and a column index β is of the form ydz, where each symbol may also be the wildcard $*$, indicating that it ranges over all possible values.*

By $M_1 \in M_2$, we mean that the row indices appearing in M_1 are a subset of the row indices appearing in M_2, and the column indices appearing in M_1 are a subset of the column indices appearing in M_2.

Also, by $M(\alpha_1 \cup \alpha_2, \beta_1 \cup \beta_2)$ we mean the submatrix obtained by restricting to all rows indexed by either α_1 or α_2, and to the columns indexed by either β_1 or β_2.

The same notation applies to $M_{\mathbf{v}}$. An example for this definition is provided in section E.2.

Based on the LP formulation, each variable $q_{y_0y_1,d_0d_1}$ is constrained by exactly two $p_{yd,z}$s. More precisely,

- $q_{y_0y_1,00}$ is constrained by $p_{y_00,0}$ and $p_{y_00,1}$,
- $q_{y_0y_1,01}$ is constrained by $p_{y_00,0}$ and $p_{y_11,1}$,
- $q_{y_0y_1,10}$ is constrained by $p_{y_00,1}$ and $p_{y_11,0}$,
- $q_{y_0y_1,11}$ is constrained by $p_{y_11,0}$ and $p_{y_11,1}$.

According to these relations between $\mathbf{p}$ and $\mathbf{q}$, the matrix M will have a specific pattern outlined in the following remark.

Remark 3. *For $i, j \in [n]$, all entries equal to 1 in M occur in the following positions and any other entry is eqaul to 0:*

- *$M(ij00, i00)$ and $M(ij00, i01)$,*
- *$M(ij01, i00)$ and $M(ij01, j11)$,*
- *$M(ij10, i01)$ and $M(ij10, j10)$,*
- *$M(ij11, j10)$ and $M(ij11, j11)$.*

B.2 STRUCTURAL PROPERTIES OF VERTICES

In this subsection, we establish some fundamental structural properties of the constraint matrix M and the active constraint matrix $M_{\mathbf{v}}$ associated with a vertex $\mathbf{v}$.

Lemma 5. *The rank of matrix M is $4n - 1$.*

Proof. By Remark 3, every row of M contains exactly one entry equal to 1 in $M(****, **0)$ and exactly one entry equal to 1 in $M(****, **1)$. Therefore, the sum of all columns of $M(****, **0)$ minus the sum of all columns of $M(****, **1)$ gives a nontrivial linear combination of columns equal to zero. Hence, M cannot have full rank, and

$$\text{rank}(M) \leq 4n - 1.$$

Now suppose that there exists a nontrivial linear combination of columns of M that is equal to zero. Then at least one column must have a nonzero coefficient. Without loss of generality, assume that the coefficient assigned to the column $M(****, i00)$ is $c_{i00} = s \neq 0$.

For any $j \in [n]$, the columns $M(****, i00)$ and $M(****, j11)$ share a 1 in some row of M. Since each row of M contains exactly two entries equal to 1, the corresponding coefficient must satisfy

$$c_{j11} = -s$$

in order for the linear combination to vanish on that row. Applying the same argument to all such rows implies that

$$c_{j11} = -s \quad \text{for all } j \in [n].$$

Next, since each column $M(****, j11)$ shares 1's with the column $M(****, j00)$, the same reasoning shows that

$$c_{j00} = s \quad \text{for all } j \in [n].$$

Similarly, for any $i \in [n]$, the columns $M(****, i00)$ and $M(****, i01)$ share n entries equal to 1 in the rows indexed by $M(i*00, i0*)$. Thus, to cancel these contributions we must have

$$c_{i01} = -s \quad \text{for all } i \in [n].$$

By the same argument, since $M(****, i01)$ and $M(****, i10)$ share 1's, we obtain

$$c_{i10} = s \quad \text{for all } i \in [n].$$

Therefore, in any nontrivial linear dependence among the columns of M, all columns appear with coefficient either s or $-s$, and this dependence is unique up to scaling. Consequently, the null space of M is one-dimensional, and hence

$$\text{rank}(M) = 4n - 1.$$

□

Remark 4. *As shown in the proof of Lemma 5, the null space of M is one-dimensional. Moreover, any vector $\mathbf{s} = (s_{ydz})_{y\in[n],\, d,z\in\{0,1\}}$ in $\ker(M)$ must satisfy*

$$s_{yd0} = s, \qquad s_{yd1} = -s \quad \textit{for all } y \in [n], d \in \{0,1\},$$

for some scalar $s \in \mathbb{R}$.

Lemma 6. *Let $\mathbf{v}$ be a feasible point satisfying $M\mathbf{v} \leq \mathbf{c}$. Then $\mathbf{v}$ is a vertex if and only if*

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

Proof. We prove both directions.

(⇐) If rank$(M_{\mathbf{v}}) < 4n - 1$, then $\mathbf{v}$ is not a vertex.

Suppose $\text{rank}(M_{\mathbf{v}}) < 4n - 1$. Since $\text{rank}(M) = 4n - 1$, we have

$$\dim(\ker(M_{\mathbf{v}})) > \dim(\ker(M)).$$

Hence, there exists a nonzero vector $\mathbf{t} \in \ker(M_{\mathbf{v}})$ such that $\mathbf{t} \notin \ker(M)$.

Since $\mathbf{t} \in \ker(M_{\mathbf{v}})$, we have

$$M_{\mathbf{v}}(\mathbf{v} + \epsilon\mathbf{t}) = M_{\mathbf{v}}\mathbf{v} = \mathbf{c}_{\mathbf{v}}$$

for all $\epsilon \in \mathbb{R}$. For sufficiently small $\epsilon > 0$, the inequalities corresponding to non-active constraints remain strict, because they are strict at $\mathbf{v}$. Therefore,

$$M(\mathbf{v} + \epsilon\mathbf{t}) \leq \mathbf{c} \quad \text{and} \quad M(\mathbf{v} - \epsilon\mathbf{t}) \leq \mathbf{c}$$

for sufficiently small ϵ.

Thus both $\mathbf{v} + \epsilon\mathbf{t}$ and $\mathbf{v} - \epsilon\mathbf{t}$ are feasible. Moreover,

$$\mathbf{v} = \tfrac{1}{2}(\mathbf{v} + \epsilon\mathbf{t}) + \tfrac{1}{2}(\mathbf{v} - \epsilon\mathbf{t}),$$

and since $\mathbf{t} \notin \ker(M)$, we have

$$M(\mathbf{v} + \epsilon\mathbf{t}) \neq M\mathbf{v},$$

so the two feasible points are M-distinct.

Hence $\mathbf{v}$ can be written as a non-trivial convex combination of two M-distinct feasible points, and therefore $\mathbf{v}$ is not a vertex by Definition 2.

($\Rightarrow$) If rank$(M_{\mathbf{v}}) = 4n - 1$, then $\mathbf{v}$ is a vertex.

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

$$\mathbf{v} = \lambda\mathbf{v}_1 + (1 - \lambda)\mathbf{v}_2, \qquad \lambda \in (0, 1),$$

for feasible points $\mathbf{v}_1, \mathbf{v}_2$.

Since $\mathbf{v}$ is active on $M_{\mathbf{v}}$, we have

$$M_{\mathbf{v}}\mathbf{v} = \mathbf{c}_{\mathbf{v}}.$$

Applying $M_{\mathbf{v}}$ to the convex combination gives

$$M_{\mathbf{v}}\mathbf{v} = \lambda M_{\mathbf{v}}\mathbf{v}_1 + (1 - \lambda)M_{\mathbf{v}}\mathbf{v}_2.$$

Because $\mathbf{v}_1$ and $\mathbf{v}_2$ are feasible,

$$M_{\mathbf{v}}\mathbf{v}_1 \leq \mathbf{c}_{\mathbf{v}}, \qquad M_{\mathbf{v}}\mathbf{v}_2 \leq \mathbf{c}_{\mathbf{v}}.$$

Since the convex combination equals $\mathbf{c}_{\mathbf{v}}$, it follows that

$$M_{\mathbf{v}}\mathbf{v}_1 = \mathbf{c}_{\mathbf{v}} \quad \text{and} \quad M_{\mathbf{v}}\mathbf{v}_2 = \mathbf{c}_{\mathbf{v}}.$$

Therefore,

$$M_{\mathbf{v}}(\mathbf{v} - \mathbf{v}_1) = \mathbf{0}, \qquad M_{\mathbf{v}}(\mathbf{v} - \mathbf{v}_2) = \mathbf{0}.$$

Because rank$(M_{\mathbf{v}}) = 4n - 1 =$ rank(M), and the fact the $M_{\mathbf{v}}$ is a submatrix of M we have

$$\ker(M_{\mathbf{v}}) = \ker(M).$$

Hence,

$$\mathbf{v} - \mathbf{v}_1 \in \ker(M), \qquad \mathbf{v} - \mathbf{v}_2 \in \ker(M).$$

By Definition 2, any two feasible points differing by a vector in $\ker(M)$ represent the same extreme point. Thus $\mathbf{v}$ cannot be written as a non-trivial convex combination of two M-distinct feasible points, and therefore $\mathbf{v}$ is a vertex. □

From now on, we assume that $\mathbf{v}$ represents a vertex. This means that $M\mathbf{v} \leq \mathbf{c}$ and $M_{\mathbf{v}}\mathbf{v} = \mathbf{c}_{\mathbf{v}}$, and rank$(M_{\mathbf{v}}) = 4n - 1$ (by Lemma 5).

Remark 5. *Since rank(M) = rank$(M_{\mathbf{v}})$ = $4n - 1$, the system $M_{\mathbf{v}}\mathbf{v} = \mathbf{c}$ has a unique solution in the affine space $\mathbf{v} + \ker(M)$. Therefore, $M_{\mathbf{v}}$ uniquely determines the M-equivalence class $[\mathbf{v}]$.*

In particular, if two active constraint matrices share the same row basis (that is, they consist of the same linearly independent rows), then they determine the same solution $\mathbf{v}$ and thus the same M-equivalence class $[\mathbf{v}]$. Consequently, the two active constraint matrices must coincide.

Lemma 7. *If $M(ij00, ***) \in M_{\mathbf{v}}$ then $j = 0$. Also if $M(ij11, ***) \in M_{\mathbf{v}}$ then $i = n - 1$.*

Proof. Suppose that $M(ij00, ***) \in M_{\mathbf{v}}$ for some $j > 0$, then

$$M(ij00, ***)\mathbf{v} = \gamma_j - \gamma_i > \gamma_0 - \gamma_i.$$

On the other hand, based on the structure of M we know that $M(ij00, ***) = M(i000, ***)$, so

$$M(i000, ***)\mathbf{v} > \gamma_0 - \gamma_i,$$

which contradicts the fact that $\mathbf{v}$ is a vertex. Hence $j = 0$.

Now suppose that $M(ij11, ***) \in M_{\mathbf{v}}$ for some $i < n - 1$. Then

$$M(ij11, ***)\mathbf{v} = \gamma_j - \gamma_i < \gamma_j - \gamma_{n-1}.$$

Again, by the structure of M we have $M(ij11, ***) = M((n-1)j11, ***)$, and therefore

$$M((n-1)j11, ***)\mathbf{v} < \gamma_j - \gamma_{n-1},$$

which contradicts the fact that $\mathbf{v}$ is a vertex. Hence $i = n - 1$. ☐

Lemma 8. *For any vertex* $\mathbf{v}$*, every column of* $M_{\mathbf{v}}$ *contains at least one entry equal to* 1*.*

Proof. We know that the null space of $M_{\mathbf{v}}$ is one-dimensional and is spanned by some non-zero vector $\mathbf{s}$ (see Remark 4). Suppose, by contradiction, that there exists a column indexed by (y, d, z) in $M_{\mathbf{v}}$ that contains no entries equal to 1. Then the corresponding standard basis vector $\mathbf{e}_{ydz}$ satisfies

$$M_{\mathbf{v}}\mathbf{e}_{ydz} = \mathbf{0}.$$

Hence $\mathbf{e}_{ydz}$ lies in the null space of $M_{\mathbf{v}}$.

Since $\mathbf{e}_{ydz}$ is linearly independent from $\mathbf{s}$, the null space of $M_{\mathbf{v}}$ has dimension at least 2. This contradicts the fact that $\mathbf{v}$ is a vertex based on Lemma 6. Hence every column of $M_{\mathbf{v}}$ must contain at least one entry equal to 1. ☐

B.3 SIGNATURE OF A VERTEX

In this section, we define the signature of a vertex and prove its fundamental properties. These results will help us show that the signature is unique for each vertex and that it can be used to reconstruct the vertex.

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

Lemma 9. *For any* y, z*, if* $b^{\mathbf{v}}_{y0z} = 0$*, then* $M(y000, ***) \in M_{\mathbf{v}}$*. Similarly, if* $b^{\mathbf{v}}_{y1z} = 0$*, then* $M((n-1)y11, ***) \in M_{\mathbf{v}}$*.*

Proof. By Lemma 8, every column of $M_{\mathbf{v}}$ must contain at least one entry equal to 1. Fix indices y, z and suppose that $b^{\mathbf{v}}_{ydz} = 0$. By definition of $\mathbf{b}^{\mathbf{v}}$, this means that there is no entry equal to 1 in $M_{\mathbf{v}}(**01 \cup **10, ydz)$. Consequently, the required 1 in column $M_{\mathbf{v}}(****, ydz)$ must lie in $M_{\mathbf{v}}(**00 \cup **11, ydz)$.

First, assume that $d = 0$. By the structure of M (see Remark 3), the only rows of $M_{\mathbf{v}}(**00 \cup **11, y0z)$ that can contain a 1 are the rows $M(yj00, ***)$. Hence, at least one such row must belong to $M_{\mathbf{v}}$. By Lemma 7, this is only possible when $j = 0$, and therefore $M(y000, ***) \in M_{\mathbf{v}}$.

The case $d = 1$ follows by a similar argument: if $b^{\mathbf{v}}_{y1z} = 0$, then the 1 in column $(y1z)$ must lie in $M_{\mathbf{v}}(**11, ***)$, and by the structure of M the only possible rows are $M(ij11, ***)$. Applying Lemma 7 yields $i = n - 1$, and hence $M((n-1)y11, ***) \in M_{\mathbf{v}}$. ☐

Lemma 10. *For any* $i, j \in [n]$*,* $b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{j11} = 1$ *if and only if* $M(ij01, ***) \in M_{\mathbf{v}}$*. Similarly, for any* $i, j \in [n]$*,* $b^{\mathbf{v}}_{i01} = b^{\mathbf{v}}_{j10} = 1$ *if and only if* $M(ij10, ***) \in M_{\mathbf{v}}$*.*

Proof. The "$\Leftarrow$" direction in both statements follows immediately from Definition 10, since the presence of a row in $M_{\mathbf{v}}$ forces a 1 in the corresponding columns.

We prove the "$\Rightarrow$" directions.

First case. Assume $b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{j11} = 1$. By Definition 10 and the fact that $M_{\mathbf{v}}(**10, *00 \cup *11)$ contains no entries equal to 1, there must exist at least one row in $M_{\mathbf{v}}(**01, ***)$ containing required 1 entries.

Suppose, by contradiction, that $M(ij01, ***) \notin M_{\mathbf{v}}$. Then there exist indices $a \neq i$ and $b \neq j$ such that

$$M(aj01, ***),\ M(ib01, ***) \in M_{\mathbf{v}}.$$

(If no such a existed, $M(**01, j11)$ would contain no 1(see Remark 3), contradicting $b^{\mathbf{v}}_{j11} = 1$; the same argument applies to b).

Since these rows belong to $M_{\mathbf{v}}$, we have

$$M(aj01, ***)\mathbf{v} = \gamma_j - \gamma_a, \qquad M(ib01, ***)\mathbf{v} = \gamma_b - \gamma_i,$$

and therefore

$$\big(M(aj01, ***) + M(ib01, ***)\big)\mathbf{v} = \gamma_j - \gamma_i + \gamma_b - \gamma_a.$$

On the other hand, since $M(ij01, ***) \notin M_{\mathbf{v}}$, we obtain

$$M(ij01, ***)\mathbf{v} < \gamma_j - \gamma_i, \qquad M(ab01, ***)\mathbf{v} \leq \gamma_b - \gamma_a,$$

and therefore,

$$\big(M(ij01, ***) + M(ab01, ***)\big)\mathbf{v} < \gamma_j - \gamma_i + \gamma_b - \gamma_a = \big(M(aj01, ***) + M(ib01, ***)\big)\mathbf{v}.$$

By the structure of M (see Remark 3),

$$M(aj01, ***) + M(ib01, ***) = M(ij01, ***) + M(ab01, ***),$$

which yields a contradiction.

Second case. Assume $b^{\mathbf{v}}_{i01} = b^{\mathbf{v}}_{j10} = 1$ and suppose $M(ij10, ***) \notin M_{\mathbf{v}}$. Since $M_{\mathbf{v}}(**01, *01 \cup *10)$ contains no 1's, the required 1's must lie in $M_{\mathbf{v}}(**10, ***)$. Thus, there exist indices $a \neq i$ and $b \neq j$ such that

$$M(aj10, ***),\ M(ib10, ***) \in M_{\mathbf{v}}.$$

As before, using the structure of M we have

$$M(aj10, ***) + M(ib10, ***) = M(ij10, ***) + M(ab10, ***),$$

and evaluating at $\mathbf{v}$ gives

$$\gamma_j - \gamma_i + \gamma_b - \gamma_a = (M(aj10, ***) + M(ib10, ***))\mathbf{v} = (M(ij10, ***) + M(ab10, ***))\mathbf{v} < \gamma_j - \gamma_i + \gamma_b - \gamma_a,$$

a contradiction.

Therefore, $M(ij10, ***) \in M_{\mathbf{v}}$. $\square$

Lemma 11. *For any $i \in [n]$, we have*

$$b^{\mathbf{v}}_{i00} + b^{\mathbf{v}}_{i01} > 0 \quad \textit{and} \quad b^{\mathbf{v}}_{i10} + b^{\mathbf{v}}_{i11} > 0.$$

Equivalently, at least one of $b^{\mathbf{v}}_{i00}, b^{\mathbf{v}}_{i01}$ and at least one of $b^{\mathbf{v}}_{i10}, b^{\mathbf{v}}_{i11}$ must be equal to 1.

Proof. We prove the first statement; the second follows by an analogous argument.

Suppose, by contradiction, that $b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = 0$. By Lemma 9, this implies that $M(i000, ***) \in M_{\mathbf{v}}$.

Moreover, $M(i000, ***)$ is the only row of $M_{\mathbf{v}}$ that can contain a 1 in either column $M_{\mathbf{v}}(****, i00)$ or $M_{\mathbf{v}}(****, i01)$. Indeed, since $b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = 0$, no row of $M_{\mathbf{v}}(**01 \cup **10, ***)$ contains a 1 in these columns. By the structure of M, the only remaining cells of M that contain a 1 in columns $M_{\mathbf{v}}(****, i00)$ or $M_{\mathbf{v}}(****, i01)$ are $M(**00, i*00 \cup i*01)$. By Lemma 7, among these rows only $M(i000, ***)$ can belong to $M_{\mathbf{v}}$.

Now consider the vector

$$\mathbf{t} = \mathbf{e}_{i00} - \mathbf{e}_{i01}.$$

Since the only row of $M_{\mathbf{v}}$ with nonzero entries in columns $M_{\mathbf{v}}(****, i00)$ or $M_{\mathbf{v}}(****, i01)$ is $M(i000, ***)$, and this row contains a 1 in both columns, we have $M_{\mathbf{v}}\mathbf{t} = \mathbf{0}$. Thus, $\mathbf{t}$ lies in the null space of $M_{\mathbf{v}}$.

Together with the vector $\mathbf{s}$, which spans the null space of $M_{\mathbf{v}}$ (see remark 4), this implies that the null space of $M_{\mathbf{v}}$ is at least two-dimensional

Consequently, $M_{\mathbf{v}}$ cannot have rank $4n - 1$, contradicting the assumption that $\mathbf{v}$ is a vertex.

Therefore, $b^{\mathbf{v}}_{i00} + b^{\mathbf{v}}_{i01} > 0$.

The proof for $b^{\mathbf{v}}_{i10} + b^{\mathbf{v}}_{i11} > 0$ is identical. ☐

Lemma 12. *We have*

$$b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = b^{\mathbf{v}}_{j10} = b^{\mathbf{v}}_{j11} = 1$$

if and only if $i = n - 1$*,* $j = 0$*, and*

$$M((n-1)000, ***),\ M((n-1)011, ***) \in M_{\mathbf{v}}.$$

Proof. We first prove the "⇒" direction.

Suppose

$$b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = b^{\mathbf{v}}_{j10} = b^{\mathbf{v}}_{j11} = 1.$$

By Lemma 10, this implies that

$$M(ij01, ***),\ M(ij10, ***) \in M_{\mathbf{v}}.$$

Hence

$$M(ij01, ***)\mathbf{v} = \gamma_j - \gamma_i, \qquad M(ij10, ***)\mathbf{v} = \gamma_j - \gamma_i,$$

and

$$(M(ij01, ***) + M(ij10, ***))\mathbf{v} = 2(\gamma_j - \gamma_i).$$

By the structure of M (see Remark 3), we have

$$M(ij01, ***) + M(ij10, ***) = M(i000, ***) + M((n-1)j11, ***).$$

Therefore,

$$2(\gamma_j - \gamma_i) = (M(i000, ***) + M((n-1)j11, ***))\mathbf{v} \le (\gamma_0 - \gamma_i) + (\gamma_j - \gamma_{n-1}).$$

Rearranging yields

$$\gamma_j - \gamma_i \le \gamma_0 - \gamma_{n-1}.$$

Since $\gamma_0 - \gamma_{n-1}$ is the maximum possible difference, equality can occur only if $j = 0$ and $i = n - 1$. In this case, equality in the above inequality forces

$$M((n-1)000, ***),\ M((n-1)011, ***) \in M_{\mathbf{v}},$$

because

$$(M((n-1)000, ***) + M((n-1)011, ***))\mathbf{v} = 2(\gamma_0 - \gamma_{n-1}).$$

We now prove the "⇐" direction.

Suppose $i = n-1$, $j = 0$, and

$$M((n-1)000, ***),\ M((n-1)011, ***) \in M_{\mathbf{v}}.$$

By the structure of M, we have

$$M((n-1)001, ***) + M((n-1)010, ***) = M((n-1)000, ***) + M((n-1)011, ***).$$

Evaluating at $\mathbf{v}$ gives

$$(M((n-1)001, ***) + M((n-1)010, ***))\mathbf{v} = 2(\gamma_0 - \gamma_{n-1}).$$

Since each of $M((n-1)001, ***)\mathbf{v}$, $M((n-1)010, ***)\mathbf{v}$ is bounded above by $\gamma_0 - \gamma_{n-1}$, both must attain this value. Thus,

$$M((n-1)001, ***),\ M((n-1)010, ***) \in M_{\mathbf{v}},$$

which implies

$$b^{\mathbf{v}}_{(n-1)00} = b^{\mathbf{v}}_{(n-1)01} = b^{\mathbf{v}}_{010} = b^{\mathbf{v}}_{011} = 1.$$

□

Remark 6. *If $b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01}$ for some $i < n-1$, then for all $j \in [n]$ we have $b^{\mathbf{v}}_{j10} \neq b^{\mathbf{v}}_{j11}$. Similarly, if $b^{\mathbf{v}}_{j10} = b^{\mathbf{v}}_{j11}$ for some $j > 0$, then for all $i \in [n]$ we have $b^{\mathbf{v}}_{i00} \neq b^{\mathbf{v}}_{i01}$.*

Proof. This follows directly from Lemma 12, together with Lemma 8. □

Lemma 13. *There exists at least one $i \in [n]$ and $d \in \{0, 1\}$ such that*

$$b^{\mathbf{v}}_{id0} = b^{\mathbf{v}}_{id1} = 1.$$

Proof. Suppose, by contradiction, that for every $i \in [n]$ and every $d \in \{0, 1\}$, we have

$$b^{\mathbf{v}}_{id0} \neq b^{\mathbf{v}}_{id1}.$$

In particular, for every i,

$$b^{\mathbf{v}}_{i00} + b^{\mathbf{v}}_{i01} = 1 \quad \text{and} \quad b^{\mathbf{v}}_{i10} + b^{\mathbf{v}}_{i11} = 1,$$

by Lemma 11.

For every i, since exactly one of $b^{\mathbf{v}}_{i00}, b^{\mathbf{v}}_{i01}$ is equal to 0, Lemma 9 implies that

$$M(i000, ***) \in M_{\mathbf{v}},$$

and, in particular for $i = n-1$,

$$M((n-1)000, ***) \in M_{\mathbf{v}}. \tag{20}$$

Similarly, since exactly one of $b^{\mathbf{v}}_{i10}, b^{\mathbf{v}}_{i11}$ is equal to 0, again by Lemma 9 we obtain

$$M((n-1)i11, ***) \in M_{\mathbf{v}},$$

and, for the choice of $i = 0$,

$$M((n-1)011, ***) \in M_{\mathbf{v}}. \tag{21}$$

Equations (20) and (21) together with Lemma 12 imply

$$b^{\mathbf{v}}_{(n-1)00} = b^{\mathbf{v}}_{(n-1)01} = b^{\mathbf{v}}_{010} = b^{\mathbf{v}}_{011} = 1,$$

which contradicts the assumption that $b^{\mathbf{v}}_{id0} + b^{\mathbf{v}}_{id1} = 1$ for all i, d.

Therefore, there must exist at least one $i \in [n]$ and $d \in \{0, 1\}$ such that

$$b^{\mathbf{v}}_{id0} = b^{\mathbf{v}}_{id1} = 1.$$

□

Lemma 14. *Let a_{dz} denote the number of indices i such that $b^{\mathbf{v}}_{idz} = 1$. Then*

$$\textit{rank}(M_{\mathbf{v}}(**01, ***)) \leq a_{00} + a_{11} - 1 \quad \textit{and} \quad \textit{rank}(M_{\mathbf{v}}(**10, ***)) \leq a_{01} + a_{10} - 1.$$

Proof. We prove the bound for $M_{\mathbf{v}}(**01, ***)$; the proof for $M_{\mathbf{v}}(**10, ***)$ is identical.

By definition of $\mathbf{b}^{\mathbf{v}}$, the submatrix $M_{\mathbf{v}}(**01, ***)$ has nonzero entries only in columns indexed by $(i00)$ and $(i11)$ with $b^{\mathbf{v}}_{i00} = 1$ or $b^{\mathbf{v}}_{i11} = 1$ (as all entries in $M_{\mathbf{v}}(**10, *00 \cup *11)$ are equal to 0). Hence, $M_{\mathbf{v}}(**01, ***)$ has at most $a_{00} + a_{11}$ nonzero columns.

Moreover, by the structure of M, every row of $M_{\mathbf{v}}(**01, ***)$ contains exactly one entry equal to 1 in a column of type $(i00)$ and exactly one entry equal to 1 in a column of type $(j11)$. Consequently, the sum of all columns of type $(i00)$ is equal to the sum of all columns of type $(j11)$. This yields a nontrivial linear dependence among the columns of $M_{\mathbf{v}}(**01, ***)$.

Therefore, the column space of $M_{\mathbf{v}}(**01, ***)$ has dimension at most $a_{00} + a_{11} - 1$, and hence

$$\text{rank}(M_{\mathbf{v}}(**01, ***)) \leq a_{00} + a_{11} - 1.$$

The same argument, applied to columns of type $(i01)$ and $(i10)$, gives

$$\text{rank}(M_{\mathbf{v}}(**10, ***)) \leq a_{01} + a_{10} - 1.$$

□

Lemma 15. *For any $d, z \in \{0, 1\}$, there exists an index $i \in [n]$ such that*

$$b^{\mathbf{v}}_{idz} = 1.$$

Proof. It suffices to prove the statement for $(d, z) = (0, 0)$; the other cases follow by symmetry.

Suppose, by contradiction, that

$$b^{\mathbf{v}}_{i00} = 0 \quad \text{for all } i \in [n].$$

By Lemma 10, this implies that no row of the form $M(i'j'01, ***)$ can belong to $M_{\mathbf{v}}$. Consequently, $M_{\mathbf{v}}(**01, ***)$ is empty and for every $j \in [n]$,

$$b^{\mathbf{v}}_{j11} = 0.$$

We now bound the rank of $M_{\mathbf{v}}$. First, submatrix $M_{\mathbf{v}}(**01, ***)$ is empty and contributes zero to the rank. By Lemma 14, the submatrix $M_{\mathbf{v}}(**10, ***)$ has rank at most

$$a_{01} + a_{10} - 1 \leq 2n - 1.$$

Next, by Lemma 7, at most $2n$ rows from $M(**00 \cup **11, ***)$ can belong to $M_{\mathbf{v}}$. Therefore, the total rank of $M_{\mathbf{v}}$ is at most

$$(2n - 1) + 2n = 4n - 1.$$

Equality can occur only if *all* such rows belong to $M_{\mathbf{v}}$.

In particular, this forces

$$M((n-1)000, ***),\ M((n-1)011, ***) \in M_{\mathbf{v}}.$$

By Lemma 12, this implies

$$b^{\mathbf{v}}_{(n-1)00} = b^{\mathbf{v}}_{(n-1)01} = b^{\mathbf{v}}_{010} = b^{\mathbf{v}}_{011} = 1,$$

which contradicts the assumption that $b^{\mathbf{v}}_{i00} = 0$ for all i.

Hence, there exists at least one $i \in [n]$ such that $b^{\mathbf{v}}_{i00} = 1$. The same argument applies to any choice of (d, z). □

Lemma 16. *Suppose that $b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = 1$ for some $i \in [n]$. Then for every j with $i < j \leq n - 1$, we have*

$$b^{\mathbf{v}}_{j00} = b^{\mathbf{v}}_{j01} = 1 \quad \textit{and} \quad M(j000, ***) \notin M_{\mathbf{v}}.$$

Similarly, if $b^{\mathbf{v}}_{i10} = b^{\mathbf{v}}_{i11} = 1$ for some $i \in [n]$, then for every j with $0 \leq j < i$,

$$b^{\mathbf{v}}_{j10} = b^{\mathbf{v}}_{j11} = 1 \quad \textit{and} \quad M((n-1)j11, ***) \notin M_{\mathbf{v}}.$$

Proof. We prove the first statement; the second follows by symmetry.

Suppose $b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = 1$ for some i. If $i = n-1$, there is nothing to prove, so assume $i < n-1$.

By Remark 6, it is impossible that $b^{\mathbf{v}}_{t10} = b^{\mathbf{v}}_{t11} = 1$ for any t. Hence, for every $t \in [n]$, by Lemma 11 exactly one of $b^{\mathbf{v}}_{t10}, b^{\mathbf{v}}_{t11}$ is equal to 0. By Lemma 9, this implies

$$M((n-1)t11, ***) \in M_{\mathbf{v}} \quad \text{for all } t \in [n].$$

By Lemma 15, there exist indices a, b such that

$$b^{\mathbf{v}}_{a10} = 1 \quad \text{and} \quad b^{\mathbf{v}}_{b11} = 1.$$

Since $b^{\mathbf{v}}_{i01} = 1$ and $b^{\mathbf{v}}_{a10} = 1$, and $b^{\mathbf{v}}_{i00} = 1$ and $b^{\mathbf{v}}_{b11} = 1$, Lemma 10 implies

$$M(ia10, ***),\ M(ib01, ***) \in M_{\mathbf{v}},$$

which yields

$$M(ia10, ***)\mathbf{v} = \gamma_a - \gamma_i, \qquad M(ib01, ***)\mathbf{v} = \gamma_b - \gamma_i.$$

Adding the sides gives

$$v_{i00} + v_{i01} + v_{a10} + v_{b11} = \gamma_a + \gamma_b - 2\gamma_i. \tag{22}$$

Also we have the inequality

$$M(i000, ***)\mathbf{v} = v_{i00} + v_{i01} \le \gamma_0 - \gamma_i,$$

which implies

$$v_{a10} + v_{b11} \ge \gamma_a + \gamma_b - \gamma_i - \gamma_0. \tag{23}$$

Now fix any j with $i < j \le n-1$. Suppose, by contradiction, that $b^{\mathbf{v}}_{j00} = 0$ or $b^{\mathbf{v}}_{j01} = 0$. Then by Lemma 9,

$$M(j000, ***) \in M_{\mathbf{v}},$$

so

$$v_{j00} + v_{j01} = \gamma_0 - \gamma_j. \tag{24}$$

Using inequalities for the rows $M(ja10, ***)$ and $M(jb01, ***)$, we obtain

$$v_{j01} + v_{a10} \le \gamma_a - \gamma_j, \qquad v_{j00} + v_{b11} \le \gamma_b - \gamma_j.$$

Adding and using (24) gives

$$v_{a10} + v_{b11} \le \gamma_a + \gamma_b - \gamma_j - \gamma_0. \tag{25}$$

Combining (23) and (25) yields

$$\gamma_a + \gamma_b - \gamma_i - \gamma_0 \le v_{a10} + v_{b11} \le \gamma_a + \gamma_b - \gamma_j - \gamma_0,$$

which is impossible since $\gamma_j > \gamma_i$. Hence,

$$b^{\mathbf{v}}_{j00} = b^{\mathbf{v}}_{j01} = 1.$$

Finally, if $M(j000, ***) \in M_{\mathbf{v}}$ while $b^{\mathbf{v}}_{j00} = b^{\mathbf{v}}_{j01} = 1$, then equality holds in (25), which again contradicts (23) because $\gamma_j > \gamma_i$. Therefore,

$$M(j000, ***) \notin M_{\mathbf{v}}.$$

This completes the proof of the first statement. The second statement follows from a symmetric argument.

□

Lemma 17. *Suppose $0 \le i < n-1$ is the smallest index such that*

$$b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = 1.$$

Then $M(i000, ***) \in M_{\mathbf{v}}$. Similarly, if $0 < j \le n-1$ is the largest index such that

$$b^{\mathbf{v}}_{j10} = b^{\mathbf{v}}_{j11} = 1,$$

then $M((n-1)j11, ***) \in M_{\mathbf{v}}$.

Proof. First, we prove the first statement.

Let $i < n-1$ be the smallest index such that $b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = 1$. By Remark 6, this implies that

$$b^{\mathbf{v}}_{y10} + b^{\mathbf{v}}_{y11} = 1 \quad \text{for all } y \in [n]. \tag{26}$$

By minimality of i and Lemma 16, we have

$$b^{\mathbf{v}}_{j00} + b^{\mathbf{v}}_{j01} = \begin{cases} 1, & 0 \le j < i, \\ 2, & i \le j \le n-1. \end{cases} \tag{27}$$

Let a_{dz} denote the number of indices k with $b^{\mathbf{v}}_{kdz} = 1$. From (27) we obtain

$$a_{00} + a_{01} = i + 2(n-i) = 2n - i, \qquad a_{10} + a_{11} = n$$

by (26). Hence, by Lemma 14,

$$\operatorname{rank}\big(M_{\mathbf{v}}(**01 \cup **10, ***)\big) \le (a_{00} + a_{01}) + (a_{10} + a_{11}) - 2 = 3n - i - 2. \tag{28}$$

By Lemma 16, no row $M(j000, ***)$ with $j > i$ can belong to $M_{\mathbf{v}}$.

Suppose, by contradiction, that $M(i000, ***) \notin M_{\mathbf{v}}$. Then the submatrix $M_{\mathbf{v}}(**00 \cup **11, ***)$ contains at most $n+i$ rows and therefore has rank at most $n+i$. Combining with (28), we obtain

$$\operatorname{rank}(M_{\mathbf{v}}) \le (3n - i - 2) + (n + i) = 4n - 2,$$

which contradicts the fact that $\mathbf{v}$ is a vertex and hence $\operatorname{rank}(M_{\mathbf{v}}) = 4n - 1$.

Therefore, $M(i000, ***) \in M_{\mathbf{v}}$.

The proof of the second statement is symmetric. Let $0 < j \le n-1$ be the largest index such that $b^{\mathbf{v}}_{j10} = b^{\mathbf{v}}_{j11} = 1$. Since $j > 0$, by Remark 6 we cannot have $b^{\mathbf{v}}_{t00} = b^{\mathbf{v}}_{t01} = 1$ for any t, and hence

$$b^{\mathbf{v}}_{t00} + b^{\mathbf{v}}_{t01} = 1 \quad \text{for all } t \in [n]. \tag{29}$$

By maximality of j and Lemma 16, we have

$$b^{\mathbf{v}}_{t10} + b^{\mathbf{v}}_{t11} = \begin{cases} 2, & t \le j, \\ 1, & t > j. \end{cases} \tag{30}$$

Let a_{dz} be as above. Then (29) gives $a_{00} + a_{01} = n$, and (30) yields

$$a_{10} + a_{11} = 2(j+1) + (n - (j+1)) = n + j + 1.$$

Therefore, by Lemma 14,

$$\operatorname{rank}\big(M_{\mathbf{v}}(**01 \cup **10, ***)\big) \le (a_{00} + a_{01}) + (a_{10} + a_{11}) - 2 = (n) + (n + j + 1) - 2 = 2n + j - 1. \tag{31}$$

By Lemma 16, no row $M((n-1)t11, ***)$ with $t < j$ can belong to $M_{\mathbf{v}}$.

If $M((n-1)j11, ***) \notin M_{\mathbf{v}}$, then $M_{\mathbf{v}}(**00 \cup **11, ***)$ contains at most $n + (n - (j+1)) = 2n - j - 1$ rows, hence has rank at most $2n - j - 1$. Combining with (31) gives

$$\operatorname{rank}(M_{\mathbf{v}}) \le (2n + j - 1) + (2n - j - 1) = 4n - 2,$$

contradicting $\operatorname{rank}(M_{\mathbf{v}}) = 4n - 1$ at a vertex. Thus $M((n-1)j11, ***) \in M_{\mathbf{v}}$.

□

Lemma 18. *The index $i = n - 1$ is the smallest index such that*

$$b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = 1$$

if and only if the index $j = 0$ is the largest index such that

$$b^{\mathbf{v}}_{j10} = b^{\mathbf{v}}_{j11} = 1.$$

Proof. It suffices to prove one direction, since the statement is symmetric.

Assume that $i = n - 1$ is the smallest index such that $b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = 1$. By Remark 6, this implies that there is no index $j > 0$ for which $b^{\mathbf{v}}_{j10} = b^{\mathbf{v}}_{j11} = 1$.

Suppose, by contradiction, that $b^{\mathbf{v}}_{010} \neq 1$ or $b^{\mathbf{v}}_{011} \neq 1$. Then $b^{\mathbf{v}}_{010} + b^{\mathbf{v}}_{011} = 1$. For all $j \in [n]$, we therefore have $b^{\mathbf{v}}_{j10} + b^{\mathbf{v}}_{j11} = 1$. By Lemma 14, this implies

$$\text{rank}(M_{\mathbf{v}}(**01 \cup **10, ***)) \leq 2n - 1.$$

Moreover, since there are at most $2n$ rows in $M_{\mathbf{v}}(**00 \cup **11, ***)$, we have

$$\text{rank}(M_{\mathbf{v}}(**00 \cup **11, ***)) \leq 2n.$$

Consequently,

$$\text{rank}(M_{\mathbf{v}}) \leq 4n - 1,$$

and equality can hold only if all such rows belong to $M_{\mathbf{v}}$. In particular, this forces

$$M((n-1)000, ***),\ M((n-1)011, ***) \in M_{\mathbf{v}},$$

which implies $b^{\mathbf{v}}_{010} = b^{\mathbf{v}}_{011} = 1$ by Lemma 12. This contradicts the assumption.

Therefore, $b^{\mathbf{v}}_{010} = b^{\mathbf{v}}_{011} = 1$, and hence $j = 0$ is the largest index such that $b^{\mathbf{v}}_{j10} = b^{\mathbf{v}}_{j11} = 1$. □

B.4 BIJECTION BETWEEN VERTICES AND THEIR SIGNATURE

Here we show that there is a bijection between vertices and their signatures. Consequently, the vertices can be enumerated via their signatures.

Proposition 3. *The active constraint matrix $M_{\mathbf{v}}$ uniquely determines $\mathbf{b}^{\mathbf{v}}$, and conversely, $\mathbf{b}^{\mathbf{v}}$ uniquely determines $M_{\mathbf{v}}$.*

Proof. One can construct $\mathbf{b}^{\mathbf{v}}$ from $M_{\mathbf{v}}$ by Definition 10.

Now suppose that we are given $\mathbf{b}^{\mathbf{v}}$ for some vertex $\mathbf{v}$ and wish to recover $M_{\mathbf{v}}$. First, by Lemma 10, the submatrix $M_{\mathbf{v}}(**01 \cup **10, ***)$ is uniquely determined by $\mathbf{b}^{\mathbf{v}}$.

It remains to determine which of the $2n$ rows in $M(**00 \cup **11, ***)$ belong to $M_{\mathbf{v}}$. By Lemma 13, there exists an index $y \in [n]$ and $d \in \{0, 1\}$ such that

$$b^{\mathbf{v}}_{yd0} = b^{\mathbf{v}}_{yd1} = 1.$$

We first consider the case $d = 0$ and let i be the smallest index such that $b^{\mathbf{v}}_{i00} = b^{\mathbf{v}}_{i01} = 1$.

If $i = n - 1$, then by Lemma 18 we have $b^{\mathbf{v}}_{010} = b^{\mathbf{v}}_{011} = 1$, and by Lemma 12 it follows that

$$M((n-1)000, ***),\ M((n-1)011, ***) \in M_{\mathbf{v}}.$$

Moreover, for any $0 \leq i' < n - 1$ and $0 < j' \leq n - 1$, we have $b^{\mathbf{v}}_{i'00} + b^{\mathbf{v}}_{i'01} = 1$ and $b^{\mathbf{v}}_{j'10} + b^{\mathbf{v}}_{j'11} = 1$ (by Remark 6), so exactly one entry in each pair is zero. By Lemma 9, this implies

$$M(i'000, ***),\ M((n-1)j'11, ***) \in M_{\mathbf{v}}.$$

Hence, $M_{\mathbf{v}}$ is uniquely determined by $\mathbf{b}^{\mathbf{v}}$ in this case.

Now suppose that $i < n-1$. By Remark 6, for every $j \in [n]$ we have $b^{\mathbf{v}}_{j10} + b^{\mathbf{v}}_{j11} = 1$, and hence exactly one of them is zero. By Lemma 9, this implies

$$M((n-1)j11, ***) \in M_{\mathbf{v}} \quad \text{for all } j \in [n].$$

Furthermore, by Lemma 16, for all $i' > i$ we have $M(i'000, ***) \notin M_{\mathbf{v}}$. By minimality of i, for all $i' < i$ we have $b^{\mathbf{v}}_{i'00} + b^{\mathbf{v}}_{i'01} = 1$, and hence $M(i'000, ***) \in M_{\mathbf{v}}$. Finally, by Lemma 17, we also have $M(i000, ***) \in M_{\mathbf{v}}$. Thus, $M_{\mathbf{v}}$ is uniquely determined by $\mathbf{b}^{\mathbf{v}}$ in this case as well.

We now consider the case $d = 1$. Let j be the largest index such that $b^{\mathbf{v}}_{j10} = b^{\mathbf{v}}_{j11} = 1$.

If $j = 0$, then by Lemma 18 we reduce to the first case considered above, and $M_{\mathbf{v}}$ can again be uniquely constructed from $\mathbf{b}^{\mathbf{v}}$.

Finally, suppose that $j > 0$. By Remark 6, for every $i \in [n]$ we have $b^{\mathbf{v}}_{i00} + b^{\mathbf{v}}_{i01} = 1$, and hence exactly one of them is zero. By Lemma 9, this implies

$$M(i000, ***) \in M_{\mathbf{v}} \quad \text{for all } i \in [n].$$

Moreover, by Lemma 16, for all $j' < j$ we have $M((n-1)j'11, ***) \notin M_{\mathbf{v}}$, while by maximality of j and Lemma 17 we have $M((n-1)j11, ***) \in M_{\mathbf{v}}$. For $j' > j$, we have $b^{\mathbf{v}}_{j'10} + b^{\mathbf{v}}_{j'11} = 1$, and hence $M((n-1)j'11, ***) \in M_{\mathbf{v}}$ by Lemma 9.

Therefore, in all cases, $M_{\mathbf{v}}$ is uniquely determined by $\mathbf{b}^{\mathbf{v}}$. This completes the proof. $\square$

Remark 7. *Since $[\mathbf{v}]$ and $M_{\mathbf{v}}$ uniquely determine each other, it follows that $[\mathbf{v}]$ uniquely determines $\mathbf{b}^{\mathbf{v}}$, and conversely $\mathbf{b}^{\mathbf{v}}$ uniquely determines $[\mathbf{v}]$.*

We introduce the following set and claim that it precisely consists of all signatures associated with vertices.

Definition 4 (Admissible signatures)**.** *The set of admissible signatures, denoted by S, is defined as $S = S_1 \cup S_2 \cup S_3$, where each S_i is a set of binary vectors $\mathbf{b} \in \mathbb{R}^{n\times 2\times 2}$ s.t.*

1. $\mathbf{b} \in S_1$ *iff*
   - $\exists\, t \in [n-1]$ *such that* $b_{i00} = b_{i01} = 1$ *for all* $i \geq t$ *and* $b_{i00} \neq b_{i01}$ *for all* $i < t$.
   - *For all* $i \in [n]$, $b_{i10} \neq b_{i11}$.
   - *There exist* $i, j \in [n]$ *such that* $b_{i10} = b_{j11} = 1$.
2. $\mathbf{b} \in S_2$ *iff*
   - $b_{(n-1)00} = b_{(n-1)01} = b_{010} = b_{011} = 1$.
   - *For all* $0 \leq i < n-1$, $b_{i00} \neq b_{i01}$.
   - *For all* $0 < j \leq n-1$, $b_{j10} \neq b_{j11}$.
3. $\mathbf{b} \in S_3$ *iff*
   - $\exists\, t \in [n-1]$ *such that* $b_{i10} = b_{i11} = 1$ *for all* $i \leq t+1$ *and* $b_{i10} \neq b_{i11}$ *for all* $i > t+1$.
   - *For all* $i \in [n]$, $b_{i00} \neq b_{i01}$.
   - *There exist* $i, j \in [n]$ *such that* $b_{i00} = b_{j01} = 1$.

Proposition 4. $\mathbf{b}^{\mathbf{v}} \in S$ *for any vertex* $\mathbf{v}$.

Proof. Suppose that some $0 \leq t < n-1$ is the smallest index such that $b^{\mathbf{v}}_{t00} = b^{\mathbf{v}}_{t01}$. Then Lemma 16, Remark 6, and Lemma 15 imply that the first, second, and third conditions of S_1 hold, respectively. Thus $\mathbf{b}^{\mathbf{v}} \in S_1$.

Suppose that some $0 < t \leq n-1$ is the largest index such that $b^{\mathbf{v}}_{t10} = b^{\mathbf{v}}_{t11}$. Then Lemma 16, Remark 6, and Lemma 15 imply that the first, second, and third conditions of S_3 hold, respectively. Thus $\mathbf{b}^{\mathbf{v}} \in S_3$.

Finally, suppose that $b^{\mathbf{v}}_{i00} \neq b^{\mathbf{v}}_{i01}$ for all $0 \leq i < n-1$ and that $b^{\mathbf{v}}_{j10} \neq b^{\mathbf{v}}_{j11}$ for all $0 < j \leq n-1$. By Lemma 13, at least one of $b^{\mathbf{v}}_{010} = b^{\mathbf{v}}_{011}$ or $b^{\mathbf{v}}_{(n-1)00} = b^{\mathbf{v}}_{(n-1)01}$ must hold. If one holds, then Lemma 18 implies the other, and hence the first condition of S_2 is satisfied. Moreover, Lemma 12 implies that the remaining two conditions of S_2 also hold. Therefore $\mathbf{b}^{\mathbf{v}} \in S_2$. $\square$

Definition 11. *A vector $\mathbf{b} \in \{0,1\}^{4n}$ is called* valid *if there exists a vertex $\mathbf{v}$ such that $\mathbf{b} = \mathbf{b}^{\mathbf{v}}$.*

Definition 5 (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:*

- $u_{i00} = -\gamma_i - \alpha$ *if* $b_{i00} = 1$, *and* $u_{i00} = \gamma_0$ *otherwise;*
- $u_{i10} = \gamma_i$ *if* $b_{i10} = 1$, *and* $u_{i10} = -\gamma_{n-1} - \alpha$ *otherwise;*
- $u_{i01} = -\gamma_i$ *if* $b_{i01} = 1$, *and* $u_{i01} = \gamma_0 + \alpha$ *otherwise;*
- $u_{i11} = \gamma_i + \alpha$ *if* $b_{i11} = 1$, *and* $u_{i11} = -\gamma_{n-1}$ *otherwise,*

where

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

Proposition 5. *Any $\mathbf{b} \in S$ is valid and its corresponding vertex is $\mathbf{u}(\mathbf{b})$.*

Proof. In this proof, for every $\mathbf{b} \in S$, we prove that vector $\mathbf{u}(\mathbf{b})$ satisfies all the required equalities and strict inequalities. In particular, we prove that

$$M\mathbf{u}(\mathbf{b}) \le \mathbf{c} \quad \text{and} \quad \mathbf{b} = \mathbf{b}^{\mathbf{u}(\mathbf{b})},$$

which results the feasibility of $\mathbf{u}(\mathbf{b})$ and its uniqueness with respect to $\mathbf{b}$. Finally, we verify that $\mathbf{u}(\mathbf{b})$ is a vertex by showing that

$$\text{rank}(M_{\mathbf{u}(\mathbf{b})}) = 4n - 1,$$

in view of Lemma 6. During this proof, $\mathbf{u}$ is used as shorthand for $\mathbf{u}(\mathbf{b})$.

Relevant constraints. Similar to Lemma 7, it is enough to check the inequality for rows

$$M(i000, ***) \quad \text{and} \quad M((n-1)j11, ***)$$

among the rows of $M(**00 \cup **11, ***)$, because all other rows correspond to strictly weaker inequalities. Therefore, it suffices to verify the following inequalities:

- $M(i000, ***)\mathbf{u} = u_{i00} + u_{i01} \le \gamma_0 - \gamma_i$;
- $M((n-1)j11, ***)\mathbf{u} = u_{j10} + u_{j11} \le \gamma_j - \gamma_{n-1}$;
- $M(ij01, ***)\mathbf{u} = u_{i00} + u_{j11} \le \gamma_j - \gamma_i$;
- $M(ij10, ***)\mathbf{u} = u_{i01} + u_{j10} \le \gamma_j - \gamma_i$,

for all relevant indices.

Equality structure. If $b_{i00} = b_{j11} = 1$, then

$$u_{i00} + u_{j11} = \gamma_j - \gamma_i,$$

so the inequality corresponding to $M(ij01, ***)$ holds with equality, independently of the value of α. If $b_{i01} = b_{j10} = 1$, then

$$u_{i01} + u_{j10} = \gamma_j - \gamma_i,$$

so the inequality corresponding to $M(ij10, ***)$ also holds with equality.

To establish strict inequality whenever at least one of the corresponding $\mathbf{b}$-entries is zero, we verify that each component u_{ydz} strictly increases when b_{ydz} takes the value 1 instead of 0. Since equality holds when both corresponding $\mathbf{b}$-entries are equal to 1, this implies that the sum is strictly smaller than $\gamma_j - \gamma_i$ whenever at least one of them is zero.

Case $\mathbf{b} \in S_1$. Let $0 \le t < n-1$ be the smallest index such that $b_{t00} = b_{t01} = 1$. Set

$$\alpha = -\gamma_0 - \gamma_t.$$

- For $i < t$, $b_{i00} \neq b_{i01}$, and hence $u_{i00} + u_{i01} = \gamma_0 - \gamma_i$. For $i \geq t$, $b_{i00} = b_{i01} = 1$, and

$$u_{i00} + u_{i01} = -2\gamma_i - \alpha = \gamma_0 - \gamma_i + (\gamma_t - \gamma_i).$$

  Thus equality holds if and only if $i = t$, and the inequality is strict for all $i > t$ (as $\gamma_t - \gamma_i < 0$). Therefore,

$$M(i000, ***) \in M_{\mathbf{u}} \iff i \leq t.$$

- For all $j \in [n]$, $b_{j10} \neq b_{j11}$. Consequently,

$$u_{j10} + u_{j11} = \gamma_j - \gamma_{n-1},$$

  so equality holds for all j, and

$$M((n-1)j11, ***) \in M_{\mathbf{u}} \quad \text{for all } j \in [n].$$

- If $b_{i00} = 0$, which occurs only for $i < t$, then

$$-\gamma_i - \alpha > \gamma_0,$$

  which shows that u_{i00} is strictly larger when $b_{i00} = 1$. If $b_{j11} = 0$, then

$$\gamma_j + \alpha > -\gamma_{n-1},$$

  which shows that u_{j11} is strictly smaller when $b_{j11} = 0$. Therefore,

$$M(ij01, ***) \in M_{\mathbf{u}} \iff b_{i00} = b_{j11} = 1.$$

- If $b_{i01} = 0$, then

$$-\gamma_i > \gamma_0 + \alpha,$$

  so u_{i01} is strictly larger when $b_{i01} = 1$. If $b_{j10} = 0$, then

$$\gamma_j > -\gamma_{n-1} - \alpha,$$

  so u_{j10} is strictly smaller when $b_{j10} = 0$. Hence,

$$M(ij10, ***) \in M_{\mathbf{u}} \iff b_{i01} = b_{j10} = 1.$$

Thus $M\mathbf{u} \leq \mathbf{c}$ and $\mathbf{b}^{\mathbf{u}} = \mathbf{b}$ for $\mathbf{b} \in S_1$.

Case $\mathbf{b} \in S_2$. Set $\alpha = -\gamma_0 - \gamma_{n-1}$.

- For $i < n-1$, $b_{i00} \neq b_{i01}$, so $u_{i00} + u_{i01} = \gamma_0 - \gamma_i$. For $i = n-1$, $b_{(n-1)00} = b_{(n-1)01} = 1$, and hence $u_{(n-1)00} + u_{(n-1)01} = \gamma_0 - \gamma_{n-1}$. Therefore,

$$M(i000, ***) \in M_{\mathbf{u}} \quad \text{for all } i \in [n].$$

- For $j > 0$, $b_{j10} \neq b_{j11}$, which gives equality. For $j = 0$, $b_{010} = b_{011} = 1$, which also gives equality. Thus,

$$M((n-1)j11, ***) \in M_{\mathbf{u}} \quad \text{for all } j \in [n].$$

- If $b_{i00} = 0$, then $i < n-1$ and

$$-\gamma_i - \alpha > \gamma_0,$$

  showing that u_{i00} is strictly larger when $b_{i00} = 1$. If $b_{j11} = 0$, then $j > 0$ and

$$\gamma_j + \alpha > -\gamma_{n-1},$$

  showing that u_{j11} is strictly smaller when $b_{j11} = 0$. Hence,

$$M(ij01, ***) \in M_{\mathbf{u}} \iff b_{i00} = b_{j11} = 1.$$

- If $b_{i01} = 0$, then $i < n-1$ and
$$-\gamma_i > \gamma_0 + \alpha,$$
showing that u_{i01} is strictly larger when $b_{i01} = 1$. If $b_{j10} = 0$, then $j > 0$ and
$$\gamma_j > -\gamma_{n-1} - \alpha,$$
showing that u_{j10} is strictly smaller when $b_{j10} = 0$. Hence,
$$M(ij10, ***) \in M_{\mathbf{u}} \iff b_{i01} = b_{j10} = 1.$$

Thus $M\mathbf{u} \le \mathbf{c}$ and $\mathbf{b}^{\mathbf{u}} = \mathbf{b}$ for $\mathbf{b} \in S_2$.

Case $\mathbf{b} \in S_3$. Let $0 < t \le n-1$ be the largest index such that $b_{t10} = b_{t11} = 1$. Set
$$\alpha = -\gamma_t - \gamma_{n-1}.$$

- For all $i \in [n]$, $b_{i00} \neq b_{i01}$, and hence $u_{i00} + u_{i01} = \gamma_0 - \gamma_i$. Therefore,
$$M(i000, ***) \in M_{\mathbf{u}} \quad \text{for all } i \in [n].$$
- For $j > t$, $b_{j10} \neq b_{j11}$ and equality holds. For $j \le t$, $b_{j10} = b_{j11} = 1$, so
$$u_{j10} + u_{j11} = \gamma_j - \gamma_{n-1} + (\gamma_j - \gamma_t).$$
Equality holds only when $j = t$. Hence,
$$M((n-1)j11, ***) \in M_{\mathbf{u}} \iff j \le t.$$
- If $b_{i00} = 0$, then
$$-\gamma_i - \alpha > \gamma_0,$$
which shows that u_{i00} is strictly larger when $b_{i00} = 1$. If $b_{j11} = 0$, then $j > t$ and
$$\gamma_j + \alpha > -\gamma_{n-1},$$
which shows that u_{j11} is strictly smaller when $b_{j11} = 0$. Therefore,
$$M(ij01, ***) \in M_{\mathbf{u}} \iff b_{i00} = b_{j11} = 1.$$
- If $b_{i01} = 0$, then
$$-\gamma_i > \gamma_0 + \alpha,$$
showing that u_{i01} is strictly larger when $b_{i01} = 1$. If $b_{j10} = 0$, then $j > t$ and
$$\gamma_j > -\gamma_{n-1} - \alpha,$$
showing that u_{j10} is strictly smaller when $b_{j10} = 0$. Hence,
$$M(ij10, ***) \in M_{\mathbf{u}} \iff b_{i01} = b_{j10} = 1.$$

Thus $M\mathbf{u} \le \mathbf{c}$ and $\mathbf{b}^{\mathbf{u}} = \mathbf{b}$ for $\mathbf{b} \in S_3$.

It remains to prove that $\operatorname{rank}(M_{\mathbf{u}}) = 4n - 1$ in all cases.

To do so, we show that $\ker(M_{\mathbf{u}})$ is one-dimensional.

Suppose that $\mathbf{s} \in \ker(M_{\mathbf{u}})$. Then
$$M_{\mathbf{u}}\mathbf{s} = 0.$$

Structure coming from $M(ij01, ***)$.

In all cases we have

$$M(ij01,***) \in M_{\mathbf{u}} \iff b_{i00} = b_{j11} = 1.$$

Hence, whenever $b_{i00} = b_{j11} = 1$, we obtain

$$M(ij01,***)\mathbf{s} = s_{i00} + s_{j11} = 0.$$

By the structure of $\mathbf{b}$, there exists some index i_0 such that $b_{i_000} = 1$. Let

$$s_{i_000} = a.$$

Then for every j such that $b_{j11} = 1$, the relation $s_{i_000} + s_{j11} = 0$ implies

$$s_{j11} = -a.$$

Substituting this back into $s_{i00} + s_{j11} = 0$ for any i with $b_{i00} = 1$ gives

$$s_{i00} = a.$$

Therefore,

$$s_{i00} = a \quad \text{if } b_{i00} = 1, \qquad s_{j11} = -a \quad \text{if } b_{j11} = 1.$$

Structure coming from $M(ij10,***)$.

Similarly, in all cases we have

$$M(ij10,***) \in M_{\mathbf{u}} \iff b_{i01} = b_{j10} = 1.$$

Hence, whenever $b_{i01} = b_{j10} = 1$, we obtain

$$M(ij10,***)\mathbf{s} = s_{i01} + s_{j10} = 0.$$

Let

$$s_{i_001} = a'$$

for some i_0 with $b_{i_001} = 1$. Then for every j such that $b_{j10} = 1$ we have

$$s_{j10} = -a',$$

and consequently for every i such that $b_{i01} = 1$,

$$s_{i01} = a'.$$

Thus,

$$s_{i01} = a' \quad \text{if } b_{i01} = 1, \qquad s_{j10} = -a' \quad \text{if } b_{j10} = 1.$$

Determination of the remaining entries.

If $b_{y0z} = 0$ for some $y \in [n]$ and $z \in \{0, 1\}$, then by construction

$$M(y000,***) \in M_{\mathbf{u}}.$$

Hence

$$M(y000,***)\mathbf{s} = s_{y00} + s_{y01} = 0.$$

Since by the structure of $\mathbf{b}$ we know that

$$b_{y00} + b_{y01} > 0,$$

at least one of s_{y00} or s_{y01} is already determined (either equal to a or a'). The equation $s_{y00} + s_{y01} = 0$ therefore uniquely determines the other entry.

Similarly, if $b_{y1z} = 0$, then

$$M((n-1)y11,***) \in M_{\mathbf{u}},$$

and therefore

$$s_{y10} + s_{y11} = 0,$$

which again uniquely determines the remaining component.

Consequently, every component of $\mathbf{s}$ is determined in terms of a and a'.

Case $\mathbf{b} \in S_1$.

In this case, we know that

$$M(t000, ***) \in M_{\mathbf{u}}$$

for the index t defined in S_1. Hence

$$s_{t00} + s_{t01} = 0.$$

Since $b_{t00} = b_{t01} = 1$, we have

$$s_{t00} = a, \qquad s_{t01} = a'.$$

Thus

$$a + a' = 0, \quad \text{so} \quad a = -a'.$$

Therefore all components of $\mathbf{s}$ are determined by a single parameter a, and the null space is one-dimensional.

Case $\mathbf{b} \in S_2$.

Here we know that

$$M((n-1)000, ***) \in M_{\mathbf{u}},$$

so

$$s_{(n-1)00} + s_{(n-1)01} = 0.$$

Since $b_{(n-1)00} = b_{(n-1)01} = 1$, we have

$$s_{(n-1)00} = a, \qquad s_{(n-1)01} = a',$$

which implies

$$a + a' = 0.$$

Again $a = -a'$, so $\mathbf{s}$ depends on a single parameter.

Case $\mathbf{b} \in S_3$.

In this case,

$$M((n-1)t11, ***) \in M_{\mathbf{u}},$$

for the index t defined in S_3. So

$$s_{t10} + s_{t11} = 0.$$

Since $b_{t10} = b_{t11} = 1$, we have

$$s_{t10} = -a', \qquad s_{t11} = -a.$$

Thus

$$-a - a' = 0, \quad \text{so again} \quad a = -a'.$$

In all three cases, $\mathbf{s}$ is uniquely determined up to a scalar multiple. Hence,

$$\dim(\ker(M_{\mathbf{u}})) = 1,$$

and therefore

$$\text{rank}(M_{\mathbf{u}}) = 4n - 1.$$

Finally, we have proven that for each $\mathbf{b} \in S$ there exists a unique active constraint matrix $M_{\mathbf{u}}$ and a unique vector $\mathbf{u}$ such that $\mathbf{u}$ is a vertex. Hence, every $\mathbf{b} \in S$ is valid, and this validity is unique. $\square$

Remark 8. *If* $\mathbf{b}$ *is valid, we can recreate its corresponding vertex in* $\mathcal{O}(n)$*, which indicates the length of the vertex.*

Proof. Given $\mathbf{b}$, we can explicitly construct the corresponding vertex $\mathbf{u}$ using Definition 5. This construction requires only $\mathcal{O}(n)$ time.

Indeed, once $\mathbf{b}$ is given, the only remaining parameter to determine is the value of α. To compute this value, it suffices to find either

- the smallest index t such that $b_{t00} = b_{t01}$, or
- the largest index t such that $b_{t10} = b_{t11}$.

Both quantities can be obtained by a single linear scan of $\mathbf{b}$, and therefore can be computed in $\mathcal{O}(n)$ time. $\square$

B.5 PROOF OF THE THEOREMS 1, 2, AND 3

Theorem 1. *The set* $\mathcal{V} = \{\mathbf{u}(\mathbf{b}) : \mathbf{b} \in S\}$ *is a set of* M*-distinct vertices of* $\mathcal{H}$*, and any other vertex of* $\mathcal{H}$ *is* M*-equivalent to a member of* $\mathcal{V}$*.*

Proof. By proposition 4, every vertex has a signature in S. Therefore, any vertex is M-equivalent to a member of $\mathcal{V}$. Conversely, by proposition 5, every $\mathbf{b} \in S$ has an associated vertex, and this vertex is unique. Indeed, no two M-distinct vectors produce the same signature $\mathbf{b}$. Thus, set $\mathcal{V}$ includes M-distinct vertices. As a result, there is a bijection between the elements of S and vertices and the proof is completed. $\square$

Proposition 6. *For every vertex* $\mathbf{v}$ *of the dual feasible set, there exists a distribution* $\mathcal{P}$ *conforming to the IV model (equivalently, a corresponding* $\mathcal{Q}$ *that the pair* $\mathcal{P}, \mathcal{Q}$ *satisfy equations in 4) such that* $\mathbf{v}$ *uniquely attains the maximum (or, in the minimization case, the minimum) value of the dual objective.*

Let $\mathbf{v}$ be a vertex of the feasible set of Equation (10). By the definition of a vertex (see Lemma 6), there exists a collection of inequalities in Equation (10) that are tight at $\mathbf{v}$ and whose corresponding equalities admit a unique vertex solution, namely $\mathbf{v}$ itself (see Remark 5). Since Equation (10) is the dual of Equation (7), each dual inequality corresponds to a primal variable $q_{kl,ij}$ for some $i, j \in \{0, 1\}$ and $k, l \in [n]$.

Proof. Fix a vertex $\mathbf{v}$, and let u denote the number of inequalities in Equation (10) that are tight at $\mathbf{v}$. By construction of the dual via the Lagrangian, each such inequality is associated with a primal variable $q_{y_0y_1,d_0d_1}$.

We define a distribution $\mathcal{Q}$ as follows: for each inequality that holds with equality at $\mathbf{v}$, set the corresponding $q_{y_0y_1,d_0d_1} = \frac{1}{u}$, and set all remaining probabilities in the support of $\mathcal{Q}$ to zero. This defines a valid distribution supported exactly on the tight inequalities at $\mathbf{v}$.

Next, rewrite the dual objective in Equation (10) in terms of $q_{y_0y_1,d_0d_1}$:

$$\begin{aligned}\sum_{d,z\in\{0,1\},\, y\in[n]} p_{yd,z}\, v_{yd,z} = &\sum_{j,l,y} q_{yl,0j}\, v_{y0,0} + \sum_{j,k,y} q_{ky,1j}\, v_{y1,0} \\ &+ \sum_{i,l,y} q_{yl,i0}\, v_{y0,1} + \sum_{i,k,y} q_{ky,i1}\, v_{y1,1},\end{aligned}$$

which can be equivalently expressed as

$$\begin{aligned}&\sum_{k,l} q_{kl,00}\big(v_{k0,0} + v_{k0,1}\big) + \sum_{k,l} q_{kl,01}\big(v_{k0,0} + v_{l1,1}\big) \\ &+ \sum_{k,l} q_{kl,10}\big(v_{l1,0} + v_{k0,1}\big) + \sum_{k,l} q_{kl,11}\big(v_{l1,0} + v_{l1,1}\big).\end{aligned} \tag{32}$$

Each term in Equation (32) is a convex combination of expressions that are individually bounded above (or below) by the inequalities defining the dual feasible set. Since all nonzero coefficients $q_{y_0y_1,d_0d_1}$ equal $\frac{1}{u} > 0$, the dual objective is maximized (or minimized) if and only if every inequality corresponding to a nonzero $q_{y_0y_1,d_0d_1}$ is tight.

By construction of M, this occurs precisely when the set of inequalities that are tight at $\mathbf{v}$ hold with equality. In other words, $\mathbf{u}$ can maximize the dual objective if and only if $M_{\mathbf{v}} \subseteq M_{\mathbf{u}}$. Since $\mathbf{v}$ is a vertex, by Remark 5, $M_{\mathbf{u}}$ has the same basis as $M_{\mathbf{v}}$ and must be equal to it. Thus, the system of equalities represented by $M_{\mathbf{v}}$ admits a unique vertex solution, namely $\mathbf{v}$ itself. Therefore, $\mathbf{v}$ uniquely attains the extreme value of the dual objective, completing the proof. □

Theorem 2. *Under Assumptions 1, 2, and 4, the ATE admits the following valid and sharp bounds:*

$$\max_{\mathbf{v}\in\mathcal{V}} \mathbf{v}^\top \mathbf{p} \;\le\; \text{ATE} \;\le\; -\max_{\mathbf{v}\in\mathcal{V}} \mathbf{v}^\top \bar{\mathbf{p}}. \tag{13}$$

Proof. We have enumerated all M-distinct vertices of the dual feasible set, which is sufficient. Moreover, by Proposition 6, each vertex provides a unique bound. Hence, every vertex contributes a necessary term to the bounds.

Therefore, we obtain

$$\max_{\mathbf{v}\in\mathcal{V}} \mathbf{v}^\top \mathbf{p} = L(\mathcal{P}) \;\le\; \text{ATE} \;\le\; -L(\bar{\mathcal{P}}) = -\max_{\mathbf{v}\in\mathcal{V}} \mathbf{v}^\top \bar{\mathbf{p}}.$$

based on Equations 7 and 8. □

Theorem 3. *The ATE bounds of Equation* (13) *are valid and sharp under Assumptions 1, 3 and 4a.*

Proof. As established in Theorem 1 of Song et al. [2024], the set of admissible pairs $(\mathcal{P}, \mathcal{Q})$ under Assumptions 1, 2, and 4 (denoted by $\mathcal{M}_1$ in Song et al. [2024]) is identical to the set of admissible pairs under Assumptions 1, 3, and 4a (denoted by $\mathcal{M}_2$ in Song et al. [2024]).

More precisely, for every pair $(\mathcal{P}, \mathcal{Q})$ for which there exists a full data law $\mathcal{Q}^{f*}$ satisfying Assumptions 1, 2, and 4, there exists a full data law $\mathcal{Q}^{f'}$ satisfying Assumptions 1, 3, and 4a that induces the same observed distribution $\mathcal{P}$, and vice versa.

Consequently, the sharp identification region for the ATE under the two sets of assumptions is identical. In particular, the bounds in (13), obtained via (7) and (8), remain valid and sharp under Assumptions 1, 3, and 4a. □

C EXTREME RAY ENUMERATION

Note that Equation (12) can be also written as follows:

$$\begin{aligned} 0 \ge \max \quad & \Sigma_{d,z\in\{0,1\},y\in[n]} p_{yd,z} x_{yd,z} \\ \text{s.t.} \quad & x_{k0,0} + x_{k0,1} \le 0 \quad \forall k \in [n] \\ & x_{k0,0} + x_{l1,1} \le 0 \quad \forall k, l \in [n] \\ & x_{l1,0} + x_{k0,1} \le 0 \quad \forall k, l \in [n] \\ & x_{l1,0} + x_{l1,1} \le 0 \quad \forall l \in [n] \end{aligned} \tag{33}$$

In the definition below, we introduce a set of vectors $\mathcal{R}$. We shall shortly prove that $\mathcal{R}$ is exactly the set of vectors that can generate the cone $\mathcal{K} := \{\mathbf{x} \in \mathbb{R}^{2\ell n} : M\mathbf{x} \le \mathbf{0}\}$.

Definition 12. *Let $\mathcal{R}$ be the set of all vectors $\mathbf{r} = \big(r_{ki,j}\big)_{k\in[n],i,j\in\{0,1\}}$ that belong to one of the following families:*

(I) The all-± 1 rays. *One vector corresponding to each $s \in \{-1, 1\}$, defined as*

$$r_{ki,1} = s, \qquad r_{ki,0} = -s, \qquad \forall\, k \in [n],\; i \in \{0, 1\}.$$

(II) Single-entry perturbations. *One vector corresponding to every $(k', i', j') \in [n] \times \{0, 1\} \times \{0, 1\}$ with $(k', i', j') \ne (n-1, 1, 1)$, defined as*

$$\begin{cases} r_{k'i',j'} = -1, \\ r_{ki,j} = 0 \qquad \forall (k, i, j) \ne (k', i', j') \end{cases}$$

(III) One special coordinate set to zero.

$$\begin{cases} r_{k1,0} = -1, \;\; r_{k0,0} = -1 & \forall k \in [n], \\ r_{k1,1} = 1 & \forall k \in [n-1], \\ r_{k0,1} = 1 & \forall k \in [n], \\ r_{(n-1)1,1} = 0. \end{cases}$$

(IV) Rays indexed by a nonzero binary vector $\mathbf{s} \in \{0,1\}^{n-1}$***.*** *One vector corresponding to every* $\mathbf{s} \in \{0,1\}^{n-1} \setminus \{\mathbf{0}\}$*, defined as*

$$\begin{cases} r_{k0,1} = -1, \;\; r_{k0,0} = 0 & \forall k \in [n], \\ r_{k1,0} = s_k & \forall k \in [n-1], \\ r_{k1,1} = -r_{k1,0} & \forall k \in [n], \\ r_{(n-1)1,0} = 0. \end{cases}$$

(V) Rays indexed by a nontrivial binary vector $\mathbf{s} \in \{0,1\}^{n}$***.*** *One vector corresponding to every* $\mathbf{s} \in \{0,1\}^{n} \setminus \{\mathbf{0}, \mathbf{1}\}$*, defined as*

$$\begin{cases} r_{k1,1} = 0, \;\; r_{k1,0} = -1 & \forall k \in [n], \\ r_{k0,1} = s_k & \forall k \in [n], \\ r_{k0,0} = -r_{k0,1} & \forall k \in [n]. \end{cases}$$

(VI) Another family indexed by $\mathbf{s} \in \{0,1\}^{n-1} \setminus \{\mathbf{0}\}$***.*** *One vector corresponding to every* $\mathbf{s} \in \{0,1\}^{n-1} \setminus \{\mathbf{0}\}$*, defined as*

$$\begin{cases} r_{k0,1} = 0, \;\; r_{k0,0} = -1 & \forall k \in [n], \\ r_{k1,1} = s_k & \forall k \in [n-1], \\ r_{k1,0} = -r_{k1,1} & \forall k \in [n], \\ r_{(n-1)1,1} = 0. \end{cases}$$

Proposition 7. *The cone* $\mathcal{K}$ *is generated by cone combination of vectors* $\mathbf{r} \in \mathcal{R}$*.*

In order to prove Proposition 7, we characterize the extreme rays. To do so, we will utilize a few preliminary results presented below.

Definition 13 (Active Constraint Matrix)**.** *For a feasible point* $\mathbf{r}$ *in* $\mathcal{K}$*, the active constraint matrix at* $\mathbf{r}$*, denoted by* $M_{\mathbf{r}}^{(0)}$*, is the maximal (row-) submatrix of* M *such that*

$$M_{\mathbf{r}}^{(0)} \mathbf{r} = \mathbf{0}.$$

Definition 14 (Positive Homogeneity)**.** *A set defined by a system of equalities is positively homogeneous if it is closed under nonnegative scaling, i.e., if* $\mathbf{r}$ *satisfies the defining equalities, then* $\lambda \mathbf{r}$ *also satisfies them for all* $\lambda \geq 0$*.*

Lemma 19. *Let* $\mathbf{r} \neq \mathbf{0}$ *be a feasible point of the cone* $\mathcal{K} := \{\mathbf{x} : M\mathbf{x} \leq \mathbf{0}\}$*. Then* $\mathbf{r}$ *is an extreme ray of* $\mathcal{K}$ *if and only if the active contain matrix at least* $4n-2$ *linearly independent constraints, i.e.,*

$$\mathit{rank}(M_{\mathbf{r}}^{(0)}) \geq 4n-2$$

Proof. $(\Rightarrow)$ Fix an extreme ray $\mathbf{r}$. Suppose by contradiction that $\operatorname{rank}(M_{\mathbf{r}}^{(0)}) < 4n-2$. Then the null space of $M_{\mathbf{r}}^{(0)}$ has dimension strictly larger than $1 + \dim(\ker(M))$. Hence there exists a nonzero vector

$$\mathbf{t} \in \ker(M_{\mathbf{r}}^{(0)}) \setminus \operatorname{span}(\mathbf{r}, \ker(M)) \,.$$

Since $\mathbf{t} \in \ker(M_{\mathbf{r}}^{(0)})$, all inequalities that are tight at $\mathbf{r}$ remain tight along the direction $\mathbf{t}$. Therefore, for sufficiently small $\epsilon > 0$,

$$M(\mathbf{r} + \epsilon \mathbf{t}) \leq \mathbf{0}, \qquad M(\mathbf{r} - \epsilon \mathbf{t}) \leq \mathbf{0},$$

so both $\mathbf{r} + \epsilon\mathbf{t}$ and $\mathbf{r} - \epsilon\mathbf{t}$ belong to $\mathcal{K}$. Moreover,

$$\mathbf{r} = \tfrac{1}{2}(\mathbf{r} + \epsilon\mathbf{t}) + \tfrac{1}{2}(\mathbf{r} - \epsilon\mathbf{t}).$$

Because $\mathbf{t} \notin \operatorname{span}(\mathbf{r}, \ker(M))$, at least one of the vectors $\mathbf{r} \pm \epsilon\mathbf{t}$ is not contained in $\operatorname{span}(\mathbf{r}, \ker(M))$, contradicting Definition 3. Hence rank$(M_{\mathbf{r}}^{(0)}) \geq 4n - 2$.

($\Leftarrow$) Suppose rank$(M_{\mathbf{r}}^{(0)}) \geq 4n - 2$. Then

$$\dim(\ker(M_{\mathbf{r}}^{(0)})) \leq 1 + \dim(\ker(M)).$$

By construction, $\mathbf{r} \in \ker(M_{\mathbf{r}}^{(0)})$ and $\ker(M) \subseteq \ker(M_{\mathbf{r}}^{(0)})$.

Let $\mathbf{r}_1, \mathbf{r}_2 \in \mathcal{K}$ such that

$$\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2.$$

Applying $M_{\mathbf{r}}^{(0)}$ gives

$$\mathbf{0} = M_{\mathbf{r}}^{(0)}\mathbf{r} = M_{\mathbf{r}}^{(0)}\mathbf{r}_1 + M_{\mathbf{r}}^{(0)}\mathbf{r}_2.$$

Since $\mathbf{r}_1, \mathbf{r}_2 \in \mathcal{K}$, we have $M_{\mathbf{r}}^{(0)}\mathbf{r}_1 \leq \mathbf{0}$ and $M_{\mathbf{r}}^{(0)}\mathbf{r}_2 \leq \mathbf{0}$, hence

$$M_{\mathbf{r}}^{(0)}\mathbf{r}_1 = M_{\mathbf{r}}^{(0)}\mathbf{r}_2 = \mathbf{0},$$

which implies $\mathbf{r}_1, \mathbf{r}_2 \in \ker(M_{\mathbf{r}}^{(0)})$.

If $\mathbf{r} \in \ker(M)$, then $M_{\mathbf{r}}^{(0)} = M$ by definition, and therefore $\mathbf{r}_1, \mathbf{r}_2 \in \ker(M)$, satisfying Definition 3. Otherwise, $\mathbf{r} \notin \ker(M)$. Since $\ker(M) \subseteq \ker(M_{\mathbf{r}}^{(0)})$ and

$$\dim(\ker(M_{\mathbf{r}}^{(0)})) \leq 1 + \dim(\ker(M)),$$

we must have

$$\ker(M_{\mathbf{r}}^{(0)}) = \operatorname{span}(\mathbf{r}, \ker(M))\,.$$

Thus any feasible decomposition of $\mathbf{r}$ lies in $\operatorname{span}(\mathbf{r}, \ker(M))$, which proves that $\mathbf{r}$ is an extreme ray. $\square$

According to Lemma 5 and Lemma 19, we have two sets of extreme rays, one set for those with rank$(M_{\mathbf{r}}^{(0)}) = 4n - 1$ and one set with rank$(M_{\mathbf{r}}^{(0)}) = 4n - 2$.

Remark 9 (Invariant directions and trivial rays)**.** *Based on the structure of the cone $\mathcal{K} := \{\mathbf{x} : M\mathbf{x} \leq \mathbf{0}\}$, if rank$(M_{\mathbf{r}}^{(0)}) =$ rank$(M) = 4n-1$ holds for a vector $\mathbf{r}$, then $M\mathbf{r} = \mathbf{0}$. Therefore, only the following vectors can satisfy rank$(M_{\mathbf{r}}^{(0)}) = 4n-1$:*

$$\{\mathbf{r}_a : \quad a \in \mathbb{R}\},$$

where $\mathbf{r}_a$ is defined as

$$r_{ki,0} = a, r_{ki,1} = -a, \forall k \in [n],\ i \in \{0,1\}.$$

On the other hand, each vector $\mathbf{r}_a$ is a scalar multiple of any other vectors of this kind. As a result, all vectors $\mathbf{r}$ that satisfy rank$(M_{\mathbf{r}}^{(0)}) = 4n - 1$, produce two extreme rays. Convenient choices for these extreme rays are the vectors $\mathbf{r}^{\mathrm{inv}_1}$ and $\mathbf{r}^{\mathrm{inv}_2}$ defined by

$$\begin{cases} r_{ki,0}^{\mathrm{inv}_1} = -1, & \forall k \in [n],\ i \in \{0,1\}, \\ r_{ki,1}^{\mathrm{inv}_1} = 1, & \forall k \in [n],\ i \in \{0,1\}. \end{cases}$$
$$\begin{cases} r_{ki,0}^{\mathrm{inv}_2} = 1, & \forall k \in [n],\ i \in \{0,1\}, \\ r_{ki,1}^{\mathrm{inv}_2} = -1, & \forall k \in [n],\ i \in \{0,1\}. \end{cases} \tag{34}$$

Moreover, since both inequalities $\mathbf{p}^\top \mathbf{r}^{\mathrm{inv}_1} \leq 0$ and

$$-\mathbf{p}^\top \mathbf{r}^{\mathrm{inv}_1} = \mathbf{p}^\top \mathbf{r}^{\mathrm{inv}_2} \leq 0$$

are required to hold, it follows that $\mathbf{p}^\top \mathbf{r}^{\mathrm{inv}_1} = 0$. As $\mathbf{r}^{\mathrm{inv}_1}$ and $\mathbf{r}^{\mathrm{inv}_2}$ span $\ker(M)$*, we have*

$$\mathbf{p}^\top \mathbf{s} = 0, \qquad \forall \mathbf{s} \in \ker(M),$$

whenever $\mathbf{p}$ satisfies Equation (14).

The two extreme rays identified in Remark 9 belong to the family **(I)** in Definition 12.

Lemma 20 (Shift invariance)**.** *Let* $\mathbf{x}$ *be feasible for Equation* (12)*, i.e.,* $M\mathbf{x} \le \mathbf{0}$*. Fix any* $a \in \mathbb{R}$*, and define* $\mathbf{y}$ *by*

$$y_{ki,0} = x_{ki,0} + a, \qquad y_{ki,1} = x_{ki,1} - a, \qquad \forall k \in [n],\ i \in \{0,1\}.$$

Then $\mathbf{y}$ *is also feasible:* $M\mathbf{y} \le \mathbf{0}$*. Moreover,* $M_{\mathbf{x}}^{(0)} = M_{\mathbf{y}}^{(0)}$ *and* $\mathbf{p}^\top\mathbf{x} = \mathbf{p}^\top\mathbf{y}$*; in particular,* $\mathbf{x}$ *and* $\mathbf{y}$ *induce the same constraints on* $\mathbf{p}$*.*

Proof. Each inequality in $M\mathbf{x} \le \mathbf{0}$ involves a sum of one $(\cdot, 0)$-coordinate and one $(\cdot, 1)$-coordinate. Under the transformation above, these sums remain unchanged, hence feasibility is preserved. The equalities $M_{\mathbf{x}}^{(0)} = M_{\mathbf{y}}^{(0)}$ and $\mathbf{p}^\top\mathbf{x} = \mathbf{p}^\top\mathbf{y}$ follow from the same cancellation (see Remark 9). ☐

Remark 10 (Normalization)**.** *By Lemma 20, for the set of extreme rays with rank*$(M_{\mathbf{r}}^{(0)}) = 4n-2$*, we may impose the normalization* $r_{(n-1)1,1} = 0$ *without loss of generality.*

C.1 CHARACTERIZING THE EXTREME RAYS

Under Remark 10, the vector $\mathbf{r}$ has $4n-1$ free (unfixed) coordinates. Throughout the remainder of this appendix, the term *coordinate* refers to these $4n-1$ unfixed entries, excluding $r_{(n-1)1,1}$. By Lemma 19, any nonzero, nontrivial extreme ray $\mathbf{r}$ of $\mathcal{K}$—excluding the case rank$(M_{\mathbf{r}}^{(0)}) = 4n-1$ already treated in Remark 9—satisfies exactly $4n-2$ linearly independent tight equalities. Consequently, the solution space of the linear equation system $M_{\mathbf{r}}^{(0)}\mathbf{r} = \mathbf{0}$ is one–dimensional.

Equivalently, $\mathbf{r}$ is unique up to scaling. In particular, at most one coordinate can remain absent from the active equalities. We therefore classify all extreme rays according to the number of unconstrained coordinates.

C.1.1 Case 1: A Single Unconstrained Coordinate

Suppose that exactly one coordinate does not appear in any equality. Then all remaining $4n-2$ variables are uniquely determined by the linear equation system. Since the system is homogeneous and closed under positive homogeneity, this unique solution should also satisfy positive homogeneity (if $\mathbf{r}$ is the unique solution then $\mathbf{r} = \lambda\mathbf{r}\ \forall\lambda \ge 0$) therefore, this solution have to set all constrained variables equal to zero.

Thus $\mathbf{r} = c\mathbf{e}_i$, where c is constant and $\mathbf{e}_i$ is the standard basis vector corresponding to the unconstrained coordinate. Feasibility of Equation (12) requires $c < 0$, as a result, a convenient choice for these extreme rays are

$$\mathbf{r} = -\mathbf{e}_i.$$

This case contributes exactly $4n-1$ extreme rays belonging to the family **(II)** in Definition 12.

In what follows, we assume that every variable appears in at least one equality of our linear equation system.

Lemma 21. *Every* M*-equivalence class* $\mathcal{E}_{\mathbf{r}}$ *(other than those in Case C.1.1 and Equation* (34)*) has a member, e.g.* $\mathbf{r}$*, that satisfies*

$$\mathbf{r} \in \{-1, 0, 1\}^{4n-1}.$$

Moreover, $\mathbf{r}$ *must follow exactly one of the two sign patterns below:*

$$\textbf{\textit{Pattern A:}}\quad \begin{cases} r_{ki,0} \in \{0, -1\}, \\ r_{ki,1} \in \{0, +1\}, \end{cases} \qquad\qquad \textbf{\textit{Pattern B:}}\quad \begin{cases} r_{k0,0} = 0, \\ r_{k1,0} \in \{0, +1\}, \\ r_{ki,1} \in \{0, -1\}, \end{cases} \tag{35}$$

for all $k \in [n]$ *and* $i \in \{0,1\}$*.*

Note that, based on the feasibility of $\mathbf{r}$, in *Pattern B*, if $r_{ki,1} = 0$ for all possible k and i, then $r_{k1,0}$ have to be zero for all k which yields to a trivial ray $\mathbf{0}$ that is a member of trivial extreme ray Equation (34). Excluding that, these two patterns are disjoint.

Before presenting the formal proof, we provide a high-level outline.

Proof outline. Let $\mathbf{r}$ be an extreme ray not covered by Case C.1.1. The defining equalities form a homogeneous linear system with $4n-1$ variables and rank $4n-2$. Hence, the solution space is one–dimensional and can be parameterized as

$$r_{ki,j} = a_{ki,j}w + b_{k,i,j},$$

for some scalar w.

Homogeneity of the system implies that if $\mathbf{r}$ is feasible, then $\lambda\mathbf{r}$ is feasible for all $\lambda \geq 0$. This forces each nonzero coordinate of $\mathbf{r}$ to have fixed magnitude, which we normalize to 1. Therefore $r_{ki,j} \in \{-1, 0, 1\}$.

Each equality has the form $r_{k_0 i_0,0} + r_{k_1 i_1,1} = 0$, which implies that all nonzero $r_{ki,0}$ share the same sign and all nonzero $r_{ki,1}$ share the opposite sign. Using Remark 10, which enforces $r_{k0,0} \leq 0$, only the two sign patterns in (35) are possible. □

Using the intuition given in the outline, we continue on with the full version of the proof.

Full version of the proof. Consider an extreme ray $\mathbf{r}$ and the system of linear equalities obtained from the inequalities of Equation (12) that are active at $\mathbf{r}$. By Assumption 10 and the feasibility of $\mathbf{r}$, we have

$$r_{(n-1)1,1} + r_{k0,0} \leq 0 \Rightarrow r_{k0,0} \leq 0 \qquad \forall k \in [n].$$

As mentioned before, the solution space of this homogeneous linear system $M\mathbf{r} \leq \mathbf{0}$ is one–dimensional. Consequently, there exists a scalar parameter $w \in \mathbb{R}$ and coefficients $a_{ki,j}, b_{ki,j} \in \mathbb{R}$ such that every solution of the system, and in particular each the extreme ray $\mathbf{r}$, can be written as

$$r_{ki,j} = a_{ki,j}w + b_{ki,j}, \qquad \forall k \in [n],\ i, j \in \{0, 1\}. \tag{36}$$

Now consider any equality of the form

$$r_{k_0 i_0,0} + r_{k_1 i_1,1} = 0$$

appearing in the system. Substituting Equation (36) yields

$$(a_{k_0 i_0,0} + a_{k_1 i_1,1})\, w + (b_{k_0 i_0,0} + b_{k_1 i_1,1}) = 0 \qquad \forall w \in \mathbb{R}.$$

Since this equality must hold for all w, we obtain

$$a_{k_0 i_0,0} + a_{k_1,i_1,1} = 0, \qquad b_{k_0 i_0,0} + b_{k_1 i_1,1} = 0.$$

To show that each nonzero coordinate has the same absolute value, we use the following crucial property of extreme rays: positive homogeneity. Therefore, for any $\lambda \geq 0$, there must exist a scalar w_λ such that

$$\lambda r_{ki,j} = a_{ki,j}w_\lambda + b_{ki,j} \qquad \forall k, i, j.$$

Substituting Equation (36) into the left-hand side gives

$$\lambda a_{ki,j}w + \lambda b_{ki,j} = a_{ki,j}w_\lambda + b_{ki,j}.$$

For indices (k, i, j) such that $a_{ki,j} \neq 0$, this implies

$$w_\lambda = \lambda w + \frac{(\lambda - 1)b_{ki,j}}{a_{ki,j}}. \tag{37}$$

Since Equation (37) holds for any indices that $a_{ki,j} \neq 0$, there exists a constant c independent of (k, i, j) such that

$$\frac{b_{ki,j}}{a_{ki,j}} = c$$

whenever $a_{ki,j} \neq 0$.

On the other hand, if $a_{ki,j} = 0$, then the above homogeneity condition reduces to

$$b_{ki,j} = \lambda b_{ki,j} \qquad \forall \lambda \geq 0,$$

which forces $b_{ki,j} = 0$.

We now justify the normalization of the nonzero coordinates of $\mathbf{r}$. Define a graph G whose vertices correspond to the coordinates (k, i, j) of $\mathbf{r}$, and where an edge connects two vertices $(k_0, i_0, 0)$ and $(k_1, i_1, 1) \neq ((n-1), 1, 1)$ whenever the equality

$$r_{k_0,i_0,0} + r_{k_1,i_1,1} = 0$$

appears in the active constraint system. Also, we have some self loops on vertices $(k_0, i_0, 0)$ whenever we have

$$r_{k_0,i_0,0} + r_{(n-1),1,1} = 0 \Rightarrow r_{k_0,i_0,0} = 0$$

in the active constraint system.

Each connected component of G corresponds to a subset of variables that are linearly linked through the equalities. Therefore, we can separate the whole linear system as several disjoint linear equations, each corresponding to a connected component of G. Since the whole linear system has rank $4n-2$ over $4n-1$ variables, exactly one connected component of G (and its corresponding linear equation) contains one degree of freedom, while all others correspond to uniquely determined variables (since their linear equation system is full ranked, it has a unique solution, which is therefore a zero vector). Equivalently, for all vertices outside this specific component we have $a_{ki,j} = 0$ and hence $r_{ki,j} = 0$.

Within the unique component containing the free variable, the equalities force all variables to be affine functions of the same scalar w. Note that there is no edge of type $r_{k_0,i_0,0} = 0$ in this connected component, due to the fact that each variable has a path with the node corresponding $r_{k_0,i_0,0}$ must have a value equal to zero and uniquely determined. Moreover, along any edge of G, the equality $r_{k_0 i_0,0} + r_{k_1 i_1,1} = 0$ implies that the corresponding coordinates must have the same absolute value. By connectivity, it follows that all nonzero coordinates in this component have equal absolute value.

Since $\mathbf{r}$ is a ray, it is defined only up to positive scaling. Without loss of generality (by rescaling w), we may normalize the ray so that

$$|a_{ki,j} w + b_{ki,j}| = 1 \qquad \text{whenever } a_{ki,j} \neq 0.$$

Consequently, each coordinate of $\mathbf{r}$ satisfies

$$r_{ki,j} \in \{-1, 0, 1\}.$$

We now analyze the sign structure of $\mathbf{r}$. Each equality in the system has the form

$$r_{k_0 i_0,0} + r_{k_1 i_1,1} = 0,$$

which implies that whenever both variables are nonzero, they must have opposite signs. Since we are excluding Case C.1.1, every variable appears in at least one equality. As a result, due to the fact that all nonzero coordinates present in one connected component, so, they all have paths to each other, all variables $r_{ki,0}$ share the same sign, and all nonzero variables $r_{ki,1}$ share the opposite sign (we treat zero as sign–neutral).

Finally, from Remark 10 we have $r_{k0,0} \leq 0$ for all $k \in [n]$. This restricts the possible sign configurations to exactly two cases:

- $r_{k0,0} = 0$ for all k, with $r_{k1,0} \geq 0$ and $r_{ki,1} \leq 0$;
- $r_{ki,0} \leq 0$ and $r_{ki,1} \geq 0$.

These correspond precisely to the two patterns stated in Equation (35), which completes the proof. □

Using Lemma 21, we now enumerate all extreme rays by determining which coordinates can be zero under each admissible sign pattern while satisfying the rank condition of Lemma 19.

C.1.2 Case 2: Pattern B

Suppose Pattern B holds, namely

$$r_{k0,0} = 0, \qquad r_{k1,0} \in \{0, 1\}, \qquad r_{ki,1} \in \{0, -1\}, \quad \forall k \in [n],\ i \in \{0, 1\}.$$

We distinguish two subcases.

Subcase 2.1: $\forall k, i : \ r_{ki,0} = 0$**.** In this subcase, the only potentially nonzero variables are $\{r_{k0,1}, r_{k1,1}\}_{k\in[n]}$. Let

$$A := \{k \in [n] : r_{k0,1} = 0\}, \qquad B := \{k \in [n] : r_{k1,1} = 0\},$$

and denote $a := |A|$ and $b := |B|$.

We prove that the rank of the active constraint matrix in this subcase equals

$$(2n-1) + a + b,$$

by constructing an independent family of this cardinality whose span contains all tight equalities.

Step 1: a spanning family. Define the following sets of equalities: Let $k' \in A$ and $l' \in B$ be arbitrary fixed values.

$$\begin{aligned}
\mathcal{L}_1 &:= \{\, r_{k0,0} + r_{k0,1} = 0 \ : \ k \in A\,\},\\
\mathcal{L}_2 &:= \{\, r_{k'0,0} + r_{l1,1} = 0 \ : \ l \in B\,\} \ \cup \ \{\, r_{k0,0} + r_{l'1,1} = 0 : \ k \in [n] \setminus \{l'\}\,\},\\
\mathcal{L}_3 &:= \{\, r_{l1,0} + r_{l1,1} = 0 \ : \ l \in B\,\},\\
\mathcal{L}_4 &:= \{\, r_{l1,0} + r_{k'0,1} = 0 \ : \ l \notin B\,\}.
\end{aligned}$$

Remark 11. *Note that choosing $k' \in A$ and $l' \in B$ implicitly assumes that both sets are nonempty. We now justify this assumption.*

First, $B \neq \varnothing$ by normalization, since $r_{(n-1)1,1} = 0$ implies $(n-1) \in B$.

Suppose next that $A = \varnothing$. Then $r_{k0,1} \neq 0$ for all $k \in [n]$, and under pattern B we have $r_{k0,1} = -1$ for every $k \in [n]$. In this case, the families $\mathcal{L}_1$ and $\mathcal{L}_4$ are empty and the only tight equalities are of types $r_{k0,0} + r_{l1,1} = 0$ and $r_{l1,0} + r_{l1,1} = 0$.

Consequently, the number of independent equalities is at most

$$(n-1+b) + b \le 2n-1,$$

which is strictly smaller than the extreme ray rank $4n-2$. Hence A must be nonempty. We may therefore fix arbitrary indices $k' \in A$ and $l' \in B$.

Let

$$\mathcal{L} := \mathcal{L}_1 \cup \mathcal{L}_2 \cup \mathcal{L}_3 \cup \mathcal{L}_4.$$

By construction,

$$|\mathcal{L}| = a + (n-1+b) + b + (n-b) = 2n-1+a+b.$$

Step 2: spanning property. We show that any remaining tight equality lies in $\operatorname{span}(\mathcal{L})$. The equalities of type $r_{k0,0}+r_{k0,1} = 0$ and $r_{l1,0} + r_{l1,1} = 0$ are already contained in $\mathcal{L}_1$ and $\mathcal{L}_3$. It is easy to see that any equality in $\{\, r_{k0,0} + r_{l1,1} = 0 \ : \ k \in [n],\ l \in B\,\}$ lies in $\operatorname{span}(\mathcal{L}_2)$. It remains to show that $\{\, r_{k0,1} + r_{l1,0} = 0 \ : \ l \in [n],\ k \in A\,\}$ lies in $\operatorname{span}(\mathcal{L})$. Fix $l \in B$ and $k \in A$. Then

$$(r_{l1,0} + r_{l1,1} = 0) - (r_{k0,0} + r_{l1,1} = 0) \quad \Rightarrow \quad r_{l1,0} - r_{k0,0} = 0.$$

Adding $(r_{k0,0} + r_{k0,1} = 0)$ yields

$$r_{l1,0} + r_{k0,1} = 0.$$

Now, fix $l \notin B$ and $k \in A \setminus \{k'\}$. Then

$$(r_{l1,0} + r_{k'0,1} = 0) - (r_{k'0,0} + r_{k'0,1} = 0) \quad \Rightarrow \quad r_{l1,0} - r_{k'0,0} = 0.$$

Adding $(r_{k'0,0} + r_{l'1,1} = 0) - (r_{k0,0} + r_{l'1,1} = 0)$ yields

$$r_{l1,0} - r_{k0,0} = 0.$$

Adding $(r_{k0,0} + r_{k0,1} = 0)$ yields

$$r_{l1,0} + r_{k0,1} = 0.$$

Hence all equalities of type $r_{l1,0} + r_{k0,1} = 0$ with $k \in A$ are linear combinations of equalities in $\mathcal{L}_1 \cup \mathcal{L}_2 \cup \mathcal{L}_3$.

Step 3: linear independence. Order the variables as

$$\{r_{k0,1} : k \in A\}, \quad \{r_{k0,0} : k \in [n]\}, \quad \{r_{l1,1} : l \in B\}, \quad \{r_{l1,0} : l \in B\}, \quad \{r_{l1,0} : l \notin B\}.$$

Each family $\mathcal{L}_1, \ldots, \mathcal{L}_4$ introduces a new pivot variable not used earlier:

- $\mathcal{L}_1$ pivots on $r_{k0,1}$ ($k \in A$),
- $\mathcal{L}_2$ pivots on $r_{k0,0}$ ($k \in [n] \setminus \{l'\}$) and $r_{l1,1}$ ($l \in B$),
- $\mathcal{L}_3$ pivots on $r_{l1,0}$ ($l \in B$),
- $\mathcal{L}_4$ pivots on $r_{l1,0}$ ($l \notin B$).

The coefficient matrix therefore contains a block upper-triangular submatrix with nonzero diagonal, implying that $\mathcal{L}$ is linearly independent.

Conclusion. The active constraint matrix has rank

$$2n - 1 + a + b.$$

Imposing the extreme ray rank condition $2n - 1 + a + b = 4n - 2$ yields $\{a, b\} = \{n - 1, n\}$, and the configuration coincides with Case C.1.1.

Subcase 2.2: $\exists k_0 : \ r_{k_0 1,0} = 1$. In this subcase, the equality constraints imply

$$r_{l0,1} = -1 \quad \forall\, l \in [n], \qquad r_{k_0 1,1} = -1.$$

Moreover, by Remark 10,

$$r_{(n-1)1,0} = 0.$$

Let

$$A := \{k \in [n] : r_{k1,0} = 1\}, \qquad B := \{k \in [n] : r_{k1,1} = 0\},$$

and denote $a := |A|$ and $b := |B|$. Since $r_{k1,1} \le -r_{k1,0}$, the above constraints imply $A \cap B = \varnothing$ and therefore

$$a + b \le n.$$

Rank count. We now count the number of independent tight equalities.

- Equalities of type $r_{k0,0} + r_{k0,1} = 0$ do not appear in this subcase.
- Equalities of type $r_{k0,0} + r_{l1,1} = 0$ contribute $n + b - 1$ independent constraints.
- Equalities of type $r_{l1,0} + r_{k0,1} = 0$ contribute $n + a - 1$ independent constraints.
- Equalities of type $r_{k1,0} + r_{k1,1} = 0$ contribute $a + b$ independent constraints.

Hence the total number of independent equalities equals

$$(n + b - 1) + (n + a - 1) + (a + b) = 2n - 2 + 2a + 2b.$$

Imposing the extreme ray rank condition $2n - 2 + 2a + 2b = 4n - 2$ yields

$$a + b = n.$$

Remark 12. *Note that A cannot be empty. Indeed, if $A = \varnothing$, then no equalities of type $r_{l1,0} + r_{k0,1} = 0$ appear. In this case the number of independent equalities is at most*

$$(n + b - 1) + b \le 3n - 1,$$

which is strictly smaller than the extreme ray rank $4n - 2$. Hence $A \neq \varnothing$.

Enumeration of rays. Thus A and B form a partition of $[n]$ with $(n-1) \notin A$ because $r_{(n-1)1,0} = 0$. Hence $A \subseteq [n-1]$ is nonempty, and every nonempty subset $A \subseteq [n-1]$ yields an extreme ray.

Equivalently, for any binary vector $\mathbf{s} \in \{0,1\}^{n-1}$ that is not identically zero, define

$$\begin{cases} r_{k0,1} = -1, & \forall k \in [n], \\ r_{k0,0} = 0, & \forall k \in [n], \\ r_{(n-1)1,0} = 0, & \\ r_{k1,0} = s_k, & \forall k \in [n-1], \\ r_{k1,1} = -r_{k1,0}, & \forall k \in [n]. \end{cases}$$

This produces $2^{n-1} - 1$ M-distinct extreme rays belonging to the family **(IV)** in Definition 12.

C.1.3 Case 3: Pattern A

Suppose Pattern A holds, namely

$$r_{ki,0} \in \{0, -1\}, \qquad r_{ki,1} \in \{0, +1\}, \quad \forall k \in [n],\ i \in \{0,1\}.$$

We consider four subcases.

Subcase 3.1: $\forall k, i:\ r_{ki,1} = 0$**.** In this subcase, the only potentially nonzero variables are $\{r_{k0,0}, r_{k1,0}\}_{k \in [n]}$. Define

$$A := \{k \in [n] : r_{k0,0} = 0\}, \qquad B := \{k \in [n] : r_{k1,0} = 0\},$$

and denote $a := |A|$ and $b := |B|$.

Rank count. We count the number of independent tight equalities.

- Equalities of type $r_{k0,0} + r_{k0,1} = 0$ contribute a independent constraints.
- Equalities of type $r_{l1,0} + r_{k0,1} = 0$ contribute $n + a - 1$ independent constraints.
- Equalities of type $r_{k1,0} + r_{k1,1} = 0$ contribute b independent constraints.
- Equalities of type $r_{k0,0} + r_{l1,1} = 0$ contribute $n - a$ independent constraints (Similar to Subcase 2.1).

Hence the total number of independent equalities equals

$$a + (n + a - 1) + (n - a) + b = 2n - 1 + a + b.$$

Imposing the extreme ray rank condition $2n - 1 + a + b = 4n - 2$ yields $\{a, b\} = \{n-1, n\}$. Therefore this configuration coincides with Case C.1.1.

Subcase 3.2: $\exists k_0, l_0:\ r_{k_00,1} = 1,\ r_{l_01,1} = 1$**.** In this case, the equality constraints force

$$r_{l0,0} = r_{l1,0} = -1 \qquad \forall l \in [n].$$

Define

$$A := \{k \in [n] : r_{k0,1} = 1\}, \qquad B := \{k \in [n] : r_{k1,1} = 1\},$$

and denote $a := |A|$ and $b := |B|$. By Remark 10, $(n-1) \notin B$, hence $b \leq n - 1$.

Rank count. As in the previous subcases, we count the number of independent equalities:

- Equalities of type $r_{k0,0} + r_{k0,1} = 0$ contribute a constraints.
- Equalities of type $r_{l1,0} + r_{k0,1} = 0$ contribute $n + a - 1$ constraints.
- Equalities of type $r_{k0,0} + r_{l1,1} = 0$ contribute $n - a$ constraints.

- Equalities of type $r_{k1,0} + r_{k1,1} = 0$ contribute b constraints.

Hence the total number of independent equalities equals

$$2n - 1 + a + b.$$

Imposing the extreme ray rank condition $2n - 1 + a + b = 4n - 2$ yields

$$a = n, \qquad b = n - 1.$$

Resulting extreme ray. These values uniquely determine the configuration, giving the extreme ray

$$\begin{cases} r_{k1,0} = -1, & \forall k \in [n], \\ r_{k0,0} = -1, & \forall k \in [n], \\ r_{(n-1)1,1} = 0, & \\ r_{k1,1} = 1, & \forall k \in [n-1], \\ r_{k0,1} = 1, & \forall k \in [n]. \end{cases}$$

which belongs to the family **(III)** in Definition 12.

Subcase 3.3: $\exists k_0 : \ r_{k_0 0,1} = 1, \ \forall l : \ r_{l1,1} = 0$**.** In this subcase, the equality constraints imply

$$r_{l1,0} = -1 \quad \forall l \in [n], \qquad r_{k_0 0,0} = -1.$$

Define

$$A := \{k \in [n] : r_{k0,1} = 1\}, \qquad B := \{k \in [n] : r_{k0,0} = 0\},$$

and denote $a := |A|$ and $b := |B|$.

Rank count. We now count the number of independent tight equalities.

- Equalities of type $r_{k1,0} + r_{k1,1} = 0$ do not appear.
- Equalities of type $r_{k0,0} + r_{l1,1} = 0$ contribute $n + b - 1$ independent constraints.
- Equalities of type $r_{l1,0} + r_{k0,1} = 0$ contribute $n + a - 1$ independent constraints.
- Equalities of type $r_{k0,0} + r_{k0,1} = 0$ (for indices where both variables appear in tight constraints) contribute $a + b$ independent constraints.

Remark 13. *If $B = \varnothing$, then no equalities of type $r_{k0,0} + r_{l1,1} = 0$ appear. In this case the number of independent equalities is at most*

$$(n + a - 1) + a = n + 2a - 1,$$

which is strictly smaller than the extreme ray rank $4n - 2$. Hence $B \neq \varnothing$.

Hence the total number of independent equalities equals

$$(n + b - 1) + (n + a - 1) + (a + b) = 2n - 2 + 2a + 2b.$$

Imposing the extreme ray rank condition

$$2n - 2 + 2a + 2b = 4n - 2$$

yields

$$a + b = n.$$

Enumeration of rays. Thus, the sets A and B form a nontrivial partition of $[n]$. Equivalently, for any binary vector $\mathbf{s} \in \{0,1\}^n$ that is neither identically zero nor identically one, define

$$\begin{cases} r_{k1,1} = 0, & \forall k \in [n], \\ r_{k1,0} = -1, & \forall k \in [n], \\ r_{k0,1} = s_k, & \forall k \in [n], \\ r_{k0,0} = -r_{k0,1}, & \forall k \in [n]. \end{cases}$$

This produces $2^n - 2$ M-distinct extreme rays belonging to the family **(V)** in Definition 12.

Subcase 3.4: $\exists k_0 : \ r_{k_0 1,1} = 1, \ \forall l : \ r_{l0,1} = 0$**.** In this subcase, the equality constraints imply

$$r_{l0,0} = -1 \quad \forall l \in [n], \qquad r_{k_0 1,0} = -1.$$

Define

$$A := \{k \in [n] : r_{k1,1} = 1\}, \qquad B := \{k \in [n] : r_{k1,0} = 0\},$$

and denote $a := |A|$ and $b := |B|$.

Rank count. We count the number of independent tight equalities.

- Equalities of type $r_{k0,0} + r_{k0,1} = 0$ do not appear.
- Equalities of type $r_{k1,0} + r_{k1,1} = 0$ contribute $a + b$ independent constraints.
- Equalities of type $r_{l1,0} + r_{k0,1} = 0$ contribute $n + b - 1$ independent constraints.
- Equalities of type $r_{k0,0} + r_{l1,1} = 0$ contribute $n + a - 1$ independent constraints.

Remark 14. *If $B = \varnothing$, then no equalities of type $r_{l1,0} + r_{k0,1} = 0$ appear. In this case the number of independent equalities is at most*

$$(n + a - 1) + a = n + 2a - 1,$$

which is strictly smaller than the extreme ray rank $4n - 2$. Hence $B \neq \varnothing$.

Hence the total number of independent equalities equals

$$(a + b) + (n + a - 1) + (n + b - 1) = 2n - 2 + 2a + 2b.$$

Imposing the extreme ray rank condition

$$2n - 2 + 2a + 2b = 4n - 2$$

yields

$$a + b = n.$$

Enumeration of rays. By Remark 10, $(n-1) \notin A$. Thus, $A \subseteq [n-1]$ is a nonempty subset, and every such subset produces an extreme ray.

Equivalently, for any nonzero binary vector $\mathbf{s} \in \{0,1\}^{n-1}$, define

$$\begin{cases} r_{k0,1} = 0, & \forall k \in [n], \\ r_{k0,0} = -1, & \forall k \in [n], \\ r_{(n-1)1,1} = 0, & \\ r_{k1,1} = s_k, & \forall k \in [n-1], \\ r_{k1,0} = -r_{k1,1}, & \forall k \in [n]. \end{cases}$$

This produces $2^{n-1} - 1$ M-distinct extreme rays belonging to the family **(VI)** in Definition 12.

The above case analysis exhausts all feasible configurations and characterizes all extreme rays of the cone. This completes the proof of Proposition 7.

C.2 NUMBER OF EXTREME RAYS

Summing the contributions from all cases,

$$2 + (4n - 1) + (2^{n-1} - 1) + 1 + (2^n - 2) + (2^{n-1} - 1) = 2^{n+1} + 4n - 2,$$

which proves the following Corollary.

Corollary 6. *According to Proposition 7, the cone $\mathcal{K}$ is generated by $2^{n+1} + 4n - 2$ M-distinct vectors r_i.*

C.3 EXTRACTING THE IV INEQUALITIES

Theorem 5. *Let $\mathcal{P}$ be the observed probability distribution, and $p_{yd,z}$ defined as in Section 2. The sharp testable implications of the IV model under Assumptions 1, 2, and 4 are the following inequalities:*

$$\begin{cases} \sum_{k\in[n]} -p_{k0,1} + \sum_{k\in T}(p_{k1,0} - p_{k1,1}) \leq 0, \\ \qquad \forall T \subseteq [n-1],\ T \neq \varnothing, \\ \sum_{k\in[n]} -p_{k1,0} + \sum_{k\in T}(p_{k0,1} - p_{k0,0}) \leq 0, \\ \qquad \forall T \subset [n],\ T \neq \varnothing, \\ \sum_{k\in[n]} -p_{k0,0} + \sum_{k\in T}(p_{k1,1} - p_{k1,0}) \leq 0, \\ \qquad \forall T \subseteq [n-1],\ T \neq \varnothing. \end{cases} \tag{17}$$

Proof. Based on the fact that $\mathcal{K}$ can be written as a cone combination of r_is, we can write the sufficient inequalities as $\mathbf{p}^\top r_i \leq 0$. As $\mathbf{p}$ is the observed probability law, the following inequalities are obviously satisfied, hence, there is no need to check them.

$$\begin{cases} \sum_{k\in[n]}(p_{k0,1} + p_{k1,1}) - \sum_{k\in[n]}(p_{k0,0} + p_{k1,0}) \leq 0, \\ \sum_{k\in[n]}(p_{k0,0} + p_{k1,0}) - \sum_{k\in[n]}(p_{k0,1} + p_{k1,1}) \leq 0, \\ -p_{k0,0} \leq 0, \quad \forall k \in [n], \\ -p_{k0,1} \leq 0, \quad \forall k \in [n], \\ -p_{k1,0} \leq 0, \quad \forall k \in [n], \\ -p_{k1,1} \leq 0, \quad \forall k \in [n-1], \\ \sum_{k\in[n-1]}(p_{k1,1} + p_{k0,1}) + p_{(n-1)0,1} \leq \sum_{k\in[n]}(p_{k1,0} + p_{k0,0}). \end{cases} \tag{38}$$

Therefore, the given inequalities in Theorem 5 are sufficient to test the validity of the observed probability vector $\mathbf{p}$ satisfying our problem assumptions and setup.

For necessity part, we need to show that each extreme ray presented in Proposition 7 cannot be derived as cone combination of the others and moreover, for each inequality in Equation (17), there exists a vector $\mathbf{p}$ such that it satisfies all inequalities except that specific one.

Fix an extreme ray representative $\mathbf{r} \in \mathcal{R}$ (where $\mathcal{R}$ is the family in Definition 12), and suppose, by contradiction, that

$$\mathbf{r} = \sum_{t=1}^{m} \lambda_t \mathbf{r}_t, \qquad \lambda_t \geq 0,$$

where each $\mathbf{r}_t \in \mathcal{R}$ are such that $\mathcal{E}_{\mathbf{r}_t} \neq \mathcal{E}_{\mathbf{r}}$ whenever $\mathbf{r}_t \notin \ker(M)$.

Let $\mathcal{I} := \{t : \lambda_t > 0\}$ and pick any $t_0 \in \mathcal{I}$. Define

$$\mathbf{r}_1 := \lambda_{t_0}\mathbf{r}_{t_0}, \qquad \mathbf{r}_2 := \sum_{t\in\mathcal{I}\setminus\{t_0\}} \lambda_t\mathbf{r}_t.$$

Then $\mathbf{r}_1, \mathbf{r}_2 \in \mathcal{K}$ and $\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2$.

Since $\mathbf{r}$ is an extreme ray, by Definition 3 we have

$$\mathcal{E}_{\mathbf{r}_1} = \mathcal{E}_{\mathbf{r}_2} = \mathcal{E}_{\mathbf{r}}.$$

Hence there exist $\alpha > 0$ and $\mathbf{s} \in \ker(M)$ such that

$$\mathbf{r} = \alpha\,\mathbf{r}_{t_0} + \mathbf{s}.$$

We distinguish two cases.

Case 1: $\mathbf{r} \notin \ker(M)$. This implies that $\mathbf{r}_{t_0} \notin \ker(M)$. By the normalization in Remark 10, $r_{(n-1)1,1} = 0$. Since $\mathbf{s} \in \ker(M)$, we have $\mathbf{s} = \mathbf{0}$ if and only if $s_{(n-1)1,1} = 0$, which implies $\mathbf{s} = \mathbf{0}$ and therefore

$$\mathbf{r} = \alpha\,\mathbf{r}_{t_0}.$$

All vectors in $\mathcal{R}$ are unique rays; hence $\mathbf{r} = \mathbf{r}_{t_0}$.

Case 2: $\mathbf{r} \in \ker(M)$. Multiplying the above equality by M yields

$$\mathbf{0} = M\mathbf{r} = \alpha\,M\mathbf{r}_{t_0},$$

so $\mathbf{r}_{t_0} \in \ker(M)$. Since both $\mathbf{r}$ and $\mathbf{r}_{t_0}$ belong to $\ker(M) = \text{cone}(\mathbf{s}_1, \cdots, \mathbf{s}_t, -\mathbf{s}_1, \cdots, -\mathbf{s}_t)$, combining the fact that $\mathbf{s}_1, \cdots, \mathbf{s}_t$ are independent bases of $\ker(M)$ with positiveness of α implies that $\mathbf{r}$ and $\mathbf{r}_{t_0}$ must coincide. Hence

$$\mathbf{r} = \mathbf{r}_{t_0}.$$

Therefore, $\mathbf{r}$ cannot be expressed as a conic combination of representatives of the other elements in $\mathcal{R}$. Since the conic hull of these representatives is a closed convex set, the strong separating hyperplane theorem guarantees the existence of a vector $\mathbf{p}$ such that

$$\mathbf{p}^\top\mathbf{r} > 0 \quad \text{and} \quad \mathbf{p}^\top\mathbf{x} \le 0$$

for all $\mathbf{x}$ in the conic hull generated by representatives of the remaining elements of $\mathcal{R}$. Moreover, since $\mathbf{r}$ does not correspond to an extreme ray inducing the inequalities in Equation (38), the vector $\mathbf{p}$ satisfies these inequalities. In particular, $\mathbf{p} \neq \mathbf{0}$, and by normalization, the vector

$$\mathbf{p}^* = \frac{2\mathbf{p}}{\|\mathbf{p}\|_1}$$

defines a valid observed probability vector such that

$$(\mathbf{p}^*)^\top\mathbf{r} > 0 \quad \text{and} \quad (\mathbf{p}^*)^\top\mathbf{x} \le 0$$

for all $\mathbf{x}$ in the conic hull generated by representatives of the other elements of $\mathcal{R}$.

As a result, for each inequality in Equation (17), there exists an observed probability vector $\mathbf{p}$ that violates that inequality while satisfying all the others. Consequently, each inequality $\mathbf{p}^\top\mathbf{r} \le 0$ associated with $\mathbf{r} \in \mathcal{R}$ is *necessary*: removing it would enlarge the feasible set of observed probability vectors $\mathbf{p}$ beyond the set induced by $\mathcal{K}$.

Combining necessity with the sufficiency argument (since $\mathcal{K} = \text{cone}(\mathcal{R})$), we conclude that the family of inequalities in Theorem 5 is necessary and sufficient. $\square$

D PROOFS OF GENERALIZED VERSIONS

Below, we provide generalized versions of our previous Lemmas. Specifically, we no longer restrict our setting to $\ell = 2$; the instrument Z can take an arbitrary number ℓ of values.

Lemma 22. *Let $\mathbf{v}$ be a feasible point satisfying $M\mathbf{v} \le \mathbf{c}$. Then $\mathbf{v}$ is a vertex if and only if*

$$\textit{rank}(M_{\mathbf{v}}) = 2n\ell - \dim(\ker(M)) = \textit{rank}(M).$$

Proof. We prove both directions.

($\Leftarrow$) If rank$(M_{\mathbf{v}}) < 2n\ell - \dim(\ker(M))$, then $\mathbf{v}$ is not a vertex.

Suppose $\mathrm{rank}(M_{\mathbf{v}}) < 2n\ell - \dim(\ker(M))$. Note that $\mathrm{rank}(M) = 2n\ell - \dim(\ker(M))$. Because $\mathrm{rank}(M_{\mathbf{v}}) < \mathrm{rank}(M)$, we have

$$\dim\ker(M_{\mathbf{v}}) > \dim\ker(M).$$

Hence, there exists a nonzero vector $\mathbf{t} \in \ker(M_{\mathbf{v}})$ such that $\mathbf{t} \notin \ker(M)$.

Since $\mathbf{t} \in \ker(M_{\mathbf{v}})$, we have

$$M_{\mathbf{v}}(\mathbf{v} + \epsilon\mathbf{t}) = M_{\mathbf{v}}\mathbf{v} = \mathbf{c}_{\mathbf{v}}$$

for all $\epsilon \in \mathbb{R}$. For sufficiently small $\epsilon > 0$, the inequalities corresponding to non-active constraints remain strict, because they are strict at $\mathbf{v}$. Therefore,

$$M(\mathbf{v} + \epsilon\mathbf{t}) \le \mathbf{c} \quad \text{and} \quad M(\mathbf{v} - \epsilon\mathbf{t}) \le \mathbf{c}$$

for sufficiently small ϵ.

Thus both $\mathbf{v} + \epsilon\mathbf{t}$ and $\mathbf{v} - \epsilon\mathbf{t}$ are feasible. Moreover,

$$\mathbf{v} = \tfrac{1}{2}(\mathbf{v} + \epsilon\mathbf{t}) + \tfrac{1}{2}(\mathbf{v} - \epsilon\mathbf{t}),$$

and since $\mathbf{t} \notin \ker(M)$, we have

$$M(\mathbf{v} + \epsilon\mathbf{t}) \neq M\mathbf{v},$$

so the two feasible points are M-distinct.

Hence $\mathbf{v}$ can be written as a non-trivial convex combination of two M-distinct feasible points, and therefore $\mathbf{v}$ is not a vertex by Definition 2.

($\Rightarrow$) If rank$(M_{\mathbf{v}}) = 2n\ell - \dim(\ker(M))$, then $\mathbf{v}$ is a vertex.

Now suppose $\mathrm{rank}(M_{\mathbf{v}}) = 2n\ell - \dim(\ker(M))$. Assume

$$\mathbf{v} = \lambda\mathbf{v}_1 + (1-\lambda)\mathbf{v}_2, \qquad \lambda \in (0,1),$$

for feasible points $\mathbf{v}_1, \mathbf{v}_2$.

Since $\mathbf{v}$ is active on $M_{\mathbf{v}}$, we have

$$M_{\mathbf{v}}\mathbf{v} = \mathbf{c}_{\mathbf{v}}.$$

Applying $M_{\mathbf{v}}$ to the convex combination gives

$$M_{\mathbf{v}}\mathbf{v} = \lambda M_{\mathbf{v}}\mathbf{v}_1 + (1-\lambda)M_{\mathbf{v}}\mathbf{v}_2.$$

Because $\mathbf{v}_1$ and $\mathbf{v}_2$ are feasible,

$$M_{\mathbf{v}}\mathbf{v}_1 \le \mathbf{c}_{\mathbf{v}}, \qquad M_{\mathbf{v}}\mathbf{v}_2 \le \mathbf{c}_{\mathbf{v}}.$$

Since the convex combination equals $\mathbf{c}_{\mathbf{v}}$, it follows that

$$M_{\mathbf{v}}\mathbf{v}_1 = \mathbf{c}_{\mathbf{v}} \quad \text{and} \quad M_{\mathbf{v}}\mathbf{v}_2 = \mathbf{c}_{\mathbf{v}}.$$

Therefore,

$$M_{\mathbf{v}}(\mathbf{v} - \mathbf{v}_1) = \mathbf{0}, \qquad M_{\mathbf{v}}(\mathbf{v} - \mathbf{v}_2) = \mathbf{0}.$$

Because $\text{rank}(M_{\mathbf{v}}) = \text{rank}(M)$, and the fact the $M_{\mathbf{v}}$ is a submatrix of M we have

$$\ker(M_{\mathbf{v}}) = \ker(M).$$

Hence,

$$\mathbf{v} - \mathbf{v}_1 \in \ker(M), \qquad \mathbf{v} - \mathbf{v}_2 \in \ker(M).$$

By Definition 2, any two feasible points differing by a vector in $\ker(M)$ represent the same extreme point. Thus $\mathbf{v}$ cannot be written as a non-trivial convex combination of two M-distinct feasible points, and therefore $\mathbf{v}$ is a vertex. $\square$

Remark 15. *Since $\text{rank}(M) = \text{rank}(M_{\mathbf{v}}) = 2n\ell - \dim(\ker(M))$, the solution set of the system $M_{\mathbf{v}}\mathbf{x} = \mathbf{c}$ is an affine space of the form $\mathbf{v} + \ker(M)$. Therefore, $M_{\mathbf{v}}$ uniquely determines the M-equivalence class $[\mathbf{v}]$.*

In particular, if two active constraint matrices share the same row basis (that is, they consist of the same linearly independent rows), then they determine the same solution $\mathbf{v}$ and thus the same M-equivalence class $[\mathbf{v}]$. Consequently, the two active constraint matrices must coincide.

Lemma 23. *Let $\mathbf{r} \neq \mathbf{0}$ be a feasible vector of the cone $\mathcal{K} = \{\mathbf{x} : M\mathbf{x} \leq 0\}$. Let $M_{\mathbf{r}}^{(0)}$ denote the matrix of all inequalities that are active at $\mathbf{r}$ as described in Definition 13. Then $\mathbf{r}$ is an extreme ray of $\mathcal{K}$ if and only if the active set contains at least $2n\ell - \dim(\ker(M)) - 1$ linearly independent constraints, i.e.,*

$$\text{rank}(M_{\mathbf{r}}^{(0)}) \geq 2n\ell - \dim(\ker(M)) - 1$$

Proof. ($\Rightarrow$) Assume $\mathbf{r}$ is an extreme ray and suppose, by contradiction, that $\text{rank}(M_{\mathbf{r}}^{(0)}) < 2n\ell - \dim(\ker(M)) - 1$. Then the null space of $M_{\mathbf{r}}^{(0)}$ has dimension strictly larger than $1 + \dim\ker(M)$. Hence there exists a nonzero vector

$$\mathbf{t} \in \ker(M_{\mathbf{r}}^{(0)}) \setminus \text{span}(\mathbf{r}, \ker(M)).$$

Since $\mathbf{t} \in \ker(M_{\mathbf{r}}^{(0)})$, all inequalities that are tight at $\mathbf{r}$ remain tight along the direction $\mathbf{t}$. Therefore, for sufficiently small $\epsilon > 0$,

$$M(\mathbf{r} + \epsilon\mathbf{t}) \leq 0, \qquad M(\mathbf{r} - \epsilon\mathbf{t}) \leq 0,$$

so both $\mathbf{r} + \epsilon\mathbf{t}$ and $\mathbf{r} - \epsilon\mathbf{t}$ belong to $\mathcal{K}$. Moreover,

$$\mathbf{r} = \tfrac{1}{2}(\mathbf{r} + \epsilon\mathbf{t}) + \tfrac{1}{2}(\mathbf{r} - \epsilon\mathbf{t}).$$

Because $\mathbf{t} \notin \text{span}(\mathbf{r}, \ker(M))$, at least one of the vectors $\mathbf{r} \pm \epsilon\mathbf{t}$ is not contained in $\text{span}(\mathbf{r}, \ker(M))$, contradicting Definition 3. Hence $\text{rank}(M_{\mathbf{r}}^{(0)}) \geq 2n\ell - \dim(\ker(M)) - 1$.

($\Leftarrow$) Assume $\text{rank}(M_{\mathbf{r}}^{(0)}) \geq 2n\ell - \dim(\ker(M)) - 1$. Then

$$\dim\ker(M_{\mathbf{r}}^{(0)}) \leq 1 + \dim\ker(M).$$

By construction, $\mathbf{r} \in \ker(M_{\mathbf{r}}^{(0)})$ and $\ker(M) \subseteq \ker(M_{\mathbf{r}}^{(0)})$.

Let $\mathbf{r}^{(1)}, \mathbf{r}^{(2)} \in \mathcal{K}$ such that

$$\mathbf{r} = \mathbf{r}^{(1)} + \mathbf{r}^{(2)}.$$

Applying $M_{\mathbf{r}}^{(0)}$ gives

$$0 = M_{\mathbf{r}}^{(0)}\mathbf{r} = M_{\mathbf{r}}^{(0)}\mathbf{r}^{(1)} + M_{\mathbf{r}}^{(0)}\mathbf{r}^{(2)}.$$

Since $\mathbf{r}^{(1)}, \mathbf{r}^{(2)} \in \mathcal{K}$, we have $M_{\mathbf{r}}^{(0)}\mathbf{r}^{(1)} \leq 0$ and $M_{\mathbf{r}}^{(0)}\mathbf{r}^{(2)} \leq 0$, hence

$$M_{\mathbf{r}}^{(0)}\mathbf{r}^{(1)} = M_{\mathbf{r}}^{(0)}\mathbf{r}^{(2)} = \mathbf{0},$$

which implies $\mathbf{r}^{(1)}, \mathbf{r}^{(2)} \in \ker(M_{\mathbf{r}}^{(0)})$.

If $\mathbf{r} \in \ker(M)$, then $M_{\mathbf{r}}^{(0)} = M$ and therefore $\mathbf{r}^{(1)}, \mathbf{r}^{(2)} \in \ker(M)$, satisfying Definition 3. Otherwise, $\mathbf{r} \notin \ker(M)$. Since $\ker(M) \subseteq \ker(M_{\mathbf{r}}^{(0)})$ and

$$\dim\ker(M_{\mathbf{r}}^{(0)}) \leq 1 + \dim\ker(M),$$

we must have

$$\ker(M_{\mathbf{r}}^{(0)}) = \text{span}(\mathbf{r}, \ker(M)).$$

Thus any feasible decomposition of $\mathbf{r}$ lies in $\text{span}(\mathbf{r}, \ker(M))$, which proves that $\mathbf{r}$ is an extreme ray. $\square$

D.1 VERTICES AND PROOFS OF SECTION 4.2

Theorem 4. *For any $a \in [\ell]$, and any $\mathbf{s} = (s_0, \dots, s_{n-1}) \in ([\ell] \setminus \{a\})^n$, define $\mathbf{w}(a, \mathbf{s}) \in \mathbb{R}^{n\times 2\times \ell}$ as*

$$\begin{aligned}
w_{y1,j} &= 0, && \forall y,\ j \neq a,\\
w_{y1,a} &= \gamma_y - \gamma_{n-1}, && \forall y,\\
w_{y0,j} &= 0, && \forall y,\ j \notin \{a, s_y\},\\
w_{y0,s_y} &= \gamma_{n-1} - \gamma_y, && \forall y,\\
w_{y0,a} &= \gamma_0 - \gamma_{n-1}, && \forall y.
\end{aligned}$$

Then $\{\mathbf{w}(a, \mathbf{s}) : a \in [\ell],\ \mathbf{s} \in ([\ell] \setminus \{a\})^n$ non-constant$\}$ is a set of M-distinct vertices of $\mathcal{H}$ (the feasible region of Equation 10). The number of these M-distinct vertices equals

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

Proof. For convenience we rewrite Equation (10) in the equivalent form

$$\begin{aligned}
\max \quad & \sum_{y\in[n]} \sum_{d\in\{0,1\}} \sum_{j\in[\ell]} p_{yd,j}\, x_{yd,j} \\
\text{s.t.} \quad & \sum_{j=0}^{\ell-1} x_{y_{i_j} i_j, j} \leq \gamma_{y_1} - \gamma_{y_0}, \qquad \forall y_0, y_1 \in [n],\ \forall (i_0, \dots, i_{\ell-1}) \in \{0,1\}^\ell.
\end{aligned} \tag{39}$$

Lemma 24 (Shift invariance for general version)**.** *Let $\mathbf{v}$ be feasible for Equation* (39)*. Fix any $t \in \mathbb{R}$ and $j \in [\ell] \setminus \{0\}$, and define $\mathbf{u}$ by*

$$u_{ki,0} = v_{ki,0} + t, \qquad u_{ki,j} = v_{ki,j} - t, \qquad \forall k \in [n],\ i \in \{0,1\}.$$

Then $\mathbf{y}$ is also feasible. Moreover, $M_{\mathbf{v}} = M_{\mathbf{u}}$ and $\mathbf{p}^\top \mathbf{u} = \mathbf{p}^\top \mathbf{v}$; in particular, $\mathbf{v}$ and $\mathbf{u}$ induce the same constraints on $\mathbf{p}$.

proof of Lemma 24. Each inequality in Equation (39) involves a sum of one $(\cdot, 0)$-coordinate and one $(\cdot, j)$-coordinate. Under the transformation above, these sums remain unchanged, hence feasibility is preserved and $M_{\mathbf{v}} = M_{\mathbf{u}}$. The equality $\mathbf{p}^\top \mathbf{v} = \mathbf{p}^\top \mathbf{u}$ follows from the same cancellation due to the fact that $\sum_{k\in[n],\ i\in\{0,1\}} p_{ki,0} = \sum_{k\in[n],\ i\in\{0,1\}} p_{ki,j} = 1$. $\square$

Remark 16 (Normalization)**.** *By Lemma 24, for any feasible point in Equation* (39)*, we may impose the normalization $v_{(n-1)1,j} = 0, \forall j \neq 0$ without loss of generality.*

In order to prove Theorem 4, we show (i) feasibility of the constructed $\mathbf{w}$, and (ii) that $\mathbf{w}$ is a vertex by exhibiting a family of constraints that are tight at $\mathbf{w}$ and whose equalities determine $\mathbf{w}$ uniquely (after fixing a normalization for the shift invariance).

(i) Feasibility. Fix $(i_0, \dots, i_{\ell-1}) \in \{0,1\}^\ell$ and let $A := \{j : i_j = 0\}$ and $A^c := \{j : i_j = 1\}$. For any $y_0, y_1 \in [n]$, the left-hand side of the corresponding constraint equals

$$\sum_{j\in A} w_{y_0 0,j} \ + \sum_{j\in A^c} w_{y_1 1,j}.$$

By construction, among the terms $w_{y_1 1,j}$ only the coordinate with $j = a$ may be nonzero, and among the terms $w_{y_0 0,j}$ only the coordinates with $j \in \{a, s_{y_0}\}$ may be nonzero. Hence the above sum contains at most two nonzero contributions, at indices a and s_{y_0}, and we obtain

$$\sum_{j\in A} w_{y_0 0,j} + \sum_{j\in A^c} w_{y_1 1,j} \leq \max\{w_{y_0 0,a}, w_{y_1 1,a}\} + \max\{w_{y_0 0,s_{y_0}}, w_{y_1 1,s_{y_0}}\}.$$

Since $w_{y_1 1,s_{y_0}} = 0$ (because $s_{y_0} \neq a$), the second maximum equals $w_{y_0 0,s_{y_0}} = \gamma_{n-1} - \gamma_{y_0}$. Moreover $w_{y_0 0,a} = \gamma_0 - \gamma_{n-1}$ and $w_{y_1 1,a} = \gamma_{y_1} - \gamma_{n-1}$, so the first maximum equals $\gamma_{y_1} - \gamma_{n-1}$. Therefore,

$$\sum_{j\in A} w_{y_0 0,j} + \sum_{j\in A^c} w_{y_1 1,j} \leq \big(\gamma_{y_1} - \gamma_{n-1}\big) + \big(\gamma_{n-1} - \gamma_{y_0}\big) = \gamma_{y_1} - \gamma_{y_0},$$

and all constraints in Equation (39) are satisfied. Thus $\mathbf{w}$ is feasible.

(ii) Vertex Property. We next show that the above construction is uniquely determined (up to the standard shift invariance) by a collection of tight constraints from Equation (39). Concretely, the construction implies that the following constraints are tight:

$$\sum_{j=0}^{\ell-1} x_{y0,j} = \gamma_0 - \gamma_y, \qquad \forall y \in [n], \tag{40}$$

$$\sum_{j=0}^{\ell-1} x_{y1,j} = \gamma_y - \gamma_{n-1}, \qquad \forall y \in [n], \tag{41}$$

$$x_{y0,s_y} + \sum_{j \neq s_y} x_{k1,j} = \gamma_k - \gamma_y, \qquad \forall y, k \in [n], \tag{42}$$

$$x_{y0,s_y} + x_{k1,a} + \sum_{j \in A} x_{y0,j} + \sum_{j \in A^c \setminus \{a, s_y\}} x_{k1,j} = \gamma_k - \gamma_y, \qquad \forall y, k,\ \forall A \subseteq [\ell] \setminus \{a, s_y\}. \tag{43}$$

After fixing the normalization

$$x_{(n-1)1,j} = 0, \qquad \forall j \neq 0,$$

we prove that the resulting linear system has a unique solution, hence $\mathbf{w}$ is a vertex of the feasible polyhedron.

From the normalization $x_{(n-1)1,j} = 0$ for all $j \neq 0$ (Remark 16) and equality (41) with $y = n-1$, we obtain

$$\sum_{j=0}^{\ell-1} x_{(n-1)1,j} = \gamma_{n-1} - \gamma_{n-1} = 0.$$

Since all summands except possibly $j = 0$ are zero, it follows that $x_{(n-1)1,0} = 0$ as well. Thus,

$$x_{(n-1)1,j} = 0, \qquad \forall j \in [\ell]. \tag{44}$$

Now apply (42) with $k = n-1$:

$$x_{y0,s_y} + \sum_{j \neq s_y} x_{(n-1)1,j} = \gamma_{n-1} - \gamma_y.$$

Using (44), the sum vanishes, and we conclude that for every y,

$$x_{y0,s_y} = \gamma_{n-1} - \gamma_y. \tag{45}$$

Substituting (45) back into (42) for general k gives

$$\gamma_{n-1} - \gamma_y + \sum_{j \neq s_y} x_{k1,j} = \gamma_k - \gamma_y,$$

hence

$$\sum_{j \neq s_y} x_{k1,j} = \gamma_k - \gamma_{n-1}, \qquad \forall y, k. \tag{46}$$

Combining (41) with (46) gives

$$x_{k1,s_y} = 0 \qquad \forall k, y. \tag{47}$$

Define the index sets

$$B := \{j : \exists k \text{ with } s_k = j\}, \qquad C := B^c.$$

By construction $s_y \neq a$ for all y, so $a \in C$.

Since there exists y such that $s_y = j$ for each $j \in B$, using (47) implies that

$$x_{k1,j} = 0, \qquad \forall k,\ \forall j \in B. \tag{48}$$

Consequently, (41) becomes

$$\sum_{j \in C} x_{k1,j} = \gamma_k - \gamma_{n-1}, \qquad \forall k. \tag{49}$$

Next, specialize (43) by choosing $A \subseteq B \setminus \{s_y\}$ (so that A contains no elements of C). Using (45) and (48), the terms involving $x_{k1,j}$ for $j \in B$ vanish, and (43) simplifies to

$$\gamma_{n-1} - \gamma_y + \sum_{j \in A} x_{y0,j} = \gamma_k - \gamma_y - x_{k1,a} - \sum_{j \in C \setminus \{a\}} x_{k1,j}. \tag{50}$$

Combining (49) with (50) gives

$$\sum_{j \in A} x_{y0,j} = 0.$$

Varying A over all subsets of $B \setminus \{s_y\}$ forces $x_{y0,j} = 0$ for all $j \in B \setminus \{s_y\}$, and combined with (45) shows that, within B, the only possibly nonzero $d = 0$ coordinate is at s_y:

$$x_{y0,j} = 0 \;\; (j \in B \setminus \{s_y\}), \qquad x_{y0,s_y} = \gamma_{n-1} - \gamma_y. \tag{51}$$

Now choose $A \subseteq C \setminus \{a\}$ in (43). With (47), the remaining terms lie entirely in C.

$$\gamma_{n-1} - \gamma_y + x_{k1,a} + \sum_{j \in A} x_{y0,j} + \sum_{j \in A^c \cap (C \setminus \{a\})} x_{k1,j} = \gamma_k - \gamma_y$$

The dependence on A then implies that for each $j \in C \setminus \{a\}$ there exists a scalar h_j such that

$$x_{k1,j} = x_{y0,j} = h_j, \qquad \forall k, y, \; \forall j \in C \setminus \{a\}. \tag{52}$$

Using (49), we obtain for all k:

$$x_{k1,a} + \sum_{j \in C \setminus \{a\}} h_j = \gamma_k - \gamma_{n-1}.$$

Also, with (40) and (51), we obtain for all y:

$$x_{y0,a} + \sum_{j \in C \setminus \{a\}} h_j = \gamma_0 - \gamma_{n-1}$$

Finally, by the normalization and (52), for every $j \in C \setminus \{a\}$,

$$h_j = x_{(n-1)1,j} = 0,$$

so $x_{k1,j} = x_{y0,j} = 0$ for all $j \in C \setminus \{a\}$. Plugging this into the two displayed equations yields

$$x_{y0,a} = \gamma_0 - \gamma_{n-1}, \qquad x_{k1,a} = \gamma_k - \gamma_{n-1}.$$

Together with (48) and (51), this determines all coordinates of $\mathbf{x}$ uniquely. Therefore, after fixing the stated normalization, the linear system (40)-(43) admits a unique solution, and hence the feasible point $\mathbf{w}$ is a vertex of the feasible polyhedral.

(iii) Counting. Fix $a \in \{0, \ldots, \ell - 1\}$. The map $\mathbf{s}$ is a length-n vector taking values in $[\ell] \setminus \{a\}$, hence there are $(\ell-1)^n$ possible maps in total. Among these, $(\ell - 1)$ maps are constant. The construction excludes constant $\mathbf{s}$, and moreover the resulting vertex is unchanged under the standard shift invariance, so the number of M-*distinct* (non-constant) choices contributing M-distinct vertices is

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

Multiplying by the l possible choices of a gives the total count

$$l\big((\ell-1)^{n-1} - (\ell-1)\big).$$

This completes the proof. □

Corollary 2 (Exponential lower bound)**.** *Under Assumptions 1, 2 and 4 (or alternatively, 1, 3 and 4a), 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.

Proof. By Theorem 4, the feasible set $\mathcal{H}$ contains at least

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

M-distinct vertices.

Moreover, according to Remark 15 and similar to Proposition 6, for each vertex $\mathbf{v} \in \mathcal{H}$, there exists a pair $(\mathcal{P}, \mathcal{Q})$ satisfying (4) such that $\mathbf{v}$ uniquely attains the optimum of the dual objective.

It follows that each such vertex $\mathbf{v}$ induces an achievable ATE bound of the form $\mathbf{p}^\top \mathbf{v}$. Therefore, any complete set of sharp ATE bounds must include at least one bound corresponding to each of these vertices.

Hence, any complete collection of sharp ATE bounds must contain at least

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

bounds. □

D.2 EXTREME RAYS AND PROOFS OF SECTION 5.2

In this setting, the dual feasibility problem Equation (12) can be written as

$$\begin{aligned} \max \quad & \mathbf{p}^\top \mathbf{x} \\ \text{s.t.} \quad & \sum_{j \in S} x_{y_0 0,j} + \sum_{j \in S^c} x_{y_1 1,j} \leq 0 \quad \forall y_0, y_1 \in \{0,\dots,n-1\},\ \forall S \subseteq \{0,\dots,\ell-1\}. \end{aligned} \tag{53}$$

Theorem 6. *Suppose Assumptions 1, 2 and 4 hold. For any $y' \in [n-1]$, $j' \in [\ell]$, and non-constant $(j_0,\dots,j_{n-1}) \in ([\ell] \setminus \{j'\})^n$, the following inequality holds:*

$$p_{y'1,j'} \ \leq \sum_{j \in \{j_0,\dots,j_{n-1}\}} p_{y'1,j} \ + \sum_{y=0}^{n-1} p_{y0,j_y}. \tag{18}$$

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

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

In order to prove Theorem 6, we introduce a set of extreme rays that produce necessary IV inequalities.

Proposition 8 (Extreme rays generating the inequalities)**.** *Fix $y' \in [n-1]$ and $j' \in [\ell]$. Let $(j_0,\dots,j_{n-1}) \in ([\ell] \setminus \{j'\})^n$ be not identically constant. Define $\mathbf{r}$ by*

$$\begin{aligned} r_{y'1,j'} &= 1, \\ r_{y0,j_y} &= -1, \quad \forall y, \\ r_{y'1,j_y} &= -1, \quad \forall y, \end{aligned}$$

and all other entries equal to zero. Then $\mathbf{r}$ is an extreme ray of the cone $\mathcal{K}$.

Proof. The proof consists of three steps.

Step 1: Feasibility. Fix any constraint indexed by (y_0, y_1, S). The left-hand side is

$$LHS = \sum_{j \in S} r_{y_0 0,j} + \sum_{j \in S^c} r_{y_1 1,j}.$$

By construction, the only nonzero entries of $\mathbf{r}$ are:

$$r_{y'1,j'} = 1, \qquad r_{y0,j_y} = -1, \qquad r_{y'1,j_y} = -1.$$

Hence, LHS contains at most one positive term, namely $r_{y'1,j'}$. If this term appears (i.e., $y_1 = y'$ and $j' \in S^c$), then since $j_{y_0} \neq j'$, either the term $r_{y_0 0,j_{y_0}} = -1$ or $r_{y'1,j_{y_0}} = -1$ appear in the sums. All remaining nonzero terms are nonpositive. Therefore,

$$LHS \leq 1 - 1 = 0,$$

and the constraint is satisfied. Hence $\mathbf{r} \in \mathcal{K}$.

Step 2: Rank of the Active Constraints. First, note that the ambient dimension is $2n\ell$, and by shift invariance (see Lemma 24), the kernel has dimension at least $\ell - 1$. Hence, by Lemma 23, $\mathbf{r}$ satisfies Definition 3 if and only if

$$\operatorname{rank}(M_{\mathbf{r}}^{(0)}) \geq 2n\ell - \dim(\ker(M)) - 1.$$

We now show that the set of tight constraints at $\mathbf{r}$ has rank at least $2n\ell - \ell$.

Consider the following family of constraints, all of which are active (tight) at $\mathbf{r}$:

$$x_{y'0,i} + \sum_{j \neq i} x_{y1,j} = 0, \qquad \forall y \neq y',\ i \neq j_{y'}, \tag{54}$$

$$\sum_{j} x_{y1,j} = 0, \qquad \forall y \neq y', \tag{55}$$

$$x_{y0,i} + \sum_{j \neq i} x_{y1,j} = 0, \qquad \forall y \neq y',\ i \neq j_y, \tag{56}$$

$$\sum_{j \neq j',i} x_{y'0,j} + x_{y'1,j'} + x_{y'1,i} = 0, \qquad \forall i \neq j_{y'}, j', \tag{57}$$

$$\sum_{j \neq j',j_{y'}} x_{y''0,j} + x_{y'1,j'} + x_{y'1,j_{y'}} = 0, \qquad \forall y'' \text{ with } j_{y''} \neq j_{y'}, \tag{58}$$

$$\sum_{j \neq j',j_y} x_{y0,j} + x_{y0,j_y} + x_{y'1,j'} = 0, \qquad \forall y. \tag{59}$$

We claim that these constraints determine $\mathbf{r}$ uniquely up to scaling and addition of a vector in $\ker(M)$.

From (55), we have

$$\sum_{j} x_{y1,j} = 0 \quad \forall y \neq y'.$$

Substituting into (56) yields, for all $i \neq j_y$,

$$x_{y0,i} + \sum_{j \neq i} x_{y1,j} = x_{y0,i} - x_{y1,i} = 0,$$

hence

$$x_{y0,i} = x_{y1,i}, \qquad \forall y \neq y',\ i \neq j_y.$$

Combining with (54) yields, for each $i \neq j_y$ there exists a scalar h_i such that,

$$h_i = x_{y'0,i} = x_{y0,i} = x_{y1,i}, \qquad \forall y \neq y',\ i \neq j_y.$$

Applying in (55) yields, for all $y \neq y'$,

$$x_{y1,j_y} = -\sum_{j \neq j_y} h_j. \tag{60}$$

Substituting into (59) yields, for all $y \neq y'$,

$$x_{y1,j_y} - x_{y0,j_y} = x_{y'1,j'} - h_{j'}. \tag{61}$$

Combining with (58) yields, for all y'' with $j_{y''} \neq j_{y'}$,

$$\sum_{j \neq j_{y'}} x_{y''1,j} + x_{y'1,j_{y'}} = 0.$$

Considering (55), we obtain

$$x_{y'1,j_{y'}} = h_{j_{y'}} = x_{y'0,j_{y'}}$$

Comparing (57) and (59) for $y = y'$, we obtain

$$x_{y'1,i} = x_{y'0,i} = h_i, \qquad \forall i \neq j', j_{y'}.$$

Applying in (57) yields

$$x_{y'1,j'} = -\sum_{j \neq j'} h_j.$$

Combining with (60) and (61) results in

$$x_{y0,j_y} = h_{j_y}.$$

Thus, by fixing values of h_i for all $i \in [\ell]$, $\mathbf{x}$ is uniquely determined as follows

$$\begin{aligned}
x_{y0,j} &= h_j, && \forall y, j;\\
x_{y1,j} &= h_j, && \forall y \neq y', j \neq j_y;\\
x_{y1,j_y} &= -\sum_{j \neq j_y} h_j && \forall y \neq y';\\
x_{y'1,j} &= h_j, && \forall j \neq j';\\
x_{y'1,j'} &= -\sum_{j \neq j'} h_j;
\end{aligned}$$

Therefore, the solution space has dimension ℓ. Since $\dim(\ker(M)) \geq \ell - 1$ and $\mathbf{r}$ belongs to the solution space,

$$\text{rank}(M_{\mathbf{r}}^{(0)}) \geq 2n\ell - \ell \geq 2n\ell - \dim(\ker(M)) - 1.$$

This shows that the active constraints have maximal rank for an extreme ray, and hence $\mathbf{x}$ defines an extreme ray.

Step 3: Counting. For each (y', j'), there are $(\ell - 1)^n - (\ell - 1)$ non-constant choices of the vector $(j_0, \ldots, j_{n-1})$. Multiplying by $(n-1)\ell$ yields the total number of M-distinct rays. □

Proof of Theorem 6. Each ray presented in Proposition 8 induces the inequality

$$\mathbf{p}^\top \mathbf{r} \leq 0,$$

which is exactly inequalities in Theorem 6. In order to prove the necessity of these inequalities, it is suffice to show that for each inequality there exists a vector $\mathbf{p}$ such that all of inequalities except this specific one is satisfied, which is obtained similarly to Section C.3 since the inequalities are derived from M-distinct extreme rays. Furthermore, as in Section C.3, the basic constraints

$$p_{yd,z} \geq 0 \qquad \forall y, d, z, \quad \text{and} \quad \sum_{y,d} p_{yd,0} = \sum_{y,d} p_{yd,z}, \qquad \forall z \in [\ell],$$

are implied by extreme rays other than those characterized in Proposition 8. Hence, it follows that the vector $\mathbf{p}$ is an observed probability vector, corresponding to some distribution $\mathcal{P}$. □

E ANALYTICAL EXAMPLE

In this section, we present an explicit example for $n = \ell = 2$. We construct the corresponding matrix M, the set S, the set of vertices $\mathcal{V}$, the ATE bounds based on the distribution $\mathcal{P}$, set of extreme rays, and the associated IV inequalities.

E.1 FORMULATION

We begin by listing all the constraints given in Proposition 2:

$$\begin{aligned}
p_{00,0} &= q_{00,00} + q_{00,01} + q_{01,00} + q_{01,01} \\
p_{10,0} &= q_{10,00} + q_{10,01} + q_{11,00} + q_{11,01} \\
p_{01,0} &= q_{00,10} + q_{00,11} + q_{10,10} + q_{10,11} \\
p_{11,0} &= q_{01,10} + q_{01,11} + q_{11,10} + q_{11,11} \\
p_{00,1} &= q_{00,00} + q_{00,10} + q_{01,00} + q_{01,10} \\
p_{10,1} &= q_{10,00} + q_{10,10} + q_{11,00} + q_{11,10} \\
p_{01,1} &= q_{00,01} + q_{00,11} + q_{10,01} + q_{10,11} \\
p_{11,1} &= q_{01,01} + q_{01,11} + q_{11,01} + q_{11,11}.
\end{aligned}$$

In addition, we impose the nonnegativity constraints

$$q_{ij,\mathbf{d}} \geq 0 \qquad \text{for all } i, j \in \{0,1\},\ \mathbf{d} \in \{0,1\}^2.$$

Our objective is to maximize or minimize the linear functional

$$\mathbf{c}^\top \mathbf{q},$$

subject to the constraints $\mathbf{p} = M^\top \mathbf{q}$ and $\mathbf{q} \geq 0$, where

$$\mathbf{p} = \begin{pmatrix} p_{00,0} \\ p_{10,0} \\ p_{01,0} \\ p_{11,0} \\ p_{00,1} \\ p_{10,1} \\ p_{01,1} \\ p_{11,1} \end{pmatrix} \qquad M = \begin{array}{c|cccccccc}
 & p_{00,0} & p_{10,0} & p_{01,0} & p_{11,0} & p_{00,1} & p_{10,1} & p_{01,1} & p_{11,1} \\ \hline
q_{00,00} & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
q_{01,00} & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
q_{10,00} & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
q_{11,00} & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
q_{00,01} & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
q_{01,01} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
q_{10,01} & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
q_{11,01} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
q_{00,10} & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
q_{01,10} & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
q_{10,10} & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
q_{11,10} & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
q_{00,11} & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
q_{01,11} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
q_{10,11} & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
q_{11,11} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{array} \qquad \mathbf{q} = \begin{pmatrix} q_{00,00} \\ q_{01,00} \\ q_{10,00} \\ q_{11,00} \\ q_{00,01} \\ q_{01,01} \\ q_{10,01} \\ q_{11,01} \\ q_{00,10} \\ q_{01,10} \\ q_{10,10} \\ q_{11,10} \\ q_{00,11} \\ q_{01,11} \\ q_{10,11} \\ q_{11,11} \end{pmatrix} \qquad \mathbf{c} = \begin{pmatrix} 0 \\ \gamma_1 - \gamma_0 \\ \gamma_0 - \gamma_1 \\ 0 \\ 0 \\ \gamma_1 - \gamma_0 \\ \gamma_0 - \gamma_1 \\ 0 \\ 0 \\ \gamma_1 - \gamma_0 \\ \gamma_0 - \gamma_1 \\ 0 \\ 0 \\ \gamma_1 - \gamma_0 \\ \gamma_0 - \gamma_1 \\ 0 \end{pmatrix}$$

E.2 EXAMPLE OF SUBMATRIX SELECTION

We illustrate the notation $M(\alpha, \beta)$ on the following matrix for Definition 9:

$$
\begin{array}{c|cccccccc}
 & p_{00,0} & p_{10,0} & p_{01,0} & p_{11,0} & p_{00,1} & p_{10,1} & p_{01,1} & p_{11,1} \\
\hline
q_{00,00} & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
q_{01,00} & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
q_{10,00} & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
q_{11,00} & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
q_{00,01} & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
q_{01,01} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
q_{10,01} & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
q_{11,01} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
q_{00,10} & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
q_{01,10} & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
q_{10,10} & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
q_{11,10} & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
q_{00,11} & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
q_{01,11} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
q_{10,11} & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
q_{11,11} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{array}
$$

The dark gray entries correspond exactly to the submatrix

$$M(0*01 \cup 0*10, **1).$$

The light grey entries correspond exactly to the submatrix

$$M(****, 110 \cup 000).$$

E.3 VERTICES AND ATE BOUNDS

We can represent each $\mathbf{b} \in \{0,1\}^{2\times2\times2} \in S$ (defined in Definition 4) as the horizontal vector ordered as

$$(b_{000},\, b_{100},\, b_{010},\, b_{11,0},\, b_{001},\, b_{101},\, b_{011},\, b_{111}).$$

in

$$
S = \left\{
\begin{array}{l}
S_1 = \left\{ \begin{array}{l} (1,1,0,1,\ 1,1,1,0), \\ (1,1,1,0,\ 1,1,0,1), \end{array} \right\} \\
S_2 = \left\{ \begin{array}{l} (0,1,1,0,\ 1,1,1,1), \\ (1,1,1,0,\ 0,1,1,1), \\ (0,1,1,1,\ 1,1,1,0), \\ (1,1,1,1,\ 0,1,1,0), \end{array} \right\} \\
S_2 = \left\{ \begin{array}{l} (0,1,1,1,\ 1,0,1,1), \\ (1,0,1,1,\ 0,1,1,1) \end{array} \right\}
\end{array}
\right\}
$$

Now based on Definition 5, we can create the vertex set $\mathcal{V}$ defined in Theorem 1:

$$
\mathcal{V} = \left\{
\begin{array}{l}
(\ \ 0,-1,-1,\ \ 1,\ \ 0,-1,\ \ 0,-1), \\
(\ \ 0,-1,\ \ 0,-1,\ \ 0,-1,-1,\ \ 1), \\
(-1,-1,-1,-1,\ \ 1,\ \ 0,\ \ 0,\ \ 1), \\
(\ \ 0,-1,-1,-1,\ \ 0,\ \ 0,\ \ 0,\ \ 1), \\
(-1,-1,-1,\ \ 0,\ \ 1,\ \ 0,\ \ 0,\ \ 0), \\
(\ \ 0,-1,-1,\ \ 0,\ \ 0,\ \ 0,\ \ 0,\ \ 0), \\
(-1,\ \ 0,-1,\ \ 0,\ \ 1,-1,-1,\ \ 0), \\
(\ \ 1,-1,-1,\ \ 0,-1,\ \ 0,-1,\ \ 0).
\end{array}
\right\}
$$

So based on Theorem 2 we have

$$\max\left\{\begin{array}{l} -p_{10,0}-p_{01,0}+p_{11,0}-p_{10,1}-p_{11,1}, \\ -p_{10,0}-p_{11,0}-p_{10,1}-p_{01,1}+p_{11,1}, \\ -p_{00,0}-p_{10,0}-p_{01,0}-p_{11,0}+p_{00,1}+p_{11,1}, \\ -p_{10,0}-p_{01,0}-p_{11,0}+p_{11,1}, \\ -p_{00,0}-p_{10,0}-p_{01,0}+p_{00,1}, \\ -p_{10,0}-p_{01,0}, \\ -p_{00,0}-p_{01,0}+p_{00,1}-p_{10,1}-p_{01,1}, \\ \;p_{00,0}-p_{10,0}-p_{01,0}-p_{00,1}-p_{01,1} \end{array}\right\} \leq \text{ATE} \leq \min\left\{\begin{array}{l} +p_{11,0}+p_{00,0}-p_{10,0}+p_{11,1}+p_{10,1}, \\ +p_{11,0}+p_{10,0}+p_{11,1}+p_{00,1}-p_{10,1}, \\ +p_{01,0}+p_{11,0}+p_{00,0}+p_{10,0}-p_{01,1}-p_{10,1}, \\ +p_{11,0}+p_{00,0}+p_{10,0}-p_{10,1}, \\ +p_{01,0}+p_{11,0}+p_{00,0}-p_{01,1}, \\ +p_{11,0}+p_{00,0}, \\ +p_{01,0}+p_{00,0}-p_{01,1}+p_{11,1}+p_{00,1}, \\ -p_{01,0}+p_{11,0}+p_{00,0}+p_{01,1}+p_{00,1} \end{array}\right\}$$

which are equivalent to the bounds introduced in Balke [1995].

E.4 EXTREME RAYS AND TESTABLE IMPLICATIONS

We now specialize Theorem 5 to the case $n = \ell = 2$. By Theorem 5, the IV model is valid if and only if the observed distribution $\mathbf{p}$ satisfies the finite family of inequalities induced by the extreme rays of $\mathcal{K}$.

We represent each non trivial extreme ray as the horizontal vector ordered with respect to

$$(p_{00,0},\, p_{10,0},\, p_{01,0},\, p_{11,0},\, p_{00,1},\, p_{10,1},\, p_{01,1},\, p_{11,1}).$$

$$\begin{array}{l} (\;\;\,0,\;\;\,0,\;\;\,1,\;\;\,0,-1,-1,-1,\;\;\,0), \\ (-1,\;\;\,0,-1,-1,\;\;\,1,\;\;\,0,\;\;\,0,\;\;\,0), \\ (\;\;\,0,-1,-1,-1,\;\;\,0,\;\;\,1,\;\;\,0,\;\;\,0), \\ (-1,-1,-1,\;\;\,0,\;\;\,0,\;\;\,0,\;\;\,1,\;\;\,0). \end{array}$$

Explicit Inequalities for $n = 2$. Instantiating above extreme rays into $\mathbf{p}^\top \mathbf{r} \leq 0$ (equivalently (17) for $n = 2$), we obtain the following system:

$$-p_{00,1} - p_{10,1} + p_{01,0} - p_{01,1} \leq 0;$$

$$\begin{array}{l} -p_{01,0} - p_{11,0} + p_{00,1} - p_{00,0} \leq 0, \\ -p_{01,0} - p_{11,0} + p_{10,1} - p_{10,0} \leq 0; \end{array}$$

$$-p_{00,0} - p_{10,0} + p_{01,1} - p_{01,0} \leq 0;$$

which by using $(p_{00,1} + p_{10,1} + p_{01,1} + p_{11,1}) = (p_{00,0} + p_{10,0} + p_{01,0} + p_{11,0}) = 1$ are equivalent to

$$\begin{array}{l} p_{01,0} + p_{11,1} \leq 1, \\ p_{00,1} + p_{10,0} \leq 1, \\ p_{10,1} + p_{00,0} \leq 1, \\ p_{01,1} + p_{11,0} \leq 1. \end{array}$$

Therefore, producing a total of $2^{n+1} - 4 = 4$ inequalities as follow:

$$\begin{array}{l} p_{01,0} + p_{11,1} \leq 1, \\ p_{00,1} + p_{10,0} \leq 1, \\ p_{10,1} + p_{00,0} \leq 1, \\ p_{01,1} + p_{11,0} \leq 1. \end{array}$$

which are equivalent to the bounds introduced in Pearl [1995].