TECHNICAL REPORT
R-96
First version: May, 2023
Last version: Dec, 2023

Towards Safe Policy Learning under Partial Identifiability: A Causal Approach

Shalmali Joshi[*]
Dept. of Biomedical Informatics
Columbia University
New York, NY

Junzhe Zhang[*]
Dept. of Computer Science
Columbia University
New York, NY

Elias Bareinboim
Dept. of Computer Science
Columbia University
New York, NY

Abstract

Learning personalized treatment policies is a formative challenge in many real-world applications, including in healthcare, econometrics, artificial intelligence. However, the effectiveness of candidate policies is not always *identifiable*, i.e., it is not uniquely computable from the combination of the available data and assumptions about the generating mechanisms. This paper studies policy learning from data collected in various non-identifiable settings, i.e., (1) observational studies with unobserved confounding; (2) randomized experiments with partial observability; and (3) their combinations. We derive sharp, closed-formed bounds from observational and experimental data over the conditional treatment effects. Based on these novel bounds, we further characterize the problem of safe policy learning and develop an algorithm that trains a policy from data guaranteed to achieve, at least, the performance of the baseline policy currently deployed. Finally, we validate our proposed algorithm on synthetic data and a large clinical trial, demonstrating that it guarantees safe behaviors and robust performance.

1 Introduction

Learning optimal personalized treatment policies that maximize a primary outcome by drawing insights from a fixed dataset is a ubiquitous challenge in many real-world applications, including in healthcare, social science, robotics. Several conditions and algorithms have been proposed to solve this problem, including reinforcement learning [59, 35, 61, 33] and causal inference [38, 39, 9]. Most of these algorithms require the critical assumption of *no unmeasured confounding* (NUC) [49], also known as *unconfoundedness*, *ignorability* [53, 52], or *backdoor* admissibility [46, Def. 3.3.1]. This requires that the treatment allocation policy that generates the data considers only the observed covariates; no unobserved confounder affects the treatment and outcome simultaneously. However, the NUC assumption could be fragile and does not necessarily hold in consequential domains with human interactions. For example, when learning personalized medicine from electronic health records (EHR), the physician might unintentionally prescribe a new drug to patients with access to better healthcare, making the drug appear more effective than it is.

A common remedy for the presence of UCs is to perform direct experimentation. The NUC assumption could be made to hold by directly controlling the treatment assignment in specific environments, as sometimes done in randomized trials [17] and online reinforcement learning [60]. Still, the challenge of policy evaluation could arise when experimental data are *partially observed*, i.e., it lacks

[*] Equal Contribution

Preprint. Under review.

critical measurements or has a mismatch in the measured covariates that will be used as input for candidate policies. For example, the Affordable Care Act required hospitals to collect demographic variables such as race that were not routinely collected before 2010 [41, 42]. Consequently, even with randomized controlled trials being performed and the NUC holds, one could not learn a personalized policy to treat patients that accounts for race using past medical data.

Broadly, causal inference provides a collection of principles and tools for evaluating the effects of policies from the combination of data and structural assumptions about the environment [46, 58, 5]. There exist conditions and algorithms to infer the effect of a new intervention from observational studies by leveraging knowledge encoded in its causal models [45, 66, 64, 57, 23]. Further, causal effects can also be inferred from randomized experiments with a mismatch in the intervened treatments [4, 31] and the measured covariates [32]. Recent advancements also lead to complete algorithmic solutions to combine observational and experimental data to identify causal effects [31]. However, when the unobserved confounders (UCs) generally exist, and critical covariates are partially observed, the effects of treatment policies are not necessarily always identifiable [46, Def. 3.2.3]. The treatment effects may not be not uniquely computable from data, despite extensive synthesis and analysis of numerous samples collected across multiple regimes.

Evaluating non-identifiable treatment effects from the combination of data and assumptions has been studied under the rubrics of *partial identification*. There is a growing body of work in causal inference [36, 50, 3, 10, 15, 47, 69, 71, 20], and more recently, in machine learning [28, 40] tackling this challenge. Among these works, one of the following approaches is employed (not exclusively): (1) bounds on the treatment effects are estimated; (2) additional parametric assumptions are invoked, and sensitivity analysis is conducted to assess how treatment effects change as parametric assumptions are perturbed. While cases exist where the partial identification analysis lead to a particular treatment recommendation [12], there is no safety guarantee for the recommended treatment's performance. Our goal in AI is to build intelligent systems that can reason and act autonomously, which means we need to move from a heuristic understanding of the interplay between partial identification and policy learning to a more principled understanding of a robust decision-making process. There are still significant challenges in policy learning under non-identifiability.

This paper aims to overcome these challenges and develop a framework for safe policy learning through causal lenses. In particular, from a fixed dataset, train a policy that is guaranteed to perform as well as a baseline policy currently deployed in the environment [63, 19, 27]. This framework supports evidence-based medicine since the learner can validate whether the treatment policy significantly improves the standard of current care without direct interaction with the patients. Closet to our work, Kallus and Zhou [27] studied the problem of confounding-robust policy improvement that optimizes a policy to achieve the best worst-case improvement relative to a baseline policy. This method computes a policy recommendation based only on observational data. Our work, instead, will account for additional data collected from controlled experiments and explore the nuanced and fundamental interplays between the observational and experimental data on policy evaluation in non-identifiable settings. For a more detailed survey of the related work, we refer readers to Appendix A.

This paper departs from existing approaches and studies safe policy learning in several non-identifiable settings, including learning from observational studies with unobserved confounders, past randomized experiments with partial observability, and combining the two. Our contributions are summarized as follows. (1) We derive closed-form bounds on effect estimates conditioned on new features that combine observational and experimental data (collected with limited context). (2) We prove these bounds are sharp (i.e., cannot be improved without additional assumptions) and identify sufficient conditions when bounds will improve over the purely observational setting. (3) We formulate two objectives and propose a new notion of safe policy learning that leverages these bounds, deviating from worst-case approaches explored in literature. Finally, the proposed approach is evaluated in the synthetic dataset and a large clinical trial. Due to space constraints, all proofs and details on the experiment setup are in Appendix B and Appendix D.

2 Preliminaries

This section will introduce basic notation and definitions used in this paper and provide a short review of related work. We use capital letters X to indicate random variables and lowercase letters x to indicate their realizations. Bold-face capital letters indicate multivariate random variables. Domain of a random variable X is denoted by Ω_X and its cardinality by $|\Omega_X|$. $P(X)$ indicates the probability

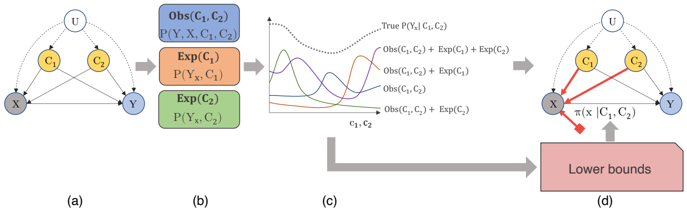

Figure 1: Pipeline of the proposed safe policy learning framework. (a) Causal diagram $\mathcal{G}$ for the underlying SCM $\mathcal{M}^*$ with confounding attributions $\mathbf{C}_1$ and $\mathbf{C}_2$. (b) Available data drawn from the observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional $P_X(Y, \mathbf{C}_1), P_X(Y, \mathbf{C}_2)$ distributions. (c) *Lower bounds* for the true treatment effect of x on Y obtained from different combinations of data. (d) A safe policy $\pi(X \mid \mathbf{C}_1, \mathbf{C}_2)$ optimizing the worst-case treatment effect.

distribution of X and $P(x)$ the probability that $X = x$. Let $\mathbf{1}_{X=x}$ denote an indicator function that takes the value 1 if X realizes to x and is 0 otherwise. We use $[K]$ to denote the set $\{1, 2, \cdots, K\}$. We indicate the event $X \neq x$ with the shorthand notation $\neg x$.

We use Structural Causal Models (SCM) as the basic semantical framework to represent data-generating mechanisms [46]. An SCM is a tuple $\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P \rangle$ where V are the observed random variables in the system, and $\mathbf{U}$ are unobserved exogenous variables that introduce stochasticity in the system. Dependence across observed variables is governed by functional relationships $\mathcal{F}$. That is, for every $V \in \mathbf{V}$, $v \leftarrow f_V(\mathbf{pa}_V, \mathbf{u}_V)$ denotes the values of V will be determined by the function $f_V \in \mathcal{F}$ taking as input a set of observed parents $\mathbf{PA}_V \subseteq \mathbf{V}$ and unobserved parents $\mathbf{U}_V \subseteq \mathbf{U}$. Values of unobserved variable $\mathbf{U}$ are drawn from an exogenous distribution $P(\mathbf{U})$. Naturally, every SCM $\mathcal{M}$ defines an *observational distribution* $P(\mathbf{V})$ over endogenous variables $\mathbf{V}$ [6, Def. 2]. The SCM can be more coarsely represented as a causal diagram $\mathcal{G}$, which is a directed acyclic graph with solid nodes representing observed variables ($\mathbf{V}$), empty nodes for unobserved variables ($\mathbf{U}$), and directed edges codifying the causal dependencies to the parents.

An intervention on a set of observed nodes $\mathbf{X} \subseteq \mathbf{V}$, denoted by $\text{do}(\mathbf{x})$, is an operation that anchors realizations of $\mathbf{X}$ to constants $\mathbf{x}$, removing the dependence on the parents (and exogenous nodes). The $\text{do}(\cdot)$ operation mechanistically allows us to measure the causal effect of the intervened variables $\mathbf{X}$ on the other observed variables $\mathbf{V} \setminus \mathbf{X}$. We will denote the original SCM by $\mathcal{M}$ and the intervened SCM (after a do operation) as $\mathcal{M}_{\mathbf{x}}$. The interventional distribution $P_{\mathbf{x}}(\mathbf{Y})$ is defined as the distribution over $\mathbf{Y}$ in the submodel $\mathcal{M}_{\mathbf{x}}$, i.e., $P_{\mathbf{x}}(\mathbf{Y}) \triangleq P_{\mathcal{M}_{\mathbf{x}}}(\mathbf{Y})$ [6, Def. 5]. We denote by $P_{\mathbf{X}}(\mathbf{Y})$ a collection of interventional distributions $\{P_{\mathbf{x}}(\mathbf{Y}) \mid \forall \mathbf{x} \in \Omega_{\mathbf{X}}\}$. Potential outcomes $\mathbf{Y}_{\mathbf{x}}(\mathbf{u})$ are solutions for a set of observed variables $\mathbf{Y} \subseteq \mathbf{V}$ evaluated in the intervened SCM $\mathcal{M}_{\mathbf{x}}$ after intervention on $\mathbf{x}$. Fix a value $\mathbf{y} \in \Omega_{\mathbf{Y}}$. Let $\mathbf{y}_{\mathbf{x}}$ denote an event $\mathbf{Y}_{\mathbf{x}} = \mathbf{y}$. For a set of variables $\mathbf{X}, \ldots \mathbf{W}, \mathbf{Y}, \ldots, \mathbf{Z}$, the counterfactual distribution $P(\mathbf{Y}_{\mathbf{x}}, \ldots, \mathbf{Z}_{\mathbf{w}})$ is a joint distribution over potential outcomes $\mathbf{Y}_{\mathbf{x}}, \ldots, \mathbf{Z}_{\mathbf{w}}$ in SCM $\mathcal{M}$, given by $P(\mathbf{y}_{\mathbf{x}}, \ldots, \mathbf{z}_{\mathbf{w}}) = \sum_{\mathbf{u}} \mathbf{1}_{\mathbf{Y}_{\mathbf{x}}(\mathbf{u}) = \mathbf{y}, \ldots, \mathbf{Z}_{\mathbf{w}}(\mathbf{u}) = \mathbf{z}} P(\mathbf{u})$ [6, Def. 7].

3 Safe Policy Learning under Partial Identifiability

We will study the problem of optimizing an action X based on values of observed covariates $\mathbf{C} = \{\mathbf{C}_1, \mathbf{C}_2\}$ to maximize a primary outcome (i.e., reward) Y in an SCM $\mathcal{M}^*$. Fig. 1 (a) shows the causal diagram $\mathcal{G}$ associated with this SCM, where unobserved confounders $\mathbf{U}$ exist affecting the action X, outcome Y, and covariates $\mathbf{C}_1, \mathbf{C}_2$, simultaneously. This class of environmental models is also referred to as contextual bandit [30] and is widely applied in reinforcement learning literature [34]. Throughout this paper, we will consistently assume domains of variables $X, Y, \mathbf{C}_1, \mathbf{C}_2$ are discrete and finite; both $\mathbf{C}_1$ and $\mathbf{C}_2$ can be high-dimensional, i.e., $|\Omega_{\mathbf{C}_i}| \gg |\Omega_X|, |\Omega_Y|$ for $i = 1, 2$.

A policy $\pi(X \mid \mathbf{C}_1, \mathbf{C}_2)$ is a function mapping from domains of covariates $\mathbf{C}_1, \mathbf{C}_2$ to the space of probability distribution over the action domain X. The collection of such policies defines a policy space Π. An intervention on action X following the policy π, denoted by $\text{do}(\pi)$, induces an interventional distribution $P_\pi(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ given by

$$P_\pi(x, y, \mathbf{c}_1, \mathbf{c}_2) = P_x(y, \mathbf{c}_1, \mathbf{c}_2)\pi(x \mid \mathbf{c}_1, \mathbf{c}_2) \tag{1}$$

The expected reward associated with a policy $\pi(X \mid \mathbf{C}_1, \mathbf{C}_2)$ is thus given by

$$\mathbb{E}_\pi[Y] = \sum_{x,y,\mathbf{c}_1,\mathbf{c}_2} y P_x(y, \mathbf{c}_1, \mathbf{c}_2)\pi(x \mid \mathbf{c}_1, \mathbf{c}_2) \tag{2}$$

The agent is interested in learning a policy π that maximizes the expected reward $\mathbb{E}_\pi[Y]$ evaluated in SCM $\mathcal{M}^*$. When detailed parametrization of the SCM $\mathcal{M}^*$ is provided, efficient planning algorithms exist to solve for an optimal policy [7, 60]. In many practical applications, however, the environment's system dynamics $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ are assumed to be unknown. Instead, the learner can access data collected from the SCM $\mathcal{M}^*$ under different regimes (observational studies or randomized experiments). We assume access to the following data sources:

1. **Observational Data Obs($\mathbf{C}_1, \mathbf{C}_2$).** An observational study is performed to collected samples Obs($\mathbf{C}_1, \mathbf{C}_2$) drawn from the observational distribution $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$. For notational convenience, since observational data always uses both sets of covariates $\mathbf{C}_1, \mathbf{C}_2$, we denote Obs($\mathbf{C}_1, \mathbf{C}_2$) $\equiv$ Obs.

2. **Experimental Data Exp($\mathbf{C}_1$), Exp($\mathbf{C}_2$).** Randomized controlled trials (RCTs) are conducted on subjects (e.g., patients) from a population $\mathbf{C}_1 = \mathbf{c}_1$ or $\mathbf{C}_2 = \mathbf{c}_2$, but never the combination of the two. This means covariates $\mathbf{C}_1, \mathbf{C}_2$ are never observed simultaneously in experimental data. Consequently, experimental data can come in two forms: Exp($\mathbf{C}_1$) $\sim P_X(Y, \mathbf{C}_1)$ or Exp($\mathbf{C}_2$) $\sim P_X(Y, \mathbf{C}_2)$.

The key challenge in evaluating a policy $\pi(X \mid \mathbf{C}_1, \mathbf{C}_2)$ is to find a function that recovers the expected reward $\mathbb{E}_\pi[Y]$ from Obs($\mathbf{C}_1, \mathbf{C}_2$), Exp($\mathbf{C}_1$), Exp($\mathbf{C}_2$) in all possible SCMs $\mathcal{M}$ generating the data. However, classic results in the causal identification suggest this is infeasible [31, 32].

Corollary 1. *The interventional distribution $P_\pi(Y)$ is not identifiable from $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, and $P_X(Y, \mathbf{C}_2)$ in contextual bandits. That is, there exists SCMs $\mathcal{M}^{(1)}, \mathcal{M}^{(2)}$ compatible with Fig. 1 (a) such that $P^{(1)}(X, Y, \mathbf{C}_1, \mathbf{C}_2) = P^{(2)}(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P^{(1)}_X(Y, \mathbf{C}_1) = P^{(2)}_X(Y, \mathbf{C}_1)$, $P^{(1)}_X(Y, \mathbf{C}_2) = P^{(2)}_X(Y, \mathbf{C}_2)$ while $P^{(1)}_\pi(Y) \neq P^{(2)}_\pi(Y)$ for some policies $\pi(X \mid \mathbf{C}_1, \mathbf{C}_2)$.*

In words, one cannot uniquely determine the expected reward $\mathbb{E}_\pi[Y]$ from any combination of the observational and interventional distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, $P_X(Y, \mathbf{C}_2)$, regardless of how many samples are collected. This result seems to suggest that when the expected reward is not identifiable from available data, it is impossible to learn a policy with satisfactory performance.

To address this challenge, we now formulate the safe policy learning problem. Instead of learning an optimal policy maximizing the expected reward, the agent attempts to obtain a robust policy to achieve a specific baseline performance τ. Let $\mathbb{M}$ be the set of SCMs $\mathcal{M}$ compatible with distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, and $P_X(Y, \mathbf{C}_2)$. Formally, a robust policy π^* is given by

$$\pi^* = \arg\max_{\pi \in \Pi} \min_{\mathcal{M} \in \mathbb{M}} \underbrace{\mathbb{E}_\pi[Y; \mathcal{M}]}_{\text{Worst-case treatment effect}} - \underbrace{\mathbb{E}[Y; \mathcal{M}^*]}_{\text{Baseline performance } \tau} \tag{3}$$

Among quantities in the above maximin program, the inner minimization in the first term computes the worst-case treatment effect of a policy $\pi(X \mid \mathbf{C}_1, \mathbf{C}_2)$. Naturally, the solution $\min_{\mathcal{M}} \mathbb{E}_\pi[Y; \mathcal{M}] \leq \mathbb{E}_\pi[Y; \mathcal{M}^*]$ is a lower bound for the expected reward for policy π evaluated in the true SCM $\mathcal{M}^*$. The second term is the baseline performance $\tau = \mathbb{E}[Y; \mathcal{M}^*]$ achieved by the behavioral policy $X \leftarrow f_X(\mathbf{C}_1, \mathbf{C}_2, \mathbf{U})$ that generates the observational data in the underlying $\mathcal{M}^*$.[2] We show in Fig. 1 a graphical representation of our problem setup.

3.1 Partial Identification from a Single Distribution

Note that in the maximin program of Eq. (3), the baseline $\mathbb{E}[Y; \mathcal{M}^*] = \mathbb{E}[Y]$ is an observational quantity and is computable from distribution $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$. Following Eq. (2) and the convexity of a minimum function, the worst-case treatment effect could be further written as:

$$\min_{\mathcal{M} \in \mathbb{M}} \mathbb{E}_\pi[Y; \mathcal{M}] \geq \sum_{x,y,\mathbf{c}_1,\mathbf{c}_2} \pi(x \mid \mathbf{c}_1, \mathbf{c}_2) y \min_{\mathcal{M} \in \mathbb{M}} P_x(y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{4}$$

It is thus sufficient to consider the problem of bounding interventional probabilities $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ from distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, and $P_X(Y, \mathbf{C}_2)$. Formally,

[2] More generally, the baseline performance $\tau \in \mathbb{R}$ could be an arbitrary real value based on the context. This paper focuses on finding a robust policy that improves over the policy f_X currently deployed in the environment.

Definition 1 (Lower Causal Bound)**.** Let $\mathcal{G}$ be a causal diagram over variables $\mathbf{V}$, $\mathcal{P}$ be a set of distributions (observational or interventional) over $\mathbf{V}$, and $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$ be (disjoint) subsets of $\mathbf{V}$. A lower bound over the causal effects $P_{\mathbf{X}}(\mathbf{Y})$ is *an expression for a function* $l(\mathbf{x}, \mathbf{y})$ in terms of $\mathcal{P}$ such that for every SCM $\mathcal{M}$ compatible with $\mathcal{G}$, $P_{\mathbf{x}}(\mathbf{y}; \mathcal{M}) \geq l(\mathbf{x}, \mathbf{y}; \mathcal{M}), \forall (\mathbf{x}, \mathbf{y}) \in \Omega_{\mathbf{X}} \times \Omega_{\mathbf{Y}}$.

First, a function of the observational distribution $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ that consistently lower bounds $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ in all SCM $\mathcal{M}$ compatible with Fig. 1 (a), called the *natural bound* [36],

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2) \geq P(x, y, \mathbf{c}_1, \mathbf{c}_2) \tag{5}$$

Interestingly, it can be shown that the marginal interventional distribution $P_X(Y, \mathbf{C}_1)$ or $P_X(Y, \mathbf{C}_2)$ does not impose any informative constraint over the joint distribution $P_X(Y, \mathbf{C}_1, \mathbf{C}_2)$. Consider, for example, $\mathbf{C}_1 = \mathbf{c}_1$. One could always construct an SCM $\mathcal{M}$ compatible with Fig. 1 (a) such that $P_x(y, \mathbf{c}_1; \mathcal{M}) = P_x(y, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M})$ and $P_x(y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = 0$. This implies a lower bound

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2) \geq 0 \tag{6}$$

So far, our analysis reveals that marginal interventional distributions $P_X(Y, \mathbf{C}_1)$ or $P_X(Y, \mathbf{C}_2)$ do not impose any meaningful constraint over the target effect $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$. This seems to suggest when computing the worst-case treatment effect, it is sufficient to consider only the observational distribution $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$. For the remainder of this paper, we will show this is not the case by investigating non-trivial interactions between the observational and interventional distributions.

4 Partial Identification from Multiple Distributions

This section will derive novel lower bounds over the target causal effects $P_X(Y, \mathbf{C}_1, \mathbf{C}_2)$ from the combination of observational and interventional distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, $P_X(Y, \mathbf{C}_2)$ in the models compatible with the causal diagram of Fig. 1 (a). We start with a novel lower bound over the target effects by combining the observational distribution $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and a marginal interventional distribution $P_X(Y, \mathbf{C}_1)$ over partial covariates $\mathbf{C}_1$.

Lemma 1 (Obs+Exp($\mathbf{C}_1$))**.** *Given distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and $P_X(Y, \mathbf{C}_1)$, the lower bound over $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ for all $(x, y, \mathbf{c}_1, \mathbf{c}_2) \in \Omega_X \times \Omega_Y \times \Omega_{\mathbf{C}_1} \times \Omega_{\mathbf{C}_2}$ is given by*

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2) \geq \max\left\{l_1(x, y, \mathbf{c}_1, \mathbf{c}_2), l_2(x, y, \mathbf{c}_1, \mathbf{c}_2)\right\} \tag{7}$$

where l_1, l_2 are functions defined as

$$l_1(x, y, \mathbf{c}_1, \mathbf{c}_2) = P(x, y, \mathbf{c}_1, \mathbf{c}_2) \tag{8}$$

$$l_2(x, y, \mathbf{c}_1, \mathbf{c}_2) = P_x(y, \mathbf{c}_1) - P(x, y, \mathbf{c}_1, \neg\mathbf{c}_2) - P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \tag{9}$$

Among the quantities in Lem. 1, $l_1(x, y, \mathbf{c}_1, \mathbf{c}_2)$ is the natural bound, but $l_2(x, y, \mathbf{c}_1, \mathbf{c}_2)$ is a function of both the observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional $P_X(Y, \mathbf{C}_1)$ distribution. It follows immediately that the bound in Lem. 1 is never inferior to the natural bound.

Definition 2. Let $\mathcal{G}$ be a causal diagram over variables $\mathbf{V}$, $\mathcal{P}$ be a set of distributions over $\mathbf{V}$, and $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$. For lower bounds $l_1(\mathbf{x}, \mathbf{y})$ and $l_2(\mathbf{x}, \mathbf{y})$ over the causal effects $P_{\mathbf{X}}(\mathbf{Y})$, $l_1(\mathbf{x}, \mathbf{y})$ is said to *consistently dominate* $l_2(\mathbf{x}, \mathbf{y})$ if the following conditions hold:

(i) For every SCM $\mathcal{M}$ compatible with $\mathcal{G}$, $l_1(\mathbf{x}, \mathbf{y}; \mathcal{M}) \geq l_2(\mathbf{x}, \mathbf{y}; \mathcal{M}), \forall (\mathbf{x}, \mathbf{y}) \in \Omega_{\mathbf{X}} \times \Omega_{\mathbf{Y}}$.
(ii) There is an SCM $\mathcal{M}$ compatible with $\mathcal{G}$ s.t. $l_1(\mathbf{x}, \mathbf{y}; \mathcal{M}) > l_2(\mathbf{x}, \mathbf{y}; \mathcal{M}), \exists (\mathbf{x}, \mathbf{y}) \in \Omega_{\mathbf{X}} \times \Omega_{\mathbf{Y}}$.

The more interesting question is whether instances exist where Lem. 1 is strictly tighter than the natural bound. For instance, consider an SCM $\mathcal{M}^*$ compatible with Fig. 1 (a) with exogenous variables $\mathbf{U} = \{U_1, U_2, U_3\}$ independently drawn over the binary domain $\{0, 1\}$ such that $P(U_1 = 0) = P(U_2 = 1) = 0.9$, $P(U_3 = 0) = 0.5$. Values of $X, Y, \mathbf{C}_1, \mathbf{C}_2$ in $\mathcal{M}^*$ are given by

$$X \leftarrow U_1 \oplus U_3, \qquad Y \leftarrow X \oplus U_1 \oplus U_2, \qquad \mathbf{C}_1 \leftarrow U_1, \qquad \mathbf{C}_2 \leftarrow U_2 \tag{10}$$

where $\oplus$ is the "xor" operator. Evaluating the causal effect $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ in SCM $\mathcal{M}^*$ gives

$$P_{X=0}(Y = 1, \mathbf{C}_1 = 0, \mathbf{C}_2 = 1) = P(U_1 = 0, U_2 = 1) = 0.81 \tag{11}$$

Evaluating the corresponding natural bound $l_1(x, y, \mathbf{c}_1, \mathbf{c}_2)$ gives

$$\begin{aligned} l_1(X = 0, Y = 1, &\mathbf{C}_1 = 0, \mathbf{C}_2 = 1) = \\ &P(U_1 = 0, U_2 = 1, U_3 = 0) = 0.405 \end{aligned} \tag{12}$$

On the other hand, evaluating the new bound $l_2(x, y, \mathbf{c}_1, \mathbf{c}_2)$ from Eq. (14) in SCM $\mathcal{M}^*$ gives

$$\begin{aligned} l_2(X = 0, Y = 1, \mathbf{C}_1 = 0, \mathbf{C}_2 = 1) = P(U_1 = 0, U_2 = 1) - \\ P(U_1 = 0, U_2 = 0, U_3 = 1) \end{aligned} \tag{13}$$

Computing Eq. (13) gives. $l_2(X = 0, Y = 1, \mathbf{C}_1 = 0, \mathbf{C}_2 = 1) = 0.765$, which is larger than the natural bound. Recall that a single marginal distribution $P_X(Y, \mathbf{C}_1)$ does not impose any (lower) constraint on $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$. Lem. 1 thus improves over the natural bound by exploring interactions between observational and interventional distributions.

Proposition 1. *Given distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and $P_X(Y, \mathbf{C}_1)$, the lower bound given in Lem. 1 consistently dominates the natural bound (Eq.* (5)*).*

We also provide a lower bound computed from marginal distributions $P_X(Y, \mathbf{C}_1), P_X(Y, \mathbf{C}_2)$, i.e., the observational distribution $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ is not available.

Lemma 2 (Exp($\mathbf{C}_1$) + Exp($\mathbf{C}_2$))**.** *Given distributions $P_X(Y, \mathbf{C}_1)$ and $P_X(Y, \mathbf{C}_2)$, the lower bound over $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ for all $(x, y, \mathbf{c}_1, \mathbf{c}_2) \in \Omega_X \times \Omega_Y \times \Omega_{\mathbf{C}_1} \times \Omega_{\mathbf{C}_2}$ is given by*

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2) \geq P_x(y, \mathbf{c}_1) - P_x(y, \neg\mathbf{c}_2) \tag{14}$$

The above bound is informative if $P_x(y, \mathbf{c}_1) > P_x(y, \neg\mathbf{c}_2)$. However, there is no clear preference between the interventional bound in Lem. 2 and other bounds computed using the observational distribution, including the one in Lem. 1. Finally, we provide a novel bound utilizing all available data, including the observational and marginal interventional distributions over covariates $\mathbf{C}_1$, $\mathbf{C}_2$.

Theorem 1 (Obs + Exp($\mathbf{C}_1$) + Exp($\mathbf{C}_2$))**.** *Given distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, and $P_X(Y, \mathbf{C}_2)$, the lower bound over $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ for all $(x, y, \mathbf{c}_1, \mathbf{c}_2) \in \Omega_X \times \Omega_Y \times \Omega_{\mathbf{C}_1} \times \Omega_{\mathbf{C}_2}$ is*

$$\begin{aligned} P_x(y, \mathbf{c}_1, \mathbf{c}_2) \geq \max\{l_1(x, y, \mathbf{c}_1, \mathbf{c}_2), l_2(x, y, \mathbf{c}_1, \mathbf{c}_2), \\ l_3(x, y, \mathbf{c}_1, \mathbf{c}_2), l_4(x, y, \mathbf{c}_1, \mathbf{c}_2)\} \end{aligned} \tag{15}$$

where l_1, l_2 are given by Eqs. (8) *and* (9)*, respectively; l_3, l_4 are functions defined as*

$$\begin{aligned} l_3(x, y, \mathbf{c}_1, \mathbf{c}_2) = P_x(y, \mathbf{c}_2) - P(x, y, \neg\mathbf{c}_1, \mathbf{c}_2) - \\ P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \end{aligned} \tag{16}$$

$$\begin{aligned} l_4(x, y, \mathbf{c}_1, \mathbf{c}_2) = P_x(y, \mathbf{c}_1) - P_x(y, \neg\mathbf{c}_2) + \\ P(x, y, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \end{aligned} \tag{17}$$

Among quantities in Thm. 1, $l_3(x, y, \mathbf{c}_1, \mathbf{c}_2)$ is symmetric to $l_2(x, y, \mathbf{c}_1, \mathbf{c}_2)$, and follows from applying Lem. 1 with input $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_2)$. The constraint in $l_4(x, y, \mathbf{c}_1, \mathbf{c}_2)$ is a function of all available distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, $P_X(Y, \mathbf{C}_2)$. One could see by inspection that Thm. 1 improves over the interventional bound in Lem. 2 if $P(x, y, \neg\mathbf{c}_1, \neg\mathbf{c}_2) > 0$. A more interesting question is how it compares with the bound given by Lem. 1. Consider again the SCM $\mathcal{M}^*$ described in Eq. (10). Evaluating the lower bound $l_4(x, y, \mathbf{c}_1, \mathbf{c}_2)$ gives

$$\begin{aligned} l_4(X = 0, Y = 1, \mathbf{C}_1 = 0, \mathbf{C}_2 = 1) = P(U_1 = 0, U_2 = 1) - \\ P(U_1 = 1, U_2 = 0, U_3 = 0) \end{aligned} \tag{18}$$

Computing the above equation gives, $l_4(X = 0, Y = 1, \mathbf{C}_1 = 0, \mathbf{C}_2 = 1) = 0.805$, which consistently dominates lower bounds l_1, l_2 evaluated in Eqs. (12) and (13).

Proposition 2. *Given distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, and $P_X(Y, \mathbf{C}_2)$, the lower bound given in Thm. 1 consistently dominates Lems. 1 and 2 and the natural bound (Eq.* (5)*).*

In Fig. 2 we summarize all the bounds derived in this paper and their relationships. A single marginal distribution $P_X(Y, \mathbf{C}_1)$ or $P_X(Y, \mathbf{C}_2)$ does not impose any constraint on the target effect $P_X(Y, \mathbf{C}_1, \mathbf{C}_2)$. One can derive meaningful bounds by incorporating the observational distribution $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ or multiple interventional distributions $P_X(Y, \mathbf{C}_1)$ and $P_X(Y, \mathbf{C}_2)$. Finally, Thm. 1 presents the most informative bounds using all available data sources.

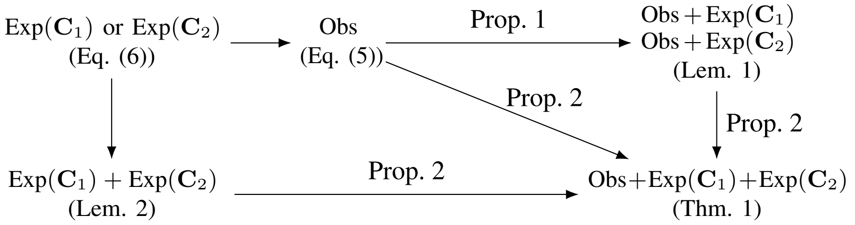
Exp($\mathbf{C}_1$) or Exp($\mathbf{C}_2$)
(Eq. (6))
Obs
(Eq. (5))
Prop. 1
Obs + Exp($\mathbf{C}_1$)
Obs + Exp($\mathbf{C}_2$)
(Lem. 1)
Prop. 2
Prop. 2
Exp($\mathbf{C}_1$) + Exp($\mathbf{C}_2$)
(Lem. 2)
Prop. 2
Obs+Exp($\mathbf{C}_1$)+Exp($\mathbf{C}_2$)
(Thm. 1)

Figure 2: Hierarchy of lower bounds derived from different data sources. A directed path from $\mathcal{D}_1$ to $\mathcal{D}_2$ indicates the bound derived from dataset $\mathcal{D}_2$ consistently dominates the one from dataset $\mathcal{D}_1$.

4.1 Sharpness Conditions of Closed-form Bounds

A natural question at this point is whether the bound provided in Thm. 1 is sharp, i.e., a tighter bound can be derived through a more refined analysis. Fortunately, we will show this is not the case.

Definition 3 (Sharp Lower Bound)**.** Let $\mathcal{G}$ be a causal diagram over variables $\mathbf{V}$, $\mathcal{P}$ be a set of distributions over $\mathbf{V}$, and $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$. A lower bound $l(\mathbf{x}, \mathbf{y})$ over the causal effects $P_{\mathbf{X}}(\mathbf{Y})$ from $\mathcal{P}$ is said to be *sharp* if there is no other lower bound $l^*(\mathbf{x}, \mathbf{y})$ that consistently dominates $l(\mathbf{x}, \mathbf{y})$.

Suppose the bound in Thm. 1, denoted by $l = \max\{l_1, l_2, l_3, l_4\}$, is not sharp, and there is a different lower bound l^* consistently dominates l. There must exist an SCM $\mathcal{M}$ compatible with Fig. 1 (a) and a realization $(x, y, \mathbf{c}_1, \mathbf{c}_2)$ such that $l^*(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) > l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})$. The key challenge is to construct an alternative SCM $\mathcal{M}^*$ from $\mathcal{M}$ so that the lower bound l^* no longer applies.

Theorem 2. *Given distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, and $P_X(Y, \mathbf{C}_2)$, Thm. 1 is a sharp lower bound over the causal effects $P_X(Y, \mathbf{C}_1, \mathbf{C}_2)$ in the causal diagram of Fig. 1 (a).*

Proof (sketch). Suppose there is an SCM $\mathcal{M}$ where $l^*(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) > l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})$ for some $(x, y, \mathbf{c}_1, \mathbf{c}_2)$. Construct an alternative SCM $\mathcal{M}^*$ such that (1) $\mathcal{M}^*$ and $\mathcal{M}$ share the same $P(X, Y, \mathbf{C}_1, \mathbf{C}_1)$, $P_X(Y, \mathbf{C}_1)$, $P_X(Y, \mathbf{C}_2)$; and (2) the counterfactual distribution $P(X, Y_x, \mathbf{C}_1, \mathbf{C}_2)$ in $\mathcal{M}^*$ satisfy the following, based on the evaluation of lower bound l in $\mathcal{M}$:

$$\begin{cases} P(y_x \mid \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) = 0 & \text{if } l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_1(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ P(y_x \mid \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}^*) = 1 & \text{if } l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_2(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ P(y_x \mid \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) = 1 & \text{if } l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_3(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ P(y_x \mid \neg x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}^*) = 0 & \text{if } l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_4(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{cases} \tag{19}$$

This construction is feasible since observational and interventional distributions are under-determined by counterfactual distributions in SCMs [6]. It is verifiable in this modified SCM $\mathcal{M}^*$, the target effect $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ matches the lower bound l given by Thm. 1,

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) = l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) \tag{20}$$

Since lower bounds l and l^* are functions of distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_1)$, $P_X(Y, \mathbf{C}_1)$, $P_X(Y, \mathbf{C}_2)$ which are shared across $\mathcal{M}$ and $\mathcal{M}^*$, we must have

$$l^*(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) = l^*(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}), \tag{21}$$

$$l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) = l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{22}$$

Since l^* consistently dominates l in SCM $\mathcal{M}$, the above equations imply

$$l^*(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) > l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) = P_x(y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) \tag{23}$$

This means that l^* is not a valid lower bound for $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ in $\mathcal{M}^*$, which is a contradiction. □

Algorithm 1 Safe Policy Learning

```
Input: Samples $(y_i, x_i, \mathbf{c}_{1,i}, \mathbf{c}_{2,i})_{i=1}^{N_{\text{obs}}}$, $(y_i, x_i, \mathbf{c}_{1,i})_{i=1}^{N_{\text{exp1}}}$, $(y_i, x_i, \mathbf{c}_{2,i})_{i=1}^{N_{\text{exp2}}}$ and learning rate $\lambda > 0$
Estimate lower bounds $l_j(x, y_1, \mathbf{c}_1, \mathbf{c}_2)$, $j = 1, \ldots, 4$, given by Thm. 1 from the observational and experimental data $(y_i, x_i, \mathbf{c}_{1,i}, \mathbf{c}_{2,i})_{i=1}^{N_{\text{obs}}}$, $(y_i, x_i, \mathbf{c}_{1,i})_{i=1}^{N_{\text{exp1}}}$, $(y_i, x_i, \mathbf{c}_{2,i})_{i=1}^{N_{\text{exp2}}}$
for $x, i \in \Omega_X \times [N_{obs}]$ do
    $w_i(x, y_1, \mathbf{c}_{1,i}, \mathbf{c}_{2,i}) \leftarrow \max_{j=1,\ldots,4} l_j(x, y_1, \mathbf{c}_{1,i}, \mathbf{c}_{2,i})$
end for
Initialize parameters of $\pi_0$ randomly.
for $e \in \{1, 2, \cdots, N_{\text{epochs}}\}$ do
    $\pi_{e+1} \leftarrow \pi_e + \lambda \nabla_\pi \mathrm{V}_{N_{obs}}(\pi)$ such that $\pi_e \in \Pi$
end for
return $\pi_{N_{\text{epochs}}+1}$
```

5 Safe Policy Learning with Partial Effects

We now apply the closed-form bounds derived so far to solve for a safe policy that outperforms the baseline policy. Without loss of generality, assume the reward $Y \in \{0, 1\}$; let y_1 denote the event $Y = 1$. By replacing the inner minimization in Eq. (4) with the lower bound given in Thm. 1, the worst-case treatment effect of policy π could be written as:

$$\min_{\mathcal{M} \in \mathbb{M}} \mathbb{E}_\pi[Y; \mathcal{M}] \geq \sum_{x, y, \mathbf{c}_1, \mathbf{c}_2} \pi(x \mid \mathbf{c}_1, \mathbf{c}_2) y \max_{j=1,\ldots,4} l_j(x, y_1, \mathbf{c}_1, \mathbf{c}_2) \tag{24}$$

The maximin objective proposed in Eq. (3) could thus be written as

$$\arg\max_{\pi \in \Pi} \underbrace{\mathbb{E}_{x \sim \pi(x|\mathbf{c}_1, \mathbf{c}_2)} \left[w(x, y_1, \mathbf{c}_1, \mathbf{c}_2) \right]}_{\text{Value function } \mathrm{V}(\pi)} - \underbrace{\mathbb{E}[Y]}_{\text{Baseline performance } \mathcal{T}} \tag{25}$$

Among the above quantities, the performance baseline $\mathbb{E}[Y]$ is estimable by computing the empirical mean of reward in the observational data; the weight w is a function defined as

$$w(x, y, \mathbf{c}_1, \mathbf{c}_2) \triangleq \max_{j=1,\ldots,4} l_j(x, y, \mathbf{c}_1, \mathbf{c}_2) \tag{26}$$

Given access to multiple distributions, the worst-case treatment effect is the best estimate of the lower bound leveraging all data sources/multiple distributions. Thus, the minimax problem in Eq. (25) reduces to weighted maximization that corresponds to the best worst-case treatment effects.

We propose a three-step algorithm to learn a safe policy π, in Alg. 1 for the case when the bounds in Thm. 1 can be estimated reliably from data. In Step 1, we use plug-in estimates of the bounds given by Thm. 1 by first deriving the bounds as a function of conditional distributions $P(y|x, \mathbf{c}_1, \mathbf{c}_2), P(x|\mathbf{c}_1, \mathbf{c}_2), P_x(y|\mathbf{c}_1), P_x(y|\mathbf{c}_2), P(x|\mathbf{c}_1)$ and $P(x|\mathbf{c}_2)$ (see Appendix C.1 for a detailed discussion), and then estimating a plug-in bound by estimating these conditional using standard supervised learning methods, such as regression and/or supervised classification, as appropriate. Steps 2 - 4 estimate the worst-case treatment effects for all instances of observed covariates $(\mathbf{c}_{1,i}, \mathbf{c}_{2,i})$ by computing the weight function w_i. Here we only consider the observational data since covariates $\mathbf{C}_1, \mathbf{C}_2$ are observed simultaneously. Steps 5 - 8 optimize for a safe policy, using estimates w_i, effectively resulting in a differentiable weighted loss function (Eq. (25)) that is maximized over a function family Π. We use policy ascent for the maximization. Alg. 1 is effectively an Oracle-based algorithm since we do not consider the statistical challenges of estimating the bounds.

6 Experiments

We evaluate the proposed method on 1) Synthetic data, and 2) the International Stroke Trial (IST) data [21, 54] and learn four policies. Each policy uses one or more derived bounds in the maximin framework: i) Alg 1 $(\mathbf{C}_1, \mathbf{C}_2)$ - l_1: Uses only Obs bound, ii) Alg 1 $(\mathbf{C}_1, \mathbf{C}_2)$ - l_1, l_2: Uses Obs and Obs+Exp($\mathbf{C}_1$) bound, iii) Alg 1 $(\mathbf{C}_1, \mathbf{C}_2)$ - l_1, l_3: Uses Obs and Obs+Exp($\mathbf{C}_2$) bound, and iv) Alg 1 $(\mathbf{C}_1, \mathbf{C}_2)$ - l_1, l_2, l_3, l_4: Uses all bounds. We compare to the following baselines: i) Random policy, ii) Behavior policy $(\mathbf{C}_1, \mathbf{C}_2)$: policy used to collect the observational data.

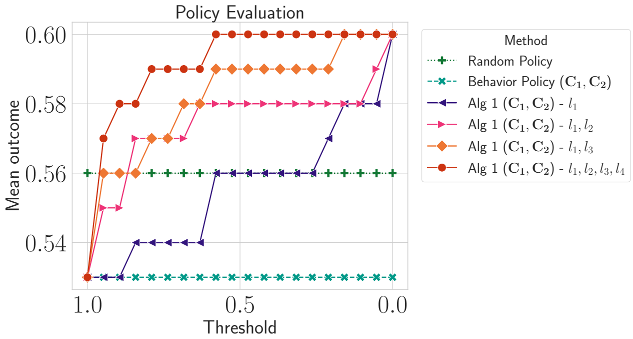

Figure 3: Synthetic Data policy evaluation at varying the threshold on policy scores, 1 (zero treated) $\to 0$ (all treated). Higher is better.

1) Synthetic Data. The generative process of the data is,

$$U \sim \mathcal{N}(0,1); \quad \mathbf{C}_1 = 0.1U + 2.05e - 04;$$

$$\mathbf{C}_2 = 0.43U + 8.97e - 05$$

$$x = \mathbf{1}\left(\sigma([-0.06, 0.39, 0.46] \cdot [U, \mathbf{C}_1, \mathbf{C}_2]^T + \mathcal{N}(0,1)) > 0.5\right)$$

$$y = \mathbf{1}\left(\sigma([-0.16, 0.41, 0.04, 0.1] \cdot [U, \mathbf{C}_1, \mathbf{C}_2, x]^T + \mathcal{N}(0,1)) > 0.5\right)$$

Three data sources are generated, each consisting of $1,000$ samples corresponding to Obs, Exp$(\mathbf{C}_1)$, and Exp$(\mathbf{C}_2)$. For experimental data, the treatment is sampled using: $x \sim$ Bernoulli(0.5). To implement Alg. 1, we first estimate all lower bounds using their simplified form derived in Appendix C.1. These require estimating intermediate conditional distributions such as $P(y|x, \mathbf{c}_1, \mathbf{c}_2), P(x|\mathbf{c}_1, \mathbf{c}_2), P(x|\mathbf{c}_1), P(x|\mathbf{c}_2)$ which require marginalization over U and/or $\mathbf{C}_1, \mathbf{C}_2$ for which numerical integration was used. These estimates are then used in Alg. 1 to learn a treatment policy π. The function family Π (see Eq. (25)) corresponds to a two-layer Multi-layer Perceptron (MLP) with 5 hidden units and the GELU activations [22].

Each policy returns a score between 0 and 1. All samples above a threshold can be chosen for treatment. We evaluate mean outcome over the data, of varying the threshold between 1 (treat no one) to 0 (treat everyone). Fig. 3 shows the mean outcome for varying thresholds averaged over 5-fold cross-validation (standard errors not visible due to low variability). The learned policy Alg 1 $(\mathbf{C}_1, \mathbf{C}_2)$ - l_1, l_2, l_3, l_4 clearly outperforms all baselines suggesting our estimates of treatment effect indeed improve using multiple data-sources and can be leveraged to learn not only safe, but improved policies. Further, the bounds Obs+ Exp$(\mathbf{C}_1)$, and Obs+ Exp$(\mathbf{C}_2)$ improve over the natural bounds Obs for some covariate values (see Fig. 5 in Appendix D) providing better policies compared to Obs. Finally, Fig. 5 suggests that bounds using Obs+ Exp$(\mathbf{C}_1)$+Exp$(\mathbf{C}_2)$ are not informative. Nonetheless, using l_1, l_2, l_3 in conjunction provide significant improvement over behavior policy, and other variants.

International Stroke Trial (IST). Broadly, the goal of this trial was to estimate the effect of Aspirin treatment on the clinical course of Ischemic Stroke. The database consists of patient age, gender, conscious state at randomization, and systolic blood pressure at randomization, among others. We study the outcome at 6 months post-treatment ($Y = 1$ if a patient survives, 0 otherwise). From this trial data, we create an observational dataset by inducing selection bias as a function of age, gender, conscious state at randomization, and systolic blood pressure (see Appendix D.3 for more details). We treat conscious state as unobserved confounding. $\mathbf{C}_1 = \{$Age, Sex$\}$ and $\mathbf{C}_2 = \{$Systolic Blood Pressure (SysBP)$\}$ constitute observed confounding attributes. We set aside 30% data as a held-out test set.

Fig. 4 demonstrates the results. On the left, we show the mean outcome for varying thresholds on the policy score (a higher threshold implies fewer patients selected for treatment). Thus higher mean outcome while selecting fewer patients is desirable. The learned policies clearly dominate the behavior policy. On the right, we show the outcome when x-fraction of the population is selected for treatment based on policy scores. Again, while the improvement over behavior policy is significant, improvement using different bounds is limited. Fig. 8 in Appendix D.3 shows the bounds obtained in each case. Notice that the effect of treatment is relatively small and only occurs for a small region of

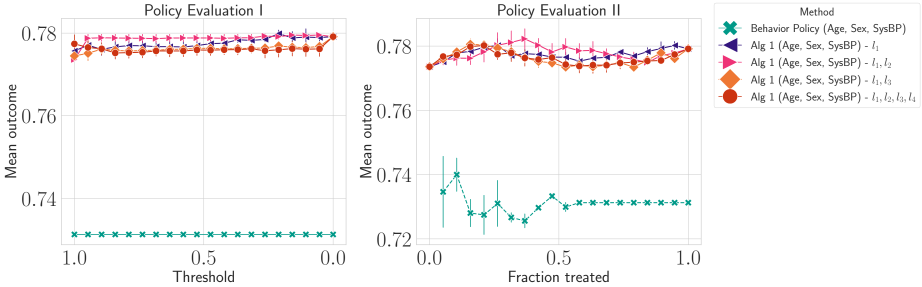

Figure 4: IST Policy Evaluation. Left: Policy evaluation with varying threshold $1\,(\text{zero treated}) \to 0\,(\text{all treated})$ on policy scores. For fewer treated patients (higher threshold), learned policies peak earlier suggesting that our learned policies are better at targeting patients over behavior policy. Right: Mean outcome when x-fraction of the population is targeted for treatment using sorted policy scores. Behavior policy trails relative to the learned policies although the difference between learned policy variants is not significant suggesting other patient features may be crucial to improve targeting.

the SysBP space. Second, our bounds require parametric assumptions on $P(\text{Age, Sex, SysBP})$. We model SysBP as a Gaussian, and Age and Sex as independent Bernoulli variables. Misspecification in our parametrization may result in biased estimates in our bounds. Additional covariates beyond Age, Sex, and SysBP may further improve the bounds and in turn the treatment policy.

7 Conclusions

We propose a safe policy learning framework in non-identifiable settings using observational studies with unobserved confounding, experimental studies with partial observability, and combinations thereof. We derive closed-form bounds over conditional treatment effects. We propose a robust policy improvement framework to train policies that maximize the worst-case treatment effect by using our lower bounds that is guaranteed to improve over a baseline policy that generate the observational data. We demonstrate utility in synthetic and real-world experimental data. We include in Appendix E more detailed discussion on limitations and broader impacts of our proposed framework.

8 Acknowledgements

This research was supported in part by the NSF, ONR, AFOSR, DoE, Amazon, JP Morgan, and The Alfred P. Sloan Foundation.

References

[1] Eitan Altman. *Constrained Markov decision processes*, volume 7. CRC press, 1999.

[2] Alexander Balke and Judea Pearl. Counterfactual probabilities: Computational methods, bounds and applications. In *Uncertainty Proceedings 1994*, pages 46–54. Elsevier, 1994.

[3] Alexander Balke and Judea Pearl. Counterfactuals and policy analysis in structural models. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 11–18, 1995.

[4] E. Bareinboim and J. Pearl. Causal inference by surrogate experiments: z-identifiability. In Nando de Freitas and Kevin Murphy, editors, *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pages 113–120, Corvallis, OR, 2012. AUAI Press.

[5] Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

[6] Elias Bareinboim, JD Correa, Duligur Ibeling, and Thomas Icard. On pearl's hierarchy and the foundations of causal inference. *ACM Special Volume in Honor of Judea Pearl (provisional title)*, 2020.

[7] Richard Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

[8] Nicolas Carrara, Edouard Leurent, Romain Laroche, Tanguy Urvoy, Odalric-Ambrym Maillard, and Olivier Pietquin. Budgeted reinforcement learning in continuous state space. *Advances in Neural Information Processing Systems*, 32, 2019.

[9] Bibhas Chakraborty and Susan A Murphy. Dynamic treatment regimes. *Annual review of statistics and its application*, 1:447–464, 2014.

[10] D.M. Chickering and J. Pearl. A clinician's apprentice for analyzing non-compliance. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, volume Volume II, pages 1269–1276. MIT Press, Menlo Park, CA, 1996.

[11] Carlos Cinelli, Daniel Kumor, Bryant Chen, Judea Pearl, and Elias Bareinboim. Sensitivity analysis of linear structural causal models. In *International conference on machine learning*, pages 1252–1261. PMLR, 2019.

[12] Jerome Cornfield, William Haenszel, E Cuyler Hammond, Abraham M Lilienfeld, Michael B Shimkin, and Ernst L Wynder. Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer institute*, 22(1):173–203, 1959.

[13] El Mahdi El Mhamdi, Rachid Guerraoui, Hadrien Hendrikx, and Alexandre Maurer. Dynamic safe interruptibility for decentralized multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 30, 2017.

[14] Y Fan and S Park. Usharp bounds on the distribution of the treatment effect and their statistical inference,% forthcoming in econometric theory. 2007.

[15] Yanqin Fan and Sang Soo Park. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26(3):931–951, 2010. ISSN 02664666, 14694360. URL `http://www.jstor.org/stable/40664510`.

[16] Mehdi Fatemi, Shikhar Sharma, Harm Van Seijen, and Samira Ebrahimi Kahou. Dead-ends and secure exploration in reinforcement learning. In *International Conference on Machine Learning*, pages 1873–1881. PMLR, 2019.

[17] R.A. Fisher. The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain*, 33:503–513, 1926.

[18] Javier Garcıa and Fernando Fernández. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(1):1437–1480, 2015.

[19] Mohammad Ghavamzadeh, Marek Petrik, and Yinlam Chow. Safe policy improvement by minimizing robust baseline regret. *Advances in Neural Information Processing Systems*, 29, 2016.

[20] Luigi Gresele, Julius Von Kügelgen, Jonas Kübler, Elke Kirschbaum, Bernhard Schölkopf, and Dominik Janzing. Causal inference through the structural causal marginal problem. In *International Conference on Machine Learning*, pages 7793–7824. PMLR, 2022.

[21] International Stroke Trial Collaborative Group et al. The international stroke trial (ist): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. *The Lancet*, 349(9065):1569–1581, 1997.

[22] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

[23] Y. Huang and M. Valtorta. Pearl's calculus of intervention is complete. In R. Dechter and T.S. Richardson, editors, *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 217–224. AUAI Press, Corvallis, OR, 2006.

[24] G Imbens and J Angrist. Estimation and identification of local average treatment effects. *Econometrica*, 62:467–475, 1994.

[25] Guido W Imbens and Charles F Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004.

[26] Andrew Jesson, Alyson Douglas, Peter Manshausen, Nicolai Meinshausen, Philip Stier, Yarin Gal, and Uri Shalit. Scalable sensitivity and uncertainty analysis for causal-effect estimates of continuous-valued interventions. *arXiv preprint arXiv:2204.10022*, 2022.

[27] Nathan Kallus and Angela Zhou. Confounding-robust policy improvement. *arXiv preprint arXiv:1805.08593*, 2018.

[28] Nathan Kallus, Aahlad Manas Puli, and Uri Shalit. Removing hidden confounding by experimental grounding. *Advances in neural information processing systems*, 31, 2018.

[29] Nathan Kallus, Xiaojie Mao, and Angela Zhou. Interval estimation of individual-level causal effects under unobserved confounding. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2281–2290. PMLR, 2019.

[30] John Langford and Tong Zhang. The epoch-greedy algorithm for multi-armed bandits with side information. In J.C. Platt, D. Koller, Y. Singer, and S.T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 817–824. Curran Associates, Inc., 2008.

[31] S. Lee, J. Correa, and E. Bareinboim. General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, Tel Aviv, Israel, 2019. AUAI Press.

[32] Sanghack Lee and Elias Bareinboim. Causal effect identifiability under partial-observability. In *International Conference on Machine Learning*, pages 5692–5701. PMLR, 2020.

[33] Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*, 2020.

[34] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670. ACM, 2010.

[35] Lihong Li, Remi Munos, and Csaba Szepesvari. Toward minimax off-policy value estimation. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, May 2015. URL `https://www.microsoft.com/en-us/research/publication/toward-minimax-off-policy-value-estimation/`.

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

[37] Charles F Manski and John V Pepper. Monotone instrumental variables with an application to the returns to schooling, 1998.

[38] Susan A Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.

[39] Susan A Murphy. A generalization error for q-learning. *Journal of machine learning research: JMLR*, 6:1073–1097, 2005.

[40] Hongseok Namkoong, Ramtin Keramati, Steve Yadlowsky, and Emma Brunskill. Off-policy policy evaluation for sequential decisions under unobserved confounding. *Advances in Neural Information Processing Systems*, 33:18819–18831, 2020.

[41] Amani M Nuru-Jeter, Elizabeth K Michaels, Marilyn D Thomas, Alexis N Reeves, Roland J Thorpe Jr, and Thomas A LaVeist. Relative roles of race versus socioeconomic position in studies of health inequalities: a matter of interpretation. *Annual review of public health*, 39: 169–188, 2018.

[42] National Academies of Sciences Engineering, Medicine, et al. Metrics that matter for population health action: workshop summary. 2016.

[43] Laurent Orseau and Stuart Armstrong. Safely interruptible agents. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, pages 557–566, 2016.

[44] J Pearl. Aspects of graphical methods connected with causality. *Proceedings of the 49th Session of the International Statistical Institute*, 1993.

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

[46] Judea Pearl. *Causality*. Cambridge university press, 2009.

[47] Amy Richardson, Michael G Hudgens, Peter B Gilbert, and Jason P Fine. Nonparametric bounds and sensitivity analysis of treatment effects. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 29(4):596, 2014.

[48] James M Robins. The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS*, pages 113–159, 1989.

[49] James M Robins. Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality*, pages 69–117. Springer, 1997.

[50] J.M. Robins. The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. In L. Sechrest, H. Freeman, and A. Mulley, editors, *Health Service Research Methodology: A Focus on AIDS*, pages 113–159. NCHSR, U.S. Public Health Service, Washington, D.C., 1989.

[51] Paul R Rosenbaum. Sensitivity analysis in observational studies. *Encyclopedia of statistics in behavioral science*, 2005.

[52] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[53] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[54] Peter AG Sandercock, Maciej Niewada, and Anna Członkowska. The international stroke trial database. *Trials*, 12(1):1–7, 2011.

[55] John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897. PMLR, 2015.

[56] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[57] Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the National Conference on Artificial Intelligence*. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[58] Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2000.

[59] Alex Strehl, John Langford, Lihong Li, and Sham M Kakade. Learning from logged implicit exploration data. In *Advances in Neural Information Processing Systems*, pages 2217–2225, 2010.

[60] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[61] Adith Swaminathan and Thorsten Joachims. Counterfactual risk minimization: Learning from logged bandit feedback. In *International Conference on Machine Learning*, pages 814–823. PMLR, 2015.

[62] Sonja A Swanson, Miguel A Hernán, Matthew Miller, James M Robins, and Thomas S Richardson. Partial identification of the average treatment effect using instrumental variables: review of methods for binary instruments, treatments, and outcomes. *Journal of the American Statistical Association*, 113(522):933–947, 2018.

[63] Philip Thomas, Georgios Theocharous, and Mohammad Ghavamzadeh. High confidence policy improvement. In *International Conference on Machine Learning*, pages 2380–2388. PMLR, 2015.

[64] J. Tian. *Studies in Causal Reasoning and Learning*. PhD thesis, Computer Science Department, University of California, Los Angeles, CA, November 2002.

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

[66] Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Aaai/iaai*, pages 567–573, 2002.

[67] David Todem, J Fine, and L Peng. A global sensitivity test for evaluating statistical hypotheses with nonidentifiable models. *Biometrics*, 66(2):558–566, 2010.

[68] Stijn Vansteelandt, Els Goetghebeur, Michael G Kenward, and Geert Molenberghs. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, pages 953–979, 2006.

[69] Junzhe Zhang and Elias Bareinboim. Bounding causal effects on continuous outcome. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

[70] Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. *arXiv preprint arXiv:2110.05690*, 2021.

[71] Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In *International Conference on Machine Learning*, pages 26548–26558. PMLR, 2022.

Appendix

A Related Work

Our work builds upon the literature on partial identification of causal effects, sensitivity analysis, and safe policy learning from offline data.

Partial Identification and Sensitivity Analysis. Seminal work of Manski [36], Imbens and Manski [25], Tian and Pearl [65], Robins [48], Balke and Pearl [2], Manski and Pepper [37], Swanson et al. [62], Fan and Park [14] developed the first bounds on causal effects in non-identifiable settings using observational data in the standard backdoor graph [44] and the instrumental variable setting [24] respectively. More recently, the bounds have been improved for applicability to continuous outcomes [69], and to establish a general framework for estimating bounds on interventional and counterfactual effects [70, 71]. While Zhang et al. [71] develop informative bounds using both observational and experimental data, they focus on general counterfactual queries by discretizing the exogenous latent space, formulating bounds as polynomial programs over this discretization and a Bayesian framework to approximately estimate bounds using MCMC. Sensitivity analysis attempts to provide intervals on causal effects by assuming the level of confounding, for example, via models such as Marginal Sensitivity analysis, which considers deviations in the propensity score in relation to the estimated propensity [51, 26, 47, 67, 68, 29]. Other approaches explore the linearity of structural functions [11]. Our work instead focuses on estimating closed-form bounds using observational and experimental data sources that are able to account for unobserved confounders and mismatched contexts.

Safe Policy Learning. Safety in reinforcement learning is an overloaded term [18], as it may be considered with respect to one of the following: uncertainty in the parametrization of the environment [63, 19], additional constraints on the optimal policy [1, 8], the ability to interrupt early before potential risks could occur [43, 13], or as exploration in a hazardous environment [55, 56, 16]. This paper focuses on the first approach, where the goal is to learn a policy from a fixed dataset that could achieve the performance of a behavioral policy, called *baseline*, that generates the observational data. Closet to our work, Kallus and Zhou [27] studied the problem of confounding-robust policy improvement that optimizes a policy to achieve the best worst-case improvement relative to a baseline treatment assignment policy. The agent is assumed to access the observational data and a sensitivity parameter describing the strength of the unobserved confounders' influence on the treatment. Our work does not invoke this untestable parametric assumption but instead requires the domains of observed variables to be computable and finite. Our work also accounts for additional data collected from controlled experiments and explores non-trivial interaction between the observational and experimental data on the bounds over the treatment effects.

B Proofs of Partial Identification

This section provides proofs for the theoretical results provided in the paper. Our proofs rely on the inference system based on the counterfactual distributions. For a set of variables $\mathbf{X}, \dots \mathbf{W}, \mathbf{Y}, \dots, \mathbf{Z}$, the counterfactual distribution $P(\mathbf{Y}_{\mathbf{x}}, \dots, , \mathbf{Z}_{\mathbf{w}})$ is a joint distribution over potential outcomes $\mathbf{Y}_{\mathbf{x}}, \dots, , \mathbf{Z}_{\mathbf{w}}$ in SCM $\mathcal{M}$ [6, Def. 7], given by

$$P(\mathbf{y}_{\mathbf{x}}, \dots, \mathbf{z}_{\mathbf{w}}) = \sum_{\mathbf{u}} \mathbf{1}_{\mathbf{Y}_{\mathbf{x}}(\mathbf{u})=\mathbf{y}, \dots, \mathbf{Z}_{\mathbf{w}}(\mathbf{u})=\mathbf{z}} P(\mathbf{u}) \tag{27}$$

Fix a value $\mathbf{y} \in \Omega_{\mathbf{Y}}$. We will consistently use $\mathbf{y}_{\mathbf{x}}$ denote an event $\mathbf{Y}_{\mathbf{x}} = \mathbf{y}$, and $\neg\mathbf{y}_{\mathbf{x}}$ for $\mathbf{Y}_{\mathbf{x}} \neq \mathbf{y}$. The counterfactual probability for a joint collection of events $\mathbf{Y}_{\mathbf{x}} \neq \mathbf{y}, \dots, \mathbf{Z}_{\mathbf{w}}(\mathbf{u})\neg\mathbf{z}$ evaluated in SCM $\mathcal{M}$ is given by

$$P(\neg\mathbf{y}_{\mathbf{x}}, \dots, \neg\mathbf{z}_{\mathbf{w}}) = \sum_{\mathbf{u}} \mathbf{1}_{\mathbf{Y}_{\mathbf{x}}(\mathbf{u})\neg\mathbf{y}, \dots, \mathbf{Z}_{\mathbf{w}}(\mathbf{u})\neg\mathbf{z}} P(\mathbf{u}) \tag{28}$$

Counterfactual variables follow three properties of composition, effectiveness, and reversibility, which hold in all structural causal models.

Theorem 3 (Counterfactual Axioms)**.** *Let $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P \rangle$ be an SCM. For any three sets of variables $\mathbf{X}, \mathbf{Y}, \mathbf{W} \subseteq \mathbf{V}$, the following properties hold*

1. ***Composition.*** $\mathbf{W}_{\mathbf{x}}(\mathbf{u}) = \mathbf{w} \Rightarrow \mathbf{Y}_{\mathbf{x},\mathbf{w}}(\mathbf{u}) = \mathbf{Y}_{\mathbf{x}}(\mathbf{u})$*;*
2. ***Effectiveness.*** $\mathbf{X}_{\mathbf{x},\mathbf{w}}(\mathbf{u}) = \mathbf{x}$*;*
3. ***Reversibility*** $\{\mathbf{Y}_{\mathbf{x},\mathbf{w}}(\mathbf{u}) = \mathbf{y}\} \,\&\, \{\mathbf{W}_{\mathbf{x},\mathbf{y}}(\mathbf{u}) = \mathbf{w}\} \Rightarrow \mathbf{Y}_{\mathbf{x}}(\mathbf{u}) = \mathbf{y}$.

B.1 Proofs of the Closed-form Bounds

We will first provide proofs for the closed-form bounds over the causal effect $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ given in Sec. 4, taking different combinations of data sources.

Lemma 1 (Obs$\,+\,$Exp($\mathbf{C}_1$))**.** *Given distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and $P_X(Y, \mathbf{C}_1)$, the lower bound over $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ for all $(x, y, \mathbf{c}_1, \mathbf{c}_2) \in \Omega_X \times \Omega_Y \times \Omega_{\mathbf{C}_1} \times \Omega_{\mathbf{C}_2}$ is given by*

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2) \geq \max\left\{l_1(x, y, \mathbf{c}_1, \mathbf{c}_2), l_2(x, y, \mathbf{c}_1, \mathbf{c}_2)\right\} \tag{7}$$

where l_1, l_2 are functions defined as

$$\begin{aligned} l_1(x, y, \mathbf{c}_1, \mathbf{c}_2) &= P(x, y, \mathbf{c}_1, \mathbf{c}_2) \qquad (8)\\ l_2(x, y, \mathbf{c}_1, \mathbf{c}_2) &= P_x(y, \mathbf{c}_1) - P(x, y, \mathbf{c}_1, \neg\mathbf{c}_2) - P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \qquad (9) \end{aligned}$$

Proof. Observe that in the contextual bandit model described in Fig. 1 (a), action X has not causal effects on the covariates $\mathbf{C}_1, \mathbf{C}_2$. We must have that given any $\mathbf{U} = \mathbf{u}$, the potential outcomes $\mathbf{C}_{1x}(\mathbf{u}) = \mathbf{C}_1(\mathbf{u})$ and $\mathbf{C}_{2x}(\mathbf{u}) = \mathbf{C}_1(\mathbf{u})$. By the definition of counterfactual and interventional distributions, the causal effect $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ could be written as a particular counterfactual probability,

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2) = P(y_x, \mathbf{c}_1, \mathbf{c}_2) \tag{29}$$

It is thus sufficient to bound the counterfactual probability $P(y_x, \mathbf{c}_1, \mathbf{c}_2)$. We will next discuss different bounding strategy based on the available distributions.

Case I: Obs($\mathbf{C}_1, \mathbf{C}_2$). The first lower bound is obtained when only observational distribution in the form of $P(Y, X, \mathbf{C}_1, \mathbf{C}_2)$ is available. By summing over the domain of the observed action X,

$$\begin{aligned} P(y_x, \mathbf{c}_1, \mathbf{c}_2) &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2) + P(y_x, \neg, x, \mathbf{c}_1, \mathbf{c}_2) \qquad (30)\\ &\geq P(y_x, x, \mathbf{c}_1, \mathbf{c}_2) \qquad (31) \end{aligned}$$

The last step holds since probability $P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2) \geq 0$. By the composition axiom (Thm. 3), $Y_x(\mathbf{u}) = Y(\mathbf{u})$ if $X(\mathbf{u}) = x$. The above equation could be written as

$$P(y_x, \mathbf{c}_1, \mathbf{c}_2) \geq P(y, x, \mathbf{c}_1, \mathbf{c}_2) \tag{32}$$

Among quantities in the above equation, the right-hand side is an observational quantity, which is also referred to as the *natural bound* in the literature [36, 48].

Case II: Obs($\mathbf{C}_1, \mathbf{C}_2$) + Exp($\mathbf{C}_1$). When the interventional distribution $P_X(Y, \mathbf{C}_1)$ is also available, applying basic probabilistic operations gives,

$$\begin{aligned} P(y_x, \mathbf{c}_1, \mathbf{c}_2) &= P(y_x, \mathbf{c}_1) - P(y_x, \mathbf{c}_1, \neg\mathbf{c}_2) \qquad (33)\\ &= P(y_x, \mathbf{c}_1) - P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2) - P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \qquad (34)\\ &= P(y_x, \mathbf{c}_1) - P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2) - P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \qquad (35) \end{aligned}$$

The last step holds since the marginal probability $P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \geq P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2)$ is lower bounded by its joint probability. Again, by the composition axiom (Thm. 3), $Y_x(\mathbf{u}) = Y(\mathbf{u})$ if $X(\mathbf{u}) = x$. The above equation could be written as

$$P(y_x, \mathbf{c}_1, \mathbf{c}_2) \geq P(y_x, \mathbf{c}_1) - P(y, x, \mathbf{c}_1, \neg\mathbf{c}_2) - P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \tag{36}$$

It follows from Eq. (29) that the above equation could be further written as

$$P(y_x, \mathbf{c}_1, \mathbf{c}_2) \geq P_x(y, \mathbf{c}_1) - P(y, x, \mathbf{c}_1, \neg\mathbf{c}_2) - P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \tag{37}$$

where the right-hand side is a function of the observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional $P_X(Y, \mathbf{C}_1)$ distributions. □

Lemma 2 (Exp($\mathbf{C}_1$)$\,+\,$Exp($\mathbf{C}_2$))**.** *Given distributions $P_X(Y, \mathbf{C}_1)$ and $P_X(Y, \mathbf{C}_2)$, the lower bound over $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ for all $(x, y, \mathbf{c}_1, \mathbf{c}_2) \in \Omega_X \times \Omega_Y \times \Omega_{\mathbf{C}_1} \times \Omega_{\mathbf{C}_2}$ is given by*

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2) \geq P_x(y, \mathbf{c}_1) - P_x(y, \neg\mathbf{c}_2) \tag{14}$$

Proof. Given both interventional distributions $P_X(Y, \mathbf{C}_1)$ and $P_X(Y, \mathbf{C}_2)$, applying Eq. (33) gives

$$P(y_x, \mathbf{c}_1, \mathbf{c}_2) = P(y_x, \mathbf{c}_1) - P(y_x, \mathbf{c}_1, \neg\mathbf{c}_2) \tag{38}$$

$$\geq P(y_x, \mathbf{c}_1) - P(y_x, \neg\mathbf{c}_2) \tag{39}$$

The last step holds since the marginal probability $P(y_x, \neg\mathbf{c}_2) \geq P(y_x, \mathbf{c}_1, \neg\mathbf{c}_2)$ is lower bounded by the joint probability. It follows from Eq. (29) that the above equation could be further written as

$$P(y_x, \mathbf{c}_1, \mathbf{c}_2) \geq P_x(y, \mathbf{c}_1) - P(y_x, \neg\mathbf{c}_2) \tag{40}$$

where the right-hand side is a function of interventional distributions $P_X(Y, \mathbf{C}_1)$ and $P_X(Y, \mathbf{C}_2)$. □

Theorem 1 (Obs + Exp($\mathbf{C}_1$) + Exp($\mathbf{C}_2$))**.** *Given distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, and $P_X(Y, \mathbf{C}_2)$, the lower bound over $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ for all $(x, y, \mathbf{c}_1, \mathbf{c}_2) \in \Omega_X \times \Omega_Y \times \Omega_{\mathbf{C}_1} \times \Omega_{\mathbf{C}_2}$ is*

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2) \geq \max\{l_1(x, y, \mathbf{c}_1, \mathbf{c}_2), l_2(x, y, \mathbf{c}_1, \mathbf{c}_2), l_3(x, y, \mathbf{c}_1, \mathbf{c}_2), l_4(x, y, \mathbf{c}_1, \mathbf{c}_2)\} \tag{15}$$

where l_1, l_2 are given by Eqs. (8) *and* (9)*, respectively; l_3, l_4 are functions defined as*

$$l_3(x, y, \mathbf{c}_1, \mathbf{c}_2) = P_x(y, \mathbf{c}_2) - P(x, y, \neg\mathbf{c}_1, \mathbf{c}_2) - P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{16}$$

$$l_4(x, y, \mathbf{c}_1, \mathbf{c}_2) = P_x(y, \mathbf{c}_1) - P_x(y, \neg\mathbf{c}_2) + P(x, y, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{17}$$

Proof. The first two bounds l_1, l_2 follow from Lem. 1. We will next focus on the last two cases.

Case III: Obs($\mathbf{C}_1, \mathbf{C}_2$) + Exp($\mathbf{C}_2$). The bounding strategy in this case is analogous to Case II. By applying basic probabilistic operations, we have

$$P(y_x, \mathbf{c}_1, \mathbf{c}_2) = P(y_x, \mathbf{c}_2) - P(y_x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{41}$$

$$= P(y_x, \mathbf{c}_2) - P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2) - P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{42}$$

$$= P(y_x, \mathbf{c}_2) - P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2) - P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{43}$$

$$= P(y_x, \mathbf{c}_2) - P(y, x, \neg\mathbf{c}_1, \mathbf{c}_2) - P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{44}$$

Eq. (43) holds since the marginal probability $P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \geq P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2)$ is lower bounded by its joint probability. Eq. (44) follows from the composition axiom (Thm. 3), i.e., $Y_x(\mathbf{u}) = Y(\mathbf{u})$ if $X(\mathbf{u}) = x$. Applying Eq. (29) gives

$$P(y_x, \mathbf{c}_1, \mathbf{c}_2) \geq P_x(y, \mathbf{c}_2) - P(y, x, \neg\mathbf{c}_1, \mathbf{c}_2) - P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{45}$$

Among quantities in the above equation, the right-hand side is a function of the observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional $P_X(Y, \mathbf{C}_2)$ distributions.

Case IV: Obs($\mathbf{C}_1, \mathbf{C}_2$) + Exp($\mathbf{C}_1$) + Exp($\mathbf{C}_2$). For the last case, all observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional $P_X(Y, \mathbf{C}_1)$ and $P_X(Y, \mathbf{C}_2)$ are available. By a telescoping sum, we have

$$P(y_x, \mathbf{c}_1, \mathbf{c}_2) = P(y_x, \mathbf{c}_1, \mathbf{c}_2) - P(y_x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) + P(y_x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{46}$$

$$\geq P(y_x, \mathbf{c}_2) - P(y_x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) + P(y_x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{47}$$

$$\geq P(y_x, \mathbf{c}_2) - P(y_x, \neg\mathbf{c}_2) + P(y_x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{48}$$

$$\geq P(y_x, \mathbf{c}_2) - P(y_x, \neg\mathbf{c}_2) + P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{49}$$

The last three steps holds due to the following inequality relationships respectively,

$$P(y_x, \mathbf{c}_2) \geq P(y_x, \mathbf{c}_1, \mathbf{c}_2) \tag{50}$$

$$P(y_x, \neg\mathbf{c}_2) \geq P(y_x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{51}$$

$$P(y_x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \geq P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{52}$$

By the composition axiom, $Y_x(\mathbf{u}) = Y(\mathbf{u})$ if $X(\mathbf{u}) = x$. Eq. (49) could be further written as

$$P(y_x, \mathbf{c}_1, \mathbf{c}_2) \geq P(y_x, \mathbf{c}_2) - P(y_x, \neg\mathbf{c}_2) + P(y, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{53}$$

$$\geq P_x(y, \mathbf{c}_2) - P_x(y, \neg\mathbf{c}_2) + P(y, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{54}$$

The last step follows from Eq. (29). Among quantities in the above equation, the right-hand side is a function of all input distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$ and $P_X(Y, \mathbf{C}_2)$. □

B.2 Proof of the Sharpness Gaurantee

This section will provide the proof for the sharpness of the lower bound given by Thm. 1. We will utilize a special parametric family of SCMs compatible with the causal diagram of Fig. 1 (a).

Definition 4 (Canonical Contextual Bandit)**.** A canonical contextual bandit model (for short, CCB) $\mathcal{M}$ is an SCM $\langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P \rangle$ compatible with the causal diagram in Fig. 1 (a) such that

(i) $\mathbf{V} = \{X, Y, \mathbf{C}_1, \mathbf{C}_2\}$ and $\mathbf{U} = \{X, \mathbf{Y}_*, \mathbf{C}_1, \mathbf{C}_2\}$;

(ii) $\mathbf{Y}_*$ is a vector of counterfactuals (i.e., potential outcomes) $(Y_x \mid \forall x \in \Omega_X)$;

(iii) Values of $X, \mathbf{C}_1, \mathbf{C}_2$ are determined by the corresponding variables in the exogenous $\mathbf{U}$;

(iv) Values of Y are determined by the counterfactual variable $Y_x \in \mathbf{Y}_*$ indexed by input x,

$$y \leftarrow f_Y(x, \mathbf{y}_*) = y_x; \tag{55}$$

(v) The exogenous distribution $P(X, \mathbf{C}_1, \mathbf{C}_2, \mathbf{Y}_*)$ decomposes as follows

$$P(x, \mathbf{c}_1, \mathbf{c}_2, \mathbf{y}_*) = P(x, \mathbf{c}_1, \mathbf{c}_2) \prod_{x' \in \Omega_X} P(y_{x'} \mid x, \mathbf{c}_1, \mathbf{c}_2) \tag{56}$$

For a general SCM graphically described in Fig. 1 (a), its structural functions $\mathcal{F}$ and the exogenous distribution $P(\mathbf{U})$ are not well-specified, and could take arbitrary forms. On the other hand, for a canonical contextual bandit model described in Def. 4, its structural functions $\mathcal{F}$ are fixed; its exogenous domains over $\mathbf{U}$ are discrete and finite, determined by the cardinality of the observed domains over $\mathbf{V}$. Moreover, its exogenous distribution $P(\mathbf{U})$ follows the independence relationships implied by the factorization in Eq. (56). Perhaps surprisingly, the parametric family of canonical bandits is sufficient in representing all possible observational distribution $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional distributions $P_X(Y, \mathbf{C}_1)$, $P_X(Y, \mathbf{C}_2)$ in the causal diagram of Fig. 1 (a).

Lemma 3. *For any SCM $\mathcal{M}$ compatible with the causal diagram in Fig. 1 (a), there is a CCB $\mathcal{N}$ such that $P(X, Y, \mathbf{C}_1, \mathbf{C}_2; \mathcal{M}) = P(X, Y, \mathbf{C}_1, \mathbf{C}_2; \mathcal{N})$, $P_X(Y, \mathbf{C}_1; \mathcal{M}) = P_X(Y, \mathbf{C}_1; \mathcal{N})$ and $P_X(Y, \mathbf{C}_2; \mathcal{M}) = P_X(Y, \mathbf{C}_2; \mathcal{N})$.*

Proof. By the composition axiom (Thm. 3), $Y_x(\mathbf{u}) = Y(\mathbf{u})$ if $X(\mathbf{u}) = x$. This implies that

$$P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{57}$$

By summing over domains of $\mathbf{C}_2$,

$$P_x(y, \mathbf{c}_1; \mathcal{M}) = \sum_{\mathbf{c}_2} P_x(y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = \sum_{\mathbf{c}_2} P(y_x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{58}$$

The last step follows from Eq. (29). By further summing over domains of observed action X,

$$P_x(y, \mathbf{c}_1; \mathcal{M}) = \sum_{x', \mathbf{c}_2} P(y_x, x', \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{59}$$

Following a similar procedure, we also have

$$P_x(y, \mathbf{c}_2; \mathcal{M}) = \sum_{x', \mathbf{c}_1} P(y_x, x', \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{60}$$

Eqs. (57), (59) and (60) imply that the observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional $P_X(Y, \mathbf{C}_1)$, $P_X(Y, \mathbf{C}_2)$ distributions in SCM $\mathcal{M}$ could be written as functions of a counterfactual distribution of the form $P(Y_x, X, \mathbf{C}_1, \mathbf{C}_2)$. It is thus sufficient to simulate this counterfactual distribution. Precisely, let the exogenous distribution in CCB $\mathcal{N}$ be defined as:

$$P(X, \mathbf{C}_1, \mathbf{C}_2; \mathcal{N}) = P(X, \mathbf{C}_1, \mathbf{C}_2; \mathcal{M}) \tag{61}$$

$$\forall x \in \Omega_X, P(Y_x \mid X, \mathbf{C}_1, \mathbf{C}_2; \mathcal{N}) = P(Y_x \mid X, \mathbf{C}_1, \mathbf{C}_2; \mathcal{M}) \tag{62}$$

We will next show that the CCB $\mathcal{N}$ constructed above generate the same observational and interventional distributions as the SCM $\mathcal{M}$.

Consistency in Obs($\mathbf{C}_1$, $\mathbf{C}_2$). By the definition of CCB (Def. 4), the observed reward Y is determined by the counterfactual Y_x indexed by the input action $X = x$ (Eq. (55)). This implies

$$\begin{aligned} P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) && (63)\\ &= P(y_x \mid x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) P(x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) && (64)\\ &= P(y_x \mid x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) P(x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) && (65)\\ &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) && (66) \end{aligned}$$

Eq. (65) follows from the construction of the CCB $\mathcal{N}$ in Eqs. (61) and (62). Applying Eq. (57) gives

$$P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) = P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{67}$$

Consistency in Exp($\mathbf{C}_1$) + Exp($\mathbf{C}_2$). Again, note that in a CCB $\mathcal{N}$, the observed reward Y is determined by the counterfactual Y_x indexed by the input action $X = x$ (Eq. (55)). This gives

$$\begin{aligned} P_x(y, \mathbf{c}_1; \mathcal{N}) &= \sum_{x', \mathbf{c}_1} P(y_x, x', \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) && (68)\\ &= \sum_{x', \mathbf{c}_1} P(y_x \mid x', \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) P(x', \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) && (69)\\ &= \sum_{x', \mathbf{c}_1} P(y_x \mid x', \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) P(x', \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) && (70)\\ &= \sum_{x', \mathbf{c}_1} P(y_x, x', \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) && (71) \end{aligned}$$

Eq. (70) follows from the construction of the CCB $\mathcal{N}$ in Eqs. (61) and (62). Applying Eq. (59) gives

$$P_x(y, \mathbf{c}_1; \mathcal{N}) = P_x(y, \mathbf{c}_1; \mathcal{M}) \tag{72}$$

Analogously, we also have

$$P_x(y, \mathbf{c}_2; \mathcal{N}) = P_x(y, \mathbf{c}_2; \mathcal{M}) \tag{73}$$

This means that SCM $\mathcal{M}$ and CCB $\mathcal{N}$ give the same evaluation of the observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional $P_X(Y, \mathbf{C}_1)$, $P_X(Y, \mathbf{C}_2)$ distributions, which completes the proof. □

Our next result identifies a set of extreme points on the exogenous probabilities of a CCB such that its evaluation of the target effect $P_x(y, \mathbf{c}_1, \mathbf{c}_1)$ matches the lower bound given by Thm. 1.

Lemma 4. *Let $\mathcal{M}$ be a CCB. Fix a realization $(x, y, \mathbf{c}_1, \mathbf{c}_2) \in \Omega_X \times \Omega_Y \times \Omega_{\mathbf{C}_1} \times \Omega_{\mathbf{C}_2}$. If its exogenous distribution $P(X, \mathbf{Y}_*, \mathbf{C}_1, \mathbf{C}_2)$ satisfies one of the following conditions, ,*

$$\begin{aligned} &P(y_x \mid \neg x, \mathbf{c}_1, \mathbf{c}_2) = 0, && P(y_x \mid \neg x, \mathbf{c}_1, \neg\mathbf{c}_2) = 1, && (74)\\ &P(y_x \mid \neg x, \neg\mathbf{c}_1, \mathbf{c}_2) = 1, && P(y_x \mid \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) = 0 && (75) \end{aligned}$$

then in this SCM $\mathcal{M}$, the causal effect $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ is equal to the lower bound $l(x, y, \mathbf{c}_1, \mathbf{c}_2)$ given by Thm. 1, i.e., $P_x(y, \mathbf{c}_1, \mathbf{c}_2) = l(x, y, \mathbf{c}_1, \mathbf{c}_2)$.

Proof. We will first evaluate the causal effect $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ in a CCB $\mathcal{M}$ at four extreme points of Eqs. (74) and (75), showing that they match the lower bound l_1, l_2, l_3, l_4 given by Thm. 1, respectively.

Case 1: $P(y_x \mid \neg x, \mathbf{c}_1, \mathbf{c}_2) = 0$**.** By summing over the domain of action X,

$$\begin{aligned} P(y_x, \mathbf{c}_1, \mathbf{c}_2) &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2) && (76)\\ &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2) && (77)\\ &= P(y, x, \mathbf{c}_1, \mathbf{c}_2) && (78)\\ &= l_1(x, y, \mathbf{c}_1, \mathbf{c}_2) && (79) \end{aligned}$$

The second step follows from the condition $P(y_x \mid \neg x, \mathbf{c}_1, \mathbf{c}_2) = 0$. The third step follows from the composition axiom (Thm. 3): $Y_x(\mathbf{u}) = Y(\mathbf{u})$ if $X(\mathbf{u}) = x$.

Case 2: $P(y_x \mid \neg x, \mathbf{c}_1, \neg\mathbf{c}_2) = 1$**.** By a telescoping sum, we have

$$\begin{aligned}
P(y_x, \mathbf{c}_1, \mathbf{c}_2) &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2) \tag{80}\\
&\quad + P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \tag{81}\\
&\quad - P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2) - P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \tag{82}\\
&= P(y_x, \mathbf{c}_1) - P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2) - P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \tag{83}\\
&= P(y_x, \mathbf{c}_1) - P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2) - P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \tag{84}\\
&= P(y_x, \mathbf{c}_1) - P(x, y, \mathbf{c}_1, \neg\mathbf{c}_2) - P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \tag{85}\\
&= P_x(y, \mathbf{c}_1) - P(x, y, \mathbf{c}_1, \neg\mathbf{c}_2) - P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2) \tag{86}\\
&= l_2(x, y, \mathbf{c}_1, \mathbf{c}_2) \tag{87}
\end{aligned}$$

The third step follows from the condition $P(y_x \mid \neg x, \mathbf{c}_1, \neg\mathbf{c}_2) = 1$. The fourth step follows from the composition axiom (Thm. 3): $Y_x(\mathbf{u}) = Y(\mathbf{u})$ if $X(\mathbf{u}) = x$. The fifth step follows from Eq. (29).

Case 3: $P(y_x \mid \neg x, \neg\mathbf{c}_1, \mathbf{c}_2) = 1$**.** By a telescoping sum, we have

$$\begin{aligned}
P(y_x, \mathbf{c}_1, \mathbf{c}_2) &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2) \tag{88}\\
&\quad + P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2) + P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{89}\\
&\quad - P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2) - P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{90}\\
&= P(y_x, \mathbf{c}_2) - P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2) - P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{91}\\
&= P(y_x, \mathbf{c}_2) - P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2) - P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{92}\\
&= P(y_x, \mathbf{c}_2) - P(x, y, \neg\mathbf{c}_1, \mathbf{c}_2) - P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{93}\\
&= P_x(y, \mathbf{c}_2) - P(x, y, \neg\mathbf{c}_1, \mathbf{c}_2) - P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2) \tag{94}\\
&= l_3(x, y, \mathbf{c}_1, \mathbf{c}_2) \tag{95}
\end{aligned}$$

The third step follows from the condition $P(y_x \mid \neg x, \neg\mathbf{c}_1, \mathbf{c}_2) = 1$. The fourth step follows from the composition axiom (Thm. 3): $Y_x(\mathbf{u}) = Y(\mathbf{u})$ if $X(\mathbf{u}) = x$. The fifth step follows from Eq. (29).

Case 4: $P(y_x \mid \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) = 0$**.** By a telescoping sum, we have

$$\begin{aligned}
P(y_x, \mathbf{c}_1, \mathbf{c}_2) &= \begin{aligned}[t] &P(y_x, x, \mathbf{c}_1, \mathbf{c}_2) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2)\\
&+ P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2)\\
&- P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2) - P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2)\\
&- P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) - P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2)\\
&+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2)\end{aligned} \tag{96}\\
&= \begin{aligned}[t] &P(y_x, \mathbf{c}_1) - P(y_x, \neg\mathbf{c}_2)\\
&+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2)\end{aligned} \tag{97}\\
&= P(y_x, \mathbf{c}_1) - P(y_x, \neg\mathbf{c}_2) + P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{98}\\
&= P(y_x, \mathbf{c}_1) - P(y_x, \neg\mathbf{c}_2) + P(x, y, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{99}\\
&= P_x(y, \mathbf{c}_1) - P_x(y, \neg\mathbf{c}_2) + P(x, y, \neg\mathbf{c}_1, \neg\mathbf{c}_2) \tag{100}\\
&= l_4(x, y, \mathbf{c}_1, \mathbf{c}_2) \tag{101}
\end{aligned}$$

The third step follows from the condition $P(y_x \mid \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2) = 0$. The fourth step follows from the composition axiom : $Y_x(\mathbf{u}) = Y(\mathbf{u})$ if $X(\mathbf{u}) = x$. The fifth step follows from Eq. (29).

At all the above extreme points, the causal effect $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ matches one of the lower bound l_i, $i = 1, \dots, 4$, given by Thm. 1. Since Thm. 1 applies to all SCMs compatible with the causal diagram of Fig. 1 (a), we must have in each of the above cases $i = 1, \dots, 4$, for any $j \neq i$,

$$l_i(x, y, \mathbf{c}_1, \mathbf{c}_2) \geq l_j(x, y, \mathbf{c}_1, \mathbf{c}_2) \tag{102}$$

This implies for all cases considered above,

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2) = l_i(x, y, \mathbf{c}_1, \mathbf{c}_2) = \max_{j=1,\dots,4} l_j(x, y, \mathbf{c}_1, \mathbf{c}_2) \tag{103}$$

In words, the causal effect $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ evaluated in CCB $\mathcal{M}$ at extreme points of Eqs. (74) and (75) matches the lower bound $l(x, y, \mathbf{c}_1, \mathbf{c}_2)$ given by Thm. 1, which completes the proof. □

A more interesting challenge at this point is to construct a canonical contextual bandit model from an SCM compatible with Fig. 1 such that its exogenous probabilities reach the extreme points given by Lem. 4 while maintaining the same observational and interventional distributions.

Lemma 5. *For any SCM $\mathcal{M}$ compatible with the causal diagram of Fig. 1 (a), fix a realization $(x, y, \mathbf{c}_1, \mathbf{c}_2) \in \Omega_X \times \Omega_Y \times \Omega_{\mathbf{C}_1} \times \Omega_{\mathbf{C}_2}$. There is an CCB $\mathcal{N}$ such that*

(i) $\mathcal{M}$ and $\mathcal{N}$ define the same observational distribution $P(X, Y, \mathbf{C}_1, \mathbf{C}_1)$ and interventional probabilities $P_x(y, \mathbf{c}_1)$, $P_x(y, \neg\mathbf{c}_1)$, $P_x(y, \mathbf{c}_2)$ and $P_x(y, \neg\mathbf{c}_2)$;

(ii) The exogenous distribution $P(X, Y_x, \mathbf{C}_1, \mathbf{C}_2)$ in $\mathcal{N}$ satisfy the following

$$\begin{cases} P(y_x \mid \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) = 0 & \textit{if } l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_1(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ P(y_x \mid \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) = 1 & \textit{if } l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_2(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ P(y_x \mid \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) = 1 & \textit{if } l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_3(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ P(y_x \mid \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) = 0 & \textit{if } l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_4(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{cases} \tag{104}$$

where $l = \max\{l_1, l_2, l_3, l_4\}$ are lower bounds over $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ given by Thm. 1.

Proof. It follows from Lem. 3 that it is sufficient to consider a CCB model $\mathcal{M}$ without loss of generality. We will next discuss the construction of an alternative CCB $\mathcal{N}$ on a case-by-case basis.

Case 1: $l_1 \geq l_2, l_3, l_4$. Construct a CCB $\mathcal{N}$ from $\mathcal{M}$ such that its exogenous distribution satisfies

$$P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) = 0 \tag{105}$$

$$\begin{aligned} P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) = {} & P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \\ & + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{106}$$

$$\begin{aligned} P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) = {} & P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ & + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{107}$$

$$\begin{aligned} P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) = {} & P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \\ & - P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{108}$$

Other exogenous probabilities $P(y_x, x, \mathbf{c}_1, \mathbf{c}_2)$ remain the same across $\mathcal{N}$ and $\mathcal{M}$. The above construction is feasible since $l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_1(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})$ implies

$$l_1(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_2(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{109}$$

$$\begin{aligned} \Rightarrow P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \leq \\ P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{110}$$

$$l_1(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_3(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{111}$$

$$\begin{aligned} \Rightarrow P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \leq \\ P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{112}$$

$$l_1(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_4(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{113}$$

$$\Rightarrow P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) - P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq 0 \tag{114}$$

We will next show that this construction of $\mathcal{N}$ maintain the same observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional probabilities $P_x(y, \mathbf{c}_1)$, $P_x(y, \neg\mathbf{c}_1)$, $P_x(y, \mathbf{c}_2)$ and $P_x(y, \neg\mathbf{c}_2)$. First,

$$\begin{aligned} & P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ = & P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ = & P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ = & P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{115}$$

The last step follows from the definition of CCBs (Def. 4). As for the interventional distribution,

$$P_x(y, \mathbf{c}_1; \mathcal{N}) \tag{116}$$
$$= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \tag{117}$$
$$+ P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) \tag{118}$$
$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + 0 \\ &+ P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{119}$$
$$= P_x(y, \mathbf{c}_1; \mathcal{M}) \tag{120}$$

and

$$P_x(y, \neg\mathbf{c}_1; \mathcal{N}) \tag{121}$$
$$\begin{aligned} &= P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) \end{aligned} \tag{122}$$
$$\begin{aligned} &= P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \neg\mathcal{M}) \\ &- P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{123}$$
$$= P_x(y, \neg\mathbf{c}_1; \mathcal{M}) \tag{124}$$

Similarly, we also have

$$P_x(y, \mathbf{c}_2; \mathcal{N}) \tag{125}$$
$$= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \tag{126}$$
$$+ P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \tag{127}$$
$$= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + 0 \tag{128}$$
$$+ P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{129}$$
$$+ P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{130}$$
$$= P_x(y, \mathbf{c}_2; \mathcal{M}) \tag{131}$$

and

$$P_x(y, \neg\mathbf{c}_2; \mathcal{N}) \tag{132}$$
$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) \end{aligned} \tag{133}$$
$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) - P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{134}$$
$$= P_x(y, \neg\mathbf{c}_2; \mathcal{M}) \tag{135}$$

Case 2: $l_2 \geq l_1, l_3, l_4$**.** Construct a CCB $\mathcal{N}$ from $\mathcal{M}$ such that its exogenous distribution satisfies

$$P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \tag{136}$$
$$\begin{aligned} &= P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \\ &- P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{137}$$
$$P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) = P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \tag{138}$$
$$\begin{aligned} &P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &= P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) - P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \\ &+ P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{139}$$
$$\begin{aligned} &P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) \\ &= P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \\ &- P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{140}$$

Other exogenous probabilities $P(y_{x'}, x, \mathbf{c}_1, \mathbf{c}_2)$ remain the same across $\mathcal{N}$ and $\mathcal{M}$. The above construction is feasible since $l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_2(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})$ implies

$$l_2(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_1(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{141}$$

$$\begin{aligned} &\Rightarrow P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \\ &- P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \geq 0 \end{aligned} \tag{142}$$

$$l_2(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_3(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{143}$$

$$\begin{aligned} &\Rightarrow P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) - P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \\ &+ P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \leq P(\neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{144}$$

$$l_2(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_4(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{145}$$

$$\begin{aligned} &\Rightarrow P(y_x, \neg x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \\ &- P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \geq 0 \end{aligned} \tag{146}$$

We will next show that this construction of $\mathcal{N}$ maintain the same observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional probabilities $P_x(y, \mathbf{c}_1)$, $P_x(y, \neg \mathbf{c}_1)$, $P_x(y, \mathbf{c}_2)$ and $P_x(y, \neg \mathbf{c}_2)$. First,

$$\begin{aligned} &P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &= P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{147}$$

The last step follows from the definition of CCBs (Def. 4). As for the interventional distribution,

$$P_x(y, \mathbf{c}_1; \mathcal{N}) \tag{148}$$

$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N}) \end{aligned} \tag{149}$$

$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) - P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) + P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{150}$$

$$= P_x(y, \mathbf{c}_1; \mathcal{M}) \tag{151}$$

and

$$P_x(y, \neg \mathbf{c}_1; \mathcal{N}) \tag{152}$$

$$\begin{aligned} &= P(y_x, x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N}) \end{aligned} \tag{153}$$

$$\begin{aligned} &= P(y_x, x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &- P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) + P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) - P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{154}$$

$$= P_x(y, \neg \mathbf{c}_1; \mathcal{M}) \tag{155}$$

Similarly, we also have

$$P_x(y, \mathbf{c}_2; \mathcal{N}) \tag{156}$$

$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \end{aligned} \tag{157}$$

$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) - P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &- P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) + P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{158}$$

$$= P_x(y, \mathbf{c}_2; \mathcal{M}) \tag{159}$$

and

$$P_x(y, \neg \mathbf{c}_2; \mathcal{N}) \tag{160}$$

$$\begin{aligned}&= P(y_x, x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N})\end{aligned} \tag{161}$$

$$\begin{aligned}&= P(y_x, x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) + P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) - P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{162}$$

$$= P_x(y, \neg \mathbf{c}_2; \mathcal{M}) \tag{163}$$

Case 3: $l_3 \geq l_1, l_2, l_4$**.** Construct a CCB $\mathcal{N}$ from $\mathcal{M}$ such that its exogenous distribution satisfies

$$P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \tag{164}$$

$$\begin{aligned}&= P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &- P(\neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{165}$$

$$P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N}) \tag{166}$$

$$\begin{aligned}&= P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) - P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &+ P(\neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{167}$$

$$P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) = P(\neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{168}$$

$$P(y_x, \neg x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N}) \tag{169}$$

$$\begin{aligned}&= P(y_x, \neg x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &- P(\neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{170}$$

Other exogenous probabilities $P(y_{x'}, x, \mathbf{c}_1, \mathbf{c}_2)$ remain the same across $\mathcal{N}$ and $\mathcal{M}$. The above construction is feasible since $l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_3(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})$ implies

$$l_3(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_1(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{171}$$

$$\begin{aligned}&\Rightarrow P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &- P(\neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq 0\end{aligned} \tag{172}$$

$$l_3(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_2(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{173}$$

$$\begin{aligned}&\Rightarrow P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) - P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &+ P(\neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \leq P(\neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{174}$$

$$l_3(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_4(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{175}$$

$$\begin{aligned}&\Rightarrow P(y_x, \neg x, \neg \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &- P(\neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq 0\end{aligned} \tag{176}$$

We will next show that this construction of $\mathcal{N}$ maintain the same observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional probabilities $P_x(y, \mathbf{c}_1)$, $P_x(y, \neg \mathbf{c}_1)$, $P_x(y, \mathbf{c}_2)$ and $P_x(y, \neg \mathbf{c}_2)$. First,

$$\begin{aligned}&P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &= P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{177}$$

The last step follows from the definition of CCBs (Def. 4). As for the interventional distribution,

$$P_x(y, \mathbf{c}_1; \mathcal{N}) \tag{178}$$

$$\begin{aligned}&= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{N})\end{aligned} \tag{179}$$

$$\begin{aligned}&= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) - P(\neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \neg \mathbf{c}_2; \mathcal{M}) \\ &- P(y_x, \neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(\neg x, \neg \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{180}$$

$$= P_x(y, \mathbf{c}_1; \mathcal{M}) \tag{181}$$

and

$$P_x(y, \neg\mathbf{c}_1; \mathcal{N}) \tag{182}$$

$$\begin{aligned}&= P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N})\\ &+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N})\end{aligned} \tag{183}$$

$$\begin{aligned}&= P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\\ &+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M})\\ &+ P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) - P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{184}$$

$$= P_x(y, \neg\mathbf{c}_1; \mathcal{M}) \tag{185}$$

Similarly, we also have

$$P_x(y, \mathbf{c}_2; \mathcal{N}) \tag{186}$$

$$\begin{aligned}&= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N})\\ &+ P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N})\end{aligned} \tag{187}$$

$$\begin{aligned}&= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\\ &+ P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) - P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\\ &+ P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{188}$$

$$= P_x(y, \mathbf{c}_2; \mathcal{M}) \tag{189}$$

and

$$P_x(y, \neg\mathbf{c}_2; \mathcal{N}) \tag{190}$$

$$\begin{aligned}&= P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N})\\ &+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N})\end{aligned} \tag{191}$$

$$\begin{aligned}&= P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M})\\ &- P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\\ &+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M})\\ &+ P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) - P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{192}$$

$$= P_x(y, \neg\mathbf{c}_2; \mathcal{M}) \tag{193}$$

Case 4: $l_4 \geq l_1, l_2, l_3$**.** Construct a CCB $\mathcal{N}$ from $\mathcal{M}$ such that its exogenous distribution satisfies

$$P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \tag{194}$$

$$= P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) - P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \tag{195}$$

$$P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) \tag{196}$$

$$= P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \tag{197}$$

$$P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \tag{198}$$

$$= P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \tag{199}$$

$$P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) = 0 \tag{200}$$

Other exogenous probabilities $P(y_{x'}, x, \mathbf{c}_1, \mathbf{c}_2)$ remain the same across $\mathcal{N}$ and $\mathcal{M}$. The above construction is feasible since $l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) = l_4(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})$ implies

$$l_4(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_1(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{201}$$

$$\Rightarrow P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) - P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \geq 0 \tag{202}$$

$$l_4(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_2(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{203}$$

$$\begin{aligned}&\Rightarrow P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M})\\ &\leq P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M})\end{aligned} \tag{204}$$

$$l_4(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \geq l_3(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \tag{205}$$

$$\begin{aligned}&\Rightarrow P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M})\\ &\leq P(\neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M})\end{aligned} \tag{206}$$

We will next show that this construction of $\mathcal{N}$ maintain the same observational $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$ and interventional probabilities $P_x(y, \mathbf{c}_1)$, $P_x(y, \neg\mathbf{c}_1)$, $P_x(y, \mathbf{c}_2)$ and $P_x(y, \neg\mathbf{c}_2)$. First,

$$\begin{aligned} &P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &= P(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{207}$$

The last step follows from the definition of CCBs (Def. 4). As for the interventional distribution,

$$P_x(y, \mathbf{c}_1; \mathcal{N}) \tag{208}$$

$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) \end{aligned} \tag{209}$$

$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &- P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{210}$$

$$= P_x(y, \mathbf{c}_1; \mathcal{M}) \tag{211}$$

and

$$P_x(y, \neg\mathbf{c}_1; \mathcal{N}) \tag{212}$$

$$\begin{aligned} &= P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) \end{aligned} \tag{213}$$

$$\begin{aligned} &= P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + 0 \end{aligned} \tag{214}$$

$$= P_x(y, \neg\mathbf{c}_1; \mathcal{M}) \tag{215}$$

Similarly, we also have

$$P_x(y, \mathbf{c}_2; \mathcal{N}) \tag{216}$$

$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{N}) \end{aligned} \tag{217}$$

$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &- P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \neg\mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{218}$$

$$= P_x(y, \mathbf{c}_2; \mathcal{M}) \tag{219}$$

and

$$P_x(y, \neg\mathbf{c}_2; \mathcal{N}) \tag{220}$$

$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) \\ &+ P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) + P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{N}) \end{aligned} \tag{221}$$

$$\begin{aligned} &= P(y_x, x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, \neg x, \mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) \\ &+ P(y_x, \neg x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + P(y_x, x, \neg\mathbf{c}_1, \neg\mathbf{c}_2; \mathcal{M}) + 0 \end{aligned} \tag{222}$$

$$= P_x(y, \neg\mathbf{c}_2; \mathcal{M}) \tag{223}$$

This completes the proof. □

Finally, we are now ready to prove the sharpness of the lower bound given by Thm. 1.

Theorem 2. *Given distributions $P(X, Y, \mathbf{C}_1, \mathbf{C}_2)$, $P_X(Y, \mathbf{C}_1)$, and $P_X(Y, \mathbf{C}_2)$, Thm. 1 is a sharp lower bound over the causal effects $P_X(Y, \mathbf{C}_1, \mathbf{C}_2)$ in the causal diagram of Fig. 1 (a).*

Proof. Suppose there is an SCM $\mathcal{M}$ where $l^*(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) > l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M})$ for some $(x, y, \mathbf{c}_1, \mathbf{c}_2)$. Construct an alternative SCM $\mathcal{M}^*$ following Lem. 5. It follows from Lem. 4 that in this modified SCM $\mathcal{M}^*$, the causal effect $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ matches the lower bound l given by Thm. 1,

$$P_x(y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) = l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) \tag{224}$$

Observe that the construction in Lem. 5 ensures that SCMs $\mathcal{M}$ and $\mathcal{M}^*$ share the same evaluation of the same observational distribution $P(X, Y, \mathbf{C}_1, \mathbf{C}_1)$ and interventional probabilities $P_x(y, \mathbf{c}_1)$, $P_x(y, \neg\mathbf{c}_1)$, $P_x(y, \mathbf{c}_2)$ and $P_x(y, \neg\mathbf{c}_2)$. Since lower bounds l and l^* are functions of these observational and interventional probabilities, we must have

$$\begin{aligned} l^*(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) &= l^*(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}), \\ l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) &= l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}) \end{aligned} \tag{225}$$

Since l^* consistently dominates l in SCM $\mathcal{M}$, the above equations imply

$$l^*(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) > l(x, y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) = P_x(y, \mathbf{c}_1, \mathbf{c}_2; \mathcal{M}^*) \tag{226}$$

This means that l^* is not a valid lower bound for $P_x(y, \mathbf{c}_1, \mathbf{c}_2)$ in $\mathcal{M}^*$, which is a contradiction. □

C Policy learning using partial effects

C.1 Simplifying bounds in terms of conditional distributions estimated using ML

Notice that the bound using Obs is simply $P(y|x, \mathbf{c}_1, \mathbf{c}_2)P(x|\mathbf{c}_1, \mathbf{c}_2)$. We consider the other cases below.

Case I: Lower bound using Obs, Exp($\mathbf{C}_1$).

$$\begin{aligned}
&\frac{(P_x(y, \mathbf{c}_1) - P(y, x, \mathbf{c}_1, \neg\mathbf{c}_2) - P(\neg x, \mathbf{c}_1, \neg c_2))}{P(\mathbf{c}_1, \mathbf{c}_2)} \\
&= \frac{P_x(y|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - \frac{(P(y, x, \mathbf{c}_1) - P(y, x, \mathbf{c}_1, \mathbf{c}_2))}{P(\mathbf{c}_1, \mathbf{c}_2)} - \frac{P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2)}{P(\mathbf{c}_1, \mathbf{c}_2)} \\
&= \frac{P_x(y|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - \frac{P(y, x, \mathbf{c}_1)}{P(\mathbf{c}_1, \mathbf{c}_2)} + \frac{P(y, x, \mathbf{c}_1, \mathbf{c}_2)}{P(\mathbf{c}_1, \mathbf{c}_2)} - \frac{P(\neg x, \mathbf{c}_1, \neg\mathbf{c}_2)}{P(\mathbf{c}_1, \mathbf{c}_2)} \\
&= \frac{P_x(y|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - \frac{P(y, x|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} + P(y|x, \mathbf{c}_1, \mathbf{c}_2)P(x|\mathbf{c}_1, \mathbf{c}_2) \\
&\qquad\qquad - \frac{P(\neg x, \mathbf{c}_1) - P(\neg x, \mathbf{c}_1, \mathbf{c}_2)}{P(\mathbf{c}_1, \mathbf{c}_2)} \\
&= \frac{P_x(y|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - \frac{P(y, x|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} + P(y|x, \mathbf{c}_1, \mathbf{c}_2)P(x|\mathbf{c}_1, \mathbf{c}_2) \\
&- \frac{P(\mathbf{c}_1) - P(\neg x, \mathbf{c}_1) - P(\neg x, \mathbf{c}_1, \mathbf{c}_2)}{P(\mathbf{c}_1, \mathbf{c}_2)} \\
&= \frac{P_x(y|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - \frac{P(y, x|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} + P(y|x, \mathbf{c}_1, \mathbf{c}_2)P(x|\mathbf{c}_1, \mathbf{c}_2) \\
&- \frac{P(\mathbf{c}_1)}{P(\mathbf{c}_1, \mathbf{c}_2)} + \frac{P(\mathbf{c}_1) - P(x, \mathbf{c}_1)}{P(\mathbf{c}_1, \mathbf{c}_2)} + \frac{P(\mathbf{c}_1, \mathbf{c}_2) - P(x, \mathbf{c}_1, \mathbf{c}_2)}{P(\mathbf{c}_1, \mathbf{c}_2)} \\
&= \frac{P_x(y|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - \frac{P(y, x|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} + P(y|x, \mathbf{c}_1, \mathbf{c}_2)P(x|\mathbf{c}_1, \mathbf{c}_2) - \frac{P(\mathbf{c}_1)}{P(\mathbf{c}_1, \mathbf{c}_2)} \\
&+ \frac{P(\mathbf{c}_1)}{P(\mathbf{c}_1, \mathbf{c}_2)} - \frac{P(x, \mathbf{c}_1)}{P(\mathbf{c}_1, \mathbf{c}_2)} + \frac{P(\mathbf{c}_1, \mathbf{c}_2)}{P(\mathbf{c}_1, \mathbf{c}_2)} - \frac{P(x, \mathbf{c}_1, \mathbf{c}_2)}{P(\mathbf{c}_1, \mathbf{c}_2)} \\
&= \frac{P_x(y|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - \frac{P(y, x|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} + P(y|x, \mathbf{c}_1, \mathbf{c}_2)P(x|\mathbf{c}_1, \mathbf{c}_2) + 1 \\
&- \frac{P(x|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - P(x|\mathbf{c}_1, \mathbf{c}_2) \\
&= \frac{P(\mathbf{c}_2|\mathbf{c}_1) + P_x(y|\mathbf{c}_1) - P(y, x|\mathbf{c}_1) - P(x|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} \\
&+ (P(y|x, \mathbf{c}_1, \mathbf{c}_2) - 1)P(x|\mathbf{c}_1, \mathbf{c}_2) \\
&= 1 + \frac{P_x(y|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - \frac{P(y|x, \mathbf{c}_1)P(x|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - \frac{P(x|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} \\
&+ (P(y|x, \mathbf{c}_1, \mathbf{c}_2) - 1)P(x|\mathbf{c}_1, \mathbf{c}_2)
\end{aligned} \tag{227}$$

The bound is analogously derived when Obs, Exp($\mathbf{C}_2$) are available.

Case II: Lower bound using Obs, Exp$(\mathbf{C}_1)$, Exp$(\mathbf{C}_2)$.

$$\begin{aligned}
&\frac{P_x(y,\mathbf{c}_2) - P_x(y,\neg\mathbf{c}_1) + P(y,x,\neg\mathbf{c}_1,\neg\mathbf{c}_2)}{P(\mathbf{c}_1,\mathbf{c}_2)} \\
&= \frac{P_x(y,\mathbf{c}_2) - P_x(y)}{P(\mathbf{c}_1,\mathbf{c}_2)} \\
&+ \frac{P_x(y,\mathbf{c}_1) + P(y,x,\neg\mathbf{c}_1) - P(y,x,\neg\mathbf{c}_1,\mathbf{c}_2)}{P(\mathbf{c}_1,\mathbf{c}_2)} \\
&= \frac{P_x(y,\mathbf{c}_2) - P_x(y) + P_x(y,\mathbf{c}_1)}{P(\mathbf{c}_1,\mathbf{c}_2)} \\
&+ \frac{P(y,x,\neg\mathbf{c}_1) - (P(y,x,\mathbf{c}_2) - P(y,x,\mathbf{c}_1,\mathbf{c}_2))}{P(\mathbf{c}_1,\mathbf{c}_2)} \\
&= \frac{P_x(y,\mathbf{c}_2) - P_x(y) + P_x(y,\mathbf{c}_1) + P(y,x,\neg\mathbf{c}_1)}{P(\mathbf{c}_1,\mathbf{c}_2)} \\
&- \frac{P(y,x,\mathbf{c}_2) + P(y,x,\mathbf{c}_1,\mathbf{c}_2)}{P(\mathbf{c}_1,\mathbf{c}_2)} \\
&= \frac{P_x(y|\mathbf{c}_2)}{P(\mathbf{c}_1|\mathbf{c}_2)} - \frac{P_x(y)}{P(\mathbf{c}_1,\mathbf{c}_2)} + \frac{P_x(y|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} + \frac{P(y,x)}{P(\mathbf{c}_1,\mathbf{c}_2)} \\
&- \frac{P(y|x,\mathbf{c}_1)P(x|\mathbf{c}_1)}{P(\mathbf{c}_2|\mathbf{c}_1)} - \frac{P(y|x,\mathbf{c}_2)P(x|\mathbf{c}_2)}{P(\mathbf{c}_1|\mathbf{c}_2)} \\
&+ P(y|x,\mathbf{c}_1,\mathbf{c}_2)p(x|\mathbf{c}_1,\mathbf{c}_2)
\end{aligned} \tag{228}$$

D Additional Empirical Analysis

D.1 Synthetic Data I

This section provides additional details on the Synthetic data experiments provided in Section 6 of the main paper.

SCM parameters

$$\begin{aligned}
&U \sim \mathcal{N}(0,1); \quad \mathbf{C}_1 = \alpha_{\mathbf{C}_1}U + \beta_{\mathbf{C}_1}; \quad \mathbf{C}_2 = \alpha_{\mathbf{C}_2}U + \beta_{\mathbf{C}_2} \\
&x = \sigma(\alpha_x^T[U,\mathbf{C}_1,\mathbf{C}_2]^T + \mathcal{N}(0,1)) > 0.5; \\
&y = \sigma(\alpha_y^T[U,\mathbf{C}_1,\mathbf{C}_2,x]^T + \mathcal{N}(0,1)) > 0.5
\end{aligned} \tag{229}$$

Parameter	Value
α_{c_1}	0.10
α_{c_2}	0.43
β_{c_1}	2.05e-04
β_{c_2}	8.9766e-05
α_x	[-0.06, 0.39, 0.46]
α_y	[-0.16, 0.41, 0.04, 0.1]

Table 1: Synthetic Data: SCM Parameters

Estimated bounds on Synthetic Data. Figure 5 shows the estimated bounds on Synthetic data with respect to covariate values.

D.2 Synthetic Data II

We present additional evaluation for a different choice of SCM parameters. The parametrization of the SCMs remains the same as above. The specific parameters used are given in Table 2. As before, the estimated bounds are shown in Figure 6. We sample 100 training points from the SCM to estimate the bounds, and learn the policy according to our safe policy learning framework. We test the learned policies on a separate sample of 100 test points. Figure 7 shows the performance of the learned

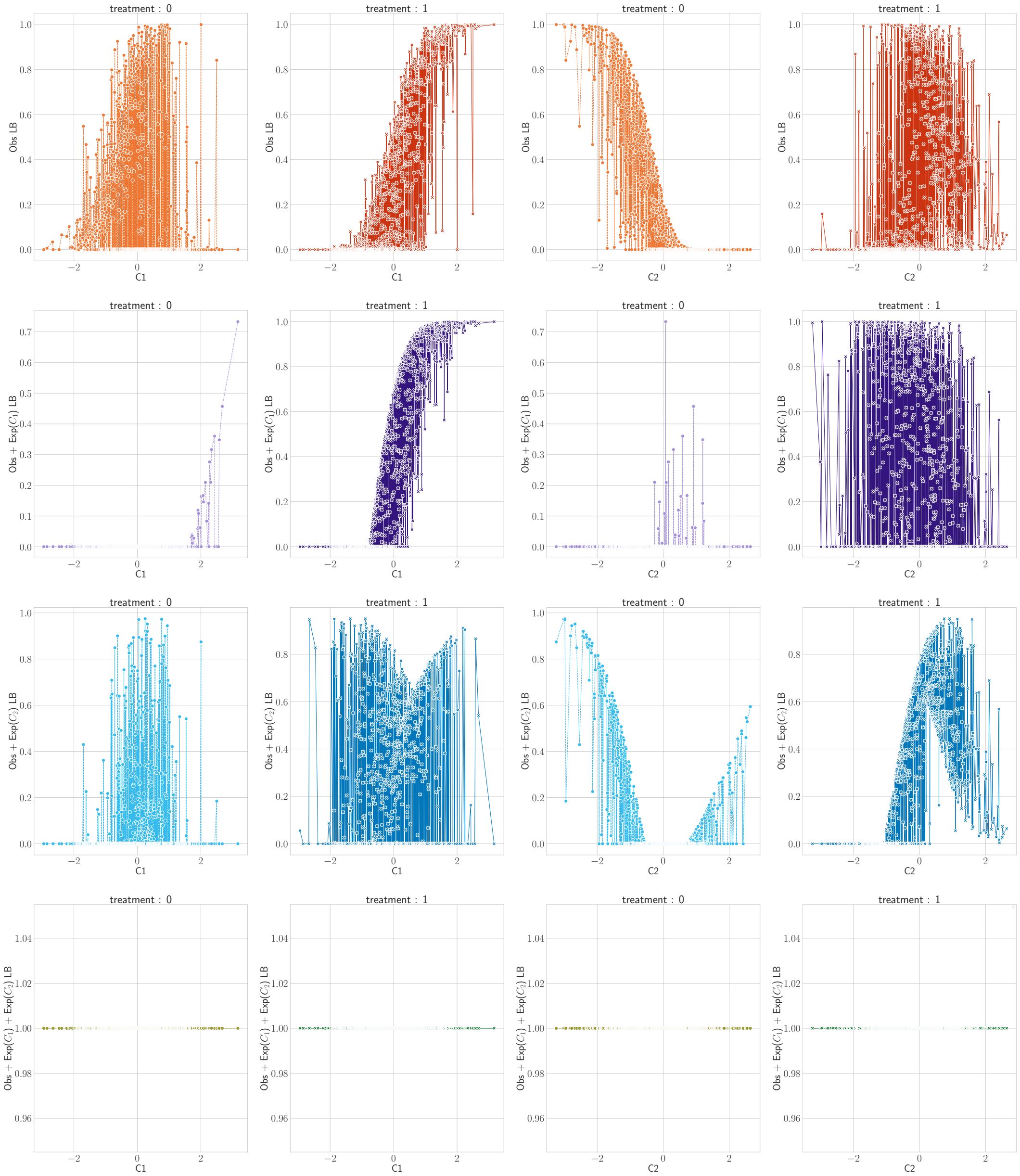

Figure 5: Synthetic Data: Lower bounds as a function of individual features C_1 and C_2. Each row corresponds to bounds obtained from different combination of data sources (see y-label). The first two columns show the corresponding lower bound for treatment $x = 0$ and $x = 1$ with respect to the first covariate c_1. Analogously, the third and fourth columns correspond to treatment $x = 0$ and $x = 1$ with respect to the second covariate c_2. Since we use plug-in estimates of the bounds, the bounds are not smooth. Nonetheless notice that for some covariate values, bounds obtained using Obs+Exp($\mathbf{C}_1$) or Obs+Exp($\mathbf{C}_2$) dominate the Obs bound providing us with better estimates of treatment effects. Finally, for this SCM setup, the bound using all data sources, specifically Obs+Exp($\mathbf{C}_1$)+Exp($\mathbf{C}_2$) saturates to 1 for $x = 0$ suggesting most samples do not need treatment to achieve improved outcome. However, it is still crucial to target patients carefully, and using the first three bounds in conjunction, allows us to learn a safe policy (see Figure 3).

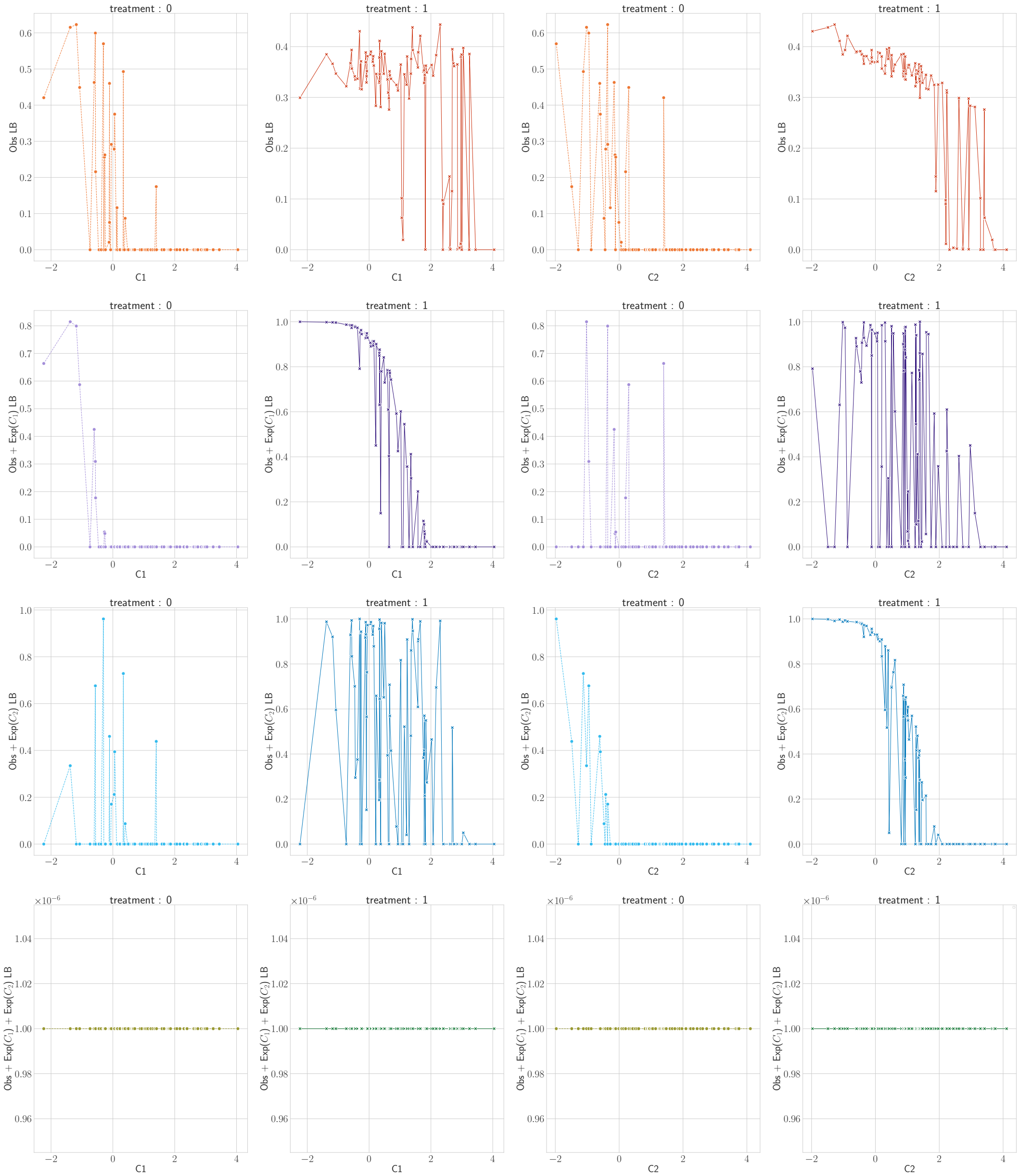

Figure 6: Synthetic Data II: Lower bounds as a function of individual features C_1 and C_2. Nonetheless notice that for some covariate values, bounds obtained using Obs+Exp($\mathbf{C}_1$) or Obs+Exp($\mathbf{C}_2$) dominate the Obs bound providing us with better estimates of treatment effects. Finally, for this SCM setup, the bound using all data sources, specifically Obs+Exp($\mathbf{C}_1$)+Exp($\mathbf{C}_2$) is 0 and unable to provide informative estimates. Nonetheless using all bounds in conjunction, we can obtain some improvements using our safe policy learning.

Parameter	Value
α_{c_1}	0.7
α_{c_2}	0.86
β_{c_1}	0.34
β_{c_2}	0.32
α_x	[-0.038, 0.012, 0.013]
α_y	[-0.72, -0.58, -0.67, 3.5]

Table 2: Synthetic Data: SCM Parameters

	Synthetic Data I	Synthetic Data II	IST Data
$p(y\|x, c)$	n/a	n/a	XGBoost
$p(y\|x, c_1), p(y\|x, c_2)$	n/a	n/a	XGBoost
$p(x\|c)$	n/a	n/a	XGBoost
$p(x\|c_1), p(x\|c_2)$	n/a	n/a	XGBoost
$p(c_1, c_2)$	Gaussian	Gaussian	Systolic BP: Gaussian Age, Sex: Bernoulli
Bounds estimation	Plugin	Plugin	Plugin
Policy Learning Rate	0.01	0.1	0.001
Policy Model	Model type: MLP # Hidden: 5 # Layers: 2 Activation: GeLU	Model type: MLP # Hidden: 5 # Layers: 2 Activation: GeLU	Model type: MLP # Hidden: 5 # Layers: 2 Activation: GeLU

Table 3: Model training details for all datasets

policies compared to behavior and random policies. On the left, we see policy value at different thresholds on the policy score. On the right is the mean outcome when x-proportion of samples are treated after sorting by policy score. As in previous experiment, we see that the learned policies dominates the behavior and random policy. In this case, since l_4 is uninformative, the performance of using all bounds overlaps with that of using l_1, l_2 and l_1, l_3. While observational bound trails the other learned policies, the consequence in selecting patients is not significant as can be seen on the right. Nonetheless all learned policies dominate the behavior policy suggesting utility of our framework to target samples for treatment.

D.3 IST Data

Generating Observational Data. IST data can be accessed at Sandercock et al. [54]. IST consists of trial data studying the effect of Aspirin treatment allocation on clinical course of Ischemic stroke. For our empirical analysis, we generate i) Obs: an observational dataset with unobserved confounding, ii) Exp($\mathbf{C}_1$), Exp($\mathbf{C}_2$): Two experimental datasets with partial observability. We choose $\mathbf{C}_1 = \{$Age, Sex$\}$ and $\mathbf{C}_2 = \{$Systolic Blood Pressure (SysBP)$\}$ constitute observed confounding attributes. We binarize Age (≥ 73 is 1, and 0 otherwise).

To generate Obs, we induce selection bias based on Age $\{0, 1\}$, Sex $\{0, 1\}$, SysBP, and Conscious state (CONSC). A person's conscious state at randomization can take 3 values (0: fully alert, 1: drowsy, 2: unconscious). We introduce a new variable Z such that:

$$\begin{aligned}
l_c &= 0.9 * \mathbf{1}(\text{CONSC} == 0) + 0.7 * \mathbf{1}(\text{CONSC} == 1) \\
&\quad - 0.6 * \mathbf{1}(\text{CONSC} == 2) \\
l_s &= 0.85 * \mathbf{1}(\text{Sex} == 0) - 0.1 * \mathbf{1}(\text{Sex} == 1) \\
l_a &= 0.7 * \mathbf{1}(\text{Age} == 0) - 0.1 * \mathbf{1}(\text{Age} == 1) \\
l_{bp} &= 0.8 * \mathbf{1}(\text{SysBP} \leq 120) + 0.5 * \mathbf{1}(120 < \text{SysBP} \leq 130) \\
&\quad - 0.01 * \mathbf{1}(130 < \text{SysBP} \leq 140) \\
&\quad - 0.3\mathbf{1}(140 < \text{SysBP} \leq 180) - 0.6\mathbf{1}(\text{SysBP} > 180) \\
Z &= \mathbf{1}(\sigma(l_c + l_s + l_a + l_{bp}) > 0.65)
\end{aligned} \tag{230}$$

Observational samples are chosen if $Z = 1$ and dropped otherwise. Exp($\mathbf{C}_1$) data drops all covariates except treatment, outcome, Age and Sex. Exp($\mathbf{C}_2$) drops all covariates except treatment, outcome, SysBP.

Estimated Lower bounds on IST Data.

D.4 Modeling Details

Table 3 provides modeling details for intermediate distributions used to estimate the plugin bounds. Bounds are estimated using the derivations in Appendix C.1. Results of the learned policies are discussed in the main paper in Section 6.

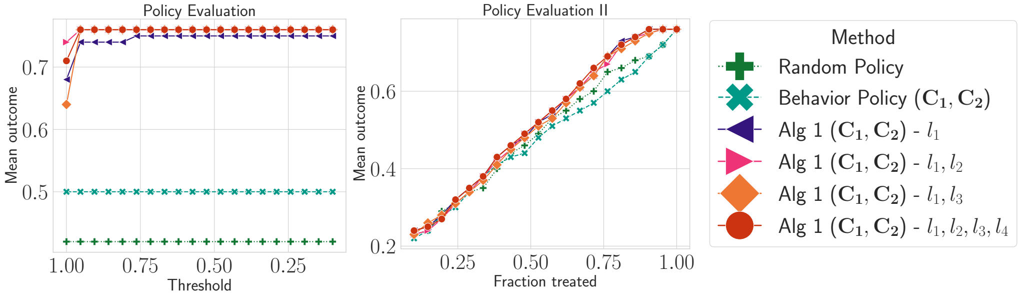

Figure 7: Synthetic Data II Policy Evaluation. Left: Policy evaluation with varying threshold 1 (zero treated) $\rightarrow 0$ (all treated) on policy scores. For fewer treated samples (higher threshold), learned policies peak much earlier suggesting that our learned policies are better at targeting patients over behavior policy. Right: Mean outcome when x-proportion of the population is targeted for treatment using sorted policy scores. Behavior policy trails relative to the learned policies. Since l_4 is uninformative, performance of using all bounds (red) is comparable to using l_1, l_2 and l_1, l_3. Learned policies are also better at selecting patients compared to random and behavior policy, though their variability in using different bounds is low.

E Additional Discussions

We discuss limitations and broader impacts of our proposed work here.

Limitations. Our bounds are derived for discrete treatments and outcomes. Our proposed algorithm is a two-step approach where the first step estimates plugin bounds and the second step learns a policy with estimated plugin bounds. Considering doubly-robust estimates of bounds, implications for policy learning, sample efficiency, and joint approaches are interesting aspects of future work.

Broader Impacts. Our work considers a highly relevant healthcare setting where observational studies are affected by unobserved confounding rendering conventional offline policy learning approaches ineffective. By leveraging relevant experimental studies with partial observability, we demonstrate that effective safe offline learning is indeed possible. Our robust policy learning framework optimizes for worst-case treatment effects using our estimated lower-bounds. Our work is a proof-of-concept and adds a novel perspective to existing body of work on robust policy learning. However, understanding of real-world constraints is required before the proposed method is ready for practical deployment in health and medicine. For example, our policy learning framework is useful for identifying patients who benefit from treatment. However, in practice we might want to focus on populations for whom no treatment is actively harmful. This requires different learning objectives and is an active area of our future work.

F Compute Resources

All experiments were conducted on a 2.3 GHz Dual-Core Intel Core i5, 8 GB RAM, MacOS Monterey 12.5.1. No GPU Requirements. Code can be found at `https://github.com/reAIM-Lab/safecrl`.

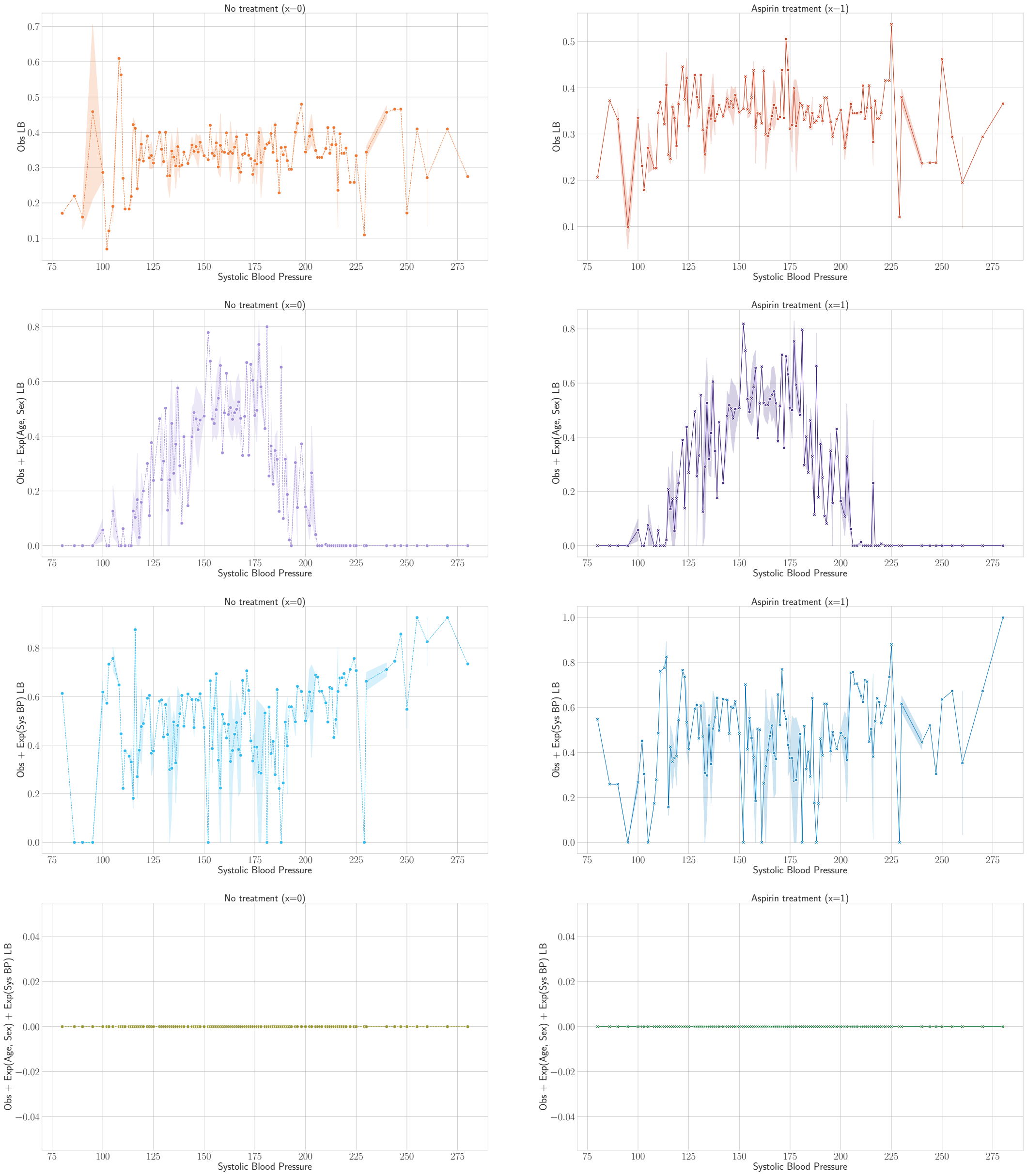

Figure 8: IST Data: Lower bounds as a function of individual SysBP features. Each row corresponds to bounds obtained from different combination of data sources (see y-label). The first two columns show the corresponding lower bound for no-treatment ($x = 0$) and Aspirin treatment ($x = 1$) respectively. We model SysBP as a Gaussian resulting in relatively smooth variation. However, since we use plug-in estimates of the bounds, the bounds are not smooth. Notice that for some covariate values, bounds obtained using Obs+Exp($\mathbf{C}_1$) or Obs+Exp($\mathbf{C}_2$) dominate the Obs bound providing us with relatively better estimates of treatment effects. However, the effect of treatment itself is not significant or uniformly beneficial with respect to SysBP.