Moments of Causal Effects

Yuta Kawakami[1] **Jin Tian**[1]

[1]Mohamed bin Zayed University of Artificial Intelligence, UAE

Abstract

The moments of random variables are fundamental statistical measures for characterizing the shape of a probability distribution, encompassing metrics such as mean, variance, skewness, and kurtosis. Additionally, the product moments, including covariance and correlation, reveal the relationships between multiple random variables. On the other hand, the primary focus of causal inference is the evaluation of causal effects, which are defined as the difference between two potential outcomes. While traditional causal effect assessment focuses on the average causal effect, this work provides definitions, identification theorems, and bounds for moments and product moments of causal effects to analyze their distribution and relationships. We conduct experiments to illustrate the estimation of the moments of causal effects from finite samples and demonstrate their practical application using a real-world medical dataset.

1 INTRODUCTION

The *moments* of random variables have been fundamental statistical measures since their introduction by Pafnuti Lvovich Chebyshev in the mid-nineteenth century [Mackey, 1980]. The m-th moment of a random variable Y is defined as $\mathbb{E}[Y^m]$, and the m-th central moment is defined as $\mathbb{E}[(Y - \mathbb{E}[Y])^m]$. These moments characterize the shape of a random variable's probability distribution, encompassing measures such as mean, variance, skewness, and kurtosis [Pearson, 1896, Joanes and Gill, 1998, Cramér, 1999, Doane and Seward, 2011, von Hippel, 2005, Westfall, 2014]. The (central) moments of random variables also play a fundamental role in various machine learning techniques [Bishop, 2006, Hastie et al., 2009, Murphy, 2022].

On the other hand, the primary focus of causal inference is the evaluation of causal effects $Y_1 - Y_0$, where Y_x denotes the potential outcome under treatment $X = x$, rather than a single random variable Y. Traditionally, to assess causal effects, researchers estimate the average causal effect (ACE), i.e., $\mathbb{E}[Y_1 - Y_0]$, which represents the first moment of causal effects [Neyman, 1923, Rubin, 1978, Holland, 1986, Balke and Pearl, 1997, Robins, 1999].

Recently, there has been increasing interest in exploring aspects of causal effects beyond their average, particularly in the distributional properties of causal effects [Ju and Geng, 2010, Wiedermann et al., 2022, Lin et al., 2023, Kennedy et al., 2023b, Post, 2023]. The shape of the distribution of causal effects uncovers causal effect heterogeneity, which is an actively researched topic in the field of statistics, causal inference, and machine learning [Athey and Imbens, 2016, Shalit et al., 2017, Athey and Imbens, 2019, Künzel et al., 2019, Wager and Athey, 2018, Singh et al., 2023, Kawakami et al., 2024b]. Causal effect heterogeneity refers to the variation in causal effects across individuals or subgroups within a population. Existing works on causal effect heterogeneity mainly examine the conditional average causal effects (CACE), i.e., $\mathbb{E}[Y_1 - Y_0|W = w]$, based on subjects' covariates W. However, CACE captures only the heterogeneity across subpopulations specified by observed covariates W, not the heterogeneity across individuals. In contrast, the shape of the distribution of causal effects reveals the heterogeneity of causal effects across individuals and provides complementary information to CACE.

Our objective is to address the following causal question:

(**Question 1**). "*How are causal effects distributed?* "

We approach this question by studying the moments of causal effects $\mathbb{E}\Big[(Y_1 - Y_0)^m\Big]$. These moments serve as measures that characterize the shape of the distribution of causal effects. Furthermore, we examine the central moments $\mathbb{E}\Big[\Big\{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\Big\}^m\Big]$. These moments quantify deviations from the ACE. They encompass

key statistical measures such as variance, standard deviation, skewness, and kurtosis, which are fundamental for characterizing the shape of the distribution of causal effects. While previous work has examined the second central moment (variance) of causal effects [Heckman et al., 1997, Hernan and Robins, 2024], this work provides a general analysis by studying arbitrary moments of causal effects.

Several studies [DiNardo et al., 1996, Robins and Rotnitzky, 2001, Rubin and van der Laan, 2006, Jung et al., 2021, Kennedy et al., 2023a, Kim et al., 2024] aim to estimate the probability density function (PDF) of Y_x. However, identifying the moments of causal effects requires the joint distribution of (Y_0, Y_1). The joint distribution of potential outcomes has been explored in the framework of probabilities of causation (PoC) [Pearl, 1999, Tian and Pearl, 2000, Li and Pearl, 2024]. We identify the (central) moments of causal effects by leveraging the recent identification results for variants of the PoC established by Kawakami et al. [2024a]. Additionally, we derive bounds for the (central) moments of causal effects under relaxed assumptions.

We further address the following causal question:

(**Question 2**). "*How are two causal effects related?* "

Researchers often consider more than two treatment options and compare multiple potential outcomes, $\{Y_1, Y_2, \dots, Y_R\}$, as discussed in [Bartholomew, 1959, Page, 1963, Imbens, 2000, Imai and van Dyk, 2004]. We then investigate the *product moments* of causal effects $\mathbb{E}\Big[(Y_i - Y_j)(Y_k - Y_h)\Big]$, as well as the central product moments (covariance and correlation) of causal effects, where $Y_i - Y_j$ represents the causal effect of changing $X = j$ to $X = i$, and $Y_k - Y_h$ represents the causal effect of changing $X = h$ to $X = k$. The product moments of causal effects reveal the association between two causal effects. When $\mathbb{E}[(Y_i - Y_j)(Y_k - Y_h)]$ is positive, subjects with larger $Y_i - Y_j$ tend to have larger $Y_k - Y_h$. When it is negative, subjects with larger $Y_i - Y_j$ tend to have smaller $Y_k - Y_h$. When it is zero, there is no linear relationship between $Y_k - Y_h$ and $Y_i - Y_j$. The product moments of causal effects provide additional insights beyond the ACE $\mathbb{E}[Y_i - Y_j]$ and $\mathbb{E}[Y_k - Y_h]$.

We establish identification theorems for the (central) product moments of causal effects and derive bounds for them under more relaxed assumptions. Finally, we conduct experiments estimating the (product) moments of causal effects from finite samples and demonstrate their practical application using a real-world medical dataset.

2 NOTATIONS AND BACKGROUNDS

We represent each variable with a capital letter (X) and its realized value with a lowercase letter (x). Let $\mathbb{I}(x)$ denote the indicator function, which takes the value 1 if x is true and 0 otherwise. We denote Ω_Y as the domain of Y, $\mathbb{E}[Y]$ as the expectation of Y, and $\mathbb{P}(Y < y)$ as the cumulative distribution function (CDF). Additionally, we denote the m-th Cartesian product of the domain Ω as Ω^m, i.e., $\Omega^m = \Omega \times \Omega \times \cdots \times \Omega$ (repeated m times).

Moments of random variables. Moments are measures related to the shape of the probability distribution of a random variable. For each $m \geq 1$, the m-th moment of a random variable Y is defined by the expectation of the m-th power of Y, i.e., $\mathbb{E}[Y^m]$, and the central moment of a random variable is defined by $C_m \triangleq \mathbb{E}[(Y - \mathbb{E}[Y])^m]$. Common statistics involving moments include mean $\mathbb{E}[Y]$, variance C_2, standard deviation $\sqrt{C_2}$, skewness $C_3/C_2^{3/2}$, and kurtosis C_4/C_2^2. They are essential statistics to capture the shape of the probability distribution of Y. The product moments of the random variables X and Y are given by $\mathbb{E}[XY]$, the covariance is $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$, and are used to define Pearson correlation coefficient $\frac{\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]}{\sqrt{\mathbb{E}[(X-\mathbb{E}[X])^2]}\sqrt{\mathbb{E}[(Y-\mathbb{E}[Y])^2]}}$.

Structural causal models. We use the language of Structural Causal Models (SCM) as our basic semantic and inferential framework [Pearl, 2009]. An SCM $\mathcal{M}$ is a tuple $\langle \boldsymbol{U}, \boldsymbol{V}, \mathcal{F}, \mathbb{P}_{\boldsymbol{U}} \rangle$, where $\boldsymbol{U}$ is a set of exogenous (unobserved) variables following a joint distribution $\mathbb{P}_{\boldsymbol{U}}$, and $\boldsymbol{V}$ is a set of endogenous (observable) variables whose values are determined by structural functions $\mathcal{F} = \{f_{V_i}\}_{V_i \in \boldsymbol{V}}$ such that $v_i := f_{V_i}(\mathbf{pa}_{V_i}, \boldsymbol{u}_{V_i})$ where $\mathbf{PA}_{V_i} \subseteq \boldsymbol{V}$ and $U^{V_i} \subseteq \boldsymbol{U}$. Each SCM $\mathcal{M}$ induces an observational distribution $\mathbb{P}_{\boldsymbol{V}}$ over $\boldsymbol{V}$. An atomic intervention of setting a set of endogenous variables $\boldsymbol{X}$ to constants $\boldsymbol{x}$, denoted by $do(\boldsymbol{x})$, replaces the original equations of $\boldsymbol{X}$ by $\boldsymbol{X} := \boldsymbol{x}$ and induces a *sub-model* $\mathcal{M}_{\boldsymbol{x}}$. We denote the potential outcome Y under intervention $do(x)$ by $Y_x(\boldsymbol{u})$, which is the solution of Y in the sub-model $\mathcal{M}_x$ given $\boldsymbol{U} = \boldsymbol{u}$.

Causal effects. Researchers usually consider the following SCM, denoted as $\mathcal{M}$:

$$Y := f_Y(X, U^Y), \;\; X := f_X(U^X), \tag{1}$$

where U^Y and U^X are latent exogenous variables. The individual causal effect (ICE) is defined as $\text{ICE}(\boldsymbol{u}) \triangleq Y_1(\boldsymbol{u}) - Y_0(\boldsymbol{u})$. The average causal effect (ACE) is defined as $\text{ACE} \triangleq \mathbb{E}[Y_1 - Y_0]$. Heckman et al. [1997] showed the identification of ICE under the rank invariance assumption stating that "for almost every subject whose potential outcomes are $(y_1, y_0) = (Y_1, Y_0)$, $\mathbb{P}(Y_0 < y_0) = \mathbb{P}(Y_1 < y_1)$ holds", which is a strong assumption. They identified the variance of causal effects by identifying ICE. [Hoshino and Takahata, 2020] studied the identification of the joint PDF of (Y_0, Y_1) under various parametric specifications, whereas this work considers a nonparametric setting.

Joint distribution of potential outcomes. Joint distributions of potential outcomes, fundamental to this work, have been employed in the framework of probabilities of cau-

sation (PoC) [Pearl, 1999, Tian and Pearl, 2000, Li et al., 2024]. PoC are a family of probabilities quantifying whether one event was the real cause of another. Recently, Kawakami et al. [2024a] defined the probability of necessity and sufficiency (PNS) for continuous treatment and outcome as $\mathbb{P}(Y_{x_0} < y \leq Y_{x_1})$, and showed that it is identified from $\mathbb{P}(X, Y)$ if there are no unmeasured confounders and the function $f_Y(x, U^Y)$ satisfies a monotonicity assumption:

Assumption 1 (Exogeneity)**.** *$Y_x \perp\!\!\!\perp X$ for all $x \in \Omega_X$.*

Assumption 2 (Monotonicity over f_Y)**.** *The function $f_Y(x, U^Y)$ is either (i) monotonic increasing on U^Y for all $x \in \Omega_X$ almost surely w.r.t. $\mathbb{P}_{U^Y}$, or (ii) monotonic decreasing on U^Y for all $x \in \Omega_X$ almost surely w.r.t. $\mathbb{P}_{U^Y}$.*

We will use the above assumptions to identify moments of causal effects in this paper.

3 MOMENTS OF CAUSAL EFFECTS

In this section, we study the (central) moments of causal effects $Y_1 - Y_0$ to address (**Question 1**).

3.1 DEFINITION OF THE MOMENTS OF CAUSAL EFFECTS

We define the moments of causal effects in the same manner as the moments of random variables.

Definition 1 (The moments of causal effects)**.** *For each $m \geq 1$, the m-th moment of causal effect $Y_1 - Y_0$ is defined as*

$$\mu^{(m)} \triangleq \mathbb{E}\Big[(Y_1 - Y_0)^m\Big]. \tag{2}$$

The m-th moment of the causal effect is defined as the expectation of the m-th power of the causal effect $Y_1 - Y_0$. The first moment, $\mu^{(1)} = \mathbb{E}[Y_1 - Y_0]$, is the ACE.

We present two examples to illustrate what the moments of causal effects specifically measure in simple SCMs.

Example 1. (Homogeneous ICE) Consider a simple linear SCM given by $Y = X + U^Y$ where $\mathbb{E}[U^Y] = 0$. In this model, the ICE is equal to 1 for every subject, meaning that the causal effect is homogeneous. Consequently, all of the m-th moments of the causal effect are also equal to 1. Note that $\mathbb{E}[Y_1^m - Y_0^m] = \mathbb{E}[(1 + U^Y)^m - (U^Y)^m]$ varies with m.

Example 2. (Heterogeneous ICE) Consider a linear SCM with an interaction term between X and U^Y, $Y = X(U^Y + 1) + 1$ where $\mathbb{E}[U^Y] = 0$. In this model, ICE $= Y_1 - Y_0 = U^Y + 1$, which varies across subjects, making the causal effect heterogeneous. The m-th moment of the causal effect is given by $\mathbb{E}[(U^Y + 1)^m]$. In comparison, $\mathbb{E}[Y_1^m - Y_0^m] = \mathbb{E}[(U^Y + 2)^m - 1]$.

The central moments of causal effects are defined as the moments of causal effects measured relative to their mean.

Definition 2 (The central moment of causal effects)**.** *For each $m \geq 1$, the m-th central moment of causal effect $Y_1 - Y_0$ is defined as*

$$\overline{\mu}^{(m)} \triangleq \mathbb{E}\Big[\Big\{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\Big\}^m\Big]. \tag{3}$$

The first central moment of causal effects is always 0 since we have $\overline{\mu}^{(1)} = \mathbb{E}[(Y_1 - Y_0)] - \mathbb{E}[(\mathbb{E}[Y_1] - \mathbb{E}[Y_0])] = 0$. When $m \geq 2$, $\overline{\mu}^{(m)}$ is not equal to $\mathbb{E}[(Y_1 - \mathbb{E}[Y_1])^m] - \mathbb{E}[(Y_0 - \mathbb{E}[Y_0])^m]$ as discussed in [Wiedermann et al., 2022].

We revisit Examples 1 and 2 in Section 3.1 to illustrate the central moments of causal effects.

Example 1 (continued). (Homogeneous ICE) In SCM given by $Y = X + U^Y$ where $\mathbb{E}[U^Y] = 0$, the m-th central moment is equal to 0 for any $m \geq 1$, indicating that the causal effect is homogeneous.

Example 2 (continued). (Heterogeneous ICE) In SCM $Y = X(U^Y + 1) + 1$ where $\mathbb{E}[U^Y] = 0$, the m-th central moment is given by $\mathbb{E}[(U^Y)^m]$ for all $m \geq 1$. The central moments of causal effects correspond to the (central) moments of the random variable U^Y.

The higher order of moments of causal effects may provide useful information on the distribution. The central moments of causal effects can be used to compute important and well-known statistics such as the variance $\overline{\mu}^{(2)}$, standard deviation $\sqrt{\overline{\mu}^{(2)}}$, skewness $\overline{\mu}^{(3)}/{\overline{\mu}^{(2)}}^{3/2}$, and kurtosis $\overline{\mu}^{(4)}/{\overline{\mu}^{(2)}}^2$ of the causal effects. Variance and standard deviation quantify the dispersion of a distribution. If the variance of causal effects is large, the causal effects may deviate significantly from ACE for some subjects. When the variance of causal effects is small, ICE is close to ACE for all subjects. Skewness is a measure of the asymmetry of a probability distribution. If the causal effect is positively skewed, the right tail of the distribution of the causal effect is longer. If the causal effect is negatively skewed, the left tail of the distribution is longer. Kurtosis is a measure of the tailedness or peakedness of a distribution. High kurtosis values indicate the presence of outliers in causal effects [Westfall, 2014].

3.2 IDENTIFICATION OF THE MOMENTS OF CAUSAL EFFECTS

Under the exogeneity assumption (Assumption 1), the first moment of causal effects is identifiable as $\mathbb{E}[Y|X = 1] - \mathbb{E}[Y|X = 0]$ [Holland, 1986]. In this section, we discuss the identification of the higher moments of causal effects.

When $m \geq 2$, $\mu^{(m)}$ is not equal to $\mathbb{E}[Y_1^m - Y_0^m]$, as discussed in [Hernan and Robins, 2024, Kuroki and Tezuka, 2024]. Under Assumption 1, $\mathbb{E}[Y_1^m - Y_0^m]$ is identifiable as

$\mathbb{E}[Y^m|X=1] - \mathbb{E}[Y^m|X=0]$. However, $\mu^{(m)}$ remains unidentifiable. For example, the second moment of the causal effect, $\mu^{(2)}$, is given by $\mathbb{E}[Y_1^2] - 2\mathbb{E}[Y_0Y_1] + \mathbb{E}[Y_0^2]$, where the term $\mathbb{E}[Y_0Y_1]$ is not identifiable.

To prepare the identification of the moments of causal effect, we first decompose $(Y_1 - Y_0)^m$ into two parts as follows:

Lemma 1. *Under SCM $\mathcal{M}$, we have*

$$\begin{aligned}(Y_1 - Y_0)^m &= (Y_1 - Y_0)^m \mathbb{I}(Y_1 > Y_0) \\ &\quad + (-1)^m (Y_0 - Y_1)^m \mathbb{I}(Y_0 > Y_1) \\ &= \int_{{\Omega_Y}^m} \mathbb{I}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, \\ &\qquad Y_0 < y_m \leq Y_1) dy_1 \ldots dy_m \\ &\quad + (-1)^m \int_{{\Omega_Y}^m} \mathbb{I}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, \\ &\qquad Y_1 < y_m \leq Y_0) dy_1 \ldots dy_m. \quad (4)\end{aligned}$$

The first part corresponds to subjects with a positive ICE, where $Y_1 - Y_0 > 0$, and the second part corresponds to subjects with a negative ICE, where $Y_1 - Y_0 < 0$.

We make the following assumption:

Assumption 3 (Finiteness of integrals)**.** *Under SCM $\mathcal{M}$, for $m \geq 1$, $\mu^{(m)} < \infty$ and $\int_{{\Omega_Y}^m} \mathbb{P}(Y_i < y_1 \leq Y_j, Y_i < y_2 \leq Y_j, \ldots, Y_i < y_m \leq Y_j) dy_1 \ldots dy_m < \infty$ hold for $(i,j) = \{(0,1),(1,0)\}$.*

Under SCM $\mathcal{M}$ and Assumption 3, taking the expectation on both sides of Eq. (4), we have

$$\begin{aligned}\mu^{(m)} &= \int_{{\Omega_Y}^m} \mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, \\ &\qquad Y_0 < y_m \leq Y_1) dy_1 \ldots dy_m \\ &\quad + (-1)^m \int_{{\Omega_Y}^m} \mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, \\ &\qquad Y_1 < y_m \leq Y_0) dy_1 \ldots dy_m. \end{aligned} \quad (5)$$

The identification of the moments of causal effects then reduces to the identification of $\mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, Y_0 < y_m \leq Y_1)$ and $\mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, Y_1 < y_m \leq Y_0)$. The identification of this type of joint distributions of potential outcomes was discussed in [Kawakami et al., 2024a], based on which we obtain the following result:

Theorem 1 (Identification of the moments of causal effect)**.** *Under SCM $\mathcal{M}$ and Assumptions 1, 2, and 3, the m-th moment of causal effect $Y_1 - Y_0$ is identifiable by $\mu^{(m)} = \sigma^{(m)}$, where*

$$\begin{aligned}\sigma^{(m)} &= \int_{{\Omega_Y}^m} \max\Big\{ \min_{p=1,\ldots,m} \{\mathbb{P}(Y < y_p|X=0)\} \\ &\quad - \max_{p=1,\ldots,m} \{\mathbb{P}(Y < y_p|X=1)\}, 0 \Big\} dy_1 \ldots dy_m \\ &\quad + (-1)^m \int_{{\Omega_Y}^m} \max\Big\{ \min_{p=1,\ldots,m} \{\mathbb{P}(Y < y_p|X=1)\} \\ &\quad - \max_{p=1,\ldots,m} \{\mathbb{P}(Y < y_p|X=0)\}, 0 \Big\} dy_1 \ldots dy_m. \quad (6)\end{aligned}$$

Theorem 1 says that the moments of causal effects can be expressed in terms of conditional CDFs. For $m=1$, Eq. (6) reduces to ACE$= \mathbb{E}[Y_1 - Y_0] = \int_{\Omega_Y} \{\mathbb{P}(Y < y_1|X=0) - \mathbb{P}(Y < y_1|X=1)\} dy_1$ [Ju and Geng, 2010], which does not require Assumption 2 to hold.

For $m=2$, the second moment of causal effects (variance) $\mu^{(2)}$ is given by $\int_{\Omega_Y} \int_{\Omega_Y} \max\{\min\{\mathbb{P}(Y < y_1|X=0), \mathbb{P}(Y < y_2|X=0)\} - \max\{\mathbb{P}(Y < y_1|X=1), \mathbb{P}(Y < y_1|X=1)\}, 0\} dy_1 dy_2 + \int_{\Omega_Y} \int_{\Omega_Y} \max\{\min\{\mathbb{P}(Y < y_1|X=1), \mathbb{P}(Y < y_2|X=1)\} - \max\{\mathbb{P}(Y < y_1|X=0), \mathbb{P}(Y < y_2|X=0), 0\} dy_1 dy_2$.

3.3 BOUNDING THE MOMENTS OF CAUSAL EFFECTS

The monotonicity Assumption 2 may sometimes be considered implausible by researchers. Therefore, we derive bounds for the moments of causal effects that do not rely on Assumption 2.

We first provide bounds of the joint distribution of the potential outcomes $\mathbb{P}(Y_i < y_1 \leq Y_j, Y_i < y_2 \leq Y_j, \ldots, Y_i < y_m \leq Y_j)$ using Fréchet inequalities [Fréchet, 1935, 1960].

Lemma 2. *Under SCM $\mathcal{M}$ and Assumptions 1 and 3, we have $l(y_1, \ldots, y_m; i, j) \leq \mathbb{P}(Y_j < y_1 \leq Y_i, Y_j < y_2 \leq Y_i, \ldots, Y_j < y_m \leq Y_i) \leq u(y_1, \ldots, y_m; i, j)$, where*

$$\begin{aligned}l(y_1, \ldots, y_m; i, j) &= \max\Big\{ \sum_{p=1,\ldots,m} \mathbb{P}(Y < y_p|X=j) \\ &\quad - \sum_{p=1,\ldots,m} \mathbb{P}(Y < y_p|X=i) - m + 1, 0 \Big\}, \end{aligned} \quad (7)$$

$$\begin{aligned}u(y_1, \ldots, y_m; i, j) &= \min\Big\{ \min_{p=1,\ldots,m} \{\mathbb{P}(Y < y_p|X=j)\}, \\ &\quad 1 - \max_{p=1,\ldots,m} \{\mathbb{P}(Y < y_p|X=i)\} \Big\} \end{aligned} \quad (8)$$

for $(i,j) \in \{(1,0),(0,1)\}$ and any $y_1, \ldots, y_m \in \Omega_Y$.

Then, we have the following theorem.

Theorem 2 (Bounds of the moments of causal effect)**.** *Under SCM $\mathcal{M}$ and Assumptions 1 and 3, we have $\sigma_L^{(m)} \leq \mu^{(m)} \leq \sigma_U^{(m)}$, where*

(A). When m is an even number,

$$\sigma_L^{(m)} = \int_{{\Omega_Y}^m} l(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m$$

$$+ \int_{{\Omega_Y}^m} l(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m, \quad (9)$$

$$\sigma_U^{(m)} = \int_{{\Omega_Y}^m} u(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m + \int_{{\Omega_Y}^m} u(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m. \quad (10)$$

(B). When m is an odd number,

$$\sigma_L^{(m)} = \int_{{\Omega_Y}^m} l(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m - \int_{{\Omega_Y}^m} u(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m, \quad (11)$$

$$\sigma_U^{(m)} = \int_{{\Omega_Y}^m} u(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m - \int_{{\Omega_Y}^m} l(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m. \quad (12)$$

If $\sigma_U^{(m)} = \infty$ and $\sigma_L^{(m)} = -\infty$, then $\mu^{(m)}$ is unbounded.

The upper bound of the Fréchet inequalities is always sharp for all $m \geq 1$ [Nelsen, 2007]; thus, the function $u(y_1, \ldots, y_m; i, j)$ in Lemma 2 is sharp for all $m \geq 1$. In contrast, the lower bound of the Fréchet inequalities is not always sharp except when $m = 1$; hence, the function $l(y_1, \ldots, y_m; i, j)$ in Lemma 2 is not sharp. As a result, only the upper bounds of the moments of causal effects are sharp when m is even. In all other cases, our bounds of the moments of causal effects are not sharp.

Remark. We present a similar identification theorem and bounds on the central moments of causal effects in Appendix B. Additionally, the skewness and kurtosis of causal effects are also bounded, as shown in Appendix B.

4 PRODUCT MOMENTS OF CAUSAL EFFECTS

In this section, we study the (central) product moments of causal effects to address (**Question 2**).

Let $\Omega_X = \{1, \ldots, R\}$. The causal effect of changing $X = j$ to $X = i$ is given by $Y_i - Y_j$ and the causal effect of changing $X = k$ to $X = h$ is given by $Y_h - Y_k$. We study the association of the two causal effects $Y_i - Y_j$ and $Y_h - Y_k$.

4.1 DEFINITION OF THE PRODUCT MOMENT OF CAUSAL EFFECTS

We define the product moment of two causal effects analogously to the product moment of two random variables.

Definition 3 (The product moment of causal effects)**.** *The product moment of causal effects is defined by*

$$\rho_{i,j;k,h} \triangleq \mathbb{E}\Big[(Y_i - Y_j)(Y_k - Y_h)\Big]. \quad (13)$$

We present three examples to illustrate the product moments of causal effects in simple SCMs.

Example 1 (continued). (Homogeneous ICE) In SCM given by $Y = X + U^Y$ where $\mathbb{E}[U^Y] = 0$, the product moment of $Y_1 - Y_0$ and $Y_0 - Y_{-1}$ is equal to 1.

Example 2 (continued). (Heterogeneous ICE) In SCM $Y = X(U^Y + 1) + 1$ where $\mathbb{E}[U^Y] = 0$, the product moment of $Y_1 - Y_0$ and $Y_0 - Y_{-1}$ is equal to $\mathbb{E}[(U^Y + 1)^2] > 0$.

Example 3. (Heterogeneous and nonlinear ICE) We consider a nonlinear SCM with an interaction term between X^2 and U^Y: $Y = X^2(U^Y + 1) + 1$ where $\mathbb{E}[U^Y] = 0$. We have $Y_1 - Y_0 = U^Y + 1$ and $Y_0 - Y_{-1} = -(U^Y + 1)$ and are heterogeneous. The product moment of $Y_1 - Y_0$ and $Y_0 - Y_{-1}$ is equal to $\mathbb{E}[-(U^Y + 1)^2] < 0$.

We examine the covariance and correlation of two causal effects.

Definition 4 (Covariance of causal effects)**.** *We define the covariance (central product moment of causal effects) as*

$$\overline{\rho}_{i,j;k,h} \triangleq \mathbb{E}\Big[\Big\{(Y_i - Y_j) - (\mathbb{E}[Y_i] - \mathbb{E}[Y_j])\Big\} \times \Big\{(Y_k - Y_h) - (\mathbb{E}[Y_k] - \mathbb{E}[Y_h])\Big\}\Big]. \quad (14)$$

Definition 5 (Correlation of causal effects)**.** *We define the correlation of causal effects as*

$$\overline{\tau}_{i,j;k,h} = \overline{\rho}_{i,j;k,h} \Big/ \Bigg\{\sqrt{\mathbb{E}\Big[\Big\{(Y_i - Y_j) - (\mathbb{E}[Y_i] - \mathbb{E}[Y_j])\Big\}^2\Big]} \times \sqrt{\mathbb{E}\Big[\Big\{(Y_k - Y_h) - (\mathbb{E}[Y_k] - \mathbb{E}[Y_h])\Big\}^2\Big]}\Bigg\}. \quad (15)$$

Correlation is a measure of association between two variables [Pearson, 1905]. Similarly, the correlation of causal effects quantifies the association between two causal effects.

For instance, when comparing three treatments $X = 0, 1, 2$, researchers often evaluate their respective averages $\mathbb{E}[Y_0], \mathbb{E}[Y_1]$, and $\mathbb{E}[Y_2]$. When $\mathbb{E}[Y_1 - Y_0] > 0$ and $\mathbb{E}[Y_2 - Y_1] > 0$, it is concluded that, on average, both changes, from $X = 0$ to $X = 1$ and from $X = 1$ to $X = 2$, have positive effects. The correlation of causal effects provides more detailed insights. When the correlation between $Y_1 - Y_0$ and $Y_2 - Y_1$ is negative, patients with larger causal effects $Y_1 - Y_0$ than the average tend to have smaller causal effects $Y_2 - Y_1$ than the average. Conversely, when patients have smaller causal effects $Y_1 - Y_0$ than the average, they tend to have larger causal effects $Y_2 - Y_1$ than the average.

We present three examples to illustrate the covariance and correlation of causal effects.

Example 1 (continued). (Homogeneous ICE) In SCM given by $Y = X + U^Y$ where $\mathbb{E}[U^Y] = 0$, the covariance of $Y_1 - Y_0$ and $Y_0 - Y_{-1}$ is equal to 0.

Example 2 (continued). (Heterogeneous ICE) In SCM $Y = X(U^Y+1)+1$ where $\mathbb{E}[U^Y] = 0$, the covariance of Y_1-Y_0 and $Y_0 - Y_{-1}$ is equal to $\mathbb{E}[(U^Y)^2] > 0$, the correlation of $Y_1 - Y_0$ and $Y_0 - Y_{-1}$ is 1, and they have a positive correlation.

Example 3 (continued). (Heterogeneous and nonlinear ICE) In SCM $Y = X^2(U^Y + 1) + 1$ where $\mathbb{E}[U^Y] = 0$, the covariance of $Y_1 - Y_0$ and $Y_0 - Y_{-1}$ is equal to $\mathbb{E}[-(U^Y)^2] < 0$, the correlation of $Y_1 - Y_0$ and $Y_0 - Y_{-1}$ is -1, and they have a negative correlation.

4.2 IDENTIFICATION OF THE PRODUCT MOMENT OF CAUSAL EFFECTS

To prepare the discussion on the identification of the product moment of causal effects, we decompose the product of causal effects $(Y_i - Y_j)$ and $(Y_k - Y_h)$ into four parts.

Lemma 3. *Under SCM $\mathcal{M}$, we have*

$$\begin{aligned}
&(Y_i - Y_j)(Y_k - Y_h)\\
&= (Y_i - Y_j)(Y_k - Y_h)\mathbb{I}(Y_i > Y_j, Y_k > Y_h)\\
&- (Y_j - Y_i)(Y_k - Y_h)\mathbb{I}(Y_j > Y_i, Y_k > Y_h)\\
&- (Y_i - Y_j)(Y_h - Y_k)\mathbb{I}(Y_i > Y_j, Y_h > Y_k)\\
&+ (Y_j - Y_i)(Y_h - Y_k)\mathbb{I}(Y_j > Y_i, Y_h > Y_k)\\
&= \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{I}(Y_j < y_1 \le Y_i, Y_h < y_2 \le Y_k)dy_1dy_2\\
&- \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{I}(Y_i < y_1 \le Y_j, Y_h < y_2 \le Y_k)dy_1dy_2\\
&- \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{I}(Y_j < y_1 \le Y_i, Y_k < y_2 \le Y_h)dy_1dy_2\\
&+ \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{I}(Y_i < y_1 \le Y_j, Y_k < y_2 \le Y_h)dy_1dy_2.
\end{aligned} \tag{16}$$

The above decomposition consists of four parts based on the signs of ICE. We make the following assumption:

Assumption 4 (Finiteness of integrals)**.** *Under SCM $\mathcal{M}$, $\rho_{i,j;k,h} < \infty$ and $\int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j < y_1 \le Y_i, Y_h < y_2 \le Y_k)dy_1dy_2 < \infty$ hold for any $i, j, k, h \in \{1, \dots, R\}$.*

Under SCM $\mathcal{M}$ and Assumption 4, taking the expectation on both sides of Eq. (16), we have

$$\begin{aligned}
&\rho_{i,j;k,h}\\
&= \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j < y_1 \le Y_i, Y_h < y_2 \le Y_k)dy_1dy_2\\
&- \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_i < y_1 \le Y_j, Y_h < y_2 \le Y_k)dy_1dy_2\\
&- \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j < y_1 \le Y_i, Y_k < y_2 \le Y_h)dy_1dy_2\\
&+ \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_i < y_1 \le Y_j, Y_k < y_2 \le Y_h)dy_1dy_2.
\end{aligned} \tag{17}$$

The identification of joint distributions of potential outcomes in the form of $\mathbb{P}(Y_j < y_1 \le Y_i, Y_h < y_2 \le Y_k)$ was discussed in [Kawakami et al., 2024a], based on which we obtain the following result:

Theorem 3 (Identification of the product moments of causal effects)**.** *Under SCM $\mathcal{M}$ and Assumptions 1, 2, and 4, the product moment of $(Y_i - Y_j)$ and $(Y_k - Y_h)$ is identifiable by $\rho_{i,j;k,h} = \sigma(i, j; k, h)$, where*

$$\begin{aligned}
&\sigma(i, j; k, h)\\
&= \int_{{\Omega_Y}^2} \max\Big\{\min\{\mathbb{P}(Y < y_1|X = j), \mathbb{P}(Y < y_2|X = h)\}\\
&\quad - \max\{\mathbb{P}(Y < y_1|X = i), \mathbb{P}(Y < y_2|X = k)\}, 0\Big\}dy_1dy_2\\
&- \int_{{\Omega_Y}^2} \max\Big\{\min\{\mathbb{P}(Y < y_1|X = i), \mathbb{P}(Y < y_2|X = h)\}\\
&\quad - \max\{\mathbb{P}(Y < y_1|X = j), \mathbb{P}(Y < y_2|X = k)\}, 0\Big\}dy_1dy_2\\
&- \int_{{\Omega_Y}^2} \max\Big\{\min\{\mathbb{P}(Y < y_1|X = j), \mathbb{P}(Y < y_2|X = k)\}\\
&\quad - \max\{\mathbb{P}(Y < y_1|X = i), \mathbb{P}(Y < y_2|X = h)\}, 0\Big\}dy_1dy_2\\
&+ \int_{{\Omega_Y}^2} \max\Big\{\min\{\mathbb{P}(Y < y_1|X = i), \mathbb{P}(Y < y_2|X = k)\}\\
&\quad - \max\{\mathbb{P}(Y < y_1|X = j), \mathbb{P}(Y < y_2|X = h)\}, 0\Big\}dy_1dy_2.
\end{aligned} \tag{18}$$

4.3 BOUNDING THE PRODUCT MOMENT OF CAUSAL EFFECTS

The monotonicity Assumption 2 may sometimes be considered implausible by researchers. Therefore, we derive bounds for the product moment of causal effects that do not rely on Assumption 2.

We first derive bounds for the joint distribution of potential outcomes $\mathbb{P}(Y_j < y_1 \le Y_i, Y_h < y_2 \le Y_k)$.

Lemma 4. *Under SCM $\mathcal{M}$ and Assumptions 1 and 4, for any $i, j, k, h \in \{1, \dots, R\}$, $y_1, y_2 \in \Omega_Y$ we have $l(y_1, y_2; i, j, k, h) \le \mathbb{P}(Y_j < y_1 \le Y_i, Y_h < y_2 \le Y_k) \le u(y_1, y_2; i, j, k, h)$, where*

$$\begin{aligned}
&l(y_1, y_2; i, j, k, h) =\\
&\max\Big\{\mathbb{P}(Y < y_1|X = j) - \mathbb{P}(Y < y_1|X = i)\\
&+ \mathbb{P}(Y < y_2|X = h) - \mathbb{P}(Y < y_2|X = k) - 1, 0\Big\},
\end{aligned} \tag{19}$$

$$u(y_1, y_2; i, j, k, h) =$$

$$\min\Big\{\min\{\mathbb{P}(Y<y_1|X=j),\mathbb{P}(Y<y_2|X=h)\},$$
$$1-\max\{\mathbb{P}(Y<y_1|X=i),\mathbb{P}(Y<y_2|X=k)\}\Big\} \quad (20)$$

Then, we have the following theorem.

Theorem 4 (Bounds of the product moments of causal effects)**.** *Under SCM $\mathcal{M}$ and Assumptions 1 and 4, for any $i,j,k,h\in\{1,\dots,R\}$, we have $\sigma_L(i,j;k,h)\le\rho_{i,j;k,h}\le\sigma_U(i,j;k,h)$, where*

$$\begin{aligned}\sigma_L(i,j;k,h)=&\int_{\Omega_Y{}^2}l(y_1,y_2;i,j,k,h)dy_1dy_2\\&-\int_{\Omega_Y{}^2}u(y_1,y_2;j,i,k,h)dy_1dy_2\\&-\int_{\Omega_Y{}^2}u(y_1,y_2;i,j,h,k)dy_1dy_2\\&+\int_{\Omega_Y{}^2}l(y_1,y_2;j,i,h,k)dy_1dy_2, \quad (21)\\ \sigma_U(i,j;k,h)=&\int_{\Omega_Y{}^2}u(y_1,y_2;i,j,k,h)dy_1dy_2\\&-\int_{\Omega_Y{}^2}l(y_1,y_2;j,i,k,h)dy_1dy_2\\&-\int_{\Omega_Y{}^2}l(y_1,y_2;i,j,h,k)dy_1dy_2\\&+\int_{\Omega_Y{}^2}u(y_1,y_2;j,i,h,k)dy_1dy_2. \quad (22)\end{aligned}$$

If $\sigma_U=\infty$ and $\sigma_L=-\infty$, then $\rho_{i,j;k,h}$ is unbounded. The above bounds for product moments are not sharp.

Remark. We present a similar identification theorem and bounds for the central product moment (covariance) of causal effects in Appendix C. Additionally, the correlation of causal effects is also bounded, as shown in Appendix C.

5 NUMERICAL EXPERIMENTS

In this section, we perform experiments to illustrate the finite-sample performance of the estimators of the moments of causal effects.

Estimation. The family of moments of causal effects $\sigma^{(m)},\sigma_L^{(m)},\sigma_U^{(m)},\sigma(i,j;k,h),\sigma_L(i,j;k,h),\sigma_U(i,j;k,h)$ in Theorems 1-4 (and the central moments of causal effects $\bar\sigma^{(m)},\bar\sigma_L^{(m)},\bar\sigma_U^{(m)},\bar\sigma(i,j,k,h),\bar\sigma_U(i,j,k,h),\bar\sigma_L(i,j,k,h)$ in Appendices B and C) are estimable by plugging in the empirical CDFs and expectations [Vaart, 1998] and calculating the integrals using the Monte Carlo integration method [Press, 2007]. Let N be the sample size of the dataset, and let N_1, N_2, N_3, and N_4 be the numbers of points for Monte Carlo integration on y_1, y_2, y_3, and y_4. We assume that the domains of Y and $Y-\mathbb{E}[Y|X=x]$ for any $x\in\Omega_X$ are bounded by $[a,b]$, which is required for Monte Carlo integration. The details of all estimators are shown in Appendix E. They are all consistent estimators as discussed in Appendix E.

Simulation for the moments of causal effects. We assume the following SCM (A):

$$Y:=-(X+1)U\mathbb{I}(XU\ge0), \quad (23)$$
$$X\sim\text{Bern}(0.8),U\sim\text{Unif}(-1,1), \quad (24)$$

where $\text{Bern}(p)$ is a Bernoulli distribution with probability p, and $\text{Unif}(-1,1)$ is a uniform distribution over $[-1,1]$. This setting satisfies Assumptions 1 and 2. We simulate 1000 times with the sample size $N=20,100,1000$, respectively. We let N_1, N_2, N_3, and N_4 all be 1000.

Results. We present the estimates obtained using our proposed methods in Table 1. All means of the estimators are close to the ground truth for $N=1000$. All ground truth values lie within the computed bounds. However, estimators for small sample sizes have large 95 % CIs, especially for high-order moments.

Simulation for the product moments of causal effects. We assume the following SCM (B):

$$Y:=X^2U,U\sim\text{Unif}(0,1), \quad (25)$$

where X takes values in $\{-1,0,1\}$ with the probabilities $\mathbb{P}(X=-1)=\mathbb{P}(X=0)=\mathbb{P}(X=1)=1/3$. The domain of Y is bounded within $[0,1]$. This setting satisfies Assumptions 1 and 2. We simulate 1000 times with the sample size $N=20,100,1000$, respectively. We let N_1 and N_2 both be 1000.

Results. We present the estimates for $\mathbb{E}[(Y_1-Y_0)(Y_0-Y_{-1})]$ in Table 2. All means of the estimators are close to the ground truth. The ground truth value lies within the computed bounds. However, estimators for small sample sizes have large 95 % CIs.

Overall, the results show that, as the sample size increases, the estimates are close to the ground truths. Figures 1 and 2 in Appendix F present the plots of the estimates obtained from the numerical experiments.

6 APPLICATION TO REAL-WORLD

We present an application to a real-world medical dataset.

Dataset. We take up a dataset "Cholesterol Reduction" [Westfall et al., 2011] (`https://search.r-project.org/CRAN/refmans/multcomp/html/cholesterol.html`). This clinical study was conducted to assess the effect of three formulations of the same drug (X) on cholesterol reduction (Y). The three formulations consisted 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

Table 1: Results of numerical experiments for SCM (A). We present the estimates of the second moments $\sigma^{(2)}$, third moments $\sigma^{(3)}$, and fourth moments $\sigma^{(4)}$ of causal effects along with their respective upper and lower bounds. Additionally, we report the means of each estimator accompanied by their 95% confidence intervals.

Estimators	$N = 20$	$N = 100$	$N = 1000$	Ground Truth
$\sigma^{(2)}$	$0.405([0.138, 0.841])$	$0.373([0.215, 0.659])$	$0.335([0.289, 0.418])$	0.333
$\sigma_U^{(2)}$	$1.548([0.804, 2.62])$	$1.582([1.127, 2.030])$	$1.647([1.485, 1.769])$	-
$\sigma_L^{(2)}$	$0.108([0.000, 0.679])$	$0.005([0.000, 0.018])$	$0.000([0.000, 0.000])$	-
$\sigma^{(3)}$	$-0.572([-1.679, 0.093])$	$-0.293([-0.750, -0.065])$	$-0.245([-0.305, -0.186])$	-0.250
$\sigma_U^{(3)}$	$0.087([-0.107, 0.292])$	$0.120([0.037, 0.234])$	$0.126([0.074, 0.177])$	-
$\sigma_L^{(3)}$	$-2.479([-5.381, -0.612])$	$-3.175([-3.999, -1.978])$	$-3.412([-3.877, -3.113])$	-
$\sigma^{(4)}$	$0.963([0.066, 6.922])$	$0.194([0.061, 0.384])$	$0.205([0.112, 0.283])$	0.200
$\sigma_U^{(4)}$	$6.837([1.157, 11.497])$	$7.712([4.909, 10.848])$	$8.093([7.214, 8.878])$	-
$\sigma_L^{(4)}$	$0.008([0.000, 0.085])$	$0.000([0.000, 0.000])$	$0.000([0.000, 0.000])$	-

Table 2: Results of numerical experiments for SCM (B). We present the estimates of the product moments of causal effects $\sigma(1, 0; 0, -1)$ along with their respective upper and lower bounds. Additionally, we report the means of each estimator accompanied by their 95% confidence intervals.

Estimators	$N = 20$	$N = 100$	$N = 1000$	Ground Truth
$\sigma(1, 0; 0, -1)$	$-0.300([-0.437, -0.131])$	$-0.323([-0.420, -0.239])$	$-0.327([-0.419, -0.260])$	-0.333
$\sigma_U(1, 0; 0, -1)$	$-0.154([-0.521, 0.000])$	$-0.105([-0.217, -0.029])$	$-0.168([-0.222, -0.112])$	-
$\sigma_L(1, 0; 0, -1)$	$-0.352([-0.559, -0.100])$	$-0.390([-0.583, -0.278])$	$-0.338([-0.409, -0.260])$	-

dataset has a sample size of 10 for each treatment group. The purpose of the study was to determine which of the formulations is efficacious. Previous studies examined multiple comparisons, specifically whether the conditions $\mathbb{E}[Y_4] > \mathbb{E}[Y_2]$, $\mathbb{E}[Y_2] > \mathbb{E}[Y_1]$, and $\mathbb{E}[Y_4] > \mathbb{E}[Y_1]$ simultaneously hold or not [Westfall et al., 2011]. The domain of Y is bounded within $[5, 19]$.

We assume the exogeneity assumption holds, as Westfall et al. [2011] did not report any potential confounding factors. We assume that the monotonicity assumption holds, meaning that increasing the formulation from 1 time to 2 times always leads to greater cholesterol reduction for all subjects, and similarly, increasing it from 2 times to 4 times consistently results in further cholesterol reduction for all subjects. We used the estimators described in Section 5. We conduct the bootstrapping [Efron, 1979] to provide the means and 95% confidence intervals (CI) for each estimator.

Results on the moments of causal effects. First, we study the causal effects of changing the formulation from $X = 1$ to $X = 2$, i.e., $Y_2 - Y_1$. The results are:

Mean: 3.432 (95%CI: $[0.914, 6.104]$).
Variance: 3.072 (95%CI: $[0.297, 8.610]$),
Standard deviation: 1.753 (95%CI: $[0.545, 2.934]$).
Skewness: 21.027 (95%CI: $[-6.747, 34.504]$),
Kurtosis: 21.312 (95%CI: $[0.000, 203.885]$).

We have that changing the treatment formulation from $X = 1$ to $X = 2$ increases the amount of cholesterol reduction, with the average causal effect being 3.432. The relatively large standard deviation suggests that the causal effect exhibits a fair degree of heterogeneity across individuals. Our results exhibit substantially greater positive skewness than that of an exponential distribution, which has a skewness of 2, and significantly higher kurtosis than that of a Gaussian distribution, which has a kurtosis of 3. The large positive skewness suggests that there may be a larger number of individuals having effects smaller than the average 3.432, rather than larger than 3.432; and it suggests the existence of a small number of individuals who have effects that are significantly larger than the average, making the average greater than the median. Finally, the kurtosis estimate has a large CI, and a larger sample size appears to be necessary to obtain more reliable estimates. If the estimate is reliable, the large positive kurtosis value indicates a high number of outliers, which, given the large positive skewness, suggests a high number of individuals with causal effects that are significantly higher than the average.

The results for $Y_4 - Y_1$ and $Y_4 - Y_2$ are presented in Appendix G, along with the estimated bounds. The bounds are all relatively wide.

Results on the product moments of causal effects. Next, we study the covariance and correlation of causal effects $Y_2 - Y_1$ and $Y_4 - Y_2$, where the estimated mean of $Y_4 - Y_2$ is 3.201 (95%CI: $[0.471, 5.811]$). The estimates are

Covariance: -2.076 (95%CI: $[-6.235, 0.297]$),

Correlation: -0.594 (95%CI: $[-1.000, 0.750]$).

The results indicate that the causal effects $Y_2 - Y_1$ and $Y_4 - Y_2$ may be negatively correlated.

Consider the change in treatment from $X = 1$ to 2 to 4. The mean of $Y_2 - Y_1$ and $Y_4 - Y_2$ is 3.432 and 3.201, respectively, indicating that the treatment changes from $X = 1$ to $X = 2$ and from $X = 2$ to $X = 4$ increase cholesterol reduction on average. The correlation result offers a more detailed insight. Patients who have larger causal effects $Y_2 - Y_1$ than the average tend to have causal effects $Y_4 - Y_2$ smaller than the average. Conversely, patients who have smaller causal effects $Y_2 - Y_1$ than the average tend to have causal effects $Y_4 - Y_2$ larger than the average. This means that if one of the changes significantly increases cholesterol reduction, then the other change has a smaller effect. However, the estimate of the correlation has a large CI. A larger sample size appears to be necessary to obtain a more reliable estimate.

Additionally, we show the estimated bounds of the covariance and correlation of causal effects $Y_2 - Y_1$ and $Y_4 - Y_2$ in Appendix G. The bounds are relatively wide.

7 CONCLUSION

This work moves beyond the traditional focus on ACE to investigate the family of moments of causal effects. They could serve as fundamental tools for understanding the shape of the distribution of causal effects and their heterogeneity. They can reveal asymmetries in the distribution, suggesting the presence of distinct response individuals that might require different treatment strategies. This understanding is crucial for tailoring interventions and designing more effective policies.

We have also studied conditional moments of causal effects, which characterize the shape of the distribution of causal effects within a subpopulation defined by subjects' covariates W. They provide complementary information to conditional ACE (CACE), i.e., $\mathbb{E}[Y_1 - Y_0|W = w]$, by capturing higher-order properties beyond the mean. The definition, identification, and bounds of the conditional moments of causal effects are discussed in Appendix D.

The exogeneity and monotonicity assumptions, while common in the causal inference literature, can restrict the practical applicability of the identification results. The assumption of exogeneity, requiring the absence of unmeasured confounders, might be plausible in some applications, especially in certain controlled settings. The monotonicity assumption can be challenging to verify in practice. In such scenarios, the bounding results that relax the monotonicity assumption and depend primarily on exogeneity provide a range of plausible causal effects, still offering valuable information for guiding decisions. In general, these assumptions require a cautious interpretation of the findings in the practical use of the results.

Future research will include sensitivity analysis and instrumental variable analysis for the moments of causal effects under unmeasured confounders, extending methods developed for the ACE [Wright, 1928, Cornfield et al., 1959, Imbens and Angrist, 1994, Vanderweele and Arah, 2011, Ding and VanderWeele, 2016]. Another direction will involve deriving narrower bounds. Deriving tighter bounds than those provided by the Fréchet inequalities remains a highly challenging open mathematical problem when $m \geq 2$. In the finance area, some studies [Lux and Papapantoleon, 2017, Bartl et al., 2022] provide improved Fréchet–Hoeffding bounds by incorporating additional information on the joint distribution. In the PoC literature, some studies [Kuroki and Cai, 2011, Dawid et al., 2017] provide narrower bounds by incorporating additional information about third variables (e.g., covariates or mediators).

ACKNOWLEDGEMENTS

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

References

Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

Susan Athey and Guido W. Imbens. Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1):685–725, 2019.

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

D. J. Bartholomew. A test of homogeneity for ordered alternatives. *Biometrika*, 46(1/2):36–48, 1959.

Daniel Bartl, Michael Kupper, Thibaut Lux, Antonis Papapantoleon, and Stephan Eckstein. Marginal and dependence uncertainty: Bounds, optimal transport, and sharpness. *SIAM J. Control Optim.*, 60(1):410–434, January 2022. ISSN 0363-0129.

Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer, January 2006.

Jerome Cornfield, William Haenszel, Hammond Ec, Abraham M. Lilienfeld, Michael Boris Shimkin, and Wynder El. Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, 22 1:173–203, 1959.

Harald Cramér. *Mathematical methods of statistics*, volume 26. Princeton university press, 1999.

A Philip Dawid, Monica Musio, and Rossella Murtas. The probability of causation1. *Law, Probability and Risk*, 16 (4):163–179, 11 2017.

John DiNardo, Nicole M. Fortin, and Thomas Lemieux. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044, 1996.

Peng Ding and Tyler J. VanderWeele. Sensitivity analysis without assumptions. *Epidemiology*, 27(3), 2016.

David P Doane and Lori E Seward. Measuring skewness: a forgotten statistic? *Journal of statistics education*, 19(2), 2011.

J. L. Doob. The Limiting Distributions of Certain Statistics. *The Annals of Mathematical Statistics*, 6(3):160 – 169, 1935.

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

Zheng Fang and Andres Santos. Inference on Directionally Differentiable Functions. *The Review of Economic Studies*, 86(1):377–412, 09 2018.

M. Fréchet. Sur les tableaux dont les marges et des bornes sont données. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 28 (1/2):10–32, 1960.

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

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer New York, 2009.

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.

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

Takahiro Hoshino and Keisuke Takahata. Parametric identification of the joint distribution of the potential outcomes. *Stat*, 9(1):e254, 2020.

Kosuke Imai and David A van Dyk. Causal inference with general treatment regimes. *Journal of the American Statistical Association*, 99(467):854–866, 2004.

Guido W. Imbens. The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3): 706–710, 2000.

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

D. N. Joanes and C. A. Gill. Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):183–189, 1998.

Chuan Ju and Zhi Geng. Criteria for surrogate end points based on causal distributions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 72(1): 129–142, 2010.

Yonghan Jung, Jin Tian, and Elias Bareinboim. Double machine learning density estimation for local treatment effects with instruments. In *Advances in Neural Information Processing Systems*, volume 34, pages 21821–21833. Curran Associates, Inc., 2021.

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)*, 2024a.

Yuta Kawakami, Manabu Kuroki, and Jin Tian. Identification and estimation of conditional average partial causal effects via instrumental variable. *Proceedings of the 40th Conference on Uncertainty in Artificial Intelligence (UAI-2024)*, 2024b.

E H Kennedy, S Balakrishnan, and L A Wasserman. Semiparametric counterfactual density estimation. *Biometrika*, 110(4):875–896, 03 2023a.

Edward H Kennedy, Sivaraman Balakrishnan, and LA Wasserman. Semiparametric counterfactual density estimation. *Biometrika*, 110(4):875–896, 2023b.

Kwangho Kim, Jisu Kim, and Edward H. Kennedy. Causal effects based on distributional distances, 2024. URL `https://arxiv.org/abs/1806.02935`.

Sören R. Künzel, Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, 2019.

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

Manabu Kuroki and Taiki Tezuka. The estimated causal effect on the variance based on the front-door criterion in gaussian linear structural equation models: an unbiased estimator with the exact variance. *Statistical Papers*, 65 (3):1285–1308, 2024.

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

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

Zhenhua Lin, Dehan Kong, and Linbo Wang. Causal inference on distribution functions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85 (2):378–398, 03 2023.

Thibaut Lux and Antonis Papapantoleon. Improved fréchet–hoeffding bounds on d-copulas and applications in model-free finance. 2017.

George W. Mackey. Harmonic analysis as the exploitation of symmetry–a historical survey. *Bulletin (New Series) of the American Mathematical Society*, 3(1.P1):543 – 698, 1980.

Kevin P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022.

R.B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer New York, 2007.

Jerzy Neyman. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes (in polish). english translation by dm dabrowska and tp speed (1990). *Statistical Science*, 5:465–480, 1923.

Ellis Batten Page. Ordered hypotheses for multiple treatments: a significance test for linear ranks. *Journal of the American Statistical Association*, 58(301):216–230, 1963.

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.

Karl Pearson. Vii. mathematical contributions to the theory of evolution.—iii. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, 1896.

Karl Pearson. "das fehlergesetz und seine verallgemeinerungen durch fechner und pearson." a rejoinder. *Biometrika*, 4(1/2):169–212, 1905.

R. A. J. Post. *Causal Effect Heterogeneity: Statistical formalization and analysis of the individual causal effect*. Phd Thesis 1 (Research TU/e / Graduation TU/e), Mathematics and Computer Science, Eindhoven University of Technology, 2023.

W.H. Press. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, 2007.

James Robins and Andrea Rotnitzky. Comment on the bickel and kwon article, 'inference for semiparametric models: Some questions and an answer'. *Statistica Sinica*, 11:920–936, 01 2001.

James M. Robins. Association, causation, and marginal structural models. *Synthese*, 121(1):151–179, 1999.

Daniel Rubin and Mark J van der Laan. Extending marginal structural models through local, penalized, and additive learning. 2006.

Donald B. Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6(1): 34–58, 1978.

Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning*, pages 3076–3085. PMLR, 2017.

R Singh, L Xu, and A Gretton. Kernel methods for causal functions: dose, heterogeneous and incremental response curves. *Biometrika*, page asad042, July 2023.

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

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

Tyler J Vanderweele and Onyebuchi A Arah. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, 22(1):42–52, Jan 2011.

Paul T. von Hippel. Mean, median, and skew: Correcting a textbook rule. *Journal of Statistics Education*, 13(2), 2005.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523): 1228–1242, 2018.

Peter H Westfall. Kurtosis as peakedness, 1905 - 2014. r.i.p. *Am Stat*, 68(3):191–195, 2014.

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

Wolfgang Wiedermann, Bixi Zhang, Wendy Reinke, Keith C Herman, and Alexander von Eye. Distributional causal effects: Beyond an "averagarian" view of intervention effects. *Psychological Methods*, 2022.

P.G. Wright. *The Tariff on Animal and Vegetable Oils*. Investigations in international commercial policies. Macmillan, 1928.

APPENDIX TO "MOMENTS OF CAUSAL EFFECTS"

We provide the proofs of the lemmas and theorems in the body of the paper in Appendix A, the discussion of the central moments of causal effects in Appendix B, the discussion of the central product moments of causal effects in Appendix C, the discussion of the conditional moments of causal effects in Appendix D, the details and consistency of all estimators in the body of the paper in Appendix E, the plots of the estimates obtained from the numerical experiments in Appendix F, and additional information about the application in the body of our paper in Appendix G.

A PROOFS

In this appendix, we provide the proofs of the lemmas and theorems in the body of the paper. We first show Fréchet inequalities [Fréchet, 1935, 1960]. If A_i are logical propositions or events, the Fréchet inequalities are

$$\max\left\{\sum_{i=1}^{n}\mathbb{P}(A_i)-(n-1),0\right\}\leq\mathbb{P}\left(\bigwedge_{i=1}^{n}A_i\right)\leq\min_{i=1,\ldots,n}\left\{\mathbb{P}(A_i)\right\}, \tag{26}$$

where $\wedge$ is a logical conjunction. Especially,

$$\max\{\mathbb{P}(A)+\mathbb{P}(B)-1,0\}\leq\mathbb{P}(A\wedge B)\leq\min\{\mathbb{P}(A),\mathbb{P}(B)\} \tag{27}$$

holds.

Lemma 1. *Under SCM $\mathcal{M}$, given $m\geq 1$, we have*

$$\begin{aligned}(Y_1-Y_0)^m&=(Y_1-Y_0)^m\mathbb{I}(Y_1>Y_0)+(-1)^m(Y_0-Y_1)^m\mathbb{I}(Y_0>Y_1)\\&=\int_{{\Omega_Y}^m}\mathbb{I}(Y_0<y_1\leq Y_1,Y_0<y_2\leq Y_1,\ldots,Y_0<y_m\leq Y_1)dy_1\ldots dy_m\\&+(-1)^m\int_{{\Omega_Y}^m}\mathbb{I}(Y_1<y_1\leq Y_0,Y_1<y_2\leq Y_0,\ldots,Y_1<y_m\leq Y_0)dy_1\ldots dy_m.\end{aligned} \tag{28}$$

Proof. Given $m\geq 1$, we have

$$\begin{aligned}&\int_{{\Omega_Y}^m}\mathbb{I}(Y_0<y_1\leq Y_1,Y_0<y_2\leq Y_1,\ldots,Y_0<y_m\leq Y_1)dy_1\ldots dy_m\\&=\int_{\Omega_Y}\mathbb{I}(Y_0<y_1\leq Y_1)dy_1\cdots\int_{\Omega_Y}\mathbb{I}(Y_0<y_m\leq Y_1)dy_m\\&=(Y_1-Y_0)^m\mathbb{I}(Y_1>Y_0).\end{aligned} \tag{29}$$

Similarly, given $m\geq 1$, we have

$$\begin{aligned}&\int_{{\Omega_Y}^m}\mathbb{I}(Y_1<y_1\leq Y_0,Y_1<y_2\leq Y_0,\ldots,Y_1<y_m\leq Y_0)dy_1\ldots dy_m\\&=\int_{\Omega_Y}\mathbb{I}(Y_1<y_1\leq Y_0)dy_1\cdots\int_{\Omega_Y}\mathbb{I}(Y_1<y_m\leq Y_0)dy_m\\&=(Y_0-Y_1)^m\mathbb{I}(Y_1>Y_0)^m\\&=(-1)^m(Y_1-Y_0)^m\mathbb{I}(Y_0>Y_1).\end{aligned} \tag{30}$$

Then, given $m\geq 1$, we have

$$\begin{aligned}(Y_1-Y_0)^m&=\int_{{\Omega_Y}^m}\mathbb{I}(Y_0<y_1\leq Y_1,Y_0<y_2\leq Y_1,\ldots,Y_0<y_m\leq Y_1)dy_1\ldots dy_m\\&+(-1)^m\int_{{\Omega_Y}^m}\mathbb{I}(Y_1<y_1\leq Y_0,Y_1<y_2\leq Y_0,\ldots,Y_1<y_m\leq Y_0)dy_1\ldots dy_m.\end{aligned} \tag{31}$$

□

Theorem 1. *Under SCM $\mathcal{M}$ and Assumptions 1, 2, and 3, given $m \geq 1$, the m-th moment of causal effect $Y_1 - Y_0$ is identifiable by $\sigma^{(m)}$, where*

$$\sigma^{(m)} = \int_{{\Omega_Y}^m} \max\Big\{ \min_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=0)\} - \max_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=1)\}, 0\Big\} dy_1 \ldots dy_m$$
$$+ (-1)^m \int_{{\Omega_Y}^m} \max\Big\{ \min_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=1)\} - \max_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=0)\}, 0\Big\} dy_1 \ldots dy_m. \tag{32}$$

Proof. $\mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, Y_0 < y_m \leq Y_1)$ and $\mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, Y_1 < y_m \leq Y_0)$ are identifiable by $\max\{\min_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=0)\} - \max_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=1), 0\}$ and $\max\{\min_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=1)\} - \max_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=0), 0\}$ respectively by Theorem 5.2 in [Kawakami et al., 2024a] under Assumptions 1 and 2. Then we have Eq. (6). □

Lemma 2. *Under SCM $\mathcal{M}$ and Assumptions 1 and 3, given $m \geq 1$, we have $l(y_1, \ldots, y_m; i, j) \leq \mathbb{P}(Y_j < y_1 \leq Y_i, Y_j < y_2 \leq Y_i, \ldots, Y_j < y_m \leq Y_i) \leq u(y_1, \ldots, y_m; i, j)$ where*

$$l(y_1, \ldots, y_m; i, j) = \max\left\{ \sum_{p=1,\ldots,m} \mathbb{P}(Y < y_p|X=j) - \sum_{p=1,\ldots,m} \mathbb{P}(Y < y_p|X=i) - m + 1, 0 \right\}, \tag{33}$$

$$u(y_1, \ldots, y_m; i, j) = \min\left\{ \min_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=j)\}, 1 - \max_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=i)\} \right\} \tag{34}$$

for $(i, j) \in \{(1, 0), (0, 1)\}$ and each $y_1, \ldots, y_m \in \Omega_Y$.

Proof. From Fréchet inequalities [Fréchet, 1935, 1960], given $m \geq 1$, we have

$$\begin{aligned}
&\mathbb{P}(Y_j < y_1 \leq Y_i, Y_j < y_2 \leq Y_i, \ldots, Y_j < y_m \leq Y_i) \\
&= \mathbb{P}(Y_j < y_1, y_1 \leq Y_i, Y_j < y_2, y_2 \leq Y_i, \ldots, Y_j < y_m, y_m \leq Y_i) \\
&\geq \max\left\{ \sum_{p=1,\ldots,m} \mathbb{P}(Y_j < y_p) + \sum_{p=1,\ldots,m} \{1 - \mathbb{P}(Y_i < y_p)\} - (2m-1), 0 \right\} \\
&= \max\left\{ \sum_{p=1,\ldots,m} \mathbb{P}(Y_j < y_p) - \sum_{p=1,\ldots,m} \mathbb{P}(Y_i < y_p) - (m-1), 0 \right\} \\
&= \max\left\{ \sum_{p=1,\ldots,m} \mathbb{P}(Y < y_p|X=j) - \sum_{p=1,\ldots,m} \mathbb{P}(Y < y_p|X=i) - (m-1), 0 \right\}
\end{aligned} \tag{35}$$

and

$$\begin{aligned}
&\mathbb{P}(Y_j < y_1 \leq Y_i, Y_j < y_2 \leq Y_i, \ldots, Y_j < y_m \leq Y_i) \\
&\leq \min\{\mathbb{P}(Y_j < y_1), \ldots, \mathbb{P}(Y_j < y_m), \mathbb{P}(Y_i \leq y_1), \ldots, \mathbb{P}(Y_i \leq y_m)\} \\
&= \min\left\{ \min_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=j)\}, 1 - \max_{p=1,\ldots,m}\{\mathbb{P}(Y < y_p|X=i)\} \right\}
\end{aligned} \tag{36}$$

for $(i, j) \in \{(1, 0), (0, 0)\}$ and each $y_1, \ldots, y_p \in \Omega_Y$. □

Theorem 2. *Under SCM $\mathcal{M}$ and Assumptions 1 and 3, given $m \geq 1$, we have $\sigma_L^{(m)} \leq \mu^{(m)} \leq \sigma_U^{(m)}$, where*

(A). When m is an even number,

$$\sigma_L^{(m)} = \int_{{\Omega_Y}^m} l(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m + \int_{{\Omega_Y}^m} l(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m, \tag{37}$$

$$\sigma_U^{(m)} = \int_{\Omega_Y{}^m} u(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m + \int_{\Omega_Y{}^m} u(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m. \tag{38}$$

(B). When m is an odd number,

$$\sigma_L^{(m)} = \int_{\Omega_Y{}^m} l(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m - \int_{\Omega_Y{}^m} u(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m, \tag{39}$$

$$\sigma_U^{(m)} = \int_{\Omega_Y{}^m} u(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m - \int_{\Omega_Y{}^m} l(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m. \tag{40}$$

Proof. When m is an even number, we have

$$\begin{aligned}
\mu^{(m)} &= \int_{\Omega_Y{}^m} \mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, Y_0 < y_m \leq Y_1) dy_1 \ldots dy_m \\
&+ (-1)^m \int_{\Omega_Y{}^m} \mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, Y_1 < y_m \leq Y_0) dy_1 \ldots dy_m \\
&= \int_{\Omega_Y{}^m} \mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, Y_0 < y_m \leq Y_1) dy_1 \ldots dy_m \\
&+ \int_{\Omega_Y{}^m} \mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, Y_1 < y_m \leq Y_0) dy_1 \ldots dy_m \\
&\geq \int_{\Omega_Y{}^m} l(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m + \int_{\Omega_Y{}^m} l(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m,
\end{aligned} \tag{41}$$

and

$$\begin{aligned}
\mu^{(m)} &= \int_{\Omega_Y{}^m} \mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, Y_0 < y_m \leq Y_1) dy_1 \ldots dy_m \\
&+ (-1)^m \int_{\Omega_Y{}^m} \mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, Y_1 < y_m \leq Y_0) dy_1 \ldots dy_m \\
&= \int_{\Omega_Y{}^m} \mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, Y_0 < y_m \leq Y_1) dy_1 \ldots dy_m \\
&+ \int_{\Omega_Y{}^m} \mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, Y_1 < y_m \leq Y_0) dy_1 \ldots dy_m \\
&\leq \int_{\Omega_Y{}^m} u(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m + \int_{\Omega_Y{}^m} u(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m.
\end{aligned} \tag{42}$$

When m is an even number, we have

$$\begin{aligned}
\mu^{(m)} &= \int_{\Omega_Y{}^m} \mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, Y_0 < y_m \leq Y_1) dy_1 \ldots dy_m \\
&+ (-1)^m \int_{\Omega_Y{}^m} \mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, Y_1 < y_m \leq Y_0) dy_1 \ldots dy_m \\
&= \int_{\Omega_Y{}^m} \mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, Y_0 < y_m \leq Y_1) dy_1 \ldots dy_m \\
&- \int_{\Omega_Y{}^m} \mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, Y_1 < y_m \leq Y_0) dy_1 \ldots dy_m \\
&\geq \int_{\Omega_Y{}^m} l(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m - \int_{\Omega_Y{}^m} u(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m,
\end{aligned} \tag{43}$$

and

$$\mu^{(m)} = \int_{\Omega_Y{}^m} \mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, Y_0 < y_m \leq Y_1) dy_1 \ldots dy_m$$

$$
\begin{aligned}
&+(-1)^m \int_{{\Omega_Y}^m} \mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, Y_1 < y_m \leq Y_0) dy_1 \ldots dy_m \\
&= \int_{{\Omega_Y}^m} \mathbb{P}(Y_0 < y_1 \leq Y_1, Y_0 < y_2 \leq Y_1, \ldots, Y_0 < y_m \leq Y_1) dy_1 \ldots dy_m \\
&- \int_{{\Omega_Y}^m} \mathbb{P}(Y_1 < y_1 \leq Y_0, Y_1 < y_2 \leq Y_0, \ldots, Y_1 < y_m \leq Y_0) dy_1 \ldots dy_m \\
&\leq \int_{{\Omega_Y}^m} u(y_1, \ldots, y_m; 1, 0) dy_1 \ldots dy_m - \int_{{\Omega_Y}^m} l(y_1, \ldots, y_m; 0, 1) dy_1 \ldots dy_m. \qquad (44)
\end{aligned}
$$

□

Lemma 3. *Under SCM $\mathcal{M}$, for any $i, j, k, h \in \{1, \ldots, R\}$, we have*

$$
\begin{aligned}
&(Y_i - Y_j)(Y_k - Y_h) \\
&= (Y_i - Y_j)(Y_k - Y_h)\mathbb{I}(Y_i > Y_j, Y_k > Y_h) - (Y_j - Y_i)(Y_k - Y_h)\mathbb{I}(Y_j > Y_i, Y_k > Y_h) \\
&- (Y_i - Y_j)(Y_h - Y_k)\mathbb{I}(Y_i > Y_j, Y_h > Y_k) + (Y_j - Y_i)(Y_h - Y_k)\mathbb{I}(Y_j > Y_i, Y_h > Y_k) \\
&= \int_{\Omega_Y} \int_{\Omega_Y} \mathbb{I}(Y_j < y_1 \leq Y_i, Y_h < y_2 \leq Y_k) dy_1 dy_2 - \int_{\Omega_Y} \int_{\Omega_Y} \mathbb{I}(Y_i < y_1 \leq Y_j, Y_h < y_2 \leq Y_k) dy_1 dy_2 \\
&- \int_{\Omega_Y} \int_{\Omega_Y} \mathbb{I}(Y_j < y_1 \leq Y_i, Y_k < y_2 \leq Y_h) dy_1 dy_2 + \int_{\Omega_Y} \int_{\Omega_Y} \mathbb{I}(Y_i < y_1 \leq Y_j, Y_k < y_2 \leq Y_h) dy_1 dy_2. \qquad (45)
\end{aligned}
$$

Proof. For any $i, j, k, h \in \{1, \ldots, R\}$, we have

$$
\begin{aligned}
&\int_{\Omega_Y} \int_{\Omega_Y} \mathbb{I}(Y_j < y_1 \leq Y_i, Y_h < y_2 \leq Y_k) dy_1 dy_2 - \int_{\Omega_Y} \int_{\Omega_Y} \mathbb{I}(Y_i < y_1 \leq Y_j, Y_h < y_2 \leq Y_k) dy_1 dy_2 \\
&- \int_{\Omega_Y} \int_{\Omega_Y} \mathbb{I}(Y_j < y_1 \leq Y_i, Y_k < y_2 \leq Y_h) dy_1 dy_2 + \int_{\Omega_Y} \int_{\Omega_Y} \mathbb{I}(Y_i < y_1 \leq Y_j, Y_k < y_2 \leq Y_h) dy_1 dy_2 \\
&= \int_{\Omega_Y} \mathbb{I}(Y_j < y_1 \leq Y_i) dy_1 \int_{\Omega_Y} \mathbb{I}(Y_h < y_2 \leq Y_k) dy_2 - \int_{\Omega_Y} \mathbb{I}(Y_i < y_1 \leq Y_j) dy_1 \int_{\Omega_Y} \mathbb{I}(Y_h < y_2 \leq Y_k) dy_2 \\
&- \int_{\Omega_Y} \mathbb{I}(Y_j < y_1 \leq Y_i) dy_1 \int_{\Omega_Y} \mathbb{I}(Y_k < y_2 \leq Y_h) dy_2 + \int_{\Omega_Y} \mathbb{I}(Y_i < y_1 \leq Y_j) dy_1 \int_{\Omega_Y} \mathbb{I}(Y_k < y_2 \leq Y_h) dy_2 \\
&= (Y_i - Y_j)(Y_k - Y_h)\mathbb{I}(Y_i > Y_j, Y_k > Y_h) - (Y_j - Y_i)(Y_k - Y_h)\mathbb{I}(Y_j > Y_i, Y_k > Y_h) \\
&- (Y_i - Y_j)(Y_h - Y_k)\mathbb{I}(Y_i > Y_j, Y_h > Y_k) + (Y_j - Y_i)(Y_h - Y_k)\mathbb{I}(Y_j > Y_i, Y_h > Y_k) \\
&= (Y_i - Y_j)(Y_k - Y_h)\mathbb{I}(Y_i > Y_j, Y_k > Y_h) + (Y_i - Y_j)(Y_k - Y_h)\mathbb{I}(Y_j > Y_i, Y_k > Y_h) \\
&+ (Y_i - Y_j)(Y_k - Y_h)\mathbb{I}(Y_i > Y_j, Y_h > Y_k) + (Y_i - Y_j)(Y_k - Y_h)\mathbb{I}(Y_j > Y_i, Y_h > Y_k) \\
&= (Y_i - Y_j)(Y_k - Y_h). \qquad (46)
\end{aligned}
$$

□

Theorem 3. *Under SCM $\mathcal{M}$ and Assumptions 1, 2, and 4, for any $i, j, k, h \in \{1, \ldots, R\}$, the product moment of $(Y_i - Y_j)$ and $(Y_k - Y_h)$ is identifiable by $\sigma(i, j; k, h)$, where*

$$
\begin{aligned}
&\sigma(i, j; k, h) \\
&= \int_{{\Omega_Y}^2} \max\Big\{ \min\{\mathbb{P}(Y < y_1 | X = j), \mathbb{P}(Y < y_2 | X = h)\} - \max\{\mathbb{P}(Y < y_1 | X = i), \mathbb{P}(Y < y_2 | X = k)\}, 0 \Big\} dy_1 dy_2 \\
&- \int_{{\Omega_Y}^2} \max\Big\{ \min\{\mathbb{P}(Y < y_1 | X = i), \mathbb{P}(Y < y_2 | X = h)\} - \max\{\mathbb{P}(Y < y_1 | X = j), \mathbb{P}(Y < y_2 | X = k)\}, 0 \Big\} dy_1 dy_2 \\
&- \int_{{\Omega_Y}^2} \max\Big\{ \min\{\mathbb{P}(Y < y_1 | X = j), \mathbb{P}(Y < y_2 | X = k)\} - \max\{\mathbb{P}(Y < y_1 | X = i), \mathbb{P}(Y < y_2 | X = h)\}, 0 \Big\} dy_1 dy_2 \\
&+ \int_{{\Omega_Y}^2} \max\Big\{ \min\{\mathbb{P}(Y < y_1 | X = i), \mathbb{P}(Y < y_2 | X = k)\} - \max\{\mathbb{P}(Y < y_1 | X = j), \mathbb{P}(Y < y_2 | X = h)\}, 0 \Big\} dy_1 dy_2. \\
&\qquad (47)
\end{aligned}
$$

Proof. $\mathbb{P}(Y_j < y_1 \leq Y_i, Y_h < y_2 \leq Y_k)$ are identifiable by $\max\{\min\{\mathbb{P}(Y < y_1|X = j), \mathbb{P}(Y < y_2|X = h)\} - \max\{\mathbb{P}(Y < y_1|X = i), \mathbb{P}(Y < y_2|X = k)\}, 0\}$ $(i, j, k, h \in \{1, \ldots, R\})$ respectively by Theorem 5.2 in [Kawakami et al., 2024a] under Assumptions 1 and 2. Then, we have Eq. (17). □

Lemma 4. *Under SCM $\mathcal{M}$ and Assumptions 1 and 4, for any $i, j, k, h \in \{1, \ldots, R\}$, we have $l(y_1, y_3; i, j, k, h) \leq \mathbb{P}(Y_j < y_1 \leq Y_i, Y_h < y_2 \leq Y_k) \leq u(y_1, y_3; i, j, k, h)$, where*

$$\begin{aligned} &l(y_1, y_2; i, j, k, h) \\ &= \max\Big\{\mathbb{P}(Y < y_1|X = j) - \mathbb{P}(Y < y_1|X = i) + \mathbb{P}(Y < y_2|X = h) - \mathbb{P}(Y < y_2|X = k) - 1, 0\Big\}, \end{aligned} \tag{48}$$

$$\begin{aligned} &u(y_1, y_2; i, j, k, h) \\ &= \min\Big\{\min\{\mathbb{P}(Y < y_1|X = j), \mathbb{P}(Y < y_2|X = h)\}, 1 - \max\{\mathbb{P}(Y < y_1|X = i), \mathbb{P}(Y < y_2|X = k)\}\Big\} \end{aligned} \tag{49}$$

for each $i, j, k, h \in \{1, \ldots, R\}$ and $y_1, y_2 \in \Omega_Y$.

Proof. From Fréchet inequalities [Fréchet, 1935, 1960], for any $i, j, k, h \in \{1, \ldots, R\}$, we have

$$\begin{aligned} &\mathbb{P}(Y_j < y_1 \leq Y_i, Y_h < y_2 \leq Y_k) = \mathbb{P}(Y_j < y_1, y_1 \leq Y_i, Y_h < y_2, y_2 \leq Y_k) \\ &\geq \max\Big\{\mathbb{P}(Y_j < y_1) + \mathbb{P}(y_1 \leq Y_i) + \mathbb{P}(Y_h < y_2) + \mathbb{P}(y_2 \leq Y_k) - 3, 0\Big\} \\ &= \max\Big\{\mathbb{P}(Y_j < y_1) - \mathbb{P}(Y_i < y_1) + \mathbb{P}(Y_h < y_2) - \mathbb{P}(Y_k < y_2) - 1, 0\Big\} \\ &= \max\Big\{\mathbb{P}(Y < y_1|X = j) - \mathbb{P}(Y < y_1|X = i) + \mathbb{P}(Y < y_2|X = h) - \mathbb{P}(Y < y_2|X = k) - 1, 0\Big\} \end{aligned} \tag{50}$$

and

$$\begin{aligned} &\mathbb{P}(Y_j < y_1 \leq Y_i, Y_h < y_2 \leq Y_k) = \mathbb{P}(Y_j < y_1, y_1 \leq Y_i, Y_h < y_2, y_2 \leq Y_k) \\ &\leq \min\Big\{\min\{\mathbb{P}(Y_j < y_1), \mathbb{P}(Y_h < y_2)\}, 1 - \max\{\mathbb{P}(Y_i < y_1), \mathbb{P}(Y_k < y_2)\}\Big\} \\ &= \min\Big\{\min\{\mathbb{P}(Y < y_1|X = j), \mathbb{P}(Y < y_2|X = h)\}, 1 - \max\{\mathbb{P}(Y < y_1|X = i), \mathbb{P}(Y < y_2|X = k)\}\Big\}. \end{aligned} \tag{51}$$

□

Theorem 4. *Under SCM $\mathcal{M}$ and Assumptions 1 and 4, for any $i, j, k, h \in \{1, \ldots, R\}$, we have $\sigma_L(i, j; k, h) \leq \rho_{i,j;k,h} \leq \sigma_U(i, j; k, h)$, where*

$$\begin{aligned} \sigma_L(i, j; k, h) = &\int_{{\Omega_Y}^2} l(y_1, y_2; i, j, k, h) dy_1 dy_2 - \int_{{\Omega_Y}^2} u(y_1, y_2; j, i, k, h) dy_1 dy_2 \\ &- \int_{{\Omega_Y}^2} u(y_1, y_2; i, j, h, k) dy_1 dy_2 + \int_{{\Omega_Y}^2} l(y_1, y_2; j, i, h, k) dy_1 dy_2, \end{aligned} \tag{52}$$

$$\begin{aligned} \sigma_U(i, j; k, h) = &\int_{{\Omega_Y}^2} u(y_1, y_2; i, j, k, h) dy_1 dy_2 - \int_{{\Omega_Y}^2} l(y_1, y_2; j, i, k, h) dy_1 dy_2 \\ &- \int_{{\Omega_Y}^2} l(y_1, y_2; i, j, h, k) dy_1 dy_2 + \int_{{\Omega_Y}^2} u(y_1, y_2; j, i, h, k) dy_1 dy_2. \end{aligned} \tag{53}$$

Proof. For any $i, j, k, h \in \{1, \ldots, R\}$, we have

$$\begin{aligned} \rho_{i,j;k,h} = &\int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j < y_1 \leq Y_i, Y_h < y_2 \leq Y_k) dy_1 dy_2 - \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_i < y_1 \leq Y_j, Y_h < y_2 \leq Y_k) dy_1 dy_2 \\ &- \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j < y_1 \leq Y_i, Y_k < y_2 \leq Y_h) dy_1 dy_2 + \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_i < y_1 \leq Y_j, Y_k < y_2 \leq Y_h) dy_1 dy_2 \end{aligned}$$

$$\geq \int_{{\Omega_Y}^2} l(y_1, y_2; i, j, k, h) dy_1 dy_2 - \int_{{\Omega_Y}^2} u(y_1, y_2; j, i, k, h) dy_1 dy_2 \\ - \int_{{\Omega_Y}^2} u(y_1, y_2; i, j, h, k) dy_1 dy_2 + \int_{{\Omega_Y}^2} l(y_1, y_2; j, i, h, k) dy_1 dy_2, \tag{54}$$

and

$$\begin{aligned}
\rho_{i,j;k,h} &= \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j < y_1 \leq Y_i, Y_h < y_2 \leq Y_k) dy_1 dy_2 - \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_i < y_1 \leq Y_j, Y_h < y_2 \leq Y_k) dy_1 dy_2 \\
&- \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j < y_1 \leq Y_i, Y_k < y_2 \leq Y_h) dy_1 dy_2 + \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_i < y_1 \leq Y_j, Y_k < y_2 \leq Y_h) dy_1 dy_2 \\
&\leq \int_{{\Omega_Y}^2} u(y_1, y_2; i, j, k, h) dy_1 dy_2 - \int_{{\Omega_Y}^2} l(y_1, y_2; j, i, k, h) dy_1 dy_2 \\
&\quad - \int_{{\Omega_Y}^2} l(y_1, y_2; i, j, h, k) dy_1 dy_2 + \int_{{\Omega_Y}^2} u(y_1, y_2; j, i, h, k) dy_1 dy_2.
\end{aligned} \tag{55}$$

□

B IDENTIFICATION AND BOUNDS OF THE CENTRAL MOMENTS OF CAUSAL EFFECTS

In this section, we discuss the central moments of causal effects.

We make the following assumption:

Assumption 5 (Existence of integrals)**.** *Under SCM $\mathcal{M}$, given $m \geq 1$, $\overline{\mu}^{(m)} < \infty$ and $\int_{{\Omega_Y}^m} \mathbb{P}(Y_i - \mathbb{E}[Y_i] < y_1 \leq Y_j - \mathbb{E}[Y_j], Y_i - \mathbb{E}[Y_i] < y_2 \leq Y_j - \mathbb{E}[Y_j], \ldots, Y_i - \mathbb{E}[Y_i] < y_m \leq Y_j - \mathbb{E}[Y_j]) dy_1 \ldots dy_m < \infty$ hold for $(i, j) = \{(0, 1), (1, 0)\}$.*

Identification of the central moment of causal effects.

Then, we have the following identification theorem.

Theorem 5 (Identification of the central moment of causal effects)**.** *Under SCM $\mathcal{M}$ and Assumptions 1, 2, and 5, given $m \geq 1$, $\overline{\mu}^{(m)}$ is identifiable by $\bar{\sigma}^{(m)}$, where*

$$\begin{aligned}
\bar{\sigma}^{(m)} &= \int_{{\Omega_Y}^m} \max\Big\{ \min_{p=1,\ldots,m} \{\mathbb{P}(Y - \mathbb{E}[Y|X=0] < y_p | X = 0)\} \\
&\qquad - \max_{p=1,\ldots,m} \{\mathbb{P}(Y - \mathbb{E}[Y|X=1] < y_p | X = 1)\}, 0 \Big\} dy_1 \ldots dy_m \\
&+ (-1)^m \int_{{\Omega_Y}^m} \max\Big\{ \min_{p=1,\ldots,m} \{\mathbb{P}(Y - \mathbb{E}[Y|X=1] < y_p | X = 1)\} \\
&\qquad - \max_{p=1,\ldots,m} \{\mathbb{P}(Y - \mathbb{E}[Y|X=0] < y_p | X = 0)\}, 0 \Big\} dy_1 \ldots dy_m.
\end{aligned} \tag{56}$$

Proof. Given $m \geq 1$, we have

$$\begin{aligned}
&\int_{{\Omega_Y}^m} \mathbb{I}(Y_0 - \mathbb{E}[Y_0] < y_1 \leq Y_1 - \mathbb{E}[Y_1], Y_0 - \mathbb{E}[Y_0] < y_2 \leq Y_1 - \mathbb{E}[Y_1], \ldots, \\
&\qquad\qquad Y_0 - \mathbb{E}[Y_0] < y_m \leq Y_1 - \mathbb{E}[Y_1]) dy_1 \ldots dy_m \\
&= \int_{\Omega_Y} \mathbb{I}(Y_0 - \mathbb{E}[Y_0] < y_1 \leq Y_1 - \mathbb{E}[Y_1]) dy_1 \cdots \int_{\Omega_Y} \mathbb{I}(Y_0 - \mathbb{E}[Y_0] < y_m \leq Y_1 - \mathbb{E}[Y_1]) dy_m \\
&= \{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\}^m \mathbb{I}(Y_1 - \mathbb{E}[Y_1] > Y_0 - \mathbb{E}[Y_0]).
\end{aligned} \tag{57}$$

Similarly, given $m \geq 1$, we have

$$\int_{{\Omega_Y}^m} \mathbb{I}(Y_1 - \mathbb{E}[Y_1] < y_1 \leq Y_0 - \mathbb{E}[Y_1], Y_1 - \mathbb{E}[Y_1] < y_2 \leq Y_0 - \mathbb{E}[Y_0], \ldots,$$

$$
\begin{aligned}
&\qquad\qquad Y_1 - \mathbb{E}[Y_1] < y_m \leq Y_0 - \mathbb{E}[Y_0])dy_1 \dots dy_m \\
&= \int_{\Omega_Y} \mathbb{I}(Y_1 - \mathbb{E}[Y_1] < y_1 \leq Y_0 - \mathbb{E}[Y_0])dy_1 \cdots \int_{\Omega_Y} \mathbb{I}(Y_1 - \mathbb{E}[Y_1] < y_m \leq Y_0 - \mathbb{E}[Y_0])dy_m \\
&= \{(Y_0 - Y_1) - (\mathbb{E}[Y_0] - \mathbb{E}[Y_1])\}^m \mathbb{I}(Y_0 - \mathbb{E}[Y_0] > Y_1 - \mathbb{E}[Y_1])^m \\
&= (-1)^m \{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\}^m \mathbb{I}(Y_0 - \mathbb{E}[Y_0] > Y_1 - \mathbb{E}[Y_1]). \qquad (58)
\end{aligned}
$$

Then, given $m \geq 1$, we have

$$
\begin{aligned}
&\{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\}^m \\
&= \int_{{\Omega_Y}^m} \mathbb{I}(Y_0 - \mathbb{E}[Y_0] < y_1 \leq Y_1 - \mathbb{E}[Y_1], Y_0 - \mathbb{E}[Y_0] < y_2 \leq Y_1 - \mathbb{E}[Y_1], \dots, \\
&\qquad\qquad Y_0 - \mathbb{E}[Y_0] < y_m \leq Y_1 - \mathbb{E}[Y_1])dy_1 \dots dy_m \\
&+ (-1)^m \int_{{\Omega_Y}^m} \mathbb{I}(Y_1 - \mathbb{E}[Y_1] < y_1 \leq Y_0 - \mathbb{E}[Y_0], Y_1 - \mathbb{E}[Y_1] < y_2 \leq Y_0 - \mathbb{E}[Y_0], \dots, \\
&\qquad\qquad Y_1 - \mathbb{E}[Y_1] < y_m \leq Y_0 - \mathbb{E}[Y_0])dy_1 \dots dy_m. \qquad (59)
\end{aligned}
$$

Taking expectations, given $m \geq 1$, we obtain

$$
\begin{aligned}
\overline{\mu}^{(m)} &= \int_{{\Omega_Y}^m} \mathbb{P}(Y_0 - \mathbb{E}[Y_0] < y_1 \leq Y_1 - \mathbb{E}[Y_1], Y_0 - \mathbb{E}[Y_0] < y_2 \leq Y_1 - \mathbb{E}[Y_1], \dots, \\
&\qquad\qquad Y_0 - \mathbb{E}[Y_0] < y_m \leq Y_1 - \mathbb{E}[Y_1])dy_1 \dots dy_m \\
&+ (-1)^m \int_{{\Omega_Y}^m} \mathbb{P}(Y_1 - \mathbb{E}[Y_1] < y_1 \leq Y_0 - \mathbb{E}[Y_0], Y_1 - \mathbb{E}[Y_1] < y_2 \leq Y_0 - \mathbb{E}[Y_0], \dots, \\
&\qquad\qquad Y_1 - \mathbb{E}[Y_1] < y_m \leq Y_0 - \mathbb{E}[Y_0])dy_1 \dots dy_m. \qquad (60)
\end{aligned}
$$

Since $\mathbb{P}(Y_0 - \mathbb{E}[Y_0] < y_1 \leq Y_1 - \mathbb{E}[Y_1], Y_0 - \mathbb{E}[Y_0] < y_2 \leq Y_1 - \mathbb{E}[Y_1], \dots, Y_0 - \mathbb{E}[Y_0] < y_m \leq Y_1 - \mathbb{E}[Y_1])$ and $\mathbb{P}(Y_1 - \mathbb{E}[Y_1] < y_1 \leq Y_0 - \mathbb{E}[Y_0], Y_1 - \mathbb{E}[Y_1] < y_2 \leq Y_0 - \mathbb{E}[Y_0], \dots, Y_1 - \mathbb{E}[Y_1] < y_m \leq Y_0 - \mathbb{E}[Y_0])$ are identifiable by re-writing f_Y as $f(x, U^Y) = Y_x - \mathbb{E}[Y_x]$ in Theorem 5.2 in [Kawakami et al., 2024a], we have, given $m \geq 1$,

$$
\begin{aligned}
\overline{\mu}^{(m)} &= \int_{{\Omega_Y}^m} \max\Big\{ \min_{p=1,\dots,m} \{\mathbb{P}(Y - \mathbb{E}[Y|X=0] < y_p | X=0)\} \\
&\qquad\qquad - \max_{p=1,\dots,m} \{\mathbb{P}(Y - \mathbb{E}[Y|X=1] < y_p | X=1)\}, 0 \Big\} dy_1 \dots dy_m \\
&+ (-1)^m \int_{{\Omega_Y}^m} \max\Big\{ \min_{p=1,\dots,m} \{\mathbb{P}(Y - \mathbb{E}[Y|X=1] < y_p | X=1)\} \\
&\qquad\qquad - \max_{p=1,\dots,m} \{\mathbb{P}(Y - \mathbb{E}[Y|X=0] < y_p | X=0)\}, 0 \Big\} dy_1 \dots dy_m. \qquad (61)
\end{aligned}
$$

□

Moments of causal effects are expressed as the combination of conditional CDFs and expectations.

Bounding the central moment of causal effects.

Assumption 2 may sometimes be considered implausible by researchers. Therefore, we derive bounds for the central moments of causal effects without relying on Assumption 2.

Lemma 5. *Under SCM $\mathcal{M}$ and Assumptions 1 and 5, given $m \geq 1$, we have $\bar{l}(y_1, \dots, y_m; i, j) \leq \mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1 \leq Y_i - \mathbb{E}[Y_i], Y_j - \mathbb{E}[Y_j] < y_2 \leq Y_i - \mathbb{E}[Y_i], \dots, Y_j - \mathbb{E}[Y_j] < y_m \leq Y_i - \mathbb{E}[Y_i]) \leq \bar{u}(y_1, \dots, y_m; i, j)$, where*

$$
\begin{aligned}
&\bar{l}(y_1, \dots, y_m; i, j) = \\
&\max\left\{ \sum_{p=1,\dots,m} \mathbb{P}(Y - \mathbb{E}[Y|X=j] < y_p | X=j) - \sum_{p=1,\dots,m} \mathbb{P}(Y - \mathbb{E}[Y|X=i] < y_p | X=i) - (m-1), 0 \right\} \qquad (62)
\end{aligned}
$$

and

$$\bar{u}(y_1,\ldots,y_m;i,j) = \min\left\{\min_{p=1,\ldots,m}\{\mathbb{P}(Y-\mathbb{E}[Y|X=j]<y_p|X=j)\}, 1-\max_{p=1,\ldots,m}\{\mathbb{P}(Y-\mathbb{E}[Y|X=i]<y_p|X=i)\}\right\} \tag{63}$$

for $(i,j)\in\{(1,0),(0,1)\}$ *and each* $y_1,\ldots,y_m\in\Omega_Y$.

Proof. From Fréchet inequalities [Fréchet, 1935, 1960], given $m\geq 1$, we have

$$\begin{aligned}
&\mathbb{P}(Y_j-\mathbb{E}[Y_j]<y_1\leq Y_i-\mathbb{E}[Y_i], Y_j-\mathbb{E}[Y_j]<y_2\leq Y_i-\mathbb{E}[Y_i],\ldots,Y_j-\mathbb{E}[Y_j]<y_m\leq Y_i-\mathbb{E}[Y_i])\\
&=\mathbb{P}(Y_j-\mathbb{E}[Y_j]<y_1, y_1\leq Y_i-\mathbb{E}[Y_i], Y_j-\mathbb{E}[Y_j]<y_2, y_2\leq Y_i-\mathbb{E}[Y_i],\ldots,Y_j-\mathbb{E}[Y_j]<y_m, y_m\leq Y_i-\mathbb{E}[Y_i])\\
&\geq \max\left\{\sum_{p=1,\ldots,m}\mathbb{P}(Y_j-\mathbb{E}[Y_j]<y_p)+\sum_{p=1,\ldots,m}\{1-\mathbb{P}(Y_i-\mathbb{E}[Y_i]<y_p)\}-(2m-1),0\right\}\\
&=\max\left\{\sum_{p=1,\ldots,m}\mathbb{P}(Y_j-\mathbb{E}[Y_j]<y_p)-\sum_{p=1,\ldots,m}\mathbb{P}(Y_i-\mathbb{E}[Y_i]<y_p)-(m-1),0\right\}\\
&=\max\left\{\sum_{p=1,\ldots,m}\mathbb{P}(Y-\mathbb{E}[Y|X=j]<y_p|X=j)-\sum_{p=1,\ldots,m}\mathbb{P}(Y-\mathbb{E}[Y|X=i]<y_p|X=i)-(m-1),0\right\}
\end{aligned} \tag{64}$$

and

$$\begin{aligned}
&\mathbb{P}(Y_j-\mathbb{E}[Y_j]<y_1\leq Y_i-\mathbb{E}[Y_i], Y_j-\mathbb{E}[Y_j]<y_2\leq Y_i-\mathbb{E}[Y_i],\ldots,Y_j-\mathbb{E}[Y_j]<y_m\leq Y_i-\mathbb{E}[Y_i])\\
&\leq \min\{\mathbb{P}(Y_j-\mathbb{E}[Y_j]<y_1),\ldots,\mathbb{P}(Y_j-\mathbb{E}[Y_j]<y_m),\mathbb{P}(Y_i-\mathbb{E}[Y_i]\leq y_1),\ldots,\mathbb{P}(Y_i-\mathbb{E}[Y_i]\leq y_m)\}\\
&=\min\left\{\min_{p=1,\ldots,m}\{\mathbb{P}(Y-\mathbb{E}[Y|X=j]<y_p|X=j)\}, 1-\max_{p=1,\ldots,m}\{\mathbb{P}(Y-\mathbb{E}[Y|X=i]<y_p|X=i)\}\right\}
\end{aligned} \tag{65}$$

for $(i,j)\in\{(1,0),(0,0)\}$ and each $y_1,\ldots,y_p\in\Omega_Y$. □

Theorem 6 (Bounds of the central moment of causal effects)**.** *Under SCM* $\mathcal{M}$ *and Assumptions 1 and 5, given* $m\geq 1$*, we have* $\bar{\sigma}_L^{(m)}\leq\overline{\mu}^{(m)}\leq\bar{\sigma}_U^{(m)}$*, where*

(A). When m *is an even number,*

$$\bar{\sigma}_L^{(m)}=\int_{{\Omega_Y}^m}\bar{l}(y_1,\ldots,y_m;1,0)dy_1\ldots dy_m+\int_{{\Omega_Y}^m}\bar{l}(y_1,\ldots,y_m;0,1)dy_1\ldots dy_m, \tag{66}$$

$$\bar{\sigma}_U^{(m)}=\int_{{\Omega_Y}^m}\bar{u}(y_1,\ldots,y_m;1,0)dy_1\ldots dy_m+\int_{{\Omega_Y}^m}\bar{u}(y_1,\ldots,y_m;0,1)dy_1\ldots dy_m. \tag{67}$$

(B). When m *is an odd number,*

$$\bar{\sigma}_L^{(m)}=\int_{{\Omega_Y}^m}\bar{l}(y_1,\ldots,y_m;1,0)dy_1\ldots dy_m-\int_{{\Omega_Y}^m}\bar{u}(y_1,\ldots,y_m;0,1)dy_1\ldots dy_m, \tag{68}$$

$$\bar{\sigma}_U^{(m)}=\int_{{\Omega_Y}^m}\bar{u}(y_1,\ldots,y_m;1,0)dy_1\ldots dy_m-\int_{{\Omega_Y}^m}\bar{l}(y_1,\ldots,y_m;0,1)dy_1\ldots dy_m. \tag{69}$$

Proof. When m is an even number, we have

$$\begin{aligned}
\overline{\mu}^{(m)}=\int_{{\Omega_Y}^m}&\mathbb{P}(Y_0-\mathbb{E}[Y_0]<y_1\leq Y_1-\mathbb{E}[Y_1], Y_0-\mathbb{E}[Y_0]<y_2\leq Y_1-\mathbb{E}[Y_1],\ldots,\\
&Y_0-\mathbb{E}[Y_0]<y_m\leq Y_1-\mathbb{E}[Y_1])dy_1\ldots dy_m
\end{aligned}$$

$$
\begin{aligned}
&+(-1)^m \int_{{\Omega_Y}^m} \mathbb{P}(Y_1-\mathbb{E}[Y_1]<y_1\le Y_0-\mathbb{E}[Y_0], Y_1-\mathbb{E}[Y_1]<y_2\le Y_0-\mathbb{E}[Y_0],\dots,\\
&\qquad\qquad Y_1-\mathbb{E}[Y_1]<y_m\le Y_0-\mathbb{E}[Y_0])dy_1\dots dy_m\\
&=\int_{{\Omega_Y}^m} \mathbb{P}(Y_0-\mathbb{E}[Y_0]<y_1\le Y_1-\mathbb{E}[Y_1], Y_0-\mathbb{E}[Y_0]<y_2\le Y_1-\mathbb{E}[Y_1],\dots,\\
&\qquad\qquad Y_0-\mathbb{E}[Y_0]<y_m\le Y_1-\mathbb{E}[Y_1])dy_1\dots dy_m\\
&+\int_{{\Omega_Y}^m} \mathbb{P}(Y_1-\mathbb{E}[Y_1]<y_1\le Y_0-\mathbb{E}[Y_0], Y_1-\mathbb{E}[Y_1-\mathbb{E}[Y_0]]<y_2\le Y_0-\mathbb{E}[Y_0],\dots,\\
&\qquad\qquad Y_1-\mathbb{E}[Y_1]<y_m\le Y_0-\mathbb{E}[Y_0])dy_1\dots dy_m\\
&\ge \int_{{\Omega_Y}^m} \bar{l}(y_1,\dots,y_m;1,0)dy_1\dots dy_m+\int_{{\Omega_Y}^m} \bar{l}(y_1,\dots,y_m;0,1)dy_1\dots dy_m
\end{aligned}
\tag{70}
$$

and

$$
\begin{aligned}
\overline{\mu}^{(m)}&=\int_{{\Omega_Y}^m} \mathbb{P}(Y_0-\mathbb{E}[Y_0]<y_1\le Y_1-\mathbb{E}[Y_1], Y_0-\mathbb{E}[Y_0]<y_2\le Y_1-\mathbb{E}[Y_1],\dots,\\
&\qquad\qquad Y_0-\mathbb{E}[Y_0]<y_m\le Y_1-\mathbb{E}[Y_1])dy_1\dots dy_m\\
&+(-1)^m \int_{{\Omega_Y}^m} \mathbb{P}(Y_1-\mathbb{E}[Y_1]<y_1\le Y_0-\mathbb{E}[Y_0], Y_1-\mathbb{E}[Y_1]<y_2\le Y_0-\mathbb{E}[Y_0],\dots,\\
&\qquad\qquad Y_1-\mathbb{E}[Y_1]<y_m\le Y_0-\mathbb{E}[Y_0])dy_1\dots dy_m\\
&=\int_{{\Omega_Y}^m} \mathbb{P}(Y_0-\mathbb{E}[Y_0]<y_1\le Y_1-\mathbb{E}[Y_1], Y_0-\mathbb{E}[Y_0]<y_2\le Y_1-\mathbb{E}[Y_1],\dots,\\
&\qquad\qquad Y_0-\mathbb{E}[Y_0]<y_m\le Y_1-\mathbb{E}[Y_1])dy_1\dots dy_m\\
&+\int_{{\Omega_Y}^m} \mathbb{P}(Y_1-\mathbb{E}[Y_1]<y_1\le Y_0-\mathbb{E}[Y_0], Y_1-\mathbb{E}[Y_1]<y_2\le Y_0-\mathbb{E}[Y_0],\dots,\\
&\qquad\qquad Y_1-\mathbb{E}[Y_1]<y_m\le Y_0-\mathbb{E}[Y_0])dy_1\dots dy_m\\
&\le \int_{{\Omega_Y}^m} \bar{u}(y_1,\dots,y_m;1,0)dy_1\dots dy_m+\int_{{\Omega_Y}^m} \bar{u}(y_1,\dots,y_m;0,1)dy_1\dots dy_m.
\end{aligned}
\tag{71}
$$

When m is an even number, we have

$$
\begin{aligned}
\overline{\mu}^{(m)}&=\int_{{\Omega_Y}^m} \mathbb{P}(Y_0-\mathbb{E}[Y_0]<y_1\le Y_1-\mathbb{E}[Y_1], Y_0-\mathbb{E}[Y_0]<y_2\le Y_1-\mathbb{E}[Y_1],\dots,\\
&\qquad\qquad Y_0-\mathbb{E}[Y_0]<y_m\le Y_1-\mathbb{E}[Y_1])dy_1\dots dy_m\\
&+(-1)^m \int_{{\Omega_Y}^m} \mathbb{P}(Y_1-\mathbb{E}[Y_1]<y_1\le Y_0-\mathbb{E}[Y_0], Y_1-\mathbb{E}[Y_1]<y_2\le Y_0-\mathbb{E}[Y_0],\dots,\\
&\qquad\qquad Y_1-\mathbb{E}[Y_1]<y_m\le Y_0-\mathbb{E}[Y_0])dy_1\dots dy_m\\
&=\int_{{\Omega_Y}^m} \mathbb{P}(Y_0-\mathbb{E}[Y_0]<y_1\le Y_1-\mathbb{E}[Y_1], Y_0-\mathbb{E}[Y_0]<y_2\le Y_1-\mathbb{E}[Y_1],\dots,\\
&\qquad\qquad Y_0-\mathbb{E}[Y_0]<y_m\le Y_1-\mathbb{E}[Y_1])dy_1\dots dy_m\\
&-\int_{{\Omega_Y}^m} \mathbb{P}(Y_1-\mathbb{E}[Y_1]<y_1\le Y_0-\mathbb{E}[Y_0], Y_1-\mathbb{E}[Y_1]<y_2\le Y_0-\mathbb{E}[Y_0],\dots,\\
&\qquad\qquad Y_1-\mathbb{E}[Y_1]<y_m\le Y_0-\mathbb{E}[Y_0])dy_1\dots dy_m\\
&\ge \int_{{\Omega_Y}^m} \bar{l}(y_1,\dots,y_m;1,0)dy_1\dots dy_m-\int_{{\Omega_Y}^m} \bar{u}(y_1,\dots,y_m;0,1)dy_1\dots dy_m,
\end{aligned}
\tag{72}
$$

and

$$
\overline{\mu}^{(m)}=\int_{{\Omega_Y}^m} \mathbb{P}(Y_0-\mathbb{E}[Y_0]<y_1\le Y_1-\mathbb{E}[Y_1], Y_0-\mathbb{E}[Y_0]<y_2\le Y_1-\mathbb{E}[Y_1],\dots,
$$

$$
\begin{aligned}
&Y_0 - \mathbb{E}[Y_0] < y_m \leq Y_1 - \mathbb{E}[Y_1])dy_1 \dots dy_m \\
&+ (-1)^m \int_{{\Omega_Y}^m} \mathbb{P}(Y_1 - \mathbb{E}[Y_1] < y_1 \leq Y_0 - \mathbb{E}[Y_0], Y_1 - \mathbb{E}[Y_1] < y_2 \leq Y_0 - \mathbb{E}[Y_0], \dots, \\
&\qquad Y_1 - \mathbb{E}[Y_1] < y_m \leq Y_0 - \mathbb{E}[Y_0])dy_1 \dots dy_m \\
&= \int_{{\Omega_Y}^m} \mathbb{P}(Y_0 - \mathbb{E}[Y_0] < y_1 \leq Y_1 - \mathbb{E}[Y_1], Y_0 - \mathbb{E}[Y_0] < y_2 \leq Y_1 - \mathbb{E}[Y_1], \dots, \\
&\qquad Y_0 - \mathbb{E}[Y_0] < y_m \leq Y_1 - \mathbb{E}[Y_1])dy_1 \dots dy_m \\
&- \int_{{\Omega_Y}^m} \mathbb{P}(Y_1 - \mathbb{E}[Y_1] < y_1 \leq Y_0 - \mathbb{E}[Y_0], Y_1 - \mathbb{E}[Y_1] < y_2 \leq Y_0 - \mathbb{E}[Y_0], \dots, \\
&\qquad Y_1 - \mathbb{E}[Y_1] < y_m \leq Y_0 - \mathbb{E}[Y_0])dy_1 \dots dy_m \\
&\leq \int_{{\Omega_Y}^m} \bar{u}(y_1, \dots, y_m; 1, 0)dy_1 \dots dy_m - \int_{{\Omega_Y}^m} \bar{l}(y_1, \dots, y_m; 0, 1)dy_1 \dots dy_m. \qquad (73)
\end{aligned}
$$

□

We can calculate the bounds of the skewness as follows:

$$
\frac{\bar{\sigma}_L^{(3)}}{\bar{\sigma}_U^{(2)\frac{3}{2}}}\mathbb{I}(\bar{\sigma}_L^{(3)} \geq 0) + \frac{\bar{\sigma}_L^{(3)}}{\bar{\sigma}_L^{(2)\frac{3}{2}}}\mathbb{I}(\bar{\sigma}_L^{(3)} < 0) \leq \frac{\overline{\mu}^{(3)}}{\overline{\mu}^{(2)\frac{3}{2}}} \leq \frac{\bar{\sigma}_U^{(3)}}{\bar{\sigma}_L^{(2)\frac{3}{2}}}\mathbb{I}(\bar{\sigma}_U^{(3)} \geq 0) + \frac{\bar{\sigma}_U^{(3)}}{\bar{\sigma}_U^{(2)\frac{3}{2}}}\mathbb{I}(\bar{\sigma}_U^{(3)} < 0), \qquad (74)
$$

and the bounds of the kurtosis as follows:

$$
\frac{\bar{\sigma}_L^{(4)}}{\bar{\sigma}_U^{(2)2}} \leq \frac{\overline{\mu}^{(4)}}{\overline{\mu}^{(2)^2}} \leq \frac{\bar{\sigma}_U^{(4)}}{\bar{\sigma}_L^{(2)2}}. \qquad (75)
$$

C IDENTIFICATION AND BOUNDS OF THE CENTRAL PRODUCT MOMENTS OF CAUSAL EFFECTS

In this section, we discuss the central product moments of causal effects. We make the following assumption:

Assumption 6 (Existence of integrals)**.** *Under SCM $\mathcal{M}$, given $m \geq 1$, $\overline{\rho}_{i,j;k,h} < \infty$ and $\int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1 \leq Y_i - \mathbb{E}[Y_i], Y_h - \mathbb{E}[Y_h] < y_2 \leq Y_k - \mathbb{E}[Y_k])dy_1dy_2 < \infty$ hold for any $i, j, k, h \in \{1, \dots, R\}$.*

Identification of the central product moment of causal effects.

Theorem 7 (Identification of the central product moment of causal effects)**.** *Under SCM $\mathcal{M}$ and Assumptions 1 2, and 6, for any $i, j, k, h \in \{1, \dots, R\}$, $\overline{\rho}_{i,j;k,h}$ is identifiable by $\bar{\sigma}(i, j; k, h)$, where*

$$
\begin{aligned}
\bar{\sigma}(i,j;k,h) = &\int_{{\Omega_Y}^2} \max\Big\{ \min\{\mathbb{P}(Y - \mathbb{E}[Y|X=j] < y_1|X=j), \mathbb{P}(Y - \mathbb{E}[Y|X=h] < y_2|X=h)\} \\
&- \max\{\mathbb{P}(Y - \mathbb{E}[Y|X=i] < y_1|X=i), \mathbb{P}(Y - \mathbb{E}[Y|X=k] < y_2|X=k)\}, 0\Big\}dy_1dy_2 \\
&- \int_{{\Omega_Y}^2} \max\Big\{ \min\{\mathbb{P}(Y - \mathbb{E}[Y|X=i] < y_1|X=i), \mathbb{P}(Y - \mathbb{E}[Y|X=h] < y_2|X=h)\} \\
&- \max\{\mathbb{P}(Y - \mathbb{E}[Y|X=j] < y_1|X=j), \mathbb{P}(Y - \mathbb{E}[Y|X=k] < y_2|X=k)\}, 0\Big\}dy_1dy_2 \\
&- \int_{{\Omega_Y}^2} \max\Big\{ \min\{\mathbb{P}(Y - \mathbb{E}[Y|X=j] < y_1|X=j), \mathbb{P}(Y - \mathbb{E}[Y|X=k] < y_2|X=k)\} \\
&- \max\{\mathbb{P}(Y - \mathbb{E}[Y|X=i] < y_1|X=i), \mathbb{P}(Y - \mathbb{E}[Y|X=h] < y_2|X=h)\}, 0\Big\}dy_1dy_2 \\
&+ \int_{{\Omega_Y}^2} \max\Big\{ \min\{\mathbb{P}(Y - \mathbb{E}[Y|X=i] < y_1|X=i), \mathbb{P}(Y - \mathbb{E}[Y|X=k] < y_2|X=k)\} \\
&- \max\{\mathbb{P}(Y - \mathbb{E}[Y|X=j] < y_1|X=j), \mathbb{P}(Y - \mathbb{E}[Y|X=h] < y_2|X=h)\}, 0\Big\}dy_1dy_2. \qquad (76)
\end{aligned}
$$

Proof. $\mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1 \leq Y_i - \mathbb{E}[Y_i], Y_h - \mathbb{E}[Y_h] < y_2 \leq Y_k - \mathbb{E}[Y_k])$ are identifiable by $\max\{\min\{\mathbb{P}(Y - \mathbb{E}[Y|X = j] < y_1|X = j), \mathbb{P}(Y - \mathbb{E}[Y|X = h] < y_2|X = h)\} - \max\{\mathbb{P}(Y - \mathbb{E}[Y|X = i] < y_1|X = i), \mathbb{P}(Y - \mathbb{E}[Y|X = k] < y_2|X = k)\}, 0\}$ $(i, j, k, h \in \{1, \ldots, R\})$ respectively by rewriting f_Y as $f(x, U^Y) = Y_x - \mathbb{E}[Y_x]$ in Theorem 5.2 in [Kawakami et al., 2024a] under Assumptions 1 and 2. Then, we have Eq. (76). □

Bounding the central product moments of causal effects.

Assumption 2 may sometimes be considered implausible by researchers. Therefore, we derive bounds for the moments of causal effects without relying on Assumption 2.

Lemma 6. *Under SCM $\mathcal{M}$ and Assumptions 1 and 6, for any $i, j, k, h \in \{1, \ldots, R\}$, we have $l(y_1, y_3; i, j, k, h) \leq \mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1 \leq Y_i - \mathbb{E}[Y_i], Y_h - \mathbb{E}[Y_h] < y_2 \leq Y_k - \mathbb{E}[Y_k]) \leq u(y_1, y_3; i, j, k, h)$, where*

$$l(y_1, y_2; i, j, k, h) = \max\Big\{\mathbb{P}(Y - \mathbb{E}[Y|X = j] < y_1|X = j) - \mathbb{P}(Y - \mathbb{E}[Y|X = i] < y_1|X = i) + \mathbb{P}(Y - \mathbb{E}[Y|X = h] < y_2|X = h) - \mathbb{P}(Y - \mathbb{E}[Y|X = k] < y_2|X = k) - 1, 0\Big\}, \quad (77)$$

$$u(y_1, y_2; i, j, k, h) = \min\Big\{\min\{\mathbb{P}(Y - \mathbb{E}[Y|X = j] < y_1|X = j), \mathbb{P}(Y - \mathbb{E}[Y|X = h] < y_2|X = h)\}, 1 - \max\{\mathbb{P}(Y - \mathbb{E}[Y|X = i] < y_1|X = i), \mathbb{P}(Y - \mathbb{E}[Y|X = k] < y_2|X = k)\}\Big\} \quad (78)$$

for each $i, j, k, h \in \{1, \ldots, R\}$ and $y_1, y_2 \in \Omega_Y$.

Proof. From Fréchet inequalities [Fréchet, 1935, 1960], for any $i, j, k, h \in \{1, \ldots, R\}$, we have

$$\begin{aligned}
&\mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1 \leq Y_i - \mathbb{E}[Y_i], Y_h - \mathbb{E}[Y_h] < y_2 \leq Y_k - \mathbb{E}[Y_k]) \\
&= \mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1, y_1 \leq Y_i - \mathbb{E}[Y_i], Y_h - \mathbb{E}[Y_h] < y_2, y_2 \leq Y_k - \mathbb{E}[Y_k]) \\
&\geq \max\{\mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1) + \mathbb{P}(y_1 \leq Y_i - \mathbb{E}[Y_i]) + \mathbb{P}(Y_h - \mathbb{E}[Y_h] < y_2) + \mathbb{P}(y_2 \leq Y_k - \mathbb{E}[Y_k]) - 3, 0\} \\
&= \max\{\mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1) - \mathbb{P}(Y_i - \mathbb{E}[Y_i] < y_1) + \mathbb{P}(Y_h - \mathbb{E}[Y_h] < y_2) - \mathbb{P}(Y_k - \mathbb{E}[Y_k] < y_2) - 1, 0\} \\
&= \max\Big\{\mathbb{P}(Y - \mathbb{E}[Y|X = j] < y_1|X = j) - \mathbb{P}(Y - \mathbb{E}[Y|X = i] < y_1|X = i) \\
&\qquad + \mathbb{P}(Y - \mathbb{E}[Y|X = h] < y_2|X = h) - \mathbb{P}(Y - \mathbb{E}[Y|X = k] < y_2|X = k) - 1, 0\Big\} \qquad (79)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1 \leq Y_i - \mathbb{E}[Y_i], Y_h - \mathbb{E}[Y_h] < y_2 \leq Y_k - \mathbb{E}[Y_k]) \\
&= \mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1, y_1 \leq Y_i - \mathbb{E}[Y_i], Y_h - \mathbb{E}[Y_h] < y_2, y_2 \leq Y_k - \mathbb{E}[Y_k]) \\
&\leq \min\Big\{\min\{\mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1), \mathbb{P}(Y_h - \mathbb{E}[Y_h] < y_2)\}, 1 - \max\{\mathbb{P}(Y_i - \mathbb{E}[Y_i] < y_1), \mathbb{P}(Y_k - \mathbb{E}[Y_k] < y_2)\}\Big\} \\
&= \min\Big\{\min\{\mathbb{P}(Y - \mathbb{E}[Y|X = j] < y_1|X = j), \mathbb{P}(Y - \mathbb{E}[Y|X = h] < y_2|X = h)\}, \\
&\qquad 1 - \max\{\mathbb{P}(Y - \mathbb{E}[Y|X = i] < y_1|X = i), \mathbb{P}(Y - \mathbb{E}[Y|X = k] < y_2|X = k)\}\Big\}. \qquad (80)
\end{aligned}$$

□

Theorem 8 (Bounds of the central product moment of causal effects)**.** *Under SCM $\mathcal{M}$ and Assumptions 1 and 6, for any $i, j, k, h \in \{1, \ldots, R\}$, we have $\bar{\sigma}_L(i, j; k, h) \leq \overline{\rho}_{i,j;k,h} \leq \bar{\sigma}_U(i, j; k, h)$, where*

$$\begin{aligned}
\bar{\sigma}_L(i, j; k, h) = &\int_{{\Omega_Y}^2} \bar{l}(y_1, y_2; i, j, k, h) dy_1 dy_2 - \int_{{\Omega_Y}^2} \bar{u}(y_1, y_2; j, i, k, h) dy_1 dy_2 \\
&- \int_{{\Omega_Y}^2} \bar{u}(y_1, y_2; i, j, h, k) dy_1 dy_2 + \int_{{\Omega_Y}^2} \bar{l}(y_1, y_2; j, i, h, k) dy_1 dy_2, \qquad (81)
\end{aligned}$$

$$\begin{aligned}
\bar{\sigma}_U(i, j; k, h) = &\int_{{\Omega_Y}^2} \bar{u}(y_1, y_2; i, j, k, h) dy_1 dy_2 - \int_{{\Omega_Y}^2} \bar{l}(y_1, y_2; j, i, k, h) dy_1 dy_2 \\
&- \int_{{\Omega_Y}^2} \bar{l}(y_1, y_2; i, j, h, k) dy_1 dy_2 + \int_{{\Omega_Y}^2} \bar{u}(y_1, y_2; j, i, h, k) dy_1 dy_2. \qquad (82)
\end{aligned}$$

Proof. For any $i, j, k, h \in \{1, \dots, R\}$, we have

$$\begin{aligned}
\overline{\rho}_{i,j;k,h} &= \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1 \le Y_i - \mathbb{E}[Y_i], Y_h - \mathbb{E}[Y_h] < y_2 \le Y_k - \mathbb{E}[Y_k])dy_1dy_2 \\
&- \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_i - \mathbb{E}[Y_i] < y_1 \le Y_j - \mathbb{E}[Y_j], Y_h - \mathbb{E}[Y_h] < y_2 \le Y_k - \mathbb{E}[Y_k])dy_1dy_2 \\
&- \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1 \le Y_i - \mathbb{E}[Y_i], Y_k - \mathbb{E}[Y_k] < y_2 \le Y_h - \mathbb{E}[Y_h])dy_1dy_2 \\
&+ \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_i - \mathbb{E}[Y_i] < y_1 \le Y_j - \mathbb{E}[Y_j], Y_k - \mathbb{E}[Y_k] < y_2 \le Y_h - \mathbb{E}[Y_h])dy_1dy_2 \\
&\ge \int_{{\Omega_Y}^2} \bar{l}(y_1, y_2; i, j, k, h)dy_1dy_2 - \int_{{\Omega_Y}^2} \bar{u}(y_1, y_2; j, i, k, h)dy_1dy_2 \\
&\qquad - \int_{{\Omega_Y}^2} \bar{u}(y_1, y_2; i, j, h, k)dy_1dy_2 + \int_{{\Omega_Y}^2} \bar{l}(y_1, y_2; j, i, h, k)dy_1dy_2
\end{aligned} \tag{83}$$

and

$$\begin{aligned}
\overline{\rho}_{i,j;k,h} &= \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1 \le Y_i - \mathbb{E}[Y_i], Y_h - \mathbb{E}[Y_h] < y_2 \le Y_k - \mathbb{E}[Y_k])dy_1dy_2 \\
&- \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_i - \mathbb{E}[Y_i] < y_1 \le Y_j - \mathbb{E}[Y_j], Y_h - \mathbb{E}[Y_h] < y_2 \le Y_k - \mathbb{E}[Y_k])dy_1dy_2 \\
&- \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_j - \mathbb{E}[Y_j] < y_1 \le Y_i - \mathbb{E}[Y_i], Y_k - \mathbb{E}[Y_k] < y_2 \le Y_h - \mathbb{E}[Y_h])dy_1dy_2 \\
&+ \int_{\Omega_Y}\int_{\Omega_Y} \mathbb{P}(Y_i - \mathbb{E}[Y_i] < y_1 \le Y_j - \mathbb{E}[Y_j], Y_k - \mathbb{E}[Y_k] < y_2 \le Y_h - \mathbb{E}[Y_h])dy_1dy_2 \\
&\le \int_{{\Omega_Y}^2} \bar{u}(y_1, y_2; i, j, k, h)dy_1dy_2 - \int_{{\Omega_Y}^2} \bar{l}(y_1, y_2; j, i, k, h)dy_1dy_2 \\
&\qquad - \int_{{\Omega_Y}^2} \bar{l}(y_1, y_2; i, j, h, k)dy_1dy_2 + \int_{{\Omega_Y}^2} \bar{u}(y_1, y_2; j, i, h, k)dy_1dy_2.
\end{aligned} \tag{84}$$

□

Denoting the bound of the moments of causal effects $Y_i - Y_j$ by $\bar{\sigma}_L^{(m)}(i,j)$ and $\bar{\sigma}_L^{(m)}(i,j)$, i.e., $\bar{\sigma}_U^{(m)}(i,j) \le \mathbb{E}[\{(Y_i - Y_j) - (\mathbb{E}[Y_i] - \mathbb{E}[Y_j])\}^m] \le \bar{\sigma}_U^{(m)}(i,j)$, we can calculate the bounds of the correlation as follows:

$$\begin{aligned}
&\frac{\bar{\sigma}_L(i,j;k,h)}{\sqrt{\bar{\sigma}_U^{(2)}(i,j)}\sqrt{\bar{\sigma}_U^{(2)}(k,h)}}\mathbb{I}(\bar{\sigma}_L(i,j;k,h) \ge 0) + \frac{\bar{\sigma}_L(i,j;k,h)}{\sqrt{\bar{\sigma}_L^{(2)}(i,j)}\sqrt{\bar{\sigma}_L^{(2)}(k,h)}}\mathbb{I}(\bar{\sigma}_L(i,j;k,h) < 0) \\
&\le \overline{\tau}_{i,j;k,h} \le \frac{\bar{\sigma}_U(i,j;k,h)}{\sqrt{\bar{\sigma}_L^{(2)}(i,j)}\sqrt{\bar{\sigma}_L^{(2)}(k,h)}}\mathbb{I}(\bar{\sigma}_U(i,j;k,h) \ge 0) + \frac{\bar{\sigma}_U(i,j;k,h)}{\sqrt{\bar{\sigma}_U^{(2)}(i,j)}\sqrt{\bar{\sigma}_U^{(2)}(k,h)}}\mathbb{I}(\bar{\sigma}_U(i,j;k,h) < 0).
\end{aligned} \tag{85}$$

D CONDITIONAL MOMENTS OF CAUSAL EFFECTS

We consider the following SCM, $\mathcal{M}$:

$$Y := f_Y(X, W, U^Y), \;\; X := f_X(W, U^X), \;\; W := f_W(U^W) \tag{86}$$

where U^Y, U^X, and U^W are latent exogenous variables. Y is a continuous variable, and X and W are binary, discrete, or continuous variables. W are the subject's covariates. They also provided the identification assumptions of the PNS for continuous treatment and outcome as below:

Assumption 7 (Conditional exogeneity)**.** $Y_x \perp\!\!\!\perp X | W$ *for all* $x \in \Omega_X$.

If X is binary, i.e., $\Omega_X = \{0, 1\}$, Assumption 7 reduces to $Y_1 \perp\!\!\!\perp X|W$ and $Y_0 \perp\!\!\!\perp X|W$.

Assumption 8 (Conditional monotonicity over f_Y)**.** *The function $f_Y(x, w, U^Y)$ is either (i) monotonic increasing on U^Y with $\leq$ for all $x \in \Omega_X$ and $w \in \Omega_W$ almost surely w.r.t. $\mathbb{P}_{U^Y}$, or (ii) monotonic decreasing on U^Y for all $x \in \Omega_X$ and $w \in \Omega_W$ almost surely w.r.t. $\mathbb{P}_{U^Y}$.*

Kawakami et al. [2024a] provided another Assumption (conditional monotonicity over Y_x), which is equivalent to Assumption 8 under Assumption 3.6 in [Kawakami et al., 2024a].

We provide similar definitions of the conditional moments of causal effects with the subjects' covariates W.

Definition 6 (The conditional moment of causal effects)**.** *For each $m \geq 1$, the conditional m-th moment of $Y_1 - Y_0$ is defined as*

$$\mathbb{E}\left[(Y_1 - Y_0)^m \middle| W = w\right]. \tag{87}$$

Under SCM $\mathcal{M}$ and Assumptions 3, 7, and 8, given $m \geq 1$, it is identified by Eq. (6) conditioned on $W = w$. Under SCM $\mathcal{M}$ and Assumptions 3 and 7, given $m \geq 1$, it is bounded by Eqs. (9) $\sim$ (12) conditioned on $W = w$.

Definition 7 (The conditional central moment of causal effects)**.** *For each $m \geq 1$, we define the conditional m-th central moment of $Y_1 - Y_0$ as*

$$\mathbb{E}\left[\Big\{(Y_1 - Y_0) - (\mathbb{E}[Y_1] - \mathbb{E}[Y_0])\Big\}^m \middle| W = w\right]. \tag{88}$$

Under SCM $\mathcal{M}$ and Assumptions 5, 7, and 8, given $m \geq 1$, it is identified by Eq. (56) conditioned on $W = w$. Under SCM $\mathcal{M}$ and Assumptions 5 and 7, given $m \geq 1$, it is bounded by Eqs. (66) $\sim$ (69) conditioned on $W = w$.

Definition 8 (The conditional product moment of causal effects)**.** *We define the conditional product moment of causal effects as*

$$\mathbb{E}\left[(Y_i - Y_j)(Y_k - Y_h) \middle| W = w\right]. \tag{89}$$

Under SCM $\mathcal{M}$ and Assumptions 4, 7, and 8, given $m \geq 1$, it is identified by Eq. (18) conditioned on $W = w$. Under SCM $\mathcal{M}$ and Assumptions 4 and 7, given $m \geq 1$, it is bounded by Eqs. (21) $\sim$ (22) conditioned on $W = w$.

Definition 9 (The conditional covariance of causal effects)**.** *We define the conditional covariance (central product moment of causal effects) as*

$$\mathbb{E}\left[\Big\{(Y_i - Y_j) - (\mathbb{E}[Y_i] - \mathbb{E}[Y_j])\Big\}\Big\{(Y_k - Y_h) - (\mathbb{E}[Y_k] - \mathbb{E}[Y_h])\Big\} \middle| W = w\right]. \tag{90}$$

Under SCM $\mathcal{M}$ and Assumptions 6, 7, and 8, given $m \geq 1$, it is identified by Eq. (76) conditioned on $W = w$. Under SCM $\mathcal{M}$ and Assumptions 6 and 7, given $m \geq 1$, it is bounded by Eqs. (81) $\sim$ (82) conditioned on $W = w$.

E CONSISTENCY OF ESTIMATORS

In this appendix, we provide the details and consistency of all estimators in the body of the paper.

Details of the estimators in the body of the paper. The empirical CDFs and expectations are given by

$$\hat{\mathbb{P}}(Y < y|X = x) = \frac{\sum_{i=1}^N \mathbb{I}(Y_i < y, X_i = x)}{\sum_{i=1}^N \mathbb{I}(X_i = x)}, \quad \hat{\mathbb{E}}[Y|X = x] = \frac{\sum_{i=1}^N Y_i \mathbb{I}(X_i = x)}{\sum_{i=1}^N \mathbb{I}(X_i = x)} \tag{91}$$

for any $x \in \Omega_X$ and $y \in \Omega_Y$. We generate $\{y^1_{k^1}\}_{k^1=1}^{N_1}, \{y^2_{k^2}\}_{k^2=1}^{N_2}, \ldots, \{y^m_{k^m}\}_{k^m=1}^{N_m}$ by i.i.d. sampling from a uniform distribution $U[\min(Y), \max(Y)]$ for Monte Carlo integration.

For $m = 1, \ldots,$ the estimators $\hat{\sigma}^{(m)}$ and $\hat{\bar{\sigma}}^{(m)}$ are

$$\hat{\sigma}^{(m)} = \frac{(b-a)^m}{N_1 \ldots N_m} \sum_{k^1=1}^{N_1} \cdots \sum_{k^m=1}^{N_m} \max\Big\{ \min_{p=1,\ldots,m} \{\hat{\mathbb{P}}(Y < y^p_{k^p}|X = 0)\} - \max_{p=1,\ldots,m} \{\hat{\mathbb{P}}(Y < y^p_{k^p}|X = 1)\}, 0 \Big\}$$

$$+(-1)^m\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=1)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=0)\},0\Big\},$$

$$\hat{\bar{\sigma}}^{(m)}=\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=0]<y^p_{k^p}|X=0)\}$$
$$-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=1]<y^p_{k^p}|X=1)\},0\Big\}$$
$$+(-1)^m\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=1]<y^p_{k^p}|X=1)\}$$
$$-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=0]<y^p_{k^p}|X=0)\},0\Big\}. \quad (92)$$

For $m=1,\dots$, the estimators $\hat{\sigma}_L^{(m)}$ and $\hat{\sigma}_U^{(m)}$ are given below.

(A). When m is an even number,

$$\hat{\sigma}_L^{(m)}=\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\Big\{\hat{l}(y^1_{k^1},\dots,y^m_{k^m};1,0)+\hat{l}(y^1_{k^1},\dots,y^m_{k^m};0,1)\Big\}, \quad (93)$$

$$\hat{\sigma}_U^{(m)}=\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\Big\{\hat{u}(y^1_{k^1},\dots,y^m_{k^m};1,0)+\hat{u}(y^1_{k^1},\dots,y^m_{k^m};0,1)\Big\}, \quad (94)$$

where

$$\hat{l}(y_1,\dots,y_m;i,j)=\max\Big\{\sum_{p=1,\dots,m}\hat{\mathbb{P}}(Y<y_p|X=j)-\sum_{p=1,\dots,m}\hat{\mathbb{P}}(Y<y_p|X=i)-m+1,0\Big\}, \quad (95)$$

$$\hat{u}(y_1,\dots,y_m;i,j)=\min\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=j)\},1-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=i)\}\Big\}. \quad (96)$$

(B). When m is an odd number,

$$\hat{\sigma}_L^{(m)}=\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\Big\{\hat{l}(y^1_{k^1},\dots,y^m_{k^m};1,0)-\hat{u}(y^1_{k^1},\dots,y^m_{k^m};0,1)\Big\}, \quad (97)$$

$$\hat{\sigma}_U^{(m)}=\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\Big\{\hat{u}(y^1_{k^1},\dots,y^m_{k^m};1,0)-\hat{l}(y^1_{k^1},\dots,y^m_{k^m};0,1)\Big\}. \quad (98)$$

For $m=1,\dots$, the estimators $\hat{\bar{\sigma}}_L^{(m)}$ and $\hat{\bar{\sigma}}_U^{(m)}$ are given below.

(A). When m is an even number,

$$\hat{\bar{\sigma}}_L^{(m)}=\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\Big\{\hat{\bar{l}}(y^1_{k^1},\dots,y^m_{k^m};1,0)+\hat{\bar{l}}(y^1_{k^1},\dots,y^m_{k^m};0,1)\Big\}, \quad (99)$$

$$\hat{\bar{\sigma}}_U^{(m)}=\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\Big\{\hat{\bar{u}}(y^1_{k^1},\dots,y^m_{k^m};1,0)+\hat{\bar{u}}(y^1_{k^1},\dots,y^m_{k^m};0,1)\Big\}, \quad (100)$$

where

$$\hat{\bar{l}}(y_1,\dots,y_m;i,j)=\max\Big\{\sum_{p=1,\dots,m}\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=j]<y_p|X=j)$$
$$-\sum_{p=1,\dots,m}\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=i]<y_p|X=i)-m+1,0\Big\}, \quad (101)$$

$$\hat{\bar{u}}(y_1,\ldots,y_m;i,j)=\min\Big\{\min_{p=1,\ldots,m}\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=j]<y_p|X=j)\},\\ 1-\max_{p=1,\ldots,m}\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=i]<y_p|X=i)\}\Big\}. \tag{102}$$

(B). When m is an odd number,

$$\hat{\bar{\sigma}}_L^{(m)}=\frac{(b-a)^m}{N_1\ldots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\Big\{\hat{\bar{l}}(y^1_{k^1},\ldots,y^m_{k^m};1,0)-\hat{\bar{u}}(y^1_{k^1},\ldots,y^m_{k^m};0,1)\Big\}, \tag{103}$$

$$\hat{\bar{\sigma}}_U^{(m)}=\frac{(b-a)^m}{N_1\ldots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\Big\{\hat{\bar{u}}(y^1_{k^1},\ldots,y^m_{k^m};1,0)-\hat{\bar{l}}(y^1_{k^1},\ldots,y^m_{k^m};0,1)\Big\}. \tag{104}$$

The estimators $\hat{\sigma}(i,j;k,h)$ and $\hat{\bar{\sigma}}(i,j;k,h)$ are

$$\begin{aligned}\hat{\sigma}(i,j;k,h)=&\frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\max\Big\{\min\{\hat{\mathbb{P}}(Y<y^1_{k^1}|X=j),\hat{\mathbb{P}}(Y<y^2_{k^2}|X=h)\}\\ &\qquad-\max\{\hat{\mathbb{P}}(Y<y^1_{k^1}|X=i),\hat{\mathbb{P}}(Y<y^2_{k^2}|X=k)\},0\Big\}\\ &-\frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\max\Big\{\min\{\hat{\mathbb{P}}(Y<y^1_{k^1}|X=i),\hat{\mathbb{P}}(Y<y^2_{k^2}|X=h)\}\\ &\qquad-\max\{\hat{\mathbb{P}}(Y<y^1_{k^1}|X=j),\hat{\mathbb{P}}(Y<y^2_{k^2}|X=k)\},0\Big\}\\ &-\frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\max\Big\{\min\{\hat{\mathbb{P}}(Y<y^1_{k^1}|X=j),\hat{\mathbb{P}}(Y<y^2_{k^2}|X=k)\}\\ &\qquad-\max\{\hat{\mathbb{P}}(Y<y^1_{k^1}|X=i),\hat{\mathbb{P}}(Y<y^2_{k^2}|X=h)\},0\Big\}\\ &+\frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\max\Big\{\min\{\hat{\mathbb{P}}(Y<y^1_{k^1}|X=i),\hat{\mathbb{P}}(Y<y^2_{k^2}|X=k)\}\\ &\qquad-\max\{\hat{\mathbb{P}}(Y<y^1_{k^1}|X=j),\hat{\mathbb{P}}(Y<y^2_{k^2}|X=h)\},0\Big\},\end{aligned} \tag{105}$$

$$\begin{aligned}\hat{\bar{\sigma}}(i,j;k,h)=&\frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\max\Big\{\min\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=j]<y^1_{k^1}|X=j),\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=h]<y^2_{k^2}|X=h)\}\\ &\qquad-\max\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=i]<y^1_{k^1}|X=i),\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=k]<y^2_{k^2}|X=k)\},0\Big\}\\ &-\frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\max\Big\{\min\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=i]<y^1_{k^1}|X=i),\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=h]<y^2_{k^2}|X=h)\}\\ &\qquad-\max\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=j]<y^1_{k^1}|X=j),\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=k]<y^2_{k^2}|X=k)\},0\Big\}\\ &-\frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\max\Big\{\min\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=j]<y^1_{k^1}|X=j),\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=k]<y^2_{k^2}|X=k)\}\\ &\qquad-\max\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=i]<y^1_{k^1}|X=i),\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=h]<y^2_{k^2}|X=h)\},0\Big\}\\ &+\frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\max\Big\{\min\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=i]<y^1_{k^1}|X=i),\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=k]<y^2_{k^2}|X=k)\}\\ &\qquad-\max\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=j]<y^1_{k^1}|X=j),\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=h]<y^2_{k^2}|X=h)\},0\Big\}.\end{aligned} \tag{106}$$

The estimators $\hat{\sigma}_L(i,j;k,h)$ and $\hat{\sigma}_U(i,j;k,h)$ are

$$\hat{\sigma}_L(i,j;k,h) = \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{l}(y^1_{k^1},y^2_{k^2};i,j,k,h) - \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{u}(y^1_{k^1},y^2_{k^2};j,i,k,h)$$
$$- \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{u}(y^1_{k^1},y^2_{k^2};i,j,h,k) + \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{l}(y^1_{k^1},y^2_{k^2};j,i,h,k), \tag{107}$$

$$\hat{\sigma}_U(i,j;k,h) = \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{u}(y^1_{k^1},y^2_{k^2};i,j,k,h) - \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{l}(y^1_{k^1},y^2_{k^2};j,i,k,h)$$
$$- \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{l}(y^1_{k^1},y^2_{k^2};i,j,h,k) + \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{u}(y^1_{k^1},y^2_{k^2};j,i,h,k), \tag{108}$$

where

$$\hat{l}(y_1,y_2;i,j,k,h)$$
$$= \max\Big\{\hat{\mathbb{P}}(Y<y_1|X=j) - \hat{\mathbb{P}}(Y<y_1|X=i) + \hat{\mathbb{P}}(Y<y_2|X=h) - \hat{\mathbb{P}}(Y<y_2|X=k) - 1, 0\Big\}, \tag{109}$$
$$\hat{u}(y_1,y_2;i,j,k,h)$$
$$= \min\Big\{\min\{\hat{\mathbb{P}}(Y<y_1|X=j), \hat{\mathbb{P}}(Y<y_2|X=h)\}, 1 - \max\{\hat{\mathbb{P}}(Y<y_1|X=i), \hat{\mathbb{P}}(Y<y_2|X=k)\}\Big\}. \tag{110}$$

The estimators $\hat{\bar{\sigma}}_L(i,j;k,h)$ and $\hat{\bar{\sigma}}_U(i,j;k,h)$ are

$$\hat{\bar{\sigma}}_L(i,j;k,h) = \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{\bar{l}}(y^1_{k^1},y^2_{k^2};i,j,k,h) - \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{\bar{u}}(y^1_{k^1},y^2_{k^2};j,i,k,h)$$
$$- \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{\bar{u}}(y^1_{k^1},y^2_{k^2};i,j,h,k) + \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{\bar{l}}(y^1_{k^1},y^2_{k^2};j,i,h,k), \tag{111}$$

$$\hat{\bar{\sigma}}_U(i,j;k,h) = \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{\bar{u}}(y^1_{k^1},y^2_{k^2};i,j,k,h) - \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{\bar{l}}(y^1_{k^1},y^2_{k^2};j,i,k,h)$$
$$- \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{\bar{l}}(y^1_{k^1},y^2_{k^2};i,j,h,k) + \frac{(b-a)^2}{N_1N_2}\sum_{k^1=1}^{N_1}\sum_{k^2=1}^{N_2}\hat{\bar{u}}(y^1_{k^1},y^2_{k^2};j,i,h,k), \tag{112}$$

where

$$\hat{\bar{l}}(y_1,y_2;i,j,k,h) = \max\Big\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=j]<y_1|X=j) - \hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=i]<y_1|X=i)$$
$$+ \hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=h]<y_2|X=h) - \hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=k]<y_2|X=k) - 1, 0\Big\}, \tag{113}$$
$$\hat{\bar{u}}(y_1,y_2;i,j,k,h) = \min\Big\{\min\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=j]<y_1|X=j), \hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=h]<y_2|X=h)\},$$
$$1 - \max\{\hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=i]<y_1|X=i), \hat{\mathbb{P}}(Y-\hat{\mathbb{E}}[Y|X=k]<y_2|X=k)\}\Big\}. \tag{114}$$

Consistency of the estimators for the moments of causal effects. First, the empirical CDFs and expectations follow $O_p\left(1/\sqrt{N}\right)$ for any $x \in \Omega_X$ and $y \in \Omega_Y$. From the delta method [Doob, 1935, Fang and Santos, 2018], both $\max\{\min_{p=1,\ldots,m}\{\hat{\mathbb{P}}(Y<y_p|X=0)\} - \max_{p=1,\ldots,m}\{\hat{\mathbb{P}}(Y<y_p|X=1)\}, 0\}$ and $\max\{\min_{p=1,\ldots,m}\{\hat{\mathbb{P}}(Y<y_p|X=1)\} - \max_{p=1,\ldots,m}\{\hat{\mathbb{P}}(Y<y_p|X=0)\}, 0\}$ follow $O_p\left(1/\sqrt{N^m}\right)$, almost surely w.r.t. Ω_Y^m. For any $m = 1, \ldots,$ we have

$$\left|\hat{\sigma}^{(m)} - \sigma^{(m)}\right|$$
$$= \left|\left\{\frac{(b-a)^m}{N_1\ldots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\max\Big\{\min_{p=1,\ldots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=0)\} - \max_{p=1,\ldots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=1)\}, 0\Big\}\right.$$

$$+(-1)^m\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=1)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=0)\},0\Big\}\Bigg\}$$

$$-\Bigg\{\sigma^{(m)}=\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=0)\}-\max_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=1)\},0\Big\}dy_1\dots dy_m$$

$$+(-1)^m\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=1)\}-\max_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=0)\},0\Big\}dy_1\dots dy_m\Bigg\}\Bigg|$$

$$=\Bigg|\Bigg\{\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=0)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=1)\},0\Big\}$$

$$+(-1)^m\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=1)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=0)\},0\Big\}\Bigg\}$$

$$-\Bigg\{\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=0)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=1)\},0\Big\}dy_1\dots dy_m$$

$$+(-1)^m\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=1)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=0)\},0\Big\}dy_1\dots dy_m\Bigg\}$$

$$+\Bigg\{\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=0)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=1)\},0\Big\}dy_1\dots dy_m$$

$$+(-1)^m\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=1)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=0)\},0\Big\}dy_1\dots dy_m\Bigg\}$$

$$-\Bigg\{\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=0)\}-\max_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=1)\},0\Big\}dy_1\dots dy_m$$

$$+(-1)^m\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=1)\}-\max_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=0)\},0\Big\}dy_1\dots dy_m\Bigg\}\Bigg|$$

$$\le\Bigg|\Bigg\{\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=0)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=1)\},0\Big\}$$

$$+(-1)^m\frac{(b-a)^m}{N_1\dots N_m}\sum_{k^1=1}^{N_1}\cdots\sum_{k^m=1}^{N_m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=1)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y^p_{k^p}|X=0)\},0\Big\}\Bigg\}$$

$$-\Bigg\{\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=0)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=1)\},0\Big\}dy_1\dots dy_m$$

$$+(-1)^m\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=1)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=0)\},0\Big\}dy_1\dots dy_m\Bigg\}\Bigg|$$

$$+\Bigg|\Bigg\{\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=0)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=1)\},0\Big\}dy_1\dots dy_m$$

$$+(-1)^m\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=1)\}-\max_{p=1,\dots,m}\{\hat{\mathbb{P}}(Y<y_p|X=0)\},0\Big\}dy_1\dots dy_m\Bigg\}$$

$$-\Bigg\{\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=0)\}-\max_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=1)\},0\Big\}dy_1\dots dy_m$$

$$+(-1)^m\int_{\Omega_Y{}^m}\max\Big\{\min_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=1)\}-\max_{p=1,\dots,m}\{\mathbb{P}(Y<y_p|X=0)\},0\Big\}dy_1\dots dy_m\Bigg\}\Bigg|$$

$$
\begin{aligned}
&= \Bigg| \Bigg\{ \frac{(b-a)^m}{N_1 \dots N_m} \sum_{k^1=1}^{N_1} \cdots \sum_{k^m=1}^{N_m} \max\Big\{ \min_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_{k^p}^p | X=0)\} - \max_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_{k^p}^p | X=1)\}, 0 \Big\} \\
&+ (-1)^m \frac{(b-a)^m}{N_1 \dots N_m} \sum_{k^1=1}^{N_1} \cdots \sum_{k^m=1}^{N_m} \max\Big\{ \min_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_{k^p}^p | X=1)\} - \max_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_{k^p}^p | X=0)\}, 0 \Big\} \Bigg\} \\
&- \Bigg\{ \int_{\Omega_Y{}^m} \max\Big\{ \min_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_p | X=0)\} - \max_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_p | X=1)\}, 0 \Big\} dy_1 \dots dy_m \\
&\qquad + (-1)^m \int_{\Omega_Y{}^m} \max\Big\{ \min_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_p | X=1)\} - \max_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_p | X=0)\}, 0 \Big\} dy_1 \dots dy_m \Bigg\} \Bigg| \\
&+ \Bigg| \int_{\Omega_Y{}^m} \Bigg\{ \max\Big\{ \min_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_p | X=0)\} - \max_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_p | X=1)\}, 0 \Big\} \\
&\qquad\qquad + (-1)^m \max\Big\{ \min_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_p | X=1)\} - \max_{p=1,\dots,m} \{\hat{\mathbb{P}}(Y < y_p | X=0)\}, 0 \Big\} \\
&- \max\Big\{ \min_{p=1,\dots,m} \{\mathbb{P}(Y < y_p | X=0)\} - \max_{p=1,\dots,m} \{\mathbb{P}(Y < y_p | X=1)\}, 0 \Big\} \\
&\qquad\qquad - (-1)^m \max\Big\{ \min_{p=1,\dots,m} \{\mathbb{P}(Y < y_p | X=1)\} - \max_{p=1,\dots,m} \{\mathbb{P}(Y < y_p | X=0)\}, 0 \Big\} \Bigg\} dy_1 \dots dy_m \Bigg| \\
&= O_p\left(\sum_{i=1}^m \frac{1}{\sqrt{N_i}} \right) + O_p\left(\frac{1}{\sqrt{N^m}} \right). \qquad (115)
\end{aligned}
$$

Then, under SCM $\mathcal{M}$ and Assumptions 1, 2, and 3, $\sigma^{(m)}$ follows $O_p\left(1/\sqrt{N^m} + \sum_{i=1}^m 1/\sqrt{N_i}\right)$. Similarly, under SCM $\mathcal{M}$ and Assumptions 1, 2, and 3, $\sigma_L^{(m)}, \sigma_U^{(m)}$ follow $O_p\left(1/\sqrt{N^m} + \sum_{i=1}^m 1/\sqrt{N_i}\right)$. Under SCM $\mathcal{M}$ and Assumptions 1, 2, and 4, $\sigma(i,j;k,h), \sigma_L(i,j;k,h), \sigma_U(i,j;k,h)$ follow $O_p\left(1/\sqrt{N^2} + \sum_{i=1}^2 1/\sqrt{N_i}\right)$. We can make $N_1, \dots, N_m$ as large as computational resources allow. Letting $N_1, \dots, N_m \to \infty$, $\sigma^{(m)}, \sigma_L^{(m)}, \sigma_U^{(m)}$ follow $O_p\left(1/\sqrt{N^m}\right)$ and $\sigma(i,j;k,h), \sigma_L(i,j;k,h), \sigma_U(i,j;k,h)$ follow $O_p\left(1/\sqrt{N^2}\right)$.

Consistency of the estimators for the central moments of causal effects. First, empirical CDFs and expectations follow $O_p\left(1/\sqrt{N}\right)$ for any $x \in \Omega_X$ and $y \in \Omega_Y$. For the central moment, we make an additional assumption.

Assumption 9. *$\mathbb{P}(Y < y | X = x)$ is differential in y for any $x \in \Omega_X$, almost surely w.r.t. Ω_Y.*

Then, under SCM $\mathcal{M}$ and Assumptions 1, 2, 5, and 9, $\sigma^{(m)}, \sigma_L^{(m)}, \sigma_U^{(m)}$ follow $O_p\left(1/\sqrt{N^{2m}} + \sum_{i=1}^m 1/\sqrt{N_i}\right)$. Under SCM $\mathcal{M}$ and Assumptions 1, 2, 6, and 9, $\sigma(i,j;k,h), \sigma_L(i,j;k,h), \sigma_U(i,j;k,h)$ follow $O_p\left(1/\sqrt{N^4} + \sum_{i=1}^2 1/\sqrt{N_i}\right)$. Letting $N_1, \dots, N_m \to \infty$, $\sigma^{(m)}, \sigma_L^{(m)}, \sigma_U^{(m)}$ follow $O_p\left(1/\sqrt{N^{2m}}\right)$ and $\sigma(i,j;k,h), \sigma_L(i,j;k,h), \sigma_U(i,j;k,h)$ follow $O_p\left(1/\sqrt{N^4}\right)$.

Thus, all estimators for the moments of causal effects are consistent.

F ADDITIONAL INFORMATION ABOUT THE NUMERICAL EXPERIMENTS

Figures 1 and 2 present the plots of the estimates obtained from the numerical experiments.

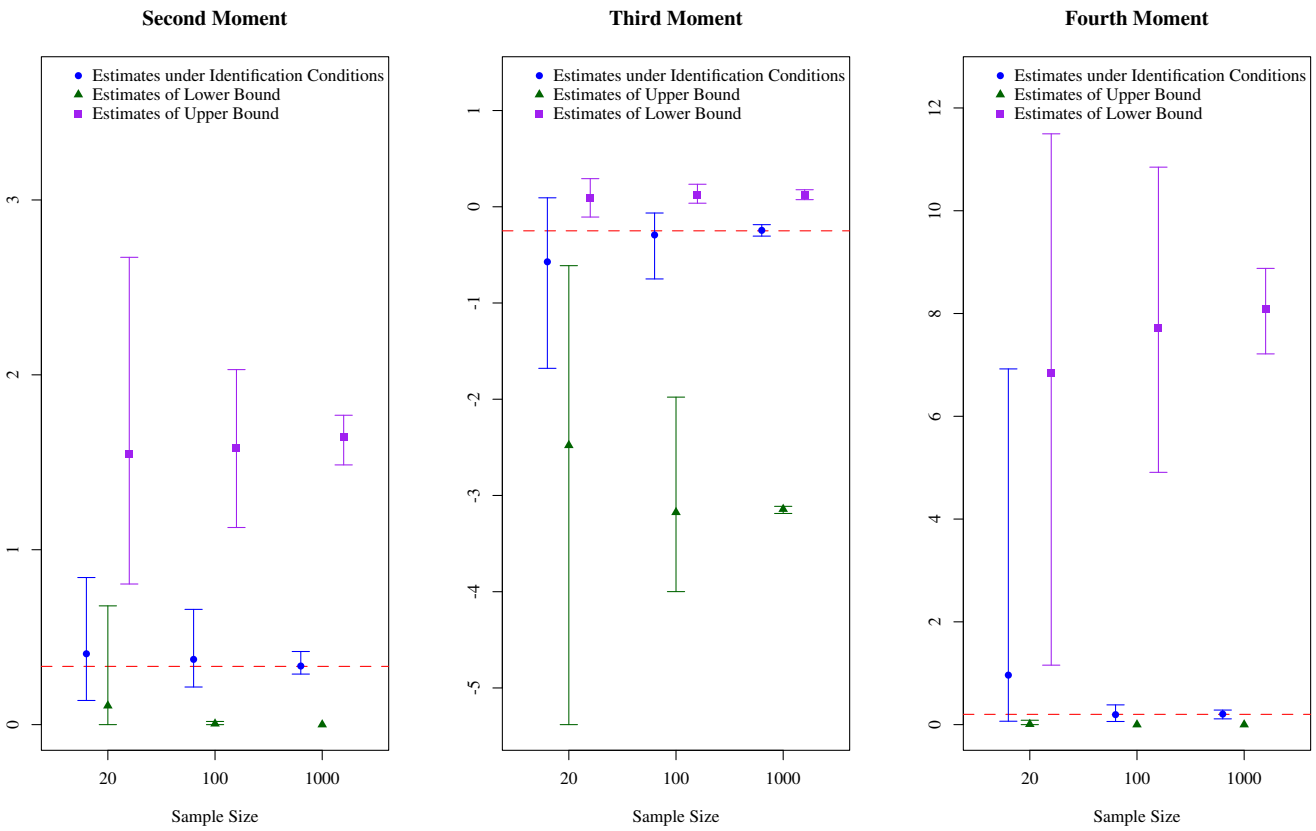

Figure 1: We present the estimates of the second, third, and fourth moments of causal effects along with their respective upper and lower bounds. Additionally, we report the means of each estimator accompanied by their 95% confidence intervals.

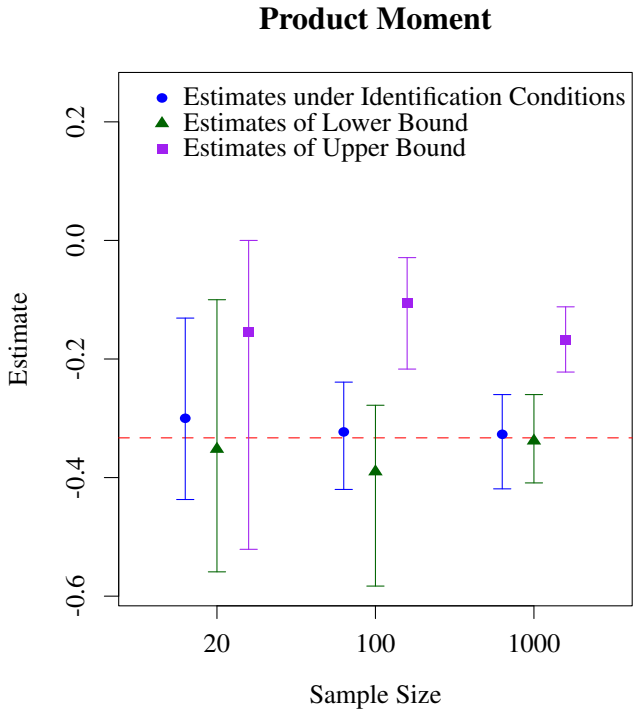

Figure 2: We present the estimate of the product moment of causal effects along with its respective upper and lower bound. Additionally, we report the means of each estimator accompanied by their 95% confidence intervals.

G ADDITIONAL INFORMATION ABOUT THE APPLICATION TO REAL-WORLD

In this section, we provide additional information about the application in the body of our paper.

Additional information about the moments of causal effects.

We provide the bounds of variance, standard deviation, skewness, and kurtosis of causal effect $Y_2 - Y_1$. When the denominator of the estimator is zero, we replace it with 0.01 to avoid numerical instability. The estimated bounds of the causal effect $Y_2 - Y_1$ are

Upper bound of variance: 29.851 (95%CI: $[12.173, 53.740]$),
Lower bound of variance: 0.000 (95%CI: $[0.000, 0.000]$),
Upper bound of standard deviation: 5.463 (95%CI: $[3.489, 7.33]$),
Lower bound of standard deviation: 0.000 (95%CI: $[0.000, 0.000]$),
Upper bound of skewness: 71296.725 (95%CI: $[0.000, 258849.124]$),
Lower bound of skewness: -13168.940 (95%CI: $[-64712.250, 0.000]$),
Upper bound of kurtosis: 6637894.937 (95%CI: $[0.000, 35261058.567]$),
Lower bound of kurtosis: 0.000 (95%CI: $[0.000, 0.000]$).

We also provide the variance, skewness, and kurtosis of causal effect $Y_4 - Y_2$. The results are:

Mean: 3.124 (95%CI: $[0.364, 5.820]$),
Variance: 3.104 (95%CI: $[0.297, 8.610]$),
Skewness: -3.373 (95%CI: $[-28.287, 8.821]$),
Kurtosis: 18.765 (95%CI: $[0.000, 203.983]$).

The estimated bounds of the causal effect $Y_4 - Y_2$ are

Upper bound of variance: 29.503 (95%CI: $[10.532, 61.318]$),
Lower bound of variance: 0.000 (95%CI: $[0.000, 0.000]$),
Upper bound of skewness: 34944.621 (95%CI: $[0.000, 129424.512]$),
Lower bound of skewness: -79272.510 (95%CI: $[-268151.400, 0.000]$),
Upper bound of kurtosis: 3261648.987 (95%CI: $[0.000, 22258543.128]$),
Lower bound of kurtosis: 0.000 (95%CI: $[0.000, 0.000]$).

We provide the variance, skewness, and kurtosis of causal effect $Y_4 - Y_1$. The results are:

Mean: 6.551 (95%CI: $[4.549, 9.034]$),
Variance: 2.078 (95%CI: $[0.000, 5.953]$),
Skewness: 5.797 (95%CI: $[-24.688, 65.521]$),
Kurtosis: 14.338 (95%CI: $[0.000, 242.5153]$).

The estimated bounds of the causal effect $Y_4 - Y_1$ are

Upper bound of variance: 25.160 (95%CI: $[9.034, 42.910]$),
Lower bound of variance: 0.000 (95%CI: $[0.000, 0.000]$),
Upper bound of skewness: 34135.710 (95%CI: $[0.000, 113246.400]$),
Lower bound of skewness: -19251.891 (95%CI: $[-97068.380, 0.000]$),
Upper bound of kurtosis: $2291969,972$ (95%CI: $[0.000, 17630519.123]$),
Lower bound of kurtosis: 0.000 (95%CI: $[0.000, 0.000]$).

Additional information about the product moments of causal effects.

We provide the bounds of covariance and correlation of causal effects $Y_2 - Y_1$ and $Y_4 - Y_2$. When the denominator of the estimator is zero, it is replaced with 0.01 to prevent numerical instability. The estimated bounds of the covariance of causal effect $Y_2 - Y_1$ and $Y_4 - Y_2$ are

$$\textbf{Upper bound of covariance}: 27.613 \text{ (95\%CI: } [9.486, 50.481]),$$
$$\textbf{Lower bound of covariance}: -29.172 \text{ (95\%CI: } [-52.263, -11.282]),$$

and the estimated bounds of the correlation of causal effect $Y_2 - Y_1$ and $Y_4 - Y_2$ are

$$\textbf{Upper bound of correlation}: 0.919 \text{ (95\%CI: } [0.000, 1.000]),$$
$$\textbf{Lower bound of correlation}: -0.976 \text{ (95\%CI: } [-1.000, -0.923]).$$