Proceedings of Machine Learning Research 343:758–779, 2026 Probabilistic Graphical Models (PGM)

Partial Identification under Causal Orders by Linear Programming

Eric Rossetto ERIC.ROSSETTO@SUPSI.CH

Alessandro Antonucci ALESSANDRO.ANTONUCCI@SUPSI.CH

Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA)
Scuola Universitaria Professionale della Svizzera Italiana (SUPSI)
Lugano, Switzerland

Editors: Gustau Camps-Valls, Manuele Leonelli and Gherardo Varando

Abstract

Non-parametric (partial) identification of counterfactual queries typically relies on a fully specified causal graph. Motivated by settings with incomplete domain knowledge, we challenge this requirement by leveraging structural assumptions that are inherently implied by the query itself. We show that any counterfactual inquiry induces a, mostly partial, topological ordering over relevant variables, which, in turn, enables an explicit query parametrisation reducing the identification task to a linear program. This allows bounding arbitrary counterfactual and nested counterfactual queries. Our work can be viewed as a generalisation of the classical bounding framework of Tian and Pearl (2000), originally developed for probabilities of causation. We also prove the *tightness* of our bounds by constructing structural causal models that attain the bounds whilst being compatible with both the observed data and the query-implied order. To assess both the generality and practical utility of the proposed bounding procedure, we revisit several case studies from the literature, demonstrating how the derived bounds can be used to yield informative insights even in the absence of an input causal graph.

Keywords: Structural Causal Models, Partial Identifiability, Linear Programming.

1. Introduction

Causal reasoning deepens our understanding of the effects of policies, actions, and treatments in domains that require accountable decision-making, such as algorithmic fairness, personalised medicine, economics, and the broader social sciences. Quantifying such causal inquiries demands a rigorous mathematical foundation. The work of Pearl (2009) formalised the systems under study using structural causal models (SCMs), which elegantly encode prior causal beliefs via directed graphs (Pearl, 1995). Ideally, these structural assumptions allow causal queries to be uniquely computed—formally *identified*—from the available set of observations. In many practical settings, however, the available assumptions are insufficient for point identification, necessitating *partial identification*: computing a range within which the target quantity is guaranteed to lie.

Early work by Manski (1990) derived closed-form bounds for simple treatment effect estimation without relying on structural assumptions. Similarly, Tian and Pearl (2000) formulated a linear program to compute sharp bounds on the *probabilities of causation* (Pearl, 1999) from experimental and observational data under either general or mild assumptions.

© 2026 E. Rossetto & A. Antonucci.

Whilst weak structural assumptions typically yield substantially wide bounds, assuming a fully specified causal graph might produce tighter intervals. Graph-based reductions to linear programs were first developed for the instrumental variable setting (Balke and Pearl, 1994a,b) and later extended to generalised instrumental constraints (Sachs et al., 2023). For general graphical assumptions, partial identification is reduced to a polynomial program via *canonicalisation*, though the exponential growth of the resulting exogenous state space ultimately necessitates approximate methods (Zhang et al., 2022; Duarte et al., 2024).

We revisit partial identification in the spirit of Tian and Pearl (2000), whose bounding method did not require a specification of a causal graph; instead, their proposal relied on a direct parametrisation of the joint counterfactual distribution. To extend this framework to arbitrary queries, we leverage the minimal structural assumptions inherently implied by the query itself, reducing the identification task to a linear program without imposing additional structural commitments. Our specific contributions are as follows: (i) we reduce partial identification under a total order of variables to a linear program, capable of handling arbitrary counterfactual, including nested, queries; (ii) we prove that the resulting bounds are *tight* by constructing witnessing structural causal models that attain them; (iii) we prove that the order-dependent optimisation task is invariant under both the marginalisation of unqueried variables and the specific choice of total order (compatible with the query); and (iv) we show that the linear program admits a lifted reformulation by exploiting symmetries induced by the observations and the query's logical evaluations. The presented results define an automated framework that enables the bounding of arbitrary queries without the need for explicit graphical assumptions. It is important to emphasise that graph-based methods generally yield narrower bounds by exploiting conditional independencies; in contrast, we pose ourselves in a regime of incomplete domain knowledge, abstaining from such graphical commitments.

After introducing the necessary background in Sect. 2, Sect. 3 reduces partial identification to a linear programme, which Sect. 4 extends to accommodate conditional queries, nested counterfactuals, and structural constraints. Sect. 5 demonstrates practical case studies, followed by conclusions in Sect. 6; proofs, additional experiments and discussions are in Apps. A, B and C, respectively.

2. Background

Notation. Let us first review some necessary background on causality. We denote variables by capital letters, whilst small and calligraphic letters are used instead for the states and the sample spaces. Thus, $x \in \mathcal{X}$ is a state of X. We consider discrete variables only. Bold is used for sets of variables. The indicator function that is one when its argument is true and zero otherwise is denoted as $[\![\cdot]\!]$, whilst $|\cdot|$ is the cardinality of the set in the argument.

Structural Causal Models (SCMs). Following Pearl (2009), an SCM $\mathcal{M} \coloneqq \langle \boldsymbol{V}, \boldsymbol{U}, \mathcal{F}, P(\boldsymbol{U}) \rangle$ is such that $\boldsymbol{V}$ is its set of endogenous variables, $\boldsymbol{U}$ the set of exogenous variables distributed according to $P(\boldsymbol{U})$, and $\mathcal{F}$ a set $\{f_V\}_{V \in \boldsymbol{V}}$ of *structural equations* (SEs). The SE of each $V \in \boldsymbol{V}$ determines $v = f_V(\mathrm{pa}_V, \boldsymbol{u}_V)$, where $\mathrm{pa}_V \in \mathrm{Pa}_V \subseteq \boldsymbol{V} \setminus \{V\}$ and $\boldsymbol{U}_V \subseteq \boldsymbol{U}$ are, respectively, the endogenous and the exogenous inputs of the SE. We may view an SE as a collection of deterministic mechanisms, or mappings, from input configurations to output states. If $P(\boldsymbol{U})$ is unavailable, $M \coloneqq \langle \boldsymbol{V}, \boldsymbol{U}, \mathcal{F} \rangle$ is termed *partially specified* SCM (PSCM).

Both SCM $\mathcal{M}$ and PSCM M induce a directed graph $\mathcal{G}$ whose nodes represent variables and an edge (X, Y) exists iff $X \in \text{Pa}_Y$. We restrict our analysis to *recursive* SCMs, i.e., such that $\mathcal{G}$ is *acyclic*. We adopt a kinship terminology (e.g., *parents*) to present graphical relations between variables. The notation $\mathbb{M}_M$ is finally used for the set of all the SCMs that share the SEs' signature, i.e., parent and child variables, with M.

Interventions. In an SCM $\mathcal{M}$, $P(\boldsymbol{U})$ and $\mathcal{F}$ induce a joint *observational* distribution $P_{\mathcal{M}}(\boldsymbol{V})$. An SCM *natural* regime can be modified by acting on its SEs. For an arbitrary set $\boldsymbol{X} \subseteq \boldsymbol{V}$, an *intervention* yields a new model $\mathcal{M}_{\boldsymbol{x}}$ (i.e., a *sub-model*) by replacing the SEs of $\boldsymbol{X}$ with constant assignments $\boldsymbol{X} \leftarrow \boldsymbol{x}$, whilst keeping $P(\boldsymbol{U})$ and all other SEs unchanged. The resulting, *interventional*, distribution is denoted as $P_{\mathcal{M}_{\boldsymbol{x}}}(\boldsymbol{V} \setminus \boldsymbol{X})$. For any $Y \in \boldsymbol{V}$, the *potential response* $Y_{\boldsymbol{x}}(\boldsymbol{u})$ denotes the value of Y in $\mathcal{M}_{\boldsymbol{x}}$ given $\boldsymbol{U} = \boldsymbol{u}$. A *counterfactual* variable (or *potential outcome*) $Y_{\boldsymbol{x}}$ is the variable induced by $Y_{\boldsymbol{x}}(\boldsymbol{u})$ when $\boldsymbol{u} \sim P(\boldsymbol{U})$.

Counterfactuals. Following the logical formalisation of a PSCM by Halpern (2000), we define an *atomic proposition* as the assignment $Y_{\boldsymbol{x}} = y$ (for short, $y_{\boldsymbol{x}}$), where $Y \in \boldsymbol{V}, \boldsymbol{X} \subseteq \boldsymbol{V}$ and $y \in \mathcal{Y}$. A *counterfactual event* γ is a conjunction of k atomic propositions, $\gamma \equiv y^{(1)}_{\boldsymbol{x}^{(1)}} \wedge \cdots \wedge y^{(k)}_{\boldsymbol{x}^{(k)}}$. A *counterfactual query* $P_{\mathcal{M}}(\gamma)$ asks for the probability of γ with respect to an SCM $\mathcal{M}$. This requires the simultaneous computation of potential responses of multiple sub-models $\{\mathcal{M}_{\boldsymbol{x}^{(i)}}\}_{i=1}^{k}$, coupled by the shared $P(\boldsymbol{U})$, i.e.,

$$P_{\mathcal{M}}(\gamma) = \sum_{\boldsymbol{u} \in \boldsymbol{\mathcal{U}}} \left[\!\!\left[\bigwedge_{i=1}^{k} Y^{(i)}_{\boldsymbol{x}^{(i)}}(\boldsymbol{u}) = y^{(i)} \right]\!\!\right] \cdot P(\boldsymbol{u})\,. \tag{1}$$

The evaluation accumulates over the exogenous instantiations $\boldsymbol{u}$ satisfying the formula γ. We denote such an *entailment* as $(\mathcal{M}, \boldsymbol{u}) \models \gamma$. When it is clear from the context which SCM we refer to, we omit specifying it and rewrite the right-hand side of Eq. (1) as $\sum_{\boldsymbol{u}} [\![\boldsymbol{u} \models \gamma]\!] P(\boldsymbol{u})$. If all the subscripts agree on $\boldsymbol{x}$, then Eq. (1) can be retrieved from the interventional distribution $P_{\mathcal{M}_{\boldsymbol{x}}}$. In general, this is not the case and we eventually characterise a *counterfactual* distribution. Finally, we denote as $\boldsymbol{V}_\gamma \subseteq \boldsymbol{V}$ the endogenous, no matter whether queried or intervened, variables in γ.

Nested Counterfactuals. We can further generalise the intervention for the counterfactual variable Y_x to allow representing settings where the variable X is set to behave as another variable, say X_z. A random variable Y in such system is represented with a potential outcome of the form $Y_{X_z} = y$, which is called a *nested counterfactual*. This expresses the statement "Y becomes y when we set X to whatever value it would have taken if we had set Z to z". Queries involving nested counterfactuals can be rewritten into a marginalisation over conjunctions of standard, atomic counterfactual events (Correa et al., 2021), i.e.,

$$(Y_{X_z} = y) \equiv \bigvee_{x \in \mathcal{X}} (Y_x = y \wedge X_z = x)\,. \tag{2}$$

(Partial) Identifiability. Eq. (1) allows to compute counterfactual queries in an SCM $\mathcal{M}$. Yet, the exogenous distribution $P(\boldsymbol{U})$ is rarely available, and one must cope with the corresponding PSCM M and a dataset $\mathcal{D}$ of endogenous observations, used to estimate an *empirical* distribution $\hat{P}(\boldsymbol{V})$. In general, there are multiple SCMs in $\mathbb{M}_M$, that could have

generated $\hat{P}(\boldsymbol{V})$ and yield different values to a query over γ. In such a *partially identifiable* setting, we can only compute bounds to the query, i.e., the minimum:

$$\min_{\mathcal{M}\in\mathbb{M}_M} P_{\mathcal{M}}(\gamma) \quad \text{s.t.} \quad P_{\mathcal{M}}(\boldsymbol{V}) = \hat{P}(\boldsymbol{V})\,, \tag{3}$$

and the maximum, which is analogously defined. In the remainder of the paper, just for the sake of brevity, we focus on the minimisation problem, noticing that the maximisation counterpart can be handled analogously.

Those optimisation tasks have been shown to be NP-hard (Zaffalon et al., 2024). Nevertheless, approximate techniques were proposed to estimate bounds efficiently (Zhang et al., 2022; Zaffalon et al., 2024; Duarte et al., 2024; Bjøru et al., 2025). Standard (point) *identifiability* emerges as a special case when the two bounds coincide.

Queries and Partial Orders. Although potential outcome semantics allow us to consider a variable Y_x for any $Y, X \in \boldsymbol{V}$ and $x \in \mathcal{X}$, an intervention on X can alter the realisation of Y only if there exists a directed path from X to Y in the causal graph induced by the assumed SCM $\mathcal{M}$ over $\boldsymbol{V}$. This is formalised as follows.

Proposition 1 (Tian, 2002) *If $Y, X \in \boldsymbol{V}$ are distinct endogenous variables of an SCM $\mathcal{M}$ such that $X \notin \mathrm{Anc}(Y)$, then, $\forall \boldsymbol{u}$, $Y_x(\boldsymbol{u}) = Y(\boldsymbol{u})$, and thus $P_{\mathcal{M}}(Y_x) = P_{\mathcal{M}}(Y)$.*

Thus, a counterfactual variable Y_x can be non-trivial only if X precedes Y in the underlying causal graph. In this paper, we restrict attention to non-trivial queries: consequently, when the graph is not available, a query over γ can be also used to induce a collection of precedence constraints, corresponding to a partial order $\prec_\gamma$ on $\boldsymbol{V}_\gamma$. A "cyclic" event such as $y_x \wedge x_y$ would be ill-posed as it induces incompatible precedence constraints. Such cases would naturally call for non-recursive models (see, for instance, Cozman et al., 2025 and Bongers et al., 2021), which fall outside the scope of this paper.

3. Bounding Framework

The main goal of this paper is to relax the optimisation task in Eq. (3) to cope with situations where the underlying PSCM M is unavailable. We start by assuming that the endogenous variables are sorted by a total order σ, corresponding to $\boldsymbol{V} \coloneqq (V^{(1)}, \ldots, V^{(k)})$. In this setting, we can replace in Eq. (3) the set $\mathbb{M}_M$ of SCMs compatible with the PSCM M with the set $\mathbb{M}_\sigma$ of *all* SCMs whose causal graph is consistent with σ, i.e.,

$$\min_{\mathcal{M}\in\mathbb{M}_\sigma} P_{\mathcal{M}}(\gamma) \quad \text{s.t.} \quad P_{\mathcal{M}}(\boldsymbol{V}) = \hat{P}(\boldsymbol{V})\,. \tag{4}$$

For the moment, we consider γ such that $\boldsymbol{V}_\gamma = \boldsymbol{V}$, i.e., all endogenous variables appear in the query. A more general account of the case $\boldsymbol{V}_\gamma \subset \boldsymbol{V}$ is addressed later.

Response Representation. Unlike Eq. (3), the optimisation task in Eq. (4) refers to a collection of SCMs possibly based on different PSCMs. Yet, given only σ and $\boldsymbol{V}$, we consider a *complete canonical* PSCM over $\boldsymbol{V}$ denoted as M_σ and such that: (i) $\mathrm{Pa}_{V^{(i)}}$ coincides with all the predecessors of $V^{(i)}$ according to σ and they are denoted as $\boldsymbol{V}_{\prec i}$; (ii) there is a single exogenous variable U, which is a parent of all the endogenous variables; (iii) the SEs of

each $V \in \boldsymbol{V}$ and the cardinality of U are such that M_σ is *canonical* in the sense of Zhang et al. (2022), this basically meaning that the exogenous states of U are indexing *all* possible structural relations between the endogenous variables and their endogenous parents (see App. C for a more detailed account).

The optimisation task in Eq. (4) can be equivalently achieved by considering the complete canonical PSCM M_σ only, as stated by the following result.

Theorem 2 *Eq. (3) with $M = M_\sigma$ gives the same minimum of Eq. (4), i.e.,*

$$\min_{\mathcal{M} \in \mathbb{M}_\sigma} P_{\mathcal{M}}(\gamma) = \min_{\mathcal{M} \in \mathbb{M}_{M_\sigma}} P_{\mathcal{M}}(\gamma)\,, \tag{5}$$

where $P_{\mathcal{M}}(\boldsymbol{V}) = \hat{P}(\boldsymbol{V})$ is required on both sides.

The above result reduces the optimisation in Eq. (4)—which considers a collection of structurally heterogenous SCMs—to span only those SCMs sharing the same causal graph as M_σ. This reduction provides a parametrisation of the task, where the optimisation variables are associated with the exogenous probabilities as in Eq. (1). To avoid explicitly modelling these latent variables, we build upon the approach proposed by Balke and Pearl (1994a,b). Here, we consider all the potential outcomes that M_σ is allowed to generate as to enumerate all possible endogenous mappings from parents to child. This is achieved by considering a *response signature* $\boldsymbol{S}_\sigma$ for M_σ, corresponding to the following collection of potential outcomes:

$$\boldsymbol{S}_\sigma \coloneqq \left\{ V^{(i)}_{\boldsymbol{v}_{\prec i}} \right\}^{i=1,\dots,k}_{\boldsymbol{v}_{\prec i} \in \boldsymbol{\mathcal{V}}_{\prec i}}\,. \tag{6}$$

The following result shows that the probability of any counterfactual query can be expressed not only as a linear combination of joint probabilities over the exogenous variables, as in Eq. (1), but also as a linear combination of joint probabilities over $\boldsymbol{S}_\sigma$.

Theorem 3 *A query $P_{\mathcal{M}}(\gamma)$ in an SCM $\mathcal{M} \in \mathbb{M}_\sigma$ can be written as a linear combination of the probabilities of the response signature, i.e.,*

$$P_{\mathcal{M}}(\gamma) = \sum_{\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma} [\![\boldsymbol{s}_\sigma \models \gamma]\!] \cdot P(\boldsymbol{s}_\sigma)\,. \tag{7}$$

By viewing observational queries as a special case of counterfactual queries in which interventions are performed on the empty set, we also obtain the following additional result.

Corollary 4 *Any observational probability $P_{\mathcal{M}}(\boldsymbol{v})$ in an SCM $\mathcal{M} \in \mathbb{M}_\sigma$ can be written as a linear combination of the probabilities of the response signature, i.e.,*

$$P_{\mathcal{M}}(\boldsymbol{v}) = \sum_{\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma} [\![\boldsymbol{s}_\sigma \models \boldsymbol{v}]\!] \cdot P(\boldsymbol{s}_\sigma)\,. \tag{8}$$

Linear Programming (LP) Formulation. Thm. 3 and Cor. 4 allow us to address the optimisation task in Eq. (4) via the signature joint states. The resulting linearity of the constraints and the objective function allows us to rewrite the bounding task as a LP task:

$$\begin{aligned} \min \quad & \sum_{\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma} [\![\boldsymbol{s}_\sigma \models \gamma]\!] \cdot q_{\boldsymbol{s}_\sigma} \\ \text{s.t.} \quad & \sum_{\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma} [\![\boldsymbol{s}_\sigma \models \boldsymbol{v}]\!] \cdot q_{\boldsymbol{s}_\sigma} = \hat{P}(\boldsymbol{v})\,, \forall \boldsymbol{v} \in \boldsymbol{\mathcal{V}}\,; \quad \sum_{\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma} q_{\boldsymbol{s}_\sigma} = 1\,; \quad q_{\boldsymbol{s}_\sigma} \geq 0\,, \forall \boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma\,. \end{aligned} \tag{9}$$

with semantics $q_{\boldsymbol{s}_\sigma} = P(\boldsymbol{s}_\sigma)$ for the optimisation variables.

The following result establishes that the above LP task solves the relaxed optimisation over SCMs consistent with σ by yielding *tight* identification bounds.

Theorem 5 *The LP task in Eq. (9) gives the solution of Eq. (4). Furthermore, there exists an SCM $\mathcal{M} \in \mathbb{M}_\sigma$ such that $P_{\mathcal{M}}(\gamma)$ coincides with the solution of the LP.*

Signature State Partitioning. The latent space reduction to finite response-function classes (Balke and Pearl, 1994a; Duarte et al., 2024) enables us to neatly solve Eq. (4) through Eq. (9). However, a further reduction can be achieved by observing that the LP in Eq. (9) depends exclusively on the Boolean evaluations of the signatures against the observations $\boldsymbol{v}$ and the counterfactual event γ. Signatures states that evaluate identically against these observations and events can be grouped into equivalence classes $\boldsymbol{t}_\sigma \in \boldsymbol{\mathcal{T}}_\sigma := \boldsymbol{\mathcal{S}}_\sigma / \sim_\gamma$, where:

$$\boldsymbol{s}_\sigma \sim_\gamma \boldsymbol{s}'_\sigma \iff \big((\boldsymbol{s}_\sigma \models \gamma) \equiv (\boldsymbol{s}'_\sigma \models \gamma)\big) \wedge \big(\exists \boldsymbol{v} \in \boldsymbol{\mathcal{V}} \text{ s.t. } (\boldsymbol{s}_\sigma \models \boldsymbol{v}) \wedge (\boldsymbol{s}'_\sigma \models \boldsymbol{v})\big)\,. \tag{10}$$

Each class $\boldsymbol{t}_\sigma \in \boldsymbol{\mathcal{T}}_\sigma$ inherits the binary evaluations from the signatures it includes, hence $[\![\boldsymbol{t}_\sigma \models \boldsymbol{v}]\!]$ and $[\![\boldsymbol{t}_\sigma \models \gamma]\!]$. By defining aggregated masses as variables, $q_{\boldsymbol{t}_\sigma} := \sum_{\boldsymbol{s}_\sigma \in \boldsymbol{t}_\sigma} q_{\boldsymbol{s}_\sigma}$, and applying Thm. 3 and Cor. 4, the LP in Eq. (9) reduces to a *lifted* LP task (Kersting et al., 2017):

$$\begin{aligned} \min \quad & \sum_{\boldsymbol{t}_\sigma \in \boldsymbol{\mathcal{T}}_\sigma} [\![\boldsymbol{t}_\sigma \models \gamma]\!] \cdot q_{\boldsymbol{t}_\sigma} \\ \text{s.t.} \quad & \sum_{\boldsymbol{t}_\sigma \in \boldsymbol{\mathcal{T}}_\sigma} [\![\boldsymbol{t}_\sigma \models \boldsymbol{v}]\!] \cdot q_{\boldsymbol{t}_\sigma} = \hat{P}(\boldsymbol{v})\,, \forall \boldsymbol{v} \in \boldsymbol{\mathcal{V}}\,; \quad \sum_{\boldsymbol{t}_\sigma \in \boldsymbol{\mathcal{T}}_\sigma} q_{\boldsymbol{t}_\sigma} = 1\,; \quad q_{\boldsymbol{t}_\sigma} \geq 0\,, \forall \boldsymbol{t}_\sigma \in \boldsymbol{\mathcal{T}}_\sigma\,. \end{aligned} \tag{11}$$

The equivalence mapping improves the tractability of the LP in Eq. (9) whilst maintaining both soundness and tightness, as it follows from the following result.

Theorem 6 *The lifted LP task in Eq. (11) gives the solution of Eq. (4). Furthermore, there exists an SCM $\mathcal{M} \in \mathbb{M}_\sigma$ such that $P_{\mathcal{M}}(\gamma)$ coincides with the solution of the LP.*

Marginalising Non-Queried Variables. So far, we assumed that all the endogenous variables to appear in the query, i.e., $\boldsymbol{V}_\gamma = \boldsymbol{V}$. We now extend our results to the more general case in which $\boldsymbol{V}_\gamma \subseteq \boldsymbol{V}$. To this end, let $\mathbb{M}_\sigma^{\boldsymbol{V}_\gamma}$ denote the set of SCMs defined only over $\boldsymbol{V}_\gamma$ and compatible with σ. Observe that $\mathbb{M}_\sigma^{\boldsymbol{V}} = \mathbb{M}_\sigma$ and, in general, $\mathbb{M}_\sigma^{\boldsymbol{V}_\gamma} \subseteq \mathbb{M}_\sigma$, where, for notational ease, we assumed that the endogenous variables might not appear explicitly in an SCM, thus being uniformly distributed. The following result holds.

Theorem 7 *The solution of Eq. (4) coincides with:*

$$\min_{\mathcal{M} \in \mathbb{M}_\sigma^{\boldsymbol{V}_\gamma}} P_{\mathcal{M}}(\gamma) \quad \textit{s.t.} \quad P_{\mathcal{M}}(\boldsymbol{V}_\gamma) = \hat{P}(\boldsymbol{V}_\gamma)\,. \tag{12}$$

where $\hat{P}(\boldsymbol{V}_\gamma)$ is obtained from $\hat{P}(\boldsymbol{V})$ by marginalisation.

As a consequence of the above result, when $\boldsymbol{V}_\gamma \subset \boldsymbol{V}$, we marginalise the empirical distribution before applying the LP formulation to solve Eq. (12) and hence Eq. (4).

Coping with Partial Orders. The optimisation task in Eq. (4) assumes a total order σ over the endogenous variables. However, as discussed in Sect. 2, a counterfactual query over γ might induce only a partial order $\prec_\gamma$. Nevertheless, a general formulation is readily obtained by considering the bound over all linear extensions $\Sigma(\prec_\gamma)$:

$$\min_{\sigma\in\Sigma(\prec_\gamma)} \min_{\mathcal{M}\in\mathbb{M}_\sigma^{\boldsymbol{V}_\gamma}} P_{\mathcal{M}}(\gamma) \quad \text{s.t.} \quad P_{\mathcal{M}}(\boldsymbol{V}_\gamma) = \hat{P}(\boldsymbol{V}_\gamma)\,. \tag{13}$$

When γ contains no interventions, $\prec_\gamma$ imposes no ordering constraints, so that all possible total orders must be considered, and we have $|\Sigma(\prec_\gamma)| = |\boldsymbol{V}_\gamma|!$. Interventions restrict this search space by ruling out inconsistent orderings; nonetheless, the number of linear extensions can still scale exponentially with respect to $|\boldsymbol{V}_\gamma|$. In spite of this, the following result establishes that evaluating an arbitrary linear extension is sufficient, thus eliminating the need for a brute-force optimisation over all the orders in $\Sigma(\prec_\gamma)$.

Theorem 8 *Consider a counterfactual query over γ inducing a partial order $\prec_\gamma$. Any total order $\sigma \in \Sigma(\prec_\gamma)$ gives the same minimum in Eq. (12) and, hence, the solution of Eq. (13).*

As a consequence of Thm. 8, we are not required to assume or enforce any specific total order. The choice of a particular σ serves only as a possible parametrisation of the task.

Complexity. Thm. 5 enables us to solve the optimisation problem in Eq. (4) using a linear solver that implements the LP defined in Eq. (9). It is straightforward to verify that, if the number of variables in the query is $k \coloneqq |\boldsymbol{V}_\gamma|$ (see Thm. 7) and $n \coloneqq \max_{i=1}^{k} |\mathcal{V}^{(i)}|$ denotes the maximum endogenous cardinality, the number of potential outcomes comprising the signature $\boldsymbol{S}_\sigma$ in Eq. (6) is bounded by $\sum_{i=0}^{k-1} n^i = \frac{n^k-1}{n-1}$. The number of joint signature states $|\boldsymbol{\mathcal{S}}_\sigma|$, which dictates the number of optimisation variables in the naïve LP of Eq. (9), is therefore double exponential, namely $\mathcal{O}(n^{n^{k-1}})$. In contrast, by Thm. 6, the size of the quotient space $|\boldsymbol{\mathcal{T}}_\sigma|$, which governs the size of the lifted LP in Eq. (11), is bounded by $\mathcal{O}\big(\prod_{i=1}^{k} |\mathcal{V}^{(i)}|\big) = \mathcal{O}(n^k)$. Thus, to partially identify any query given the empirical marginal $\hat{P}(\boldsymbol{V}_\gamma)$, we only need to solve Eq. (11), whose optimisation space scales linearly with the size of the sample space of $\boldsymbol{V}_\gamma$.

4. Extensions

In the current section, we present some important extensions of the LP mapping introduced in the previous section.

Conditional Queries. To bound a conditional query $P(\gamma \mid \delta) \coloneqq P(\gamma \wedge \delta)/P(\delta)$, it suffices to note that both terms decompose linearly over $\boldsymbol{\mathcal{S}}_\sigma$ via Thm. 3 and Cor. 4. This reformulates the bounding task as a *linear-fractional* programme, which is reduced to a standard LP via the Charnes-Cooper transformation (Charnes and Cooper, 1962). We incorporate evidence by refining the equivalence relation in Eq. (10) to evaluate both $\gamma \wedge \delta$ and δ:

$$\boldsymbol{s}_\sigma \sim_{\gamma|\delta} \boldsymbol{s}'_\sigma \iff \boldsymbol{s}_\sigma \sim_{\gamma\wedge\delta} \boldsymbol{s}'_\sigma \wedge \big((\boldsymbol{s}_\sigma \models \delta) \equiv (\boldsymbol{s}'_\sigma \models \delta)\big)\,. \tag{14}$$

Letting $b_\psi \coloneqq [\![\boldsymbol{s}_\sigma \models \psi]\!]$, the logical constraint $\gamma \wedge \delta \Rightarrow \delta$ enforces $b_{\gamma\wedge\delta} \leq b_\delta \in \{0,1\}$. This restricts evaluations to three valid Boolean states: $(b_{\gamma\wedge\delta}, b_\delta) \in \{(0,0),(0,1),(1,1)\}$. The optimisation space retains the same asymptotic size discussed in Sect. 3.

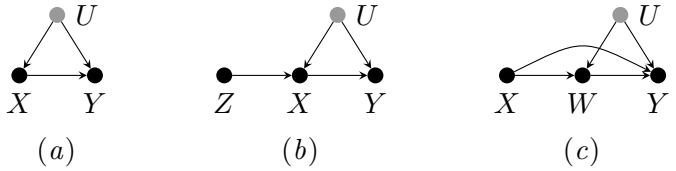


Figure 1: Causal diagrams considered in the examples. (a) Confounded treatment-outcome pair. (b) Instrumental Variable (IV) setting, where Z serves as an instrument for X. (c) Mediation model with a direct effect and confounding between the mediator W and the outcome. Grey nodes denote unobserved exogenous variables.

Nested Queries. The LP mapping in Eq. (9), and by extension the lifted LP in Eq. (11), accommodates nested counterfactuals. Although nested variables do not appear directly in the response signature $\boldsymbol{S}_\sigma$, their evaluation remains a linear optimisation task. For a single nested counterfactual, it is sufficient to consider the probabilistic version of Eq. (2), i.e.,

$$P_{\mathcal{M}}(Y_{X_z} = y) = \sum_{x \in \mathcal{X}} P_{\mathcal{M}}(Y_x = y, X_z = x)\,. \tag{15}$$

In the general case, it is sufficient to perform such un-nesting on all the nested terms to obtain a linear function of standard counterfactual queries and thus a linear objective function.

Experimental Data. Beyond observational data, it is straightforward to incorporate experimental distributions obtained directly from randomised controlled trials. By Thm. 3, experimental probabilities translate directly into linear constraints on the signature probabilities in Eq. (9). To preserve the validity of the reduction in Eq. (11), the equivalence relation is refined by further requiring that equivalent signatures agree on the Boolean evaluation of the relevant interventional events. For instance, given an experimental probability $P(y_x)$, the relation in Eq. (10) is augmented with the logical condition $(\boldsymbol{s}_\sigma \models y_x) \equiv (\boldsymbol{s}'_\sigma \models y_x)$. Tracking these additional constraints partitions the signature space more finely, thereby scaling the quotient space by a factor determined by the number of interventional events considered. Furthermore, when multiple studies or heterogeneous data sources are available, one can seamlessly incorporate all evidence by appending the constraints induced by each trial in the same LP. Such a procedure natively performs automated data fusion, analogously to the approach developed by Zaffalon et al. (2023) for PSCMs.

Monotonicity and Weak Exogeneity. For ordinal variables, *monotonicity* (Angrist et al., 1996; Manski, 1997) posits a directional effect (e.g., $Y_{x_1} \geq Y_{x_0}$ if $x_1 > x_0$). *Weak exogeneity* (Tian and Pearl, 2000) requires that a variable is unaffected by unobserved confounders. Both assumptions translate expert knowledge into logical restrictions on the response signature $\boldsymbol{S}_\sigma$: monotonicity excludes states violating the inequality by fixing relevant optimisation variables to zero; exogeneity requires that causal effects are identified from data, i.e., $P(y_x) = P(y \mid x)$, functionally mirroring the embedding of experimental data.

5. Examples

Let us validate the application of our LP-based bounding scheme across a range of scenarios. We compare our bounds against analytical formulae and results from the literature.[1]

Probabilities of Causation. As a first illustrative application of our method, we derive bounds for the well-known *probability of necessity and sufficiency* (PNS, Pearl, 1999). This query refers to the counterfactual event $\gamma := (Y_{X=0} = 0) \wedge (Y_{X=1} = 1)$, which is intended to capture the causal relationship between two Boolean endogenous variables, X and Y. The only total order σ compatible with $\prec_\gamma$ is $\sigma := (X, Y)$. The associated response signature is therefore $\boldsymbol{S}_\sigma = \{X, Y_{X=0}, Y_{X=1}\}$. Accordingly, PNS bounds are obtained by solving LPs with eight variables, corresponding to the joint states of $\boldsymbol{S}_\sigma$; and four linear constraints induced by the observational distribution $\hat{P}(X, Y)$. The resulting solutions coincide with the classical closed-form bounds derived by Tian and Pearl (2000). The equivalence relation in Eq. (10) prunes the parameter space by identifying empty equivalence classes. Specifically, the candidate classes defined by the Boolean tuples $\langle (X = x, Y = y), [\![\gamma]\!] \rangle$ for $(x, y) \in \{(1, 0), (0, 1)\}$ contain no valid signatures. This is because an observational state of $(X = 1, Y = 0)$ forces $Y_{X=1} = 0$ via consistency (Galles and Pearl, 1998), directly contradicting the requirement in γ that $Y_{X=1} = 1$. We similarly proceed for $(X = 0, Y = 1)$.

The *tightness* (referred to as *sharpness* in the original work) of these bounds follows from Thm. 5: one can construct two complete canonical SCMs under which the query attains the lower and upper bounds (see also the constructive argument in the proof).

Finally, by considering the extension to conditional queries discussed in Sect. 4, analogous results can be readily obtained for the other *probabilities of causation* (PN and PS, Pearl, 1999). Moreover, unlike the analytical results of Tian and Pearl (2000), which are restricted to binary variables, our approach naturally generalises to the non-binary case. By simply adjusting the cardinality of the variables in the response signature $\boldsymbol{S}_\sigma$, we can derive tight bounds for probabilities of causation involving multi-valued treatments and outcomes (Sun et al., 2025) without requiring new algebraic derivations. From this perspective, our work is a genuine generalisation of the analysis in Tian and Pearl (2000), with a clear semantics, and allowing to bound arbitrary counterfactual queries, whether nested or not.

Average Causal Effect (ACE). The ACE measures the average change due to treatment:

$$\mathrm{ACE}_{x,x'}(y) := P(Y_x = y) - P(Y_{x'} = y)\,. \tag{16}$$

For an outcome $Y \in \{y, y'\}$ and treatment $|\mathcal{X}| = 3$, the unique order $\sigma = (X, Y)$ induces an LP with 24 signature variables and six empirical constraints in Eq.(9). The resulting bounds, $[P(x, y) + P(x', y') - 1, 1 - P(x, y') - P(x', y)]$ (Balke and Pearl, 1997), recover those of Sachs et al. (2023, Sect. 6.1) whose causal assumptions are in Fig. 1(*a*). Crucially, our method achieves this without explicitly parametrising a fully specified causal graph nor the SEs. A different account of the ACE bounding can be found in App. B.

Controlled Direct Effect (CDE). Consider a setting in which the effect of a treatment X on an outcome Y is *mediated* by a variable W. In this context, the ACE in Eq. (16) can

1. A Python implementation of our method is available at `www.github.com/Erhtric/copid`. The experimental setup employed the `COIN-OR CBC` open-source solver via the PuLP modeller (`www.coin-or.org`).

Assumption Set	$\mathrm{CDE}^{w_0}_{x_0,x_1}(y_1)$		$\mathrm{CDE}^{w_1}_{x_0,x_1}(y_1)$		$\lvert\boldsymbol{\mathcal{S}}_\sigma\rvert$	$\lvert\boldsymbol{\mathcal{T}}_\sigma\rvert$
	Lower	Upper	Lower	Upper		
No additional assumptions						
Cai et al. (2008) (Fig. 1(*c*))	-0.200	0.385	-0.781	0.634	–	–
This paper	-0.599	0.694	-0.891	0.817	128	20
Monotonicity						
Cai et al. (2008)	0.000	0.036	0.000	0.582	–	–
This paper	0.000	0.519	0.000	0.791	36	13
Weak exogeneity						
This paper	-0.200	0.385	-0.782	0.633	128	64
Weak exogeneity + monotonicity						
This paper	0.000	0.035	0.000	0.580	36	24

Table 1: Comparison of CDE bounds on the (binary) Lipid data (Cai et al., 2008, Tab. 1) against analytical bounds under varying assumptions. In this experiment, *Monotonicity* forces positive effects of X on W, X on Y, and W on Y; whilst, *weak exogeneity* imposes no-interaction between X and W on Y.

be refined into a CDE to measure the change due to treatment at different mediator *strata*:

$$\mathrm{CDE}^{w}_{x_0,x_1}(y) \coloneqq P(Y_{x_1,w} = y) - P(Y_{x_0,w} = y)\,. \tag{17}$$

The query induces a partial order, namely $\{X, W\} \prec Y$. By Thm. 8, we can choose an arbitrary order among its set of linear extensions, together with an empirical distribution $\hat{P}(X, W, Y)$, to solve Eq. (11). We validate the method against the problem proposed by Cai et al. (2008) of bounding Eq. (17) under the assumptions encoded in Fig. 1(*c*). In such a setting, we first consider all binary variables and a randomised treatment (i.e., no confounders between X and Y), examining data from the *Coronary Primary Prevention Trial* (Cai et al., 2008).

A summary of the computed bounds is reported in Tab. 1. One can observe that when no assumptions are made, or when solely monotonicity is assumed, the minimal structure inherently implied by the query yields valid, but wide, intervals. Notably, the tight bounds reported by Cai et al. (2008) rely on a fully specified causal graph. To mirror their structural assumptions within our framework, one must incorporate a weak exogeneity constraint—also referred to as the *no-interaction* assumption—into the signature space $\boldsymbol{\mathcal{S}}_\sigma$. Specifically, enforcing the independence $\{Y, W\} \perp\!\!\!\perp X$ recovers Cai et al. (2008) analytical bounds. As a special instance, one can also require $Y \perp\!\!\!\perp \{X, W\}$ to collapse the bound to a point estimate, since $P(y_{x,w}) = P(y \mid x, w)$. As expected, introducing these independence constraints expands the number of optimisation parameters $\lvert\boldsymbol{\mathcal{T}}_\sigma\rvert$, whilst introducing monotonicity constraints shrinks the space (cf. Tab. 1). A similar discussion holds for categorical variables, where the LP reduction can be systematically constructed and compared against the equations presented in Cai et al. (2008).

Natural Direct Effect (NDE). The CDE in Eq. (17) refers to a fixed state of the mediator. By contrast, the NDE evaluates the effect of a treatment when the mediator is

	Lower	Upper	$\lvert\boldsymbol{\mathcal{S}}_\sigma\rvert$	$\lvert\boldsymbol{\mathcal{T}}_\sigma\rvert$
Assumptions for $\mathrm{NDE}_{x_0,x_1}(y_1)$				
Query-induced Order	-0.691	0.688	294,912	60
$Y \perp\!\!\!\perp X$	-0.503	0.497	294,912	96
$\{Y,W\} \perp\!\!\!\perp X$	-0.503	0.497	294,912	528
$Y \perp\!\!\!\perp \{X,W\}$	-0.628	0.679	294,912	89,172
$Y \perp\!\!\!\perp \{X,W\}$ & $W \perp\!\!\!\perp X$	-0.489	0.637	294,912	294,912
Query				
$P(Y_{x_1,W=w_0}=y_1, W_{x_0}=w_0)$	0.000	0.610	294,912	37
$P(Y_{x_1,W=w_1}=y_1, W_{x_0}=w_1)$	0.000	0.493	294,912	37
$P(Y_{x_1,W=w_2}=y_1, W_{x_0}=w_2)$	0.000	0.365	294,912	37
$P(Y_{x_1,W=w_3}=y_1, W_{x_0}=w_3)$	0.000	0.415	294,912	37
$P(Y_{x_1,W=w_4}=y_1, W_{x_0}=w_4)$	0.000	0.357	294,912	37
$P(Y_{x_1,W=w_5}=y_1, W_{x_0}=w_5)$	0.000	0.391	294,912	37
$P(Y_{x_0}=y_1)$	0.312	0.691	8	6

Table 2: UC Berkeley dataset: NDE bounds under various weak exogeneity assumptions (top) and individual unnested query terms under query-induced assumptions (bottom).

set to the value it would have attained in the absence of treatment. This leads to the nested counterfactual query:

$$\mathrm{NDE}_{x_0,x_1}(y_1) \coloneqq P(Y_{x_1,W_{x_0}}=y_1) - P(Y_{x_0}=y_1)\,. \tag{18}$$

As a matter of fact, the nested syntactical structure of the query induces a unique complete order $X \prec W \prec Y$. We proceed to apply the un-nesting procedure in Eq. (15), which, in turn, allows one to rewrite nested counterfactuals like in Eq. (18) as sums of standard (non-nested) counterfactual queries. This transformation preserves the linear structure of the objective function in our LP formulation and one can consider the PSCM whose endogenous structure is like in Fig. 1(*c*) but with one exogenous U parent of all variables. As discussed previously, by Thm. 6, the optimisation space's size is bounded by $O(|\mathcal{X}|\cdot|\mathcal{W}|\cdot|\mathcal{Y}|)$ (cf. Tab. 2), making our approach computationally feasible for modern LP solvers. To the best of our knowledge, no general bounding techniques are currently available for queries of this kind without imposing additional structural assumptions about the underlying causal graph.

Berkeley Admissions Dataset. For a numerical illustration, we consider the popular *UC Berkeley admissions* dataset (Bickel et al., 1975). In this dataset, the binary variables X and Y represent *gender* (with x_1 indicating *female*) and *admission outcome* (with y_1 denoting *acceptance*). The mediator W is a categorical variable with six states corresponding to the department to which the applicant applied. We omit monotonicity assumptions, as deterministic effects are tricky to justify in social contexts.

Tab. 2 summarises the total NDE bounds under varying weak exogeneity assumptions, alongside the decomposed counterfactual terms from Eq. (15). In the context of university admissions, enforcing $Y \perp\!\!\!\perp X$ assumes no confounding jointly affecting gender and admission. The stricter condition $\{Y,W\} \perp\!\!\!\perp X$ treats gender as if it were completely randomised, assuming department choice does not depend on gender. Conversely, $Y \perp\!\!\!\perp \{X,W\}$ dictates

that the admission committee's decision is entirely independent of unobserved background factors influencing either the applicant's gender or department preference. It is worth noting that when all possible exogeneity assumptions are employed, the equivalence reduction from Sect. 2 fails to compress the optimisation space, as enforcing mutual independence across all response mechanisms requires tracking each signature state individually. Whilst these assumptions narrow the bounds, at the cost of expanding the optimisation space $|\mathcal{T}_\sigma|$, they risk misrepresenting complex dynamics.

Unlike bespoke analytical methods that require graphical constraints to bound nested counterfactuals, our LP formulation automatically evaluates the NDE using only the minimal assumptions inherent to the query itself. For reference, under the *standard fairness model* (Plečko and Bareinboim, 2024), the NDE evaluates to a point estimate of 0.043. Concluding a lack of discriminative policies based on this point estimate might be completely inaccurate, as the stark contrast with our wider, assumption-free intervals demonstrates how such conclusions rely heavily on rigid structural assumptions. Furthermore, it is interesting to notice that directly bounding the full NDE yields substantially tighter intervals than combining separately bounded decomposed components via interval arithmetic.

6. Conclusions

We presented a linear programming framework for partially identifying arbitrary, including nested, counterfactual queries, relying exclusively on structural assumptions inherent to the query itself. In doing so, we successfully extended the classical bounding results of Tian and Pearl (2000) for probabilities of causation. We proved that partial identification under these minimal assumptions yields tight intervals, and established their invariance with respect to the marginalisation of the variables not appearing in the query. Rather than requiring an input causal graph, an SCM attaining the bounds emerges constructively as a by-product. Furthermore, exploiting query symmetries drastically reduces the optimisation space, rendering the LP tractable. We acknowledge an inherent trade-off in the proposed method: reducing the reliance on prior domain knowledge, and explicit conditional independencies, produces intervals that are naturally wider than those derived from a fully specified causal graph. Nevertheless, as demonstrated in our case studies, this framework serves as a general and computationally efficient method to bound arbitrary inquiries under minimal assumptions. Future work could investigate extending this framework to cyclic causal structures, examining the impact of different valid causal orders, and investigate finer quotient-space reductions to further enhance scalability.

Acknowledgements

The research of Eric Rossetto was funded by Swiss Post AG as part of a Doctoral Studies Grant on *AI and Robustness*. The authors thank the anonymous reviewers for their constructive comments, and in particular the first reviewer for insightful remarks on the underlying assumptions of our approach and its possible extension to cyclic models.

References

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434): 444–455, 1996.

Alexander Balke and Judea Pearl. Counterfactual probabilities: Computational methods, bounds and applications. In *Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence*, pages 46–54, 1994a.

Alexander Balke and Judea Pearl. Probabilistic evaluation of counterfactual queries. In *Probabilistic and Causal Inference*, volume Proceedings of the Twelfth National Conference on Artificial Intelligence of *AAAI '94*, pages 230–237. American Association for Artificial Intelligence, 1994b.

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.

Peter Bickel, Eugene Hammel, and William O'Connell. Sex bias in graduate admissions: data from Berkeley. *Science*, 187(4175):398–404, 1975.

Anna Rodum Bjøru, Rafael Cabañas, Helge Langseth, and Antonio Salmerón. Divide and conquer for causal computation. *International Journal of Approximate Reasoning*, 186: 109520, 2025.

Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M Mooij. Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5):2885–2915, 2021.

Zhihong Cai, Manabu Kuroki, Judea Pearl, and Jin Tian. Bounds on direct effects in the presence of confounded intermediate variables. *Biometrics*, 64(3):695–701, 2008.

Abraham Charnes and William W. Cooper. Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 9(3-4):181–186, 1962.

Juan D. Correa, Sanghack Lee, and Elias Bareinboim. Nested counterfactual identification from arbitrary surrogate experiments. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2021.

Fabio G. Cozman, Radu Marinescu, Junkyu Lee, Alexander Gray, and Denis D. Mauá. Dealing with cycles in graph-based probabilistic models: the case of logical credal networks. In *Proceedings of the Fourteenth International Symposium on Imprecise Probabilities:*

Theories and Applications, volume 290 of *Proceedings of Machine Learning Research*, pages 93–102. PMLR, 15–18 Jul 2025.

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.

Robin J. Evans. Graphs for margins of Bayesian networks. *Scandinavian Journal of Statistics*, 43(3):625–648, 2016.

David Galles and Judea Pearl. An Axiomatic Characterization of Causal Counterfactuals. *Foundations of Science*, 3(1):151–182, 1998.

Joseph Y. Halpern. Axiomatizing Causal Reasoning. *Journal of Artificial Intelligence Research*, 12:317–337, 2000.

Yuta Kawakami, Manabu Kuroki, and Jin Tian. Probabilities of causation for continuous and vector variables. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pages 1901–1921. PMLR, 15–19 Jul 2024.

Kristian Kersting, Martin Mladenov, and Pavel Tokmakov. Relational linear programming. *Artificial Intelligence*, 244:188–216, 2017.

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

Charles F. Manski. Monotone Treatment Response. *Econometrica*, 65(6):1311, 1997.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

Judea Pearl. Probabilities of causation: Three counterfactual interpretations and their identification. *Synthese*, 121(1):93–149, 1999.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

Drago Plečko and Elias Bareinboim. Causal Fairness Analysis: A Causal Toolkit for Fair Machine Learning. *Foundations and Trends® in Machine Learning*, 17(3):304–589, 2024.

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.

Ilya Shpitser and Judea Pearl. What Counterfactuals Can Be Tested. 2007.

Hanmei Sun, Chengfeng Shi, and Qiang Zhao. Bounding the probability of causation under ordinal outcomes. *Communications in Statistics-Theory and Methods*, 54(24):8121–8132, 2025.

Jin Tian. *Studies in causal reasoning and learning*. PhD thesis, University of California, Los Angeles, 2002.

Jin Tian and Judea Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1-4):287–313, 2000.

Marco Zaffalon, Alessandro Antonucci, Rafael Cabañas, and David Huber. Approximating counterfactual bounds while fusing observational, biased and randomised data sources. *International Journal of Approximate Reasoning*, 162:109023, 2023.

Marco Zaffalon, Alessandro Antonucci, Rafael Cabañas, David Huber, and Dario Azzimonti. Efficient computation of counterfactual bounds. *International Journal of Approximate Reasoning*, 171:109111, 2024.

Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 26548–26558, 2022.

Appendix A. Proofs

Proof of Thm. 2 Let us separately prove the non-strict inequalities between the two sides of Eq. (5). By definition, M_σ is based on a causal graph compatible with σ. This implies $\mathbb{M}_{M_\sigma} \supseteq \mathbb{M}_\sigma$ and hence the $\leq$ inequality.

In order to prove the opposite inequality, let $\mathcal{M}^*$ be an SCM that solves Eq. (4). By definition of $\mathbb{M}_\sigma$, the causal graph of $\mathcal{M}^*$ is consistent with the total order σ. Therefore, for each $V^{(i)} \in \boldsymbol{V}$, the endogenous parents of $V^{(i)}$ in $\mathcal{M}^*$ form a subset of its predecessors $\boldsymbol{V}_{\prec i}$ under σ. We can enlarge each SE of $V^{(i)}$ by including as inputs all variables in $\boldsymbol{V}_{\prec i}$ that are not already endogenous parents of $V^{(i)}$, and by similarly incorporating all exogenous variables not already among its exogenous parents. In this way, the collection $\boldsymbol{U}$ of exogenous variables can be treated as a single aggregated variable U. Consequently, $\mathcal{M}^*$ may be regarded as an SCM that is consistent with the same causal graph as the PSCM M_σ. Since M_σ is canonical, there exists an equivalent SCM in $\mathbb{M}_{M_\sigma}$. This implies the $\geq$ inequality. ■

Lemma 9 *For any atomic counterfactual proposition $V_{\boldsymbol{x}} = v$ and $\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma$, the truth value of the entailment $\boldsymbol{s}_\sigma \models (V_{\boldsymbol{x}} = v)$ is determined. Furthermore, if γ is a conjunction of such propositions, the truth value of $\boldsymbol{s}_\sigma \models \gamma$ is also determined.*

Proof Consider an arbitrary intervention $\boldsymbol{x}$. We proceed by induction on the topological order σ to show that the potential response of every variable $V^{(i)}$ is uniquely determined by $\boldsymbol{s}_\sigma$ and $\boldsymbol{x}$.

If $i = 1$, the variable $V^{(1)}$ has no endogenous parents. If $V^{(1)} \in \boldsymbol{X}$, its value is fixed to the assignment specified in $\boldsymbol{x}$. If $V^{(1)} \notin \boldsymbol{X}$, its value is uniquely determined by the realization encoded in $\boldsymbol{s}_\sigma$.

Assume now that for some $i > 1$, the interventional values of all preceding variables $\boldsymbol{V}_{\prec i}$ are fixed to $(\boldsymbol{v}_{\prec i})_{\boldsymbol{x}}$. We determine the value of $V^{(i)}$ as follows:

1. If $V^{(i)} \in \boldsymbol{X}$, its value is determined by the assignment in $\boldsymbol{x}$.
2. If $V^{(i)} \notin \boldsymbol{X}$, its value is uniquely determined by $\boldsymbol{s}_\sigma$ given the values of its parents, which are fixed to $(\boldsymbol{v}_{\prec i})_{\boldsymbol{x}}$ by the induction hypothesis.

Since $|\boldsymbol{V}| < \infty$, every counterfactual variable $V_{\boldsymbol{x}}$ evaluates to a unique constant under $\boldsymbol{s}_\sigma$. Consequently, the atomic entailment is well-defined. By extension, the entailment for a conjunction $\gamma = \bigwedge_{j=1}^{k}(Y^{(j)}_{\boldsymbol{x}^{(j)}} = y^{(j)})$ is determined by:

$$(\boldsymbol{s}_\sigma \models \gamma) \iff \bigwedge_{j=1}^{k} \left(\boldsymbol{s}_\sigma \models Y^{(j)}_{\boldsymbol{x}^{(j)}} = y^{(j)}\right). \tag{19}$$

■

Lemma 10 *For any arbitrary SCM $\mathcal{M} \in \mathbb{M}_{M_\sigma}$, there exists a bijection $h : \mathcal{U} \to \boldsymbol{\mathcal{S}}_\sigma$ between its exogenous domain $\mathcal{U}$ and the response signature domain $\boldsymbol{\mathcal{S}}_\sigma$.*

Proof Let $h : \mathcal{U} \to \boldsymbol{\mathcal{S}}_\sigma$ map each exogenous state u to its induced signature $\boldsymbol{s}_\sigma$. Because SEs are deterministic, each u induces exactly one signature, making h a well-defined function. We establish that h is a *bijection*.

By definition, $\boldsymbol{\mathcal{S}}_\sigma$ comprises all valid configurations of potential outcomes which M_σ generates. Because $\mathcal{M} \in \mathbb{M}_{M_\sigma}$, it follows a canonical specification (Zhang et al., 2022, Def. 2.3); its exogenous domain $\mathcal{U}$ explicitly indexes the Cartesian product of every possible deterministic function mapping from endogenous parents to children. Thus, for any arbitrary signature $\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma$, there is at least one state $u \in \mathcal{U}$ that generates it. We conclude that h is surjective.

By the same canonical specification of $\mathcal{M}$, the cardinality of $\mathcal{U}$ is exactly equal to the cardinality of the Cartesian product of all possible functional mappings—which coincides with $|\boldsymbol{\mathcal{S}}_\sigma|$ by construction in Eq. (6). Because h is a surjective function mapping between two finite sets of identical cardinality, it must necessarily be injective. Therefore, every state u corresponds uniquely to a signature $\boldsymbol{s}_\sigma$. ■

Proof of Thm. 3 By Thm. 2, without loss of generality, we can restrict our analysis to an arbitrary SCM $\mathcal{M} \in \mathbb{M}_{M_\sigma}$.

By Lem. 10, there exists a bijection h mapping each exogenous state $u \in \mathcal{U}$ to a unique response signature $\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma$. Consequently, the logical equivalence $(u \models \gamma) \iff (\boldsymbol{s}_\sigma \models \gamma)$ holds true for $\boldsymbol{s}_\sigma = h(u)$. The left-hand side is well-defined by the soundness axiomatisation of SCMs (Halpern, 2000), and the right-hand side is determined by Lem. 9.

Because h is bijective, the probability measure over $\mathcal{U}$ directly translates to $\boldsymbol{\mathcal{S}}_\sigma$, yielding $P(U = u) = P(\boldsymbol{S}_\sigma = \boldsymbol{s}_\sigma)$. Substituting u with $\boldsymbol{s}_\sigma$ in Eq. (1), we obtain:

$$P_{\mathcal{M}}(\gamma) = \sum_{\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma} [\![\boldsymbol{s}_\sigma \models \gamma]\!] \cdot P(\boldsymbol{S}_\sigma = \boldsymbol{s}_\sigma).$$

■

Proof of Thm. 5 Let L^* be the minimum of the optimisation problem in Eq. (4), and let L be the minimum of the LP task defined in Eq. (9). To prove the theorem, we first show that $L \leq L^*$, and then $L \geq L^*$.

Assume the optimal value L^* is achieved by some SCM $\mathcal{M}^* \in \mathbb{M}_\sigma$ for query $P_{\mathcal{M}^*}(\gamma)$. By Thm. 2, we can consider $\mathcal{M}^*$ to be an SCM parametrising the canonical PSCM M_σ. By construction of $\boldsymbol{\mathcal{S}}_\sigma$, the exogenous distribution of $\mathcal{M}^*$ maps to a valid joint probability distribution $\mathbf{q}^*$ over the states $\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma$. By Thm. 3, the objective query can be expressed as a linear combination of the probabilities associated with such states. Furthermore, since $\mathcal{M}^*$ is a feasible solution to Eq. (4), it must satisfy the observational constraint $P_{\mathcal{M}^*}(\boldsymbol{V}) = \hat{P}(\boldsymbol{V})$. Correspondingly, $\mathbf{q}^*$ satisfies the LP's constraint imposed by observational data as by Cor. 4. Consequently, $\mathbf{q}^*$ constitutes a feasible solution within the constraints of the LP formulation in Eq. (9). Because the LP identifies the global minimum L over all feasible distributions, it must be that $L \leq L^*$.

Conversely, let $\mathbf{q}^*$ be the optimal solution that minimises the LP task in Eq. (9), such that the objective evaluates to L. Consider a specific SCM $\mathcal{M} \in \mathbb{M}_\sigma$; our argument will equivalently consider $\mathcal{M} \in \mathbb{M}_{M_\sigma}$ (after Thm. 2). We define its exogenous distribution by assigning $P_{\mathcal{M}}(u) = q^*_{h(u)}$, where h is the bijection established in Lem. 10. Additionally, because the LP enforces the observational constraints, $\mathcal{M}$ matches the empirical distribution, then it is also a feasible candidate for the optimisation task in Eq. (4). Since L^* represents the minimum over *all* feasible SCMs in $\mathbb{M}_\sigma$, and we have found a specific feasible SCM $\mathcal{M}$ that achieves L, it must hold that $L^* \leq L$.

Since $L \leq L^*$ and $L^* \leq L$, we conclude that $L = L^*$. Furthermore, the explicit construction in the second half of the proof demonstrates that there exists an SCM $\mathcal{M} \in \mathbb{M}_{M_\sigma}$ that precisely attains the LP bounds. ■

Proof of Thm. 6 Let Q_{LP} and Q_{LLP} denote the feasible regions of the LP in Eq. (9) and of the lifted LP in Eq. (11), respectively.

For any feasible solution $\boldsymbol{q}^* \in Q_{LP}$, by definition of the equivalence relation in Eq. (10), $q_{\boldsymbol{t}_\sigma} = \sum_{\boldsymbol{s}_\sigma \in \boldsymbol{t}_\sigma} q^*_{\boldsymbol{s}_\sigma}$ trivially satisfies the constraints of Eq. (11) whilst preserving the objective value, as the Boolean evaluations are constant within a class $\boldsymbol{t}_\sigma \in \boldsymbol{\mathcal{T}}_\sigma$.

Conversely, consider any feasible solution $\boldsymbol{q}^*_{\boldsymbol{t}_\sigma} \in Q_{LLP}$. Let $k_{\boldsymbol{t}_\sigma} \coloneqq |\boldsymbol{t}_\sigma|$ be the number of signatures $\boldsymbol{s}_\sigma$ contained within a given equivalence class $\boldsymbol{t}_\sigma$. Note that by construction every class $\boldsymbol{t}_\sigma$ is non-empty, hence $k_{\boldsymbol{t}_\sigma} > 0$. We construct a valid distribution in the original space by assigning a uniform mass $q_{\boldsymbol{s}_\sigma} = q^*_{\boldsymbol{t}_\sigma}/k_{\boldsymbol{t}_\sigma}$ to every signature $\boldsymbol{s}_\sigma \in \boldsymbol{t}_\sigma$, for all $\boldsymbol{t}_\sigma \in \boldsymbol{\mathcal{T}}_\sigma$. Because the logical constraints evaluate identically across $\boldsymbol{t}_\sigma$, summing over this uniformly distributed mass recovers the constraints and the objective value of the lifted LP, satisfying Eq. (9).

Since these mappings strictly preserve both feasibility and the objective value in both directions, the extrema evaluated over Q_{LLP} must exactly match those over Q_{LP}. Following the constructive argument detailed in the proof of Thm. 5, there exists a witness SCM that attains these bounds, ensuring the tightness is inherited by the lifted LP in Eq. (11). ■

Proof of Thm. 7 Let $\mathcal{M}^* \in \mathbb{M}_\sigma$ and $\tilde{\mathcal{M}}^* \in \mathbb{M}_\sigma^{\boldsymbol{V}_\gamma}$ denote the SCMs giving the optima in, respectively, Eq. (4) and Eq. (12). Let also $L \coloneqq P_{\mathcal{M}^*}(\gamma)$ and $L' \coloneqq P_{\tilde{\mathcal{M}}^*}(\gamma)$. Following, for instance, Bongers et al. (2021), we can marginalise out from $\mathcal{M}^*$ the variables in $\boldsymbol{V} \setminus \boldsymbol{V}_\gamma$ and

obtain a well-defined SCM over $\boldsymbol{V}_\gamma$. Moreover, by construction, we have $P^*_{\mathcal{M}}(\boldsymbol{V}) = \hat{P}(\boldsymbol{V})$, and this implies, in the marginal SCM, $P^*_{\mathcal{M}}(\boldsymbol{V}_\gamma) = \hat{P}(\boldsymbol{V}_\gamma)$. Overall, we have an SCM in $\mathbb{M}_\sigma^{\boldsymbol{V}_\gamma}$ giving L, and this proves $L \geq L'$.

To prove the inverse inequality, let $P(\boldsymbol{V}' \mid \boldsymbol{V}_\gamma) := \hat{P}(\boldsymbol{V})/\hat{P}(\boldsymbol{V}_\gamma)$, where $\boldsymbol{V}' := \boldsymbol{V} \setminus \boldsymbol{V}_\gamma$. We augment the SCM $\tilde{\mathcal{M}}^*$ with the variables in $\boldsymbol{V}'$. By setting all the variables in $\boldsymbol{V}_\gamma$ as endogenous parents of those in $\boldsymbol{V}'$, we can easily obtain $P(\boldsymbol{V}) = \hat{P}(\boldsymbol{V})$. Thus we have an SCM in $\mathbb{M}_\sigma$ such that $P(\boldsymbol{V}) = \hat{P}(\boldsymbol{V})$ and giving L'. This proves $L \leq L'$, and hence the thesis. ■

Proof of Thm. 8 Let $\sigma \in \Sigma(\prec_\gamma)$ be an arbitrary total order compatible with $\prec_\gamma$, and consider the lifted LP in Eq. (11) constructed over σ. Each equivalence class $\boldsymbol{t}_\sigma \in \boldsymbol{\mathcal{T}}_\sigma$ is uniquely characterised by an observational assignment $\boldsymbol{v} \in \boldsymbol{\mathcal{V}}$ and a Boolean query evaluation $b = [\![\boldsymbol{t}_\sigma \models \gamma]\!] \in \{0, 1\}$.

An equivalence class indexed by $\langle \boldsymbol{v}, b \rangle$ is non-empty (i.e., contains at least one signature $\boldsymbol{s} \in \boldsymbol{\mathcal{S}}_\sigma$) if and only if $\boldsymbol{v}$ and b do not contradict the consistency axiom (Galles and Pearl, 1998). Specifically, for any atomic proposition $Y_x = y^*$ in γ, if $\boldsymbol{v}$ assigns $X = x$, consistency forces $Y_x = v_Y$. Hence, if $v_Y \neq y^*$, consistency prevents γ from being satisfied, ruling out $b = 1$; conversely, if $\boldsymbol{v}$ forces every atomic term in γ to hold, consistency rules out $b = 0$.

Crucially, whether a pair $\langle \boldsymbol{v}, b \rangle$ satisfies the consistency axiom depends solely on the assignment $\boldsymbol{v}$ and the syntax of γ, independently of the topological order σ. Moreover, because the canonical specification contains all possible functional responses, every pair $\langle \boldsymbol{v}, b \rangle$ that is compatible with consistency is realised by at least one signature $\boldsymbol{s} \in \boldsymbol{\mathcal{S}}_\sigma$: the observational response is fixed to $\boldsymbol{v}$, while the counterfactual responses under unobserved interventions can be set to satisfy γ (yielding $b = 1$) or violate at least one atom of γ (yielding $b = 0$).

We can conclude that the quotient space $\boldsymbol{\mathcal{T}}_\sigma$ of non-empty equivalence classes is invariant to σ. Indexing the lifted optimisation variables as $q_{\langle \boldsymbol{v}, b \rangle}$, the LP objective function is $\sum_{\boldsymbol{v}} q_{\langle \boldsymbol{v}, 1 \rangle}$ and the empirical constraints are $q_{\langle \boldsymbol{v}, 0 \rangle} + q_{\langle \boldsymbol{v}, 1 \rangle} = \hat{P}(\boldsymbol{v})$ for each $\boldsymbol{v} \in \boldsymbol{\mathcal{V}}$. Because the optimisation variables, constraints, and objective function are entirely decoupled from σ, the resulting LP is identical for all linear extensions $\sigma \in \Sigma(\prec_\gamma)$. Solving the LP for any arbitrary compatible σ therefore yields the same bounds as Eq. (12) and solves the global optimisation task in Eq. (13). ■

Appendix B. Additional Experiments

ACE with Imperfect Compliance. We consider a model with three Boolean variables. In addition to the outcome Y, we distinguish between treatment exposure X and treatment assignment Z. To evaluate the causal effect of the treatment when assigned, we focus on the query $\mathrm{ACE}_{X=0, X=1}(Y = 1)$, defined as in Eq. (16). Observational information is available only in the form of a *conditional* empirical distribution $\hat{P}(X, Y|Z)$. In our framework, consistency with the empirical distribution is expressed as:

$$\frac{\sum_{\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma} [\![\boldsymbol{s}_\sigma \models (x, y, z)]\!] \cdot q_{\boldsymbol{s}_\sigma}}{\sum_{\boldsymbol{s}_\sigma \in \boldsymbol{\mathcal{S}}_\sigma} [\![\boldsymbol{s}_\sigma \models z]\!] \cdot q_{\boldsymbol{s}_\sigma}} = \hat{P}(x, y|z)\,, \tag{20}$$

Method	Lower	Upper
Vitamin A		
Balke and Pearl (1997)	-0.195	0.005
This paper	-0.587	0.412
Coronary		
Balke and Pearl (1997)	0.262	0.868
This paper	-0.685	0.971

Table 3: Comparison of ACE bounds. Lower and upper bounds for the ACE grouped by Study.

which yields linear constraints over the optimisation variables $q_{\boldsymbol{s}_\sigma}$, thus preserving the LP structure of our method. Although the query involves only X and Y, the conditional nature of the empirical distribution prevents a marginalisation of Z. Consequently, the response signature must be based on all three Boolean variables. By Thm. 8 we choose an arbitrary admissible order σ such that $X \prec Y$.

We compare our bounds with those obtained from the LP formulation introduced by Balke and Pearl (1997) and later generalised by Sachs et al. (2023). That approach requires an explicit causal graph. In particular, the model in Fig. 1(*b*) assumes that Z affects Y only through X, and that Z is independent of an unobserved confounder U affecting both X and Y. Tab. 3 reports the resulting comparison for two observational studies: the *vitamin A* supplementation study in northern Sumatra and the *coronary* primary prevention trial. Our bounds remain informative, but are substantially wider than those of Balke and Pearl (1997). This difference reflects the fact that our framework relies only on the order restrictions induced by the query, whereas the graphical model in Fig. 1(*b*) imposes additional structure, most notably the exclusion of a direct effect from Z to Y. To model these effects with such a granularity one would either impose additional (non-linear) constraints or eventually require a causal graph, but this lies outside manuscript's scope. By Thm. 5, the more extreme values admitted by our procedure are attainable under SCMs compatible with our weaker assumptions, including models in which Z may directly affect Y.

As pointed out by Balke and Pearl (1994a), bounds can be further tightened if additional assumptions are made about subject behaviours: imposing a monotonic constraint, e.g. a positive response on treatment, $Y_{x_1,z} \geq Y_{x_0,z}$ for every $z \in \mathcal{Z}$. As a consequence, negative lower bounds on the effects are raised to zero. Nevertheless, even when combined with additional assumptions such as weak exogeneity, these restrictions are still insufficient to recover the analytical bounds implied by the graphical model, as discussed in Sect. 5.

PNS Generalisation for Non-Binary Variables To assess the practical scalability of the lifted LP in Eq. (11), we measured the end-to-end time required to construct the signature equivalence classes, aggregate the lifted variables, and solve the resulting LP. The benchmark query is a multi-hypothetical generalisation of PNS for ordinal treatment and outcome variables, in the spirit of Kawakami et al. (2024). Let X and Y be ordinal treatment and effect variables, respectively, defined over finite domains $\mathcal{X} = \{x^{(0)}, \ldots, x^{(n-1)}\}$ and $\mathcal{Y} = \{y^{(0)}, \ldots, y^{(n-1)}\}$ where n is an experimental parameter. The GPNS of order r (where

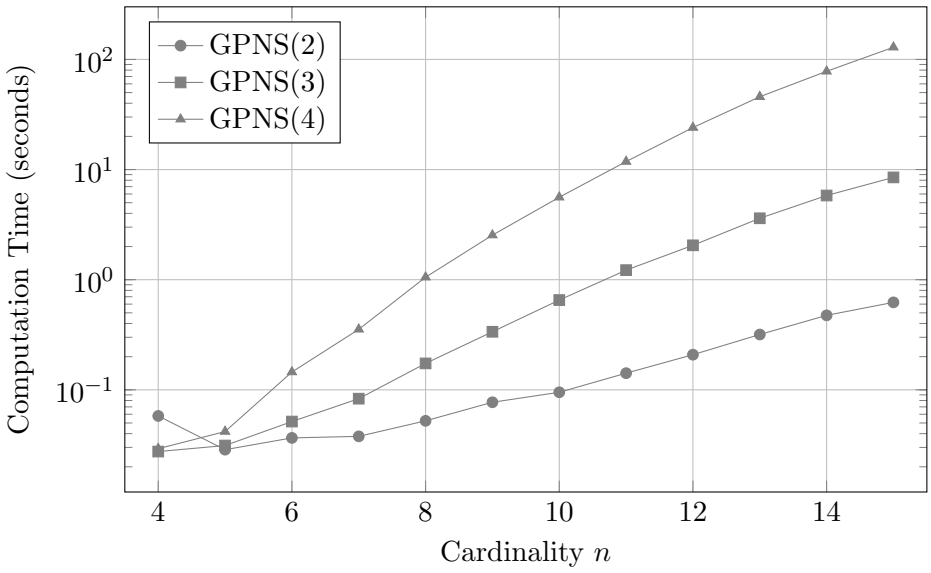


Figure 2: End-to-end computation time for GPNS(r) bounds, on a semi-logarithmic scale, as a function of the variable cardinalities ($n = |\mathcal{X}| = |\mathcal{Y}|$).

$2 \leq r \leq n$) evaluates the joint probability of strict compliance across r distinct interventions, defined as:

$$\text{GPNS}(r) \coloneqq P\left(\bigwedge_{i=0}^{r-1} Y_{x^{(i)}} = y^{(i)}\right). \tag{21}$$

For binary X and Y, GPNS(2) reduces to the standard PNS event. For each $n \in \{4, \ldots, 15\}$, we randomly parametrise a Bayesian network with the endogenous structure in Fig. 1(*a*), estimated $\hat{P}(X, Y)$ from simulated observational data, and computed bounds for GPNS(r) with $r \in \{2, 3, 4\}$. Figure 2 reports the resulting end-to-end computation times.

For this benchmark, $\boldsymbol{V}_\gamma = \{X, Y\}$, so the lifted LP has at most $|\mathcal{X}||\mathcal{Y}| = n^2$ variables, as derived in Sect. 3. Results are consistent with the polynomial scaling of the lifted formulation, although we must consider the overhead introduced by the symmetry detection, LP construction, and numerical optimisation. As expected, solving the LP remains tractable for standard solvers as the domain expands.

Appendix C. Canonical Specifications

We expand the discussion of *canonicalisation* of SCMs and PSCMs briefly introduced in Sect. 3. We recall here that the canonical PSCM used in the proof of Thm. 2 builds upon the theoretical contributions of Zhang et al. (2022) and Balke and Pearl (1994a). Generally speaking, an SCM takes its structural equations (SEs) as given. When these are unknown and the endogenous variables take on finitely many values, the literature shows that one may still specify a PSCM from the causal graph alone by adopting a *canonical* representation.

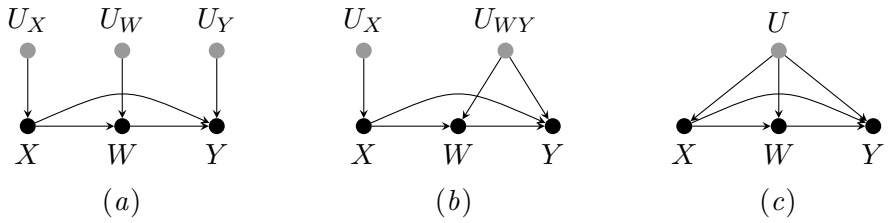


Figure 3: Examples of canonical PSCM specifications compatible with the endogenous order $\sigma \coloneqq (X, W, Y)$. (a) Markovian specification. (b) Specification with one exogenous variable shared by W and Y. (c) Specification with one exogenous variable shared by all endogenous variables. Grey nodes denote exogenous variables.

For an endogenous variable $V \in \boldsymbol{V}$ with parents Pa_V, a state of its exogenous parent U_V acts as a selector for a deterministic mapping (or *response function*) from the domain of Pa_V to V. There are exactly $|\mathcal{V}|^{|\mathrm{Pa}_V|}$ such possible functions, where $|\mathrm{Pa}_V| \coloneqq \prod_{Z \in \mathrm{Pa}_V} |\mathcal{Z}|$. In the absence of unobserved confounding, as in Markovian models, each U_V is independent of all other exogenous variables and can be replaced without loss of generality by a discrete canonical variable of finite cardinality $|\mathcal{U}_V| = |\mathcal{V}|^{|\mathrm{Pa}_V|}$. When latent confounding is present, as in semi-Markovian models, unobserved variables may jointly affect multiple endogenous variables. The graph is then partitioned into *confounded components* (or *c-components*) (Tian, 2002), defined as the maximal sets of endogenous variables connected by paths of latent confounders (see *confounded paths* in Shpitser and Pearl, 2007). In a canonical specification, all latent confounding within each c-component $\boldsymbol{C} \subseteq \boldsymbol{V}$ is replaced without loss of generality by a single discrete exogenous variable $U_{\boldsymbol{C}}$ that jointly selects the response functions for every variable in $\boldsymbol{C}$, whilst the exogenous variables associated with distinct c-components remain mutually independent. Because the state space of $U_{\boldsymbol{C}}$ must encompass all joint assignments of response functions, its canonical space $\mathcal{U}_{\boldsymbol{C}}$ indexes all possible combinations of individual mappings, yielding $|\mathcal{U}_{\boldsymbol{C}}| = \prod_{V \in \boldsymbol{C}} |\mathcal{V}|^{|\mathrm{Pa}_V|}$. Canonicalisation guarantees that any discrete SCM (or PSCM) admits a discrete equivalent inducing an identical set of observational and counterfactual distributions (Evans, 2016; Duarte et al., 2024).

To ground the ideas, we proceed by means of an example. Consider $\boldsymbol{V} \coloneqq \{X, W, Y\}$ with binary domains $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and ternary domain $\mathcal{W} = \{0, 1, 2\}$. We assume the causal order $X \prec W \prec Y$. The three panels in Fig. 3 share the same endogenous structure compatible with this order. A canonical specification of these causal diagrams is obtained by enumerating the deterministic response functions for each endogenous variable. First, consider the unconfounded setting in Fig. 3(*a*), where every variable forms a singleton c-component. Since $\mathrm{Pa}_X = \emptyset$, there are two constant mappings, giving $|\mathcal{U}_X| = 2$. For W, whose parent is X, the mechanism $w \leftarrow f_W(x, u_W)$ assigns a state in $\mathcal{W}$ to each $x \in \mathcal{X}$; hence, $|\mathcal{U}_W| = |\mathcal{W}|^{|\mathcal{X}|} = 3^2 = 9$. Finally, the mapping $y \leftarrow f_Y(x, w, u_Y)$ assigns a value in $\mathcal{Y}$ to every configuration $(x, w) \in \mathcal{X} \times \mathcal{W}$, yielding $|\mathcal{U}_Y| = |\mathcal{Y}|^{|\mathcal{X}| \cdot |\mathcal{W}|} = 2^{2 \cdot 3} = 64$. Next, consider Fig. 3(*b*), where unobserved confounding between W and Y induces the partition into c-components $\{X\}$ and $\{W, Y\}$. The canonical confounder U_{WY} jointly indexes the mappings of both W and Y, yielding $|\mathcal{U}_{WY}| = |\mathcal{U}_W| \cdot |\mathcal{U}_Y| = 9 \cdot 64 = 576$, whilst $|\mathcal{U}_X| = 2$.

Lastly, in the fully confounded model of Fig. 3(*c*), all endogenous variables belong to a single c-component $\boldsymbol{C} = \boldsymbol{V}$. A single global exogenous parent U governs all mechanisms, with cardinality $|\mathcal{U}| = |\mathcal{U}_X| \cdot |\mathcal{U}_W| \cdot |\mathcal{U}_Y| = 2 \cdot 9 \cdot 64 = 1{,}152$. The canonical PSCM constructed in the proof of Thm. 2 matches this full-confounding specification.