Identification and Bounding of Joint Expectation over Potential Outcomes

Yuta Kawakami and Jin Tian
Mohamed bin Zayed University of Artificial Intelligence
{Yuta.Kawakami,Jin.Tian}@mbzuai.ac.ae

Abstract

The joint expectation of functions of potential outcomes (Y_1, Y_0) is a fundamental quantity in causal inference, particularly when investigating the variation or heterogeneity of causal effects. This paper provides novel assumptions for their identification: comonotonicity and countercomonotonicity. These assumptions offer a unified framework for their identification across discrete, continuous, and mixed outcome variables. In the absence of these assumptions, we derive sharp bounds for two broad classes of functions, yielding new results for bounding moments of individual causal effects, which are central to measuring effect heterogeneity. We present corresponding estimation methods and illustrate them in simulation studies and real-world datasets.

1 Introduction

The joint expectation of a function $g(Y_1, Y_0)$ over potential outcomes (POs) is a fundamental quantity in causal inference, particularly for investigating the variation and heterogeneity of causal effects between an outcome (Y) and a binary treatment (X), where Y_x denotes PO $Y_{X=x}$. The joint expectation over POs is generally defined as $\mathbb{E}[g(Y_1, Y_0)] \triangleq \int g(y_1, y_0) dF_{(Y_1,Y_0)}(y_1, y_0)$, where $F_{(Y_1,Y_0)}(y_1, y_0)$ represents the joint cumulative distribution function (CDF) of POs. The joint expectation over POs plays a central role in characterizing various statistical aspects of causal effects such as moments, CDF, and quantiles, including those defined by the individual causal difference effect (ICDE), i.e., $Y_1 - Y_0$, and the individual causal ratio effect (ICRE), i.e., Y_1/Y_0. Recently, the moments of ICDE have been studied by Kawakami and Tian [2025], who established that the moments of ICDE are identifiable under a monotonicity assumption and also provided corresponding bounds when the monotonicity is violated. On the other hand, although ICRE has been historically discussed [Rosenbaum and Rubin, 1983, Greenland et al., 1999, Hernan and Robins, 2024], its moments—even the mean—have been rarely studied due to the difficulty in identifying and estimating them. These forms of counterfactual reasoning are substantially more complex, as they reside in the highest layer—$\mathcal{L}_3$—of Pearl's causal hierarchy [Bareinboim et al., 2022, Pearl and Mackenzie, 2018].

The joint expectations over POs may be identified by identifying the value of unobserved counterfactual outcomes. This can be achieved based on rank preservation assumptions [Heckman et al., 1997, Athey and Imbens, 2006, Wu et al., 2025], or strict monotonicity plus homogeneity assumptions [Xie et al., 2023, Nasr-Esfahany et al., 2023]. However, all these assumptions are quite restrictive.

In this paper, we present new assumptions, comonotonicity and countercomonotonicity, for identifying the joint expectations over POs. These assumptions offer a unified identification framework for joint expectations over POs across discrete, continuous, and mixed outcome variables. When the outcome variable is continuous, the comonotonicity reduces to the rank preservation assumption in [Heckman et al., 1997]. In general, the comonotonicity assumption is weaker than the rank preservation assumptions [Heckman et al., 1997, Wu et al., 2025] and the monotonicity assumptions [Xie et al., 2023, Kawakami and Tian, 2025]. When Y is binary, the comonotonicity assumption is

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Table 1: Summary of Identification Assumptions and Their Objectives.

Research	Outcome Type	Assumption	Objective
p. 504 in [Heckman et al., 1997]	Continuous	Rank Preservation	Values of Unobserved POs
Assumption 4.6 in [Wu et al., 2025]	Mixed	Rank Preservation with $\tilde{\rho}$	Values of Unobserved POs
Assumption of Theorem 1 in [Xie et al., 2023]	Continuous	Strict monotonicity over f_Y	Values of Unobserved POs
Definition 14 in [Tian and Pearl, 2000]	Binary	Monotonicity	PoC
Assumption 2 in [Kawakami and Tian, 2025]	Continuous	Monotonicity over f_Y	Moments of ICDE
Assumptions 1 and 2 in this paper	Mixed	Comonotonicity and Countercomonotonicity	Joint Expectations over POs

weaker than the monotonicity assumption in [Pearl, 1999, Tian and Pearl, 2000]. Table 1 summarizes the related identification assumptions and the corresponding targets of each study.

Our assumptions, much like those of monotonicity and rank preservation, may sometimes be implausible for researchers. Therefore, we next study bounding the joint expectations over POs in the absence of these assumptions. We obtain sharp bounds for joint expectation of two classes of g functions, and show that the results provide tighter bounds for moments of ICDE and, to the best of our knowledge, the first bounds for moments of ICRE.

Finally, we present methods for estimating the joint expectation over POs and their bounds from finite samples, and illustrate them in simulation studies and two real-world datasets.

The main contributions of the paper are as follows:

- We introduce two new identification assumptions, comonotonicity and countercomonotonicity, for $\mathbb{E}[g(Y_1, Y_0)]$. These assumptions are often weaker than existing alternatives and provide a unified framework applicable to discrete, continuous, and mixed outcome Y.
- We derive novel sharp bounds for $\mathbb{E}[g(Y_1, Y_0)]$ for two broad classes of functions g, leading to tighter bounds for moments of ICDE and the first bounds for moments of ICRE.
- We develop corresponding estimation procedures for implementation with finite samples.

2 Notation 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. Denote Ω_Y as the domain of Y, $\mathbb{E}[Y]$ as the expectation of Y, and $F_Y(y) = \mathbb{P}(Y \leq y)$ as the cumulative distribution function (CDF) of Y. The generalized inverse of the CDF is defined as $F_Y^{-1}(p) = \inf\{y; F_Y(y) \geq p\}$ for $p \in [0, 1]$.

Comonotonicity and countercomonotonicity. Comonotonicity and countercomonotonicity are foundational concepts in the study of dependence structures between random variables [Nelsen, 2006]. A pair of random variables (X, Y) is said to be comonotonic if there exists a random variable $U \sim \text{Unif}(0, 1)$ and monotonic increasing functions α and β such that $X = \alpha(U)$ and $Y = \beta(U)$. A pair of random variables (X, Y) is said to be countercomonotonic if there exists a random variable $U \sim \text{Unif}(0, 1)$ and monotonic functions α and β, one decreasing and the other increasing, such that $X = \alpha(U)$ and $Y = \beta(U)$.

Causal Effects. We use the language of Structural Causal Models (SCM) as our basic semantic and inferential framework [Pearl, 2009]. We denote the PO of Y under intervention $do(X = x)$ by Y_x. Researchers often consider the following SCM, $\mathcal{M}$:

$$Y := f_Y(X, U_Y), X := f_X(U_X), \tag{1}$$

where U_Y and U_X are latent exogenous variables following a joint distribution $\mathbb{P}(U_Y, U_X)$. Y is a discrete, continuous, or mixed variable, and X is a binary variable taking values in $\{0, 1\}$. The individual causal difference effect (ICDE) is defined as [Holland, 1986]:

$$\text{ICDE} \triangleq Y_1 - Y_0, \tag{2}$$

and the average causal difference effect is defined as ACDE $\triangleq \mathbb{E}[Y_1 - Y_0]$. For outcome variables that take strictly positive values, the individual causal ratio effect (ICRE) is defined as [Rosenbaum and Rubin, 1983, Greenland et al., 1999, Hernan and Robins, 2024]:

$$\text{ICRE} \triangleq Y_1/Y_0. \tag{3}$$

The average causal ratio effect (ACRE) is given by ACRE $\triangleq \mathbb{E}\left[Y_1/Y_0\right]$. The identification of ACRE is more difficult than the identification of ACDE because ACRE is not the causal risk ratio [Hernán and Robins, 2006], i.e., $\mathbb{E}\left[Y_1/Y_0\right] \neq \mathbb{E}[Y_1]/\mathbb{E}[Y_0]$.

3 Joint Expectations over POs

First, we show the formal definition of the joint expectation of a function g over POs.

Definition 1. *The joint CDF of* (Y_1, Y_0) *is defined as* $F_{(Y_1,Y_0)}(y_1, y_0) \triangleq \mathbb{P}(Y_1 \leq y_1, Y_0 \leq y_0)$.

The target of this paper is the joint expectation of a function g evaluated over the POs (Y_1, Y_0).

Definition 2. *The joint expectation of a measurable function* g *over POs* (Y_1, Y_0) *is defined as* $\mathbb{E}[g(Y_1, Y_0)] \triangleq \int g(y_1, y_0) dF_{(Y_1,Y_0)}(y_1, y_0)$.

When the function $g(y_1, y_0)$ takes the separable form $g_1(y_1) - g_0(y_0)$, as in the case of the ACDE, computing its expectation does not require the joint distribution of (Y_1, Y_0). Specifically, the expectation simplifies as $\mathbb{E}[g(Y_1, Y_0)] = \int g_1(y_1) dF_{Y_1}(y_1) - \int g_0(y_0) dF_{Y_0}(y_0)$, which depends only on the marginals Y_1 and Y_0.

3.1 Why do joint expectations over POs matter?

Joint expectations over POs play a central role in capturing various statistical aspects of the variation or heterogeneity in causal effects for both ICDE and ICRE. Firstly, we can represent the CDF of ICDE and ICRE in the form of a joint expectation over POs as follows:

$$\mathbb{P}(Y_1 - Y_0 < y) = \int \mathbb{I}(y_1 - y_0 < y) dF_{(Y_1,Y_0)}(y_1, y_0), \tag{4}$$

$$\mathbb{P}(Y_1/Y_0 < y) = \int \mathbb{I}(y_1/y_0 < y) dF_{(Y_1,Y_0)}(y_1, y_0). \tag{5}$$

Additionally, the moments of ICDE and ICRE are special cases of joint expectations over POs with $g(Y_1, Y_0) = (Y_1 - Y_0)^m$ and $g(Y_1, Y_0) = (Y_1/Y_0)^m$. So are their central moments with $g(Y_1, Y_0) = \{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\}^m$, and $g(Y_1, Y_0) = \{(Y_1/Y_0) - (\mathbb{E}[Y_1/Y_0])\}^m$. These moments encompass statistical measures such as mean, variance, skewness, and kurtosis, and have recently been studied by Kawakami and Tian [2025] in the context of ICDE. They go beyond the average causal effect (which represents the first moment) and characterize the shape of the distribution of causal effects. The shape of the distribution of causal effects uncovers causal effect heterogeneity. In the context of ICDE, existing studies on effect heterogeneity focus on the conditional average treatment effect (CATE), defined as $\mathbb{E}[Y_1 - Y_0|W]$, where W denotes observed covariates. However, CATE captures heterogeneity only at the level of subpopulations defined by W, while these joint expectations over POs capture heterogeneity at the individual level, providing insights beyond subgroup averages.

Furthermore, the probability of necessity and sufficiency (PNS) for a continuous outcome [Kawakami et al., 2024] is encompassed within the joint expectation over POs, i.e., $\mathbb{P}(Y_0 < y \leq Y_1) = \int \mathbb{I}(y_0 < y \leq y_1) dF_{(Y_1,Y_0)}(y_1, y_0)$. PNS reveals another aspect of the causal relationship. As another example, the causal parameter $\mathbb{P}(Y_1 > Y_0)$ studied by Fay et al. [2018], which is the probability that a randomly selected individual will have a larger response under action than under control, is also included within the joint expectation over POs, i.e., $\mathbb{P}(Y_1 > Y_0) = \int \mathbb{I}(y_1 > y_0) dF_{(Y_1,Y_0)}(y_1, y_0)$. This causal parameter has received considerable attention in the field of causal inference [Gadbury et al., 2004, Fan and Park, 2010, Huang et al., 2016, Lu et al., 2019, Gabriel et al., 2024].

4 Identification of Joint Expectations over POs

We begin by presenting the Fréchet–Hoeffding bounds on the joint CDF of POs.

Lemma 1. *Given F_{Y_0} and F_{Y_1}, the joint CDF of POs $F_{(Y_1,Y_0)}(y_1, y_0)$ is bounded by $F^L_{(Y_1,Y_0)}(y_1, y_0) \leq F_{(Y_1,Y_0)}(y_1, y_0) \leq F^U_{(Y_1,Y_0)}(y_1, y_0)$, where*

$$F^U_{(Y_1,Y_0)}(y_1, y_0) = \min\{F_{Y_0}(y_0), F_{Y_1}(y_1)\}, F^L_{(Y_1,Y_0)}(y_1, y_0) = \max\{F_{Y_0}(y_0) + F_{Y_1}(y_1) - 1, 0\}. \tag{6}$$

The upper bound is attained when (Y_1, Y_0) are comonotonic, while the lower bound is attained when (Y_1, Y_0) are countercomonotonic. These bounds are sharp given F_{Y_0} and F_{Y_1}.

We note that $F^U_{(Y_1,Y_0)}$ and $F^L_{(Y_1,Y_0)}$ defined by Eq. (6) are valid joint CDFs of (y_1, y_0) [Nelsen, 2006]. In addition, the marginal distributions of $F^U_{(Y_1,Y_0)}$ and $F^L_{(Y_1,Y_0)}$ over Y_0 are both equal to $F^U_{Y_0} = F^L_{Y_0} = F_{Y_0}$ [Nelsen, 2006].

We then introduce new identification assumptions for the joint expectation over POs, corresponding to comonotonicity and countercomonotonicity.

Assumption 1 (Comonotonicity)**.** *Potential outcomes (Y_1, Y_0) are comonotonic.*

Assumption 2 (Countercomonotonicity)**.** *Potential outcomes (Y_1, Y_0) are countercomonotonic.*

In terms of (Y_1, Y_0), comonotonicity means that whenever individual A would have a higher outcome than B under $do(X = 0)$, A would not have a lower outcome than B under $do(X = 1)$, almost surely. In practice, comonotonicity is plausible when the interventions differ only in intensity, dosage, or delivery mode—such as alternative dosing schedules of the same drug, short versus extended program duration, or online versus in-person delivery of the same service—so that individuals who benefit more under one condition also benefit more under the other. Comonotonicity is also reasonable when treatment effects are additive or proportional, or when both POs are increasing functions of a shared latent trait such as baseline severity or responsiveness. We refer to Appendix E for real-world applications that concretely illustrate the practical interpretation of comonotonicity.

In contrast, countercomonotonicity means that whenever individual A would have a higher outcome than B under $do(X = 0)$, A would not have a higher outcome than B under $do(X = 1)$, and vice versa. Countercomonotonicity is plausible when the two interventions benefit different types of individuals in systematically opposing ways, so that those who would perform better under one intervention tend to perform worse under the other. This can arise in remedial education programs that disproportionately benefit low-ability students, in targeted subsidies that primarily support the least productive firms, or in assistive technologies that compensate for physical or cognitive limitations.

We note that the comonotonicity and countercomonotonicity assumptions are applicable to any type of outcome variables—discrete, continuous, or mixed.

We define

$$\mathbb{E}^U[g(Y_1, Y_0)] \triangleq \int g(y_1, y_0) dF^U_{(Y_1,Y_0)}(y_1, y_0), \tag{7}$$

$$\mathbb{E}^L[g(Y_1, Y_0)] \triangleq \int g(y_1, y_0) dF^L_{(Y_1,Y_0)}(y_1, y_0). \tag{8}$$

Then, the joint expectation over POs is identifiable as below:

Theorem 1. *Under SCM $\mathcal{M}$, given F_{Y_0} and F_{Y_1}, we have the following two identification statements:*

(A). Under Assumption 1, $\mathbb{E}[g(Y_1, Y_0)] = \mathbb{E}^U[g(Y_1, Y_0)]$.

(B). Under Assumption 2, $\mathbb{E}[g(Y_1, Y_0)] = \mathbb{E}^L[g(Y_1, Y_0)]$.

4.1 Comparison with previous identification assumptions

We compare comonotonicity (Assumption 1) with previous identification assumptions in the literature. We note that countercomonotonicity has rarely been studied in prior work.

Strict monotonicity. Strict monotonicity is used for the identification of values of unobserved POs [Lu et al., 2020, Xie et al., 2023, Nasr-Esfahany et al., 2023].

Assumption 3 (Strict Monotonicity)**.** *The functions $f_Y(1, U_Y)$ and $f_Y(0, U_Y)$ are smooth and strictly monotone in U_Y, with the same direction of monotonicity.*

Proposition 1. *Assumption 3 implies Assumption 1; however, the converse implication does not hold.*

The strict monotonicity assumption is stronger than comonotonicity and is violated if the outcome variable is discrete or mixed.

Monotonicity for binary outcome. For the binary outcome Y, the following monotonicity assumption is often utilized in the probabilities of causation (PoC) literature [Pearl, 1999, Tian and Pearl, 2000]:

Assumption 4 (Monotonicity for Binary Outcome)**.** $\mathbb{P}(Y_0 = 1, Y_1 = 0) = 0$ *holds.*

Proposition 2. *When Y is binary, Assumption 1 is equivalent to the condition that either $\mathbb{P}(Y_0 = 1, Y_1 = 0) = 0$ or $\mathbb{P}(Y_0 = 0, Y_1 = 1) = 0$ holds, which is weaker than Assumption 4.*

Monotonicity over f_Y. Monotonicity over f_Y is used to identify PoC and moments of ICDE [Kawakami et al., 2024, Kawakami and Tian, 2025].

Assumption 5 (Monotonicity over f_Y)**.** *The function $f_Y(x, U_Y)$ is either (i) monotonic increasing on U_Y for all $x \in \{0, 1\}$ almost surely w.r.t. $\mathbb{P}_{U_Y}$, or (ii) monotonic decreasing on U_Y for all $x \in \{0, 1\}$ almost surely w.r.t. $\mathbb{P}_{U_Y}$.*

Proposition 3. *Assumption 5 implies Assumption 1; however, the converse need not hold for a given structural representation $Y = f_Y(X, U_Y)$. Nevertheless, if (Y_1, Y_0) are comonotonic, then there exists a latent variable U'_Y and a structural function $f'_Y(X, U'_Y)$ satisfying Assumption 5 such that $Y_x = f'_Y(x, U'_Y)$ for $x \in \{0, 1\}$.*

Therefore, Assumption 1 is weaker than Assumption 5. Kawakami et al. [2024] also provide an equivalent assumption (monotonicity over Y_x) to Assumption 5 under their Assumption 3.6.

Rank preservation assumptions. The rank preservation assumption is often used for the identification of the values of the unobserved POs [Heckman et al., 1997].

Assumption 6 (Rank preservation assumption)**.** *For continuous CDFs F_{Y_1} and F_{Y_0}, $F_{Y_1}(y_1) = F_{Y_0}(y_0)$ holds for almost every subject whose POs are $(Y_1, Y_0) = (y_1, y_0)$.*

The rank preservation assumption is closely related to comonotonicity. In fact, they are equivalent when the outcome variable is continuous. The causal inference literature has often employed the rank preservation assumption, but has rarely explicitly considered comonotonicity.

Proposition 4. *For continuous outcome Y, Assumption 1 is equivalent to Assumption 6.*

The rank preservation assumption is only applicable to a continuous outcome. However, the comonotonicity assumption is applicable to any type of variables - discrete, continuous, or mixed. Wu et al. [2025] showed a weaker rank preservation assumption that is applicable to any type of variables using Kendall's rank correlation coefficient [Kendall, 1938]:

Definition 3 (Modified Kendall Rank Correlation Coefficient)**.** *Let $(x_1, y_1), \ldots, (x_n, y_n)$ be the observations of two random variables (X, Y); the modified Kendall rank correlation coefficient is defined as $\tilde{\rho}_n(X, Y) = \sum_{1 \le i < j \le n} \frac{\text{sign}\{(x_i - x_j)(y_i - y_j)\}}{\sqrt{n(n-1)/2 - T_x}\sqrt{n(n-1)/2 - T_y}}$, where T_x is the number of tied pairs in $\{x_1, \ldots, x_n\}$, and T_y is the number of tied pairs in $\{y_1, \ldots, y_n\}$.*

Assumption 7 (Rank Preservation on Kendall's $\tilde{\rho}$)**.** $\tilde{\rho}_n(Y_1, Y_0) = 1$ *holds for all integers $n \ge 2$.*

The condition $\tilde{\rho}(Y_1, Y_0) = 1$ indicates that Y_1 and Y_0 are perfectly positively correlated. Rank preservation with Kendall's $\tilde{\rho}$ is weaker than rank preservation. Rank preservation with Kendall's $\tilde{\rho}$ accommodates both discrete and continuous outcomes. Both identification assumptions lead to the same identified results: Under SCM $\mathcal{M}$ and Assumptions 6 (or 7), given continuous CDF F_{Y_1} and F_{Y_0}, for any subjects whose POs are $(Y_1, Y_0) = (y_1, y_0)$, y_1 is identifiable by

$$y_1 = F^{-1}_{Y_1}(F_{Y_0}(y_0)). \tag{9}$$

Therefore, the joint expectations over POs are identified by

$$\mathbb{E}[g(Y_1, Y_0)] = \mathbb{E}[g(F^{-1}_{Y_1}(F_{Y_0}(Y_0)), Y_0)]. \tag{10}$$

We note that the right-hand side of Eq. (10) is equivalent to the right-hand side of Eq. (7) under Assumption 6 (or 7). We have the following relationship between Assumption 7 and Assumption 1:

Proposition 5. *Assumption 7 implies Assumption 1; however, the converse implication does not hold.*

Hence, Assumption 1 is weaker than Assumption 7. We note that the values of the unobserved POs are *not* identified in general under our comonotonicity assumption, and Eq. (10) does not hold in general under Assumption 1.

In summary, by Propositions 1-5, *our comonotonicity assumption (Assumption 1) offers a weaker identifying assumption for joint expectations over POs than existing assumptions in the literature.* Moreover, *it is applicable to any type of outcome variables - discrete, continuous, or mixed.*

Furthermore, we establish the following new identification theorem for unobserved POs under the countercomonotonicity assumption:

Theorem 2. *Under SCM $\mathcal{M}$ and Assumption 2, given F_{Y_0} and F_{Y_1}, when Y is continuous, for any subjects whose POs are $(Y_1, Y_0) = (y_1, y_0)$, y_1 is identifiable by*

$$y_1 = F^{-1}_{Y_1}(1 - F_{Y_0}(y_0)). \tag{11}$$

5 Bounding Joint Expectations over POs

Like monotonicity and rank preservation assumptions, our assumptions may sometimes be implausible to researchers. Therefore, we next study bounding the joint expectations over POs in the absence of these assumptions. We present two situations in which $\mathbb{E}^U[g(Y_1, Y_0)]$ and $\mathbb{E}^L[g(Y_1, Y_0)]$ provide sharp bounds for the joint expectation over POs $\mathbb{E}[g(Y_1, Y_0)]$. Specifically, we consider the following two classes of g functions:

Assumption 8 (Supermodularity of g)**.** *The function $g(y_1, y_0)$ is bounded and supermodular, i.e., the inequality $g(y_1', y_0') - g(y_1, y_0') \geq g(y_1', y_0) - g(y_1, y_0)$ holds for every $y_0, y_1, y_0', y_1' \in \Omega_Y$ such that $y_1 \leq y_1'$ and $y_0 \leq y_0'$.*

If g is twice differentiable, supermodularity means $\frac{\partial^2 g(y_1, y_0)}{\partial y_1 \partial y_0} \geq 0$ for any $y_1, y_0 \in \Omega_Y$.

Assumption 9 (Submodularity of g)**.** *The function $g(y_1, y_0)$ is bounded and submodular, i.e., the inequality $g(y_1', y_0') - g(y_1, y_0') \leq g(y_1', y_0) - g(y_1, y_0)$ holds for every $y_0, y_1, y_0', y_1' \in \Omega_Y$ such that $y_1 \leq y_1'$ and $y_0 \leq y_0'$.*

If g is twice differentiable, submodularity means $\frac{\partial^2 g(y_1, y_0)}{\partial y_1 \partial y_0} \leq 0$ for any $y_1, y_0 \in \Omega_Y$. Then, we have the following bounding theorem:

Theorem 3. *Under SCM $\mathcal{M}$, given F_{Y_0} and F_{Y_1}, we have the following two statements:*

(A). Under Assumption 8, $\mathbb{E}[g(Y_1, Y_0)]$ is bounded by $\mathbb{E}^L[g(Y_1, Y_0)] \leq \mathbb{E}[g(Y_1, Y_0)] \leq \mathbb{E}^U[g(Y_1, Y_0)]$.

(B). Under Assumption 9, $\mathbb{E}[g(Y_1, Y_0)]$ is bounded by $\mathbb{E}^U[g(Y_1, Y_0)] \leq \mathbb{E}[g(Y_1, Y_0)] \leq \mathbb{E}^L[g(Y_1, Y_0)]$.

These bounds are sharp given the marginals F_{Y_0} and F_{Y_1}.

Theorem 3 does not provide fully general bounds for an arbitrary function g. Still, it leads to new bounding results for the moments and central moments of ICDE and ICRE, as discussed in the sequel.

5.1 Bounds for the moments of ICDE

First, both the functions $g(y_1, y_0) = (y_1 - y_0)^m$ and $g(y_1, y_0) = \{(y_1 - y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\}^m$ are submodular when m is even.

Corollary 1. *Under SCM $\mathcal{M}$, if Y_1 and Y_0 are bounded, the even-order moments of the ICDE are bounded by $\mathbb{E}^U[(Y_1 - Y_0)^m] \leq \mathbb{E}[(Y_1 - Y_0)^m] \leq \mathbb{E}^L[(Y_1 - Y_0)^m]$, and the bounds are sharp.*

Corollary 2. *Under SCM $\mathcal{M}$, if Y_1 and Y_0 are bounded, the even-order central moments of the ICDE are bounded by $\mathbb{E}^U[\{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\}^m] \leq \mathbb{E}[\{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\}^m] \leq \mathbb{E}^L[\{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\}^m]$, and the bounds are sharp.*

The sharp upper bound above coincides with that provided by Kawakami and Tian [2025]. Our sharp lower bound is tighter than theirs, as their lower bound is not sharp.

On the other hand, when m is odd, the m-th moment of ICDE can be given by $\mathbb{E}[(Y_1 - Y_0)^m] = \sum_{l=0}^{m} \binom{m}{l}(-1)^l \mathbb{E}[Y_1^{m-l} Y_0^l]$. If Y_1 and Y_0 take non-negative values, then for any k, l, each $\mathbb{E}[Y_1^k Y_0^l]$ is bounded as $\mathbb{E}^L[Y_1^k Y_0^l] \leq \mathbb{E}[Y_1^k Y_0^l] \leq \mathbb{E}^U[Y_1^k Y_0^l]$ since the function $g(y_1, y_0) = y_1^k y_0^l$ is supermodular as $\frac{\partial^2 (y_1^k y_0^l)}{\partial y_1 \partial y_0} = kl y_1^{k-1} y_0^{l-1} \geq 0$.

Corollary 3. *Under SCM $\mathcal{M}$, if Y_1 and Y_0 are bounded and take non-negative values, the odd-order moments of the ICDE are bounded by*

$$\begin{aligned}
&\sum_{l=0}^{m} \binom{m}{l}(-1)^l \big\{\mathbb{I}((-1)^l = -1)\mathbb{E}^U[Y_1^{m-l} Y_0^l] + \mathbb{I}((-1)^l = 1)\mathbb{E}^L[Y_1^{m-l} Y_0^l]\big\} \leq \mathbb{E}[(Y_1 - Y_0)^m] \\
&\leq \sum_{l=0}^{m} \binom{m}{l}(-1)^l \big\{\mathbb{I}((-1)^l = -1)\mathbb{E}^L[Y_1^{m-l} Y_0^l] + \mathbb{I}((-1)^l = 1)\mathbb{E}^U[Y_1^{m-l} Y_0^l]\big\}.
\end{aligned} \tag{12}$$

The bounds for the odd-order central moments of ICDE are presented in Appendix B. These bounds are not sharp.

5.2 Bounds for the moments of ICRE

Theorem 3 provides bounds for the moments and central moments of ICRE, where the outcome variable is assumed to take strictly positive values (otherwise the ratio effect is not meaningful). To our knowledge, these bounds are novel and represent a new contribution to the literature.

First, $g(y_1, y_0) = (y_1/y_0)^m$ is submodular for any m when Y_1/Y_0 takes positive values. We have

Corollary 4. *Under SCM $\mathcal{M}$, if Y_1/Y_0 is bounded and takes positive values, the moments of the ICRE are bounded by $\mathbb{E}^U[(Y_1/Y_0)^m] \leq \mathbb{E}[(Y_1/Y_0)^m] \leq \mathbb{E}^L[(Y_1/Y_0)^m]$, and the bounds are sharp.*

The central moments of ICRE are bounded as follows:

Corollary 5. *Under SCM $\mathcal{M}$, if Y_1/Y_0 is bounded and takes positive values, the central moments of the ICRE are bounded by*

$$\begin{aligned}
&\sum_{l=0}^{m} \binom{m}{l} \big(\mathbb{I}((-1)^l = 1)\left(\mathbb{E}^U[Y_1/Y_0]\right)^l \mathbb{E}^U[(Y_1/Y_0)^{m-l}] - \mathbb{I}((-1)^l = -1) \\
&\qquad \times \left(\mathbb{E}^L[Y_1/Y_0]\right)^l \mathbb{E}^L[(Y_1/Y_0)^{m-l}]\big) \\
&\leq \mathbb{E}[\{(Y_1/Y_0) - (\mathbb{E}[Y_1/Y_0])\}^m] \\
&\leq \sum_{l=0}^{m} \binom{m}{l} \big(\mathbb{I}((-1)^l = 1)\left(\mathbb{E}^L[Y_1/Y_0]\right)^l \mathbb{E}^L[(Y_1/Y_0)^{m-l}] - \mathbb{I}((-1)^l = -1) \\
&\qquad \times \left(\mathbb{E}^U[Y_1/Y_0]\right)^l \mathbb{E}^U[(Y_1/Y_0)^{m-l}]\big).
\end{aligned} \tag{13}$$

The above bounds are not sharp.

Finally, given the bounds for the 2nd, 3rd, and 4th central moments of ICRE, the skewness and kurtosis of ICRE can be bounded as shown in Appendix B.

Comparison with existing bounding results. There is a substantial literature on deriving closed-form bounds for causal quantities under various assumptions (e.g., [Manski, 1997, Dawid et al., 2016, Sachs et al., 2023, Zhang et al., 2025]). Relative to [Kawakami and Tian, 2025], which represents the current state of the art for moments of ICDE, we obtain sharper bounds for these quantities. To the best of our knowledge, our work provides the first set of bounds for moments of ICRE, which have not been directly addressed in the prior literature. Related approaches compute bounds via numerical optimization, including methods based on optimal transport [Gao et al., 2025], linear programming [Hashimoto et al., 2026], and polynomial programming in general causal graphs [Zhang et al., 2022, Zaffalon et al., 2024, Duarte et al., 2024].

6 Estimating Joint Expectations over POs

In this section, we illustrate estimation methods for $\mathbb{E}^U[g(Y_1, Y_0)]$ and $\mathbb{E}^L[g(Y_1, Y_0)]$, which are used to identify and bound the joint expectation over POs. We leverage inverse transform sampling (ITS) [McNeil et al., 2015], a method for generating samples from a target distribution by applying the inverse of its CDF to samples drawn from the uniform distribution on $[0, 1]$.

From the Radon–Nikodym theorem [Billingsley, 2017], Eqs. (7) and (8) can be rewritten as follows:

$$\mathbb{E}^U[g(Y_1, Y_0)] = \int\int g(y_1, y_0) dF^U_{Y_1|Y_0}(y_1, y_0) dF_{Y_0}(y_0), \tag{14}$$

$$\mathbb{E}^L[g(Y_1, Y_0)] = \int\int g(y_1, y_0) dF^L_{Y_1|Y_0}(y_1, y_0) dF_{Y_0}(y_0), \tag{15}$$

where $F^U_{Y_1|Y_0}(y_1, y_0)$ and $F^L_{Y_1|Y_0}(y_1, y_0)$ denote conditional CDFs, and we have used $F^U_{Y_0} = F^L_{Y_0} = F_{Y_0}$. These hold for discrete, continuous, and mixed-type outcomes.

First, we derive the explicit expression of the conditional CDFs $F^U_{Y_1|Y_0}(y_1, y_0)$ and $F^L_{Y_1|Y_0}(y_1, y_0)$ from the joint CDFs given by Eq. (6). We use F' to denote CDFs without the point y_0 as follows:

$$F'_{Y_0}(y_0) \triangleq \mathbb{P}(Y_0 < y_0). \tag{16}$$

We obtain the following lemma:

Lemma 2. *The conditional CDFs of $F^U_{Y_1|Y_0}(y_1, y_0)$ and $F^L_{Y_1|Y_0}(y_1, y_0)$ are given as follows:*

(A). When $F_{Y_0}(y_0) \neq F'_{Y_0}(y_0)$, which means y_0 is a mass point of F_{Y_0}, we have

$$F^U_{Y_1|Y_0}(y_1, y_0) = \max\left\{\frac{\min\{F_{Y_1}(y_1), F_{Y_0}(y_0)\} - F'_{Y_0}(y_0)}{F_{Y_0}(y_0) - F'_{Y_0}(y_0)}, 0\right\}, \tag{17}$$

$$F^L_{Y_1|Y_0}(y_1, y_0) = 1 - \max\left\{\frac{\min\{1 - F_{Y_1}(y_1), F_{Y_0}(y_0)\} - F'_{Y_0}(y_0)}{F_{Y_0}(y_0) - F'_{Y_0}(y_0)}, 0\right\}. \tag{18}$$

(B). When $F_{Y_0}(y_0) = F'_{Y_0}(y_0)$, which means y_0 is not a mass point of F_{Y_0}, we have

$$F^U_{Y_1|Y_0}(y_1, y_0) = \mathbb{I}(y_1 \geq F^{-1}_{Y_1}(F_{Y_0}(y_0))), F^L_{Y_1|Y_0}(y_1, y_0) = \mathbb{I}(y_1 \geq F^{-1}_{Y_1}(1 - F_{Y_0}(y_0))). \tag{19}$$

To apply ITS, we compute the inverse functions of the above conditional CDFs and obtain the following theorem:

Theorem 4. *The inverse of $F^U_{Y_1|Y_0}(y_1, y_0)$ given y_0, for any $u \in (0, 1)$, is*

$$F^{U-1}_{Y_1|Y_0}(u, y_0) = F^{-1}_{Y_1}(u(F_{Y_0}(y_0) - F'_{Y_0}(y_0)) + F'_{Y_0}(y_0)), \tag{20}$$

and the inverse of $F^L_{Y_1|Y_0}(y_1, y_0)$ given y_0 is

$$F^{L-1}_{Y_1|Y_0}(u, y_0) = F^{-1}_{Y_1}(1 - (1 - u)(F_{Y_0}(y_0) - F'_{Y_0}(y_0)) - F'_{Y_0}(y_0)), \tag{21}$$

where $F^{-1}_{Y_1}$ is the generalized inverse of F_{Y_1}.

Although $F^{U-1}_{Y_1|Y_0}(u, y_0)$ and $F^{L-1}_{Y_1|Y_0}(u, y_0)$ are not defined at points $u = 0$ and $u = 1$ in Theorem 4, the uniform random variable $U \sim \text{Unif}[0, 1]$ takes these values with probability zero. Consequently, this does not pose a problem.

When F_{Y_0} is continuous, i.e., $F_{Y_0}(y_0) = F'_{Y_0}(y_0)$, Eq. (20) reduces to $F^{U-1}_{Y_1|Y_0}(u, y_0) = F^{-1}_{Y_1}(F_{Y_0}(y_0))$, which is equal to the right-hand side of Eq. (9). In this case, our identification result under the comonotonicity assumption coincides with the identification result of [Heckman et al., 1997] in Eq. (10); this aligns with Proposition 4. Also, when F_{Y_0} is continuous, Eq. (21) reduces to $F^{L-1}_{Y_1|Y_0}(u, y_0) = F^{-1}_{Y_1}(1 - F_{Y_0}(y_0))$, which is equal to the right-hand side of Eq. (11). This result aligns with Theorem 2.

6.1 Estimation of joint expectations over POs from finite samples

We assume we are given experimental data from $\mathbb{P}(Y_x)$. We denote the samples of Y_1 from the treatment group as $\mathcal{D}^T = \{Y_1^T, \ldots, Y_{N^T}^T\}$ and the samples of Y_0 from the control group as $\mathcal{D}^C = \{Y_1^C, \ldots, Y_{N^C}^C\}$.

To estimate $\mathbb{E}^U[g(Y_1, Y_0)]$ and $\mathbb{E}^L[g(Y_1, Y_0)]$ from Eqs. (14) and (15), we use (i) i.i.d. samples from the distribution F_{Y_0}, which are from $\mathcal{D}^C$, and (ii) for each $y_0 = Y_i^C$, draw samples $\tilde{Y}_i^{U,T}$ and $\tilde{Y}_i^{L,T}$ from the distributions $F^U_{Y_1|Y_0}$ and $F^L_{Y_1|Y_0}$ respectively. The latter samples are generated via ITS as follows: (i) Draw a sample U_i from the uniform distribution Unif$[0, 1]$, and (ii) Compute $\tilde{Y}_i^{U,T}$ and $\tilde{Y}_i^{L,T}$ by substituting $u = U_i$ and $y_0 = Y_i^C$ into Eqs. (20) and (21), respectively. The expressions of the empirical CDFs used in the computation are given in Appendix C. Finally, we use Eqs. (14) and (15) to calculate the empirical expectation of $g(Y_1, Y_0)$ using generated samples $\{\tilde{Y}_i^{U,T}, Y_i^C\}_{i=1}^{N^C}$ and $\{\tilde{Y}_i^{L,T}, Y_i^C\}_{i=1}^{N^C}$ as follows:

$$\hat{\mathbb{E}}^U[g(Y_1, Y_0)] = \frac{1}{N^C}\sum_{i=1}^{N^C} g(\tilde{Y}_i^{U,T}, Y_i^C), \quad \hat{\mathbb{E}}^L[g(Y_1, Y_0)] = \frac{1}{N^C}\sum_{i=1}^{N^C} g(\tilde{Y}_i^{L,T}, Y_i^C). \tag{22}$$

These are consistent estimators as follows:

Theorem 5. *Suppose that g is Lipschitz continuous and that Y_0 and Y_1 have compact supports. Then $\hat{\mathbb{E}}^U[g(Y_1, Y_0)]$ and $\hat{\mathbb{E}}^L[g(Y_1, Y_0)]$ converge in probability to $\mathbb{E}^U[g(Y_1, Y_0)]$ and $\mathbb{E}^L[g(Y_1, Y_0)]$, respectively, at the rate $\mathcal{O}_p\big((N^C)^{-1/2} + (N^T)^{-1/2}\big)$.*

6.2 Estimation from observational data

Our results extend immediately to settings where, instead of experimental data from $\mathbb{P}(Y_x)$, we observe only observational data from $\mathbb{P}(X, Y)$ under the *exogeneity assumption*, $Y_x \perp\!\!\!\perp X$. Under exogeneity, we have $F_{Y_x} = F_{Y|X=x}$ and $\mathbb{P}(Y_x) = \mathbb{P}(Y|x)$. Consequently, the observed samples of Y corresponding to $X = 1$ and $X = 0$ can be treated as samples from $\mathbb{P}(Y_1)$ and $\mathbb{P}(Y_0)$, respectively.

Because our identification and bounding results are expressed entirely in terms of $F_{Y_x}(y)$, they extend naturally to settings with observed covariates W under the *conditional exogeneity* assumption, $Y_x \perp\!\!\!\perp X|W$. We further assume positivity alongside conditional exogeneity: for each treatment level $x \in \{0, 1\}$, $\mathbb{P}(X = x \mid W = w) > 0$ for almost every w in the support of W. In finite samples, inadequate overlap may lead to unstable estimation. Under these assumptions, each instance of $F_{Y_x}(y)$ can be replaced by the standard covariate adjustment $\int F_{Y|X=x,W=w}(y)dF_W(w)$. Estimation of the resulting expressions can proceed via plug-in methods. However, developing effective estimators when W is high-dimensional remains nontrivial.

7 Numerical Experiments and Applications to Real-World Datasets

The code is available at https://github.com/NA679723/NeurIPS2026_Joint_Expectation.git.

Numerical experiments. We conduct numerical experiments to evaluate the finite-sample properties of our estimators. The results indicate that the estimators perform reasonably well at a sample size of $N = 100$ and are stable at $N = 1000$. For even-order moments, our bounds are tighter than those obtained from the baseline Monte Carlo integration (MCI) of Kawakami and Tian [2025]. Moreover, our estimator based on closed-form expressions is significantly more computationally efficient than the numerical optimization-based causal optimal transport (COT) estimator by Gao et al. [2025]. The details are provided in Appendix D.

Applications. We apply our methods to two real-world datasets "Cholesterol Reduction" [Westfall et al., 2011] and the Lalonde dataset [LaLonde, 1986]. We estimate the variance, skewness, and kurtosis of ICDE and ICRE. In both cases, the results reveal substantial heterogeneity in treatment effects, with high tail risk. These findings suggest caution in making treatment decisions based solely on the mean ICDE. Complete results are reported in Appendix E.

8 Conclusion

We study the identification and bounding of the joint expectation $\mathbb{E}[g(Y_1, Y_0)]$ over POs. This quantity encompasses a range of important statistical characteristics of the ICDE and the ICRE. We establish new identification results under the comonotonicity and countercomonotonicity assumptions (Theorems 1, 2). Notably, comonotonicity is a weaker identifying assumption than existing assumptions in the literature (Propositions 1-5). In the absence of these assumptions, we present novel sharp bounds for two classes of g functions (Theorem 3). These results lead to tighter bounds for moments of ICDE (Corollaries 1-3) and new bounds for moments of ICRE (Corollaries 4-5). Importantly, these results apply to any type of outcome variable Y—discrete, continuous, or mixed. Overall, our results provide new tools for investigating causal effect heterogeneity and enable analyses that were previously inaccessible. These results have important implications for real-world decision-making: reliance on average causal effects alone may be misleading, whereas information on causal effect heterogeneity provides a more comprehensive and reliable basis for informed decisions by practitioners.

The comonotonicity and countercomonotonicity assumptions are generally not testable from observed data. While we discuss their practical plausibility (following Assumptions 1 and 2), their credibility must ultimately be assessed based on domain knowledge and the scientific context. When these assumptions are not credible, bounding approaches offer a principled alternative. Although we obtain sharp bounds for certain moments of ICDE and ICRE, our general results are restricted to specific classes of g functions. Extending sharp bounds to more general function classes is left for future work.

Although our identification and bounding results extend to settings with observed covariates (Section 6.2), developing effective and statistically efficient estimators in the presence of *high-dimensional* covariates remains nontrivial and would require substantial additional methodological advances. We therefore leave the development of robust estimation procedures for high-dimensional covariate settings to future work. Another promising direction is to extend the analysis to multivariate outcomes, as considered in Torous et al. [2024].

Acknowledgments

The authors thank the anonymous reviewers for their time and thoughtful comments. This work received no external funding. The authors declare no competing interests.

References

Susan Athey and Guido W Imbens. Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497, 2006.

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.

Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2017.

Stamatis Cambanis, Gordon Simons, and William Stout. Inequalities for e k (x, y) when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 36(4):285–294, 1976.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

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.

Yanqin Fan and Sang Soo Park. Sharp bounds on the distribution of the treatment effect and their statistical inference. *Econometric Theory*, 26(3):931–951, 2010.

Michael P Fay, Erica H Brittain, Joanna H Shih, Dean A Follmann, and Erin E Gabriel. Causal estimands and confidence intervals associated with wilcoxon-mann-whitney tests in randomized experiments. *Stat Med*, 37(20):2923–2937, Sep 2018.

Erin E Gabriel, Michael C Sachs, and Andreas Kryger Jensen. Sharp symbolic nonparametric bounds for measures of benefit in observational and imperfect randomized studies with ordinal outcomes. *Biometrika*, 111(4):1429–1436, 04 2024.

Gary L. Gadbury, Hari K. Iyer, and Jeffrey M. Albert. Individual treatment effects in randomized trials with binary outcomes. *Journal of Statistical Planning and Inference*, 121(2):163–174, 2004.

Zijun Gao, Shu Ge, and Jian Qian. Bridging multiple worlds: Multi-marginal optimal transport for causal partial-identification problem. In Yingzhen Li, Stephan Mandt, Shipra Agrawal, and Emtiyaz Khan, editors, *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 721–729. PMLR, 03–05 May 2025.

Sander Greenland, Judea Pearl, and James M. Robins. Confounding and Collapsibility in Causal Inference. *Statistical Science*, 14(1):29 – 46, 1999.

Naoya Hashimoto, Yuta Kawakami, and Jin Tian. Bounds and identification of joint probabilities of potential outcomes and observed variables under monotonicity assumptions. In *The 29th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2026.

James J. Heckman, Jeffrey Smith, and Nancy Clements. Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, 64(4):487–535, 10 1997.

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

Miguel A Hernán and James M Robins. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*, 60(7):578–586, Jul 2006.

Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81 (396):945–960, 1986.

Emily J. Huang, Ethan X. Fang, Daniel F. Hanley, and Michael Rosenblum. Inequality in treatment benefits: Can we determine if a new treatment benefits the many or the few? *Biostatistics*, 18(2): 308–324, 11 2016.

Yuta Kawakami and Jin Tian. Moments of causal effects. *Proceedings of the 41st Conference on Uncertainty in Artificial Intelligence (UAI-2025)*, 2025. URL https://arxiv.org/abs/2505.04971.

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.

M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

Robert J LaLonde. Evaluating the econometric evaluations of training programs with experimental data. *The American economic review*, pages 604–620, 1986.

Chaochao Lu, Biwei Huang, Ke Wang, José Miguel Hernández-Lobato, Kun Zhang, and Bernhard Schölkopf. Sample-efficient reinforcement learning via counterfactual-based data augmentation. *arXiv preprint arXiv:2012.09092*, 2020.

Jiannan Lu, Yunshu Zhang, and Peng Ding. Sharp bounds on the relative treatment effect for ordinal outcomes. *Biometrics*, 76(2):664–669, 11 2019.

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

Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

Arash Nasr-Esfahany, Mohammad Alizadeh, and Devavrat Shah. Counterfactual identifiability of bijective causal models. In *International Conference on Machine Learning*, pages 25733–25754. PMLR, 2023.

Roger B Nelsen. *An introduction to copulas*. Springer, 2006.

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.

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

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.

André H Tchen. Inequalities for distributions with given marginals. *The Annals of Probability*, pages 814–827, 1980.

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

William Torous, Florian Gunsilius, and Philippe Rigollet. An optimal transport approach to estimating causal effects via nonlinear difference-in-differences. *Journal of Causal Inference*, 12(1):20230004, 2024.

Peter H Westfall, Randall D Tobias, and Russell D Wolfinger. *Multiple comparisons and multiple tests using SAS*. SAS Institute, 2011.

Peng Wu, Haoxuan Li, Chunyuan Zheng, Yan Zeng, Jiawei Chen, Yang Liu, Ruocheng Guo, and Kun Zhang. Learning counterfactual outcomes under rank preservation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Shaoan Xie, Biwei Huang, Bin Gu, Tongliang Liu, and Kun Zhang. Advancing counterfactual inference through nonlinear quantile regression. *arXiv preprint arXiv:2306.05751*, 2023.

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

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.

Appendix to "Identification and Bounding of Joint Expectation over Potential Outcomes"

A Proof

In this appendix, we provide proofs of lemmas, theorems, and propositions in the body of the paper.

Lemma 1. *Given F_{Y_0} and F_{Y_1}, the joint CDF of POs $F_{(Y_1,Y_0)}(y_1,y_0)$ is bounded by $F^L_{(Y_1,Y_0)}(y_1,y_0) \leq F_{(Y_1,Y_0)}(y_1,y_0) \leq F^U_{(Y_1,Y_0)}(y_1,y_0)$, where*

$$F^U_{(Y_1,Y_0)}(y_1,y_0) = \min\{F_{Y_0}(y_0), F_{Y_1}(y_1)\}, F^L_{(Y_1,Y_0)}(y_1,y_0) = \max\{F_{Y_0}(y_0) + F_{Y_1}(y_1) - 1, 0\}. \tag{6}$$

The upper bound is attained when (Y_1,Y_0) are comonotonic, while the lower bound is attained when (Y_1,Y_0) are countercomonotonic. These bounds are sharp given F_{Y_0} and F_{Y_1}.

Proof. The upper bound is attained when (Y_1,Y_0) are comonotonic, while the lower bound is attained when (Y_1,Y_0) are countercomonotonic [Nelsen, 2006]. We have

$$F^L_{(Y_1,Y_0)}(y_1,y_0) = \max\{F_{Y_0}(y_0) + F_{Y_1}(y_1) - 1, 0\}, \tag{23}$$

$$F^U_{(Y_1,Y_0)}(y_1,y_0) = \min\{F_{Y_0}(y_0), F_{Y_1}(y_1)\}. \tag{24}$$

Then, Lemma 1 holds. □

Theorem 1. *Under SCM $\mathcal{M}$, given F_{Y_0} and F_{Y_1}, we have the following two identification statements:*

(A). Under Assumption 1, $\mathbb{E}[g(Y_1,Y_0)] = \mathbb{E}^U[g(Y_1,Y_0)]$.

(B). Under Assumption 2, $\mathbb{E}[g(Y_1,Y_0)] = \mathbb{E}^L[g(Y_1,Y_0)]$.

Proof. Under SCM $\mathcal{M}$ and Assumption 1, we have $F_{(Y_1,Y_0)}(y_1,y_0) = F^U_{(Y_1,Y_0)}(y_1,y_0)$ since Assumption 1 implies $F_{(Y_1,Y_0)}(y_1,y_0)$ attaining the upper bound of Fréchet–Hoeffding bounds. Also, under SCM $\mathcal{M}$ and Assumption 2, we have $F_{(Y_1,Y_0)}(y_1,y_0) = F^L_{(Y_1,Y_0)}(y_1,y_0)$ since Assumption 2 implies $F_{(Y_1,Y_0)}(y_1,y_0)$ attaining the lower bound of Fréchet–Hoeffding bounds. Since $F^U_{(Y_1,Y_0)}(y_1,y_0)$ and $F^L_{(Y_1,Y_0)}(y_1,y_0)$ are identified through Lemma 1, we have Theorem 1. □

Proposition 1. *Assumption 3 implies Assumption 1; however, the converse implication does not hold.*

Proof. This is a corollary of Proposition 3 since Assumption 3 implies Assumption 5. The discrete or mixed outcome is a counterexample to the converse implication. □

Proposition 2. *When Y is binary, Assumption 1 is equivalent to the condition that either $\mathbb{P}(Y_0 = 1, Y_1 = 0) = 0$ or $\mathbb{P}(Y_0 = 0, Y_1 = 1) = 0$ holds, which is weaker than Assumption 4.*

Proof. **(Assumption 1** $\Rightarrow \mathbb{P}(Y_0 = 1, Y_1 = 0) = 0$ **or** $\mathbb{P}(Y_0 = 0, Y_1 = 1) = 0$**)** Assumption 1 implies $\mathbb{P}(Y_0 = 0, Y_1 = 1) = \max\{\mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1), 0\}$ and $\mathbb{P}(Y_0 = 1, Y_1 = 0) = \max\{\mathbb{P}(Y_0 = 1) - \mathbb{P}(Y_1 = 1), 0\}$. If $\mathbb{P}(Y_1 = 1) \geq \mathbb{P}(Y_0 = 1)$, then $\mathbb{P}(Y_0 = 1, Y_1 = 0) = 0$. If $\mathbb{P}(Y_1 = 1) < \mathbb{P}(Y_0 = 1)$, then $\mathbb{P}(Y_0 = 0, Y_1 = 1) = 0$. Thus, the condition $\mathbb{P}(Y_0 = 1, Y_1 = 0) = 0$ or $\mathbb{P}(Y_0 = 0, Y_1 = 1) = 0$ holds.

($\mathbb{P}(Y_0 = 1, Y_1 = 0) = 0$ **or** $\mathbb{P}(Y_0 = 0, Y_1 = 1) = 0 \Rightarrow$ **Assumption 1)** When $\mathbb{P}(Y_0 = 1, Y_1 = 0) = 0$, we consider the random variable U', which takes three values depending on POs as: $1 = (Y_0 = 0, Y_1 = 0), 2 = (Y_0 = 0, Y_1 = 1), 3 = (Y_0 = 1, Y_1 = 1)$. Then, Y_1 follows $\mathbb{I}(U' \geq 2)$, and Y_0 follows $\mathbb{I}(U' \geq 3)$ and they are monotonic increasing functions on U'. Since the random variable U' can be represented by a uniform distribution, then Assumption 1 holds. When $\mathbb{P}(Y_0 = 0, Y_1 = 1) = 0$, we consider the random variable U', which takes three values depending on POs as: $1 = (Y_0 = 0, Y_1 = 0), 2 = (Y_0 = 1, Y_1 = 0), 3 = (Y_0 = 1, Y_1 = 1)$. Then, Y_0 follows $\mathbb{I}(U' \geq 2)$, and Y_1 follows $\mathbb{I}(U' \geq 3)$ and they are monotonic increasing functions on U'. Since the random variable U' can be represented by a uniform distribution, then Assumption 1 holds. □

Proposition 3. *Assumption 5 implies Assumption 1; however, the converse need not hold for a given structural representation $Y = f_Y(X, U_Y)$. Nevertheless, if (Y_1, Y_0) are comonotonic, then there exists a latent variable U'_Y and a structural function $f'_Y(X, U'_Y)$ satisfying Assumption 5 such that $Y_x = f'_Y(x, U'_Y)$ for $x \in \{0, 1\}$.*

Proof. When f_Y is monotonic increasing on U_Y for all $x \in \{0, 1\}$ almost surely w.r.t. $\mathbb{P}_{U_Y}$, Assumption 1 holds since any random variable can be represented by a uniform distribution, i.e., $U_Y = F^{-1}_{U_Y}(U)$, and $F^{-1}_{U_Y}$ is monotonic increasing. When f_Y is monotonic decreasing on U_Y for all $x \in \{0, 1\}$ almost surely w.r.t. $\mathbb{P}_{U_Y}$, Assumption 1 also holds by considering $-U_Y$ in place of U_Y.

The converse, however, need not hold for a given structural representation. For example, let $U_Y \sim \mathrm{Unif}(0, 1)$ and $f_Y(x, U_Y) = |U_Y - 1/2|$ for both $x \in \{0, 1\}$. Then $Y_1 = Y_0$ almost surely, so (Y_1, Y_0) is trivially comonotonic and Assumption 1 holds. However, $u \mapsto |u - 1/2|$ is decreasing on $(0, 1/2)$ and increasing on $(1/2, 1)$, so $f_Y(x, \cdot)$ is not monotonic on any set of full P_{U_Y}-measure, and Assumption 5 does not hold.

Nevertheless, if (Y_1, Y_0) are comonotonic, there exists $U'_Y \sim \mathrm{Unif}(0, 1)$ such that $Y_x = F^{-1}_{Y_x}(U'_Y)$ for $x \in \{0, 1\}$. Setting $f'_Y(x, u) = F^{-1}_{Y_x}(u)$, which is non-decreasing in u, the structural function $f'_Y(X, U'_Y)$ satisfies Assumption 5 and $Y_x = f'_Y(x, U'_Y)$ for $x \in \{0, 1\}$. □

Proposition 4. *For continuous outcome Y, Assumption 1 is equivalent to Assumption 6.*

Proof. (Assumption 6 ⇒ Assumption 1) Assumption 6 represents $U = F_{Y_1}(Y_1) = F_{Y_0}(Y_0)$, where $U \sim \mathrm{Unif}[0, 1]$. Since F_{Y_1} and F_{Y_0} are both invertible and monotonic increasing, $Y_1 = F^{-1}_{Y_1}(U)$ and $Y_0 = F^{-1}_{Y_0}(U)$. Then, Assumption 1 holds.
(Assumption 1 ⇒ Assumption 6) Under Assumption 1, we have $Y_1 = F^{-1}_{Y_1}(U)$ and $Y_0 = F^{-1}_{Y_0}(U)$. Then, we have $F_{Y_1}(Y_1) = U = F_{Y_0}(Y_0)$ and Assumption 6. □

Proposition 5. *Assumption 7 implies Assumption 1; however, the converse implication does not hold.*

Proof. If $\tilde{\rho}(Y_1, Y_0) = 1$, every i.i.d. pair $(Y_1, Y_0), (Y'_1, Y'_0)$ is concordant and none is discordant: $(Y_1 - Y'_1)(Y_0 - Y'_0) \geq 0$ almost surely. Since every pair of independent draws is concordant, the rank variables $U = F_{Y_1}(Y_1)$ and $V = F_{Y_0}(Y_0)$ satisfy $U = V$ almost surely. Then, we have $Y_0 = F^{-1}_{Y_0}(U)$ and $Y_1 = F^{-1}_{Y_1}(U)$, realizing both variables as non-decreasing transforms of the same latent U. Thus, (Y_1, Y_0) are comonotonic.

We present a counterexample showing that Assumption 1 does not imply Assumption 7:

$$\textbf{Counterexample: } (Y_1, Y_0) = \{(1, 1), (1, 0), (1, 0), (0, 0)\}. \tag{25}$$

In this case, $\tilde{\rho}_4 = 1/3$ and is not 1. This setting does not satisfy Assumption 7. However, our comonotonicity assumption holds. □

Theorem 2. *Under SCM $\mathcal{M}$ and Assumption 2, given F_{Y_0} and F_{Y_1}, when Y is continuous, for any subjects whose POs are $(Y_1, Y_0) = (y_1, y_0)$, y_1 is identifiable by*

$$y_1 = F^{-1}_{Y_1}(1 - F_{Y_0}(y_0)). \tag{11}$$

Proof. Under SCM $\mathcal{M}$ and Assumption 2, when Y is continuous, there exists a random variable $U \sim \mathrm{Unif}(0, 1)$ such that: $Y_0 = F^{-1}_{Y_0}(U)$ and $Y_1 = F^{-1}_{Y_1}(1-U)$. Then, we have $Y_1 = F^{-1}_{Y_1}(1-F_{Y_0}(Y_0))$. Then, given $Y = y_0$ and $X = 0$ (⇒ $Y_0 = y_0$), the value of y_1 is identifiable by

$$y_1 = F^{-1}_{Y_1}(1 - F_{Y_0}(y_0)). \tag{26}$$

Then, the joint expectations over POs are identified by

$$\mathbb{E}[g(Y_1, Y_0)] = \mathbb{E}[g(F^{-1}_{Y_1}(1 - F_{Y_0}(Y_0)), Y_0)]. \tag{27}$$

□

Theorem 3. *Under SCM $\mathcal{M}$, given F_{Y_0} and F_{Y_1}, we have the following two statements:*

(A). Under Assumption 8, $\mathbb{E}[g(Y_1, Y_0)]$ is bounded by $\mathbb{E}^L[g(Y_1, Y_0)] \leq \mathbb{E}[g(Y_1, Y_0)] \leq \mathbb{E}^U[g(Y_1, Y_0)]$.

(B). Under Assumption 9, $\mathbb{E}[g(Y_1, Y_0)]$ is bounded by $\mathbb{E}^U[g(Y_1, Y_0)] \leq \mathbb{E}[g(Y_1, Y_0)] \leq \mathbb{E}^L[g(Y_1, Y_0)]$.

These bounds are sharp given the marginals F_{Y_0} and F_{Y_1}.

Proof. The result is closely related to classical results on extremal couplings [Cambanis et al., 1976, Tchen, 1980]; we include a proof for completeness. Let F be any bivariate distribution that has the marginals, F_{Y_1} and F_{Y_0}. Let a function g that is supermodular and bounded, and $\Omega_Y = (y^{min}, y^{max})$, where $y^{min}, y^{max} \in [-\infty, \infty]$. We approximate the function g as $\tilde{g}(y_1, y_0) = g(y^{min}, y^{min}) + \sum_{i=1}^{I}(g(a_i, y^{min}) - g(a_{i-1}, y^{min}))\mathbb{I}(y_1 > a_i) + \sum_{j=1}^{J}(g(y^{min}, b_j) - g(y^{min}, b_{j-1}))\mathbb{I}(y_0 > b_j) + \sum_{i=1}^{I}\sum_{j=1}^{J}(g(a_i, b_j) - g(a_{i-1}, b_j) - g(a_i, b_{j-1}) + g(a_{i-1}, b_{j-1}))\mathbb{I}(y_1 > a_i, y_0 > b_j)$, which converges to g as $I, J \to \infty$. By taking expectation, $\mathbb{E}[\tilde{g}(Y_1, Y_0)] = g(y^{min}, y^{min}) + \sum_{i=1}^{I}(g(a_i, y^{min}) - g(a_{i-1}, y^{min}))\mathbb{P}(Y_1 > a_i) + \sum_{j=1}^{J}(g(y^{min}, b_j) - g(y^{min}, b_{j-1}))\mathbb{P}(Y_0 > b_j) + \sum_{i=1}^{I}\sum_{j=1}^{J}(g(a_i, b_j) - g(a_{i-1}, b_j) - g(a_i, b_{j-1}) + g(a_{i-1}, b_{j-1}))\mathbb{P}(Y_1 > a_i, Y_0 > b_j)$. Since the coefficient of $\mathbb{P}(Y_1 > a_i, Y_0 > b_j)$, i.e., $g(a_i, b_j) - g(a_{i-1}, b_j) - g(a_i, b_{j-1}) + g(a_{i-1}, b_{j-1})$, is nonnegative due to the supermodularity of g even for approximation, the upper bound of $\mathbb{E}[\tilde{g}(Y_1, Y_0)]$ is achieved when Y_1 and Y_0 are comonotonic, and the lower bound is achieved when Y_1 and Y_0 are countercomonotonic. Then, taking limits $I, J \to \infty$, we have $\mathbb{E}^L[g(Y_1, Y_0)] \leq \mathbb{E}[g(Y_1, Y_0)] \leq \mathbb{E}^U[g(Y_1, Y_0)]$. Similarly, when g is submodular, the coefficient of $\mathbb{P}(Y_1 > a_i, Y_0 > b_j)$, i.e., $g(a_i, b_j) - g(a_{i-1}, b_j) - g(a_i, b_{j-1}) + g(a_{i-1}, b_{j-1})$, is nonpositive for all i, j. Then, we have $\mathbb{E}^U[g(Y_1, Y_0)] \leq \mathbb{E}[g(Y_1, Y_0)] \leq \mathbb{E}^L[g(Y_1, Y_0)]$. □

Lemma 2. *The conditional CDFs of $F^U_{Y_1|Y_0}(y_1, y_0)$ and $F^L_{Y_1|Y_0}(y_1, y_0)$ are given as follows:*

(A). When $F_{Y_0}(y_0) \neq F'_{Y_0}(y_0)$, which means y_0 is a mass point of F_{Y_0}, we have

$$F^U_{Y_1|Y_0}(y_1, y_0) = \max\left\{\frac{\min\{F_{Y_1}(y_1), F_{Y_0}(y_0)\} - F'_{Y_0}(y_0)}{F_{Y_0}(y_0) - F'_{Y_0}(y_0)}, 0\right\}, \tag{17}$$

$$F^L_{Y_1|Y_0}(y_1, y_0) = 1 - \max\left\{\frac{\min\{1 - F_{Y_1}(y_1), F_{Y_0}(y_0)\} - F'_{Y_0}(y_0)}{F_{Y_0}(y_0) - F'_{Y_0}(y_0)}, 0\right\}. \tag{18}$$

(B). When $F_{Y_0}(y_0) = F'_{Y_0}(y_0)$, which means y_0 is not a mass point of F_{Y_0}, we have

$$F^U_{Y_1|Y_0}(y_1, y_0) = \mathbb{I}(y_1 \geq F^{-1}_{Y_1}(F_{Y_0}(y_0))), F^L_{Y_1|Y_0}(y_1, y_0) = \mathbb{I}(y_1 \geq F^{-1}_{Y_1}(1 - F_{Y_0}(y_0))). \tag{19}$$

Proof. We consider two random variables (A, B) whose joint distribution follows $F^U_{Y_1,Y_0}$. There exists a random variable $U \sim \text{Unif}(0, 1)$ and non-decreasing functions f_1 and f_0 such that $A = f_1(U)$ and $B = f_0(U)$, where the marginal distributions of A and B are equal to the distribution of Y given $X = 1$ and $X = 0$, respectively. We denote $u_{y_0} = \sup\{u \in [0, 1]; f_0(u) \leq y_0\}$, $u_{y_1} = \sup\{u \in [0, 1]; f_1(u) \leq y_1\}$, and $u'_{y_0} = \sup\{u \in [0, 1]; f_0(u) < y_0\}$. When $F_{Y_0}(y_0) \neq F'_{Y_0}(y_0)$, we have

$$\mathbb{P}(A \leq y_1 | B = y_0) = \frac{\mathbb{P}(A \leq y_1, B = y_0)}{\mathbb{P}(B = y_0)} = \frac{\mathbb{P}(U \leq u_{y_1}, u'_{y_0} < U \leq u_{y_0})}{\mathbb{P}(u'_{y_0} < U \leq u_{y_0})} \tag{28}$$

$$= \max\left\{\frac{\min\{\mathbb{P}(U \leq u_{y_1}), \mathbb{P}(U \leq u_{y_0})\} - \mathbb{P}(U \leq u'_{y_0})}{\mathbb{P}(U \leq u_{y_0}) - \mathbb{P}(U \leq u'_{y_0})}, 0\right\} \tag{29}$$

$$= \max\left\{\frac{\min\{F_{Y_1}(y_1), F_{Y_0}(y_0)\} - F'_{Y_0}(y_0)}{F_{Y_0}(y_0) - F'_{Y_0}(y_0)}, 0\right\}. \tag{30}$$

When $F_{Y_0}(y_0) = F'_{Y_0}(y_0)$, we have

$$\mathbb{P}(A \leq y_1 | B = y_0) = \mathbb{P}(U \leq u_{y_1} | U = u_{y_0}) \tag{31}$$

$$= \mathbb{I}(u_{y_0} \leq u_{y_1}) \tag{32}$$
$$= \mathbb{I}(F_{Y_1}(y_1) \geq F_{Y_0}(y_0)) \tag{33}$$
$$= \mathbb{I}(y_1 \geq F_{Y_1}^{-1}(F_{Y_0}(y_0))). \tag{34}$$

We consider two random variables (A, B) whose joint distribution follows $F^L_{Y_1,Y_0}$. There exists a random variable $U \sim \text{Unif}(0,1)$ and a non-increasing function f_0 and a non-decreasing function f_1 (or a non-decreasing function f_0 and a non-increasing function f_1) such that $A = f_1(U)$ and $B = f_0(U)$, where the marginal distributions of A and B are equal to the distribution of Y given $X = 1$ and $X = 0$, respectively. We denote $u_{y_0} = \sup\{u \in [0,1]; f_0(u) \leq y_0\}$, $u_{y_1} = \inf\{u \in [0,1]; f_1(1-u) \leq y_1\}$, and $u'_{y_0} = \sup\{u \in [0,1]; f_0(u) < y_0\}$. When $F_{Y_0}(y_0) \neq F'_{Y_0}(y_0)$, we have

$$\mathbb{P}(A \leq y_1 | B = y_0) = 1 - \mathbb{P}(A > y_1 | B = y_0) \tag{35}$$
$$= 1 - \frac{\mathbb{P}(A > y_1, B = y_0)}{\mathbb{P}(B = y_0)} \tag{36}$$
$$= 1 - \frac{\mathbb{P}(U \leq u_{y_1}, u'_{y_0} < U \leq u_{y_0})}{\mathbb{P}(u'_{y_0} < U \leq u_{y_0})} \tag{37}$$
$$= 1 - \max\left\{\frac{\min\{\mathbb{P}(U \leq u_{y_1}), \mathbb{P}(U \leq u_{y_0})\} - \mathbb{P}(U \leq u'_{y_0})}{\mathbb{P}(U \leq u_{y_0}) - \mathbb{P}(U \leq u'_{y_0})}, 0\right\} \tag{38}$$
$$= 1 - \max\left\{\frac{\min\{1 - F_{Y_1}(y_1), F_{Y_0}(y_0)\} - F'_{Y_0}(y_0)}{F_{Y_0}(y_0) - F'_{Y_0}(y_0)}, 0\right\}. \tag{39}$$

When $F_{Y_0}(y_0) = F'_{Y_0}(y_0)$, we have

$$\mathbb{P}(A \leq y_1 | B = y_0) = \mathbb{P}(1 - U \leq u_{y_1} | U = u_{y_0}) \tag{40}$$
$$= \mathbb{P}(u_{y_1} \geq 1 - u_{y_0}) \tag{41}$$
$$= \mathbb{I}(F_{Y_1}(y_1) \geq 1 - F_{Y_0}(y_0)) \tag{42}$$
$$= \mathbb{I}(y_1 \geq F_{Y_1}^{-1}(1 - F_{Y_0}(y_0))). \tag{43}$$

□

Theorem 4. *The inverse of $F^U_{Y_1|Y_0}(y_1, y_0)$ given y_0, for any $u \in (0,1)$, is*

$$F^{U-1}_{Y_1|Y_0}(u, y_0) = F^{-1}_{Y_1}(u(F_{Y_0}(y_0) - F'_{Y_0}(y_0)) + F'_{Y_0}(y_0)), \tag{20}$$

and the inverse of $F^L_{Y_1|Y_0}(y_1, y_0)$ given y_0 is

$$F^{L-1}_{Y_1|Y_0}(u, y_0) = F^{-1}_{Y_1}(1 - (1-u)(F_{Y_0}(y_0) - F'_{Y_0}(y_0)) - F'_{Y_0}(y_0)), \tag{21}$$

where $F^{-1}_{Y_1}$ is the generalized inverse of F_{Y_1}.

Proof. **(A). When $F_{Y_0}(y_0) \neq F'_{Y_0}(y_0)$.** If $\frac{\min\{F_{Y_1}(y_1), F_{Y_0}(y_0)\} - F'_{Y_0}(y_0)}{F_{Y_0}(y_0) - F'_{Y_0}(y_0)} \leq 0$, then the right-hand side of Eq. (17) is equal to 0. If $F_{Y_1}(y_1) \geq F_{Y_0}(y_0)$, then the right-hand side of Eq. (17) is equal to 1. Thus, for any $u \in (0,1)$, we have

$$F^{U-1}_{Y_1|Y_0}(u, y_0) = \inf\left\{y_1; \frac{F_{Y_1}(y_1) - F'_{Y_0}(y_0)}{F_{Y_0}(y_0) - F'_{Y_0}(y_0)} \geq u\right\} \tag{44}$$
$$= \inf\{y_1; F_{Y_1}(y_1) \geq u(F_{Y_0}(y_0) - F'_{Y_0}(y_0)) + F'_{Y_0}(y_0)\} \tag{45}$$
$$= F^{-1}_{Y_1}(u(F_{Y_0}(y_0) - F'_{Y_0}(y_0)) + F'_{Y_0}(y_0)). \tag{46}$$

If $\frac{\min\{1 - F_{Y_1}(y_1), F_{Y_0}(y_0)\} - F'_{Y_0}(y_0)}{F_{Y_0}(y_0) - F'_{Y_0}(y_0)} \leq 0$, then the right-hand side of Eq. (18) is equal to 1. If $1 - F_{Y_1}(y_1) \geq F_{Y_0}(y_0)$, then the right-hand side of Eq. (18) is equal to 0. Thus, for any $u \in (0,1)$, we have

$$F^{L-1}_{Y_1|Y_0}(u, y_0) = \inf\left\{y_1; 1 - \frac{1 - F_{Y_1}(y_1) - F'_{Y_0}(y_0)}{F_{Y_0}(y_0) - F'_{Y_0}(y_0)} \geq u\right\} \tag{47}$$

$$= \inf\{y_1; F_{Y_1}(y_1) \geq 1 - (1-u)(F_{Y_0}(y_0) - F'_{Y_0}(y_0)) - F'_{Y_0}(y_0)\} \tag{48}$$

$$= F_{Y_1}^{-1}(1 - (1-u)(F_{Y_0}(y_0) - F'_{Y_0}(y_0)) - F'_{Y_0}(y_0)) \tag{49}$$

(B). When $F_{Y_0}(y_0) = F'_{Y_0}(y_0)$. From Lemma 2, we have $F^{U-1}_{Y_1|Y_0}(u, y_0) = F^{-1}_{Y_1}(F_{Y_0}(y_0))$. and $F^{L-1}_{Y_1|Y_0}(u, y_0) = F^{-1}_{Y_1}(1 - F_{Y_0}(y_0))$. □

Theorem 5. *Suppose that g is Lipschitz continuous and that Y_0 and Y_1 have compact supports. Then $\hat{\mathbb{E}}^U[g(Y_1, Y_0)]$ and $\hat{\mathbb{E}}^L[g(Y_1, Y_0)]$ converge in probability to $\mathbb{E}^U[g(Y_1, Y_0)]$ and $\mathbb{E}^L[g(Y_1, Y_0)]$, respectively, at the rate $\mathcal{O}_p(N^{C^{-1/2}} + N^{T^{-1/2}})$.*

Proof. We prove the claim for $\hat{\mathbb{E}}^U$; the proof for $\hat{\mathbb{E}}^L$ is identical. Under comonotonicity, $\mathbb{E}^U[g(Y_1, Y_0)] = \int_0^1 g(F^{-1}_{Y_1}(u), F^{-1}_{Y_0}(u))\, du$. The error of $\hat{\mathbb{E}}^U$ splits into two parts.

(i) Sampling error from U_i. Conditional on the data, each generated pair $(\tilde{Y}_i^{U,T}, Y_i^C)$ equals $(\hat{F}^{-1}_{Y_1}(V_i), \hat{F}^{-1}_{Y_0}(V_i))$ for a uniform draw V_i within the rank interval of Y_i^C. Hence the conditional mean of $\hat{\mathbb{E}}^U$ is $\int_0^1 g(\hat{F}^{-1}_{Y_1}(u), \hat{F}^{-1}_{Y_0}(u))\, du$. Since g is bounded, its conditional variance is at most $\|g\|^2_\infty / N^C$, so this error is $\mathcal{O}_p(N^{C^{-1/2}})$.

(ii) Error from estimating the marginals. Since g is Lipschitz, the remaining error is bounded by a constant times $\sum_{x\in\{0,1\}} \int_0^1 |\hat{F}^{-1}_{Y_x}(u) - F^{-1}_{Y_x}(u)|\, du$. Because Y_x has compact support $[a_x, b_x]$, each integral is at most $(b_x - a_x) \sup_y |\hat{F}_{Y_x}(y) - F_{Y_x}(y)|$. This sup-norm error is $\mathcal{O}_p(N^{T^{-1/2}})$ for $x = 1$ and $\mathcal{O}_p(N^{C^{-1/2}})$ for $x = 0$ by the Dvoretzky–Kiefer–Wolfowitz inequality.

Combining (i) and (ii) gives the rate $\mathcal{O}_p(N^{C^{-1/2}} + N^{T^{-1/2}})$. □

B Additional Bounding Corollaries

The m-th central moments of ICDE and ICRE are given as follows:

$$\mathbb{E}[\{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\}^m] \tag{50}$$

$$= \sum_{l=0}^{m} \binom{m}{l} (-1)^l \mathbb{E}[(Y_1 - Y_0)^{m-l}] \left(\mathbb{E}[Y_1 - Y_0]\right)^l \tag{51}$$

$$= \sum_{l=0}^{m} \sum_{k=0}^{m-l} \binom{m}{l} \binom{m-l}{k} (-1)^{l+k} \mathbb{E}[Y_1^{m-l-k} Y_0^k] \left(\mathbb{E}[Y_1 - Y_0]\right)^l, \tag{52}$$

$$\mathbb{E}[\{(Y_1/Y_0) - (\mathbb{E}[Y_1/Y_0])\}^m] = \sum_{l=0}^{m} \binom{m}{l} (-1)^l \mathbb{E}[(Y_1/Y_0)^{m-l}] \left(\mathbb{E}[Y_1/Y_0]\right)^l. \tag{53}$$

Then, we have the following corollaries:

Corollary 6 (The odd-order central moments of ICDE)**.** *Under SCM $\mathcal{M}$, if Y_1 and Y_0 take non-negative values, the odd-order central moments of the ICDE are bounded by*

$$\begin{aligned}
&\sum_{l=0}^{m} \sum_{k=0}^{m-l} \binom{m}{l} \binom{m-l}{k} \Big(\mathbb{I}((-1)^{l+k} \left(\mathbb{E}[Y_1 - Y_0]\right)^l \leq 0)(-1)^{l+k} \left(\mathbb{E}[Y_1 - Y_0]\right)^l \mathbb{E}^U[Y_1^{m-l-k} Y_0^k] \\
&+ \mathbb{I}((-1)^{l+k} \left(\mathbb{E}[Y_1 - Y_0]\right)^l > 0)(-1)^{l+k} \left(\mathbb{E}[Y_1 - Y_0]\right)^l \mathbb{E}^L[Y_1^{m-l-k} Y_0^k] \Big) \\
&\leq \mathbb{E}[\{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\}^m] \\
&\leq \sum_{l=0}^{m} \sum_{k=0}^{m-l} \binom{m}{l} \binom{m-l}{k} \Big(\mathbb{I}((-1)^{l+k} \left(\mathbb{E}[Y_1 - Y_0]\right)^l \leq 0)(-1)^{l+k} \left(\mathbb{E}[Y_1 - Y_0]\right)^l \mathbb{E}^L[Y_1^{m-l-k} Y_0^k] \\
&+ \mathbb{I}((-1)^{l+k} \left(\mathbb{E}[Y_1 - Y_0]\right)^l > 0)(-1)^{l+k} \left(\mathbb{E}[Y_1 - Y_0]\right)^l \mathbb{E}^U[Y_1^{m-l-k} Y_0^k] \Big).
\end{aligned} \tag{54}$$

Similarly, the skewness and kurtosis of ICDE and ICRE are also bounded. We denote the lower and upper bounds of variance, the central third and fourth moments of ICDE (or ICRE) as $\overline{\mu}_L^{(2)}$, $\overline{\mu}_U^{(2)}$, $\overline{\mu}_L^{(3)}$, $\overline{\mu}_U^{(3)}$, $\overline{\mu}_L^{(4)}$, and $\overline{\mu}_U^{(4)}$. We can calculate the bounds of the skewness of ICDE (or ICRE) as follows:

$$\begin{aligned}\frac{\overline{\mu}_L^{(3)}}{\left(\overline{\mu}_U^{(2)}\right)^{3/2}}\mathbb{I}\big(\overline{\mu}_L^{(3)} \geq 0\big) + \frac{\overline{\mu}_L^{(3)}}{\left(\overline{\mu}_L^{(2)}\right)^{3/2}}\mathbb{I}\big(\overline{\mu}_L^{(3)} < 0\big) &\leq \frac{\overline{\mu}^{(3)}}{\left(\overline{\mu}^{(2)}\right)^{3/2}} \\ &\leq \frac{\overline{\mu}_U^{(3)}}{\left(\overline{\mu}_L^{(2)}\right)^{3/2}}\mathbb{I}\big(\overline{\mu}_U^{(3)} \geq 0\big) + \frac{\overline{\mu}_U^{(3)}}{\left(\overline{\mu}_U^{(2)}\right)^{3/2}}\mathbb{I}\big(\overline{\mu}_U^{(3)} < 0\big),\end{aligned} \tag{55}$$

and the bounds of the kurtosis of ICDE (or ICRE) are calculated as follows:

$$\frac{\overline{\mu}_L^{(4)}}{\left(\overline{\mu}_U^{(2)}\right)^2} \leq \frac{\overline{\mu}^{(4)}}{\left(\overline{\mu}^{(2)}\right)^2} \leq \frac{\overline{\mu}_U^{(4)}}{\left(\overline{\mu}_L^{(2)}\right)^2}. \tag{56}$$

These bounds are not sharp.

C Details of Estimation of Joint Expectation over POs

The following empirical CDFs will be used in the computation:

$$\hat{F}_{Y_0}(y_0) = \frac{1}{N^C}\sum\nolimits_{j=1}^{N^C}\mathbb{I}(Y_j^C \leq y_0), \tag{57}$$

$$\hat{F}'_{Y_0}(y_0) = \frac{1}{N^C}\sum\nolimits_{j=1}^{N^C}\mathbb{I}(Y_j^C < y_0), \tag{58}$$

$$\hat{F}_{Y_1}(y_1) = \frac{1}{N^T}\sum\nolimits_{i=1}^{N^T}\mathbb{I}(Y_i^T \leq y_1), \tag{59}$$

$$\hat{F}_{Y_1}^{-1}(p) = \min_{i=1,\ldots,N^T}\{Y_i^T; \hat{F}_{Y_1}(Y_i^T) \geq p\}. \tag{60}$$

We have

$$\tilde{Y}_i^{U,T} = \hat{F}_{Y_1}^{-1}(U_i(\hat{F}_{Y_0}(Y_i^C) - \hat{F}'_{Y_0}(Y_i^C)) + \hat{F}'_{Y_0}(Y_i^C)), \tag{61}$$

$$\tilde{Y}_i^{L,T} = \hat{F}_{Y_1}^{-1}(1-(1-U_i)(\hat{F}_{Y_0}(Y_i^C) - \hat{F}'_{Y_0}(Y_i^C)) - \hat{F}'_{Y_0}(Y_i^C)). \tag{62}$$

D Numerical Experiments

We illustrate the finite-sample properties of our estimators. We pick up the moments of ICDE and ICRE as examples.

D.1 Experiments for Expectations over ICDE

Baselines. We compare our estimators with the following baselines.
(i) The estimator based on the rank preservation assumptions (RP) (Assumptions 6 and 7) [Heckman et al., 1997, Wu et al., 2025]. RP corresponds to setting $u = 0$ in our estimator for $\mathbb{E}^U[g(Y_1, Y_0)]$ instead of drawing u from Unif$(0, 1)$; at $u = 0$, the right-hand side of Eq. (20) becomes $F_{Y_1}^{-1}(F'_{Y_0}(y_0))$, which coincides with that of Eq. (9) when F_{Y_0} is continuous (i.e., $F_{Y_0} = F'_{Y_0}$).
(ii) The Monte Carlo integration-based estimator (MCI) in [Kawakami and Tian, 2025]. The MCI estimator requires Assumption 5 for identification and can also compute bounds. It requires specifying the outcome domain range, $\Omega_Y = [a, b]$, which we used the minimum and maximum values in the dataset.
(iii) The causal-optimal-transportation-based estimator (COT) for causal bounding in [Gao et al., 2025]. COT relies on linear programming.
(iv) COT with Sinkhorn regularization. Sinkhorn regularization is an entropic regularization technique for optimal transport that enables efficient computation.

Table 2: Mean estimates with 95% CI of moments (M.) of ICDE under SCM (A). The estimates reported as "Ours" are obtained under the comonotonicity assumption, which the SCM satisfies.

Statistics	$N = 20$	$N = 100$	$N = 1000$	Ground Truth
2nd M. (**Ours**)	$1.033[0.134, 2.609]$	$0.869[0.501, 1.189]$	$0.836[0.698, 0.984]$	0.833
2nd M. (RP)	$2.081[0.867, 4.144]$	$2.036[1.271, 3.040]$	$2.137[1.843, 2.418]$	0.833
2nd M. (MCI)	$0.890[0.333, 1.542]$	$0.882[0.577, 1.192]$	$0.843[0.716, 0.980]$	0.833
3rd M. (**Ours**)	$-0.924[-4.586, 0.873]$	$-0.817[-1.817, -0.058]$	$-0.903[-1.162, -0.657]$	-0.875
3rd M. (RP)	$-3.896[-10.148, -0.862]$	$-3.648[-5.662, -2.321]$	$-3.820[-4.355, -3.428]$	-0.875
3rd M. (MCI)	$-0.665[-2.337, 0.569]$	$-0.806[-1.361, -0.237]$	$-0.888[-1.114, -0.637]$	-0.875
4th M. (**Ours**)	$2.125[0.026, 5.906]$	$1.813[0.427, 4.022]$	$1.734[1.303, 2.263]$	1.700
4th M. (RP)	$8.436[1.166, 32.412]$	$7.451[4.375, 10.298]$	$8.026[6.905, 8.973]$	1.700
4th M. (MCI)	$1.824[0.213, 3.884]$	$1.687[0.952, 2.883]$	$1.685[1.218, 2.447]$	1.700

Table 3: Mean estimates with 95% CIs for the upper (UB) and lower (LB) bounds of the moments (M.) of ICDE under SCM (B). A checkmark (✓) indicates that the estimator is based on sharp bounds. COT is a numerical approach and does not use closed-form expressions for the bounds. The Sinkhorn regularization parameter is 100. COT results for $N = 1000$ are omitted because of computational resource constraints.

Statistics	$N = 20$	$N = 100$	$N = 1000$	True M. value
UB of 2nd M. (**Ours**) ✓	$4.393[1.602, 7.648]$	$4.342[3.180, 5.742]$	$4.347[3.958, 4.783]$	4.100
LB of 2nd M. (**Ours**) ✓	$3.656[0.912, 6.574]$	$3.382[2.079, 4.807]$	$3.359[2.899, 3.809]$	4.100
UB of 2nd M. (MCI) ✓	$4.359[1.585, 7.557]$	$4.374[3.154, 5.870]$	$4.348[3.831, 4.822]$	4.100
LB of 2nd M. (MCI)	$0.984[0.053, 4.734]$	$0.642[0.232, 1.342]$	$0.572[0.394, 0.772]$	4.100
UB of 2nd M. (COT)	$4.420[2.460, 7.815]$	$4.260[3.073, 5.519]$	–	4.100
LB of 2nd M. (COT)	$3.634[1.813, 6.705]$	$3.319[2.253, 4.475]$	–	4.100
UB of 2nd M. (COT; Sinkhorn)	$3.986[2.078, 7.202]$	$3.714[2.578, 4.948]$	$3.748[3.342, 4.164]$	4.100
LB of 2nd M. (COT; Sinkhorn)	$3.980[2.071, 7.189]$	$3.706[2.572, 4.938]$	$3.740[3.335, 4.155]$	4.100
UB of 3rd M. (**Ours**)	$16.261[4.434, 32.235]$	$16.501[11.386, 22.740]$	$17.358[15.527, 19.211]$	13.500
LB of 3rd M. (**Ours**)	$8.466[-0.247, 22.496]$	$6.951[3.084, 11.547]$	$7.349[6.007, 8.695]$	13.500
UB of 3rd M. (MCI)	$14.467[3.791, 29.534]$	$14.895[9.526, 20.092]$	$15.309[13.789, 17.401]$	13.500
LB of 3rd M. (MCI)	$0.933[0.000, 7.114]$	$0.163[0.000, 0.637]$	$0.106[0.009, 0.259]$	13.500
UB of 3rd M. (COT)	$14.244[3.196, 26.965]$	$15.296[9.470, 23.007]$	–	13.500
LB of 3rd M. (COT)	$9.363[1.254, 17.763]$	$9.310[5.436, 13.841]$	–	13.500
UB of 3rd M. (COT; Sinkhorn)	$11.474[3.244, 23.400]$	$11.592[6.942, 18.224]$	$11.450[9.996, 12.907]$	13.500
LB of 3rd M. (COT; Sinkhorn)	$11.222[3.165, 22.961]$	$11.249[6.681, 17.709]$	$11.108[9.715, 12.544]$	13.500
UB of 4th M. (**Ours**) ✓	$49.656[4.782, 121.923]$	$57.574[30.873, 84.743]$	$58.409[50.830, 66.497]$	48.944
LB of 4th M. (**Ours**) ✓	$27.307[1.071, 61.272]$	$28.798[11.022, 49.474]$	$27.877[21.527, 34.861]$	48.944
UB of 4th M. (MCI) ✓	$50.697[6.849, 106.510]$	$59.138[29.202, 89.327]$	$59.425[51.387, 67.160]$	48.944
LB of 4th M. (MCI)	$0.329[0.000, 2.484]$	$0.026[0.000, 0.278]$	$0.010[0.000, 0.137]$	48.944
UB of 4th M. (COT)	$52.338[11.757, 111.960]$	$57.791[30.035, 95.728]$	–	48.944
LB of 4th M. (COT)	$29.518[3.053, 68.991]$	$29.173[15.233, 53.132]$	–	48.944
UB of 4th M. (COT; Sinkhorn)	$37.118[3.098, 99.440]$	$42.780[25.074, 71.245]$	$43.257[36.782, 50.880]$	48.944
LB of 4th M. (COT; Sinkhorn)	$30.258[2.921, 81.995]$	$33.744[19.564, 54.594]$	$33.707[28.965, 39.999]$	48.944

Settings. We assume the following SCMs (A) and (B):

$$\text{(A): } Y := -2U_Y\mathbb{I}(X = 1) - U_Y\mathbb{I}(U_Y < 0, X = 0), \tag{63}$$

$$\text{(B): } Y := 4(U_Y)^2\mathbb{I}(X = 1) - (U_Y)^2\mathbb{I}(U_Y > 0, X = 0)+1, \tag{64}$$

where $U_Y \sim \text{Unif}(-1, 1)$ and $X \sim \text{Bern}(0.5)$. $\text{Bern}(p)$ represents a Bernoulli distribution with probability p. Y_1 is a continuous variable, whereas Y_0 is a mixed variable under both SCM (A) and (B). Y takes nonnegative values under SCM (B). Both SCMs satisfy the exogeneity assumption, and violate the strictly monotonicity (Assumption 3), the rank preservation (Assumption 6), and the rank preservation on Kendall's $\tilde{\rho}$ (Assumption 7) assumptions. SCM (A) satisfies the comonotonicity (Assumption 1) and the monotonicity over f_Y (Assumption 5) whereas SCM (B) does not. In setting (A), we focus on point identification. In setting (B), we compute bounds. We simulate 100 times with the total sample size $N^C + N^T = 20, 100, 1000$, respectively, and report the mean along with the 95% confidence intervals (CI) for each estimate. The experiments are performed using an Apple M4 (24GB).

Results. Table 2 reports the mean estimates with 95% CI of the second, third, and fourth moments (2nd M., 3rd M., and 4th M.) of the ICDE for each estimator under SCM (A), together with the corresponding ground truths. Both our estimator and MCI perform well (their identification assumptions are satisfied in SCM (A)), while RP exhibits severe bias (its identification assumption is violated).

Table 3 reports the estimates of the upper (UB) and lower (LB) bounds of the second, third, and fourth moments of the ICDE for each estimator under SCM (B), together with the corresponding true

Table 4: Runtime Comparison for a Single Simulation. COT for $N^C + N^T = 2000$ and 5000 exceeds the available memory. The regularization parameter of Sinkhorn is set to 100.

Time	$N^C + N^T = 20$	100	1000	2000	5000
Ours	0.051 sec.	0.055 sec.	0.134 sec.	0.344 sec.	1.110 sec.
COT	0.001 sec.	0.243 sec.	203.476 sec.	-	-
COT (Sinkhorn)	0.001 sec.	0.003 sec.	0.155 sec.	0.782 sec.	8.230 sec.

moment values. The true values fall within the estimated bounds. For the 2nd and 4th moments, our bounds are tighter than the MCI bounds, particularly for the lower bounds. This is consistent with our theoretical results that for the even-order moments of ICDE, our bounds are sharp, while MCI's upper bounds are sharp but lower bounds are not. For the 3rd moment, neither method's bounds are sharp, and our upper bound can be wider than that of MCI.

Table 4 reports the runtime for a single simulation of our estimator and the COT estimator. COT relies on linear programming, which can be computationally demanding, $\mathcal{O}((N^T + N^C)^3)$ time and $\mathcal{O}((N^T + N^C)^2)$ memory. In contrast, our estimator requires $\mathcal{O}((N^T + N^C)\log(N^T + N^C))$ time and $\mathcal{O}(N^T + N^C)$ memory. COT estimators are very similar to ours for $N = 20$ and $N = 100$. However, results for $N = 1000$ could not be obtained; nonetheless, our estimator significantly reduces computational cost while achieving similar confidence intervals compared to COT. This reduction in computational resources arises from our closed-form expressions for the bounds. Regularization methods for COT, such as Sinkhorn [Cuturi, 2013], greatly reduce computational cost; however, they can introduce substantial bias in the estimates and may fail to cover the ground truth, as shown in Table 3.

D.2 Experiments for Expectations over ICRE

Baseline. We compare our estimators with (i) the RP estimator and (ii) the causal-optimal-transportation (COT) estimator. MCI does not work for ICRE.

Settings. We assume the following SCM (C) and (D):

$$\text{(C): } Y := 6U_Y\mathbb{I}(X = 1) + U_Y\mathbb{I}(U_Y > 0.75, X = 0) + 2, U_Y \sim \text{Unif}(0, 1), \tag{65}$$

$$\text{(D): } Y := 6(U_Y)^2\mathbb{I}(X = 1) + (U_Y)^2\mathbb{I}(U_Y > 0.75, X = 0) + 2, U_Y \sim \text{Unif}(-1, 1), \tag{66}$$

where $X \sim \text{Bern}(0.5)$. Y_1 is a continuous variable, whereas Y_0 is a mixed variable under both SCM (C) and (D). Y takes positive values under both (C) and (D). Both SCMs satisfy the exogeneity, and violate the rank preservation (Assumption 6) and the rank preservation on Kendall's $\tilde{\rho}$ (Assumption 7). SCM (C) satisfies the comonotonicity (Assumption 1), whereas SCM (D) does not. In setting (C), we focus on point identification. In setting (D), we compute bounds. We simulate 100 times with the sample size $N^C + N^T = 20, 100, 1000$, respectively.

Results. Table 5 reports the mean estimates with 95% CI of the first, second, third, and fourth moments (1st M., 2nd M., 3rd M., and 4th M.) of the ICRE for each estimator under SCM (C), together with the corresponding ground truths. Our estimator performs well, while RP exhibits severe bias (its identification assumption is violated).

Table 6 reports the estimates of the bounds of the first, second, third, and fourth moments of the ICRE under SCM (D), together with the corresponding true moment values. These bounds are theoretically sharp.

Similarly to ICDE, COT estimators are very similar to ours for $N = 20$ and $N = 100$. However, results for $N = 1000$ could not be obtained due to computational constraints. Regularization methods for COT, such as Sinkhorn [Cuturi, 2013], can greatly reduce computational cost; however, they may introduce substantial bias and fail to cover the ground truth, as shown in Table 6.

D.3 Experiments Under the Rank Preservation Assumption

We provide additional numerical experiments.

Settings. We assume the following SCMs (E) and (F):

$$\text{(E): } Y := -(X + 1)U_Y, U_Y \sim \text{Unif}(-1, 1), \tag{67}$$

Table 5: Mean estimates with 95% CI of moments (M.) of ICRE under SCM (C). The estimates reported as "Ours" are obtained under the comonotonicity assumption, which the SCM satisfies.

Statistics	$N = 20$	$N = 100$	$N = 1000$	Ground Truth
1st M. (**Ours**)	$2.173[1.591, 2.794]$	$2.212[1.942, 2.527]$	$2.231[2.144, 2.342]$	2.223
1st M. (RP)	$1.523[1.154, 1.999]$	$1.417[1.293, 1.584]$	$1.378[1.337, 1.429]$	2.223
2nd M. (**Ours**)	$5.615[2.891, 9.210]$	$5.425[4.171, 6.554]$	$5.318[4.825, 5.896]$	5.292
2nd M. (RP)	$2.578[1.418, 4.289]$	$2.341[1.842, 2.849]$	$2.335[2.197, 2.507]$	5.292
3rd M. (**Ours**)	$14.252[6.384, 26.138]$	$13.392[8.298, 18.450]$	$13.233[11.594, 15.262]$	13.227
3rd M. (RP)	$4.808[1.763, 8.546]$	$4.675[3.507, 5.853]$	$4.759[4.267, 5.228]$	13.227
4th M. (**Ours**)	$45.743[9.299, 106.961]$	$34.699[20.084, 54.742]$	$34.826[29.713, 42.228]$	34.519
4th M. (RP)	$9.529[1.264, 21.594]$	$10.152[6.984, 13.769]$	$10.922[9.998, 11.896]$	34.519

Table 6: Mean estimates with 95% CI of upper (UB) and lower (LB) bounds of moments (M.) of ICRE under SCM (D). A checkmark (✓) indicates whether the estimator is based on sharp bounds (for our method). We do not report COT results for $N = 1000$ due to computational resource constraints. The regularization parameter of Sinkhorn is set to 100.

Statistics	$N = 20$	$N = 100$	$N = 1000$	True M. value
UB of 1st M. (**Ours**) (✓)	$1.985[1.417, 2.839]$	$1.985[1.727, 2.275]$	$1.959[1.867, 1.965]$	1.884
LB of 1st M. (**Ours**) (✓)	$1.875[1.386, 2.631]$	$1.874[1.644, 2.193]$	$1.873[1.789, 2.196]$	1.884
UB of 1st M. (COT)	$1.993[1.454, 2.528]$	$1.965[1.709, 2.247]$	-	1.884
LB of 1st M. (COT)	$1.912[1.355, 2.451]$	$1.875[1.598, 2.166]$	-	1.884
UB of 1st M. (COT; Sinkhorn)	$1.913[1.449, 2.444]$	$1.926[1.700, 2.175]$	$1.937[1.870, 1.999]$	1.884
LB of 1st M. (COT; Sinkhorn)	$1.913[1.449, 2.444]$	$1.926[1.700, 2.175]$	$1.937[1.870, 1.999]$	1.884
UB of 2nd M. (**Ours**) (✓)	$4.572[1.824, 8.662]$	$4.736[3.382, 6.263]$	$4.770[4.322, 5.256]$	4.124
LB of 2nd M. (**Ours**) (✓)	$4.077[1.667, 8.637]$	$4.002[2.980, 5.415]$	$4.021[3.600, 4.445]$	4.124
UB of 2nd M. (COT)	$4.519[2.425, 7.374]$	$4.746[3.557, 6.085]$	-	4.124
LB of 2nd M. (COT)	$4.017[2.065, 6.696]$	$3.986[2.995, 5.401]$	-	4.124
UB of 2nd M. (COT; Sinkhorn)	$4.404[2.210, 6.406]$	$4.553[3.666, 5.742]$	$4.548[4.184, 4.914]$	4.124
LB of 2nd M. (COT; Sinkhorn)	$4.397[2.208, 6.391]$	$4.544[3.657, 5.728]$	$4.539[4.176, 4.906]$	4.124
UB of 3rd M. (**Ours**) (✓)	$13.111[3.629, 22.645]$	$13.447[7.029, 20.343]$	$12.915[10.515, 14.949]$	10.251
LB of 3rd M. (**Ours**) (✓)	$9.578[2.926, 25.494]$	$9.900[5.703, 15.309]$	$9.338[7.772, 10.831]$	10.251
UB of 3rd M. (COT)	$12.530[3.934, 22.876]$	$13.375[9.299, 17.724]$	-	10.251
LB of 3rd M. (COT)	$9.918[3.263, 20.067]$	$9.752[6.556, 13.875]$	-	10.251
UB of 3rd M. (COT; Sinkhorn)	$12.282[4.443, 20.953]$	$12.183[8.529, 16.304]$	$12.362[11.011, 13.879]$	10.251
LB of 3rd M. (COT; Sinkhorn)	$12.110[4.414, 20.933]$	$11.977[8.397, 16.067]$	$12.144[10.807, 13.670]$	10.251
UB of 4th M. (**Ours**)(✓)	$39.128[5.569, 100.039]$	$40.786[22.117, 65.312]$	$39.770[33.264, 48.310]$	28.132
LB of 4th M. (**Ours**) (✓)	$27.363[5.082, 80.050]$	$25.602[12.818, 45.459]$	$23.712[19.036, 29.0653]$	28.132
UB of 4th M. (COT)	$37.648[11.288, 67.875]$	$38.829[24.517, 55.395]$	-	28.132
LB of 4th M. (COT)	$27.073[7.380, 64.114]$	$24.258[13.207, 38.913]$	-	28.132
UB of 4th M. (COT; Sinkhorn)	$37.120[12.387, 72.289]$	$37.963[22.942, 52.821]$	$38.267[32.890, 43.705]$	28.132
LB of 4th M. (COT; Sinkhorn)	$33.442[11.887, 68.291]$	$33.860[20.124, 48.042]$	$34.026[29.083, 38.746]$	28.132

$$\text{(F): } Y := X + 1 + U_Y, U_Y \sim \text{Unif}(0, 1). \tag{68}$$

These SCMs satisfy the rank preservation assumption (Assumption 6). In settings (E) and (F), we compare our estimator with RP and MCI under the assumption that all estimators are unbiased.

Results. Table 7 shows the results of the numerical experiments for ICDE under SCM (E) and Table 8 shows the results of the numerical experiments for ICRE under SCM (F). All mean estimates are close to the ground truth, particularly when the sample size is large. Our estimator, RP, and MCI have almost the same 95% CIs.

Table 7: Results of numerical experiments for ICDE under SCM (E). The estimates reported as "Ours" are obtained under the comonotonicity assumption, which the SCM satisfies.

Statistics	$N = 20$	$N = 100$	$N = 1000$	Ground Truth
2nd M. (**Ours**)	$0.627[0.177, 1.631]$	$0.380[0.203, 0.632]$	$0.338[0.274, 0.397]$	0.333
2nd M. (RP)	$0.664[0.176, 1.536]$	$0.407[0.208, 0.651]$	$0.343[0.286, 0.415]$	0.333
2nd M. (MCI)	$0.616[0.139, 1.205]$	$0.382[0.198, 0.610]$	$0.338[0.265, 0.407]$	0.333
3rd M. (**Ours**)	$-0.038[-1.116, 1.013]$	$-0.000[-0.289, 0.359]$	$-0.000[-0.106, 0.075]$	0.000
3rd M. (RP)	$-0.403[-1.866, 0.641]$	$-0.040[-0.280, 0.255]$	$-0.008[-0.093, 0.066]$	0.000
3rd M. (MCI)	$0.002[-1.070, 1.176]$	$0.023[-0.345, 0.295]$	$-0.001[-0.112, 0.099]$	0.000
4th M. (**Ours**)	$0.789[0.047, 2.714]$	$0.263[0.106, 0.496]$	$0.205[0.153, 0.272]$	0.200
4th M. (RP)	$0.774[0.035, 2.275]$	$0.272[0.105, 0.601]$	$0.206[0.151, 0.262]$	0.200
4th M. (MCI)	$0.870[0.054, 4.082]$	$0.269[0.086, 0.556]$	$0.205[0.080, 0.336]$	0.200

Table 8: Results of numerical experiments for ICRE under SCM (F). The estimates reported as "Ours" are obtained under the comonotonicity assumption, which the SCM satisfies.

Statistics	$N = 20$	$N = 100$	$N = 1000$	Ground Truth
1st M. (**Ours**)	$1.703[1.473, 1.926]$	$1.693[1.586, 1.797]$	$1.696[1.668, 1.728]$	1.693
1st M. (RP)	$1.685[1.446, 1.927]$	$1.695[1.607, 1.798]$	$1.695[1.664, 1.723]$	1.693
2nd M. (**Ours**)	$2.977[2.114, 3.910]$	$2.867[2.508, 3.259]$	$2.883[2.773, 2.981]$	2.886
2nd M. (RP)	$2.821[2.139, 3.790]$	$2.883[2.526, 3.283]$	$2.884[2.783, 2.998]$	2.886
3rd M. (**Ours**)	$5.233[3.395, 7.428]$	$5.020[4.227, 5.981]$	$4.967[4.692, 5.315]$	4.961
3rd M. (RP)	$4.904[2.953, 7.017]$	$4.983[4.121, 5.962]$	$4.978[4.632, 5.252]$	4.961
4th M. (**Ours**)	$8.726[4.956, 13.613]$	$8.806[7.016, 10.951]$	$8.595[8.010, 9.305]$	8.575
4th M. (RP)	$8.312[5.237, 12.498]$	$8.592[6.513, 10.973]$	$8.607[8.004, 9.308]$	8.575

E Applications to Real-World Datasets

E.1 Cholesterol Reduction Dataset

Dataset. We analyze a medical dataset "Cholesterol Reduction" [Westfall et al., 2011] (`https://search.r-project.org/CRAN/refmans/multcomp/html/cholesterol.html`). This study investigates the effect of three formulations of the same drug (X) on cholesterol reduction (Y). The three formulations consist of 20 mg taken once daily ("1 time", $X = 1$), 10 mg taken twice daily ("2 times", $X = 2$), and 5 mg taken four times daily ("4 times", $X = 4$). The dataset has a sample size of 10 for each treatment group. The domain of Y is bounded within $[5, 19]$, and all values are strictly positive. The variance, skewness, and kurtosis of ICDE for this dataset were analyzed by Kawakami and Tian [2025].

We assume the exogeneity condition holds, as Westfall et al. [2011] reported no potential confounding variables. We evaluate the variance, skewness, and kurtosis of the ICDE for the effects from $X = 1$ to $X = 2$, from $X = 2$ to $X = 4$, and from $X = 1$ to $X = 4$, respectively, and compare the results with those reported in [Kawakami and Tian, 2025]. In addition, we assess the mean, variance, skewness, and kurtosis of the ICRE for the same pairs of treatment levels. Comonotonicity is plausible when, at the same total exposure, splitting the dose does not materially alter efficacy profiles, so that patients who respond well under one dosing schedule also respond well under the other. For example, comonotonicity of (Y_1, Y_2) means that patients who would experience a larger cholesterol reduction on 20 mg once daily would also experience a larger reduction on 10 mg twice daily. We conduct the bootstrapping [Efron, 1979] to provide the means and 95% confidence intervals (CIs) for each estimator.

Results. The complete results are reported in Table 9. For ICDE, our method and MCI agree on the estimated variance, but differ substantially on the estimated skewness and kurtosis. Their skewness and kurtosis estimates are likely unreliable given their wide CIs. In general, our estimates have significantly narrower CIs than MCI. Still, these CIs are relatively wide due to the small number of samples. If these mean estimates are to be trusted, under the comonotonicity assumption, the estimated skewness indicates that the distribution of the ICDE is approximately symmetric, meaning that the number of individuals with effects greater than the average is roughly equal to the number of individuals with effects less than the average, in contrast to the large positive skewness reported by Kawakami and Tian [2025]. The estimated kurtosis suggests heavier tails than the Gaussian distribution (kurtosis = 3), albeit to a lesser extent than that reported by Kawakami and Tian [2025], which indicates a high number of outliers with causal effects that differ significantly from the average. Given these results under the comonotonicity assumption, practitioners may hesitate to implement the treatment based solely on the mean ICDE, as doing so could result in substantially positive or negative impacts that deviate from the mean for a meaningful minority of individuals.

The bound for the variance (and skewness and kurtosis) of ICDE is quite wide, indicating that the results obtained under comonotonicity could be misleading if, in fact, comonotonicity does not hold.

Under the comonotonicity assumption, the estimated mean of ICRE indicates a positive effect (greater than 1), which is consistent with the positive mean of ICDE. The estimated bound for the mean of ICRE is relatively narrow, and confirms the positive effect without assuming comonotonicity. The estimated variance suggests that the heterogeneity of the ICRE is relatively limited. The estimated skewness values indicate that the ICRE distributions are approximately symmetric. The kurtosis

estimates suggest lighter tails than those of a Gaussian distribution for Y_2/Y_1 and Y_4/Y_1, but slightly heavier tails for Y_4/Y_2.

We provide further information about the application to the cholesterol reduction data. We evaluate the variance, skewness, and kurtosis of the ICDE for the effects from $X = 1$ to $X = 2$, from $X = 2$ to $X = 4$, and from $X = 1$ to $X = 4$, respectively, and compare the results with those reported in Kawakami and Tian [2025]. In addition, we assess the mean, variance, skewness, and kurtosis of the ICRE for the same pairs of treatment levels.

Table 9 shows all the results for ICDE $Y_2 - Y_1$ and ICRE Y_2/Y_1, Table 10 shows the results for ICDE $Y_4 - Y_2$ and ICRE Y_4/Y_2, and Table 11 shows the results for ICDE $Y_4 - Y_1$ and ICRE Y_4/Y_1. The results of MCI were reported by [Kawakami and Tian, 2025]. Under the comonotonicity assumption, the estimates suggest the following. The ICDE $Y_4 - Y_2$ exhibits a similar mean and variance to $Y_2 - Y_1$, but is slightly negatively skewed and displays heavier tails. The ICDE $Y_4 - Y_1$ exhibits a larger mean and smaller variance compared to $Y_2 - Y_1$, with an approximately symmetric distribution and relatively lighter tails. The ICRE Y_4/Y_2 exhibits a smaller mean and variance compared to Y_2/Y_1, but shows greater skewness and heavier tails. The ICRE Y_4/Y_1 exhibits a greater mean and variance compared to Y_2/Y_1, with slightly lower skewness and slightly lighter tails. All bounds for skewness and kurtosis are considerably wide.

These conclusions, which rely on the comonotonicity assumption, may change if the potential outcomes have a different dependence structure.

E.2 Lalonde Dataset

Dataset. We analyze the Lalonde dataset [LaLonde, 1986] (`https://users.nber.org/~rdehejia/nswdata2.html`). This dataset contains both experimental and observational data and is commonly used as a benchmark to assess whether observational inference methods can recover results consistent with experimental findings. In contrast to this line of research, in this paper, we evaluate the shape of the distribution of causal effects, which cannot be directly computed even from experimental data. The sample size of this dataset is 722. The treatment variable is a binary indicator (1 = treated, 0 = control). We use real earnings in 1978 as the outcome variable, which ranges from 0 to 6.031×10^4. Most individuals have positive, continuous income, while 196 subjects report zero income; thus, the outcome is mixed. For the calculation of the ICRE statistic, we add 1 to the outcome to avoid division by zero. We select the regularization parameter by increasing it as $1, 10, 100, \ldots$, and stop when the computation finishes within one hour. We do not report COT results due to computational constraints. We conduct the bootstrapping [Efron, 1979] to provide the means and 95% CIs for each estimator.

We assume the exogeneity condition holds, as we use experimental data from the National Supported Work (NSW) demonstration [LaLonde, 1986]. We evaluate the variance, skewness, and kurtosis of the ICDE, and the mean, variance, skewness, and kurtosis of the ICRE for $Y_1 - Y_0$ and Y_1/Y_0, respectively. Comonotonicity of (Y_0, Y_1) means that individuals who would earn more under the control condition would also earn more under the treatment condition. This is plausible if the job training program benefits participants in proportion to their pre-existing earning potential, so that individuals who would perform better without the program also perform better with it. Comonotonicity would be less plausible if the program were designed as a remedial intervention that disproportionately benefits the lowest earners, since in that case individuals who rank low under control might rank higher under treatment.

Results. The complete results are also reported in Table 12. The mean of ICDE is positive, indicating that the treatment has a positive effect on average. Researchers may conclude that the treatment should be implemented. However, statistics based on higher moments of ICDE reveal risks associated with the treatment. The variance of ICDE is non-negligible relative to the scale of the outcome $[0, 6.031 \times 10^4]$. This implies substantial heterogeneity in the causal effect of the treatment on earnings. Furthermore, the distribution is positively skewed and highly kurtotic, which together imply the presence of substantial tail risk, including both large upside gains and potential downside losses for some individuals.

The mean of ICRE is large, suggesting that the treatment increases the outcome by a factor of about 50 on average. However, similarly to ICDE, ICRE exhibits substantial heterogeneity in causal effects.

ICRE is positively skewed and highly kurtotic. ICRE is also highly kurtotic. They suggest the presence of tail risks in ICRE.

We also report results under the rank preservation assumption. These results differ from our estimates, particularly for the ICRE statistics, suggesting that imposing rank preservation in the presence of mixed variables can introduce bias.

Similarly, COT with Sinkhorn regularization may exhibit substantial regularization bias. Such biases can lead to severely misleading scientific conclusions.

Table 9: Estimates of the statistics for $Y_2 - Y_1$ and Y_2/Y_1. We report the means of the estimates along with their 95% CIs. The results of MCI were reported by [Kawakami and Tian, 2025]. MCI is not applicable to ICRE. A checkmark (✓) indicates whether the estimator is based on sharp bounds (for our method and MCI).

Statistics for $Y_2 - Y_1$ and Y_2/Y_1	Estimates
Mean of ICDE	3.432 [0.914, 6.104]
Variance of ICDE under comonotonicity (Ours)	2.920 [0.551,6.978]
Variance of ICDE under monotonicity over f_Y (MCI)	3.072 [0.297,8.610]
Skewness of ICDE under comonotonicity (Ours)	-0.184 [-2.612,2.713]
Skewness of ICDE under monotonicity over f_Y (MCI)	21.027 [-6.747,34.504]
Kurtosis of ICDE under comonotonicity (Ours)	6.332 [1.264,15.078]
Kurtosis of ICDE under monotonicity over f_Y (MCI)	21.312 [0.000,203.885]
UB of Variance of ICDE (Ours) (✓)	30.508 [11.422,54.639]
LB of Variance of ICDE (Ours) (✓)	2.920 [0.551,6.978]
UB of Variance of ICDE (MCI) (✓)	29.851 [12.173, 53.740]
LB of Variance of ICDE (MCI)	0.000 [0.000,0.000]
UB of Variance of ICDE (COT)	29.732 [13.317, 47.979]
LB of Variance of ICDE (COT)	2.858 [0.873, 6.150]
UB of Skewness of ICDE (Ours)	5.580 [3.020,11.677]
LB of Skewness of ICDE (Ours)	-1.204 [-6.722,15.312]
UB of Skewness of ICDE (MCI)	71296.725 [0.000,258849.124]
LB of Skewness of ICDE (MCI)	-13168.940 [-64712.250,0.000]
UB of Skewness of ICDE (COT)	29.057 [1.762, 90.612]
LB of Skewness of ICDE (COT)	-27.294 [-98.295, -2.509]
UB of Kurtosis of ICDE (Ours)	155.876 [73.433,298.713]
LB of Kurtosis of ICDE (Ours)	0.000 [0.000,0.000]
UB of Kurtosis of ICDE (MCI)	6.638×10^6 [0.000, 3.526×10^7]
LB of Kurtosis of ICDE (MCI)	0.000 [0.000,0.000]
UB of Kurtosis of ICDE (COT)	473.925 [31.378, 1817.645]
LB of Kurtosis of ICDE (COT)	0.031 [0.002, 0.141]
Mean of ICRE under comonotonicity	1.765 [1.210,2.442]
Variance of ICRE under comonotonicity	0.267 [0.041,0.577]
Skewness of ICRE under comonotonicity	0.378[-0.916,1.778]
Kurtosis of ICRE under comonotonicity	2.328[1.282,4.785]
UB of Mean of ICRE (Ours) (✓)	2.314 [1.559,3.215]
LB of Mean of ICRE (Ours) (✓)	1.765 [1.210,2.442]
UB of Mean of ICRE (COT)	2.271 [1.479, 3.071]
LB of Mean of ICRE (COT)	1.741 [1.174, 2.508]
UB of Variance of ICRE (Ours)	5.540 [2.059,9.854]
LB of Variance of ICRE (Ours)	0.000 [0.000,0.000]
UB of Variance of ICRE (COT)	5.258 [1.541, 9.517]
LB of Variance of ICRE (COT)	0.000 [0.000,0.000]
UB of Skewness of ICRE (Ours)	56382.41 [6343.13,111289.8]
LB of Skewness of ICRE (Ours)	-50319.46 [-102465.4,-5773.89]
UB of Skewness of ICRE (COT)	1.101×10^{13} [9.568×10^{11} , 2.567×10^{13}]
LB of Skewness of ICRE (COT)	-7.236×10^{10} [-4.137×10^{11}, 0.048]
UB of Kurtosis of ICRE (Ours)	488100.6 [39205.61,1430178]
LB of Kurtosis of ICRE (Ours)	0.000 [0.000,0.000]
UB of Kurtosis of ICRE (COT)	4.829×10^{13} [2.407×10^{12}, 1.214×10^{14}]
LB of Kurtosis of ICRE (COT)	0.013 [0.000, 0.074]

Table 10: Estimates of the statistics for $Y_4 - Y_2$ and Y_4/Y_2. We report the means of the estimates along with their 95% CIs. The results of MCI were reported by [Kawakami and Tian, 2025]. MCI is not applicable to ICRE. A checkmark (✓) indicates whether the estimator is based on sharp bounds (for our method and MCI).

Statistics for $Y_4 - Y_2$ and Y_4/Y_2	Estimates
Mean of ICDE	3.124 [0.364, 5.820]
Variance of ICDE under comonotonicity (Ours)	2.891 [0.613,7.143]
Variance of ICDE under monotonicity over f_Y (MCI)	3.104 [0.297, 8.610]
Skewness of ICDE under comonotonicity (Ours)	-0.762 [-6.468,2.089]
Skewness of ICDE under monotonicity over f_Y (MCI)	-3.373 [-28.287, 8.821]
Kurtosis of ICDE under comonotonicity (Ours)	8.737 [1.661,23.059]
Kurtosis of ICDE under monotonicity over f_Y (MCI)	18.765 [0.000, 203.983]
UB of Variance of ICDE (Ours) (✓)	30.508 [9.285,51.361]
LB of Variance of ICDE (Ours) (✓)	2.891 [0.613,7.143]
UB of Variance of ICDE (MCI) (✓)	29.503 [10.532, 61.318]
LB of Variance of ICDE (MCI)	0.000 [0.000,0.000]
UB of Variance of ICDE (COT)	30.965 [13.196, 53.903]
LB of Variance of ICDE (COT)	3.373 [0.824, 9.508]
UB of Skewness of ICDE (Ours)	7.487 [3.984,14.582]
LB of Skewness of ICDE (Ours)	-4.051 [-10.147,-0.0718]
UB of Skewness of ICDE (MCI)	34944.621 [0.000, 129424.512]
LB of Skewness of ICDE (MCI)	-79272.510 [-268151.400, 0.000]
UB of Skewness of ICDE (COT)	32.322 [1.246, 131.101]
LB of Skewness of ICDE (COT)	-33.339 [-143.721, -1.297]
UB of Kurtosis of ICDE (Ours)	259.648 [166.984,430.6644]
LB of Kurtosis of ICDE (Ours)	0.000 [0.000,0.000]
UB of Kurtosis of ICDE (MCI)	3.262×10^{6} [0.000, 2.226×10^{7}]
LB of Kurtosis of ICDE (MCI)	0.000 [0.000,0.000]
UB of Kurtosis of ICDE (COT)	665.578 [28.824, 3107.350]
LB of Kurtosis of ICDE (COT)	0.000 [0.000,0.000]
Mean of ICRE under comonotonicity	1.446 [1.166,1.879]
Variance of ICRE under comonotonicity	0.151 [0.010,0.441]
Skewness of ICRE under comonotonicity	0.776[-1.083,2.089]
Kurtosis of ICRE under comonotonicity	3.600[1.502,7.317]
UB of Mean of ICRE (Ours) (✓)	1.655 [1.244,2.263]
LB of Mean of ICRE (Ours) (✓)	1.446 [1.166,1.879]
UB of Mean of ICRE (COT)	1.676 [1.227, 2.249]
LB of Mean of ICRE (COT)	1.461 [1.127, 1.913]
UB of Variance of ICRE (Ours)	1.735 [0.244,4.392]
LB of Variance of ICRE (Ours)	0.000 [0.000,0.000]
UB of Variance of ICRE (COT)	1.883 [0.369, 4.472]
LB of Variance of ICRE (COT)	0.000 [0.000,0.000]
UB of Skewness of ICRE (Ours)	15346.36 [1207.602,38685.48]
LB of Skewness of ICRE (Ours)	-13089.57 [-33060.27,-1196.08]
UB of Skewness of ICRE (COT)	3.019×10^{12} [4.039×10^{10}, 9.351×10^{12}]
LB of Skewness of ICRE (COT)	-1.781×10^{10} [-8.175×10^{10}, 0.148]
UB of Kurtosis of ICRE (Ours)	83329.95 [1994.425,239344.1]
LB of Kurtosis of ICRE (Ours)	0.000 [0.000,0.000]
UB of Kurtosis of ICRE (COT)	1.026×10^{13} [4.790×10^{10}, 3.433×10^{13}]
LB of Kurtosis of ICRE (COT)	0.041 [0.001, 0.254]

Table 11: Estimates of the statistics for $Y_4 - Y_1$ and Y_4/Y_1. We report the means of the estimates along with their 95% CIs. The results of MCI were reported by [Kawakami and Tian, 2025]. MCI is not applicable to ICRE. A checkmark (✓) indicates whether the estimator is based on sharp bounds (for our method and MCI).

Statistics for $Y_4 - Y_1$ and Y_4/Y_1	Estimates
Mean of ICDE	6.551 [4.549, 9.034]
Variance of ICDE under comonotonicity (Ours)	1.700 [0.286,3.287]
Variance of ICDE under monotonicity over f_Y (MCI)	2.078 [0.000, 5.953]
Skewness of ICDE under comonotonicity (Ours)	0.004 [-2.323,1.860]
Skewness of ICDE under monotonicity over f_Y (MCI)	5.797 [-24.688, 65.521]
Kurtosis of ICDE under comonotonicity (Ours)	4.338 [1.155,11.044]
Kurtosis of ICDE under monotonicity over f_Y (MCI)	14.338 [0.000, 242.5153]
UB of Variance of ICDE (Ours) (✓)	24.353 [12.546,42.149]
LB of Variance of ICDE (Ours) (✓)	1.700 [0.286,3.287]
UB of Variance of ICDE (MCI) (✓)	25.160 [9.034, 42.910]
LB of Variance of ICDE (MCI)	0.000 [0.000,0.000]
UB of Variance of ICDE (COT)	25.949 [14.145, 43.069]
LB of Variance of ICDE (COT)	2.152 [0.486, 5.057]
UB of Skewness of ICDE (Ours)	58.819 [-3.814,375.223]
LB of Skewness of ICDE (Ours)	10.499 [6.123,17.745]
UB of Skewness of ICDE (MCI)	34135.710 [0.000, 113246.400]
LB of Skewness of ICDE (MCI)	-19251.891 [-97068.380, 0.000]
UB of Skewness of ICDE (COT)	36.701 [1.288, 140.127]
LB of Skewness of ICDE (COT)	-36.297 [-136.666, -3.202]
UB of Kurtosis of ICDE (Ours)	265.288 [164.641,436.681]
LB of Kurtosis of ICDE (Ours)	0.000 [0.000,0.000]
UB of Kurtosis of ICDE (MCI)	2.292×10^6 [0.000, 1.763×10^7]
LB of Kurtosis of ICDE (MCI)	0.000 [0.000,0.000]
UB of Kurtosis of ICDE (COT)	664.012 [43.686, 2992.521]
LB of Kurtosis of ICDE (COT)	0.024 [0.001, 0.121]
Mean of ICRE under comonotonicity (Ours)	2.554 [1.777,3.681]
Variance of ICRE under comonotonicity (Ours)	0.640 [0.099,1.902]
Skewness of ICRE under comonotonicity (Ours)	0.294[-0.777,1.377]
Kurtosis of ICRE under comonotonicity (Ours)	2.209[1.229,4.653]
UB of Mean of ICRE (Ours) (✓)	3.028 [2.087,4.164]
LB of Mean of ICRE (Ours) (✓)	2.554 [1.777,3.681]
UB of Mean of ICRE (COT)	2.974 [2.039, 3.994]
LB of Mean of ICRE (COT)	2.502 [1.772, 3.462]
UB of Variance of ICRE (Ours)	6.540 [2.534,12.705]
LB of Variance of ICRE (Ours)	0.019 [0.000,0.157]
UB of Variance of ICRE (COT)	6.454 [2.351, 11.704]
LB of Variance of ICRE (COT)	0.000 [0.000, 0.000]
UB of Skewness of ICRE (Ours)	94626 [17344.85,199385.5]
LB of Skewness of ICRE (Ours)	-87341.18 [-191424.8,-16479.25]
UB of Skewness of ICRE (COT)	1.201×10^{13} [1.645×10^{12}, 2.910×10^{13}]
LB of Skewness of ICRE (COT)	-2.376×10^{11} [-1.557×10^{12}, 0.051]
UB of Kurtosis of ICRE (Ours)	882022.3 [218388.7,2450092]
LB of Kurtosis of ICRE (Ours)	0.000 [0.000,0.000]
UB of Kurtosis of ICRE (COT)	5.752×10^{13} [4.749×10^{12}, 1.541×10^{14}]
LB of Kurtosis of ICRE (COT)	0.026 [0.001, 0.110]

Table 12: Estimates of the statistics for $Y_1 - Y_0$ and Y_1/Y_0. We report the means of the estimates along with their 95% CIs. MCI is not applicable to ICRE. A checkmark (✓) indicates whether the estimator is based on sharp bounds (for our method and MCI). The regularization parameter of Sinkhorn is set to 10^9.

Statistics for $Y_1 - Y_0$ and Y_1/Y_0	Estimates
Mean of ICDE	840.603 [-261.218,1782.534]
Variance of ICDE under comonotonicity (Ours)	4.299×10^6 [4.688×10^5 ,1.236×10^7]
Variance of ICDE (RP)	2.929×10^6 [2.993×10^5 ,1.049×10^7]
Skewness of ICDE under comonotonicity (Ours)	5.903 [-2.422 ,11.735]
Skewness of ICDE (RP)	4.983 [-7.903,13.555]
Kurtosis of ICDE under comonotonicity (Ours)	76.837[5.483,204.682]
Kurtosis of ICDE (RP)	63.362 [4.905,227.056]
UB of Variance of ICDE (Ours)(✓)	1.288×10^8 [1.007×10^8,1.680×10^8]
LB of Variance of ICDE (Ours) (✓)	4.299×10^6 [4.688×10^5 ,1.236×10^7]
UB of Variance of ICDE (MCI)(✓)	1.851×10^8 [1.446×10^8, 2.383×10^8]
LB of Variance of ICDE (MCI)	0.000 [0.000 ,0.000]
UB of Variance of ICDE (COT; Sinkhorn)	1.282×10^8 [9.665×10^7 , 1.760×10^8]
LB of Variance of ICDE (COT; Sinkhorn)	1.018×10^8 [8.113×10^7, 1.331×10^8]
UB of Skewness of ICDE (Ours)	1.923×10^3 [96.144,1.015×10^4]
LB of Skewness of ICDE (Ours)	-1.734×10^3 [-1.001×10^4,-43.391]
UB of Skewness of ICDE (MCI)	7.864×10^{16} [3.100×10^{16}, 1.676×10^{17}]
LB of Skewness of ICDE (MCI)	-5.458×10^{16}[-7.178×10^{16}, -4.009×10^{16}]
UB of Skewness of ICDE (COT; Sinkhorn)	-0.841 [-1.488, -0.144]
LB of Skewness of ICDE (COT; Sinkhorn)	-3.756 [-6.349, -0.913]
UB of Kurtosis of ICDE (Ours)	1.056×10^8 [6.861×10^6,4.086×10^8]
LB of Kurtosis of ICDE (Ours)	0.000 [0.000,0.000]
UB of Kurtosis of ICDE (MCI)	1.318×10^{23} [5.100×10^{22}, 2.844×10^{23}]
LB of Kurtosis of ICDE (MCI)	0.000 [0.000,0.000]
UB of Kurtosis of ICDE (COT; Sinkhorn)	20.641 [6.582, 30.991]
LB of Kurtosis of ICDE (COT; Sinkhorn)	0.473 [1.295×10^{-4}, 1.040]
Mean of ICRE under comonotonicity (Ours)	50.782 [5.957,139.223]
Mean of ICRE (RP)	1.353[1.054,1.680]
Variance of ICRE under comonotonicity (Ours)	4.903×10^3 [2.898×10^3 ,2.221×10^5]
Variance of ICRE (RP)	2.420 [0.041,9.0247]
Skewness of ICRE under comonotonicity (Ours)	4.497 [2.741,8.634]
Skewness of ICRE (RP)	9.415 [4.685,14.411]
Kurtosis of ICRE under comonotonicity (Ours)	29.678 [9.915,89.070]
Kurtosis of ICRE (RP)	100.260 [25.990,202.785]
UB of Mean of ICRE (Ours) (✓)	4.271×10^3 [3.532×10^3,5.044×10^3]
LB of Mean of ICRE (Ours) (✓)	50.782 [5.957,139.223]
UB of Mean of ICRE (COT; Sinkhorn)	1.496×10^3 [1.194×10^3, 1.863×10^3]
LB of Mean of ICRE (COT; Sinkhorn)	1.496×10^3 [1.194×10^3, 1.863×10^3]
UB of Variance of ICRE (Ours)	5.523×10^7 [3.488×10^7 ,8.883×10^7]
LB of Variance of ICRE (Ours)	4.903×10^3 [2.898×10^3 ,2.221×10^5]
UB of Variance of ICRE (COT; Sinkhorn)	4.338×10^7 [1.844×10^7, 8.878×10^7]
LB of Variance of ICRE (COT; Sinkhorn)	1.302×10^7 [1.007×10^7, 1.622×10^7]
UB of Skewness of ICRE (Ours)	2.180×10^{15} [9.447×10^{14},5.022×10^{15}]
LB of Skewness of ICRE (Ours)	-1.058×10^{15} [-1.986×10^{15},-5.929×10^{14}]
UB of Skewness of ICRE (COT; Sinkhorn)	29.881 [3.410, 88.351]
LB of Skewness of ICRE (COT; Sinkhorn)	-0.069 [-0.105, -0.047]
UB of Kurtosis of ICRE (Ours)	8.690×10^{19} [1.926×10^{19},2.453×10^{20}]
LB of Kurtosis of ICRE (Ours)	0.000 [0.000,0.000]
UB of Kurtosis of ICRE (COT; Sinkhorn)	4.807×10^2 [1.893×10, 1.481×10^3]
LB of Kurtosis of ICRE (COT; Sinkhorn)	3.541×10^{-3} [8.277×10^{-4}, 9.519×10^{-3}]

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our abstract and introduction accurately reflect the paper's contributions and scope.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of this paper in Section 8.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Each theorem, lemma, and corollary has a corresponding assumption.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper fully discloses all information necessary to reproduce the main experimental results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The code is publicly available at https://github.com/NA679723/NeurIPS2026_Joint_Expectation.git.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide all experimental details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: All estimates have confidence intervals.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All experiments were conducted on an Apple M4 system with 24 GB of memory.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: We comply with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: This work provides theoretical tools for identifying the joint expectation of potential outcomes; misuse risk is mitigated by providing bounds.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We cite the original paper that produced the code package or dataset.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.