Bounds and Identification of Joint Probabilities of Potential Outcomes and Observed Variables Under Monotonicity Assumptions

Naoya Hashimoto[1] **Yuta Kawakami**[1] **Jin Tian**
Mohamed bin Zayed University of Artificial Intelligence

Abstract

Evaluating joint probabilities of potential outcomes and observed variables, and their linear combinations, is a fundamental challenge in causal inference. This paper addresses the bounding and identification of these probabilities in settings with discrete treatment and discrete ordinal outcome. We propose new families of monotonicity assumptions and formulate the bounding problem as a linear programming problem. We further introduce a new monotonicity assumption specifically to achieve identification. Finally, we present numerical experiments to validate our methods and demonstrate their application using real-world datasets.

1 INTRODUCTION

One of the fundamental challenges in causal inference is the evaluation of joint probabilities of potential outcomes (POs) and observed variables (OVs) and their linear combinations. Let the treatment (X) take values in $\{0, 1, \dots, d_X - 1\}$ and the outcome (Y) take values in $\{0, 1, \dots, d_Y - 1\}$, and let Y_x denote the PO under treatment x. The joint POs/OVs probabilities $\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \dots, Y_{d_X-1} = y_{d_X-1}, X = x, Y = y)$ and their linear combinations encompass various causal quantities of practical interest, including, among others, (i) the probabilities of causation (PoC) [Pearl, 1999, Tian and Pearl, 2000, Li and Pearl, 2024], which measure the necessity and sufficiency of the treatment; (ii) the posterior causal effects [Lu et al., 2022], which measure causal effects given evidence on post-treatment variables; and (iii) the moments of causal effects [Kawakami and Tian, 2025a], which capture the shape of the distribution of causal effects.

[1]Equal contribution.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

However, the joint POs/OVs probabilities are not identifiable from observational data without additional assumptions; while bounds can be derived, they are often too wide to be informative. Their evaluation positions the problem at the highest third layer of Pearl's causal hierarchy [Pearl and Mackenzie, 2018, Bareinboim et al., 2022]. To address this issue, various monotonicity assumptions (MAs) on POs have been actively studied.

Pearl [1999] and Tian and Pearl [2000] provided an identification theorem for PoC under the MA that no subjects satisfy $(Y_0 = 1, Y_1 = 0)$ (combined with the exogeneity assumption) when the treatment and outcome are binary. This MA is a natural and reasonable assumption in a wide range of empirical studies [Hernan and Robins, 2024]. In the instrumental variable (IV) literature, this MA is often invoked to describe the relationship between the binary IV (Z) and treatment (X) for the identification of local average treatment effects (LATE) [Imbens and Angrist, 1994, Angrist and Imbens, 1995]. Recently, Mueller and Pearl [2023], Mueller [2025] introduced a probability-limited MA, termed ϵ-limited harm, that allows subjects with $(Y_0 = 1, Y_1 = 0)$ with probability at most ϵ, under which they provided closed-form bounds for PoC.

For binary treatment and discrete ordinal outcome, Zhang et al. [2025] studied the bounds for the probability of necessity (PN) under an extended MA for discrete outcome, $Y_0 \leq Y_1$, and further proposed a new increment-limited MA $0 \leq Y_1 - Y_0 \leq 1$, termed the monotonic incremental treatment effect (MITE), for identifying PN. For discrete treatment settings, Manski [1997] has introduced "Monotone Treatment Response" (MTR) assumption ($Y_0 \leq Y_1 \leq \cdots \leq Y_{d_X-1}$) to bound the average causal effect (ACE) $\mathbb{E}[Y_i - Y_j]$.

In this paper, we introduce new families of MAs in settings with discrete treatment and ordinal outcome. Our work unifies the existing MAs into a single general MA, which accommodates multiple probabilistic and increment-limited relationships. Table 1 summarizes the MAs in the literature.

Under our MAs, we address the problem of bounding the joint POs/OVs probabilities and their linear combina-

Table 1: Summary of monotonicity assumptions (MAs) in causal inference. The symbol (✓) indicates that the objective quantity is identified under the MA. $\lambda(\mathcal{P})$ represents an arbitrary linear combination of the joint POs/OVs probabilities.

Research	Treatment Type	Outcome Type	MA	Objective	Identification
Pearl [1999]; Tian and Pearl [2000]	Binary	Binary	$\mathbb{P}(Y_0 = 1, Y_1 = 0) = 0$	PoC	✓
Mueller and Pearl [2023]	Binary	Binary	$0 \leq \mathbb{P}(Y_0 = 1, Y_1 = 0) \leq \epsilon$	PoC	
Zhang et al. [2025]	Binary	Ordinal	$Y_0 \leq Y_1$	PN	
Zhang et al. [2025]	Binary	Ordinal	$0 \leq Y_1 - Y_0 \leq 1$	PN	✓
Manski [1997]	Discrete	General	$Y_0 \leq Y_1 \leq \cdots \leq Y_{d_X - 1}$	ACE	
This paper	Discrete	Ordinal	Assumptions 7 $\sim$ 11	$\lambda(\mathcal{P})$	
This paper	Discrete	Ordinal	Assumption 12	$\lambda(\mathcal{P})$	✓

tions. We formulate the bounding problem using a linear programming (LP) approach [Balke, 1995, Balke and Pearl, 1997, Tian and Pearl, 2000, Duarte et al., 2024], incorporating new linear constraints implied by our new MAs. We discuss the estimation of the bounds from finite-sample data and show numerical experiments to validate our method.

We further provide an MA assumption for identifying joint POs/OVs probabilities (and their linear combinations) from observational or experimental data. We show numerical experiments to validate their estimation from finite-sample data.

Finally, we demonstrate the application of our methods on real-world datasets.

2 MOTIVATING EXAMPLE

We motivate our study using a real-world example.

Dataset. We consider an open-access dataset on student performance in two Portuguese schools, available at (https://archive.ics.uci.edu/dataset/320/student+performance). Students in secondary education are tested annually over three years, yielding three total assessments. The observational dataset includes attributes related to demographics, social factors, school-related features, and student academic performance. The sample size is 649, with no missing values.

Variables. We define study hours as the treatment variable X, where $X = 0, 1, 2, 3$ denotes studying less than 2 hours, 2 to 5 hours, 5 to 10 hours, or more than 10 hours weekly, respectively. We consider the Portuguese language score in the final period as the outcome variable (Y), and categorize it as $Y = 0$ if the score is between 0 and 8, $Y = 1$ if the score is between 9 and 15, and $Y = 2$ if the score is between 16 and 20, corresponding to low, medium, and high score categories, respectively.

To assess the causal effect of study hours on student performance, existing work typically computes ACEs, such as $\mathbb{E}[Y_1 - Y_0]$, $\mathbb{E}[Y_2 - Y_1]$. However, ACEs capture only one aspect of the causal relationship. Researchers are often interested in more nuanced causal questions that necessitate the use of the joint POs/OVs probabilities. For example, consider the following two causal questions:

"*For a student who studied more than 10 hours and obtained a score of more than 15 in reality, would their score be less than 16 had they studied less than 10 hours?*"

This is a situation in which studying more than 10 hours is necessary to obtain a score above 15.

"*For a student who studied less than 2 hours and obtained a score of no more than 8 in reality, would their score be more than 8 had they studied at least 2 hours?*"

This is a situation in which studying at least 2 hours is sufficient to obtain a score above 8.

Probabilities representing such two situations can be formally expressed in terms of the joint POs/OVs probabilities as follows, respectively:

$$\begin{aligned} &\mathbb{P}(Y_0 \leq 1, Y_1 \leq 1, Y_2 \leq 1 | X = 3, Y = 2), \\ &\mathbb{P}(Y_1 \geq 1, Y_2 \geq 1, Y_3 \geq 1 | X = 0, Y = 0). \end{aligned} \tag{1}$$

However, counterfactual probabilities in (1) are not identifiable from observational data. In this paper, we show how to derive bounds for these probabilities using linear programming. We illustrate that the bounds can be tightened by incorporating plausible MAs based on domain knowledge. For example, one might assume that "a student would obtain a higher score had they studied 2 to 5 hours weekly rather than less than 2 hours" ($Y_0 \leq Y_1$) or "99.5% of the students would have obtained a higher score had they studied more than 10 hours weekly rather than 5 to 10 hours" ($\mathbb{P}(Y_2 \leq Y_3) \geq 0.995$). Further, we show that the probabilities in (1) are identifiable if we impose stronger MAs $0 \leq Y_i - Y_j \leq 1$ for all $i > j$.

3 NOTATIONS AND BACKGROUND

We represent each variable with a capital letter (X) and its realized value with a small letter (x). Let $\mathbb{I}(x)$ be an indicator function that takes 1 if x is true and 0 if x is false. The symbol $\wedge$ represents logical conjunction.

Structural Causal Models. We use the language of Structural Causal Models (SCM) as our basic semantic and in-

ferential framework [Pearl, 2009]. An SCM $\mathcal{M}$ is a tuple $\langle \boldsymbol{U}, \boldsymbol{V}, \mathcal{F}, \mathbb{P}_{\boldsymbol{U}} \rangle$, where $\boldsymbol{U}$ is a set of exogenous (unobserved) variables following joint probabilities $\mathbb{P}_{\boldsymbol{U}}$, and $\boldsymbol{V}$ is a set of endogenous (observable) variables whose values are determined by structural functions $\mathcal{F} = \{f_{V_i}\}_{V_i \in \boldsymbol{V}}$ such that $v_i := f_{V_i}(\mathbf{pa}_{V_i}, \boldsymbol{u}_{V_i})$ where $\mathbf{PA}_{V_i} \subseteq \boldsymbol{V}$ and $U_{V_i} \subseteq \boldsymbol{U}$. An atomic intervention of setting a set of endogenous variables X to constants x, denoted by $do(x)$, replaces the original equations of X by $X := x$ and induces a *sub-model* $\mathcal{M}_x$. We denote the potential outcome (PO) Y under intervention $do(x)$ by $Y_x(\boldsymbol{u})$, which is the solution of Y with $\boldsymbol{U} = \boldsymbol{u}$ in the sub-model $\mathcal{M}_x$.

In this paper, we consider a discrete treatment variable X and an *ordinal* outcome variable Y. Y must be ordinal since monotonicity assumptions presuppose an ordering among the POs. By contrast, X need not be ordinal.

Assumptions in Previous Works. The exogeneity is a commonly used assumption stated as follows:

Assumption 1 (Exogeneity)**.** *$Y_x \perp\!\!\!\perp X$ for any x.*

Assumption 1 implies $\mathbb{P}(Y_x) = \mathbb{P}(Y|X = x)$ for any x.

Let X and Y be discrete random variables with supports $\{0, 1, \ldots, d_X - 1\}$ and $\{0, 1, \ldots, d_Y - 1\}$, respectively. We list the MAs used in previous works.

Assumption 2 (Monotonicity for Binary Treatment and Outcome)**.** *[Pearl, 1999, Tian and Pearl, 2000]* $\mathbb{P}(Y_0 = 1, Y_1 = 0) = 0$.

Assumption 3 (Monotonicity for Discrete Outcome)**.** *[Zhang et al., 2025]* $Y_0 \leq Y_1$*, almost surely (a.s.).*

Assumption 4 (Monotonic Incremental Treatment Effect (MITE))**.** *[Zhang et al., 2025]* $0 \leq Y_1 - Y_0 \leq 1$*, a.s.*

Assumption 5 (ϵ-limited harm)**.** *[Mueller and Pearl, 2023]* $0 \leq \mathbb{P}(Y_0 = 1, Y_1 = 0) \leq \epsilon$.

Assumption 6 (Monotone Treatment Response (MTR))**.** *[Manski, 1997]* $Y_0 \leq Y_1 \leq \cdots \leq Y_{d_X - 1}$*, a.s.*

We note that Assumptions 2 and 5 apply to binary treatment and binary outcome, while Assumptions 3 and 4 apply to binary treatment. A detailed discussion of these MAs is provided in Appendix A.

4 MONOTONICITY ASSUMPTIONS FOR DISCRETE TREATMENT AND OUTCOME

We introduce new families of MAs by extending the existing MAs to discrete treatment and discrete outcome settings and generalizing them to accommodate more complex PO relationships.

MAs for Binary Treatment and Discrete Outcome. We introduce a new assumption for binary treatment that captures both probability-limited and increment-limited forms of monotonicity.

Assumption 7 (Probability-limited and Increment-limited Monotonicity for Binary Treatment)**.** *There exist known constants $\underline{D}$, $\overline{D}$, L, and U, such that the following condition holds:*

$$L \leq \mathbb{P}\left(\underline{D} \leq Y_1 - Y_0 \leq \overline{D}\right) \leq U. \tag{2}$$

These values $\underline{D}$, $\overline{D}$, L, and U are pre-specified by the researchers as prior knowledge. This MA generalizes Assumption 4 by allowing both the increment condition and the probability condition simultaneously. It reduces to Assumption 3 when $L = U = 1$, $\underline{D} = 0$, $\overline{D} = \infty$, and equals Assumption 4 when $L = U = 1$, $\underline{D} = 0$, $\overline{D} = 1$.

MAs for Discrete Treatment and Outcome. Next, we present MAs for discrete treatment and discrete outcome.

We introduce a relaxation of Assumption 6 by allowing for probability-limited restrictions.

Assumption 8 (Probability-limited MTR)**.** *There exist known constants L and U, such that*

$$L \leq \mathbb{P}(Y_0 \leq Y_1 \leq \cdots \leq Y_{d_X - 1}) \leq U. \tag{3}$$

This condition implies that the POs for each subject increase monotonically with the treatment level with probability in $[L, U]$. It coincides with Assumption 6 when $L = U = 1$.

We introduce a relaxation of the MITE assumption (Assumption 4) that allows for increment-limited restrictions in the case of discrete treatments.

Assumption 9 (Increment-limited Monotonicity for Discrete Treatment)**.** *There exist known constants $\underline{D}_{st}$ and $\overline{D}_{st}$ for $s, t = 0, \ldots, d_X - 1$, such that the following condition holds:*

$$\bigwedge_{s,t=0,\ldots,d_X-1; s>t} (\underline{D}_{st} \leq Y_s - Y_t \leq \overline{D}_{st}),\ \textit{a.s.} \tag{4}$$

This condition requires that the difference between any two potential outcomes Y_s and Y_t lies within the interval $[\underline{D}_{st}, \overline{D}_{st}]$ a.s. whenever $s > t$. Note that we can avoid imposing a restrictive condition on $Y_s - Y_t$ by setting $\underline{D}_{st} = -\infty$ and $\overline{D}_{st} = \infty$.

We relax Assumption 9 with probability-limited restrictions.

Assumption 10 (Probability-limited and Increment-limited Monotonicity with a Single Term)**.** *There exist known constants L, U, and $\underline{D}_{st}$ and $\overline{D}_{st}$ for $s, t = 0, \ldots, d_X - 1$, such that the following condition holds:*

$$L \leq \mathbb{P}\left(\bigwedge_{s,t=0,\ldots,d_X-1; s>t} (\underline{D}_{st} \leq Y_s - Y_t \leq \overline{D}_{st})\right) \leq U. \tag{5}$$

The above condition bounds the probability that the difference between any two potential outcomes Y_s and Y_t with $s > t$ lies within the interval $[\underline{D}_{st}, \overline{D}_{st}]$ by $[L, U]$. Assumption 9 is the special case with $L = U = 1$.

Researchers may impose multiple probability-limited and increment-limited MAs simultaneously, e.g., $0.5 \leq \mathbb{P}(0 \leq Y_1 - Y_0 \leq 1)$, $0.5 \leq \mathbb{P}(0 \leq Y_2 - Y_1 \leq 1)$, and $0.5 \leq \mathbb{P}(0 \leq Y_2 - Y_0 \leq 1)$, each in the form of Assumption 10. For convenience, we introduce the following single MA to represent multiple simultaneous requirements in the form of Assumption 10:

Assumption 11 (Probability-limited and Increment-limited Monotonicity with Multiple Terms)**.** *There exist known constants $\underline{D}_{st}^{(w)}$ and $\overline{D}_{st}^{(w)}$ for $s, t = 0, \ldots, d_X - 1$ and $w = 1, \ldots, W$, and L^w and U^w for $w = 1, \ldots, W$, such that the following condition holds, for $w = 1, \ldots, W$:*

$$L^w \leq \mathbb{P}\left(\bigwedge_{s,t=0,\ldots,d_X-1; s>t} (\underline{D}_{st}^{(w)} \leq Y_s - Y_t \leq \overline{D}_{st}^{(w)}) \right) \leq U^w. \tag{6}$$

In Assumption 11, W represents the number of simultaneous monotonicity conditions imposed.

Relationships among MAs.

Proposition 1. *We have the following relationships among the MAs presented thus far: (1) Assumptions 2 through 10 are all special cases of Assumption 11. (2) Assumption 4 is a special case of Assumptions 7, 9, and 10. (3) Assumption 5 is a special case of Assumptions 7, 8, and 10. (4) Assumption 6 is a special case of Assumptions 8, 9, and 10.*

Assumption 11 encompasses existing MAs in the literature and represents the most general class of monotonicity among all the MAs considered in this paper.

5 BOUNDING WITH MONOTONICITY ASSUMPTIONS

In this section, we formulate the bounding problem for joint POs/OVs probabilities given observed distributions as a linear programming (LP), incorporating additional MAs.

We denote $[d] \coloneqq \{0, \ldots, d-1\}$, $Y_{[d_X]} = (Y_0, \ldots, Y_{d_X-1})$, and $y_{[d_X]} = (y_0, \ldots, y_{d_X-1})$.

5.1 Objective of the Study

We denote the set of the joint POs/OVs probabilities as

$$\mathcal{P} \triangleq \{\mathbb{P}(Y_{[d_X]} = y_{[d_X]}, X = x, Y = y) : y_{[d_X]} \in [d_Y]^{d_X}, x \in [d_X], y \in [d_Y]\}. \tag{7}$$

The objective of this study is a linear combination of the joint POs/OVs probabilities, expressed as $\lambda(\mathcal{P})$, where λ is an arbitrary linear function mapping the parameters $\mathcal{P}$ to $\mathbb{R}$. $\lambda(\mathcal{P})$ encompasses a wide range of causal quantities studied in the literature, including the PoC families in [Li and Pearl, 2024], the moments of causal effect [Kawakami and Tian, 2025a], and the posterior causal effects [Lu et al., 2022]. A detailed illustration is provided in Appendix B.

5.2 Bounding Problem Setup

We set up the bounding problem by LP. We denote the following parameters

$$p_{y_{[d_X]},x} = \mathbb{P}(Y_{[d_X]} = y_{[d_X]}, X = x). \tag{8}$$

We denote the set of parameters as $\mathfrak{p} \triangleq \{p_{y_{[d_X]},x} : y_{[d_X]} \in [d_Y]^{d_X}, x \in [d_X]\}$. The total number of parameters is $d_Y^{d_X} d_X$. The number of parameters grows exponentially in d_X and polynomially in d_Y. Note that we do not include Y in the parameterization of the bounding problem since we have $\mathbb{P}(Y_{[d_X]} = y_{[d_X]}, X = x, Y = y) = \mathbb{P}(Y_{[d_X]} = y_{[d_X]}, X = x)\mathbb{I}(y_x = y)$ from the counterfactual consistency property [Pearl, 2009], i.e., $X = x \Rightarrow Y_x = Y$. Then, $\lambda(\mathcal{P})$ can be represented as a linear combination of the parameters in $\mathfrak{p}$ and denote it $\phi(\mathfrak{p})$.

Constraints from Observed Distributions. The general constraints are

$$\begin{aligned} &p_{y_{[d_X]},x} \geq 0 \text{ for all } y_{[d_X]} \text{ and } x, \\ &\sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} p_{y_{[d_X]},x} = 1. \end{aligned} \tag{9}$$

Next, we present the constraints imposed by three types of distributions: (i) experimental distribution only, (ii) observational distribution only, and (iii) both experimental and observational distributions. First, if we have the experimental distribution $\mathbb{P}(Y_x)$, it imposes

$$\begin{aligned} &\sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{k-1}=0}^{d_Y-1} \sum_{y_{k+1}=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \\ &\qquad p_{y_0,\ldots y_{k-1},j,y_{k+1}\ldots,y_{d_X-1},x} = \mathbb{P}(Y_k = j) \end{aligned} \tag{10}$$

for $k \in [d_X]$ and $j \in [d_Y - 1]$. We do not include $j = d_Y - 1$ since Eq. (10) for $j = d_Y - 1$ can be expressed as a linear combination of Eqs. (9) and (10).

Next, if we have the observational distribution $\mathbb{P}(X, Y)$, it imposes

$$\begin{aligned} &\sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{l-1}=0}^{d_Y-1} \sum_{y_{l+1}=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \\ &\qquad p_{y_0,\ldots,y_{l-1},m,y_{l+1},\ldots,y_{d_X-1},l} = \mathbb{P}(X = l, Y = m) \end{aligned} \tag{11}$$

for $l \in [d_X]$, $m \in [d_Y]$, $(l, m) \neq (d_X - 1, d_Y - 1)$. We excluded $(l, m) = (d_X - 1, d_Y - 1)$ since Eq. (11) in that

case can be expressed as a linear combination of Eqs. (9) and (11).

Finally, when we have both experimental and observational distributions, they impose the constraints in both Eqs. (10) and (11).

All of the above constraints from distributions are linear in $\mathfrak{p}$.

Remark. Introducing exogeneity (Assumption 1) can be helpful in deriving tighter bounds. $Y_x \perp\!\!\!\perp X$ imposes the constraints $\mathbb{P}(Y_x = y_x, X = l) = \mathbb{P}(Y_x = y_x)\mathbb{P}(X = l)$ on $p_{y_{[d_X]},x}$, which are linear given either observational or experimental distributions.

Additional Linear Constraints by MAs. Next, we present the additional linear constraints imposed by our introduced MAs.

Proposition 2. *The constraints implied by Assumption 11 are given by*

$$L^w \leq \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \left(\prod_{s,t\in[d_X];s>t} \mathbb{I}(\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)}) \right) p_{y_{[d_X]},x} \leq U^w \tag{12}$$

for $w = 1, \ldots, W$. The above constraints are all linear constraints in $\mathcal{P}$.

The explicit constraints implied by Assumptions 6-10 are provided in Appendix D.2.

Bounding Problems Under MA. Under Assumption 11, the bounds of $\lambda(\mathcal{P})$ are obtained by solving the following constrained optimization problems:

$$\underline{\lambda}(\mathcal{P}) = \min_{\mathfrak{p}} \phi(\mathfrak{p}), \;\; \overline{\lambda}(\mathcal{P}) = \max_{\mathfrak{p}} \phi(\mathfrak{p}),$$
$$\text{subject to the constraints from observed distributions and Eq. (12)} \tag{13}$$

The resulting interval $[\underline{\lambda}(\mathcal{P}), \overline{\lambda}(\mathcal{P})]$ represents the bounds of $\lambda(\mathcal{P})$ under Assumption 11. These bounds are guaranteed to be sharp under Assumption 11. The bound $[\underline{\lambda}(\mathcal{P}), \overline{\lambda}(\mathcal{P})]$ is said to be *sharp* if there exist $\underline{\mathfrak{p}}, \overline{\mathfrak{p}} \in \mathcal{P}$ that satisfy the imposed constraints and achieve $\phi(\underline{\mathfrak{p}}) = \underline{\lambda}(\mathcal{P})$ and $\phi(\overline{\mathfrak{p}}) = \overline{\lambda}(\mathcal{P})$. We provide the explicit formulation of the LPs in Appendix D.3.

Therefore, we have the following theorem:

Theorem 1. *Given observed distributions (experimental $\mathbb{P}(Y_x)$, observational $\mathbb{P}(X, Y)$, or both), under Assumption 11, solving the LP problem in* (13) *yields sharp bounds for $\lambda(\mathcal{P})$.*

Remark. There may be a compatibility issue in LPs, that is, the constraints implied by the MAs conflict with those supported by the observed distributions. If the feasible region of the LP—that is, the set of all points satisfying all constraints simultaneously—is empty, then the LP is infeasible. In such cases, the researcher should withdraw or revise the assumptions.

5.3 Bounding Joint POs/OVs Probabilities and Their Linear Combinations From Finite Samples

We use a plug-in estimator to compute the bounds from finite samples. That is, we use the empirical probabilities of $\mathbb{P}(Y_x)$ and $\mathbb{P}(X, Y)$ in solving the LP in (13). The resulting plug-in estimators of the lower and upper bounds of $\lambda(\mathcal{P})$ are denoted by $[\hat{\underline{\lambda}}(\mathcal{P}), \hat{\overline{\lambda}}(\mathcal{P})]$. $\hat{\underline{\lambda}}(\mathcal{P})$ and $\hat{\overline{\lambda}}(\mathcal{P})$ are consistent estimators of $\underline{\lambda}(\mathcal{P})$ and $\overline{\lambda}(\mathcal{P})$. A detailed description is given in Appendix D.4.

A discussion of the computational complexity is provided in Appendix D.5. The computational complexity of solving our bounding problems is polynomial when using interior-point methods [Karmarkar, 1984]. If we use the simplex method to solve LP, the worst-case computational complexity of solving our bounding problems is exponential [RinnooyKan and Telgen, 1981]. However, in practice, the simplex method rarely exhibits its exponential worst-case behavior and is often faster than interior-point algorithms [Vanderbei, 2001].

Numerical Experiments. We conduct numerical experiments to illustrate how MAs can tighten the bounds. All details are stated in Appendix E.1. We simulated i.i.d. samples of experimental and observational datasets and computed bounds of the joint POs probability $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$, the posterior causal effect $\mathbb{E}[Y_1 - Y_0 | X = 2, Y = 2]$, and the second moment of causal effect $\mathbb{E}[(Y_1 - Y_0)^2]$ under various MAs. The results are reported in Tables 10-21 in Appendix E.1. The results show that our methods can provide informative bounds. We observe that introducing stronger assumptions leads to narrower bounds, and that the 95% confidence intervals (CIs) for the estimates become narrower as the sample size increases.

6 IDENTIFICATION WITH MONOTONICITY ASSUMPTIONS

In this section, we introduce an MA that leads to the identification of the joint POs/OVs probabilities and their linear combinations.

6.1 Identifying Joint POs/OVs Probabilities and Their Linear Combinations

We present an assumption for identifying the joint POs/OVs probabilities, which is a special case of Assumption 11 and an extension of MITE (Assumption 4).

Assumption 12 (MITE for Discrete Treatment and Outcome)**.** *For any $s > t$,*

$$0 \leq Y_s - Y_t \leq 1, \text{ a.s.} \tag{14}$$

This assumption means that raising any treatment level can never make the outcome worse and can only improve it slightly, by at most one level.

Assumption 12 is a special case of Assumption 9, where $\underline{D}_{st} = 0$ and $\overline{D}_{st} = 1$ for any $s > t$. When Y is binary, Assumption 12 is equivalent to Assumption 6 and is further equivalent to $\mathbb{P}(Y_t = 1, Y_s = 0) = 0$ for any $s > t$.

Remark. Similar to the bounding problem, the feasible region defined by the linear constraints from data and the restriction imposed by Assumption 12 must be non-empty.

Experimental Data. We discuss the identification of the joint POs probabilities, i.e., $\mathbb{P}(Y_0, \ldots, Y_{d_X-1})$, and their linear combinations, using experimental data. Because the experimental data does not contain (X, Y), we focus solely on $\mathbb{P}(Y_0, \ldots, Y_{d_X-1})$.

We denote the set of the joint probabilities of POs as

$$\mathcal{P}' \triangleq \{\mathbb{P}(Y_{[d_X]} = y_{[d_X]}) | y_{[d_X]} \in [d_Y]^{d_X}\}. \tag{15}$$

The objective is a linear combination of the joint POs probabilities, expressed as $\lambda'(\mathcal{P}')$, where λ' is an arbitrary linear function mapping the parameters $\mathcal{P}'$ to $\mathbb{R}$.

We have the following identification theorem.

Theorem 2 (Identifiability under MA)**.** *Under Assumption 12, given experimental distributions $\mathbb{P}(Y_x)$, $\mathbb{P}(Y_0, \ldots, Y_{d_X-1})$ is identifiable as*

$$\begin{aligned} &\mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0) \\ &= \sum\nolimits_{j=0}^{y_0} \mathbb{P}(Y_{d_X-1} = j) - \sum\nolimits_{j=0}^{y_0-1} \mathbb{P}(Y_0 = j) \end{aligned} \tag{16}$$

for $y_0 \in [d_Y]$, and

$$\begin{aligned} &\mathbb{P}(Y_0 = \cdots = Y_k = y_0,\ Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1) \\ &= \sum\nolimits_{j=0}^{y_0} \mathbb{P}(Y_k = j) - \sum\nolimits_{j=0}^{y_0} \mathbb{P}(Y_{k+1} = j) \end{aligned} \tag{17}$$

for $k \in [d_X - 1]$ and $y_0 \in [d_Y - 1]$; otherwise,

$$\mathbb{P}(Y_0 = y_0, \ldots, Y_{d_X-1} = y_{d_X-1}) = 0. \tag{18}$$

Then, any $\lambda'(\mathcal{P}')$ is identifiable.

Observational Data. Next, we discuss the identification of the joint POs/OVs probabilities, i.e., $\mathbb{P}(Y_0, \ldots, Y_{d_X-1}, X, Y)$, and their linear combinations, using observational data. This requires exogeneity (Assumption 1), i.e. $Y_x \perp\!\!\!\perp X$, which implies $\mathbb{P}(Y_x) = P(Y|x)$.

We have the following:

Theorem 3 (Identifiability under MA)**.** *Under Assumptions 1 and 12, given observational distributions $\mathbb{P}(X, Y)$, $\mathbb{P}(Y_0, \ldots, Y_{d_X-1}, X, Y)$ is identifiable as*

$$\begin{aligned} &\mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0, X = x, Y = y) = \\ &\left(\sum_{m=0}^{y_0} \mathbb{P}(Y = m|X = d_X - 1) - \sum_{m=0}^{y_0-1} \mathbb{P}(Y = m|X = 0)\right) \\ &\times \mathbb{P}(X = x) \end{aligned} \tag{19}$$

for $y_0 \in [d_Y]$, $x \in [d_X]$, and $y = y_0$, and

$$\begin{aligned} &\mathbb{P}(Y_0 = \cdots = Y_k = y_0, \\ &\quad Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1, X = x, Y = y) \\ &= \left(\sum_{l=0}^{y_0} \mathbb{P}(Y = l|X = k) - \sum_{l=0}^{y_0} \mathbb{P}(Y = l|X = k+1)\right) \\ &\times \mathbb{P}(X = x) \end{aligned} \tag{20}$$

for $k \in [d_X - 1]$, $y_0 \in [d_Y - 1]$, $y = y_0$ for $x \leq k$, and $y = y_0 + 1$ for $x = k + 1, \ldots, d_X - 1$; otherwise,

$$\mathbb{P}(Y_0 = y_0, \ldots, Y_{d_X-1} = y_{d_X-1}, X = x, Y = y) = 0. \tag{21}$$

Then, any $\lambda(\mathcal{P})$ is identifiable.

Remark. We have the following characterization of identification (when restricted to linear constraints):

- Given experimental distributions $\mathbb{P}(Y_x)$, $\mathbb{P}(Y_0, \ldots, Y_{d_X-1})$ is identifiable if and only if we have $d_Y^{d_X} - d_X(d_Y - 1) - 1$ additional linearly independent constraints.
- Under Assumption 1, given observational distributions $\mathbb{P}(X, Y)$, $\mathbb{P}(Y_0, \ldots, Y_{d_X-1}, X, Y)$ is identifiable if and only if we have $d_X(d_Y^{d_X} - \{d_Y + (d_X - 1)(d_Y - 1)\})$ additional linearly independent constraints.

Assumption 12 imposes the minimal number of additional linear constraints needed to achieve identification in the sense that it provides exactly the required number of additional linearly independent constraints in both the experimental and observational settings.

6.2 Estimating Joint POs/OVs Probabilities and Their Linear Combinations

We estimate the joint POs/OVs probabilities and their linear combinations from finite samples by plugging the empirical probabilities of $\mathbb{P}(Y_x)$ and $\mathbb{P}(X, Y)$ into the identification results in Theorems 2 and 3. These are consistent estimators. A detailed discussion is given in Appendix D.4.

Numerical Experiments. We conduct numerical experiments to illustrate properties of the estimates under identification assumptions. All details are stated in Appendix

E.2. We simulated experimental and observational datasets and estimated $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$, $\mathbb{E}[Y_1 - Y_0 | X = 2, Y = 2]$, and $\mathbb{E}[(Y_1 - Y_0)^2]$. The results are reported in Tables 23-25 in Appendix E.2. We observe that all 95% CIs of the estimates become narrower as the sample size increases, and that the means of the estimators are very close to the ground truth when the sample size exceeds 100.

7 APPLICATION TO REAL-WORLD DATASETS

7.1 Experimental Data

Dataset. We use the Project STAR randomized trial [Stock and Watson, 2007], publicly available at (https://search.r-project.org/CRAN/refmans/AER/html/STAR.html). The experiment was conducted in Tennessee in the late 1980s across 79 schools, and the experiment randomized students into three class-size treatments: regular classes with one teacher ("regular"; $X = 0$), regular classes with a teacher and a teacher's aide ("regular+aide"; $X = 1$), and small classes ("small"; $X = 2$). Teachers and students were randomly assigned, and outcomes were standardized math and reading scores. After removing missing values, the sample size is 5967. Previous analyses of STAR by Stock and Watson [2007] reported that smaller classes have positive effects on academic performance.

Analysis. We take the sum of third-grade math and reading scores as the outcome Y. We group the scores into quartiles based on their empirical distribution, $[1009, 1180]$, $[1181, 1230]$, $[1231, 1282]$, and $[1283, 1527]$, and label them 0, 1, 2, and 3, respectively. We conduct 100 times bootstrapping [Efron, 1979] to provide the mean and 95% confidence interval (CI) for the estimator.

Assumptions. We assume the following MAs:

(A1). No MAs,

(A2). $Y_0 \leq Y_1$ holds a.s. (Assumption 9),

(A3). $Y_1 \leq Y_2$ holds a.s. (Assumption 9),

(A4). $\mathbb{P}(Y_0 \leq Y_1) \geq 0.95$ and $\mathbb{P}(Y_1 \leq Y_2) \geq 0.95$ (Assumption 11),

(A5). $0 \leq Y_i - Y_j \leq 1$ holds for all $i > j$ a.s. (Assumption 12).

(A2) states that, for almost all students, the potential score if assigned to a regular class with both a teacher and a teacher's aide is greater than the potential score if assigned to a regular class with only a teacher. (A3) states that, for almost all students, the potential score if assigned to a small class is greater than the potential score if assigned to a regular class with both a teacher and a teacher's aide. (A4) combines (A2) and (A3) and permits a 5% violation of each condition. (A5) states that, for almost all students, increasing the treatment level never decreases the score level, and can raise it by at most one level. By Theorem 2, the joint probabilities of POs are identifiable under Assumption (A5). These MAs may be regarded as reasonable.

Table 2: The means and 95% CIs of the estimates of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 \geq 1)$. LB stands for "lower bound" and UB stands for "upper bound."

Assumption	LB	UB
(A1)	0.000 [0.000,0.000]	0.261 [0.247,0.274]
(A2)	0.064 [0.042,0.089]	0.261 [0.247,0.274]
(A3)	0.000 [0.000,0.000]	0.064 [0.042,0.089]
(A4)	0.014 [0.000,0.039]	0.114 [0.092,0.139]
Assumption	Identification	
(A5)	0.064 [0.042,0.089]	

Results. We first estimate the mean POs (directly from the experimental data) with 95% CIs and obtain

$$\mathbb{E}[Y_0]\text{: 1228.725 [1225.898,1231.136]}$$
$$\mathbb{E}[Y_1]\text{: 1228.357 [1225.792,1231.461]}$$
$$\mathbb{E}[Y_2]\text{: 1244.199 [1240.400,1248.306]}$$

The results show that small classes lead to slightly higher average scores. Then one may be motivated to implement small class sizes.

We next evaluate the probability $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 \geq 1)$, which denotes the probability of the following counterfactual scenario:

"*A student would score in the bottom 25% if placed in a regular class with either a teacher alone or with both a teacher and a teacher's aide, and would score above the bottom 25% if enrolled in a small class.*"

This is a situation in which being assigned to a small class is necessary and sufficient to obtain a score above the bottom 25%.

Table 2 shows that this probability lies between 0% and 26.1% (without making any assumptions), has a higher lower bound 6.4% under assumption (A2), a tighter upper bound 6.4% under (A3), a narrower interval of [1.4%,11.4%] under (A4), and is point-identified at 6.4% under (A5). The results mean that, for instance, if one believes that assumption (A3) or (A5) is plausible, then only for a small number of subjects (no more than 6.4%), assignment to a small class is both necessary and sufficient to achieve a score above the bottom 25%. This finding may discourage one from implementing small classes. Note that in experimental results such as Table 2, the bootstrap distribution of the estimator may become distorted when the true probability is close to 0 or 1 due to boundary problems.

7.2 Observational Data

Dataset. We use the dataset introduced in Section 2.

Cortez and Silva [2008], Helwig [2017] focused on predicting students' academic performance from their attributes. Kawakami et al. [2024] assessed the causal relationship between student performance, study time, and extra paid classes by estimating PoC. They treat the outcome as a continuous variable and focus only on identifying the PoC. Kawakami and Tian [2025b] further performed mediation analysis of the PoC.

Analysis. The treatment (X) and outcome (Y) variables are defined in Section 2. We restrict our analysis to the subpopulation defined by the covariates (sex = F, schoolsup = no, famsup = yes). We assume that these covariates cover all the confounders such that the exogeneity (Assumption 1) holds in this subpopulation. The sample size of this subpopulation is 219. We conduct 100 times bootstrapping to provide the mean and 95% CI for the estimator.

Assumptions. We assume the following MAs:

(B1). No MAs,

(B2). $Y_0 \leq Y_1$ holds a.s. (Assumption 9),

(B3). $Y_2 \leq Y_3$ holds a.s. (Assumption 9),

(B4). $\mathbb{P}(Y_0 \leq Y_1) \geq 0.995$ and $\mathbb{P}(Y_2 \leq Y_3) \geq 0.995$ (Assumption 11),

(B5). $0 \leq Y_i - Y_j \leq 1$ holds for all $i > j$ a.s. (Assumption 12).

(B2) (resp. (B3)) states that, for almost all students, the potential score if they studied 2 to 5 hours (resp. more than 10 hours) is greater than the potential score if they studied less than 2 hours (resp. 5 to 10 hours). (B4) combines (B2) and (B3) and permits a 0.5% violation of each condition. (B5) states that, for almost all students, increasing the study time alone never decreases the score level and can raise it by at most one level. By Theorem 3, the joint POs/OVs probabilities and their linear combinations are identifiable under (B5) and the exogeneity assumption. These MAs may be regarded as reasonable.

Results. First, we estimate the POs based on $\mathbb{E}[Y_x] = \mathbb{E}[Y|x]$ by the exogeneity assumption. The mean estimates with 95% CIs are as follows:

$\mathbb{E}[Y_0]$: 11.154 [10.662,11.663]
$\mathbb{E}[Y_1]$: 12.457 [11.940,12.959]
$\mathbb{E}[Y_2]$: 12.979 [12.357,13.708]
$\mathbb{E}[Y_3]$: 15.732 [14.220,16.909]

These estimates indicate that the average score increases as the level of study time increases. The results may motivate students to study more in order to obtain higher scores.

We then evaluate $\mathbb{P}(Y_0 \leq 1, Y_1 \leq 1, Y_2 \leq 1|X = 3, Y = 2)$, whose interpretation is stated in Section 2. This is a situation in which studying more than 10 hours is necessary to obtain a score above 15.

Table 3: The means and 95% CIs of the estimates of $\mathbb{P}(Y_0 \leq 1, Y_1 \leq 1, Y_2 \leq 1|X = 3, Y = 2)$ under exogeneity.

Assumption	LB	UB
(B1)	0.203 [0.000,0.489]	0.997 [0.956,1.000]
(B2)	0.293 [0.000,0.561]	0.997 [0.956,1.000]
(B3)	0.203 [0.000,0.489]	0.575 [0.192,0.740]
(B4)	0.207 [0.000,0.489]	0.783 [0.531,1.000]
Assumption	Identification	
(B5)	0.575 [0.192,0.740]	

Table 4: The means and 95% CIs of the estimates of $\mathbb{P}(Y_1 \geq 1, Y_2 \geq 1, Y_3 \geq 1|X = 0, Y = 0)$ under exogeneity.

Assumption	LB	UB
(B1)	0.044 [0.000,0.280]	1.000 [1.000,1.000]
(B2)	0.044 [0.000,0.280]	0.251 [0.013,0.612]
(B3)	0.044 [0.000,0.280]	1.000 [1.000,1.000]
(B4)	0.044 [0.000,0.280]	0.404 [0.165,0.665]
Assumption	Identification	
(B5)	0.251 [0.013,0.612]	

Table 3 reports the estimates under exogeneity. (B2), (B3), and (B4) yield tighter bounds than without MAs (B1). A combination of (B5) and exogeneity yields a point estimate of 57.5%, indicating that, for 57.5% of the students, studying more than 10 hours is necessary to obtain a score above 15. This may be strong evidence that studying long hours is necessary to obtain a high score for half of the students, and may support the view that students' study time is not wasted. We note that this finding is not captured by the average results about $\mathbb{E}[Y_x]$.

We also evaluate $\mathbb{P}(Y_1 \geq 1, Y_2 \geq 1, Y_3 \geq 1|X = 0, Y = 0)$, whose interpretation is stated in Section 2. This is a situation in which studying at least 2 hours is sufficient to obtain a score above 8.

Table 4 reports the estimates under exogeneity. (B2) and (B4) yield tighter bounds than without MAs (B1), while the assumption (B3) does not tighten the bounds. A combination of (B5) and exogeneity yields a point estimate of 25.1%, indicating that 25.1% of the students who studied less than 2 hours and obtained a score of no more than 8 would have obtained a score above 8 had they studied more than 2 hours. That is, assuming (B5) and exogeneity, studying at least 2 hours is sufficient to obtain a score above 8 for a quarter of the students. This finding is not captured by the average results about $\mathbb{E}[Y_x]$.

Additionally, we conduct evaluations under each MA without the exogeneity assumption. All bounds without exogeneity are uninformative, ranging from 0% to 100%.

8 CONCLUSION

We provide methods for the bounding and identification of joint probabilities of potential outcomes and observed variables in settings involving a discrete treatment and an ordinal outcome. We introduce a broad class of flexible monotonicity assumptions, which empower researchers to integrate nuanced domain knowledge into the analysis. Our approach derives bounds that are provably as tight as the available domain knowledge allows. This bounding methodology is particularly valuable in practice because it yields informative results under plausible assumptions based on prior knowledge. The tools developed here move beyond standard Average Causal Effects (ACEs), providing a rigorous framework for evaluating a broad class of nuanced causal quantities. This includes families of effects like the Probability of Causation (PoC) and moments of causal effects, which ultimately offer crucial, valuable support for informed decision-making across various applied fields.

Balke [1995], Sachs et al. [2023] derived closed-form symbolic bounds for causal effects or joint POs probabilities involving binary variables. However, their algorithm becomes computationally infeasible when the domain sizes d_X and d_Y get larger. We therefore adopt a numerical approach to solve the bounding problem. Our source code is available at (https://github.com/Nmath997/AISTATS2026_bounding.git).

If additional variables such as covariates or mediators are included, tighter bounds can be obtained, as shown in [Kuroki and Cai, 2011, Dawid et al., 2016]. Extending our results to incorporate covariates or mediators represents an interesting direction for future work. However, such an extension would involve a larger number of parameters and may become computationally difficult to solve.

One limitation of this work is that we do not incorporate constraints implied by causal graphs. The problem of bounding counterfactual probabilities in a general causal graph with discrete variables has been studied in [Cozman, 2000, Zhang et al., 2022, Zaffalon et al., 2020, 2024, Duarte et al., 2024], including multi-linear programming for causal bounds in quasi-Markovian models [Shridharan and Iyengar, 2023a,b, Arroyo et al., 2025]. Additionally, Maiti et al. [2025] extended the binary MA to identify counterfactual probabilities in causal graphs. Investigating MAs for bounding counterfactual probabilities in a general causal graph is a future research direction.

Acknowledgements

The authors thank the anonymous reviewers for their time and thoughtful comments.

References

Joshua D. Angrist and Guido W. Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442, 1995.

João Pedro Arroyo, Denis Mauá, João Gabriel Barreto Rodrigues, Daniel AE Lawand, Junkyu Lee, Radu Marinescu, Alexander G Gray, Eduardo Rocha Laurentino, and Fabio Cozman. Multilinear and linear programs for partially identifiable queries in quasi-markovian structural causal models. In *UAI 2025 Workshop on Causal Abstractions and Representations*, 2025.

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

Alexander Abraham Balke. *Probabilistic counterfactuals: semantics, computation, and applications*. University of California, Los Angeles, 1995.

Elias Bareinboim, Juan D. Correa, Duligur Ibeling, and Thomas Icard. *On Pearl's Hierarchy and the Foundations of Causal Inference*, page 507–556. Association for Computing Machinery, New York, NY, USA, 1 edition, 2022.

P. Cortez and A. M. Gonçalves Silva. Using data mining to predict secondary school student performance. In *Proceedings of 5th Annual Future Business Technology Conference*, pages 5–12, 2008.

Fabio G Cozman. Credal networks. *Artificial intelligence*, 120(2):199–233, 2000.

A Philip Dawid, Rossella Murtas, and Monica Musio. Bounding the probability of causation in mediation analysis. In *Topics on Methodological and Applied Statistical Inference*, pages 75–84. Springer, 2016.

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

B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26, 1979.

Nathaniel E Helwig. Adding bias to reduce variance in psychological results: A tutorial on penalized regression. *The Quantitative Methods for Psychology*, 13(1):1–19, 2017.

M.A. Hernan and J.M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC Monographs on Statistics & Applied Probab. CRC Press, 2024.

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

Narendra Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 302–311, 1984.

Yuta Kawakami and Jin Tian. Moments of causal effects. *Proceedings of the 41st Conference on Uncertainty in Artificial Intelligence (UAI-2025)*, 2025a.

Yuta Kawakami and Jin Tian. Decomposition of probabilities of causation with two mediators. *Proceedings of the 41st Conference on Uncertainty in Artificial Intelligence (UAI-2025)*, 2025b.

Yuta Kawakami, Manabu Kuroki, and Jin Tian. Probabilities of causation for continuous and vector variables. *Proceedings of the 40th Conference on Uncertainty in Artificial Intelligence (UAI-2024)*, 2024.

Manabu Kuroki and Zhihong Cai. Statistical analysis of 'probabilities of causation' using co-variate information. *Scandinavian Journal of Statistics*, 38(3):564–577, 2011.

Ang Li and Judea Pearl. Probabilities of causation with nonbinary treatment and effect. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(18): 20465–20472, Mar. 2024.

Ting Li, Chengchun Shi, Jianing Wang, Fan Zhou, and Hongtu Zhu. Optimal treatment allocation for efficient policy evaluation in sequential decision making. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2024. Curran Associates Inc.

Zitong Lu, Zhi Geng, Wei Li, Shengyu Zhu, and Jinzhu Jia. Evaluating causes of effects by posterior effects of causes. *Biometrika*, 110(2):449–465, 07 2022.

Aurghya Maiti, Drago Plecko, and Elias Bareinboim. Counterfactual identification under monotonicity constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39(25), pages 26841–26850, 2025.

Charles Manski. Monotone treatment response. *Econometrica*, 65(6):1311–1334, 1997.

Scott Mueller. *Personalized and Situation-Specific Decision Making*. University of California, Los Angeles, 2025.

Scott Mueller and Judea Pearl. Monotonicity: Detection, refutation, and ramification, 2023.

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

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition, 2009.

Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic books, 2018.

AHG RinnooyKan and Jan Telgen. The complexity of linear programming. *Statistica Neerlandica*, 35(2):91–107, 1981.

Michael C Sachs, Gustav Jonzon, Arvid Sjölander, and Erin E Gabriel. A general method for deriving tight symbolic bounds on causal effects. *Journal of Computational and Graphical Statistics*, 32(2):567–576, 2023.

Madhumitha Shridharan and Garud Iyengar. Causal bounds in quasi-markovian graphs. In *International Conference on Machine Learning*, pages 31675–31692. PMLR, 2023a.

Madhumitha Shridharan and Garud Iyengar. Scalable computation of causal bounds. *Journal of Machine Learning Research*, 24(237):1–35, 2023b.

James H Stock and Mark W Watson. Econometrics, 2007.

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

Robert J. Vanderbei. *Linear programming. Foundations and extensions*, volume 37 of *Int. Ser. Oper. Res. Manag. Sci.* Dordrecht: Kluwer Academic Publishers, 2nd ed. edition, 2001.

Marco Zaffalon, Alessandro Antonucci, and Rafael Cabañas. Structural causal models are (solvable by) credal networks. In *International Conference on Probabilistic Graphical Models*, pages 581–592. PMLR, 2020.

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

Chao Zhang, Zhi Geng, Wei Li, and Peng Ding. Identifying and bounding the probability of necessity for causes of effects with ordinal outcomes. *Biometrika*, page asaf049, 2025.

Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 26548–26558. PMLR, 17–23 Jul 2022.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]

(c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes] The source code is available at (https://github.com/Nmath997/AISTATS2026_bounding.git).

2. For any theoretical claim, check if you include:

   (a) Statements of the full set of assumptions of all theoretical results. [Yes] All theorems are presented with their corresponding assumptions.

   (b) Complete proofs of all theoretical results. [Yes]

   (c) Clear explanations of any assumptions. [Yes] In Appendix C.

3. For all figures and tables that present empirical results, check if you include:

   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]

   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes] We have provided 95% confidence intervals for all reported estimates.

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] We use a computer with AMD Ryzen 7 8840U 3.30GHz processor and 16GB of RAM.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. [Not Applicable]

   (b) The license information of the assets, if applicable. [Not Applicable]

   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

   (d) Information about consent from data providers/curators. [Not Applicable]

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Appendix to "Bounds and Identification of Joint Probabilities of Potential Outcomes and Observed Variables under Monotonicity Assumptions"

A DETAILS OF MONOTONICITY ASSUMPTIONS IN THE PREVIOUS WORKS

In this appendix, we provide details of the monotonicity assumptions used in previous works.

(**Assumption 2**.) For binary treatment and outcome ($d_X = d_Y = 2$), Pearl [1999] and Tian and Pearl [2000] established bounds for PoC without additional assumptions and provided an identification theorem under the exogeneity assumption, i.e., $Y_x \perp\!\!\!\perp X$ for any $x \in \{0, 1\}$, combined with the monotonicity assumption, i.e.,

$$\mathbb{P}(Y_0 = 1, Y_1 = 0) = 0. \tag{22}$$

The monotonicity is equivalently stated as

$$\mathbb{P}(Y_0 \leq Y_1) = 1, \tag{23}$$

or

$$Y_0 \leq Y_1, \text{ almost surely}, \tag{24}$$

where a statement holds almost surely if it is true with probability one. This condition represents that the POs are monotonically increasing with the treatment level for almost every subject.

The monotonicity assumption is a natural and reasonable assumption in a wide range of empirical studies [Hernan and Robins, 2024]. In the instrumental variables (IV) literature, the monotonicity assumption is often invoked to describe the relationship between binary IV (Z) and treatment (X) in the identification of local average treatment effects (LATE) [Imbens and Angrist, 1994, Angrist and Imbens, 1995]. They provide an identification theorem for PoC under the proposed monotonicity condition together with the exogeneity assumption.

(**Assumptions 3 and 4**.) Zhang et al. [2025] studied PN for a binary treatment and ordinal outcome, i.e., $\mathbb{P}(Y_0 < y | X = 1, Y = y)$, and derived bounds for PN under the following monotonicity assumption:

$$Y_0 \leq Y_1, \text{ almost surely}. \tag{25}$$

This is a straightforward extension of the monotonicity assumption in [Pearl, 1999, Tian and Pearl, 2000]. They provided closed-form bounds for PN under this assumption.

This paper also introduced an assumption termed the monotonic incremental treatment effect (MITE) assumption for identifying PN, which states

$$0 \leq Y_1 - Y_0 \leq 1, \text{ almost surely}. \tag{26}$$

This assumption is stronger than Assumption 3. This assumption serves as the identification condition for PN. This assumption states that the treatment X does not decrease the level of Y and increases it by at most one level. They established an identification lemma for $\mathbb{P}(Y_1, Y_0 | X = 1)$ given $\mathbb{P}(Y_0 | X = 1)$ under the above monotonicity condition. Then, they established the identification theorem for PN through the lemma.

(**Assumption 5**.) The relaxed assumption, i.e., ϵ-limited harm assumption, for binary treatment and outcome by Mueller and Pearl [2023] states that:

$$0 \leq \mathbb{P}(Y_0 = 1, Y_1 = 0) \leq \epsilon. \tag{27}$$

This assumption relaxes the monotonicity assumption in [Pearl, 1999, Tian and Pearl, 2000] by allowing violations in which subjects have POs $Y_0 = 1$ and $Y_1 = 0$ for at most an ϵ proportion of the population. This is equivalent to $l \leq \mathbb{P}(Y_0 \leq Y_1) \leq 1$, where $l = 1 - \epsilon$. They provided closed-form bounds for PNS with this assumption. When $\epsilon = 0$ or $l = 1$, Assumption 5 reduces to Assumption 2.

(**Assumption 6**.) The above studies have been limited to settings with binary treatment. Manski [1997] has introduced the following assumption:

$$Y_0 \leq Y_1 \leq \cdots \leq Y_{d_X-1}, \text{ almost surely.} \tag{28}$$

This condition implies that the POs are monotonically increasing with the treatment level for almost every subject. This is equivalently stated as $\mathbb{P}(Y_0 \leq Y_1 \leq \cdots \leq Y_{d_X-1}) = 1$. Their target is the average causal effect (ACE) for discrete, continuous, or mixed outcomes. The joint POs/OVs probabilities are not their target.

B CAUSAL QUANTITIES REPRESENTED BY LINEAR COMBINATIONS OF JOINT POs/OVs PROBABILITIES

In this appendix, we show more examples that can be expressed as linear combinations of $\mathcal{P}$ (Section 5.1).

PoC for discrete treatment and outcome. Recently, Li and Pearl [2024] have introduced new families of PoC for non-binary discrete treatment—including the probability of preservation, the probability of replacement, the probability of substitution, the probability of necessity, and the probability of necessity and sufficiency—extending the classical PoC framework to nonbinary and discrete treatments and outcomes, using the joint probabilities of POs and OVs $\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{d_X-1} = y_{d_X-1}, X = x, Y = y)$.

PoC with a single hypothetical term for non-binary treatment and outcome is given as

(i) The probability of preservation $\mathbb{P}(Y_j = y, Y = y)$,

(ii) The probability of replacement $\mathbb{P}(Y_j = y_j, Y = y)$ $(y_j \neq y)$,

(iii) The probability of substitute $\mathbb{P}(Y_j = y_j, X = x)$ $(x \neq j)$, and

(iv) The probability of necessity $\mathbb{P}(Y_j = y_j, Y = y, X = x)$ $(x \neq j)$.

PoC with multiple hypothetical terms for non-binary treatment and outcome are given as

(i) The probability of necessity and sufficiency $\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{d_X-1} = y_{d_X-1})$,

(ii) The probability of substitute $\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{d_X-1} = y_{d_X-1}, X = x)$,

(iii) The probability of replacement $\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{d_X-1} = y_{d_X-1}, Y = y)$, and

(iv) The probability of necessity $\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{d_X-1} = y_{d_X-1}, X = x, Y = y)$.

They are given as marginals of the joint POs/OVs probabilities. Li and Pearl [2024] provided closed forms of their bounds without additional assumptions, and the bounds for PoC with a single hypothetical term are sharp.

Joint expectation of a function over POs and OVs. The joint expectation of a function over POs and OVs can be expressed as linear combinations of $\mathcal{P}$:

$$\begin{aligned}&\mathbb{E}[g(Y_0, \ldots, Y_{d_X-1}, X, Y)] \\ &= \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \sum_{y=0}^{d_Y-1} g(y_0, \ldots, y_{d_X-1}, x, y) \times \mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{d_X-1} = y_{d_X-1}, X = x, Y = y).\end{aligned} \tag{29}$$

For example, the moments of causal effects [Kawakami and Tian, 2025a], such as

$$\begin{aligned}&\mathbb{E}[(Y_1 - Y_0)^m] \\ &= \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \sum_{y=0}^{d_Y-1} (y_1 - y_0)^m \times \mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{d_X-1} = y_{d_X-1}, X = x, Y = y),\end{aligned} \tag{30}$$

and the product moments of causal effects [Kawakami and Tian, 2025a], such as

$$\begin{aligned}&\mathbb{E}[(Y_1 - Y_0)(Y_2 - Y_1)] \\ &= \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \sum_{y=0}^{d_Y-1} (y_1 - y_0)(y_2 - y_1) \times \mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{d_X-1} = y_{d_X-1}, X = x, Y = y).\end{aligned} \tag{31}$$

can be expressed as linear combinations of $\mathcal{P}$.

When $\mathbb{P}(X = l, Y = m)$ is given, the conditional probabilities given X and Y can be expressed as the following linear relationship:

$$\mathbb{P}(Y_{[d_X]} = y_{[d_X]} | X = l, Y = m) = \frac{\mathbb{P}(Y_{[d_X]} = y_{[d_X]}, X = l, Y = m)}{\mathbb{P}(X = l, Y = m)}, \tag{32}$$

and this quantity is also bounded if the joint POs/OVs probabilities are bounded, e.g., PN [Zhang et al., 2025].

Furthermore, when $\mathbb{P}(X = l, Y = m)$ is given, the joint expectation of a function λ over POs, conditional on X and Y, i.e., $\mathbb{E}[g(Y_0, \ldots, Y_{d_X-1}, X, Y)|X = l, Y = m]$, can be expressed as the following linear relationship:

$$\begin{aligned}&\mathbb{E}[g(Y_0, \ldots, Y_{d_X-1}, X, Y)|X = l, Y = m] \\ &= \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} g(y_0, \ldots, y_{d_X-1}, l, m) \times \frac{\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{d_X-1} = y_{d_X-1}, X = l, Y = m)}{\mathbb{P}(X = l, Y = m)}.\end{aligned} \tag{33}$$

The causal effect given $(X = l, Y = m)$,

$$\begin{aligned}&\mathbb{E}[Y_x - Y_{x'}|X = l, Y = m] \\ &= \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} (y_x - y_{x'}) \times \frac{\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{d_X-1} = y_{d_X-1}, X = l, Y = m)}{\mathbb{P}(X = l, Y = m)}.\end{aligned} \tag{34}$$

is a member of the class of posterior causal effects, a concept introduced by Lu et al. [2022], who define such effects as expectations conditional on evidence observed for a post-treatment variable. It evaluates the counterfactual average causal effect of changing the intervention from $X = x'$ to $X = x$ for subjects who, in reality, had $Y = m$ under $X = l$.

C PROOFS

In this section, we provide the proofs of the propositions and theorems stated in the main body of the paper.

Proposition 1. *We have the following relationships among the MAs presented thus far:*

(1) Assumptions 2 through 10 are all special cases of Assumption 11.

(2) Assumption 4 is a special case of Assumptions 7, 9, and 10.

(3) Assumption 5 is a special case of Assumptions 7, 8, and 10.

(4) Assumption 6 is a special case of Assumptions 8, 9, and 10.

Proof. First, we recall that Assumptions 2 and 3 are special cases of Assumption 4 (Appendix A), and that Assumption 10 is a special case of Assumption 11 (Section 4). We also remark that Assumptions 7, 8, 9 are special cases of Assumption 10. Indeed, Assumption 10 reduces to Assumption 7 (resp.8, 9) when $d_X = 2$ and $\underline{D}_{st} = \underline{D}$ (resp. when $\underline{D}_{st} = -\infty$, $\overline{D}_{st} = \infty$ for all $s > t$ with $s = t + 1$, when $L = U = 1$).

Then we can prove each assertion of the proposition as follows:

(2): We obtain Assumption 4 from Assumption 7 (resp. 9) by putting $\underline{D} = 0, \overline{D} = 1, L = -\infty, U = \infty$ (resp. by setting $d_X = 2$ and putting $\underline{D}_{10} = 0, \overline{D}_{10} = 1$). Since these are special cases of Assumption 10, the assertion (2) holds.

(3): We obtain Assumption 5 from Assumption 7 (resp. 8) by setting $d_Y = 2$ and putting $\underline{D} = 0, \overline{D} = 1, L = 0, U = \epsilon$ (resp. by setting $d_X = d_Y = 2$ and putting $L = 1 - \epsilon, U = 1$). Since these are special cases of Assumption 10, the assertion (3) holds.

(4): We obtain Assumption 6 from Assumption 8 (resp. 9) by putting $L = -\infty, U = \infty$ (resp. by putting $\underline{D}_{st} = -\infty$, $\overline{D}_{st} = \infty$ for all $s > t$ with $s = t + 1$). Since these are special cases of Assumption 10, the assertion (4) holds.

Now we also have the assertion (1) from (2)–(4) and remarks in the beginning of this proof. □

Proposition 2. *The constraints implied by Assumption 11 are given by*

$$L^w \leq \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \left(\prod_{s,t\in[d_X];s>t} \mathbb{I}(\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)}) \right) p_{y_{[d_X]},x} \leq U^w \tag{35}$$

for $w = 1, \ldots, W$. The above constraints are all linear constraints in $\mathcal{P}$.

Remark. $[d_X]$ represents the set $\{0, 1, \ldots, d_X - 1\}$.

Proof. We have the following relationships:

$$\begin{aligned} &\bigwedge_{s,t\in[d_X];s>t} (\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)}) \\ &\iff \underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)} \text{ holds for any } s, t(s > t) \\ &\iff \mathbb{I}(\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)}) = 1 \text{ for any } s, t(s > t) \\ &\iff \prod_{s>t} \mathbb{I}(\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)}) = 1. \end{aligned} \tag{36}$$

Then, the following relationship holds:

$$\begin{aligned} &\mathbb{P}\left(\bigwedge_{s,t\in[d_X];s>t} (\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)}) \right) \\ &= \mathbb{P}\left(\prod_{s>t} \mathbb{I}(\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)}) = 1 \right) \\ &= \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \left(\prod_{s>t} \mathbb{I}(\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)}) \right) \times p_{y_{[d_X]},x}. \end{aligned} \tag{37}$$

They are all linear constraints in $\mathcal{P}$ since $\prod_{s>t} \mathbb{I}(\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)})$ is a constant 0 or 1. Then, we have Proposition 2. □

Theorem 1. *Given data (experimental, observational, or both), under Assumption 11, solving the LP problem in* (13) *yields sharp bounds for $\lambda(\mathcal{P})$.*

Proof. In general, optima obtained by solving LP problems are global optima. In addition, since constraints from data and Eq. (12) restricts $\mathfrak{p}$ to a closed subset of $[0,1]^{d_Y^{d_X} d_X}$, LP (13) has the minimum $\underline{\lambda}(\mathcal{P})$ and the maximum $\overline{\lambda}(\mathcal{P})$. Then $[\underline{\lambda}(\mathcal{P}), \overline{\lambda}(\mathcal{P})]$ obtained by solving (12) is a bound of $\lambda(\mathcal{P})$, and it is sharp since $\underline{\lambda}(\mathcal{P})$ and $\overline{\lambda}(\mathcal{P})$ are achieved by some $\mathfrak{p}$ which satisfies constraints from data and Eq. (12). □

Theorem 2. *Under Assumption 12, given experimental distributions $\mathbb{P}(Y_x)$, $\mathbb{P}(Y_0, \ldots, Y_{d_X-1})$ is identifiable as*

$$\mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0) = \sum_{j=0}^{y_0} \mathbb{P}(Y_{d_X-1} = j) - \sum_{j=0}^{y_0-1} \mathbb{P}(Y_0 = j) \tag{38}$$

for $y_0 \in [d_Y]$, and

$$\mathbb{P}(Y_0 = \cdots = Y_k = y_0,\ Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1) = \sum_{j=0}^{y_0} \mathbb{P}(Y_k = j) - \sum_{j=0}^{y_0} \mathbb{P}(Y_{k+1} = j) \tag{39}$$

for $k \in [d_X - 1]$ and $y_0 \in [d_Y - 1]$; otherwise,

$$\mathbb{P}(Y_0 = y_0, \ldots, Y_{d_X-1} = y_{d_X-1}) = 0. \tag{40}$$

Then, any $\lambda'(\mathcal{P}')$ is identifiable.

Proof. Eq. (40) follows from Assumption 12. Denote

$$\theta_{y_0} = \mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0), \tag{41}$$

for $y_0 \in [d_Y]$, and

$$\varphi_{y_0,k} = \mathbb{P}(Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1) \tag{42}$$

for $y_0 \in [d_Y - 1], k \in [d_X - 1]$. Then, we have

$$\mathbb{P}(Y_i = j) = \theta_j + \sum_{k=0}^{i-1} \varphi_{j-1,k} + \sum_{k=i}^{d_X-2} \varphi_{j,k}, \tag{43}$$

for $i \in [d_X]$ and $j \in [d_Y]$, where $\varphi_{-1,k} = \varphi_{d_Y-1,k} = 0$. This implies

$$\mathbb{P}(Y_{i+1} = j) - \mathbb{P}(Y_i = j) = \varphi_{j-1,i} - \varphi_{j,i} \tag{44}$$

for $i \in [d_X - 1]$. Then, for each $k \in [d_X - 1]$, we can identify $\varphi_{y_0,k}$ as

$$\varphi_{y_0,k} = \sum_{j=0}^{y_0} (\varphi_{j,k} - \varphi_{j-1,k}) = \sum_{j=0}^{y_0} (\mathbb{P}(Y_k = j) - \mathbb{P}(Y_{k+1} = j)). \tag{45}$$

θ_{y_0} are identified as

$$\begin{aligned}
\theta_{y_0} &= \mathbb{P}(Y_0 = y_0) - \sum_{k=0}^{d_X-2} \varphi_{y_0,k} \\
&= \mathbb{P}(Y_0 = y_0) - \sum_{l=0}^{y_0} \sum_{k=0}^{d_X-2} (\mathbb{P}(Y_k = l) - \mathbb{P}(Y_{k+1} = l)) \\
&= \mathbb{P}(Y_0 = y_0) - \sum_{l=0}^{y_0} (\mathbb{P}(Y_0 = l) - \mathbb{P}(Y_{d_X-1} = l)) \\
&= \sum_{l=0}^{y_0} \mathbb{P}(Y_{d_X-1} = l) - \sum_{l=0}^{y_0-1} \mathbb{P}(Y_0 = l).
\end{aligned} \tag{46}$$

for $y_0 \in [d_Y]$. Then, we have Theorem 2. □

We have the following three lemmas for the proof of Theorem 3.

Lemma 1. *$\mathbb{P}(Y_0, \ldots, Y_{d_X-1}, X, Y)$ can be represented by $\mathbb{P}(Y_0, \ldots, Y_{d_X-1}|X)$ and $\mathbb{P}(X, Y)$ as*

$$\begin{aligned}
&\mathbb{P}(Y_0, \ldots, Y_{x-1}, Y_x = y_x, Y_{x+1}, \ldots, Y_{d_X-1}, X = x, Y = y) \\
&= \frac{\mathbb{P}(Y_0, \ldots, Y_{x-1}, Y_x = y_x, Y_{x+1}, \ldots, Y_{d_X-1}|X = x)}{\mathbb{P}(Y = y|X = x)} \times \mathbb{I}(y_x = y)\mathbb{P}(X = x, Y = y).
\end{aligned} \tag{47}$$

Proof. The joint probabilities $\mathbb{P}(Y_0, \ldots, Y_{d_X-1}, X, Y)$ can be represented by $\mathbb{P}(Y_0, \ldots, Y_{d_X-1}|X = x)$ and $\mathbb{P}(X, Y)$ as follows:

$$\begin{aligned}
&\mathbb{P}(Y_0, \ldots, Y_{x-1}, Y_x = y_x, Y_{x+1}, \ldots, Y_{d_X-1}, X = x, Y = y) \\
&= \mathbb{P}(Y_0, \ldots, Y_{x-1}, Y_x = y_x, Y_{x+1}, \ldots, Y_{d_X-1}|X = x, Y = y) \times \mathbb{P}(X = x, Y = y) \\
&= \frac{\mathbb{P}(Y_0, \ldots, Y_{x-1}, Y_x = y_x, Y_{x+1}, \ldots, Y_{d_X-1}, Y = y|X = x)}{\mathbb{P}(Y = y|X = x)} \times \mathbb{P}(X = x, Y = y) \\
&= \frac{\mathbb{P}(Y_0, \ldots, Y_{x-1}, Y_x = y_x, Y_{x+1}, \ldots, Y_{d_X-1}, Y_x = y|X = x)}{\mathbb{P}(Y = y|X = x)} \times \mathbb{P}(X = x, Y = y) \\
&= \frac{\mathbb{P}(Y_0, \ldots, Y_{x-1}, Y_x = y_x, Y_{x+1}, \ldots, Y_{d_X-1}|X = x)}{\mathbb{P}(Y = y|X = x)} \times \mathbb{I}(y_x = y)\mathbb{P}(X = x, Y = y).
\end{aligned} \tag{48}$$

Then, we have Lemma 1. □

Lemma 2. *Under Assumptions 12, with observational data,* $\mathbb{P}(Y_0, \dots, Y_{d_X-1}|X = x)$ *is given as*

$$\mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0|X = x) = \sum_{l=0}^{y_0} \mathbb{P}(Y_{d_X-1} = l|X = x) - \sum_{l=0}^{y_0-1} \mathbb{P}(Y_0 = l|X = x) \tag{49}$$

for $y_0 \in [d_Y]$ *and* $x \in [d_X]$*, and*

$$\mathbb{P}(Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1|X = x) = \sum_{j=0}^{y_0} (\mathbb{P}(Y_k = j|X = x) - \mathbb{P}(Y_{k+1} = j|X = x)) \tag{50}$$

for $k \in [d_X - 1]$*,* $y_0 \in [d_Y - 1]$*, and* $x \in [d_X]$*; otherwise,*

$$\mathbb{P}(Y_0 = y_0, \dots, Y_{d_X-1} = y_{d_X-1}|X = x) = 0. \tag{51}$$

Proof. Eq. (51) holds from Assumption 12. Denote

$$\theta^x_{y_0} = \mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0|X = x), \tag{52}$$

$$\varphi^x_{y_0,k} = \mathbb{P}(Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1|X = x) \tag{53}$$

for $k \in [d_X - 1]$, $y_0 \in [d_Y - 1]$, and $x \in [d_X]$. Then we have

$$\mathbb{P}(Y_i = j|X = x) = \theta^x_j + \sum_{k=0}^{i-1} \varphi^x_{j-1,k} + \sum_{k=i}^{d_X-2} \varphi^x_{j,k} \tag{54}$$

for $i \in [d_X]$ and $j \in [d_Y]$, where $\varphi^x_{-1,k} = \varphi^x_{d_Y-1,k} = 0$. This implies

$$\mathbb{P}(Y_{i+1} = j|X = x) - \mathbb{P}(Y_i = j|X = x) = \varphi^x_{j-1,i} - \varphi^x_{j,i} \tag{55}$$

for $i \in [d_X - 1]$. Then, for each $k \in [d_X - 1]$, we can rewrite $\varphi^x_{y_0,k}$ as

$$\varphi^x_{y_0,k} = \sum_{j=0}^{y_0} (\varphi^x_{j,k} - \varphi^x_{j-1,k}) = \sum_{j=0}^{y_0} (\mathbb{P}(Y_k = j|X = x) - \mathbb{P}(Y_{k+1} = j|X = x)). \tag{56}$$

$\theta^x_{y_0}$ are rewritten as

$$\begin{aligned}
\theta^x_{y_0} &= \mathbb{P}(Y_0 = y_0|X = x) - \sum_{k=0}^{d_X-2} \varphi^x_{y_0,k} \\
&= \mathbb{P}(Y_0 = y_0|X = x) - \sum_{l=0}^{y_0} \sum_{k=0}^{d_X-2} (\mathbb{P}(Y_k = l|X = x) - \mathbb{P}(Y_{k+1} = l|X = x)) \\
&= \mathbb{P}(Y_0 = y_0|X = x) - \sum_{l=0}^{y_0} (\mathbb{P}(Y_0 = l|X = x) - \mathbb{P}(Y_{d_X-1} = l|X = x)) \\
&= \sum_{l=0}^{y_0} \mathbb{P}(Y_{d_X-1} = l|X = x) - \sum_{l=0}^{y_0-1} \mathbb{P}(Y_0 = l|X = x)
\end{aligned} \tag{57}$$

for $y_0 \in [d_Y]$. □

Lemma 3. *Under Assumptions 1 and 12, with observational data,* $\mathbb{P}(Y_0, \dots, Y_{d_X-1}|X = x)$ *is identifiable as*

$$\mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0|X = x) = \sum_{m=0}^{y_0} \mathbb{P}(Y = m|X = d_X - 1) - \sum_{m=0}^{y_0-1} \mathbb{P}(Y = m|X = 0) \tag{58}$$

for $y_0 \in [d_Y]$ *and* $x \in [d_X]$*, and*

$$\mathbb{P}(Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1 | X = x) = \sum_{m=0}^{y_0} \mathbb{P}(Y = m | X = k) - \sum_{m=0}^{y_0} \mathbb{P}(Y = m | X = k+1) \tag{59}$$

for $k \in [d_X - 1]$*,* $y_0 \in [d_Y - 1]$*, and* $x \in [d_X]$*; otherwise,*

$$\mathbb{P}(Y_0 = y_0, \ldots, Y_{d_X-1} = y_{d_X-1} | X = x) = 0. \tag{60}$$

Remark. This lemma reduces to Lemma 1 in [Zhang et al., 2025] when $d_X = 2$ and $X = 1$.

Proof. Assumption 1 implies

$$\mathbb{P}(Y_j = l | X = x) = \mathbb{P}(Y_j = l | X = j) = \mathbb{P}(Y = l | X = j) \tag{61}$$

for arbitrary $j \in [d_X],\ x \in [d_X],\ l \in [d_Y]$. Hence, from Lemma 2, we have

$$\theta^x_{y_0} = \sum_{l=0}^{y_0} \mathbb{P}(Y = l | X = d_X - 1) - \sum_{l=0}^{y_0-1} \mathbb{P}(Y = l | X = 0) \tag{62}$$

and

$$\varphi^x_{y_0,k} = \sum_{l=0}^{y_0} \mathbb{P}(Y = l | X = k) - \sum_{l=0}^{y_0} \mathbb{P}(Y = l | X = k+1). \tag{63}$$

Then, we have Lemma 3. □

Theorem 3. *Under Assumptions 1 and 12, given observational distributions* $\mathbb{P}(X, Y)$*,* $\mathbb{P}(Y_0, \ldots, Y_{d_X-1}, X, Y)$ *is identifiable as*

$$\begin{aligned}\mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0, X = x, Y = y) = &\left(\sum_{m=0}^{y_0} \mathbb{P}(Y = m | X = d_X - 1) - \sum_{m=0}^{y_0-1} \mathbb{P}(Y = m | X = 0) \right) \\ &\times \mathbb{P}(X = x)\end{aligned} \tag{64}$$

for $y_0 \in [d_Y]$*,* $x \in [d_X]$*, and* $y = y_0$*, and*

$$\begin{aligned}&\mathbb{P}(Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1, X = x, Y = y) \\ &= \left(\sum_{l=0}^{y_0} \mathbb{P}(Y = l | X = k) - \sum_{l=0}^{y_0} \mathbb{P}(Y = l | X = k+1) \right) \mathbb{P}(X = x)\end{aligned} \tag{65}$$

for $k \in [d_X - 1]$*,* $y_0 \in [d_Y - 1]$*,* $y = y_0$ *for* $x \le k$*, and* $y = y_0 + 1$ *for* $x = k+1, \ldots, d_X - 1$*; otherwise,*

$$\mathbb{P}(Y_0 = y_0, \ldots, Y_{d_X-1} = y_{d_X-1}, X = x, Y = y) = 0. \tag{66}$$

Then, any $\lambda(\mathcal{P})$ *is identifiable.*

Proof. $\mathbb{P}(Y_0, \ldots, Y_{d_X-1}, X, Y)$ is categorized into the following cases:

$$\begin{cases} \text{(A): } Y_0 = \cdots = Y_{d_X-1} = y_0 & \text{for some } y_0 \in [d_Y]. \\ \text{(B): } Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1 & \text{for some } y_0 \in [d_Y - 1] \text{ and } k \in [d_X - 1]. \\ \text{(C): Otherwise.} \end{cases} \tag{67}$$

In case (C), we have $\mathbb{P}(Y_0, \ldots, Y_{d_X-1}, X, Y) = 0$. In case (A), we have $\mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0, X = x, Y = y) = 0$ if $y \neq y_0$. When $y = y_0$, by Lemmas 1 and 3 , we have

$$\begin{aligned}&\mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0, X = x, Y = y) \\ &= \frac{\mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0 | X = x)}{\mathbb{P}(Y = y | X = x)} \times \mathbb{I}(y_0 = y) \mathbb{P}(X = x, Y = y) \\ &= \mathbb{P}(Y_0 = \cdots = Y_{d_X-1} = y_0 | X = x) \mathbb{P}(X = x) \\ &= \left(\sum_{l=0}^{y_0} \mathbb{P}(Y = l | X = d_X - 1) - \sum_{l=0}^{y_0-1} \mathbb{P}(Y = l | X = 0) \right) \mathbb{P}(X = x).\end{aligned} \tag{68}$$

In case (B), by Lemmas 1 and 3, we have

$$\begin{aligned}&\mathbb{P}(Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1, X = x, Y = y)\\ &= \frac{\mathbb{P}(Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1|X = x)}{\mathbb{P}(Y = y|X = x)} \times \mathbb{I}(Y_x = y)\mathbb{P}(X = x, Y = y).\end{aligned} \tag{69}$$

If $x \leq k, y = y_0$ or $x \geq k + 1, y = y_0 + 1$, it can be calculated as

$$\mathbb{P}(Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1|X = x)\mathbb{P}(X = x) \tag{70}$$

$$= \left(\sum_{l=0}^{y_0} \mathbb{P}(Y = l|X = k) - \sum_{l=0}^{y_0} \mathbb{P}(Y = l|X = k + 1)\right) \mathbb{P}(X = x). \tag{71}$$

Otherwise, we have $\mathbb{P}(Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0+1, X = x, Y = y) = 0$. Then, we have Theorem 3. □

D TECHNICAL DETAILS

This section presents the technical details that support the method introduced in the main body of the paper.

D.1 Preliminaries on LPs

We provide some important properties of LP.

Formulation of LPs. LPs in our paper are formulated as

$$\text{maximize / minimize } c_1x_1 + \cdots + c_nx_n$$
$$\text{subject to } \begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \\ a_{t1}x_1 + \cdots + a_{tn}x_n = b_t \end{cases}, \quad \begin{cases} a'_{11}x_1 + \cdots + a'_{1n}x_n \leq b'_1 \\ \vdots \\ a'_{t'1}x_1 + \cdots + a'_{t'n}x_n \leq b'_{t'} \end{cases}, \quad x_1, \ldots, x_n \geq 0 \tag{72}$$

We can represent the same LP in one line by using matrices. Defining

$$c = (c_1 \ \ldots \ c_n)^T, \quad x = (x_1 \ \ldots \ x_n)^T,$$
$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{t1} & \cdots & a_{tn} \end{pmatrix}, \quad A' = \begin{pmatrix} a'_{11} & \cdots & a'_{1n} \\ \vdots & \ddots & \vdots \\ a'_{t'1} & \cdots & a'_{t'n} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_t \end{pmatrix}, \quad b = \begin{pmatrix} b'_1 \\ \vdots \\ b'_t \end{pmatrix}, \tag{73}$$

we can represent Eq. (72) by

$$\text{maximize /minimize } c^Tx \text{ subject to } Ax = b, A'x \leq b', x \geq 0. \tag{74}$$

Note that, for vectors $x = (x_1, \ldots x_n)$ and $y = (y_1, \ldots, y_n)$, $x \leq y$ represents $x_1 \leq y_1, \ldots, x_n \leq y_n$.

Reformulation of Linear Programming. Some references on LPs used in the proof discuss only equality constraints or only inequality constraints. Then, we show that mixed constraints can be represented as only equality constraints or only inequality constraints.

We can reformulate the inequality and equality constraints in our bounding problems as equality constraints. We can rewrite

$$\begin{cases} a'_{11}x_1 + \cdots + a'_{1u}x_n \leq b'_1 \\ \vdots \\ a'_{t'1}x_1 + \cdots + a'_{tu}x_n \leq b'_{t'} \end{cases}, \ x_1, \ldots, x_n \geq 0 \tag{75}$$

into an equivalent constraint system

$$\begin{cases} a'_{11}x_1 + \cdots + a'_{1n}x_n + w_1 = b'_1 \\ a'_{21}x_1 + \cdots + a'_{2n}x_n + w_2 = b'_2 \\ \vdots \\ a'_{t'1}x_1 + \cdots + a'_{tn}x_n + w_t = b'_{t'} \end{cases}, \ x_1, \ldots, x_n, w_1, \ldots, w_{t'} \geq 0. \tag{76}$$

We can also reformulate equality constraints

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \\ a_{t1}x_1 + \cdots + a_{tn}x_n = b_t \end{cases}, \ x_1, \ldots, x_n \geq 0 \tag{77}$$

into

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n \leq b_1 \\ \vdots \\ a_{t1}x_1 + \cdots + a_{tn}x_n \leq b_t \end{cases}, \quad \begin{cases} -a_{11}x_1 - \cdots - a_{1n}x_n \leq -b_1 \\ \vdots \\ -a_{t1}x_1 + \cdots - a_{tn}x_n \leq -b_t \end{cases}, \quad x_1, \ldots, x_n \geq 0, \tag{78}$$

since we have $x = b \iff x \leq b,\, b \leq x \iff x \leq b,\, -x \leq -b$.

Dual problem of Linear Programming. We use the dual problem of LPs in the proof and provide a brief explanation of LP duality. The dual problem of the LP

$$\text{maximize } c^T x \text{ subject to } Ax = b, A'x \leq b', x \geq 0 \tag{79}$$

is an LP

$$\text{minimize } b^T y + (b')^T z \text{ subject to } A^T y + (A')^T z \geq c, z \geq 0 \tag{80}$$

where y and z are vectors. If (79) is feasible and has an optimum (maximum), (80) is also feasible, the minimum of (80) exists and is equal to the maximum of (79).

D.2 Explicit constraints implied by Assumptions 6-10

The constraint by strong non-decreasing monotonicity (Assumption 6) is given as

$$\sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \mathbb{I}(y_0 \leq y_1 \leq \cdots \leq y_{d_X-1}) \times p_{y_0,\ldots,y_{d_X-1},x} = 1, \tag{81}$$

the constraint by weak non-decreasing monotonicity for binary treatment (Assumption 7) is given as

$$L \leq \sum_{y_0=0}^{d_Y-1} \sum_{y_1=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \mathbb{I}(y_0 \leq y_1) \times p_{y_0,y_1,x} \leq U, \tag{82}$$

the constraint by weak non-decreasing monotonicity (Assumption 8) is given as

$$L \leq \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \mathbb{I}(y_0 \leq y_1 \leq \cdots \leq y_{d_X-1}) \times p_{y_0,\ldots,y_{d_X-1},x} \leq U, \tag{83}$$

the constraint by single strong monotonicity (Assumption 9) is given as

$$\sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \left(\prod_{s>t} \mathbb{I}(\underline{D}_{st} \leq y_s - y_t \leq \overline{D}_{st}) \right) \times p_{y_0,\ldots,y_{d_X-1},x} = 1, \tag{84}$$

and the constraint by single weak monotonicity (Assumption 10) is given as

$$L \leq \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \left(\prod_{s>t} \mathbb{I}(\underline{D}_{st} \leq y_s - y_t \leq \overline{D}_{st}) \right) \times p_{y_0,\ldots,y_{d_X-1},x} \leq U. \tag{85}$$

D.3 Formulation of Our LPs in Matrix Form

Since the implementation of LP is formulated in matrix form, we discuss the formulation of our bounding problems in matrix form.

We denote the tensor of parameters as $\mathcal{T} \in \mathbb{R}^{d_Y^{d_X} \times d_X}$, where the element of $\mathcal{T}$ at index $(y_0, \dots, y_{d_X-1}, x)$ is given by $p_{y_0,\dots,y_{d_X-1},x}$, i.e.,

$$\mathcal{T} = \{p_{y_0,\dots,y_{d_X-1},x}\}_{y_0,\dots,y_{d_X-1},x}. \tag{86}$$

We denote vec($\mathcal{T}$) as the vectorization of tensor $\mathcal{T}$ in lexicographic order, i.e., $(p_{0,\dots,0,0,0}, \dots,)$. The length of vec($\mathcal{T}$) is $d_Y^{d_X} d_X$. Let $\mathbf{1}$ be a $d_Y^{d_X} d_X$ vector, whose elements are all 1. Let $\mathcal{C}_\alpha \in \mathbb{R}^{d_Y^{d_X} \times d_X}$, where α is a function from $\mathbb{R}^{d_Y^{d_X} \times d_X}$ to $\{0, 1\}$, denote the tensor whose element at $(y_0, \dots, y_{d_X-1}, x)$ is given by the function $\alpha(y_0, \dots, y_{d_X-1}, x)$, i.e.,

$$\mathcal{C}_\alpha = \{\alpha(y_0, \dots, y_{d_X-1}, x)\}_{y_0,\dots,y_{d_X-1},x}. \tag{87}$$

(1) Total of Probabilities. The constraint

$$\sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} p_{y_0,\dots,y_{d_X-1},x} = 1. \tag{88}$$

can be represented as

$$\mathbf{1}^T \text{vec}(\mathcal{T}) = 1. \tag{89}$$

This is a single constraint.

(2) Experimental Data. The constraint by experimental data

$$\sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{l-1}=0}^{d_Y-1} \sum_{y_{l+1}=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} p_{y_0,\dots y_{l-1},m,y_{l+1}\dots,y_{d_X-1},x} = \mathbb{P}(Y_l = m) \tag{90}$$

for $l = 0, \dots, d_X - 1, m = 0, \dots, d_Y - 2$. This can be represented as

$$\text{vec}(\mathcal{C}_{\mathbb{I}(y_l=m)})^T \text{vec}(\mathcal{T}) = \mathbb{P}(Y_l = m) \tag{91}$$

for $l = 0, \dots, d_X - 1, m = 0, \dots, d_Y - 2$. These are $d_X(d_Y - 1)$ constraints.

(3) Observational Data. The constraint by observational data

$$\sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{l-1}=0}^{d_Y-1} \sum_{y_{l+1}=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} p_{y_0,\dots,y_{l-1},m,y_{l+1},\dots,y_{d_X-1},l} = \mathbb{P}(X = l, Y = m) \tag{92}$$

for $l = 0, \dots, d_X - 1, m = 0, \dots, d_Y - 1$. This can be represented as

$$\text{vec}(\mathcal{C}_{\mathbb{I}(y_l=m,x=l)})^T \text{vec}(\mathcal{T}) = \mathbb{P}(X = l, Y = m) \tag{93}$$

for $l = 0, \dots, d_X - 1, m = 0, \dots, d_Y - 1$. However, we need to exclude the case $(l, m) = (d_X - 1, d_Y - 1)$ since we have $\mathbb{P}(X = d_X - 1, Y = d_Y - 1) = 1 - \sum_{(l,m)\neq(d_X-1,d_Y-1)} \mathbb{P}(X = l, Y = m)$. Therefore, observational data derive $d_X d_Y - 1$ constraints.

(4) Monotonicity Assumption. The constraints derived from Assumption 11, i.e.,

$$L^w \leq \sum_{y_0=0}^{d_Y-1} \cdots \sum_{y_{d_X-1}=0}^{d_Y-1} \sum_{x=0}^{d_X-1} \left(\Big(\prod_{s,t=0,\dots,d_X-1; s>t} \mathbb{I}(\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)}) \Big) \times p_{y_0,\dots,y_{d_X-1},x} \right) \leq U^w \tag{94}$$

for $w = 1, \dots, W$. This can be expressed as

$$\begin{aligned} &\text{vec}(\mathcal{C}_{\prod_{s>t} \mathbb{I}(\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)})})^T \cdot \text{vec}(\mathcal{T}) \geq L^w \ (w = 1, \dots, W), \\ &\text{vec}(\mathcal{C}_{\prod_{s>t} \mathbb{I}(\underline{D}_{st}^{(w)} \leq y_s - y_t \leq \overline{D}_{st}^{(w)})})^T \cdot \text{vec}(\mathcal{T}) \leq U^w \ (w = 1, \dots, W). \end{aligned} \tag{95}$$

These are $2W$ inequality constraints.

Let

$$\begin{aligned} A' = \Big(&\mathrm{vec}(\mathcal{C}_{\prod_{s>t}\mathbb{I}(\underline{D}_{st}^{(1)}\le y_s - y_t \le \overline{D}_{st}^{(1)})}) \ \cdots \ \mathrm{vec}(\mathcal{C}_{\prod_{s>t}\mathbb{I}(\underline{D}_{st}^{(W)}\le y_s - y_t \le \overline{D}_{st}^{(W)})}) \\ &- \mathrm{vec}(\mathcal{C}_{\prod_{s>t}\mathbb{I}(\underline{D}_{st}^{(1)}\le y_s - y_t \le \overline{D}_{st}^{(1)})}) \ \cdots \ - \mathrm{vec}(\mathcal{C}_{\prod_{s>t}\mathbb{I}(\underline{D}_{st}^{(W)}\le y_s - y_t \le \overline{D}_{st}^{(W)})})\Big)^T, \\ b' = (&U^1, \ldots, U^W, -L^1, \ldots, -L^W)^T. \end{aligned} \tag{96}$$

A' is a $(2W) \times (d_X d_Y^{d_X})$-matrix. Then our LP is formulated in a matrix form in the following way with respect to the marginal data we have:

Formulation of our LP when using experimental data. Let

$$\begin{aligned} A_{\text{exp}} &= \Big(\mathbf{1} \quad \mathrm{vec}(\mathcal{C}_{\mathbb{I}(y_0=0)}) \ \ \mathrm{vec}(\mathcal{C}_{\mathbb{I}(y_0=1)}) \ \ \cdots \ \ \mathrm{vec}(\mathcal{C}_{\mathbb{I}(y_0=d_Y-1)}) \ \ \mathrm{vec}(\mathcal{C}_{\mathbb{I}(y_1=0)}) \cdots \ \ \mathrm{vec}(\mathcal{C}_{\mathbb{I}(y_{d_X-1}=d_Y-1)})\Big)^T, \\ b &= (1 \ \ \mathbb{P}(Y_0=0) \ \ \mathbb{P}(y_0=1) \ \ \cdots \ \ \mathbb{P}(y_0=d_Y-1) \ \ \mathbb{P}(y_1=0) \cdots \ \ \mathbb{P}(y_{d_X-1}=d_Y-1))^T. \end{aligned} \tag{97}$$

A_{exp} is a $(d_X d_Y - d_X + 1) \times (d_X d_Y^{d_X})$-matrix.

Then our LP is formulated as

$$\text{maximize / minimize } \lambda(\mathcal{P}) \text{ subject to } A_{\text{exp}}\mathrm{vec}(\mathcal{T}) = b, A'\mathrm{vec}(\mathcal{T}) \le b', \mathrm{vec}(\mathcal{T}) \ge 0. \tag{98}$$

We can reformulate it only using equality constraints or only using inequality constraints. Let $\mathcal{W} = (w_1, \ldots, w_{2W})^T$ be a sequence of slack variables. Then, the problem (98) is equivalent to

$$\text{maximize / minimize } \lambda(\mathcal{P}) \text{ subject to } \begin{bmatrix} A_{\text{exp}} & O \\ A' & E_{2W} \end{bmatrix} \begin{bmatrix} \mathrm{vec}(\mathcal{T}) \\ \mathcal{W} \end{bmatrix} = \begin{bmatrix} b \\ b' \end{bmatrix}, \quad \mathrm{vec}(\mathcal{T}) \ge 0, \ \mathcal{W} \ge 0 \tag{99}$$

or

$$\text{maximize / minimize } \lambda(\mathcal{P}) \text{ subject to } \begin{bmatrix} A_{\text{exp}} \\ -A_{\text{exp}} \\ A' \end{bmatrix} \mathrm{vec}(\mathcal{T}) \ \le \ \begin{bmatrix} b \\ -b \\ b' \end{bmatrix}, \quad \mathrm{vec}(\mathcal{T}) \ge 0. \tag{100}$$

where $\mathcal{W} = (w_1, \ldots, w_{2W})$ is a vector consisting of slack variables.

Formulation of our LP when using observational data. Let

$$\begin{aligned} A_{\text{obs}} = \Big(&\mathbf{1} \ \ \mathrm{vec}(\mathcal{C}_{\mathbb{I}(X=0,Y=0)}) \ \ \mathrm{vec}(\mathcal{C}_{\mathbb{I}(X=0,Y=1)}) \ \ \cdots \ \ \mathrm{vec}(\mathcal{C}_{\mathbb{I}(X=0,Y=d_Y-1)}) \\ &\mathrm{vec}(\mathcal{C}_{\mathbb{I}(X=1,Y=0)}) \cdots \ \ \mathrm{vec}(\mathcal{C}_{\mathbb{I}(X=d_X-1,Y=d_Y-1)})\Big)^T, \\ b = \Big(&1 \quad \mathbb{P}(X=0,Y=0) \ \ \mathbb{P}(X=0,Y=1) \ \ \cdots \ \ \mathbb{P}(X=0,Y=d_Y-1) \\ &\mathbb{P}(X=1,Y=0) \cdots \ \ \mathbb{P}(X=d_X-1,Y=d_Y-1)\Big)^T. \end{aligned} \tag{101}$$

A_{obs} is a $(d_X d_Y) \times (d_X d_Y^{d_X})$-matrix.

Then our LP is formulated as

$$\text{maximize / minimize } \lambda(\mathcal{P}) \text{ subject to } A_{\text{obs}}\mathrm{vec}(\mathcal{T}) = b, A'\mathrm{vec}(\mathcal{T}) \le b', \mathrm{vec}(\mathcal{T}) \ge 0. \tag{102}$$

We can reformulate it only using equality constraints or only using inequality constraints. Let $\mathcal{W} = (w_1, \ldots, w_{2W})^T$ be a sequence of slack variables. Then, the problem (102) is equivalent to

$$\text{maximize / minimize } \lambda(\mathcal{P}) \text{ subject to } \begin{bmatrix} A_{\text{obs}} & O \\ A' & E_{2W} \end{bmatrix} \begin{bmatrix} \mathrm{vec}(\mathcal{T}) \\ \mathcal{W} \end{bmatrix} = \begin{bmatrix} b \\ b' \end{bmatrix}, \quad \mathrm{vec}(\mathcal{T}) \ge 0, \ \mathcal{W} \ge 0 \tag{103}$$

or

$$\text{maximize / minimize } \lambda(\mathcal{P}) \text{ subject to } \begin{bmatrix} A_{\text{obs}} \\ -A_{\text{obs}} \\ A' \end{bmatrix} \text{vec}(\mathcal{T}) \leq \begin{bmatrix} b \\ -b \\ b' \end{bmatrix}, \quad \text{vec}(\mathcal{T}) \geq 0. \tag{104}$$

where $\mathcal{W} = (w_1, \ldots, w_{2W})$ is a vector consisting of slack variables.

Formulation of our LP when using both experimental and observational data. Let

$$\begin{aligned} A_{\text{both}} = \Big(& \mathbf{1} \;\; \text{vec}(\mathcal{C}_{\mathbb{I}(y_0=0)}) \;\; \text{vec}(\mathcal{C}_{\mathbb{I}(y_0=1)}) \;\; \cdots \;\; \text{vec}(\mathcal{C}_{\mathbb{I}(y_0=d_Y-1)}) \;\; \text{vec}(\mathcal{C}_{\mathbb{I}(y_1=0)}) \cdots \;\; \text{vec}(\mathcal{C}_{\mathbb{I}(y_{d_X-1}=d_Y-1)}) \\ & \text{vec}(\mathcal{C}_{\mathbb{I}(X=0,Y=0)}) \;\; \text{vec}(\mathcal{C}_{\mathbb{I}(X=0,Y=1)}) \;\; \cdots \;\; \text{vec}(\mathcal{C}_{\mathbb{I}(X=0,Y=d_Y-1)}) \\ & \text{vec}(\mathcal{C}_{\mathbb{I}(X=1,Y=0)}) \cdots \;\; \text{vec}(\mathcal{C}_{\mathbb{I}(X=d_X-1,Y=d_Y-1)}) \Big)^T, \\ b = (1 \;\; & \mathbb{P}(Y_0=0) \;\; \mathbb{P}(y_0=1) \;\; \cdots \;\; \mathbb{P}(y_0=d_Y-1) \;\; \mathbb{P}(y_1=0) \cdots \;\; \mathbb{P}(y_{d_X-1}=d_Y-1) \\ & \mathbb{P}(X=0,Y=0) \;\; \mathbb{P}(X=0,Y=1) \;\; \cdots \;\; \mathbb{P}(X=0,Y=d_Y-1) \\ & \mathbb{P}(X=1,Y=0) \cdots \;\; \mathbb{P}(X=d_X-1,Y=d_Y-1) \Big)^T. \end{aligned} \tag{105}$$

A_{both} is a $(2d_X d_Y - d_X) \times (d_X d_Y^{d_X})$-matrix.

Then our LP is formulated as

$$\text{maximize / minimize } \lambda(\mathcal{P}) \text{ subject to } A_{\text{both}}\text{vec}(\mathcal{T}) = b, A'\text{vec}(\mathcal{T}) \leq b', \text{vec}(\mathcal{T}) \geq 0. \tag{106}$$

We can reformulate it only using equality constraints or only using inequality constraints. Let $\mathcal{W} = (w_1, \ldots, w_{2W})^T$ be a sequence of slack variables. Then, the problem (106) is equivalent to

$$\text{maximize / minimize } \lambda(\mathcal{P}) \text{ subject to } \begin{bmatrix} A_{\text{both}} & O \\ A' & E_{2W} \end{bmatrix} \begin{bmatrix} \text{vec}(\mathcal{T}) \\ \mathcal{W} \end{bmatrix} = \begin{bmatrix} b \\ b' \end{bmatrix}, \quad \text{vec}(\mathcal{T}) \geq 0, \; \mathcal{W} \geq 0 \tag{107}$$

or

$$\text{maximize / minimize } \lambda(\mathcal{P}) \text{ subject to } \begin{bmatrix} A_{\text{both}} \\ -A_{\text{both}} \\ A' \end{bmatrix} \text{vec}(\mathcal{T}) \leq \begin{bmatrix} b \\ -b \\ b' \end{bmatrix}, \quad \text{vec}(\mathcal{T}) \geq 0, \tag{108}$$

where $\mathcal{W} = (w_1, \ldots, w_{2W})$ is a vector consisting of slack variables.

D.4 Details of estimation problem

We provide the details of plug-in estimators.

Bounding Joint POs/OVs Probabilities and Their Linear Combinations From Finite Samples We consider two types of datasets: (i) d_X experimental datasets $\mathcal{D}^k = \{Y^k_{(1)}, Y^k_{(2)}, \ldots, Y^k_{(N^k)}\}$, for $k = 0, \ldots, d_X - 1$, where N^k denotes the sample size of the k-th experimental dataset $\mathcal{D}^k$, obtained under the intervention $do(X = k)$. (ii) one observational dataset $\mathcal{D} = \{(X_{(1)}, Y_{(1)}), (X_{(2)}, Y_{(2)}), \ldots, (X_{(N)}, Y_{(N)})\}$, where N denotes the sample size for the observational dataset $\mathcal{D}$. To bound the joint POs/OVs probabilities, we use the empirical probabilities of $\mathbb{P}(Y_x)$ and $\mathbb{P}(X, Y)$, i.e.,

$$\hat{\mathbb{P}}(Y_k = j) = {N^k}^{-1} \sum_{i=1}^{N^k} \mathbb{I}(Y^k_{(i)} = j), \tag{109}$$

$$\hat{\mathbb{P}}(X = l, Y = m) = N^{-1} \sum_{i=1}^{N} \mathbb{I}(X_{(i)} = l, Y_{(i)} = m). \tag{110}$$

Substituting the empirical probabilities in Eqs. (109) and (110) into the LP constraints yields the plug-in estimators of the lower and upper bounds of $\lambda(\mathcal{P})$, i.e., $[\hat{\underline{\lambda}}(\mathcal{P}), \hat{\overline{\lambda}}(\mathcal{P})]$.

Note that even if Assumption 11 holds, sampling errors in the empirical probabilities may render the LPs infeasible. Throughout, we assume that the LPs constructed from empirical probabilities are feasible.

Consistency. The estimates of the bounds on $\lambda(\mathcal{P})$ are consistent estimators as follows.

Theorem 4 (Consistency)**.** *Under Assumption 11, if the plug-in LPs are feasible, $\hat{\underline{\lambda}}(\mathcal{P})$ and $\hat{\overline{\lambda}}(\mathcal{P})$ converge in probability to $\underline{\lambda}(\mathcal{P})$ and $\overline{\lambda}(\mathcal{P})$. The convergence rate is $\mathcal{O}_p(\sum_{k=0}^{d_X-1} N^{k^{-1/2}})$ for experimental data, $\mathcal{O}_p(N^{-1/2})$ for observational data, and $\mathcal{O}_p(\sum_{k=0}^{d_X-1} N^{k^{-1/2}} + N^{-1/2})$ for combined experimental and observational data.*

Estimating Joint POs/OVs Probabilities and Their Linear Combinations. By plugging the empirical probabilities in Eqs. (109) and (110) into the identification results in Theorems 2 and 3, we can estimate the joint POs/OVs probabilities and their linear combinations from finite samples. Explicitly, from Theorem 2, the estimates of $\mathbb{P}(Y_0, \dots, Y_{d_X-1})$ is

$$\hat{\mathbb{P}}(Y_0 = \cdots = Y_{d_X-1} = y_0) = \sum_{j=0}^{y_0} \hat{\mathbb{P}}(Y_{d_X-1} = j) - \sum_{j=0}^{y_0-1} \hat{\mathbb{P}}(Y_0 = j) \tag{111}$$

for $y_0 \in [d_Y]$, and

$$\hat{\mathbb{P}}(Y_0 = \cdots = Y_k = y_0,\ Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1) = \sum_{j=0}^{y_0} \hat{\mathbb{P}}(Y_k = j) - \sum_{j=0}^{y_0} \hat{\mathbb{P}}(Y_{k+1} = j) \tag{112}$$

for $k \in [d_X - 1]$ and $y_0 \in [d_Y - 1]$; otherwise,

$$\hat{\mathbb{P}}(Y_0 = y_0, \dots, Y_{d_X-1} = y_{d_X-1}) = 0. \tag{113}$$

The estimate of $\lambda'(\mathcal{P}')$ is given by the linear combination of them. From Theorem 3, the estimates of $\mathbb{P}(Y_0, \dots, Y_{d_X-1}, X, Y)$ are

$$\begin{aligned}\hat{\mathbb{P}}(Y_0 = \cdots = Y_{d_X-1} = y_0, X = x, Y = y) = &\left(\sum_{m=0}^{y_0} \hat{\mathbb{P}}(Y = m | X = d_X - 1) - \sum_{m=0}^{y_0-1} \hat{\mathbb{P}}(Y = m | X = 0) \right) \\ &\times \hat{\mathbb{P}}(X = x)\end{aligned} \tag{114}$$

for $y_0 \in [d_Y]$, $x \in [d_X]$, and $y = y_0$, and

$$\begin{aligned}&\hat{\mathbb{P}}(Y_0 = \cdots = Y_k = y_0, Y_{k+1} = \cdots = Y_{d_X-1} = y_0 + 1, X = x, Y = y) \\ &= \left(\sum_{l=0}^{y_0} \hat{\mathbb{P}}(Y = l | X = k) - \sum_{l=0}^{y_0} \hat{\mathbb{P}}(Y = l | X = k + 1) \right) \hat{\mathbb{P}}(X = x)\end{aligned} \tag{115}$$

for $k \in [d_X - 1]$, $y_0 \in [d_Y - 1]$, $y = y_0$ for $x \le k$, and $y = y_0 + 1$ for $x = k + 1, \dots, d_X - 1$; otherwise,

$$\mathbb{P}(Y_0 = y_0, \dots, Y_{d_X-1} = y_{d_X-1}, X = x, Y = y) = 0. \tag{116}$$

The estimate of $\lambda(\mathcal{P})$ is given by the linear combination of them.

These are consistent estimators.

Theorem 5 (Consistency)**.** *Under Assumption 12, $\lambda'(\hat{\mathcal{P}}')$ converges in probability to $\lambda'(\mathcal{P}')$ at the rate $\mathcal{O}_p(\sum_{k=0}^{d_X-1} N^{k^{-1/2}})$.*

Theorem 6 (Consistency)**.** *Under Assumptions 1 and 12, $\lambda(\hat{\mathcal{P}})$ converges in probability to $\lambda(\mathcal{P})$ at the rate $\mathcal{O}_p(N^{-1/2})$.*

Proof. We provide proofs of theorems in this subsection.

We first show the following lemma to prove Theorem 4.

Lemma 4. *Let $V = \{x \in \mathbb{R}^d : Ax \ge c\}$ be a nonempty polyhedron. For any $h \in \mathbb{R}^d$ such that $\alpha := \max_{x \in V} h^T x$ is finite and attained on V, there exists a finite set $S \subseteq V$, depending only on V (not on h), such that $\max_{x \in V} h^T x = \max_{a \in S} h^T a$.*

Proof. Fix h with finite attained optimum α. Define the maximizer set

$$F(h) := \{x \in V \mid h^T x = \alpha\}. \tag{117}$$

Then $F(h)$ is the nonempty intersection of V with the hyperplane $\{x \mid h^T x = \alpha\}$; hence $F(h)$ is an exposed face of V. Choose one representative point $r(F)$ from each nonempty face F of V, and set $S = \{r(F) \mid F \text{ is a nonempty face of } V\}$. Then S is finite since a polyhedron has only finitely many faces, and S is independent of h. $F(h) \cap V (\neq \emptyset)$ is one of those faces, since V is contained in the closed half-space $\{x \mid h^T x \leq \alpha\}$. Then $z \coloneqq r(F(h)) \in S$ and $\alpha = h^T z$. This proves the lemma. □

Proof of Theorem 4. We remark that the feasible regions of our LPs are subsets of $[0,1]^{\mathcal{X}}$, where $\mathcal{X}$ is the number of parameters. As a consequence, since we assumed the plug-in LPs are feasible, their optima always exist and are bounded.

Passing from the mixed system in (75) to the equality form in (76) by introducing nonnegative slack variables preserves feasibility, the right-hand-side vector and optimum. Therefore, we may assume that the plug-in LPs are represented in the equality form.

Let us denote the corresponding LP whose optima are $\underline{\lambda}(\mathcal{P})$ and $\overline{\lambda}(\mathcal{P})$ by

$$\text{maximize/minimize } c^T x \text{ with respect to } Ax = b, x \geq 0. \tag{118}$$

We also denote the corresponding LP whose optima are $\hat{\underline{\lambda}}(\mathcal{P})$ and $\hat{\overline{\lambda}}(\mathcal{P})$ by

$$\text{maximize/minimize } c^T x \text{ with respect to } Ax = \hat{b}, x \geq 0. \tag{119}$$

Note that A, c are common in (118) and (119).

First, we will show the statement on UB. The feasible region of the dual of (118) and that of (119) are $\{y \mid A^T y \geq c\}$ when the primal LP is the maximization, and do not depend on b or $\hat{b}$. It is not empty, since primal LPs have optima. We remark that any dual LP of (118) or (119) has an optimum. Therefore, by Lemma 4, we can take a ***finite*** set $\mathcal{C}$ such that the optima of any dual LP of (118) or (119) are achieved at some point of $\mathcal{C}$. We take $M = \max_{a \in \mathcal{C}} \|a\|$.

Let $u(b) \in \mathcal{C}$ be a solution of the dual of (118). By properties of dual LP, we have $\overline{\lambda}(\mathcal{P}) = b^T u(b)$. We also have $\hat{\overline{\lambda}}(\mathcal{P}) = \min_{y; A^T y \geq c} \hat{b}^T y \leq \hat{b}^T u(b)$. Then we obtain

$$\hat{\overline{\lambda}}(\mathcal{P}) - \overline{\lambda}(\mathcal{P}) \leq (\hat{b} - b)^T u(b) \leq \|\hat{b} - b\| \|u(b)\| \leq M \|\hat{b} - b\|. \tag{120}$$

Changing the roles of b and $\hat{b}$, we also obtain $\overline{\lambda}(\mathcal{P}) - \hat{\overline{\lambda}}(\mathcal{P}) \leq M \|\hat{b} - b\|$. Summing up, we have

$$|\overline{\lambda}(\mathcal{P}) - \hat{\overline{\lambda}}(\mathcal{P})| \leq M \|\hat{b} - b\|. \tag{121}$$

Then, for any $\varepsilon > 0$, we have

$$\mathbb{P}\big(|\hat{\overline{\lambda}}(\mathcal{P}) - \overline{\lambda}(\mathcal{P})| > \varepsilon\big) \leq \mathbb{P}\big(\|\hat{b} - b\| > \varepsilon / M\big)\big). \tag{122}$$

Since $\hat{b}$ converges in probability to b, $\hat{\overline{\lambda}}(\mathcal{P})$ converges in probability to $\overline{\lambda}(\mathcal{P})$. The convergence rate of $\hat{\overline{\lambda}}(\mathcal{P})$ is the same as that of $\hat{b}$ since (121) holds. In our case, the convergence rate is $\mathcal{O}_p(\sum_{k=0}^{d_X - 1} \frac{1}{\sqrt{N^k}})$ for experimental data, $\mathcal{O}_p(\frac{1}{\sqrt{N}})$ for observational data, and $\mathcal{O}_p(\sum_{k=0}^{d_X - 1} \frac{1}{\sqrt{N^k}} + \frac{1}{\sqrt{N}})$ when both are combined.

Next we show the statement on LB. For any vector z, minimizing $c^T x$ with respect to $Ax = b, x \geq 0$ is equivalent to maximizing $-c^T x$ with respect to $Ax = b, x \geq 0$. Therefore, we see that the feasible region of the dual of (118) and that of (119) are $\{y \mid A^T y \geq -c\}$ when the primal LP is the maximization, and do not depend on b or $\hat{b}$. It is not empty, since primal LPs have optima. We remark that any dual LP of (118) or (119) has an optimum. Therefore, by Lemma 4, we can take a ***finite*** set $\mathcal{C}'$ such that the optima of any dual LP of (118) or (119) are achieved at some point of $\mathcal{C}'$. We take $M' = \max_{a \in \mathcal{C}'} \|a\|$.

Let $v(b) \in \mathcal{C}'$ be a solution of the dual of (118). By properties of dual LP, we have $\underline{\lambda}(\mathcal{P}) = b^T v(b)$. We also have $\hat{\underline{\lambda}}(\mathcal{P}) = \max_{y; A^T y \geq -c} \hat{b}^T y \geq \hat{b}^T v(b)$. Then we obtain

$$\underline{\lambda}(\mathcal{P}) - \hat{\underline{\lambda}}(\mathcal{P}) \leq (b - \hat{b})^T v(b) \leq \|\hat{b} - b\| \|v(b)\| \leq M' \|\hat{b} - b\|. \tag{123}$$

Changing the roles of b and $\hat{b}$, we also obtain $\hat{\underline{\lambda}}(\mathcal{P}) - \underline{\lambda}(\mathcal{P}) \leq M'\|\hat{b} - b\|$. Summing up, we have

$$|\underline{\lambda}(\mathcal{P}) - \hat{\underline{\lambda}}(\mathcal{P})| \leq M'\|\hat{b} - b\|. \tag{124}$$

Then, for any $\varepsilon > 0$, we have

$$\mathbb{P}\big(|\hat{\underline{\lambda}}(\mathcal{P}) - \underline{\lambda}(\mathcal{P})| > \varepsilon\big) \leq \mathbb{P}\big(\|\hat{b} - b\| > \varepsilon/M'\big). \tag{125}$$

Since $\hat{b}$ converges in probability to b, $\hat{\underline{\lambda}}(\mathcal{P})$ converges in probability to $\underline{\lambda}(\mathcal{P})$. The convergence rate of $\hat{\underline{\lambda}}(\mathcal{P})$ is the same as that of $\hat{b}$ since (124) holds. In our case, the convergence rate is $\mathcal{O}_p(\sum_{k=0}^{d_X-1} \frac{1}{\sqrt{N^k}})$ for experimental data, $\mathcal{O}_p(\frac{1}{\sqrt{N}})$ for observational data, and $\mathcal{O}_p(\sum_{k=0}^{d_X-1} \frac{1}{\sqrt{N^k}} + \frac{1}{\sqrt{N}})$ when both are combined. □

Proof of Theorem 5. $\lambda'(\hat{\mathcal{P}}')$ is a linear function of $\mathbb{P}(Y|X)$. Then, by the delta method, $\lambda'(\hat{\mathcal{P}}')$ converges in probability to $\lambda'(\mathcal{P}')$ at the rate $\mathcal{O}_p\left(\sum_{k=0}^{d_X-1} \frac{1}{\sqrt{N^k}}\right)$. Then, we have Theorem 5. □

Proof of Theorem 6. $\lambda(\hat{\mathcal{P}})$ is a linear function of $\mathbb{P}(Y|X)$ and $\mathbb{P}(X)$, respectively. Then, by the delta method, $\lambda(\hat{\mathcal{P}})$ converges in probability to $\lambda(\mathcal{P})$ at the rate $\mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right)$. Then, we have Theorem 6. □

D.5 Computational complexity to solve our bounding problems

Next, we discuss the computational complexity. There are two main approaches for solving LPs: the interior-point methods [Karmarkar, 1984] and the simplex method [RinnooyKan and Telgen, 1981].

Theorem 7. *The computational complexity of solving our bounding problems is polynomial when using interior-point methods, on the order of* $\mathcal{O}((d_X d_Y^{d_X} + 2W)^{3.5}(d_X d_Y^{d_X} + 2W + 1)(d_X d_Y - d_X + 1 + 2W) + (\sum_{k=0}^{d_X-1} N^k)d_Y)$ *when using experimental data,* $\mathcal{O}(d_X d_Y^{d_X} + 2W)^{3.5}(d_X d_Y^{d_X} + 2W + 1)d_X d_Y + N d_X d_Y)$ *when using observational data,* $\mathcal{O}((d_X d_Y^{d_X} + 2W)^{3.5}(d_X d_Y^{d_X} + 2W + 1)(2d_X d_Y - d_X + 2W) + (\sum_{k=0}^{d_X-1} N^k)d_Y + N d_X d_Y)$ *when using both experimental and observational data.*

Therefore, our bounding problems can be solved in polynomial time. The simplex method has exponential worst-case complexity as follows:

Theorem 8. *If we use the simplex method to solve LP, the worst-case computational complexity of solving our bounding problems is on the order of* $\mathcal{O}((d_X d_Y^{d_X})^{2(d_X d_Y - d_X + 1 + W)} + (\sum_{k=0}^{d_X-1} N^k)d_Y)$ *when using experimental data,* $\mathcal{O}((d_X d_Y^{d_X})^{2(d_X d_Y + W)} + N d_X d_Y)$ *when using observational data,* $\mathcal{O}((d_X d_Y^{d_X})^{2(2d_X d_Y - d_X + W)} + (\sum_{k=0}^{d_X-1} N^k)d_Y + N d_X d_Y)$ *when using both experimental and observational data.*

However, in practice, the simplex method rarely exhibits its exponential worst-case behavior and is often faster than interior-point algorithms [Vanderbei, 2001].

The calculation of $\lambda'(\hat{\mathcal{P}}')$ using Theorem 2 requires $\mathcal{O}(d_Y \sum_{k=0}^{d_X-1} N^k + d_X d_Y)$ computations, while $\lambda(\hat{\mathcal{P}})$ using Theorem 3 requires $\mathcal{O}(d_X d_Y N + d_X^2 d_Y^2)$ computations.

Proof. We provide proofs of theorems in this subsection.

Proof of Theorem 7. We note that our LP has $d_X d_Y^{d_X}$ variables, $d_X d_Y - d_X + 1$ (resp. $d_X d_Y$, $2d_X d_Y - d_X$) equality constraints and $2W$ inequality constraints when we have experimental (resp. observational, both experimental and observational) data. As stated in Appendix D, we can reformulate our LP to an LP with $d_X d_Y^{d_X} + 2W$ variables and $d_X d_Y - d_X + 1 + 2W$ (resp. $d_X d_Y + 2W$, $2d_X d_Y - d_X + 2W$) equality constraints we have experimental (resp. observational, both experimental and observational) data.

For an LP with $\mathfrak{n}$ variables and $\mathfrak{m}$ equality constraints, the computational complexity of solving LP using the interior-point method is on the order of $\mathcal{O}(\mathfrak{n}^{3.5}\ell)$, where ℓ denotes the length of input [Karmarkar, 1984]. As for bounding of joint distribution of POs, since the coefficients of constraints and objectives take 0 or 1, ℓ is on the order of $\mathcal{O}(\mathfrak{n}\mathfrak{m} + \mathfrak{n})$. Therefore, the computational complexity of solving our LP is on the order of $\mathcal{O}((d_X d_Y^{d_X} + 2W)^{3.5}(d_X d_Y^{d_X} + 2W + 1)(d_X d_Y - d_X + 1 + 2W))$ when using experimental data, $\mathcal{O}((d_X d_Y^{d_X} + 2W)^{3.5}(d_X d_Y^{d_X+1} + 2W + 1)(d_X d_Y))$ when using observational data, $\mathcal{O}((d_X d_Y^{d_X} + 2W)^{3.5}(d_X d_Y^{d_X} + 2W + 1)(2d_X d_Y - d_X + 2W))$ when using both experimental and observational data.

We also need a calculation to estimate $\hat{\mathbb{P}}(Y_k)$ or $\hat{\mathbb{P}}(X,Y)$ so that we can impose marginal constraints. (109) and (110) imply that the amount of calculation to impose marginal constraints is on the order of $(\sum_{k=0}^{d_X-1} N^k)d_Y$, Nd_Xd_Y, and $(\sum_{k=0}^{d_X-1} N^k)d_Y + Nd_Xd_Y$, when using experimental data, observational data and both experimental and observational data, respectively. Since these calculations are independent from solving LPs, the computational complexity of solving our bounding problems is a sum of that of solving LP and that of estimating marginal probabilities. This proves theorem 7. □

Proof of Theorem 8. We note that our LP has $d_Xd_Y^{d_X}$ variables, $d_Xd_Y - d_X + 1$ (resp. $d_Xd_Y,\ 2d_Xd_Y - d_X$) equality constraints and $2W$ inequality constraints when we have experimental (resp. observational, both experimental and observational) data. As stated in Appendix D, we can reformulate our LP to an LP with $d_Xd_Y^{d_X}$ variables and $2(d_Xd_Y - d_X + 1) + 2W$ (resp. $2d_Xd_Y + 2W,\ 2(2d_Xd_Y - d_X) + 2W$) inequality constraints.

For an LP with $\mathfrak{n}$ variables, and $\mathfrak{m}$ inequality constraints, the computational complexity of solving LP using the simplex method is on the order of $\mathcal{O}(\mathfrak{n}^{\mathfrak{m}})$ in the worst case [RinnooyKan and Telgen, 1981]. Then, in our case, the computational complexity of solving our LP is on the order of $\mathcal{O}((d_Xd_Y^{d_X})^{2(d_Xd_Y-d_X+1+W)})$ when using experimental data, $\mathcal{O}((d_Xd_Y^{d_X})^{2(d_Xd_Y+W)})$ when using observational data, $\mathcal{O}((d_Xd_Y^{d_X})^{2(2d_Xd_Y-d_X+W)})$ when using both experimental and observational data.

We also need a calculation to estimate $\hat{\mathbb{P}}(Y_k)$ or $\hat{\mathbb{P}}(X,Y)$ so that we can impose marginal constraints. (109) and (110) imply that the amount of calculation to impose marginal constraints is on the order of $(\sum_{k=0}^{d_X-1} N^k)d_Y$, Nd_Xd_Y, and $(\sum_{k=0}^{d_X-1} N^k)d_Y + Nd_Xd_Y$, when using experimental data, observational data, and both experimental and observational data, respectively. Since these calculations are independent from solving LPs, the computational complexity of solving our bounding problems is a sum of that of solving LP and that of estimating marginal probabilities. This proves theorem 8. □

E NUMERICAL EXPERIMENTS

In this section, we provide details of the numerical experiments.

We illustrate how MAs can tighten the bounds on the joint POs/OVs probabilities and their linear combinations, and how point estimates can be obtained under additional identification assumptions. We use a computer with AMD Ryzen 7 8840U 3.30GHz processor and 16GB of RAM. The source code is available at (https://github.com/Nmath997/AISTATS2026_bounding.git).

E.1 Bounding Under Monotonicity Assumption

Baselines. Our bounds for the joint POs/OVs probabilities are compared against those in Li and Pearl [2024] when incorporating both experimental and observational data. Their bounds are given in closed form but are not sharp for PoC involving multiple hypothetical terms (multiple PO terms).

Settings. We set $d_X = d_Y = 3$ and the number of parameters is 243. We set the values of $\mathcal{P}$ as presented in Table 5. We evaluate (a) the joint POs probability $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$ using experimental, observational, and combined data, (b) the posterior causal effects $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ using observational and combined data, and (c) the second moment of causal effects $\mathbb{E}[(Y_1 - Y_0)^2]$ using experimental, observational, and combined data. We solve the LPs using the Simplex algorithm implemented in the R package "Rglpk" (https://cran.r-project.org/web/packages/Rglpk/index.html), which provides an R interface to the GNU Linear Programming Kit (GLPK). We set the true joint POs/OVs probabilities as specified in Table 5, and then draw an i.i.d. sample of size N for the experimental and observational datasets, respectively. We perform 100 simulation runs and report the sample mean along with 95% confidence intervals (CIs). We exclude infeasible simulations and report the results of 100 feasible simulations. Tables 8 and 9 report the running times (in milliseconds) for a single simulation. The running times are quite fast across all situations.

To compare the "almost surely" assumptions, we first consider the following MAs:

(i). No Additional Assumption,
(ii). $Y_0 \leq Y_1$ almost surely (Assumption 9),
(iii). $Y_1 \leq Y_2$ almost surely (Assumption 9),
(iv). $Y_0 \leq Y_1 \leq Y_2$ almost surely (Assumption 6),

The setting in Table 5 satisfies all of the MAs listed above. Assumptions (ii) and (iii) are both stronger than Assumption (i) but weaker than Assumption (iv); moreover, there is no ordering relationship between Assumptions (ii) and (iii). We next additionally compare the probability-limited assumption (Assumption 8) varying threshold L, i.e.,

(v). $L \leq \mathbb{P}(Y_0 \leq Y_1 \leq Y_2)$.

The assumption becomes stronger as L grows.

We finally compare multiple probability-limited assumptions (Assumption 11) varying thresholds L_1, L_2, and L_3,

(vi). $L_1 \leq \mathbb{P}(0 \leq Y_1 - Y_0 \leq 1)$, $L_2 \leq \mathbb{P}(0 \leq Y_2 - Y_1 \leq 1)$, $L_3 \leq \mathbb{P}(0 \leq Y_2 - Y_0 \leq 1)$,

under the setting in Table 7. We have $0 \leq Y_1 - Y_0 \leq 1$, $0 \leq Y_2 - Y_1 \leq 1$, and $0 \leq Y_2 - Y_0 \leq 1$ under this setting. As L_1, L_2, and L_3 increase, the assumptions become more restrictive.

Results. First, Table 10 reports the estimates of the bounds of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$, while Table 11 presents the estimates of the bounds of $\mathbb{E}[Y_1 - Y_0 | X = 2, Y = 2]$ under almost surely monotonicities with sample size $N = 10, 100, 1000, 10000$. Table 12 reports the estimates of the bounds of $\mathbb{E}[(Y_1 - Y_0)^2]$ under Assumptions (i)–(iv). In Table 10, when using experimental data, Assumption (ii) does not yield tighter bounds than those obtained without additional assumptions. In contrast, Assumptions (iii) and (iv) provide tighter bounds compared to the no-assumption case. When using observational data, Assumptions (i)–(iv) do not yield tighter bounds than those obtained without additional assumptions. When using experimental and observational data, Assumption (ii) does not yield tighter bounds than those obtained without additional assumptions. In contrast, Assumptions (iii) and (iv) provide tighter bounds compared to the no-assumption case. The bounds by [Li et al., 2024] are the same as our bounds without additional assumptions. However, these bounds are not sharp, as demonstrated in Table 13. In Table 11, when using only observational data, all assumptions do not yield tighter bounds than those obtained without additional assumptions. When using experimental and observational data, Assumption (iii) does not yield tighter bounds than those obtained without additional assumptions. In contrast, Assumptions (ii) and (iv) provide tighter bounds compared to the no-assumption case. In Table 12, we also observe that Assumptions (ii) and (iv) provide tighter bounds, while Assumption (iii) does not yield tighter bounds than the case with no additional MAs. All 95% CIs, both for the estimates under the identification assumptions and for the bounds, become narrower as the sample size increases. When $N = 10000$, all the estimates are accurate.

Next, Table 14 (resp. Table 15, Table 16) reports the estimates of the values of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$ (resp. $\mathbb{E}[Y_1 - Y_0 | X = 2, Y = 2]$, $\mathbb{E}[(Y_1 - Y_0)^2]$), varying the threshold L of Assumption (v) when $N = 1000$. The bounds become tight as the threshold L grows large, and it is visualized in Figure 1 (resp. Figure 2, Figure 3). Table 17 reports the estimates of the values of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$, varying the thresholds L_1, L_2 and L_3 of Assumption (vi) when $N = 1000$. Tables 18 and 19 (resp. Tables 20 and 21) report the estimates of the values of $\mathbb{E}[Y_1 - Y_0 | X = 2, Y = 2]$ (resp. $\mathbb{E}[(Y_1 - Y_0)^2]$), varying the thresholds L_1, L_2 and L_3 of Assumption (vi) when $N = 1000$. As L_1, L_2, and L_3 increase, the bounds become tight.

We provide details of the experimental setting for the bounding problem under MAs. We set the value $\mathbb{P}(Y_0, Y_1, Y_2, X, Y)$ as in Table 5. These values satisfy the condition $Y_0 \leq Y_1 \leq Y_2$.

Table 5: Setting of simulations for Assumptions (i)-(v) in Section E.1

$X=0,Y=0$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	1/40	0	0	0	0	0	0	0	0
$Y_2=1$	1/40	1/40	0	0	0	0	0	0	0
$Y_2=2$	1/40	1/40	1/40	0	0	0	0	0	0

$X=0,Y=1$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	1/30	0	0	0	0
$Y_2=2$	0	0	0	0	1/30	1/30	0	0	0

$X=0,Y=2$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	1/10

$X=1,Y=0$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	1/30	0	0	0	0	0	0	0	0
$Y_2=1$	1/30	0	0	0	0	0	0	0	0
$Y_2=2$	1/30	0	0	0	0	0	0	0	0

$X=1,Y=1$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	1/20	0	0	1/20	0	0	0	0
$Y_2=2$	0	1/20	0	0	1/20	0	0	0	0

$X=1,Y=2$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	1/30	0	0	1/30	0	0	1/30

$X=2,Y=0$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	1/20	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	0

$X=2,Y=1$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	1/30	1/30	0	0	1/30	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	0

$X=2,Y=2$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	1/60	1/60	1/60	0	1/60	1/60	0	0	1/60

Table 6: Settings of $\mathbb{P}(Y_1, Y_0, Y_2, X, Y)$ for comparing bounds based on [Li and Pearl, 2024] and LP solver. When $x \neq y$, we set $\mathbb{P}(Y_0, Y_1, Y_2, X = x, Y = y) = 0$.

$X=0,Y=0$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0.011	0.010	0.042	0	0	0	0	0	0
$Y_2=1$	0.045	0.059	0.069	0	0	0	0	0	0
$Y_2=2$	0.017	0.041	0.047	0	0	0	0	0	0
$X=1,Y=1$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0.030	0	0	0.000	0	0	0.000	0
$Y_2=1$	0	0.112	0	0	0.013	0	0	0.033	0
$Y_2=2$	0	0.111	0	0	0.017	0	0	0.003	0
$X=2,Y=2$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0.060	0.062	0.052	0.007	0.018	0.006	0.121	0.004	0.009

Table 7: Setting of simulations for Assumption (vi) in Section E.1

$X=0,Y=0$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	1/20	0	0	0	0	0	0	0	0
$Y_2=1$	1/20	1/20	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	0

$X=0,Y=1$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	1/30	0	0	0	0
$Y_2=2$	0	0	0	0	1/30	1/30	0	0	0

$X=0,Y=2$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	1/10

$X=1,Y=0$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	1/20	0	0	0	0	0	0	0	0
$Y_2=1$	1/20	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	0

$X=1,Y=1$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	1/15	0	0	1/15	0	0	0	0
$Y_2=2$	0	0	0	0	1/15	0	0	0	0

$X=1,Y=2$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	1/20	0	0	1/20

$X=2,Y=0$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	1/20	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	0

$X=2,Y=1$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	1/30	1/30	0	0	1/30	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	0

$X=2,Y=2$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	1/30	1/30	0	0	1/30

Table 8: Running time of solving maximization and minimization problem of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$ for one time in milliseconds. Each entry is the mean over 100 repetitions.

Exp.

MAs	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	3	3	3	3
(ii)	3	3	3	3
(iii)	3	3	3	3
(iv)	3	4	3	3

Obs.

MAs	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	3	3	2	3
(ii)	3	3	3	2
(iii)	3	2	3	2
(iv)	2	2	4	2

Both

MAs	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	3	3	3	3
(ii)	4	3	3	3
(iii)	3	3	3	3
(iv)	3	3	3	2

Table 9: Running time of solving maximization and minimization problem of $\mathbb{E}[Y_1 - Y_0 | X = 2, Y = 2]$ for one time in milliseconds. Each entry is the mean over 100 repetitions.

Exp.

MAs	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	2	3	2	3
(ii)	3	3	3	3
(iii)	3	3	2	2
(iv)	3	2	2	2

Obs.

MAs	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	2	2	2	2
(ii)	2	2	2	2
(iii)	2	2	2	2
(iv)	2	2	2	3

Both

MAs	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	2	3	3	3
(ii)	3	3	3	2
(iii)	3	3	3	3
(iv)	3	3	3	3

Table 10: Results of the estimates of the bounds of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$. We report the means and 95% confidence intervals (CIs). The true value of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$ is 0.092.

MAs	Bounds (Exp.)	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(i)	UB	0.224 [0.100,0.452]	0.265 [0.200,0.340]	0.274 [0.245,0.304]	0.275 [0.268,0.284]
(ii)	LB	0.002 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(ii)	UB	0.224 [0.100,0.452]	0.265 [0.200,0.340]	0.274 [0.245,0.304]	0.275 [0.268,0.284]
(iii)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(iii)	UB	0.163 [0.000,0.400]	0.167 [0.105,0.235]	0.166 [0.144,0.186]	0.167 [0.161,0.175]
(iv)	LB	0.027 [0.000,0.152]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(iv)	UB	0.163 [0.000,0.400]	0.167 [0.105,0.235]	0.166 [0.144,0.186]	0.167 [0.161,0.175]

MAs	Bounds (Obs.)	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(i)	UB	0.350 [0.100,0.700]	0.350 [0.265,0.435]	0.349 [0.318,0.383]	0.350 [0.339,0.358]
(ii)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(ii)	UB	0.350 [0.100,0.700]	0.350 [0.265,0.435]	0.349 [0.318,0.383]	0.350 [0.339,0.358]
(iii)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(iii)	UB	0.350 [0.100,0.700]	0.350 [0.265,0.435]	0.349 [0.318,0.383]	0.350 [0.339,0.358]
(iv)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(iv)	UB	0.350 [0.100,0.700]	0.350 [0.265,0.435]	0.349 [0.318,0.383]	0.350 [0.339,0.358]

MAs	Bounds (Both)	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i) Li–Pearl	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(i) Li–Pearl	UB	0.205 [0.000,0.400]	0.263 [0.200,0.335]	0.274 [0.245,0.304]	0.275 [0.268,0.284]
(i)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(i)	UB	0.205 [0.000,0.400]	0.262 [0.200,0.335]	0.274 [0.245,0.304]	0.275 [0.268,0.284]
(ii)	LB	0.015 [0.000,0.100]	0.000 [0.000,0.005]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(ii)	UB	0.204 [0.000,0.400]	0.262 [0.200,0.335]	0.274 [0.245,0.304]	0.275 [0.268,0.284]
(iii)	LB	0.003 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(iii)	UB	0.156 [0.000,0.400]	0.166 [0.105,0.235]	0.166 [0.144,0.186]	0.167 [0.161,0.175]
(iv)	LB	0.047 [0.000,0.200]	0.002 [0.000,0.030]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(iv)	UB	0.151 [0.000,0.400]	0.166 [0.105,0.235]	0.166 [0.144,0.186]	0.167 [0.161,0.175]

Table 11: Results of the estimates of the bounds of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$. We report the means and 95% confidence intervals (CIs). The true value of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ is 0.667.

MAs	Bounds (Obs.)	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	LB	-2.000 [-2.000,-2.000]	-2.000 [-2.000,-2.000]	-2.000 [-2.000,-2.000]	-2.000 [-2.000,-2.000]
(i)	UB	2.000 [2.000,2.000]	2.000 [2.000,2.000]	2.000 [2.000,2.000]	2.000 [2.000,2.000]
(ii)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(ii)	UB	2.000 [2.000,2.000]	2.000 [2.000,2.000]	2.000 [2.000,2.000]	2.000 [2.000,2.000]
(iii)	LB	-2.000 [-2.000,-2.000]	-2.000 [-2.000,-2.000]	-2.000 [-2.000,-2.000]	-2.000 [-2.000,-2.000]
(iii)	UB	2.000 [2.000,2.000]	2.000 [2.000,2.000]	2.000 [2.000,2.000]	2.000 [2.000,2.000]
(iv)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(iv)	UB	2.000 [2.000,2.000]	2.000 [2.000,2.000]	2.000 [2.000,2.000]	2.000 [2.000,2.000]

MAs	Bounds (Both)	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	LB	-1.073 [-2.000,0.417]	-1.515 [-2.000,-0.956]	-1.508 [-1.812,-1.251]	-1.501 [-1.582,-1.425]
(i)	UB	1.879 [1.000,2.000]	1.996 [2.000,2.000]	2.000 [2.000,2.000]	2.000 [2.000,2.000]
(ii)	LB	0.030 [0.000,0.417]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(ii)	UB	1.083 [0.000,2.000]	1.822 [1.116,2.000]	1.963 [1.702,2.000]	1.999 [1.987,2.000]
(iii)	LB	-0.743 [-2.000,1.000]	-1.423 [-2.000,-0.703]	-1.496 [-1.806,-1.215]	-1.501 [-1.582,-1.425]
(iii)	UB	1.879 [1.000,2.000]	1.996 [2.000,2.000]	2.000 [2.000,2.000]	2.000 [2.000,2.000]
(iv)	LB	0.128 [0.000,1.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(iv)	UB	0.976 [0.000,2.000]	1.709 [1.032,2.000]	1.961 [1.702,2.000]	1.999 [1.987,2.000]

Table 12: Results of the estimates of the bounds of $\mathbb{E}[(Y_1 - Y_0)^2]$. We report the means and 95% confidence intervals (CIs). The true value of $\mathbb{E}[(Y_1 - Y_0)^2]$ is 0.583.

MAs	Bounds (Exp.)	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	LB	0.417 [0.047,0.952]	0.441 [0.325,0.570]	0.432 [0.393,0.472]	0.432 [0.422,0.443]
(i)	UB	2.003 [0.900,3.052]	2.176 [1.803,2.485]	2.189 [2.077,2.313]	2.197 [2.151,2.240]
(ii)	LB	0.417 [0.047,0.952]	0.441 [0.325,0.570]	0.432 [0.393,0.472]	0.432 [0.422,0.443]
(ii)	UB	0.629 [0.048,1.300]	0.764 [0.539,1.016]	0.745 [0.659,0.816]	0.747 [0.725,0.771]
(iii)	LB	0.417 [0.047,0.952]	0.441 [0.325,0.570]	0.432 [0.393,0.472]	0.432 [0.422,0.443]
(iii)	UB	2.003 [0.900,3.052]	2.176 [1.803,2.485]	2.189 [2.077,2.313]	2.197 [2.151,2.240]
(iv)	LB	0.417 [0.047,0.952]	0.441 [0.325,0.570]	0.432 [0.393,0.472]	0.432 [0.422,0.443]
(iv)	UB	0.629 [0.048,1.300]	0.764 [0.539,1.016]	0.745 [0.659,0.816]	0.747 [0.725,0.771]
MAs	Bounds (Obs.)	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(i)	UB	3.136 [2.200,3.700]	3.094 [2.814,3.370]	3.100 [2.998,3.175]	3.102 [3.078,3.125]
(ii)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(ii)	UB	2.284 [1.300,3.100]	2.278 [1.965,2.625]	2.299 [2.193,2.414]	2.301 [2.274,2.330]
(iii)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(iii)	UB	3.136 [2.200,3.700]	3.094 [2.814,3.370]	3.100 [2.998,3.175]	3.102 [3.078,3.125]
(iv)	LB	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
(iv)	UB	1.848 [0.947,2.857]	1.806 [1.460,2.201]	1.800 [1.718,1.904]	1.801 [1.769,1.831]
MAs	Bounds (Both)	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(i)	LB	0.435 [0.047,1.100]	0.442 [0.325,0.570]	0.432 [0.393,0.472]	0.432 [0.422,0.443]
(i)	UB	1.913 [0.848,2.952]	2.170 [1.803,2.485]	2.189 [2.077,2.313]	2.197 [2.151,2.240]
(ii)	LB	0.435 [0.048,1.100]	0.442 [0.325,0.570]	0.432 [0.393,0.472]	0.432 [0.422,0.443]
(ii)	UB	0.617 [0.047,1.300]	0.762 [0.539,1.016]	0.745 [0.659,0.816]	0.747 [0.725,0.771]
(iii)	LB	0.447 [0.047,1.100]	0.445 [0.325,0.590]	0.432 [0.393,0.472]	0.432 [0.422,0.443]
(iii)	UB	1.811 [0.848,2.700]	2.125 [1.784,2.425]	2.186 [2.073,2.307]	2.197 [2.151,2.240]
(iv)	LB	0.447 [0.047,1.100]	0.445 [0.325,0.590]	0.432 [0.393,0.472]	0.432 [0.422,0.443]
(iv)	UB	0.613 [0.047,1.300]	0.762 [0.519,1.016]	0.745 [0.659,0.816]	0.747 [0.725,0.771]

Table 13: The bound of $\mathbb{P}(Y_0 = 0, Y_1 = 1, Y_2 = 2)$ based on Theorem 11 in [Li and Pearl, 2024] and LP solver under the setting in Table 6. The true value of $\mathbb{P}(Y_0 = 0, Y_1 = 1, Y_2 = 2)$ is 0.214.

Bounds (Both)	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
LB ([Li and Pearl, 2024])	0.056 [0.000,0.300]	0.003 [0.000,0.025]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
UB ([Li and Pearl, 2024])	0.426 [0.100,0.700]	0.501 [0.410,0.570]	0.517 [0.485,0.549]	0.514 [0.505,0.526]
LB	0.056 [0.000,0.300]	0.003 [0.000,0.025]	0.000 [0.000,0.000]	0.000 [0.000,0.000]
UB	0.389 [0.100,0.600]	0.414 [0.305,0.508]	0.426 [0.387,0.451]	0.426 [0.411,0.435]

Table 14: Results of the estimates of the bounds of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$ under assumption $L \leq \mathbb{P}(Y_0 \leq Y_1 \leq Y_2)$ and $N = 1000$. We report the means and 95% confidence intervals (CIs). The true value of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$ is 0.092.

Data	L	LB	UB
Obs	(All)	0.000 [0.000, 0.000]	0.351 [0.326, 0.374]
Exp. & Both	0.00	0.000 [0.000, 0.000]	0.274 [0.250, 0.301]
Exp. & Both	0.70	0.000 [0.000, 0.000]	0.274 [0.250, 0.301]
Exp. & Both	0.75	0.000 [0.000, 0.000]	0.274 [0.250, 0.301]
Exp. & Both	0.80	0.000 [0.000, 0.000]	0.274 [0.250, 0.301]
Exp. & Both	0.85	0.000 [0.000, 0.000]	0.274 [0.250, 0.301]
Exp. & Both	0.86	0.000 [0.000, 0.000]	0.274 [0.250, 0.301]
Exp. & Both	0.87	0.000 [0.000, 0.000]	0.274 [0.250, 0.300]
Exp. & Both	0.88	0.000 [0.000, 0.000]	0.273 [0.250, 0.294]
Exp. & Both	0.89	0.000 [0.000, 0.000]	0.270 [0.250, 0.291]
Exp. & Both	0.90	0.000 [0.000, 0.000]	0.264 [0.245, 0.286]
Exp. & Both	0.91	0.000 [0.000, 0.000]	0.255 [0.235, 0.276]
Exp. & Both	0.92	0.000 [0.000, 0.000]	0.245 [0.225, 0.266]
Exp. & Both	0.93	0.000 [0.000, 0.000]	0.235 [0.215, 0.256]
Exp. & Both	0.94	0.000 [0.000, 0.000]	0.225 [0.205, 0.246]
Exp. & Both	0.95	0.000 [0.000, 0.000]	0.215 [0.195, 0.236]
Exp. & Both	0.96	0.000 [0.000, 0.000]	0.205 [0.185, 0.226]
Exp. & Both	0.97	0.000 [0.000, 0.000]	0.195 [0.175, 0.216]
Exp. & Both	0.98	0.000 [0.000, 0.000]	0.185 [0.165, 0.206]
Exp. & Both	0.99	0.000 [0.000, 0.000]	0.175 [0.155, 0.196]
Exp. & Both	1.00	0.000 [0.000, 0.000]	0.165 [0.145, 0.186]

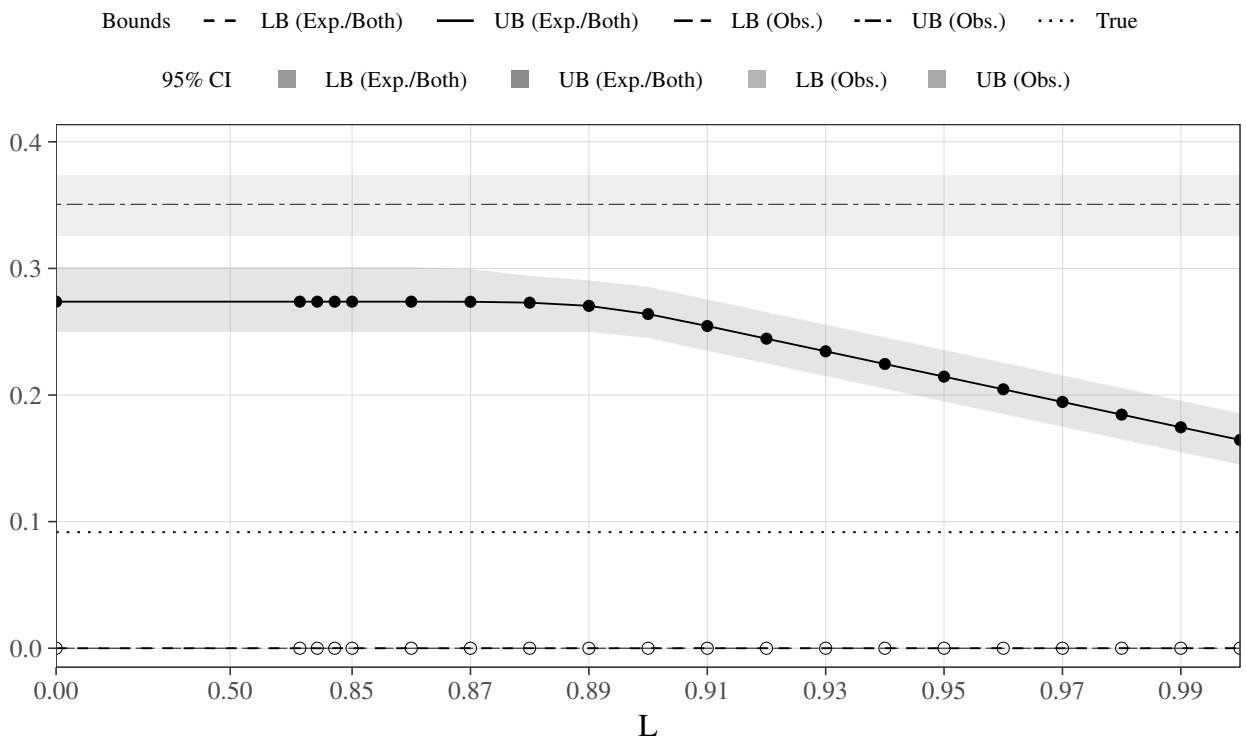

Figure 1: Results of the estimates of the bounds of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$ as L varies. The x-axis shows the value of L, and the y-axis shows the mean estimates of the bounds of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$.

Table 15: Results of the estimates of the bounds of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ under assumption $L \leq \mathbb{P}(Y_0 \leq Y_1 \leq Y_2)$ and $N = 1000$. We report the means and 95% confidence intervals (CIs). The true value of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ is 0.667.

Data	L	LB	UB
Obs	(All)	-2.000 [-2.000, -2.000]	2.000 [2.000, 2.000]
Both	0.00	-1.509 [-1.850, -1.267]	2.000 [2.000, 2.000]
Both	0.70	-1.509 [-1.850, -1.267]	2.000 [2.000, 2.000]
Both	0.75	-1.509 [-1.850, -1.267]	2.000 [2.000, 2.000]
Both	0.80	-1.509 [-1.850, -1.267]	2.000 [2.000, 2.000]
Both	0.85	-1.509 [-1.850, -1.267]	2.000 [2.000, 2.000]
Both	0.86	-1.509 [-1.850, -1.267]	2.000 [2.000, 2.000]
Both	0.87	-1.509 [-1.850, -1.267]	2.000 [2.000, 2.000]
Both	0.88	-1.508 [-1.850, -1.255]	2.000 [2.000, 2.000]
Both	0.89	-1.503 [-1.850, -1.216]	2.000 [2.000, 2.000]
Both	0.90	-1.481 [-1.821, -1.174]	2.000 [2.000, 2.000]
Both	0.91	-1.414 [-1.794, -1.112]	2.000 [2.000, 2.000]
Both	0.92	-1.318 [-1.696, -1.020]	2.000 [2.000, 2.000]
Both	0.93	-1.210 [-1.565, -0.922]	2.000 [2.000, 2.000]
Both	0.94	-1.095 [-1.356, -0.825]	2.000 [2.000, 2.000]
Both	0.95	-0.956 [-1.160, -0.726]	2.000 [2.000, 2.000]
Both	0.96	-0.792 [-0.958, -0.631]	1.999 [2.000, 2.000]
Both	0.97	-0.605 [-0.728, -0.515]	1.997 [1.977, 2.000]
Both	0.98	-0.405 [-0.485, -0.351]	1.995 [1.930, 2.000]
Both	0.99	-0.203 [-0.243, -0.175]	1.988 [1.836, 2.000]
Both	1.00	0.000 [0.000, 0.000]	1.971 [1.743, 2.000]

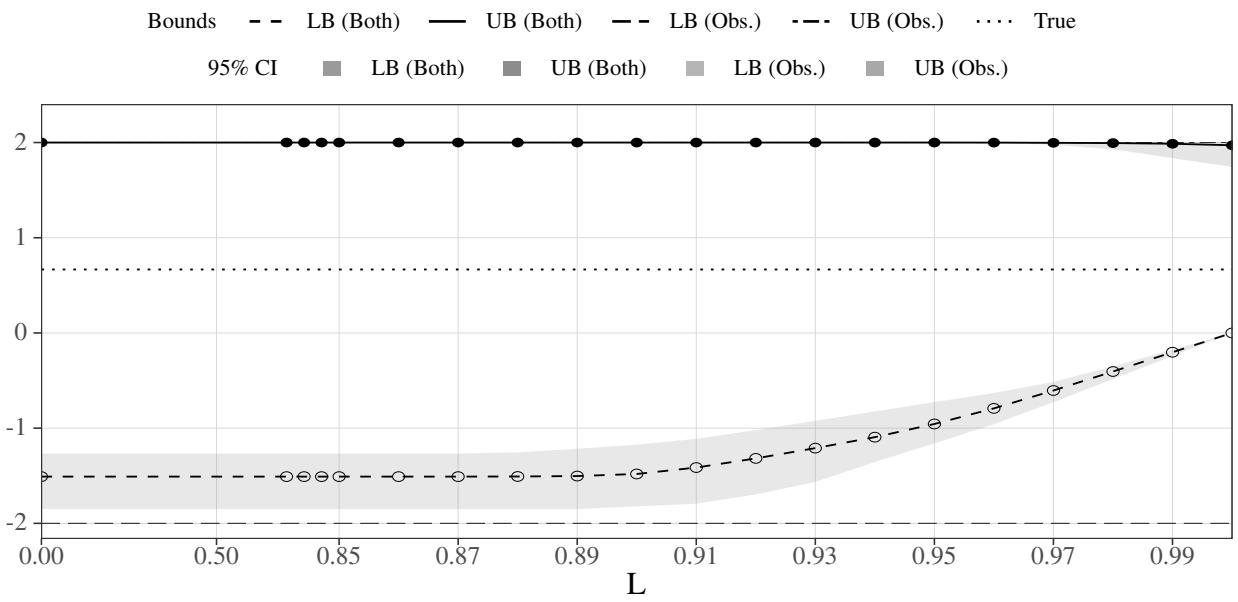

Figure 2: Results of the estimates of the bounds of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ as L varies. The x-axis shows the value of L, and the y-axis shows the mean estimates of the bounds of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$.

Table 16: Results of the estimates of the bounds of $\mathbb{E}[(Y_1 - Y_0)^2]$ under assumption $L \leq \mathbb{P}(Y_0 \leq Y_1 \leq Y_2)$ and $N = 1000$. We report the means and 95% confidence intervals (CIs). The true value of $\mathbb{E}[(Y_1 - Y_0)^2]$ is 0.583.

Data	L	LB	UB
Exp.	0.00	0.435 [0.392, 0.474]	2.196 [2.073, 2.339]
Exp.	0.70	0.435 [0.392, 0.474]	2.195 [2.073, 2.331]
Exp.	0.75	0.435 [0.392, 0.474]	2.141 [2.003, 2.272]
Exp.	0.80	0.435 [0.392, 0.474]	2.041 [1.903, 2.172]
Exp.	0.85	0.435 [0.392, 0.474]	1.921 [1.803, 2.024]
Exp.	0.86	0.435 [0.392, 0.474]	1.863 [1.775, 1.952]
Exp.	0.87	0.435 [0.392, 0.474]	1.791 [1.698, 1.876]
Exp.	0.88	0.435 [0.392, 0.474]	1.713 [1.621, 1.796]
Exp.	0.89	0.435 [0.392, 0.474]	1.633 [1.541, 1.716]
Exp.	0.90	0.435 [0.392, 0.474]	1.553 [1.461, 1.636]
Exp.	0.91	0.435 [0.392, 0.474]	1.473 [1.381, 1.556]
Exp.	0.92	0.435 [0.392, 0.474]	1.393 [1.301, 1.476]
Exp.	0.93	0.435 [0.392, 0.474]	1.313 [1.221, 1.396]
Exp.	0.94	0.435 [0.392, 0.474]	1.233 [1.141, 1.316]
Exp.	0.95	0.435 [0.392, 0.474]	1.153 [1.061, 1.236]
Exp.	0.96	0.435 [0.392, 0.474]	1.073 [0.981, 1.156]
Exp.	0.97	0.435 [0.392, 0.474]	0.993 [0.901, 1.076]
Exp.	0.98	0.435 [0.392, 0.474]	0.913 [0.821, 0.996]
Exp.	0.99	0.435 [0.392, 0.474]	0.833 [0.741, 0.916]
Exp.	1.00	0.435 [0.392, 0.474]	0.753 [0.661, 0.836]

Data	L	LB	UB
Obs.	0.00	0.000 [0.000, 0.000]	3.104 [3.026, 3.190]
Obs.	0.70	0.000 [0.000, 0.000]	2.945 [2.864, 3.030]
Obs.	0.75	0.000 [0.000, 0.000]	2.788 [2.691, 2.880]
Obs.	0.80	0.000 [0.000, 0.000]	2.593 [2.491, 2.695]
Obs.	0.85	0.000 [0.000, 0.000]	2.393 [2.291, 2.495]
Obs.	0.86	0.000 [0.000, 0.000]	2.353 [2.251, 2.455]
Obs.	0.87	0.000 [0.000, 0.000]	2.313 [2.211, 2.415]
Obs.	0.88	0.000 [0.000, 0.000]	2.273 [2.171, 2.375]
Obs.	0.89	0.000 [0.000, 0.000]	2.233 [2.131, 2.335]
Obs.	0.90	0.000 [0.000, 0.000]	2.193 [2.091, 2.295]
Obs.	0.91	0.000 [0.000, 0.000]	2.153 [2.051, 2.255]
Obs.	0.92	0.000 [0.000, 0.000]	2.113 [2.011, 2.215]
Obs.	0.93	0.000 [0.000, 0.000]	2.073 [1.971, 2.175]
Obs.	0.94	0.000 [0.000, 0.000]	2.033 [1.931, 2.135]
Obs.	0.95	0.000 [0.000, 0.000]	1.993 [1.891, 2.095]
Obs.	0.96	0.000 [0.000, 0.000]	1.953 [1.851, 2.055]
Obs.	0.97	0.000 [0.000, 0.000]	1.913 [1.811, 2.015]
Obs.	0.98	0.000 [0.000, 0.000]	1.873 [1.771, 1.975]
Obs.	0.99	0.000 [0.000, 0.000]	1.833 [1.731, 1.935]
Obs.	1.00	0.000 [0.000, 0.000]	1.793 [1.691, 1.895]

Data	L	LB	UB
Both	0.00	0.435 [0.392, 0.474]	2.196 [2.073, 2.339]
Both	0.70	0.435 [0.392, 0.474]	2.195 [2.073, 2.331]
Both	0.75	0.435 [0.392, 0.474]	2.139 [2.003, 2.265]
Both	0.80	0.435 [0.392, 0.474]	2.039 [1.903, 2.165]
Both	0.85	0.435 [0.392, 0.474]	1.918 [1.803, 2.003]
Both	0.86	0.435 [0.392, 0.474]	1.863 [1.775, 1.948]
Both	0.87	0.435 [0.392, 0.474]	1.791 [1.698, 1.876]
Both	0.88	0.435 [0.392, 0.474]	1.713 [1.621, 1.796]
Both	0.89	0.435 [0.392, 0.474]	1.633 [1.541, 1.716]
Both	0.90	0.435 [0.392, 0.474]	1.553 [1.461, 1.636]
Both	0.91	0.435 [0.392, 0.474]	1.473 [1.381, 1.556]
Both	0.92	0.435 [0.392, 0.474]	1.393 [1.301, 1.476]
Both	0.93	0.435 [0.392, 0.474]	1.313 [1.221, 1.396]
Both	0.94	0.435 [0.392, 0.474]	1.233 [1.141, 1.316]
Both	0.95	0.435 [0.392, 0.474]	1.153 [1.061, 1.236]
Both	0.96	0.435 [0.392, 0.474]	1.073 [0.981, 1.156]
Both	0.97	0.435 [0.392, 0.474]	0.993 [0.901, 1.076]
Both	0.98	0.435 [0.392, 0.474]	0.913 [0.821, 0.996]
Both	0.99	0.435 [0.392, 0.474]	0.833 [0.741, 0.916]
Both	1.00	0.435 [0.392, 0.474]	0.753 [0.661, 0.836]

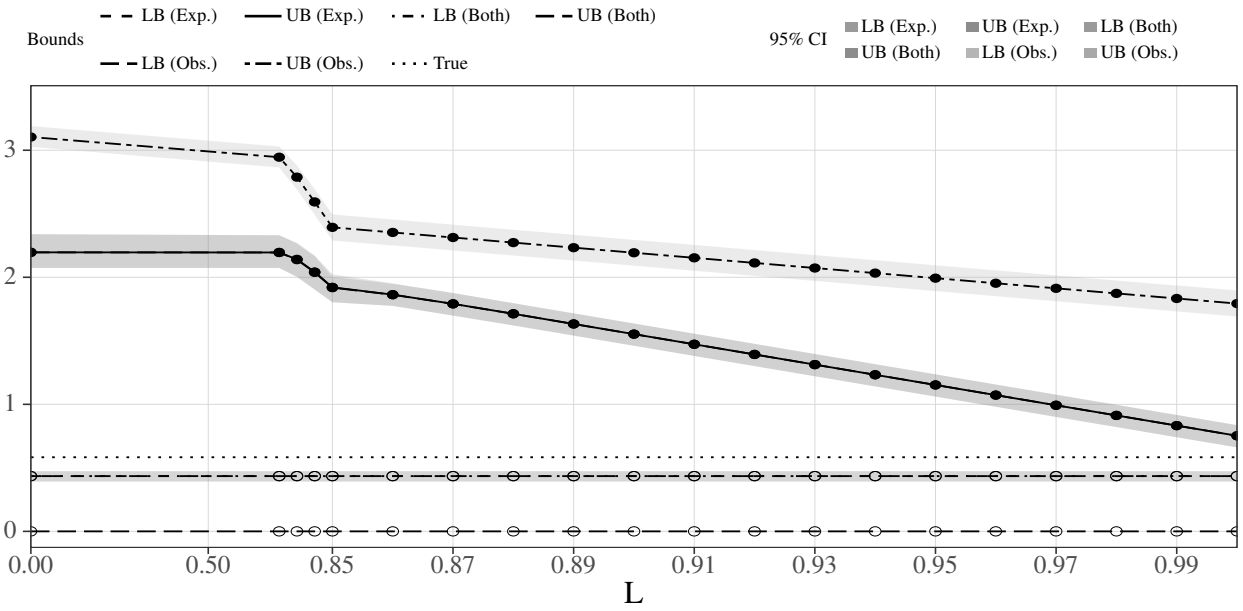

Figure 3: Results of the estimates of the bounds of $\mathbb{E}[(Y_1 - Y_0)^2]$ as L varies. The x-axis shows the value of L, and the y-axis shows the mean estimates of the bounds of $\mathbb{E}[(Y_1 - Y_0)^2]$.

Table 17: Results of the estimates of bounds of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$ under Assumption (vi) and $N = 1000$ with experimental data / with both experimental and observational data. The true value of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$ is 0.133.

$L_3 = 0.900$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.001 [0.000, 0.008]	0.009 [0.000, 0.033]	0.033 [0.012, 0.058]
$L_1 = 0.900$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.925$	LB	0.000 [0.000, 0.000]	0.001 [0.000, 0.008]	0.009 [0.000, 0.033]	0.033 [0.012, 0.058]	0.058 [0.037, 0.083]
$L_1 = 0.925$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.950$	LB	0.001 [0.000, 0.008]	0.009 [0.000, 0.033]	0.033 [0.012, 0.058]	0.058 [0.037, 0.083]	0.083 [0.062, 0.108]
$L_1 = 0.950$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.975$	LB	0.009 [0.000, 0.033]	0.033 [0.012, 0.058]	0.058 [0.037, 0.083]	0.083 [0.062, 0.108]	0.108 [0.087, 0.133]
$L_1 = 0.975$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 1.000$	LB	0.033 [0.012, 0.058]	0.058 [0.037, 0.083]	0.083 [0.062, 0.108]	0.108 [0.087, 0.133]	0.133 [0.112, 0.158]
$L_1 = 1.000$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]

$L_3 = 0.925$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.001 [0.000, 0.008]	0.009 [0.000, 0.033]	0.033 [0.012, 0.058]
$L_1 = 0.900$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.925$	LB	0.001 [0.000, 0.008]	0.001 [0.000, 0.008]	0.009 [0.000, 0.033]	0.033 [0.012, 0.058]	0.058 [0.037, 0.083]
$L_1 = 0.925$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.950$	LB	0.009 [0.000, 0.033]	0.009 [0.000, 0.033]	0.033 [0.012, 0.058]	0.058 [0.037, 0.083]	0.083 [0.062, 0.108]
$L_1 = 0.950$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.975$	LB	0.033 [0.012, 0.058]	0.033 [0.012, 0.058]	0.058 [0.037, 0.083]	0.083 [0.062, 0.108]	0.108 [0.087, 0.133]
$L_1 = 0.975$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 1.000$	LB	0.058 [0.037, 0.083]	0.058 [0.037, 0.083]	0.083 [0.062, 0.108]	0.108 [0.087, 0.133]	0.133 [0.112, 0.158]
$L_1 = 1.000$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]

$L_3 = 0.950$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.001 [0.000, 0.008]	0.001 [0.000, 0.008]	0.001 [0.000, 0.008]	0.009 [0.000, 0.033]	0.033 [0.012, 0.058]
$L_1 = 0.900$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.925$	LB	0.009 [0.000, 0.033]	0.009 [0.000, 0.033]	0.009 [0.000, 0.033]	0.033 [0.012, 0.058]	0.058 [0.037, 0.083]
$L_1 = 0.925$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.950$	LB	0.033 [0.012, 0.058]	0.033 [0.012, 0.058]	0.033 [0.012, 0.058]	0.058 [0.037, 0.083]	0.083 [0.062, 0.108]
$L_1 = 0.950$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.975$	LB	0.058 [0.037, 0.083]	0.058 [0.037, 0.083]	0.058 [0.037, 0.083]	0.083 [0.062, 0.108]	0.108 [0.087, 0.133]
$L_1 = 0.975$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 1.000$	LB	0.083 [0.062, 0.108]	0.083 [0.062, 0.108]	0.083 [0.062, 0.108]	0.108 [0.087, 0.133]	0.133 [0.112, 0.158]
$L_1 = 1.000$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]

$L_3 = 0.975$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.009 [0.000, 0.033]	0.009 [0.000, 0.033]	0.009 [0.000, 0.033]	0.009 [0.000, 0.033]	0.033 [0.012, 0.058]
$L_1 = 0.900$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.925$	LB	0.033 [0.012, 0.058]	0.033 [0.012, 0.058]	0.033 [0.012, 0.058]	0.033 [0.012, 0.058]	0.058 [0.037, 0.083]
$L_1 = 0.925$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.950$	LB	0.058 [0.037, 0.083]	0.058 [0.037, 0.083]	0.058 [0.037, 0.083]	0.058 [0.037, 0.083]	0.083 [0.062, 0.108]
$L_1 = 0.950$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.975$	LB	0.083 [0.062, 0.108]	0.083 [0.062, 0.108]	0.083 [0.062, 0.108]	0.083 [0.062, 0.108]	0.108 [0.087, 0.133]
$L_1 = 0.975$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 1.000$	LB	0.108 [0.087, 0.133]	0.108 [0.087, 0.133]	0.108 [0.087, 0.133]	0.108 [0.087, 0.133]	0.133 [0.112, 0.158]
$L_1 = 1.000$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]

$L_3 = 1.000$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.033 [0.012, 0.058]	0.033 [0.012, 0.058]	0.033 [0.012, 0.058]	0.033 [0.012, 0.058]	0.033 [0.012, 0.058]
$L_1 = 0.900$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.925$	LB	0.058 [0.037, 0.083]	0.058 [0.037, 0.083]	0.058 [0.037, 0.083]	0.058 [0.037, 0.083]	0.058 [0.037, 0.083]
$L_1 = 0.925$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.950$	LB	0.083 [0.062, 0.108]	0.083 [0.062, 0.108]	0.083 [0.062, 0.108]	0.083 [0.062, 0.108]	0.083 [0.062, 0.108]
$L_1 = 0.950$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 0.975$	LB	0.108 [0.087, 0.133]	0.108 [0.087, 0.133]	0.108 [0.087, 0.133]	0.108 [0.087, 0.133]	0.108 [0.087, 0.133]
$L_1 = 0.975$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]
$L_1 = 1.000$	LB	0.133 [0.112, 0.158]	0.133 [0.112, 0.158]	0.133 [0.112, 0.158]	0.133 [0.112, 0.158]	0.133 [0.112, 0.158]
$L_1 = 1.000$	UB	0.233 [0.212, 0.258]	0.208 [0.187, 0.233]	0.183 [0.162, 0.208]	0.158 [0.137, 0.183]	0.133 [0.112, 0.158]

Obs.: LB: 0.000 [0.000,0.000] UB: 0.350 [0.322, 0.380]

Table 18: Results of the estimates of bounds of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ under Assumption (vi) and $N = 1000$ with observational data. The true value of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ is 0.333.

$L_3 = 0.900$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	-1.929 [-2.000, -1.667]	-1.713 [-1.909, -1.458]	-1.464 [-1.606, -1.250]	-1.214 [-1.303, -1.042]	-0.965 [-1.000, -0.833]
$L_1 = 0.900$	UB	1.965 [1.833, 2.000]	1.965 [1.833, 2.000]	1.965 [1.833, 2.000]	1.965 [1.833, 2.000]	1.965 [1.833, 2.000]
$L_1 = 0.925$	LB	-1.498 [-1.819, -1.250]	-1.498 [-1.819, -1.250]	-1.248 [-1.516, -1.042]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]
$L_1 = 0.925$	UB	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]
$L_1 = 0.950$	LB	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]	-0.499 [-0.606, -0.417]
$L_1 = 0.950$	UB	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]
$L_1 = 0.975$	LB	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.250 [-0.303, -0.208]
$L_1 = 0.975$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]

$L_3 = 0.925$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	-1.929 [-2.000, -1.667]	-1.713 [-1.909, -1.458]	-1.464 [-1.606, -1.250]	-1.214 [-1.303, -1.042]	-0.965 [-1.000, -0.833]
$L_1 = 0.900$	UB	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]
$L_1 = 0.925$	LB	-1.498 [-1.819, -1.250]	-1.498 [-1.819, -1.250]	-1.248 [-1.516, -1.042]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]
$L_1 = 0.925$	UB	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]	1.749 [1.625, 1.909]
$L_1 = 0.950$	LB	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]	-0.499 [-0.606, -0.417]
$L_1 = 0.950$	UB	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]
$L_1 = 0.975$	LB	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.250 [-0.303, -0.208]
$L_1 = 0.975$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]

$L_3 = 0.950$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	-1.929 [-2.000, -1.667]	-1.713 [-1.909, -1.458]	-1.464 [-1.606, -1.250]	-1.214 [-1.303, -1.042]	-0.965 [-1.000, -0.833]
$L_1 = 0.900$	UB	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]
$L_1 = 0.925$	LB	-1.498 [-1.819, -1.250]	-1.498 [-1.819, -1.250]	-1.248 [-1.516, -1.042]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]
$L_1 = 0.925$	UB	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]
$L_1 = 0.950$	LB	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]	-0.499 [-0.606, -0.417]
$L_1 = 0.950$	UB	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]
$L_1 = 0.975$	LB	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.250 [-0.303, -0.208]
$L_1 = 0.975$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]

$L_3 = 0.975$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	-1.929 [-2.000, -1.667]	-1.713 [-1.909, -1.458]	-1.464 [-1.606, -1.250]	-1.214 [-1.303, -1.042]	-0.965 [-1.000, -0.833]
$L_1 = 0.900$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]
$L_1 = 0.925$	LB	-1.498 [-1.819, -1.250]	-1.498 [-1.819, -1.250]	-1.248 [-1.516, -1.042]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]
$L_1 = 0.925$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]
$L_1 = 0.950$	LB	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]	-0.499 [-0.606, -0.417]
$L_1 = 0.950$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]
$L_1 = 0.975$	LB	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.250 [-0.303, -0.208]
$L_1 = 0.975$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]

$L_3 = 1.000$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	-1.929 [-2.000, -1.667]	-1.713 [-1.909, -1.458]	-1.464 [-1.606, -1.250]	-1.214 [-1.303, -1.042]	-0.965 [-1.000, -0.833]
$L_1 = 0.900$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]
$L_1 = 0.925$	LB	-1.498 [-1.819, -1.250]	-1.498 [-1.819, -1.250]	-1.248 [-1.516, -1.042]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]
$L_1 = 0.925$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]
$L_1 = 0.950$	LB	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]	-0.499 [-0.606, -0.417]
$L_1 = 0.950$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]
$L_1 = 0.975$	LB	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.250 [-0.303, -0.208]
$L_1 = 0.975$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]

Table 19: Results of the estimates of bounds of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ under Assumption (vi) and $N = 1000$ with both experimental and observational data. The true value of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ is 0.333.

$L_3 = 0.900$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	-1.758 [-2.000, -1.436]	-1.556 [-1.835, -1.241]	-1.307 [-1.546, -1.022]	-1.057 [-1.268, -0.795]	-0.808 [-1.000, -0.564]
$L_1 = 0.900$	UB	1.926 [1.767, 2.000]	1.926 [1.767, 2.000]	1.926 [1.767, 2.000]	1.926 [1.767, 2.000]	1.917 [1.752, 2.000]
$L_1 = 0.925$	LB	-1.475 [-1.769, -1.248]	-1.461 [-1.769, -1.198]	-1.216 [-1.467, -0.998]	-0.966 [-1.164, -0.788]	-0.716 [-0.862, -0.564]
$L_1 = 0.925$	UB	1.744 [1.625, 1.909]	1.744 [1.625, 1.909]	1.744 [1.625, 1.909]	1.744 [1.625, 1.909]	1.727 [1.551, 1.909]
$L_1 = 0.950$	LB	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]	-0.499 [-0.606, -0.417]
$L_1 = 0.950$	UB	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.482 [1.321, 1.606]
$L_1 = 0.975$	LB	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.250 [-0.303, -0.208]
$L_1 = 0.975$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.249 [1.208, 1.303]	1.232 [1.092, 1.303]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	0.999 [1.000, 1.000]	0.983 [0.844, 1.000]

$L_3 = 0.925$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	-1.758 [-2.000, -1.436]	-1.556 [-1.835, -1.241]	-1.307 [-1.546, -1.022]	-1.057 [-1.268, -0.795]	-0.808 [-1.000, -0.564]
$L_1 = 0.900$	UB	1.745 [1.625, 1.909]	1.745 [1.625, 1.909]	1.745 [1.625, 1.909]	1.745 [1.625, 1.909]	1.745 [1.625, 1.909]
$L_1 = 0.925$	LB	-1.475 [-1.769, -1.248]	-1.461 [-1.769, -1.198]	-1.216 [-1.467, -0.998]	-0.966 [-1.164, -0.788]	-0.716 [-0.862, -0.564]
$L_1 = 0.925$	UB	1.704 [1.540, 1.909]	1.704 [1.540, 1.909]	1.704 [1.540, 1.909]	1.704 [1.540, 1.909]	1.695 [1.530, 1.909]
$L_1 = 0.950$	LB	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]	-0.499 [-0.606, -0.417]
$L_1 = 0.950$	UB	1.496 [1.416, 1.606]	1.496 [1.416, 1.606]	1.496 [1.416, 1.606]	1.495 [1.413, 1.606]	1.479 [1.321, 1.606]
$L_1 = 0.975$	LB	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.250 [-0.303, -0.208]
$L_1 = 0.975$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.249 [1.208, 1.303]	1.232 [1.092, 1.303]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	0.999 [1.000, 1.000]	0.983 [0.844, 1.000]

$L_3 = 0.950$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	-1.758 [-2.000, -1.436]	-1.556 [-1.835, -1.241]	-1.307 [-1.546, -1.022]	-1.057 [-1.268, -0.795]	-0.808 [-1.000, -0.564]
$L_1 = 0.900$	UB	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]	1.499 [1.417, 1.606]
$L_1 = 0.925$	LB	-1.475 [-1.769, -1.248]	-1.461 [-1.769, -1.198]	-1.216 [-1.467, -0.998]	-0.966 [-1.164, -0.788]	-0.716 [-0.862, -0.564]
$L_1 = 0.925$	UB	1.496 [1.416, 1.606]	1.496 [1.416, 1.606]	1.496 [1.416, 1.606]	1.496 [1.416, 1.606]	1.495 [1.413, 1.606]
$L_1 = 0.950$	LB	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]	-0.499 [-0.606, -0.417]
$L_1 = 0.950$	UB	1.454 [1.314, 1.606]	1.454 [1.314, 1.606]	1.454 [1.314, 1.606]	1.454 [1.314, 1.606]	1.446 [1.292, 1.606]
$L_1 = 0.975$	LB	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.250 [-0.303, -0.208]
$L_1 = 0.975$	UB	1.246 [1.200, 1.303]	1.246 [1.200, 1.303]	1.246 [1.200, 1.303]	1.246 [1.186, 1.303]	1.229 [1.092, 1.303]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	0.999 [1.000, 1.000]	0.983 [0.844, 1.000]

$L_3 = 0.975$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	-1.758 [-2.000, -1.436]	-1.556 [-1.835, -1.241]	-1.307 [-1.546, -1.022]	-1.057 [-1.268, -0.795]	-0.808 [-1.000, -0.564]
$L_1 = 0.900$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]
$L_1 = 0.925$	LB	-1.475 [-1.769, -1.248]	-1.461 [-1.769, -1.198]	-1.216 [-1.467, -0.998]	-0.966 [-1.164, -0.788]	-0.716 [-0.862, -0.564]
$L_1 = 0.925$	UB	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]	1.250 [1.208, 1.303]
$L_1 = 0.950$	LB	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]	-0.499 [-0.606, -0.417]
$L_1 = 0.950$	UB	1.246 [1.200, 1.303]	1.246 [1.200, 1.303]	1.246 [1.200, 1.303]	1.246 [1.200, 1.303]	1.246 [1.186, 1.303]
$L_1 = 0.975$	LB	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.250 [-0.303, -0.208]
$L_1 = 0.975$	UB	1.205 [1.077, 1.303]	1.205 [1.077, 1.303]	1.205 [1.077, 1.303]	1.205 [1.077, 1.303]	1.196 [1.046, 1.303]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	0.993 [0.908, 1.000]	0.993 [0.908, 1.000]	0.993 [0.908, 1.000]	0.993 [0.908, 1.000]	0.976 [0.836, 1.000]

$L_3 = 1.000$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	-1.730 [-2.000, -1.384]	-1.506 [-1.813, -1.165]	-1.256 [-1.501, -0.941]	-1.007 [-1.229, -0.730]	-0.757 [-0.959, -0.504]
$L_1 = 0.900$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]
$L_1 = 0.925$	LB	-1.475 [-1.769, -1.248]	-1.454 [-1.769, -1.165]	-1.204 [-1.467, -0.941]	-0.954 [-1.164, -0.730]	-0.705 [-0.862, -0.504]
$L_1 = 0.925$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]
$L_1 = 0.950$	LB	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.998 [-1.213, -0.833]	-0.749 [-0.909, -0.625]	-0.499 [-0.606, -0.417]
$L_1 = 0.950$	UB	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]
$L_1 = 0.975$	LB	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.499 [-0.606, -0.417]	-0.250 [-0.303, -0.208]
$L_1 = 0.975$	UB	0.993 [0.908, 1.000]	0.993 [0.908, 1.000]	0.993 [0.908, 1.000]	0.993 [0.908, 1.000]	0.993 [0.908, 1.000]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	0.911 [0.651, 1.000]	0.911 [0.651, 1.000]	0.911 [0.651, 1.000]	0.911 [0.651, 1.000]	0.911 [0.651, 1.000]

Table 20: Results of the estimates of bounds of $\mathbb{E}[(Y_1 - Y_0)^2]$ under Assumption (vi) and $N = 1000$ with experimental data / with both experimental and observational data. The true value of $\mathbb{E}[(Y_1 - Y_0)^2]$ is 0.267.

$L_3 = 0.900$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.900$	UB	0.869 [0.842, 0.898]	0.869 [0.842, 0.898]	0.869 [0.842, 0.898]	0.869 [0.842, 0.898]	0.869 [0.842, 0.898]
$L_1 = 0.925$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.925$	UB	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]
$L_1 = 0.950$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.950$	UB	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]
$L_1 = 0.975$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.975$	UB	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]
$L_1 = 1.000$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 1.000$	UB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]

$L_3 = 0.925$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.900$	UB	0.869 [0.842, 0.898]	0.869 [0.842, 0.898]	0.869 [0.842, 0.898]	0.869 [0.842, 0.898]	0.769 [0.742, 0.798]
$L_1 = 0.925$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.925$	UB	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]
$L_1 = 0.950$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.950$	UB	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]
$L_1 = 0.975$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.975$	UB	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]
$L_1 = 1.000$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 1.000$	UB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]

$L_3 = 0.950$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.900$	UB	0.869 [0.842, 0.898]	0.869 [0.842, 0.898]	0.869 [0.842, 0.898]	0.769 [0.742, 0.798]	0.669 [0.642, 0.698]
$L_1 = 0.925$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.925$	UB	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.619 [0.592, 0.648]
$L_1 = 0.950$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.950$	UB	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]
$L_1 = 0.975$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.975$	UB	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]
$L_1 = 1.000$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 1.000$	UB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]

$L_3 = 0.975$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.900$	UB	0.869 [0.842, 0.898]	0.869 [0.842, 0.898]	0.769 [0.742, 0.798]	0.669 [0.642, 0.698]	0.569 [0.542, 0.598]
$L_1 = 0.925$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.925$	UB	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.619 [0.592, 0.648]	0.519 [0.492, 0.548]
$L_1 = 0.950$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.950$	UB	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.469 [0.442, 0.498]
$L_1 = 0.975$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.975$	UB	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]
$L_1 = 1.000$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 1.000$	UB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]

$L_3 = 1.000$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.900$	UB	0.869 [0.842, 0.898]	0.769 [0.742, 0.798]	0.669 [0.642, 0.698]	0.569 [0.542, 0.598]	0.469 [0.442, 0.498]
$L_1 = 0.925$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.925$	UB	0.719 [0.692, 0.748]	0.719 [0.692, 0.748]	0.619 [0.592, 0.648]	0.519 [0.492, 0.548]	0.419 [0.392, 0.448]
$L_1 = 0.950$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.950$	UB	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.569 [0.542, 0.598]	0.469 [0.442, 0.498]	0.369 [0.342, 0.398]
$L_1 = 0.975$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 0.975$	UB	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.419 [0.392, 0.448]	0.319 [0.292, 0.348]
$L_1 = 1.000$	LB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]
$L_1 = 1.000$	UB	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]	0.269 [0.242, 0.298]

Table 21: Results of the estimates of bounds of $\mathbb{E}[(Y_1 - Y_0)^2]$ under Assumption (vi) and $N = 1000$ with observational data. The true value of $\mathbb{E}[(Y_1 - Y_0)^2]$ is 0.267.

$L_3 = 0.900$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.900$	UB	1.200 [1.175, 1.219]	1.200 [1.175, 1.219]	1.197 [1.172, 1.218]	1.175 [1.149, 1.198]	1.150 [1.124, 1.173]
$L_1 = 0.925$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.925$	UB	1.100 [1.075, 1.119]	1.100 [1.075, 1.119]	1.097 [1.072, 1.118]	1.075 [1.049, 1.098]	1.050 [1.024, 1.073]
$L_1 = 0.950$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.950$	UB	1.000 [0.975, 1.019]	1.000 [0.975, 1.019]	0.997 [0.972, 1.018]	0.975 [0.949, 0.998]	0.950 [0.924, 0.973]
$L_1 = 0.975$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.975$	UB	0.900 [0.875, 0.919]	0.900 [0.875, 0.919]	0.897 [0.872, 0.918]	0.875 [0.849, 0.898]	0.850 [0.824, 0.873]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	0.800 [0.775, 0.819]	0.800 [0.775, 0.819]	0.797 [0.772, 0.818]	0.775 [0.749, 0.798]	0.750 [0.724, 0.773]

$L_3 = 0.925$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.900$	UB	1.200 [1.175, 1.219]	1.197 [1.172, 1.218]	1.175 [1.149, 1.198]	1.150 [1.124, 1.173]	1.075 [1.049, 1.098]
$L_1 = 0.925$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.925$	UB	1.100 [1.075, 1.119]	1.100 [1.075, 1.119]	1.097 [1.072, 1.118]	1.075 [1.049, 1.098]	1.050 [1.024, 1.073]
$L_1 = 0.950$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.950$	UB	1.000 [0.975, 1.019]	1.000 [0.975, 1.019]	0.997 [0.972, 1.018]	0.975 [0.949, 0.998]	0.950 [0.924, 0.973]
$L_1 = 0.975$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.975$	UB	0.900 [0.875, 0.919]	0.900 [0.875, 0.919]	0.897 [0.872, 0.918]	0.875 [0.849, 0.898]	0.850 [0.824, 0.873]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	0.800 [0.775, 0.819]	0.800 [0.775, 0.819]	0.797 [0.772, 0.818]	0.775 [0.749, 0.798]	0.750 [0.724, 0.773]

$L_3 = 0.950$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.900$	UB	1.197 [1.172, 1.218]	1.175 [1.149, 1.198]	1.150 [1.124, 1.173]	1.075 [1.049, 1.098]	1.000 [0.974, 1.023]
$L_1 = 0.925$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.925$	UB	1.100 [1.075, 1.119]	1.097 [1.072, 1.118]	1.075 [1.049, 1.098]	1.050 [1.024, 1.073]	0.975 [0.949, 0.998]
$L_1 = 0.950$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.950$	UB	1.000 [0.975, 1.019]	1.000 [0.975, 1.019]	0.997 [0.972, 1.018]	0.975 [0.949, 0.998]	0.950 [0.924, 0.973]
$L_1 = 0.975$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.975$	UB	0.900 [0.875, 0.919]	0.900 [0.875, 0.919]	0.897 [0.872, 0.918]	0.875 [0.849, 0.898]	0.850 [0.824, 0.873]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	0.800 [0.775, 0.819]	0.800 [0.775, 0.819]	0.797 [0.772, 0.818]	0.775 [0.749, 0.798]	0.750 [0.724, 0.773]

$L_3 = 0.975$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.900$	UB	1.175 [1.149, 1.198]	1.150 [1.124, 1.173]	1.075 [1.049, 1.098]	1.000 [0.974, 1.023]	0.925 [0.899, 0.948]
$L_1 = 0.925$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.925$	UB	1.097 [1.072, 1.118]	1.075 [1.049, 1.098]	1.050 [1.024, 1.073]	0.975 [0.949, 0.998]	0.900 [0.874, 0.923]
$L_1 = 0.950$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.950$	UB	1.000 [0.975, 1.019]	0.997 [0.972, 1.018]	0.975 [0.949, 0.998]	0.950 [0.924, 0.973]	0.875 [0.849, 0.898]
$L_1 = 0.975$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.975$	UB	0.900 [0.875, 0.919]	0.900 [0.875, 0.919]	0.897 [0.872, 0.918]	0.875 [0.849, 0.898]	0.850 [0.824, 0.873]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	0.800 [0.775, 0.819]	0.800 [0.775, 0.819]	0.797 [0.772, 0.818]	0.775 [0.749, 0.798]	0.750 [0.724, 0.773]

$L_3 = 1.000$

		$L_2 = 0.900$	$L_2 = 0.925$	$L_2 = 0.950$	$L_2 = 0.975$	$L_2 = 1.000$
$L_1 = 0.900$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.900$	UB	1.150 [1.124, 1.173]	1.075 [1.049, 1.098]	1.000 [0.974, 1.023]	0.925 [0.899, 0.948]	0.850 [0.824, 0.873]
$L_1 = 0.925$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.925$	UB	1.075 [1.049, 1.098]	1.050 [1.024, 1.073]	0.975 [0.949, 0.998]	0.900 [0.874, 0.923]	0.825 [0.799, 0.848]
$L_1 = 0.950$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.950$	UB	0.997 [0.972, 1.018]	0.975 [0.949, 0.998]	0.950 [0.924, 0.973]	0.875 [0.849, 0.898]	0.800 [0.774, 0.823]
$L_1 = 0.975$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 0.975$	UB	0.900 [0.875, 0.919]	0.897 [0.872, 0.918]	0.875 [0.849, 0.898]	0.850 [0.824, 0.873]	0.775 [0.749, 0.798]
$L_1 = 1.000$	LB	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$L_1 = 1.000$	UB	0.800 [0.775, 0.819]	0.800 [0.775, 0.819]	0.797 [0.772, 0.818]	0.775 [0.749, 0.798]	0.750 [0.724, 0.773]

E.2 Under Identification Assumptions

Next, we illustrate the properties of the point estimator under an identification assumption. There are no baselines because no other identification assumptions exist for the joint POs/OVs probabilities with discrete treatments and outcomes.

Settings. We set the values of $\mathcal{P}$ as presented in Table 22. For this purpose, we impose the following identification assumptions.

(I). $0 \leq Y_s - Y_t \leq 1$ almost surely for any $s > t$ (Assumption 12) for experimental data,

(II). $0 \leq Y_s - Y_t \leq 1$ almost surely for any $s > t$ (Assumption 12) and the exogeneity (Assumption 1) for observational data.

The setting in Table 22 satisfies the above identification assumptions.

Results. Table 23 (resp. 24, 25) reports the estimates of the values of $\mathbb{P}(Y_0 = 0, Y_1 = 0, Y_2 = 1)$ (resp. $\mathbb{E}[Y_1 - Y_0 | X = 2, Y = 2]$, $\mathbb{E}[(Y_1 - Y_0)^2]$) with sample size $N = 10, 100, 1000, 10000$. All estimators converge to the ground truth, and their 95% CIs become tighter as the sample size increases. When the sample size is relatively large (e.g., $N = 1000$ or $N = 10000$), the estimates are reliable.

Table 22: Setting of Simulations

$X=0,Y=0$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	1/21	0	0	0	0	0	0	0	0
$Y_2=1$	1/21	1/21	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	0
$X=0,Y=1$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	1/21	0	0	0	0
$Y_2=2$	0	0	0	0	1/21	1/21	0	0	0
$X=0,Y=2$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	1/21
$X=1,Y=0$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	1/21	0	0	0	0	0	0	0	0
$Y_2=1$	1/21	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	0
$X=1,Y=1$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	1/21	0	0	1/21	0	0	0	0
$Y_2=2$	0	0	0	0	1/21	0	0	0	0
$X=1,Y=2$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	1/21	0	0	1/21
$X=2,Y=0$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	1/21	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	0
$X=2,Y=1$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	1/21	1/21	0	0	1/21	0	0	0	0
$Y_2=2$	0	0	0	0	0	0	0	0	0
$X=2,Y=2$	$Y_0=0$			$Y_0=1$			$Y_0=2$		
	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$	$Y_1=0$	$Y_1=1$	$Y_1=2$
$Y_2=0$	0	0	0	0	0	0	0	0	0
$Y_2=1$	0	0	0	0	0	0	0	0	0
$Y_2=2$	0	0	0	0	1/21	1/21	0	0	1/21

Table 23: Results of the estimates of $\mathbb{P}(Y_0=0,Y_1=0,Y_2=1)$ under identification assumptions. We report the means and 95% confidence intervals (CIs). The ground truth of $\mathbb{P}(Y_0=0,Y_1=0,Y_2=1)$ is 0.143.

MAs	Data	$N=10$	$N=100$	$N=1000$	$N=10000$
(I)	Exp.	0.142 [0.000,0.353]	0.145 [0.085,0.225]	0.142 [0.120,0.166]	0.143 [0.135,0.150]
(II)	Obs.	0.190 [0.000,0.500]	0.143 [0.020,0.309]	0.152 [0.085,0.226]	0.144 [0.128,0.157]

Table 24: Results of the estimates of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ under identification assumptions. We report the means and 95% confidence intervals (CIs). The ground truth of $\mathbb{E}[Y_1 - Y_0|X = 2, Y = 2]$ is 0.333.

MAs	Data	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(II)	Obs.	0.431 [0.000,1.000]	0.332 [0.042,0.700]	0.321 [0.202,0.465]	0.334 [0.288,0.382]

Table 25: Results of the estimates of $\mathbb{E}[(Y_1 - Y_0)^2]$ under identification assumptions. We report the means and 95% confidence intervals (CIs). The ground truth of $\mathbb{E}[(Y_1 - Y_0)^2]$ is 0.286.

MAs	Data	$N = 10$	$N = 100$	$N = 1000$	$N = 10000$
(I)	Exp.	0.313 [0.000,0.600]	0.279 [0.200,0.350]	0.288 [0.264,0.314]	0.285 [0.278,0.294]
(II)	Obs.	0.395 [0.000,1.000]	0.303 [0.077,0.561]	0.285 [0.169,0.388]	0.282 [0.238,0.319]