
Identification and Bounding of Central Moments of Causal Effects Using Marginal Moments Information

Naoya Hashimoto^{1,*}

Yuta Kawakami^{1,*}

Jin Tian¹

¹Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, UAE

Abstract

Evaluating the causal effect of a treatment on an outcome is a central objective in causal inference. While the average causal effect summarizes the mean impact of treatment, the central moments of the individual causal effect (ICE) characterize the shape of the ICE distribution, thereby revealing the extent and structure of treatment effect heterogeneity across individuals. This paper investigates the identification and bounding of the central moments of the ICE using only the marginal central moments of each potential outcome (PO). Compared with existing approaches that require knowledge of the full marginal distributions of the POs, marginal moment information is often substantially easier to obtain in empirical applications. Finally, we illustrate the practical relevance of our results through two empirical case studies.

1 INTRODUCTION

Evaluating the causal effect of a treatment (X) on an outcome (Y) from observed data is a central objective in causal inference. Researchers often focus on the average causal effect (ACE) $\mathbb{E}[Y_1 - Y_0]$, where Y_x denotes the potential outcome (PO) of an intervention $do(X = x)$. However, the ACE ignores heterogeneity in causal effects. To reveal the heterogeneity of causal effects across individuals, the distribution of individual causal effects (ICE) $Y_1 - Y_0$ has become an increasingly important focus in causal inference. Fan and Park [2010] provided sharp bounds for the distribution function of ICE, i.e., $\mathbb{P}(Y_1 - Y_0 < y)$. Post and Van Den Heuvel [2025] provided an identification result for the distribution function of ICE under the assumption, named "Independent Effect Deviation" (IED), that the ICE $Y_1 - Y_0$ is independent of the baseline Y_0 , i.e., $(Y_1 - Y_0) \perp\!\!\!\perp Y_0$. Re-

cently, Kawakami and Tian [2025] instead studied the central moments of ICE, $\bar{\mu}^{(m)} \triangleq \mathbb{E}[(Y_1 - Y_0 - \mathbb{E}[Y_1 - Y_0])^m]$, to characterize the shape of the ICE distribution. They provided identification and bounding results for $\bar{\mu}^{(m)}$. Summary statistics derived from the central moments—including variance, skewness, and kurtosis—are widely used in statistics. In particular, variance captures the extent of treatment effect heterogeneity; skewness indicates asymmetry in treatment response, helping identify whether a subset of individuals benefits (or is harmed) disproportionately; and kurtosis reflects tail behavior, which is important in applications where extreme responses (e.g., adverse effects) are of concern.

These previous studies [Post, 2023, Post and Van Den Heuvel, 2025, Kawakami and Tian, 2025] assume access to the full marginal distributions of each PO, namely $\mathbb{P}(Y_1 < y_1)$ and $\mathbb{P}(Y_0 < y_0)$. In many empirical settings, however, only summary statistics—such as mean, variance, skewness, or kurtosis—are reported, while the underlying distributions and individual-level data are unavailable. This may occur due to privacy or confidentiality restrictions, or in historically accumulated studies that report only aggregated statistics rather than the original datasets. In such settings, the marginal central moments of each PO, $\bar{\sigma}_1^{(k)} \triangleq \mathbb{E}[(Y_1 - \mathbb{E}[Y_1])^k]$ and $\bar{\sigma}_0^{(k)} \triangleq \mathbb{E}[(Y_0 - \mathbb{E}[Y_0])^k]$, are typically more accessible in practice, even though they contain less information than the full marginal distributions.

This paper provides methods for evaluating the central moments of ICE $\bar{\mu}^{(m)}$ from the marginal central moments $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$. $\bar{\mu}^{(m)}$ characterize the shape of the distribution of ICE and reveal the heterogeneity of causal effects across individuals. They can be used to compute key statistical measures of causal effects heterogeneity such as standard deviation (SD) $\sqrt{\bar{\mu}^{(2)}}$, skewness $\bar{\mu}^{(3)}/(\bar{\mu}^{(2)})^{3/2}$, and kurtosis $\bar{\mu}^{(4)}/(\bar{\mu}^{(2)})^2$. SD captures dispersion: a large SD indicates substantial heterogeneity, with ICEs potentially far from the ACE, whereas a small SD implies that ICEs cluster tightly around the ACE. Skewness measures distributional asymmetry; positive skewness corresponds to a longer right

*Equal contribution.

Table 1: Summary table of the theorems in this paper for bounding $\bar{\mu}^{(m)}$ using $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ for a single k . All bounds reported in this table are sharp. Nontrivial bounds for even m mean that the resulting bounds are tighter than $[0, \infty]$, whereas nontrivial bounds for odd m mean that the resulting bounds are tighter than $[-\infty, \infty]$. LB denotes lower bound, and UB denotes upper bound.

Thm.	Target m	Available k	Nontrivial bounds?
3	2	2	✓
5	2	3	×
6	2	4	✓ (only UB)
8	3	2	×
9	3	3	×
10	3	4	✓
12	4	2	✓ (only LB)
13	4	3	×
14	4	4	✓
16	Even m	m	✓
17	Even m	Even $k < m$	✓ (only LB)
17	Even m	Odd $k < m$	×
18	Odd m	$k \leq m$	×

tail, and negative skewness to a longer left tail. Kurtosis reflects tail thickness and concentration, with higher values signaling a greater likelihood of extreme causal effects.

We study identifying and bounding the central moments of ICE $\bar{\mu}^{(m)}$ using *only* the marginal central moments $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$. In contrast to prior approaches, our framework does not require access to the full marginal distributions of Y_1 and Y_0 ; it relies solely on their marginal central moments. Our theoretical contributions are as follows:

- We establish identification results for the central moments of ICE using only marginal central moment information from Y_1 and Y_0 under the IED condition.
- We derive bounds for the central moments of ICE using only marginal central moment information from Y_1 and Y_0 . Table 1 summarizes our results for bounding $\bar{\mu}^{(m)}$ using $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ for a single k . All bounds reported in the table are sharp. In addition, Theorems 7, 11, and 15, together with Corollary 3, provide bounds for $\bar{\mu}^{(m)}$ using multiple $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ by intersecting the single- k bounds in Table 1; these bounds are generally not sharp.

To our knowledge, these identification and bounding results are novel. The problem of evaluating central moments of the ICE using only marginal central moments of Y_1 and Y_0 has not previously been studied in a systematic manner.

Finally, we show the practical relevance of our theoretical results through two empirical case studies.

2 MOTIVATING EXAMPLES

We begin by motivating the problem of bounding the central moments of the ICE using the marginal central moments of each PO. In many experimental studies, datasets containing individual-level information are not publicly released due to privacy or confidentiality concerns. By contrast, moment information—such as the variance, skewness, or kurtosis—of treatment and control groups in randomized controlled trials (RCTs) is frequently reported as basic summary statistics.

For example, Moseley et al. [2002] conducted a randomized, placebo-controlled trial of arthroscopic surgery for osteoarthritis of the knee. A total of 180 patients with osteoarthritis of the knee were randomly assigned to arthroscopic débridement ($X = 1$), arthroscopic lavage, or placebo surgery ($X = 0$) (we chose two of the three groups for our analysis). Moseley et al. [2002] reported the Physical Functioning Scale (PFS) score at 24 months (outcome Y), which records the time in seconds that a patient requires to walk 30 m and to climb up and down a flight of stairs as quickly as possible, where longer times indicate poorer functioning. The article reports the mean and standard deviation (SD) of PFS for each group, given in Table 2.

Špirtović et al. [2023] reported an experimental study on the influence of a 12-week preventive program of mixed aerobics on body composition in healthy adult women. A total of 64 women were randomly divided into two groups and completed both the initial and final measurements. The experimental group ($X = 1$) took part in the exercise program of mixed aerobics, whereas the control group ($X = 0$) did not participate in any form of organized physical activity. Muscle mass percentage (MM%) in the final state (outcome Y) is calculated and expressed as a percentage of total body mass. The study reports the mean, SD, skewness, and kurtosis for MM% in each group, given in Table 3.

Because the authors of the two studies did not make the underlying dataset publicly available or provide the full distribution functions for each group, it is impossible to evaluate the central moments of the ICE using the methods based on marginal distributions of each group in [Post, 2023, Post and Van Den Heuvel, 2025, Kawakami and Tian, 2025].

Additionally, several experimental studies (e.g., Hébert et al. [1999], Van den Berghe et al. [2001], Rivers et al. [2001], Investigators [2009]) reported the means and standard deviations for each group. Other studies (e.g., Hinojo Lucena et al. [2020], Goble et al. [2021], Ulaş and Seğer [2022], Özdemir Cihan and Dilekmen [2024], Stefanescu et al. [2024]) provided higher-order moment information, such as skewness and kurtosis, even though the underlying individual-level datasets are not publicly available. These studies compared the differences in marginal summary statistics between the two groups (control and treatment). Our results in this paper enable re-analysis of these

Table 2: Summary statistics reported by [Moseley et al., 2002]. SD: standard deviation.

Group	Sample Size	Mean	SD
Placebo surgery ($X = 0$)	44	47.7	12.0
Arthroscopic débridement ($X = 1$)	44	52.6	16.4

studies to infer properties of the *distribution* of causal effects using only the reported marginal moment information.

3 NOTATION AND BACKGROUND

We represent each variable with a capital letter (X) and its realized value with a lowercase letter (x). We denote $\mathbb{E}[Y]$ as the expectation of Y . Let $\overline{\mathbb{R}} = [-\infty, \infty] = \mathbb{R} \cup \{\pm\infty\}$ be the extended real line.

We use the language of structural causal models (SCM) as our basic semantic and inferential framework [Pearl, 2009]. We denote the potential outcome (PO) Y under intervention $do(x)$ by Y_x . The individual causal effect (ICE) is defined as $Y_1 - Y_0$ [Holland, 1986], and the central moments of ICE are defined as follows [Kawakami and Tian, 2025]:

Definition 1 (The central moments of ICE). *The m -th central moment of ICE is defined as*

$$\bar{\mu}^{(m)} \triangleq \mathbb{E}[(Y_1 - Y_0 - \mathbb{E}[Y_1 - Y_0])^m]. \quad (1)$$

Post and Van Den Heuvel [2025] showed that, under the "Independent Effect Deviation" (IED) assumption¹, $(Y_1 - Y_0) \perp\!\!\!\perp Y_0$, the distribution of ICE is identified from the marginal distributions of Y_1 and Y_0 . Kawakami and Tian [2025] derived identification results for the central moments $\bar{\mu}^{(m)}$ of ICE in terms of the marginals of Y_1 and Y_0 under a monotonicity assumption, and in the absence of monotonicity, established bounds for $\bar{\mu}^{(m)}$ in terms of the marginals of Y_1 and Y_0 . In this paper, we study the identification and bounding of $\bar{\mu}^{(m)}$ using only the central moments of the marginals of Y_1 and Y_0 , without assuming full knowledge of their marginal distributions.

¹IED assumes that the individual causal effect $Y_1 - Y_0$ is independent of the baseline outcome Y_0 . Intuitively, this means that how much a unit benefits from treatment is unrelated to its baseline level. This assumption may be plausible in settings where the mechanisms generating the treatment effect are independent of those determining the baseline outcome. Conversely, the assumption is less plausible in scenarios where treatment effects depend on baseline levels—for example, when individuals with higher (or lower) baseline outcomes tend to benefit more (or less) from the treatment.

4 IDENTIFICATION OF CENTRAL MOMENTS OF ICE UNDER IED CONDITION

We study the evaluation of the central moments of ICE, $\bar{\mu}^{(m)}$, through the marginal central moments of each PO,

$$\begin{aligned} \bar{\sigma}_1^{(k)} &\triangleq \mathbb{E}[(Y_1 - \mathbb{E}[Y_1])^k], \\ \bar{\sigma}_0^{(k)} &\triangleq \mathbb{E}[(Y_0 - \mathbb{E}[Y_0])^k]. \end{aligned} \quad (2)$$

In RCTs, $\bar{\sigma}_1^{(k)}$ and $\bar{\sigma}_0^{(k)}$ correspond to the k -th central moment of the outcome for the treatment and control groups, respectively. In observational studies with no confounders, i.e., $Y_x \perp\!\!\!\perp X$, $\bar{\sigma}_1^{(k)}$ and $\bar{\sigma}_0^{(k)}$ correspond to the k -th central moment of the outcome for subgroups conditioning on $X = 1$ and $X = 0$, respectively. In observational studies with observed covariates C and no unmeasured confounding, i.e., $Y_x \perp\!\!\!\perp X|C$, the quantities $\bar{\sigma}_1^{(k)}$ and $\bar{\sigma}_0^{(k)}$ can be identified via covariate adjustment. For discrete C ,

$$\begin{aligned} \bar{\sigma}_1^{(k)} &= \sum_{c \in \Omega_C} \mathbb{E}[(Y - \mathbb{E}[Y_1])^k | X = 1, C = c] \mathbb{P}(C = c), \\ \bar{\sigma}_0^{(k)} &= \sum_{c \in \Omega_C} \mathbb{E}[(Y - \mathbb{E}[Y_0])^k | X = 0, C = c] \mathbb{P}(C = c), \end{aligned} \quad (3)$$

where

$$\begin{aligned} \mathbb{E}[Y_1] &= \sum_{c \in \Omega_C} \mathbb{E}[Y | X = 1, C = c] \mathbb{P}(C = c), \\ \mathbb{E}[Y_0] &= \sum_{c \in \Omega_C} \mathbb{E}[Y | X = 0, C = c] \mathbb{P}(C = c). \end{aligned} \quad (4)$$

For continuous C , the sums are replaced by integrals.

The same arguments can be applied within strata of observed covariates. If the marginal central moments of $Y_x | C = c$ are available, the identification and bounding results below apply after replacing Y_x by $Y_x | C = c$. The resulting stratum-specific moments describe the distribution of the ICE within the subpopulation $C = c$, while the conditional average causal effect $\mathbb{E}[Y_1 - Y_0 | C = c]$ describes only the corresponding mean.

In this section, we discuss the identification of the central moments of the ICE. Post and Van Den Heuvel [2025] provided identification of the distribution of the ICE using the marginal distributions of each PO. In contrast, we provide identification of the central moments of the ICE using only the central moments of each PO.

Identification of the variance $\bar{\mu}^{(2)}$ of ICE. We first discuss the identification of the variance of ICE using the variance of each PO. The variance of ICE quantifies the degree of dispersion in ICE and serves as a key measure of causal effect heterogeneity [Kawakami and Tian, 2025]. A large variance indicates substantial heterogeneity in causal effects, whereas a small variance indicates relative homogeneity.

Table 3: Summary statistics reported by [Špirtović et al., 2023]. SD: standard deviation, Skew: skewness, Kurt: kurtosis, 2nd M: second central moments, 3rd M: third central moments, 4th M: fourth central moments. (We calculated 2nd M, 3rd M, and 4th M from their SD, Skew, and Kurt.)

Group	Sample Size	Mean	SD	Skew	Kurt	2nd M	3rd M	4th M
Control ($X = 0$)	30	45.469	4.573	0.648	3.652	20.912	61.970	1597.113
Experimental ($X = 1$)	34	44.125	5.262	0.915	3.273	27.689	133.313	2509.282

Post and Van Den Heuvel [2025] showed that the variance of ICE is identifiable under the IED condition. We restate their results formally in the following proposition.

Proposition 1. Assume $\bar{\sigma}_1^{(2)}$ and $\bar{\sigma}_0^{(2)}$ exist and are available. If the IED condition $(Y_1 - Y_0) \perp\!\!\!\perp Y_0$ holds, then the variance of ICE is identified as

$$\bar{\mu}^{(2)} = \bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)}. \quad (5)$$

The variance of ICE is given by the difference between the variances of the two POs.

Remark. Under the IED, $\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)} \geq 0$ holds by Proposition 1, which provides a testable condition to falsify the IED assumption.

Kawakami and Tian [2025] identified the central moments of ICE from the marginal distributions of Y_1 and Y_0 under a monotonicity assumption (their Assumption 2). However, $\bar{\mu}^{(2)}$ is not identifiable from $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ alone, even under this monotonicity assumption. A counterexample is provided in Appendix A.1.

Identification of the higher-order central moments of the ICE. Next, we provide an identification theorem for the higher-order central moments of the ICE through the central moments of each PO $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$, which is a generalization of Proposition 1.

Theorem 1. Assume $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$, $k = 1, 2, \dots, m-3, m-2, m$, exist and are available. If the IED condition $(Y_1 - Y_0) \perp\!\!\!\perp Y_0$ holds, then the m -th central moment of ICE is recursively identified as

$$\bar{\mu}^{(m)} = \bar{\sigma}_1^{(m)} - \bar{\sigma}_0^{(m)} - \sum_{\ell=2}^{m-2} \binom{m}{\ell} \bar{\mu}^{(m-\ell)} \bar{\sigma}_0^{(\ell)}. \quad (6)$$

Theorem 1 shows that the central moments of ICE are identified from marginal central moments of each PO under IED, without requiring the complete knowledge of the marginal distributions used in [Post and Van Den Heuvel, 2025].

We have the following important corollaries, as $\bar{\mu}^{(3)}$ and $\bar{\mu}^{(4)}$ are used to compute skewness and kurtosis:

Corollary 1. If $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available, and the IED condition $(Y_1 - Y_0) \perp\!\!\!\perp Y_0$ holds, then $\bar{\mu}^{(3)}$ is identified as follows:

$$\bar{\mu}^{(3)} = \bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)}. \quad (7)$$

Then the skewness of the ICE, i.e., $\bar{\mu}^{(3)}/(\bar{\mu}^{(2)})^{3/2}$, is identified from $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$, $k = 2, 3$, under IED.

Corollary 2. If $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available, and the IED condition $(Y_1 - Y_0) \perp\!\!\!\perp Y_0$ holds, then $\bar{\mu}^{(4)}$ is identified as follows:

$$\bar{\mu}^{(4)} = \bar{\sigma}_1^{(4)} - \bar{\sigma}_0^{(4)} - 6(\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)})\bar{\sigma}_0^{(2)}. \quad (8)$$

Then the kurtosis of the ICE, i.e., $\bar{\mu}^{(4)}/(\bar{\mu}^{(2)})^2$, is identified from $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$, $k = 2, 4$, under IED.

5 BOUNDING 2ND, 3RD, AND 4TH CENTRAL MOMENTS OF ICE

We next study the bounding of the central moments $\bar{\mu}^{(m)}$ of ICE from the marginal central moments $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ without imposing the IED condition. The marginal means $\mathbb{E}[Y_1]$ and $\mathbb{E}[Y_0]$ are also assumed to be available; however, they do not provide additional power for bounding $\bar{\mu}^{(m)}$. In practice, interest in the moments of causal effects primarily concerns the variance, skewness, and kurtosis. These can be calculated through the second, third, and fourth central moments of the ICE, i.e., $\bar{\mu}^{(2)}$, $\bar{\mu}^{(3)}$, and $\bar{\mu}^{(4)}$, which are the focus of this section.

Sharpness. The notion of sharpness varies across contexts [Zhang and Richardson, 2025, Jonzon et al., 2025]. We derive bounds for $\bar{\mu}^{(m)}$ in the form of a closed interval $[l, u]$, where $l, u \in \mathbb{R}$, given available marginal central moments $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$. We say that $[l, u]$ is a *sharp* bound for $\bar{\mu}^{(m)}$ if there exist no strictly tighter closed bounds for $\bar{\mu}^{(m)}$ under the given moment information. Our sharpness means that the closed interval cannot be further tightened without additional information (e.g., the marginal distributions of Y_1 and Y_0) or assumptions.

Note that we restrict the bounds to a closed interval. Some researchers instead define the bounds using an open interval. However, the interpretation is essentially unchanged whether the interval is written as $[l, u]$ or (l, u) , since the distinction concerns only the inclusion or exclusion of the boundary points and does not affect the substantive properties under consideration.

5.1 BOUNDS OF $\bar{\mu}^{(2)}$

We first study bounds for the variance of ICE by analyzing the consequences of violating the IED condition. We denote the correlation of $Y_1 - Y_0$ and Y_0 as λ :

$$\lambda \triangleq \frac{\mathbb{E}[(Y_1 - Y_0 - \mathbb{E}[Y_1 - Y_0])(Y_0 - \mathbb{E}[Y_0])]}{\sqrt{\text{Var}(Y_1 - Y_0)}\sqrt{\text{Var}(Y_0)}}. \quad (9)$$

We define $\lambda = 0$ if $\text{Var}(Y_1 - Y_0)\text{Var}(Y_0) = 0$. When the IED condition holds, we have $\lambda = 0$. A positive correlation implies that units with a higher baseline outcome Y_0 exhibit a larger ICE, whereas units with a lower baseline outcome exhibit a smaller ICE; conversely, a negative correlation implies that higher baseline outcomes are associated with smaller ICEs and lower baseline outcomes with larger ICEs.

We examine how the variance of the ICE is influenced by the parameter λ . If the correlation of $Y_1 - Y_0$ and Y_0 is fixed, $\bar{\mu}^{(2)}$ is determined as follows:

Theorem 2. (1). Assume $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$. If $\bar{\sigma}_0^{(2)} = 0$, then $\lambda = 0$ and $\bar{\mu}^{(2)} = \bar{\sigma}_1^{(2)}$. If $\bar{\sigma}_0^{(2)} > 0$, then for any $-1 \leq \lambda \leq 1$, $\bar{\mu}^{(2)}$ is uniquely determined as

$$\bar{\mu}^{(2)} = \left(-\sqrt{\bar{\sigma}_0^{(2)}}\lambda + \sqrt{\bar{\sigma}_1^{(2)} - (1 - \lambda^2)\bar{\sigma}_0^{(2)}} \right)^2. \quad (10)$$

(2). Assume $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)}$. Then $\lambda \leq 0$, and $\bar{\mu}^{(2)}$ is uniquely determined as

$$\bar{\mu}^{(2)} = 4\lambda^2 \bar{\sigma}_0^{(2)}. \quad (11)$$

(3). Assume $\bar{\sigma}_1^{(2)} < \bar{\sigma}_0^{(2)}$. We have $-1 \leq \lambda \leq -\sqrt{1 - \bar{\sigma}_1^{(2)}/\bar{\sigma}_0^{(2)}}$, and

$$\bar{\mu}^{(2)} \in \left\{ \left(-\sqrt{\bar{\sigma}_0^{(2)}}\lambda - \sqrt{\bar{\sigma}_1^{(2)} - (1 - \lambda^2)\bar{\sigma}_0^{(2)}} \right)^2, \left(-\sqrt{\bar{\sigma}_0^{(2)}}\lambda + \sqrt{\bar{\sigma}_1^{(2)} - (1 - \lambda^2)\bar{\sigma}_0^{(2)}} \right)^2 \right\}. \quad (12)$$

The number of possible values that $\bar{\mu}^{(2)}$ can take given λ depends on the relative magnitudes of $\bar{\sigma}_0^{(2)}$ and $\bar{\sigma}_1^{(2)}$. In addition, when $\bar{\sigma}_1^{(2)} \leq \bar{\sigma}_0^{(2)}$, the feasible range of λ is restricted.

By Theorem 2, we obtain the following bound on $\bar{\mu}^{(2)}$:

Theorem 3. Assume that $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ exist and are available. $\bar{\mu}^{(2)}$ is bounded as follows:

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^2 \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}} \right)^2. \quad (13)$$

If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ are available, then this bound is sharp.

If we know whether $\lambda \geq 0$ or $\lambda \leq 0$ a priori, we can obtain tighter bounds:

Theorem 4. Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ exist and are available.

(1). If $\lambda \geq 0$, then we have $\bar{\sigma}_1^{(2)} \geq \bar{\sigma}_0^{(2)}$. If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)}$, then $\bar{\mu}^{(2)} = 0$; if $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$, we have

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^2 \leq \bar{\mu}^{(2)} \leq \bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)}. \quad (14)$$

(2). Assume $\lambda \leq 0$. If $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$, we have

$$\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)} \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}} \right)^2. \quad (15)$$

If $\bar{\sigma}_1^{(2)} \leq \bar{\sigma}_0^{(2)}$, Eq. (13) holds.

These bounds are sharp.

Can $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ or $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$, which may be available in practice, provide information about $\bar{\mu}^{(2)}$?

Theorem 5. Assume that $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available. Then we have $0 \leq \bar{\mu}^{(2)} \leq \infty$. This bound is sharp if only $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available.

$(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ alone does not provide a nontrivial bound for $\bar{\mu}^{(2)}$. On the other hand, $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ provides a nontrivial upper bound as follows:

Theorem 6. Assume that $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$0 \leq \bar{\mu}^{(2)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^2. \quad (16)$$

This bound is sharp if only $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available.

Furthermore, by taking the intersection of the bounds in Theorems 3, 5, and 6, we have the following results:

Theorem 7. (1). Assume that $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available. Then we have

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^2 \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}} \right)^2. \quad (17)$$

If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available, then this bound is sharp.

(2). Assume that $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available, we have

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^2 \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}} \right)^2. \quad (18)$$

If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available, then this bound is sharp.

(3). Assume that $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$0 \leq \bar{\mu}^{(2)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^2. \quad (19)$$

(4). Assume that $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$, and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}\right)^2 \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}}\right)^2. \quad (20)$$

In general, taking the intersection of sharp bounds does not preserve the sharpness. Indeed, the bounds in (3) and (4) of Theorem 7 are not sharp: we provide their counterexamples in Appendix A.2. As an exception, the bounds in (1) and (2) are sharp, which means that, given $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, additionally knowing either $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ alone or $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ alone does not further tighten the bound for $\bar{\mu}^{(2)}$. On the other hand, jointly knowing both $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$, in addition to $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, provides additional identifying information about $\bar{\mu}^{(2)}$.

The upper bound on $\bar{\mu}^{(2)}$ can be translated into a bound on the probability that the ICE has the opposite sign from the ACE. Let $\tau := \mathbb{E}[Y_1 - Y_0]$ and suppose that $\bar{\mu}^{(2)} \leq U_2$. If $\tau > 0$, then by Cantelli's inequality,

$$\mathbb{P}(Y_1 - Y_0 \leq 0) \leq \frac{U_2}{U_2 + \tau^2}. \quad (21)$$

If $\tau < 0$, then

$$\mathbb{P}(Y_1 - Y_0 \geq 0) \leq \frac{U_2}{U_2 + \tau^2}. \quad (22)$$

5.2 BOUNDS OF $\bar{\mu}^{(3)}$

We next study the bounding of $\bar{\mu}^{(3)}$ using $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$, or $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$. First we consider the cases where only one of these is available.

Theorem 8. Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ exist and are available.

(1). If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$, we have $\bar{\mu}^{(3)} = 0$.

(2). If $\bar{\sigma}_1^{(2)} > 0$ or $\bar{\sigma}_0^{(2)} > 0$, we have $-\infty \leq \bar{\mu}^{(3)} \leq \infty$. If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ are available, then this bound is sharp.

Theorem 9. Assume that $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available. Then we have $-\infty \leq \bar{\mu}^{(3)} \leq \infty$. If only $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available, then this bound is sharp.

By Theorems 8 and 9, $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ or $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ alone does not help to obtain a nontrivial bound for $\bar{\mu}^{(3)}$. The condition $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$ corresponds to an extreme case where both Y_1 and Y_0 are constant with probability one.

On the other hand, $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ is helpful in bounding $\bar{\mu}^{(3)}$:

Theorem 10. Assume that $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\begin{aligned} -\frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^3 &\leq \bar{\mu}^{(3)} \\ &\leq \frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^3. \end{aligned} \quad (23)$$

If only $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available, then this bound is sharp.

Moreover, if $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ ($k = 2, 3, 4$) are jointly available, by taking the intersection of the bounds in Theorems 8-10, we have the following results:

Theorem 11. (1). Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available.

(i). If $\bar{\sigma}_0^{(2)} > 0$ and $\bar{\sigma}_1^{(2)} > 0$, we have $-\infty \leq \bar{\mu}^{(3)} \leq \infty$. If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available, then this bound is sharp.

(ii). If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$, we have $\bar{\mu}^{(3)} = 0$. Otherwise, if $\bar{\sigma}_1^{(2)} = 0$, we have $\bar{\mu}^{(3)} = -\bar{\sigma}_0^{(3)}$; if $\bar{\sigma}_0^{(2)} = 0$, we have $\bar{\mu}^{(3)} = \bar{\sigma}_1^{(3)}$.

(2) Assume that $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available, and that at least one of $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available. Then we have

$$\begin{aligned} -\frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^3 &\leq \bar{\mu}^{(3)} \\ &\leq \frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^3. \end{aligned} \quad (24)$$

The bound in (2) of Theorem 11 is not sharp: we provide a counterexample in Appendix A.2. On the other hand, the bound in (1) of Theorem 11 is sharp, which means that jointly knowing $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ does not provide a nontrivial bound for $\bar{\mu}^{(3)}$ (except in extreme cases). However, once $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ is given, both $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ provide additional identifying information about $\bar{\mu}^{(3)}$.

Bounding of skewness. To obtain bounds on the skewness of ICE, i.e., $\bar{\mu}^{(3)}/(\bar{\mu}^{(2)})^{3/2}$, we first compute bounds on $\bar{\mu}^{(k)}$ ($k = 2, 3$). Then, we compute the upper bound for the skewness as the ratio of the upper bound of $\bar{\mu}^{(3)}$ to the lower bound of $\bar{\mu}^{(2)}$ raised to the power 3/2, and the lower bound is the negative of this value. Note that the obtained bounds may not be sharp.

5.3 BOUNDS OF $\bar{\mu}^{(4)}$

We consider the bounding of $\bar{\mu}^{(4)}$ using $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$, or $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$.

$(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ are helpful in getting a lower bound for $\bar{\mu}^{(4)}$:

Theorem 12. Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ exist and are available.

(1). If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$, we have $\bar{\mu}^{(4)} = 0$.

(2). If $\bar{\sigma}_1^{(2)} > 0$ or $\bar{\sigma}_0^{(2)} > 0$, we have

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^4 \leq \bar{\mu}^{(4)} \leq \infty. \quad (25)$$

If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ are available, then this bound is sharp.

On the other hand, $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are not helpful in obtaining a nontrivial bound for $\bar{\mu}^{(4)}$:

Theorem 13. Assume that $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available. Then we have $0 \leq \bar{\mu}^{(4)} \leq \infty$. If only $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available, then this bound is sharp.

When $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available, we can bound $\bar{\mu}^{(4)}$ as:

Theorem 14. Assume that $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\left(\sqrt[4]{\bar{\sigma}_1^{(4)}} - \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^4 \leq \bar{\mu}^{(4)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^4. \quad (26)$$

If only $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available, then this bound is sharp.

If $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ ($k = 2, 3, 4$) are jointly available, by taking the intersection of the bounds in Theorems 12-14, we have:

Theorem 15. (1). Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available.

(i). If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$, then $\bar{\mu}^{(4)} = 0$.

(ii). If $\bar{\sigma}_1^{(2)} > 0$ and $\bar{\sigma}_0^{(2)} > 0$, then

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^4 \leq \bar{\mu}^{(4)} \leq \infty. \quad (27)$$

(iii). If $\bar{\sigma}_1^{(2)} > 0$ and $\bar{\sigma}_0^{(2)} = 0$, then $\bar{\sigma}_0^{(3)} = 0$, and

$$(\bar{\sigma}_1^{(2)})^2 + \frac{(\bar{\sigma}_1^{(3)})^2}{\bar{\sigma}_1^{(2)}} \leq \bar{\mu}^{(4)} \leq \infty. \quad (28)$$

(iv). If $\bar{\sigma}_1^{(2)} = 0$ and $\bar{\sigma}_0^{(2)} > 0$, then $\bar{\sigma}_1^{(3)} = 0$, and

$$(\bar{\sigma}_0^{(2)})^2 + \frac{(\bar{\sigma}_0^{(3)})^2}{\bar{\sigma}_0^{(2)}} \leq \bar{\mu}^{(4)} \leq \infty. \quad (29)$$

If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available, then these bounds are sharp.

(2). Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\max \left\{ \left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^4, \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} - \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^4 \right\} \leq \bar{\mu}^{(4)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^4. \quad (30)$$

(3). Assume $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\left(\sqrt[4]{\bar{\sigma}_1^{(4)}} - \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^4 \leq \bar{\mu}^{(4)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^4. \quad (31)$$

(4). Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$, and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\max \left\{ \left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^4, \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} - \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^4 \right\} \leq \bar{\mu}^{(4)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^4. \quad (32)$$

The bounds in (2)-(4) of Theorem 15 are not sharp: we provide counterexamples in Appendix A.2. On the other hand, the bounds in (1) of Theorem 15 are sharp. Therefore, when $\bar{\sigma}_1^{(2)}\bar{\sigma}_0^{(2)} > 0$, additionally knowing $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ does not further tighten the bound for $\bar{\mu}^{(4)}$ obtained from $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ alone. However, given $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$, both $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ provide additional identifying information about $\bar{\mu}^{(4)}$.

Bounding of kurtosis. To obtain bounds on the kurtosis of ICE, $\bar{\mu}^{(4)}/(\bar{\mu}^{(2)})^2$, we first compute bounds on $\bar{\mu}^{(k)}$ ($k = 2, 4$). Then, we compute the upper (resp. lower) bound for the kurtosis as given by the ratio of the upper (resp. lower) bound of $\bar{\mu}^{(4)}$ to the square of the lower (resp. upper) bound of $\bar{\mu}^{(2)}$. The bounds may not be sharp.

Remark. If we have access to variance, skewness and kurtosis, then we can compute 3rd and 4th central moments. In the settings where we have only marginal skewness and kurtosis of POs but not marginal variance, we are not able to derive nontrivial bounds for $\bar{\mu}^{(2)}$, $\bar{\mu}^{(3)}$, or $\bar{\mu}^{(4)}$, unless we have range information about $\bar{\sigma}_1^{(2)}$ and $\bar{\sigma}_0^{(2)}$. A discussion is given in Appendix B.

6 BOUNDING HIGHER CENTRAL MOMENTS OF ICE

In this section, we study the bounding of higher orders of the central moments $\bar{\mu}^{(m)}$ of the ICE for arbitrary m .

6.1 BOUNDS OF EVEN-ORDER $\bar{\mu}^{(m)}$

As a generalization of Theorems 3 and 14, we have bounds of $\bar{\mu}^{(m)}$ for arbitrary even m if $(\bar{\sigma}_1^{(m)}, \bar{\sigma}_0^{(m)})$ are available.

Theorem 16. Let $m \geq 2$ be an even integer. Assume $(\bar{\sigma}_1^{(m)}, \bar{\sigma}_0^{(m)})$ exist and are available. Then we have

$$\left(\sqrt[m]{\bar{\sigma}_1^{(m)}} - \sqrt[m]{\bar{\sigma}_0^{(m)}} \right)^m \leq \bar{\mu}^{(m)} \leq \left(\sqrt[m]{\bar{\sigma}_1^{(m)}} + \sqrt[m]{\bar{\sigma}_0^{(m)}} \right)^m. \quad (33)$$

If only $(\bar{\sigma}_1^{(m)}, \bar{\sigma}_0^{(m)})$ are available, then this bound is sharp.

As a generalization of Theorems 12 and 13, when $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ for some $k < m$ are available, we have:

Theorem 17. Let $m \geq 2$ be an even integer. Assume that $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ exist and are available for a single $k < m$.

(1). If k is odd, we have $0 \leq \bar{\mu}^{(m)} \leq \infty$.

(2). If k is even, we have the following:

(i). If $\bar{\sigma}_1^{(k)} = \bar{\sigma}_0^{(k)} = 0$, we have $\bar{\mu}^{(m)} = 0$.

(ii). If $\bar{\sigma}_1^{(k)} > 0$ or $\bar{\sigma}_0^{(k)} > 0$, we have

$$\left(\sqrt[k]{\bar{\sigma}_1^{(k)}} - \sqrt[k]{\bar{\sigma}_0^{(k)}} \right)^m \leq \bar{\mu}^{(m)} \leq \infty. \quad (34)$$

If only $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ are available, then these bounds are sharp.

6.2 BOUNDS OF ODD-ORDER $\bar{\mu}^{(m)}$

As a generalization of Theorems 8 and 9, we have:

Theorem 18. Let $m \geq 3$ be an odd number. Assume that $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ exist and are available for a single $k \leq m$.

(1). If k is even and $\bar{\sigma}_1^{(k)} = \bar{\sigma}_0^{(k)} = 0$, we have $\bar{\mu}^{(m)} = 0$.

(2). Otherwise, we have $-\infty \leq \bar{\mu}^{(m)} \leq \infty$. If only $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ are available, then this bound is sharp.

Theorem 18 means that, when m is odd, knowing $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ for a single $k \leq m$ does not provide a nontrivial bound for $\bar{\mu}^{(m)}$ (except in extreme cases). In particular, unlike the cases where m is even, we cannot obtain a nontrivial bound for $\bar{\mu}^{(m)}$ even if $(\bar{\sigma}_1^{(m)}, \bar{\sigma}_0^{(m)})$ are available.

6.3 BOUNDS OF $\bar{\mu}^{(m)}$ GIVEN MULTIPLE $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$

We obtain bounds for $\bar{\mu}^{(m)}$ when multiple $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ are jointly available by intersecting bounds in Theorems 16–18.

Corollary 3. Assume $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ exist and are available for all $k \leq m$, and $\bar{\sigma}_1^{(k)} + \bar{\sigma}_0^{(k)} > 0$ for all even k . We have:

(1). If m is even, we have

$$\begin{aligned} \max_{k \leq m, k \text{ is even}} \left(\sqrt[k]{\bar{\sigma}_1^{(k)}} - \sqrt[k]{\bar{\sigma}_0^{(k)}} \right)^m \\ \leq \bar{\mu}^{(m)} \leq \left(\sqrt[m]{\bar{\sigma}_1^{(m)}} + \sqrt[m]{\bar{\sigma}_0^{(m)}} \right)^m. \end{aligned} \quad (35)$$

(2). If m is odd, we have $-\infty \leq \bar{\mu}^{(m)} \leq \infty$.

These bounds are obtained by simply intersecting the individual bounds and are not sharp.

Remark. We do not investigate the general setting in which $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ are available for $k > m$, of which Theorem 10 is a special instance. The setting is very challenging.

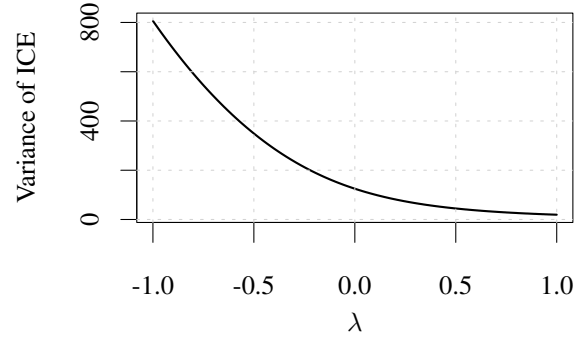


Figure 1: The variance of the ICE as a function of λ for the case study in [Moseley et al., 2002].

7 CASE STUDIES

In this section, we analyze the two empirical case studies described in Section 2. To support our findings in finite-sample settings, we also conducted simulated numerical experiments, the results of which are reported in Appendix D.

7.1 CASE STUDY IN [MOSELEY ET AL. 2002]

Summary of the study. We analyze the study by Moseley et al. [2002] introduced in Section 2. The mean and standard deviation (SD) of the PFS score (Y) for the placebo surgery group ($X = 0$) and the arthroscopic débridement group ($X = 1$) are shown in Table 2. ACE is reported as 4.9, which means arthroscopic débridement increases PFS.

Results. Figure 1 presents the variance of the ICE $\bar{\mu}^{(2)}$ as a function of λ , as given in Theorem 2 case (1). $\bar{\mu}^{(2)}$ is monotonically decreasing in λ . This implies that $\bar{\mu}^{(2)}$ decreases as the correlation between the PO under placebo surgery and the ICE increases. The lower bound of $\bar{\mu}^{(2)}$ is 19.360, and the upper bound is 806.560. If researchers are willing to assume that $\lambda \geq 0$ (i.e., a positive correlation between Y_0 and $Y_1 - Y_0$), we obtain a tighter bound from 19.360 to 124.960. The estimated variance of the ICE under the IED condition ($\lambda = 0$) is 124.960. We also obtain a lower bound of 5.762×10^{-4} for the kurtosis of the ICE. Note that we cannot compute confidence intervals of the estimates from the reported point estimates in [Moseley et al., 2002].

Our results indicate relatively large heterogeneity in the ICE (the variance is substantial relative to the ACE). These findings further highlight the risk of arthroscopic débridement, as there are a substantial number of patients whose ICE differs markedly from the ACE, including potentially adverse effects.

7.2 CASE STUDY IN [ŠPIRTOVIC ET AL. 2023]

Summary of the study. We analyze the study by [Špir-

Table 4: Estimated bounds and IED-based estimates of the variance, skewness, and kurtosis of the ICE for the case study in [Špirtović et al., 2023].

Target	Estimates	
Variance	Under IED	6.776
	UB	96.727
	LB	0.475
Skewness	Under IED	4.045
	UB	4563.173
	LB	-4563.173
Kurtosis	Under IED	1.348
	UB	1.430×10^5
	LB	3.490×10^{-5}

Špirtović et al., 2023] introduced in Section 2. The mean, SD, skewness, and kurtosis of MM% (Y) in the control group ($X = 0$) and the experimental group ($X = 1$) are shown in Table 3. ACE is reported as -1.344 .

Results. Table 4 reports the estimated bounds for the variance, skewness, and kurtosis of the ICE, together with their estimates under the IED condition. The bounds for the variance, skewness, and kurtosis are quite wide and therefore less informative.

For this study, the IED condition states that the treatment-induced change in the outcome is independent of the outcome under the control treatment. This assumption is plausible if the mechanisms generating the treatment effect are distinct from those determining the outcome under control. Under the IED condition, the distribution of the ICE exhibits large dispersion relative to the mean, is positively skewed, and is less kurtotic than a Gaussian distribution. Thus, a substantial number of participants have ICEs that differ markedly from the ACE, a smaller subset of participants may benefit substantially more than the ACE, and extreme outliers are rare.

8 CONCLUSION

This paper studies the identification and bounding of the central moments of the ICE using only the marginal central moments of each PO. We establish new identification results and derive sharp bounds when identification is not attainable. By relying solely on marginal central moment information, our framework broadens the applicability of moment-based analyses of causal effects. In particular, our results enable researchers to conduct heterogeneity analysis using historically accumulated studies that report only summary statistics rather than full datasets.

Our framework can be extended to study stratum-specific moments of the ICE if marginal moments are available within strata of observed covariates C : our identification

and bounding arguments can be applied within each stratum defined by $C = c$. These stratum-specific moments of the ICE describe the heterogeneity within a subpopulation defined by $C = c$, while standard conditional average causal effects (CACE), $\mathbb{E}[Y_1 - Y_0 | C = c]$, represent the average effects in the subpopulation $C = c$.

The identification results rely on a strong, untestable IED assumption, whose appropriateness cannot be verified directly from observed data, but must instead be assessed based on domain knowledge and the scientific context of the study. In practice, researchers must evaluate whether such an assumption is plausible or acceptable for their application, and thereby determine whether the resulting identification is credible. The bounds without IED can be wide, especially for higher-order moments. The width of sharp bounds reflects fundamental non-identifiability given the limited information available. A contribution of our work is to explicitly characterize what can and cannot be learned under moment-only information. In this sense, wide bounds are still informative: they quantify the extent to which higher-order features of treatment effect heterogeneity are inherently underdetermined given only moment information without stronger assumptions. Still, information on variance can often be learned and yield meaningful insights. In addition, our bounds from moment-only information can, in some cases, be tighter than those obtained by existing methods that assume access to full marginal distributions. This is illustrated in the numerical experiments in Appendix D: in Table 6, the lower bounds for $\bar{\mu}^{(2)}$ and $\bar{\mu}^{(4)}$ given $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ are substantially sharper than the corresponding bounds by the method in [Kawakami and Tian, 2025].

A limitation of this study is that we do not establish sharp bounds for $\bar{\mu}^{(m)}$ when multiple $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ pairs are jointly available. Moreover, the bounds derived for the skewness and kurtosis of the ICE are not sharp. Deriving sharp bounds in these settings appears to be substantially more challenging and remains a future research direction.

The bounds derived in this paper may be further tightened under additional structural assumptions such as a causal graph [Cozman, 2000, Zaffalon et al., 2020, Zhang et al., 2022, Shridharan and Iyengar, 2023a,b, Zaffalon et al., 2024, Duarte et al., 2024, Arroyo et al., 2025, Maiti et al., 2025] or monotonicity assumptions on POs [Manski, 1997, Pearl, 1999, Tian and Pearl, 2000, Mueller and Pearl, 2023, Mueller, 2025, Zhang et al., 2025, Hashimoto et al., 2026]. Assessing how these additional assumptions can sharpen the bounds is an important direction for future research.

Acknowledgements

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

References

- João Pedro Arroyo, Denis Mauá, João Gabriel Barreto Rodrigues, Daniel AE Lawand, Junkyu Lee, Radu Marinescu, Alexander G Gray, Eduardo Rocha Laurentino, and Fabio Cozman. Multilinear and linear programs for partially identifiable queries in quasi-markovian structural causal models. In *UAI 2025 Workshop on Causal Abstractions and Representations*, 2025.
- Fabio G Cozman. Credal networks. *Artificial intelligence*, 120(2):199–233, 2000.
- Guilherme Duarte, Noam Finkelstein, Dean Knox, Jonathan Mummolo, and Ilya Shpitser and. An automated approach to causal inference in discrete settings. *Journal of the American Statistical Association*, 119(547):1778–1793, 2024.
- Yanqin Fan and Sang Soo Park. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26(3):931–951, 2010.
- Priscilla Goble, Toria Flynn, Cambrian Nauman, Pond Al-mendarez, and Meagan Linstrom. Intervention implementation of tools of the mind for preschool children’s executive functioning. *Frontiers in psychology*, 12:624140, 2021.
- Naoya Hashimoto, Yuta Kawakami, and Jin Tian. Bounds and identification of joint probabilities of potential outcomes and observed variables under monotonicity assumptions. In *The 29th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2026.
- Paul C Hébert, George Wells, Morris A Blajchman, John Marshall, Claudio Martin, Giuseppe Pagliarello, Martin Tweeddale, Irwin Schweitzer, Elizabeth Yetisir, and Transfusion Requirements in Critical Care Investigators for the Canadian Critical Care Trials Group. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. *New England Journal of Medicine*, 340(6):409–417, 1999.
- Francisco Javier Hinojo Lucena, Jesus Lopez Belmonte, Arturo Fuentes Cabrera, Juan Manuel Trujillo Torres, and Santiago Pozo Sanchez. Academic effects of the use of flipped learning in physical education. *International journal of environmental research and public health*, 17(1):276, 2020.
- Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.
- Nice-Sugar Study Investigators. Intensive versus conventional glucose control in critically ill patients. *New England Journal of Medicine*, 360(13):1283–1297, 2009.
- Gustav Jonzon, Erin E Gabriel, Arvid Sjölander, and Michael C Sachs. Adding covariates to bounds: what is the question? *Journal of Causal Inference*, 13(1):20250004, 2025.
- Yuta Kawakami and Jin Tian. Moments of causal effects. *Proceedings of the 41st Conference on Uncertainty in Artificial Intelligence (UAI-2025)*, 2025.
- Aurghya Maiti, Drago Plecko, and Elias Bareinboim. Counterfactual identification under monotonicity constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39(25), pages 26841–26850, 2025.
- Charles Manski. Monotone treatment response. *Econometrica*, 65(6):1311–1334, 1997.
- J Bruce Moseley, Kimberly O’malley, Nancy J Petersen, Terri J Menke, Baruch A Brody, David H Kuykendall, John C Hollingsworth, Carol M Ashton, and Nelda P Wray. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *New England Journal of Medicine*, 347(2):81–88, 2002.
- Scott Mueller. *Personalized and Situation-Specific Decision Making*. University of California, Los Angeles, 2025.
- Scott Mueller and Judea Pearl. Monotonicity: Detection, refutation, and ramification, 2023.
- Meryem Özdemir Cihan and Mühahit Dilekmen. Emotional intelligence training for pre-service primary school teachers: a mixed methods research. *Frontiers in Psychology*, 15:1326082, 2024.
- 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.
- 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.
- Richard AJ Post and Edwin R Van Den Heuvel. Beyond conditional averages: Estimating the individual causal effect distribution. *Journal of Causal Inference*, 13(1):20240007, 2025.
- Emanuel Rivers, Bryant Nguyen, Suzanne Havstad, Julie Ressler, Alexandria Muzzin, Bernhard Knoblich, Edward Peterson, and Michael Tomlanovich. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *New England journal of medicine*, 345(19):1368–1377, 2001.

- Madhumitha Shridharan and Garud Iyengar. Causal bounds in quasi-markovian graphs. In *International Conference on Machine Learning*, pages 31675–31692. PMLR, 2023a.
- Madhumitha Shridharan and Garud Iyengar. Scalable computation of causal bounds. *Journal of Machine Learning Research*, 24(237):1–35, 2023b.
- Omer Špirtović, Ilma Čaprić, Mima Stanković, Dušan Đorđević, Benin Murić, Izet Kahrović, Rifat Mujanović, Raid Mekić, Borko Katanić, Igor Jelaska, et al. The effects of preventive aerobics mix on body composition in healthy adult women. *Frontiers in physiology*, 14: 1132619, 2023.
- Cristina Stefanescu, Aurel Nechita, and Claudia Iuliana Iacob. Brief acceptance and commitment therapy for children and adolescents with type 1 diabetes. *Frontiers in Psychology*, 15:1382509, 2024.
- Jin Tian and Judea Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1):287–313, 2000.
- Sümeyye Ulaş and İsmail Seğer. Developing a cbt-based intervention program for reducing school burnout and investigating its effectiveness with mixed methods research. *Frontiers in Psychology*, 13:884912, 2022.
- Greet Van den Berghe, Pieter Wouters, Frank Weekers, Charles Verwaest, Frans Bruyninckx, Miet Schetz, Dirk Vlasselaers, Patrick Ferdinande, Peter Lauwers, and Roger Bouillon. Intensive insulin therapy in critically ill patients. *New England journal of medicine*, 345(19): 1359–1367, 2001.
- Marco Zaffalon, Alessandro Antonucci, and Rafael Cabañas. Structural causal models are (solvable by) credal networks. In *International Conference on Probabilistic Graphical Models*, pages 581–592. PMLR, 2020.
- Marco Zaffalon, Alessandro Antonucci, Rafael Cabañas, David Huber, and Dario Azzimonti. Efficient computation of counterfactual bounds. *International Journal of Approximate Reasoning*, 171:109111, 2024.
- Chao Zhang, Zhi Geng, Wei Li, and Peng Ding. Identifying and bounding the probability of necessity for causes of effects with ordinal outcomes. *Biometrika*, page asaf049, 2025.
- Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 26548–26558. PMLR, 17–23 Jul 2022.
- Zhehao Zhang and Thomas S Richardson. Bounds on the distribution of a sum of two random variables: Revisiting a problem of kolmogorov with application to individual treatment effects. *arXiv preprint arXiv:2405.08806*, 2025.

Identification and Bounding of Central Moments of Causal Effects Using Marginal Moments Information (Supplementary Material)

Naoya Hashimoto^{1,*}

Yuta Kawakami^{1,*}

Jin Tian¹

¹Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, UAE

A COUNTEREXAMPLES

A.1 COUNTEREXAMPLE MENTIONED IN SECTION 4

Counterexample for identifiability using $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ under Assumption 2 in [Kawakami and Tian, 2025]. We provide two SCMs that satisfy Assumption 2 in [Kawakami and Tian, 2025] and have the same $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, while $\bar{\mu}^{(2)}$ is different.

Let $U \sim \text{Unif}[-1, 1]$ and consider the following two SCMs:

$$\begin{aligned} \text{(A)} : Y &:= U, \\ \text{(B)} : Y &:= (1 - X)U + cXU^3, \quad c := \sqrt{7/3}. \end{aligned} \tag{36}$$

Assumption 2 in [Kawakami and Tian, 2025] requires that, in an SCM with $Y = f_Y(X, U)$, the function $u \mapsto f_Y(x, u)$ is monotonic in u for every $x \in \{0, 1\}$ almost surely with respect to $\mathbb{P}_U$. We remark that both SCM (A) and (B) satisfy this assumption.

In both SCMs we have $\bar{\sigma}_0^{(2)} = \text{Var}(U) = 1/3$ and $\bar{\sigma}_1^{(2)} = \text{Var}(cU^3) = c^2\mathbb{E}[U^6] = (7/3) \cdot (1/7) = 1/3$. However, $\bar{\mu}^{(2)} = \text{Var}(Y_1 - Y_0)$ equals 0 in SCM (A), while in SCM (B),

$$\begin{aligned} \bar{\mu}^{(2)} &= \mathbb{E}[(cU^3 - U)^2] = c^2\mathbb{E}[U^6] - 2c\mathbb{E}[U^4] + \mathbb{E}[U^2] \\ &= \frac{2}{3} - \frac{2}{5}\sqrt{\frac{7}{3}}. \end{aligned} \tag{37}$$

Therefore, under Assumption 2, identifying $\bar{\mu}^{(2)}$ generally requires more information than $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, e.g., the marginal distributions of each PO.

A.2 COUNTEREXAMPLES SUPPORTING THE SHARPNESS DISCUSSION IN SECTION 5

Throughout this section, we put $\tilde{Y}_x := Y_x - \mathbb{E}[Y_x]$ ($x \in \{0, 1\}$) and $\tilde{\Delta} := (Y_1 - Y_0) - \mathbb{E}[Y_1 - Y_0] = \tilde{Y}_1 - \tilde{Y}_0$. Then $\bar{\sigma}_x^{(m)} = \mathbb{E}[\tilde{Y}_x^m]$ and $\bar{\mu}^{(m)} = \mathbb{E}[\tilde{\Delta}^m]$.

Note that, to show that a proposed closed bound $[L, U]$ is *not* sharp, we need to obtain an example of moment information such that $L' \leq \bar{\mu}^{(m)} \leq U'$ under that condition, where $L' > L$ or $U' < U$. We remark that to show the bounds themselves are unattainable is *not* sufficient to conclude that they are not sharp. Indeed, even if a bound is not attained, there remains a possibility that arbitrarily close value to the bound is attainable, whence the bound is still sharp.

In this section, for each non-sharp bound in the main part, we provide an example with $L' \leq \bar{\mu}^{(m)} \leq U'$, where $L' > L$ and/or $U' < U$.

We use the following lemma:

Lemma 1. (1) Let Z be a random variable with $\mathbb{E}[Z] = 0$, $\mathbb{E}[Z^3] = \sqrt{2}$, and $\mathbb{E}[Z^4] = 3$. Then we have $\mathbb{E}[Z^2] = 1$, and the distribution of Z is given by

$$\mathbb{P}\left(Z = \frac{\sqrt{2} + \sqrt{6}}{2}\right) = \frac{1}{2} - \frac{\sqrt{3}}{6}, \quad \mathbb{P}\left(Z = \frac{\sqrt{2} - \sqrt{6}}{2}\right) = \frac{1}{2} + \frac{\sqrt{3}}{6}. \quad (38)$$

(2) Let Z be a random variable with $\mathbb{E}[Z] = 0$, $\mathbb{E}[Z^3] = -\sqrt{2}$, and $\mathbb{E}[Z^4] = 3$. Then we have $\mathbb{E}[Z^2] = 1$, and the distribution of Z is given by

$$\mathbb{P}\left(Z = -\frac{\sqrt{2} + \sqrt{6}}{2}\right) = \frac{1}{2} - \frac{\sqrt{3}}{6}, \quad \mathbb{P}\left(Z = -\frac{\sqrt{2} - \sqrt{6}}{2}\right) = \frac{1}{2} + \frac{\sqrt{3}}{6}. \quad (39)$$

Proof. (1) Write $v := \mathbb{E}[Z^2]$. Then $\mathbb{E}[Z^3] = \mathbb{E}[Z(Z^2 - v)]$. By the Cauchy-Schwarz inequality, we have

$$2 = \mathbb{E}[Z^3]^2 \leq \mathbb{E}[Z^2]\mathbb{E}[(Z^2 - v)^2] = v(\mathbb{E}[Z^4] - v^2) = v(3 - v^2). \quad (40)$$

For $v \geq 0$, $v(3 - v^2) - 2 = -(v - 1)^2(v + 2) \geq 0$ holds only at $v = 1$. Therefore (40) forces $v = 1$ and equality in the Cauchy-Schwarz inequality. Hence Z and $Z^2 - 1$ are linearly dependent almost surely, and $\mathbb{E}[Z(Z^2 - 1)] = \mathbb{E}[Z^3] = \sqrt{2}$ implies $Z^2 - 1 = \sqrt{2}Z$ almost surely. The roots of $z^2 - \sqrt{2}z - 1 = 0$ are $\frac{\sqrt{2} \pm \sqrt{6}}{2}$, and Z takes these values almost surely. The probabilities in (38) follow from $\mathbb{E}[Z] = 0$. Conversely, assume that (38) holds. Then $\mathbb{E}[Z] = 0$. Moreover, since Z takes values in the set of roots of $z^2 - \sqrt{2}z - 1 = 0$, we have $Z^2 = \sqrt{2}Z + 1$ almost surely, and hence

$$Z^3 = ZZ^2 = \sqrt{2}Z^2 + Z = 3Z + \sqrt{2}, \quad Z^4 = (Z^2)^2 = (\sqrt{2}Z + 1)^2 = 4\sqrt{2}Z + 3$$

almost surely. Then we obtain $\mathbb{E}[Z^3] = \sqrt{2}$ and $\mathbb{E}[Z^4] = 3$.

(2) Write $v := \mathbb{E}[Z^2]$. Eq. (40) holds, and it forces $v = 1$ and equality in the Cauchy-Schwarz inequality. Hence Z and $Z^2 - 1$ are linearly dependent almost surely, and $\mathbb{E}[Z(Z^2 - 1)] = \mathbb{E}[Z^3] = -\sqrt{2}$ implies $Z^2 - 1 = -\sqrt{2}Z$ almost surely.

The roots of $z^2 + \sqrt{2}z - 1 = 0$ are $\frac{-\sqrt{2} \pm \sqrt{6}}{2}$, and Z takes these values almost surely. The probabilities in (39) follow from $\mathbb{E}[Z] = 0$. Conversely, assume that (39) holds. Then $\mathbb{E}[Z] = 0$. Moreover, since Z takes values in the set of roots of $z^2 + \sqrt{2}z - 1 = 0$, we have $Z^2 = -\sqrt{2}Z + 1$ almost surely, and hence

$$Z^3 = ZZ^2 = -\sqrt{2}Z^2 + Z = 3Z - \sqrt{2}, \quad Z^4 = (Z^2)^2 = (-\sqrt{2}Z + 1)^2 = -4\sqrt{2}Z + 3 \quad (41)$$

almost surely. Then we have $\mathbb{E}[Z^3] = -\sqrt{2}$ and $\mathbb{E}[Z^4] = 3$. $\square$

Counterexample for sharpness of the bound (19). Let $\bar{\sigma}_1^{(3)} = \sqrt{2}$, $\bar{\sigma}_0^{(3)} = -\sqrt{2}$, $\bar{\sigma}_1^{(4)} = \bar{\sigma}_0^{(4)} = 3$. Let $(\tilde{Y}_1, \tilde{Y}_0)$ be a distribution compatible with it. By Lemma 1, $\tilde{Y}_1$ and $\tilde{Y}_0$ have a distribution in (38) and (39), respectively. Let $x := \mathbb{P}\left(\tilde{Y}_1 = \frac{\sqrt{2} + \sqrt{6}}{2}, \tilde{Y}_0 = -\frac{\sqrt{2} - \sqrt{6}}{2}\right)$. Then $0 \leq x \leq \frac{1}{2} - \frac{\sqrt{3}}{6}$, and

$$\begin{aligned} \mathbb{E}[\tilde{Y}_1 \tilde{Y}_0] &= -\left(\frac{1}{2} - \frac{\sqrt{3}}{6} - x\right)\left(\frac{\sqrt{2} + \sqrt{6}}{2}\right)^2 - 2x\frac{\sqrt{2} + \sqrt{6}}{2}\frac{\sqrt{2} - \sqrt{6}}{2} - \left(\frac{1}{2} + \frac{\sqrt{3}}{6} - x\right)\left(\frac{\sqrt{2} - \sqrt{6}}{2}\right)^2 \\ &= -\left(\frac{1}{2} - \frac{\sqrt{3}}{6} - x\right)(2 + \sqrt{3}) + 2x - \left(\frac{1}{2} + \frac{\sqrt{3}}{6} - x\right)(2 - \sqrt{3}) = -1 + 6x, \\ \bar{\mu}^{(2)} &= \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^2] = \mathbb{E}[\tilde{Y}_1^2] - 2\mathbb{E}[\tilde{Y}_1 \tilde{Y}_0] + \mathbb{E}[\tilde{Y}_0^2] = 2 - 2\mathbb{E}[\tilde{Y}_1 \tilde{Y}_0] = 2 - 2(-1 + 6x) = 4 - 12x. \end{aligned} \quad (42)$$

Therefore $2\sqrt{3} - 2 \leq \bar{\mu}^{(2)} \leq 4$. On the other hand, (19) gives a bound $0 \leq \bar{\mu}^{(2)} \leq 4\sqrt{3}$. Hence (19) is not sharp.

Counterexample for sharpness of the bound (20).

Lower bound. Let $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 1$, $\bar{\sigma}_1^{(3)} = \sqrt{2}$, $\bar{\sigma}_0^{(3)} = -\sqrt{2}$ and $\bar{\sigma}_1^{(4)} = \bar{\sigma}_0^{(4)} = 3$. Let $(\tilde{Y}_1, \tilde{Y}_0)$ be a distribution compatible with it. By Lemma 1, such an $(\tilde{Y}_1, \tilde{Y}_0)$ exists, and $\tilde{Y}_1$ and $\tilde{Y}_0$ have the distribution in (38) and (39), respectively.

Eq. (20) gives a bound $0 \leq \bar{\mu}^{(2)} \leq 4$. However, we have $2\sqrt{3} - 2 \leq \bar{\mu}^{(2)} \leq 4$, so the lower bound in (20) is not sharp.

Upper bound. Let $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 1$, $\bar{\sigma}_1^{(3)} = \bar{\sigma}_0^{(3)} = \sqrt{2}$ and $\bar{\sigma}_1^{(4)} = \bar{\sigma}_0^{(4)} = 3$. Let $(\tilde{Y}_1, \tilde{Y}_0)$ be a distribution compatible with it. By Lemma 1, such an $(\tilde{Y}_1, \tilde{Y}_0)$ exists, and both $\tilde{Y}_1$ and $\tilde{Y}_0$ have distribution in (38). Eq. (20) gives a bound $0 \leq \bar{\mu}^{(2)} \leq 4$.

Let $x := \mathbb{P}\left(\tilde{Y}_1 = \frac{\sqrt{2}+\sqrt{6}}{2}, \tilde{Y}_0 = \frac{\sqrt{2}+\sqrt{6}}{2}\right)$. Then $0 \leq x \leq \frac{1}{2} - \frac{\sqrt{3}}{6}$, and

$$\begin{aligned} \bar{\mu}^{(2)} &= \mathbb{E}[\tilde{\Delta}^2] = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^2] \\ &= \left(\frac{\sqrt{2}+\sqrt{6}}{2} - \frac{\sqrt{2}-\sqrt{6}}{2}\right)^2 \left\{ \mathbb{P}\left(\tilde{Y}_1 = \frac{\sqrt{2}+\sqrt{6}}{2}, \tilde{Y}_0 = \frac{\sqrt{2}-\sqrt{6}}{2}\right) + \mathbb{P}\left(\tilde{Y}_1 = \frac{\sqrt{2}-\sqrt{6}}{2}, \tilde{Y}_0 = \frac{\sqrt{2}+\sqrt{6}}{2}\right) \right\} \\ &= 6 \cdot 2 \left(\frac{1}{2} - \frac{\sqrt{3}}{6} - x\right) = 12 \left(\frac{1}{2} - \frac{\sqrt{3}}{6} - x\right) \\ &\leq 12 \left(\frac{1}{2} - \frac{\sqrt{3}}{6}\right) = 6 - 2\sqrt{3}. \end{aligned} \tag{43}$$

Since $6 - 2\sqrt{3} < 4$, the upper bound in (20) is not sharp.

Counterexample for sharpness of the bound (24).

The case $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available. Let $\bar{\sigma}_1^{(2)} = \bar{\sigma}_1^{(4)} = 1$ and $\bar{\sigma}_0^{(2)} = \bar{\sigma}_0^{(4)} = 0$. (These are realized by, e.g., $\tilde{Y}_0 = 0$ and $\tilde{Y}_1$ with $\mathbb{P}(\tilde{Y}_1 = 1) = \mathbb{P}(\tilde{Y}_1 = -1) = 1/2$.)

Then $\bar{\sigma}_0^{(4)} = 0$ implies $\tilde{Y}_0 = 0$ almost surely. Moreover, by the Cauchy-Schwarz inequality,

$$\mathbb{E}[\tilde{Y}_1^2]^2 \leq \mathbb{E}[\tilde{Y}_1^4]. \tag{44}$$

Since $\mathbb{E}[\tilde{Y}_1^2] = \mathbb{E}[\tilde{Y}_1^4] = 1$, equality holds and hence $\tilde{Y}_1^2 = 1$ almost surely. Since $\mathbb{E}[\tilde{Y}_1] = 0$, this implies $\mathbb{P}(\tilde{Y}_1 = 1) = \mathbb{P}(\tilde{Y}_1 = -1) = 1/2$. Therefore $\tilde{\Delta} = \tilde{Y}_1$ almost surely and $\bar{\mu}^{(3)} = \mathbb{E}[\tilde{\Delta}^3] = \mathbb{E}[\tilde{Y}_1^3] = 0$. In contrast, the bound (24) reduces to $-\frac{\sqrt{2}}{\sqrt[3]{27}} \leq \bar{\mu}^{(3)} \leq \frac{\sqrt{2}}{\sqrt[3]{27}}$. In particular, neither bound in (24) is sharp.

The case $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available. Let $\bar{\sigma}_1^{(3)} = 0$, $\bar{\sigma}_1^{(4)} = 1$, $\bar{\sigma}_0^{(3)} = 0$, and $\bar{\sigma}_0^{(4)} = 0$. (These are realized by, e.g., $\tilde{Y}_0 = 0$ and $\tilde{Y}_1$ with $\mathbb{P}(\tilde{Y}_1 = 1) = \mathbb{P}(\tilde{Y}_1 = -1) = 1/2$.)

Then $\bar{\sigma}_0^{(4)} = 0$ implies $\tilde{Y}_0 = 0$ almost surely and hence $\tilde{\Delta} = \tilde{Y}_1$ almost surely. Therefore $\bar{\mu}^{(3)} = \mathbb{E}[\tilde{\Delta}^3] = \mathbb{E}[\tilde{Y}_1^3] = \bar{\sigma}_1^{(3)} = 0$. In contrast, the bound (24) reduces to $-\frac{\sqrt{2}}{\sqrt[3]{27}} \leq \bar{\mu}^{(3)} \leq \frac{\sqrt{2}}{\sqrt[3]{27}}$. In particular, neither bound in (24) is sharp.

The case $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available. Let $\bar{\sigma}_1^{(2)} = \bar{\sigma}_1^{(4)} = 1$, $\bar{\sigma}_1^{(3)} = 0$, and $\bar{\sigma}_0^{(2)} = \bar{\sigma}_0^{(3)} = \bar{\sigma}_0^{(4)} = 0$. (These are realized by, e.g., $\tilde{Y}_0 = 0$ and $\tilde{Y}_1$ with $\mathbb{P}(\tilde{Y}_1 = 1) = \mathbb{P}(\tilde{Y}_1 = -1) = 1/2$.)

Then $\bar{\sigma}_0^{(4)} = 0$ implies $\tilde{Y}_0 = 0$ almost surely and hence $\tilde{\Delta} = \tilde{Y}_1$ almost surely. Therefore $\bar{\mu}^{(3)} = \mathbb{E}[\tilde{\Delta}^3] = \mathbb{E}[\tilde{Y}_1^3] = \bar{\sigma}_1^{(3)} = 0$. In contrast, the bound (24) reduces to $-\frac{\sqrt{2}}{\sqrt[3]{27}} \leq \bar{\mu}^{(3)} \leq \frac{\sqrt{2}}{\sqrt[3]{27}}$. In particular, neither bound in (24) is sharp.

Counterexample for sharpness of the bound (31).

Lower bound. Let $\bar{\sigma}_1^{(3)} = \sqrt{2}$, $\bar{\sigma}_0^{(3)} = -\sqrt{2}$ and $\bar{\sigma}_1^{(4)} = \bar{\sigma}_0^{(4)} = 3$, and let $(\tilde{Y}_1, \tilde{Y}_0)$ be a distribution compatible with it. By Lemma 1, $\tilde{Y}_1$ and $\tilde{Y}_0$ have the distribution in (38) and (39), respectively. Eq. (31) gives $0 \leq \bar{\mu}^{(4)} \leq 48$.

Let $x := \mathbb{P}\left(\tilde{Y}_1 = \frac{\sqrt{2}+\sqrt{6}}{2}, \tilde{Y}_0 = -\frac{\sqrt{2}-\sqrt{6}}{2}\right)$. Then

$$\begin{aligned} \bar{\mu}^{(4)} &= \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^4] \\ &= (\sqrt{2} + \sqrt{6})^4 \left(\frac{1}{2} - \frac{\sqrt{3}}{6} - x\right) + (\sqrt{2})^4 \cdot 2x + (\sqrt{2} - \sqrt{6})^4 \left(\frac{1}{2} + \frac{\sqrt{3}}{6} - x\right) \\ &= (112 + 64\sqrt{3}) \left(\frac{1}{2} - \frac{\sqrt{3}}{6} - x\right) + 8x + (112 - 64\sqrt{3}) \left(\frac{1}{2} + \frac{\sqrt{3}}{6} - x\right) = 48 - 216x. \end{aligned} \tag{45}$$

Since $0 \leq x \leq \frac{1}{2} - \frac{\sqrt{3}}{6}$, we have $\bar{\mu}^{(4)} \geq 48 - 216\left(\frac{1}{2} - \frac{\sqrt{3}}{6}\right) = 36\sqrt{3} - 60 > 0$. Thus the lower bound in (31) is not sharp.

Upper bound. Let $\bar{\sigma}_1^{(3)} = \bar{\sigma}_0^{(3)} = \sqrt{2}$ and $\bar{\sigma}_1^{(4)} = \bar{\sigma}_0^{(4)} = 3$, and let $(\tilde{Y}_1, \tilde{Y}_0)$ be a distribution compatible with it. By Lemma 1, such an $(\tilde{Y}_1, \tilde{Y}_0)$ exists, and both $\tilde{Y}_1$ and $\tilde{Y}_0$ have the distribution in (38).

Eq. (31) gives $0 \leq \bar{\mu}^{(4)} \leq 48$. With $x := \mathbb{P}\left(\tilde{Y}_1 = \frac{\sqrt{2}+\sqrt{6}}{2}, \tilde{Y}_0 = \frac{\sqrt{2}+\sqrt{6}}{2}\right)$, we have $0 \leq x \leq \frac{1}{2} - \frac{\sqrt{3}}{6}$. Then we have

$$\begin{aligned} \bar{\mu}^{(4)} &= \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^4] = (\sqrt{6})^4 \mathbb{P}(\tilde{Y}_1 \neq \tilde{Y}_0) = 36 \cdot 2\left(\frac{1}{2} - \frac{\sqrt{3}}{6} - x\right) = 72\left(\frac{1}{2} - \frac{\sqrt{3}}{6} - x\right) \\ &\leq 72\left(\frac{1}{2} - \frac{\sqrt{3}}{6}\right) = 36 - 12\sqrt{3}. \end{aligned} \quad (46)$$

Since $36 - 12\sqrt{3} < 48$, the upper bound in (31) is not sharp.

Counterexample for sharpness of the bounds (30) and (32). Let $\bar{\sigma}_1^{(2)} = 1, \bar{\sigma}_0^{(2)} = \frac{1}{10}, \bar{\sigma}_1^{(4)} = \bar{\sigma}_0^{(4)} = 1$. Let $(\tilde{Y}_1, \tilde{Y}_0)$ be any distribution which satisfies $\bar{\sigma}_1^{(2)} = 1, \bar{\sigma}_0^{(2)} = \frac{1}{10}, \bar{\sigma}_1^{(4)} = \bar{\sigma}_0^{(4)} = 1$. Such an $(\tilde{Y}_1, \tilde{Y}_0)$ exists, e.g., $(\tilde{Y}_1, \tilde{Y}_0)$ with $\mathbb{P}(\tilde{Y}_1 = 1) = \mathbb{P}(\tilde{Y}_1 = -1) = \frac{1}{2}$ and

$$\mathbb{P}(\tilde{Y}_0 = 0) = \frac{99}{100}, \quad \mathbb{P}(\tilde{Y}_0 = \sqrt{10}) = \mathbb{P}(\tilde{Y}_0 = -\sqrt{10}) = \frac{1}{200}. \quad (47)$$

Since $\bar{\sigma}_1^{(4)} = (\bar{\sigma}_1^{(2)})^2$, the equality in the Cauchy-Schwarz inequality $(\bar{\sigma}_1^{(2)})^2 = \mathbb{E}[\tilde{Y}_1^2]^2 \leq \mathbb{E}[(\tilde{Y}_1^2)^2] \mathbb{E}[1^2] = \bar{\sigma}_1^{(4)}$ holds. Therefore, we have $\tilde{Y}_1^2 = c$ almost surely for some $c \in \mathbb{R}$, and we obtain $c = 1$ since $\bar{\sigma}_1^{(2)} = 1$. Thus $\tilde{Y}_1^3 = \tilde{Y}_1$ and $\mathbb{E}[\tilde{Y}_1^2 \tilde{Y}_0^2] = \mathbb{E}[\tilde{Y}_0^2] = \frac{1}{10}$.

Now we have

$$\bar{\mu}^{(4)} = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^4] = \frac{13}{5} - 4\mathbb{E}[\tilde{Y}_1 \tilde{Y}_0 (1 + \tilde{Y}_0^2)]. \quad (48)$$

By the Cauchy-Schwarz inequality,

$$(\mathbb{E}[\tilde{Y}_1 \tilde{Y}_0 (1 + \tilde{Y}_0^2)])^2 \leq \mathbb{E}[\tilde{Y}_1^2 \tilde{Y}_0^2] \mathbb{E}[(1 + \tilde{Y}_0^2)^2] = \frac{1}{10} (1 + 2\mathbb{E}[\tilde{Y}_0^2] + \mathbb{E}[\tilde{Y}_0^4]) = \frac{11}{50} \quad (49)$$

holds. Thus (48) implies

$$\frac{13}{5} - 4\sqrt{\frac{11}{50}} \leq \bar{\mu}^{(4)} \leq \frac{13}{5} + 4\sqrt{\frac{11}{50}}. \quad (50)$$

On the other hand, (30) gives a bound $(1 - \frac{1}{\sqrt{10}})^4 \leq \bar{\mu}^{(4)} \leq 16$. Since $\sqrt{11/50} < 1$ and $11/50 < 1/4$, (50) implies $\bar{\mu}^{(4)} < \frac{33}{5} < 16$ and $\bar{\mu}^{(4)} > \frac{3}{5}$. Moreover, since $\sqrt{10} < 4$ we have $(1 - \frac{1}{\sqrt{10}})^4 < (3/4)^4 < \frac{3}{5}$. Thus neither bound in (30) is sharp.

Also, (32) coincides with (30), so (32) is not sharp either.

B BOUNDING $\bar{\mu}^{(m)}$ USING ONLY MARGINAL SKEWNESS AND KURTOSIS

If we have access to variance, skewness and kurtosis, then we can compute 3rd and 4th central moments. In some cases, researchers may have access only to the skewness and kurtosis, but not to the marginal central moments $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ for $k = 2, 3, 4$. In this section, we then assume that only skewness γ_x and kurtosis κ_x , $\{(\gamma_x, \kappa_x)\}_{x \in \{0,1\}}$, exist and are available and none of $\{\bar{\sigma}_x^{(k)}\}_{x \in \{0,1\}, k \in \{2,3,4\}}$ are available. We discuss the evaluation of the central moments $\bar{\mu}^{(m)}$ for $m = 2, 3, 4$ under this setting. The marginal skewness γ_x and kurtosis κ_x are given by

$$\gamma_x := \frac{\bar{\sigma}_x^{(3)}}{(\bar{\sigma}_x^{(2)})^{3/2}}, \quad \kappa_x := \frac{\bar{\sigma}_x^{(4)}}{(\bar{\sigma}_x^{(2)})^2}. \quad (51)$$

If $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ are fixed, $\bar{\sigma}_x^{(3)}$ and $\bar{\sigma}_x^{(4)}$ are uniquely determined as

$$\bar{\sigma}_x^{(3)} = \gamma_x (\bar{\sigma}_x^{(2)})^{3/2}, \quad \bar{\sigma}_x^{(4)} = \kappa_x (\bar{\sigma}_x^{(2)})^2, \quad (52)$$

and we have the following bounds from Theorems 7(4), 11(2), and 15(4):

Proposition 2. Assume that $\gamma_1, \kappa_1, \gamma_0, \kappa_0$ exist and are available, and $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ are fixed. Then we have

- (1). $\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}\right)^2 \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}}\right)^2$.
- (2). $-\frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{(\bar{\sigma}_1^{(2)})^2 \kappa_1} + \sqrt[4]{(\bar{\sigma}_0^{(2)})^2 \kappa_0}\right)^3 \leq \bar{\mu}^{(3)} \leq \frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{(\bar{\sigma}_1^{(2)})^2 \kappa_1} + \sqrt[4]{(\bar{\sigma}_0^{(2)})^2 \kappa_0}\right)^3$.
- (3). $\max \left\{ \left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}\right)^4, \left(\sqrt[4]{(\bar{\sigma}_1^{(2)})^2 \kappa_1} - \sqrt[4]{(\bar{\sigma}_0^{(2)})^2 \kappa_0}\right)^4 \right\} \leq \bar{\mu}^{(4)} \leq \left(\sqrt[4]{(\bar{\sigma}_1^{(2)})^2 \kappa_1} + \sqrt[4]{(\bar{\sigma}_0^{(2)})^2 \kappa_0}\right)^4$.

Proof. Substituting $\bar{\sigma}_x^{(3)} = \gamma_x(\bar{\sigma}_x^{(2)})^{3/2}$ and $\bar{\sigma}_x^{(4)} = (\bar{\sigma}_x^{(2)})^2 \kappa_x$ into the bounds in Theorems 7(4), 11(2), and 15(4), we have the proposition. $\square$

If we know the bounds of $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, we have the following bounds:

Corollary 4. Assume that $\gamma_1, \kappa_1, \gamma_0, \kappa_0$ exist and are available, and $0 \leq l_1^{(2)} \leq \bar{\sigma}_1^{(2)} \leq u_1^{(2)}, 0 \leq l_0^{(2)} \leq \bar{\sigma}_0^{(2)} \leq u_0^{(2)}$ for some $l_1^{(2)}, u_1^{(2)}, l_0^{(2)}, u_0^{(2)} \in \mathbb{R}$. Then we have

- (1). $\min_{l_1^{(2)} \leq x_1 \leq u_1^{(2)}, l_0^{(2)} \leq x_0 \leq u_0^{(2)}} \left(\sqrt{x_1} - \sqrt{x_0}\right)^2 \leq \bar{\mu}^{(2)} \leq \left(\sqrt{u_1^{(2)}} + \sqrt{u_0^{(2)}}\right)^2$.
- (2). $-\frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{(u_1^{(2)})^2 \kappa_1} + \sqrt[4]{(u_0^{(2)})^2 \kappa_0}\right)^3 \leq \bar{\mu}^{(3)} \leq \frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{(u_1^{(2)})^2 \kappa_1} + \sqrt[4]{(u_0^{(2)})^2 \kappa_0}\right)^3$.
- (3). $\min_{l_1^{(2)} \leq x_1 \leq u_1^{(2)}, l_0^{(2)} \leq x_0 \leq u_0^{(2)}} \max \left\{ \left(\sqrt{x_1} - \sqrt{x_0}\right)^4, \left(\sqrt[4]{(x_1)^2 \kappa_1} - \sqrt[4]{(x_0)^2 \kappa_0}\right)^4 \right\} \leq \bar{\mu}^{(4)} \leq \left(\sqrt[4]{(u_1^{(2)})^2 \kappa_1} + \sqrt[4]{(u_0^{(2)})^2 \kappa_0}\right)^4$.

However, if we do not have the bounds of $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, we cannot derive nontrivial bounds by this approach.

C PROOFS

In this section, we provide proofs of the theorems stated in the main part. Throughout this section, we put $\tilde{Y}_x := Y_x - \mathbb{E}[Y_x]$ ($x = 0, 1$), $\Delta := Y_1 - Y_0$ and $\tilde{\Delta} := \Delta - \mathbb{E}[\Delta]$.

C.1 LEMMAS

In this subsection, we present several lemmas used in the proofs of the theorems.

Lemma 2. Let (Z_1, Z_0) be any pair of real-valued random variables, and let $(c_1, c_0) \in \mathbb{R}^2$ be any constants. Then there exists an SCM $\mathcal{M}$ with endogenous variables $\mathbf{V} = \{X, Y\}$, where $X \in \{0, 1\}$, such that the induced potential outcomes (Y_1, Y_0) satisfy

$$(Y_1, Y_0) \stackrel{d}{=} (Z_1 + c_1, Z_0 + c_0). \quad (53)$$

Proof. Let $\mathbf{U} = \{U_X, U_0, U_1\}$ be exogenous variables such that (U_0, U_1) has the same joint distribution as (Z_0, Z_1) , and $U_X \in \{0, 1\}$ is arbitrary. Define the structural equations by

$$\begin{aligned} X &:= U_X, \\ Y &:= (U_0 + c_0) + X\{(U_1 + c_1) - (U_0 + c_0)\}. \end{aligned} \quad (54)$$

Consider the sub-model $\mathcal{M}_x$ induced by the intervention $do(X = x)$. In $\mathcal{M}_x$, (54) yields

$$Y = (U_0 + c_0) + x\{(U_1 + c_1) - (U_0 + c_0)\} = U_x + c_x, \quad (55)$$

where $c_x \in \{c_0, c_1\}$ corresponds to $x \in \{0, 1\}$. Therefore $(Y_1, Y_0) = (U_1 + c_1, U_0 + c_0)$ almost surely, and hence $(Y_1, Y_0) \stackrel{d}{=} (Z_1 + c_1, Z_0 + c_0)$. $\square$

Lemma 3. Let $\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)} \geq 0$, and let $f_{\pm}$ be a function defined by

$$f_{\pm}(\lambda) = -\sqrt{\bar{\sigma}_0^{(2)}}\lambda \pm \sqrt{\bar{\sigma}_0^{(2)}\lambda^2 + (\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)})}, \quad (56)$$

where the domain is the set of λ such that $-1 \leq \lambda \leq 1$ and $\bar{\sigma}_0^{(2)}\lambda^2 + \bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)} \geq 0$. Then we have the following:

(i) If $\bar{\sigma}_1^{(2)} \geq \bar{\sigma}_0^{(2)}$, f_+ and f_- are both defined in $[-1, 1]$ and non-increasing.

(ii) If $\bar{\sigma}_1^{(2)} < \bar{\sigma}_0^{(2)}$, f_+ (resp. f_-) is defined in $\left[-1, -\sqrt{1 - \bar{\sigma}_1^{(2)}/\bar{\sigma}_0^{(2)}}\right] \cup \left[\sqrt{1 - \bar{\sigma}_1^{(2)}/\bar{\sigma}_0^{(2)}}, 1\right]$, decreasing (resp. increasing) if $\lambda < 0$ and increasing (resp. decreasing) if $\lambda > 0$, within its domain.

Proof. If $\bar{\sigma}_0^{(2)} = 0$, $f_{\pm}(\lambda) = \pm\sqrt{\bar{\sigma}_1^{(2)}}$ and (i) holds. The assumption of (ii) does not hold when $\bar{\sigma}_0^{(2)} = 0$.

In the following, we assume $\bar{\sigma}_0^{(2)} > 0$. For simplicity of notation, we put $S(\lambda) := \sqrt{\bar{\sigma}_0^{(2)}\lambda^2 + (\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)})}$ and $T(\lambda) := \frac{\bar{\sigma}_0^{(2)}\lambda}{S(\lambda)}$.

Domain. If $\bar{\sigma}_1^{(2)} \geq \bar{\sigma}_0^{(2)}$, we have $\bar{\sigma}_0^{(2)}\lambda^2 + (\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)}) \geq 0$ for any λ , hence $f_{\pm}$ is defined for all $\lambda \in [-1, 1]$. If $\bar{\sigma}_1^{(2)} < \bar{\sigma}_0^{(2)}$, $\bar{\sigma}_0^{(2)}\lambda^2 + (\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)}) \geq 0$ holds if and only if $\lambda^2 \geq 1 - \bar{\sigma}_1^{(2)}/\bar{\sigma}_0^{(2)}$. Therefore $f_{\pm}$ is defined in $\left[-1, -\sqrt{1 - \bar{\sigma}_1^{(2)}/\bar{\sigma}_0^{(2)}}\right] \cup \left[\sqrt{1 - \bar{\sigma}_1^{(2)}/\bar{\sigma}_0^{(2)}}, 1\right]$.

Increase/Decrease. We remark that we have

$$f'_{\pm}(\lambda) = -\sqrt{\bar{\sigma}_0^{(2)}} \pm \frac{\bar{\sigma}_0^{(2)}\lambda}{\sqrt{\bar{\sigma}_0^{(2)}\lambda^2 + (\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)})}}. \quad (57)$$

(i): If $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$, then $T(0) = 0$, and if we further assume $\lambda \neq 0$, we have $S(\lambda) > \sqrt{\bar{\sigma}_0^{(2)}}|\lambda|$ so

$$|T(\lambda)| = \frac{\bar{\sigma}_0^{(2)}|\lambda|}{S(\lambda)} < \frac{\bar{\sigma}_0^{(2)}|\lambda|}{\sqrt{\bar{\sigma}_0^{(2)}}|\lambda|} = \sqrt{\bar{\sigma}_0^{(2)}}. \quad (58)$$

Thus $|T(\lambda)| \leq \sqrt{\bar{\sigma}_0^{(2)}}$ for all $\lambda \in [-1, 1]$. Then we have $f'_{\pm}(\lambda) \leq 0$, therefore f_+, f_- are both non-increasing.

If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)}$, $f_{\pm}(\lambda) = -\sqrt{\bar{\sigma}_0^{(2)}}\lambda \pm \sqrt{\bar{\sigma}_0^{(2)}}|\lambda|$, so we have

$$f_+(\lambda) = \begin{cases} 0 & (\lambda \geq 0) \\ -2\sqrt{\bar{\sigma}_0^{(2)}}\lambda & (\lambda \leq 0) \end{cases}, \quad f_-(\lambda) = \begin{cases} -2\sqrt{\bar{\sigma}_0^{(2)}}\lambda & (\lambda \geq 0) \\ 0 & (\lambda \leq 0) \end{cases}. \quad (59)$$

Then f_+, f_- are both non-increasing.

(ii): If $\bar{\sigma}_1^{(2)} < \bar{\sigma}_0^{(2)}$, we have $|T(\lambda)| > \sqrt{\bar{\sigma}_0^{(2)}}$. Therefore we have

$$f'_+(\lambda) \begin{cases} < 0 & (\lambda < 0) \\ > 0 & (\lambda > 0) \end{cases}, \quad f'_-(\lambda) \begin{cases} > 0 & (\lambda < 0) \\ < 0 & (\lambda > 0) \end{cases} \quad (60)$$

(in the domain of $f_{\pm}$). Hence we have the claim. $\square$

Lemma 4. Let Z be a random variable with $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^4] < \infty$. Then

$$|\mathbb{E}[Z^3]| \leq \frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\mathbb{E}[Z^4]} \right)^3. \quad (61)$$

Moreover, there exists a random variable Z_0 with $\mathbb{E}[Z_0] = 0$, $\mathbb{E}[Z_0^4] = 1$, and $\mathbb{E}[Z_0^3] = \sqrt{2}/\sqrt[4]{27}$.

Proof. If $Z = 0$ almost surely, we have (61). In the following, we assume $\mathbb{P}(Z = 0) < 1$.

Let $s := \mathbb{E}[Z^2]$ and $\kappa := \mathbb{E}[Z^4]$. By the Cauchy–Schwarz inequality, $s^2 \leq \kappa$. Since $\mathbb{E}[Z] = 0$, $\mathbb{E}[Z^3] = \mathbb{E}[Z(Z^2 - s)]$, and the Cauchy–Schwarz inequality yields $|\mathbb{E}[Z^3]|^2 \leq \mathbb{E}[Z^2] \cdot \mathbb{E}[(Z^2 - s)^2] = s(\kappa - s^2)$. Hence

$$|\mathbb{E}[Z^3]| \leq \sqrt{s(\kappa - s^2)} = \left(\sqrt[4]{\kappa}\right)^3 \sqrt{u(1 - u^2)}, \quad u := \frac{s}{\sqrt{\kappa}} \in [0, 1]. \quad (62)$$

The function $u \mapsto u(1 - u^2)$ is maximized over $[0, 1]$ at $u = 1/\sqrt{3}$, with

$$\max_{u \in [0, 1]} \sqrt{u(1 - u^2)} = \sqrt{\frac{2}{3\sqrt{3}}} = \frac{\sqrt{2}}{\sqrt[4]{27}}. \quad (63)$$

Substituting this bound gives (61).

For any random variable Z with probabilities $\mathbb{P}(Z = a) = b/(a + b)$ and $\mathbb{P}(Z = -b) = a/(a + b)$ ($a, b \geq 0$, $a + b > 0$), one checks

$$\mathbb{E}[Z] = 0, \quad \mathbb{E}[Z^2] = ab, \quad \mathbb{E}[Z^3] = ab(a - b), \quad \mathbb{E}[Z^4] = ab(a^2 - ab + b^2). \quad (64)$$

We define a random variable Z_0 by

$$\mathbb{P}(Z_0 = a_0) := \frac{b_0}{a_0 + b_0}, \quad \mathbb{P}(Z_0 = -b_0) := \frac{a_0}{a_0 + b_0} \quad \text{where} \quad a_0 := \sqrt{\frac{3 + 2\sqrt{3}}{3}}, \quad b_0 := \sqrt{\frac{2\sqrt{3} - 3}{3}}. \quad (65)$$

Then $\mathbb{E}[Z_0^4] = a_0 b_0 ((a_0^2 + b_0^2) - a_0 b_0) = \frac{1}{\sqrt{3}} \left(\frac{4\sqrt{3}}{3} - \frac{1}{\sqrt{3}} \right) = 1$. Also, $(a_0 - b_0)^2 = (a_0^2 + b_0^2) - 2a_0 b_0 = \frac{4\sqrt{3}}{3} - \frac{2}{\sqrt{3}} = \frac{2}{\sqrt{3}}$, so $\mathbb{E}[Z_0^3] = a_0 b_0 (a_0 - b_0) = \frac{1}{\sqrt{3}} \sqrt{\frac{2}{\sqrt{3}}} = \frac{\sqrt{2}}{\sqrt[4]{27}}$. $\square$

Lemma 5. For any $v > 0$ and any $t \in \mathbb{R}$, there exists a random variable W such that $\mathbb{E}[W] = 0$, $\mathbb{E}[W^2] = v$ and $\mathbb{E}[W^3] = t$.

Proof. Let $\kappa := t/v^{3/2} \in \mathbb{R}$ and define a function $\phi : (0, 1) \rightarrow \mathbb{R}$ by

$$\phi(p) := \frac{1 - 2p}{\sqrt{p(1 - p)}}. \quad (66)$$

Then ϕ is continuous, $\lim_{p \downarrow 0} \phi(p) = +\infty$, and $\lim_{p \uparrow 1} \phi(p) = -\infty$. Hence there exists $p \in (0, 1)$ such that $\phi(p) = \kappa$. Define W by

$$W = \begin{cases} \sqrt{v(1 - p)/p} & \text{with probability } p, \\ -\sqrt{vp/(1 - p)} & \text{with probability } 1 - p. \end{cases} \quad (67)$$

Then we have $\mathbb{E}[W] = 0$, $\mathbb{E}[W^2] = v$ and $\mathbb{E}[W^3] = v^{3/2} \phi(p) = t$. $\square$

Lemma 6. Assume $\bar{\sigma}_x^{(4)} \geq (\bar{\sigma}_x^{(2)})^2$ for $x \in \{0, 1\}$. There exist random variables $A, B \geq 0$ such that

$$\begin{aligned} \mathbb{E}[A] &= \bar{\sigma}_1^{(2)}, & \mathbb{E}[A^2] &= \bar{\sigma}_1^{(4)}, & \mathbb{E}[B] &= \bar{\sigma}_0^{(2)}, & \mathbb{E}[B^2] &= \bar{\sigma}_0^{(4)}, \\ \mathbb{E}[AB] &= \bar{\sigma}_1^{(2)} \bar{\sigma}_0^{(2)} + \sqrt{\bar{\sigma}_1^{(4)} - (\bar{\sigma}_1^{(2)})^2} \sqrt{\bar{\sigma}_0^{(4)} - (\bar{\sigma}_0^{(2)})^2} =: \beta_{2,2}^{\text{CS}}. \end{aligned} \quad (68)$$

Moreover, there exist random signs S independent of (A, B) such that, defining $(\tilde{Y}_1, \tilde{Y}_0) := (S\sqrt{A}, S\sqrt{B})$, we have $\mathbb{E}[\tilde{Y}_x] = 0$ and $\mathbb{E}[\tilde{Y}_x^2] = \bar{\sigma}_x^{(2)}$, $\mathbb{E}[\tilde{Y}_x^4] = \bar{\sigma}_x^{(4)}$ for $x \in \{0, 1\}$.

Proof. By definition of $\bar{\sigma}_x^{(k)}$, we have $\bar{\sigma}_x^{(2)} = 0 \iff \bar{\sigma}_x^{(4)} = 0$.

If $\bar{\sigma}_1^{(2)} = 0$ and $\bar{\sigma}_0^{(2)} > 0$, take $A = 0$ almost surely. If $\bar{\sigma}_0^{(4)} = (\bar{\sigma}_0^{(2)})^2$, take $B = \bar{\sigma}_0^{(2)}$ almost surely. Otherwise, choose $t > 0$ such that $t < \frac{\bar{\sigma}_0^{(2)}}{\sqrt{\bar{\sigma}_0^{(4)} - (\bar{\sigma}_0^{(2)})^2}}$, let Z be a random variable with

$$\mathbb{P}(Z = -t) = \frac{1}{1 + t^2}, \quad \mathbb{P}(Z = 1/t) = \frac{t^2}{1 + t^2}. \quad (69)$$

whence $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^2] = 1$. Then, set $B = \bar{\sigma}_0^{(2)} + \sqrt{\bar{\sigma}_0^{(4)} - (\bar{\sigma}_0^{(2)})^2} Z$. Then $A, B \geq 0$, $\mathbb{E}[A] = \mathbb{E}[A^2] = 0$, $\mathbb{E}[B] = \bar{\sigma}_0^{(2)}$, $\mathbb{E}[B^2] = \bar{\sigma}_0^{(4)}$, and $\mathbb{E}[AB] = 0 = \beta_{2,2}^{\text{CS}}$. The case $\bar{\sigma}_0^{(2)} = 0$ and $\bar{\sigma}_1^{(2)} > 0$ is symmetric.

If further $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$, we can take $A = B = 0$ almost surely.

Assume now that $\bar{\sigma}_1^{(2)} > 0$ and $\bar{\sigma}_0^{(2)} > 0$. Take $t \in \mathbb{R}$ with $0 < t < \min \left\{ \frac{\bar{\sigma}_1^{(2)}}{\sqrt{\bar{\sigma}_1^{(4)} - (\bar{\sigma}_1^{(2)})^2}}, \frac{\bar{\sigma}_0^{(2)}}{\sqrt{\bar{\sigma}_0^{(4)} - (\bar{\sigma}_0^{(2)})^2}} \right\}$, where we interpret $a/0 := +\infty$ for $a > 0$. Set $p := 1/(1+t^2)$. Let Z be a random variable with (69). Define

$$A := \bar{\sigma}_1^{(2)} + \sqrt{\bar{\sigma}_1^{(4)} - (\bar{\sigma}_1^{(2)})^2} Z, \quad B := \bar{\sigma}_0^{(2)} + \sqrt{\bar{\sigma}_0^{(4)} - (\bar{\sigma}_0^{(2)})^2} Z. \quad (70)$$

Since $Z \in \{-t, 1/t\}$, we have $A \geq \bar{\sigma}_1^{(2)} - \sqrt{\bar{\sigma}_1^{(4)} - (\bar{\sigma}_1^{(2)})^2} t \geq 0$, $B \geq \bar{\sigma}_0^{(2)} - \sqrt{\bar{\sigma}_0^{(4)} - (\bar{\sigma}_0^{(2)})^2} t \geq 0$. By definition and $\mathbb{E}[Z] = 0$, $\mathbb{E}[A] = \bar{\sigma}_1^{(2)}$ and $\mathbb{E}[B] = \bar{\sigma}_0^{(2)}$. Further, by $\mathbb{E}[Z^2] = 1$,

$$\mathbb{E}[A^2] = (\bar{\sigma}_1^{(2)})^2 + (\bar{\sigma}_1^{(4)} - (\bar{\sigma}_1^{(2)})^2) = \bar{\sigma}_1^{(4)}, \quad \mathbb{E}[B^2] = \bar{\sigma}_0^{(4)}. \quad (71)$$

Also,

$$\mathbb{E}[AB] = \bar{\sigma}_1^{(2)} \bar{\sigma}_0^{(2)} + \sqrt{\bar{\sigma}_1^{(4)} - (\bar{\sigma}_1^{(2)})^2} \sqrt{\bar{\sigma}_0^{(4)} - (\bar{\sigma}_0^{(2)})^2} \mathbb{E}[Z^2] = \beta_{2,2}^{\text{CS}}. \quad (72)$$

Finally, take S independent of Z with $\mathbb{P}(S=1) = \mathbb{P}(S=-1) = 1/2$. Since (A, B) is a function of Z , S is independent of (A, B) . Set $(\tilde{Y}_1, \tilde{Y}_0) = (S\sqrt{A}, S\sqrt{B})$. Then $\mathbb{E}[\tilde{Y}_x] = 0$ and $\mathbb{E}[\tilde{Y}_x^2] = \mathbb{E}[A]$ or $\mathbb{E}[B]$, $\mathbb{E}[\tilde{Y}_x^4] = \mathbb{E}[A^2]$ or $\mathbb{E}[B^2]$, as required. $\square$

Lemma 7. Let $k \geq 3$ be an odd integer and let $t \in \mathbb{R}$. There exists a random variable Z such that $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^k] = t$.

Proof. If $t = 0$, take $Z = 0$. Otherwise, define a random variable T by $\mathbb{P}(T=2) = 1/3$ and $\mathbb{P}(T=-1) = 2/3$. Then $\mathbb{E}[T] = 0$ and $\mathbb{E}[T^k] = (2^k - 2)/3$. Set $Z := \sqrt[k]{t/\mathbb{E}[T^k]} T$. Then $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^k] = t$. $\square$

C.2 PROOFS

Theorem 1. Assume $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$, $k = 1, 2, \dots, m-3, m-2, m$, exist and are available. If the IED condition $(Y_1 - Y_0) \perp\!\!\!\perp Y_0$ holds, then the m -th central moment of ICE is recursively identified as

$$\bar{\mu}^{(m)} = \bar{\sigma}_1^{(m)} - \bar{\sigma}_0^{(m)} - \sum_{\ell=2}^{m-2} \binom{m}{\ell} \bar{\mu}^{(m-\ell)} \bar{\sigma}_0^{(\ell)}. \quad (6)$$

Proof. Let $\Delta := Y_1 - Y_0$. Since $Y_1 - \mathbb{E}[Y_1] = (Y_0 - \mathbb{E}[Y_0]) + (\Delta - \mathbb{E}[\Delta])$, by the binomial theorem

$$\bar{\sigma}_1^{(m)} = \sum_{\ell=0}^m \binom{m}{\ell} \mathbb{E}[(Y_0 - \mathbb{E}[Y_0])^{m-\ell} (\Delta - \mathbb{E}[\Delta])^\ell]. \quad (73)$$

Under the condition $Y_0 \perp\!\!\!\perp \Delta$, the expectation factorizes into $\mathbb{E}[(Y_0 - \mathbb{E}[Y_0])^{m-\ell} (\Delta - \mathbb{E}[\Delta])^\ell] = \mathbb{E}[(Y_0 - \mathbb{E}[Y_0])^{m-\ell}] \mathbb{E}[(\Delta - \mathbb{E}[\Delta])^\ell] = \bar{\sigma}_0^{(m-\ell)} \bar{\mu}^{(\ell)}$. Using $\bar{\sigma}_0^{(1)} = 0$ and $\bar{\mu}^{(1)} = 0$, we obtain

$$\bar{\sigma}_1^{(m)} = \bar{\sigma}_0^{(m)} + \bar{\mu}^{(m)} + \sum_{\ell=2}^{m-2} \binom{m}{\ell} \bar{\sigma}_0^{(m-\ell)} \bar{\mu}^{(\ell)}, \quad (74)$$

which is equivalent to (6). $\square$

Theorem 2. (1). Assume $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$. If $\bar{\sigma}_0^{(2)} = 0$, then $\lambda = 0$ and $\bar{\mu}^{(2)} = \bar{\sigma}_1^{(2)}$. If $\bar{\sigma}_0^{(2)} > 0$, then for any $-1 \leq \lambda \leq 1$, $\bar{\mu}^{(2)}$ is uniquely determined as

$$\bar{\mu}^{(2)} = \left(-\sqrt{\bar{\sigma}_0^{(2)}} \lambda + \sqrt{\bar{\sigma}_1^{(2)} - (1 - \lambda^2) \bar{\sigma}_0^{(2)}} \right)^2. \quad (10)$$

(2). Assume $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)}$. Then $\lambda \leq 0$, and $\bar{\mu}^{(2)}$ is uniquely determined as

$$\bar{\mu}^{(2)} = 4\lambda^2 \bar{\sigma}_0^{(2)}. \quad (11)$$

(3). Assume $\bar{\sigma}_1^{(2)} < \bar{\sigma}_0^{(2)}$. We have $-1 \leq \lambda \leq -\sqrt{1 - \bar{\sigma}_1^{(2)}/\bar{\sigma}_0^{(2)}}$, and

$$\bar{\mu}^{(2)} \in \left\{ \left(-\sqrt{\bar{\sigma}_0^{(2)}}\lambda - \sqrt{\bar{\sigma}_1^{(2)} - (1 - \lambda^2)\bar{\sigma}_0^{(2)}} \right)^2, \right. \\ \left. \left(-\sqrt{\bar{\sigma}_0^{(2)}}\lambda + \sqrt{\bar{\sigma}_1^{(2)} - (1 - \lambda^2)\bar{\sigma}_0^{(2)}} \right)^2 \right\}. \quad (12)$$

Proof. If $\bar{\sigma}_0^{(2)} = 0$, then $\tilde{Y}_0 = 0$ almost surely. Hence $\bar{\mu}^{(2)} = \bar{\sigma}_1^{(2)}$ and $\lambda = 0$ by definition. This proves (1) when $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)} = 0$ and (2) when $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$.

In the following, assume $\bar{\sigma}_0^{(2)} > 0$. Since $\mathbb{E}[(Y_1 - Y_0 - \mathbb{E}[Y_1 - Y_0])(Y_0 - \mathbb{E}[Y_0])] = \text{Cov}(Y_1 - Y_0, Y_0) = \lambda\sqrt{\bar{\mu}^{(2)}}\sqrt{\bar{\sigma}_0^{(2)}}$, we have $\bar{\mu}^{(2)} = \bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)} - 2\lambda\sqrt{\bar{\mu}^{(2)}}\sqrt{\bar{\sigma}_0^{(2)}}$. In other words, $\sqrt{\bar{\mu}^{(2)}}$ is a solution to the equation

$$x^2 + 2\lambda\sqrt{\bar{\sigma}_0^{(2)}}x - (\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)}) = 0. \quad (75)$$

The solutions to this equation are given by $f_{\pm}(\lambda)$, which are defined in Lemma 3. Since $\sqrt{\bar{\mu}^{(2)}} \geq 0$, at least one of $f_{\pm}(\lambda)$ is nonnegative. Hence λ satisfies $f_+(\lambda) \geq 0$. We remark that $f_{\pm}(-1) = \sqrt{\bar{\sigma}_0^{(2)}} \pm \sqrt{\bar{\sigma}_1^{(2)}}$ and $f_{\pm}(1) = -\sqrt{\bar{\sigma}_0^{(2)}} \pm \sqrt{\bar{\sigma}_1^{(2)}}$.

(1): If $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$, $f_-(\lambda) < 0$ for all $\lambda \in [-1, 1]$ since f_- is decreasing (Lemma 3) and $f_-(-1) < 0$. Moreover, $f_+(\lambda) \geq 0$ for all $\lambda \in [-1, 1]$ since f_+ is decreasing and $f_+(1) \geq 0$. Therefore, λ can take any value of $[-1, 1]$ and $\sqrt{\bar{\mu}^{(2)}}$ is equal to $f_+(\lambda) = -\sqrt{\bar{\sigma}_0^{(2)}}\lambda + \sqrt{\bar{\sigma}_1^{(2)} - (1 - \lambda^2)\bar{\sigma}_0^{(2)}}$.

(2): Assume $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)}$. Then $f_{\pm}(\lambda) = \sqrt{\bar{\sigma}_0^{(2)}}(-\lambda \pm |\lambda|)$. If $\lambda > 0$, we have $f_+(\lambda) = 0$, $f_-(\lambda) < 0$ and hence we get $\bar{\mu}^{(2)} = 0$. However, $\text{Var}(Y_1 - Y_0) = 0$ in this case, so by definition we have $\lambda = 0$; hence $\lambda > 0$ is infeasible. If $\lambda = 0$, we have $\bar{\mu}^{(2)} = 0$. If $\lambda < 0$, $f_+(\lambda) = 2\sqrt{\bar{\sigma}_0^{(2)}}|\lambda|$ and $f_-(\lambda) = 0$. Since $\bar{\mu}^{(2)} = 0$ contradicts with $\lambda < 0$, we have $\sqrt{\bar{\mu}^{(2)}} = 2\sqrt{\bar{\sigma}_0^{(2)}}|\lambda|$. By summarizing the above results, we have the claim.

(3): Assume $\bar{\sigma}_1^{(2)} < \bar{\sigma}_0^{(2)}$, and let $s := \sqrt{1 - \bar{\sigma}_1^{(2)}/\bar{\sigma}_0^{(2)}}$. Both f_+ and f_- are defined in $[-1, -s] \cup [s, 1]$, and we see

$$\begin{aligned} f_-(-s) &= \sqrt{\bar{\sigma}_0^{(2)} - \bar{\sigma}_1^{(2)}} > 0, & f_-(s) &= -\sqrt{\bar{\sigma}_0^{(2)} - \bar{\sigma}_1^{(2)}} < 0, \\ f_+(-s) &= \sqrt{\bar{\sigma}_0^{(2)} - \bar{\sigma}_1^{(2)}} > 0, & f_+(1) &= \sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} < 0. \end{aligned} \quad (76)$$

By Lemma 3, $f_{\pm}(\lambda) < 0$ for all $\lambda \in [s, 1]$, and $f_{\pm}(\lambda) > 0$ for all $\lambda \in [-1, -s]$. Therefore, the set of λ compatible with $\bar{\sigma}_1^{(2)} < \bar{\sigma}_0^{(2)}$ is $[-1, -s]$, and for λ in this range, $\sqrt{\bar{\mu}^{(2)}}$ can be equal to $f_+(\lambda)$ or $f_-(\lambda)$. Hence we have the claim. $\square$

Theorem 3. Assume that $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ exist and are available. $\bar{\mu}^{(2)}$ is bounded as follows:

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^2 \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}} \right)^2. \quad (13)$$

If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ are available, then this bound is sharp.

Proof. Let $f_{\pm}$ be as in Lemma 3. Note that $f_{\pm}$ are continuous in their domain. To prove the inequality, it suffices to show

$$\left| \sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right| \leq \sqrt{\bar{\mu}^{(2)}} \leq \sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}}. \quad (77)$$

If $\sqrt{\bar{\sigma}_1^{(2)}} > \sqrt{\bar{\sigma}_0^{(2)}}$, we have $\sqrt{\bar{\mu}^{(2)}} \leq \max_{\lambda} f_+(\lambda) = f_+(-1) = \sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}}$ and $\sqrt{\bar{\mu}^{(2)}} \geq \min_{\lambda} f_+(\lambda) = f_+(1) = \sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}$ by Lemma 3.

If $\sqrt{\bar{\sigma}_1^{(2)}} = \sqrt{\bar{\sigma}_0^{(2)}}$, we see $0 \leq \sqrt{\bar{\mu}^{(2)}} \leq 2\sqrt{\bar{\sigma}_0^{(2)}}$ as a consequence of (59) since $\sqrt{\bar{\mu}^{(2)}}$ is nonnegative.

If $\sqrt{\bar{\sigma}_1^{(2)}} < \sqrt{\bar{\sigma}_0^{(2)}}$, we see $f_-(-1) = \sqrt{\bar{\sigma}_0^{(2)}} - \sqrt{\bar{\sigma}_1^{(2)}}$ and $f_+(-1) = \sqrt{\bar{\sigma}_0^{(2)}} + \sqrt{\bar{\sigma}_1^{(2)}}$. Combining these with (76), we see that the range of $\sqrt{\bar{\mu}^{(2)}}$ is $\left[\sqrt{\bar{\sigma}_0^{(2)}} - \sqrt{\bar{\sigma}_1^{(2)}}, \sqrt{\bar{\sigma}_0^{(2)}} + \sqrt{\bar{\sigma}_1^{(2)}}\right] \cup \left[\sqrt{\bar{\sigma}_0^{(2)}} - \sqrt{\bar{\sigma}_1^{(2)}}, \sqrt{\bar{\sigma}_0^{(2)}} + \sqrt{\bar{\sigma}_1^{(2)}}\right] = \left[\sqrt{\bar{\sigma}_0^{(2)}} - \sqrt{\bar{\sigma}_1^{(2)}}, \sqrt{\bar{\sigma}_0^{(2)}} + \sqrt{\bar{\sigma}_1^{(2)}}\right]$ by Lemma 3, which implies (77).

Take a random variable S with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$. Set $\tilde{Y}_0 = \sqrt{\bar{\sigma}_0^{(2)}}S$ and $\tilde{Y}_1 = \sqrt{\bar{\sigma}_1^{(2)}}S$. Then $\bar{\mu}^{(2)} = (\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}})^2$. If we set $\tilde{Y}_1 = -\sqrt{\bar{\sigma}_1^{(2)}}S$ instead, then $\bar{\mu}^{(2)} = (\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}})^2$. By Lemma 2, these distributions are realized by some SCMs, so the bound (13) is sharp. $\square$

Theorem 4. Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ exist and are available.

(1). If $\lambda \geq 0$, then we have $\bar{\sigma}_1^{(2)} \geq \bar{\sigma}_0^{(2)}$. If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)}$, then $\bar{\mu}^{(2)} = 0$; if $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$, we have

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}\right)^2 \leq \bar{\mu}^{(2)} \leq \bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)}. \quad (14)$$

(2). Assume $\lambda \leq 0$. If $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$, we have

$$\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)} \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}}\right)^2. \quad (15)$$

If $\bar{\sigma}_1^{(2)} \leq \bar{\sigma}_0^{(2)}$, Eq. (13) holds.

These bounds are sharp.

Proof. Let $f_{\pm}$ be as in Lemma 3.

Bounds. (1): If $\bar{\sigma}_1^{(2)} < \bar{\sigma}_0^{(2)}$, the proof of Theorem 2 implies $\lambda < 0$. Hence $\bar{\sigma}_1^{(2)} \geq \bar{\sigma}_0^{(2)}$. If $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$, by Lemma 3 we have $f_+(1) \leq \sqrt{\bar{\mu}^{(2)}} \leq f_+(0)$, which is equivalent to (14). The case $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)}$ is Theorem 2(2).

(2): If $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$, by Lemma 3, we have $f_+(0) \leq \sqrt{\bar{\mu}^{(2)}} \leq f_+(-1)$, which is equivalent to (15). If $\bar{\sigma}_1^{(2)} \leq \bar{\sigma}_0^{(2)}$, we see (13) since $\lambda > 0$ does not hold in this case.

Sharpness. Take random variables S, T with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = \mathbb{P}(T = 1) = \mathbb{P}(T = -1) = 1/2$ and assume that S and T are independent. Set $\tilde{Y}_0 := \sqrt{\bar{\sigma}_0^{(2)}}S$.

Set $\tilde{Y}_1 := \sqrt{\bar{\sigma}_1^{(2)}}S$. Then $\text{Var}(\tilde{Y}_x) = \bar{\sigma}_x^{(2)}$ and $\bar{\mu}^{(2)} = (\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}})^2$. If $\bar{\sigma}_0^{(2)} > 0$ and $\bar{\sigma}_1^{(2)} \neq \bar{\sigma}_0^{(2)}$, then $\lambda = 1$ when $\bar{\sigma}_1^{(2)} > \bar{\sigma}_0^{(2)}$ and $\lambda = -1$ when $\bar{\sigma}_1^{(2)} < \bar{\sigma}_0^{(2)}$. If $\text{Var}(Y_1 - Y_0)\text{Var}(Y_0) = 0$, then $\lambda = 0$ by definition.

If $\bar{\sigma}_1^{(2)} \geq \bar{\sigma}_0^{(2)}$, set $\tilde{Y}_1 := \sqrt{\bar{\sigma}_0^{(2)}}S + \sqrt{\bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)}}T$. Then $\text{Var}(\tilde{Y}_x) = \bar{\sigma}_x^{(2)}$, $\bar{\mu}^{(2)} = \bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)}$, and $\lambda = 0$.

Set $\tilde{Y}_1 := -\sqrt{\bar{\sigma}_1^{(2)}}S$. Then $\text{Var}(\tilde{Y}_x) = \bar{\sigma}_x^{(2)}$, $\bar{\mu}^{(2)} = (\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}})^2$, and $\lambda = -1$ if $\bar{\sigma}_0^{(2)} > 0$, while $\lambda = 0$ if $\bar{\sigma}_0^{(2)} = 0$ by definition.

Under $\lambda \geq 0$, the first and second constructions attain both bounds of (14). Under $\lambda \leq 0$, the second and third constructions attain both bounds of (15). If $\bar{\sigma}_1^{(2)} \leq \bar{\sigma}_0^{(2)}$, the first and third constructions attain both bounds of (13) and satisfy $\lambda \leq 0$. Lemma 2 yields an SCM attaining each bound, hence we have the sharpness. $\square$

Theorem 5. Assume that $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available. Then we have $0 \leq \bar{\mu}^{(2)} \leq \infty$. This bound is sharp if only $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available.

Proof. We have $0 \leq \bar{\mu}^{(2)} \leq \infty$ by definition.

Sharpness (lower bound). Fix $\varepsilon > 0$. Choose a random variable B with $\mathbb{E}[B] = 0$ and $\mathbb{E}[B^3] = 2\bar{\sigma}_0^{(3)}$. Take a constant $M_0 > 0$ such that $M_0^2 > \mathbb{E}[B^2]$ and $|\delta| < \sqrt{\varepsilon}$, where $\delta := \frac{2(\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)})}{3(M_0^2 - \mathbb{E}[B^2])}$. Then define $\tilde{Y}_0$ by

$$\tilde{Y}_0 := \begin{cases} B & \text{with probability } 1/2, \\ M_0 & \text{with probability } 1/4, \\ -M_0 & \text{with probability } 1/4 \end{cases} \quad (78)$$

and define $\tilde{Y}_1$ by

$$\tilde{Y}_1 := \begin{cases} \tilde{Y}_0 - \delta & \text{if } \tilde{Y}_0 = B, \\ \tilde{Y}_0 + \delta & \text{if } \tilde{Y}_0 \in \{M_0, -M_0\}. \end{cases} \quad (79)$$

Then $\mathbb{E}[\tilde{Y}_0] = \mathbb{E}[\tilde{\Delta}] = 0$ (recall $\tilde{\Delta} = \tilde{Y}_1 - \tilde{Y}_0$), hence $\mathbb{E}[\tilde{Y}_1] = 0$. Moreover $\tilde{\Delta}^2 \equiv \delta^2$ and $\mathbb{E}[\tilde{\Delta}^3] = 0$, so

$$\mathbb{E}[\tilde{Y}_1^3] = \mathbb{E}[\tilde{Y}_0^3] + 3\mathbb{E}[\tilde{Y}_0^2 \tilde{\Delta}]. \quad (80)$$

Since $\mathbb{E}[\tilde{Y}_0^3] = \bar{\sigma}_0^{(3)}$ and $\mathbb{E}[\tilde{Y}_0^2 \tilde{\Delta}] = -\frac{1}{2}\mathbb{E}[B^2]\delta + \frac{1}{2}M_0^2\delta = \frac{\delta}{2}(M_0^2 - \mathbb{E}[B^2])$, we have $\mathbb{E}[\tilde{Y}_1^3] = \bar{\sigma}_1^{(3)}$. We also see $\bar{\mu}^{(2)} = \mathbb{E}[\tilde{\Delta}^2] = \delta^2 < \varepsilon$. By Lemma 2, these are realized by an SCM.

Sharpness (upper bound). Fix $M > 0$. Let B_0, B_1 be random variables with $\mathbb{E}[B_x] = 0$, $\mathbb{E}[B_x^3] = 2\bar{\sigma}_x^{(3)}$ ($x \in \{0, 1\}$). Take a constant $A > 0$ and define $(\tilde{Y}_1, \tilde{Y}_0)$ by

$$(\tilde{Y}_1, \tilde{Y}_0) := \begin{cases} (B_1, B_0) & \text{with probability } 1/2, \\ (A, -A) & \text{with probability } 1/4, \\ (-A, A) & \text{with probability } 1/4. \end{cases} \quad (81)$$

Then $\mathbb{E}[\tilde{Y}_1^3] = \bar{\sigma}_1^{(3)}$ and $\mathbb{E}[\tilde{Y}_0^3] = \bar{\sigma}_0^{(3)}$. We have $\tilde{\Delta} = \tilde{Y}_1 - \tilde{Y}_0 = \pm 2A$ when $(\tilde{Y}_1, \tilde{Y}_0) = (\pm A, \mp A)$, and we have $\bar{\mu}^{(2)} = \mathbb{E}[\tilde{\Delta}^2] \geq \frac{1}{2}(2A)^2 = 2A^2$. Taking $A > \sqrt{M/2}$ yields $\bar{\mu}^{(2)} > M$. By Lemma 2, these are realized by an SCM. $\square$

Theorem 6. Assume that $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$0 \leq \bar{\mu}^{(2)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^2. \quad (16)$$

This bound is sharp if only $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available.

Proof. Bounds. By the Cauchy–Schwarz inequality, for each $x \in \{0, 1\}$ we have $\mathbb{E}[\tilde{Y}_x^2] \leq \sqrt{\mathbb{E}[\tilde{Y}_x^4]} \sqrt{1^2} = \sqrt{\bar{\sigma}_x^{(4)}}$. Hence $\|\tilde{Y}_x\|_2 \leq \sqrt[4]{\bar{\sigma}_x^{(4)}}$. By Minkowski's inequality in L^2 , $\|\tilde{Y}_1 - \tilde{Y}_0\|_2 \leq \|\tilde{Y}_1\|_2 + \|\tilde{Y}_0\|_2 \leq \sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}}$. Hence $\bar{\mu}^{(2)} \leq (\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}})^2$. The lower bound $0 \leq \bar{\mu}^{(2)}$ follows from the definition.

Sharpness (lower bound). Fix $\varepsilon > 0$. Choose $p_x \in (0, 1)$ ($x \in \{0, 1\}$) with $\sqrt{p_1 \bar{\sigma}_1^{(4)}} + \sqrt{p_0 \bar{\sigma}_0^{(4)}} < \varepsilon$. Let $\tilde{Y}_1$ and $\tilde{Y}_0$ be independent random variables which satisfy

$$\mathbb{P}(\tilde{Y}_x = \pm a_x) = p_x/2, \quad \mathbb{P}(\tilde{Y}_x = 0) = 1 - p_x, \quad \text{where } a_x := \sqrt[4]{\bar{\sigma}_x^{(4)}}/p_x \quad (x \in \{0, 1\}). \quad (82)$$

Then $\mathbb{E}[\tilde{Y}_x] = 0$ and $\mathbb{E}[\tilde{Y}_x^4] = \bar{\sigma}_x^{(4)}$ hold for $x = 0, 1$. Moreover, by independence and $\mathbb{E}[\tilde{Y}_x] = 0$,

$$\bar{\mu}^{(2)} = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^2] = \mathbb{E}[\tilde{Y}_1^2] + \mathbb{E}[\tilde{Y}_0^2] = \sqrt{p_1 \bar{\sigma}_1^{(4)}} + \sqrt{p_0 \bar{\sigma}_0^{(4)}} < \varepsilon. \quad (83)$$

By Lemma 2, these are realized by an SCM.

Sharpness (upper bound). We show that the upper bound is attainable. Take a random variable S with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$, and set

$$\tilde{Y}_1 := \sqrt[4]{\bar{\sigma}_1^{(4)}} S, \quad \tilde{Y}_0 := -\sqrt[4]{\bar{\sigma}_0^{(4)}} S. \quad (84)$$

Then $\mathbb{E}[\tilde{Y}_x^4] = \bar{\sigma}_x^{(4)}$ and $\bar{\mu}^{(2)} = (\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}})^2$. By Lemma 2, these are realized by an SCM. $\square$

Theorem 7. (1). Assume that $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available. Then we have

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}\right)^2 \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}}\right)^2. \quad (17)$$

If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available, then this bound is sharp.

(2). Assume that $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available, we have

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}\right)^2 \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}}\right)^2. \quad (18)$$

If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available, then this bound is sharp.

(3). Assume that $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$0 \leq \bar{\mu}^{(2)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}}\right)^2. \quad (19)$$

(4). Assume that $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$, and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}\right)^2 \leq \bar{\mu}^{(2)} \leq \left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}}\right)^2. \quad (20)$$

Proof. (1): The inequality follows from Theorem 3. In the following, we show the sharpness.

If $\bar{\sigma}_1^{(2)}\bar{\sigma}_0^{(2)} = 0$, then the lower and upper endpoints in (17) coincide, and the sharpness is clear.

Fix $\varepsilon > 0$ and assume $\bar{\sigma}_1^{(2)}\bar{\sigma}_0^{(2)} > 0$. By Lemma 5, choose a random variable Z_0 with $\mathbb{E}[Z_0] = 0$, $\mathbb{E}[Z_0^2] = 1$ and $\mathbb{E}[Z_0^3] = \bar{\sigma}_0^{(3)}/(\bar{\sigma}_0^{(2)})^{3/2}$. Set $\tilde{Y}_0 := \sqrt{\bar{\sigma}_0^{(2)}} Z_0$. Pick $\delta \in (0, 1)$ with $2\delta\sqrt{\bar{\sigma}_1^{(2)}\bar{\sigma}_0^{(2)}} < \varepsilon$ and set $\alpha := \sqrt{\bar{\sigma}_1^{(2)}}(1 - \delta)$.

Sharpness (lower bound). Take a random variable W independent of Z_0 such that $\mathbb{E}[W] = 0$, $\mathbb{E}[W^2] = \bar{\sigma}_1^{(2)} - \alpha^2$ and $\mathbb{E}[W^3] = \bar{\sigma}_1^{(3)} - \alpha^3\mathbb{E}[Z_0^3]$ (Lemma 5). Let $\tilde{Y}_1 := \alpha Z_0 + W$. Then $\mathbb{E}[\tilde{Y}_1^2] = \bar{\sigma}_1^{(2)}$ and $\mathbb{E}[\tilde{Y}_1^3] = \bar{\sigma}_1^{(3)}$. Moreover, $\mathbb{E}[\tilde{Y}_1\tilde{Y}_0] = \alpha\sqrt{\bar{\sigma}_0^{(2)}}$, so

$$\bar{\mu}^{(2)} = \bar{\sigma}_1^{(2)} + \bar{\sigma}_0^{(2)} - 2\alpha\sqrt{\bar{\sigma}_0^{(2)}} = (\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}})^2 + 2\delta\sqrt{\bar{\sigma}_1^{(2)}\bar{\sigma}_0^{(2)}}. \quad (85)$$

By the choice of δ , we obtain $\bar{\mu}^{(2)} < (\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}})^2 + \varepsilon$. By Lemma 2, these are realized by an SCM.

Sharpness (upper bound). Take a random variable W' independent of Z_0 such that $\mathbb{E}[W'] = 0$, $\mathbb{E}[W'^2] = \bar{\sigma}_1^{(2)} - \alpha^2$ and $\mathbb{E}[W'^3] = \bar{\sigma}_1^{(3)} + \alpha^3\mathbb{E}[Z_0^3]$ (Lemma 5). Let $\tilde{Y}_1 := -\alpha Z_0 + W'$. Then $\mathbb{E}[\tilde{Y}_1^2] = \bar{\sigma}_1^{(2)}$ and $\mathbb{E}[\tilde{Y}_1^3] = \bar{\sigma}_1^{(3)}$. Moreover, $\mathbb{E}[\tilde{Y}_1\tilde{Y}_0] = -\alpha\sqrt{\bar{\sigma}_0^{(2)}}$, so

$$\bar{\mu}^{(2)} = \bar{\sigma}_1^{(2)} + \bar{\sigma}_0^{(2)} + 2\alpha\sqrt{\bar{\sigma}_0^{(2)}} = (\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}})^2 - 2\delta\sqrt{\bar{\sigma}_1^{(2)}\bar{\sigma}_0^{(2)}}. \quad (86)$$

By the choice of δ , we obtain $\bar{\mu}^{(2)} > (\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}})^2 - \varepsilon$. By Lemma 2, these are realized by an SCM.

(2): The inequality follows from Theorem 3: note that we have $\left(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}}\right)^2 \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}}\right)^2$ since $\bar{\sigma}_x^{(2)} \leq \sqrt{\bar{\sigma}_x^{(4)}}$ holds by the Cauchy-Schwarz inequality.

Sharpness. If $\bar{\sigma}_1^{(2)}\bar{\sigma}_0^{(2)} = 0$, then the lower and upper endpoints in (18) coincide, and $\bar{\mu}^{(2)}$ is forced to be this common endpoint. Thus (18) is sharp.

Fix $\varepsilon > 0$ and assume $\bar{\sigma}_0^{(2)}\bar{\sigma}_1^{(2)} > 0$. Choose $\eta \in (0, 1)$ such that $2\eta\sqrt{\bar{\sigma}_1^{(2)}\bar{\sigma}_0^{(2)}} < \varepsilon$. Let S be a random variable with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$. For $x \in \{0, 1\}$, define

$$m_x := \frac{\bar{\sigma}_x^{(4)} - (1 - \eta)(\bar{\sigma}_x^{(2)})^2}{\eta}. \quad (87)$$

Let U_x be a random variable with $\mathbb{P}(U_x = \pm\sqrt{m_x/\bar{\sigma}_x^{(2)}}) = (\bar{\sigma}_x^{(2)})^2/2m_x$ and $\mathbb{P}(U_x = 0) = 1 - ((\bar{\sigma}_x^{(2)})^2/m_x)$. Then $\mathbb{E}[U_x] = 0$, $\mathbb{E}[U_x^2] = \bar{\sigma}_x^{(2)}$, and $\mathbb{E}[U_x^4] = m_x$. Assume U_1 and U_0 are independent of each other and independent of S . For $s \in \{1, -1\}$, define

$$(\tilde{Y}_1, \tilde{Y}_0) = \begin{cases} (\sqrt{\bar{\sigma}_1^{(2)}}sS, \sqrt{\bar{\sigma}_0^{(2)}}S) & \text{with probability } 1 - \eta, \\ (U_1, U_0) & \text{with probability } \eta. \end{cases} \quad (88)$$

Then $\mathbb{E}[\tilde{Y}_x^2] = \bar{\sigma}_x^{(2)}$ and $\mathbb{E}[\tilde{Y}_x^4] = \bar{\sigma}_x^{(4)}$, and $\mathbb{E}[\tilde{Y}_1\tilde{Y}_0] = (1 - \eta)s\sqrt{\bar{\sigma}_1^{(2)}\bar{\sigma}_0^{(2)}}$. Hence

$$\bar{\mu}^{(2)} = \bar{\sigma}_1^{(2)} + \bar{\sigma}_0^{(2)} - 2(1 - \eta)s\sqrt{\bar{\sigma}_1^{(2)}\bar{\sigma}_0^{(2)}}. \quad (89)$$

For $s = 1$ this is at most $(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}})^2 + \varepsilon$, and for $s = -1$ it is at least $(\sqrt{\bar{\sigma}_1^{(2)}} + \sqrt{\bar{\sigma}_0^{(2)}})^2 - \varepsilon$. By Lemma 2, these are realized by an SCM. Thus we have the sharpness since we can take ε arbitrarily small.

(3): This follows from Theorem 6.

(4): This follows by intersecting Theorems 3–6. By the proof of (2), this intersection coincides with the bound in Theorem 3. $\square$

Theorem 8. Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ exist and are available.

(1). If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$, we have $\bar{\mu}^{(3)} = 0$.

(2). If $\bar{\sigma}_1^{(2)} > 0$ or $\bar{\sigma}_0^{(2)} > 0$, we have $-\infty \leq \bar{\mu}^{(3)} \leq \infty$. If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ are available, then this bound is sharp.

Proof. (1): If $\bar{\sigma}_0^{(2)} = \bar{\sigma}_1^{(2)} = 0$, then $\tilde{Y}_0 = \tilde{Y}_1 = 0$ almost surely, and hence $\bar{\mu}^{(3)} = 0$.

(2): By swapping the roles of 0 and 1 if necessary, we assume $\bar{\sigma}_1^{(2)} > 0$.

Fix $M \in \mathbb{R}$. By Lemma 5, choose $\tilde{Y}_1$ such that $\mathbb{E}[\tilde{Y}_1] = 0$, $\mathbb{E}[\tilde{Y}_1^2] = \bar{\sigma}_1^{(2)}$ and $\mathbb{E}[\tilde{Y}_1^3] = M$. Take a random variable S with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$, independent of $\tilde{Y}_1$, and set $\tilde{Y}_0 := \sqrt{\bar{\sigma}_0^{(2)}}S$. Then $\mathbb{E}[\tilde{Y}_0] = 0$, $\mathbb{E}[\tilde{Y}_0^2] = \bar{\sigma}_0^{(2)}$ and $\mathbb{E}[\tilde{Y}_0^3] = 0$. By independence and $\mathbb{E}[\tilde{Y}_1] = \mathbb{E}[\tilde{Y}_0] = 0$,

$$\bar{\mu}^{(3)} = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^3] = \mathbb{E}[\tilde{Y}_1^3] = M. \quad (90)$$

Since M is arbitrary, $\bar{\mu}^{(3)}$ can take any value in $\mathbb{R}$. By Lemma 2, these $(\tilde{Y}_1, \tilde{Y}_0)$ can be realized by some SCM. Then we have the claim. $\square$

Theorem 9. Assume that $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available. Then we have $-\infty \leq \bar{\mu}^{(3)} \leq \infty$. If only $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available, then this bound is sharp.

Proof. Fix $M > 0$. Using Lemma 5, choose W_1, W_0 such that $\mathbb{E}[W_x] = 0$, $\mathbb{E}[W_x^2] = 1$ and $\mathbb{E}[W_x^3] = 2\bar{\sigma}_x^{(3)}$ for each $x \in \{0, 1\}$, and independent of each other.

Fix $c \in \mathbb{R}$. Let J be a random variable such that $\mathbb{P}(J = 0) = 1/2$, $\mathbb{P}(J = j) = 1/6$ ($j = 1, 2, 3$), and independent of (W_1, W_0) . We define

$$(\tilde{Y}_1, \tilde{Y}_0) := \begin{cases} (W_1, W_0) & \text{if } J = 0, \\ (c, -c) & \text{if } J = 1, \\ (-c, 0) & \text{if } J = 2, \\ (0, c) & \text{if } J = 3. \end{cases} \quad (91)$$

Then for each $x \in \{0, 1\}$ we have $\mathbb{E}[\tilde{Y}_x^3] = \bar{\sigma}_x^{(3)}$. Moreover, since W_1 and W_0 are independent and $\mathbb{E}[W_1] = \mathbb{E}[W_0] = 0$,

$$\mathbb{E}[(W_1 - W_0)^3] = \mathbb{E}[W_1^3] - \mathbb{E}[W_0^3] = 2(\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)}). \quad (92)$$

Therefore,

$$\bar{\mu}^{(3)} = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^3] = (\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)}) + c^3. \quad (93)$$

Taking c with $|c|$ large yields $\bar{\mu}^{(3)} > M$ and $\bar{\mu}^{(3)} < -M$. By Lemma 2, these are realized by some SCMs. $\square$

Theorem 10. Assume that $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\begin{aligned} -\frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^3 &\leq \bar{\mu}^{(3)} \\ &\leq \frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^3. \end{aligned} \quad (23)$$

If only $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available, then this bound is sharp.

Proof. By Lemma 4 applied to $\tilde{Y}_1 - \tilde{Y}_0$ and Minkowski's inequality,

$$|\bar{\mu}^{(3)}| = |\mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^3]| \leq \frac{\sqrt{2}}{\sqrt[4]{27}} \|\tilde{Y}_1 - \tilde{Y}_0\|_4^3 \leq \frac{\sqrt{2}}{\sqrt[4]{27}} (\|\tilde{Y}_1\|_4 + \|\tilde{Y}_0\|_4)^3 = \frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^3. \quad (94)$$

Sharpness. Set Z_0 as in Lemma 4, then taking $\tilde{Y}_1 = \sqrt[4]{\bar{\sigma}_1^{(4)}} Z_0$ and $\tilde{Y}_0 = -\sqrt[4]{\bar{\sigma}_0^{(4)}} Z_0$, we have $\bar{\mu}^{(3)} = \frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^3$. Replacing Z_0 by $-Z_0$ attains the lower bound. By Lemma 2, these can be realized by an SCM. $\square$

Theorem 11. (1). Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available.

(i). If $\bar{\sigma}_0^{(2)} > 0$ and $\bar{\sigma}_1^{(2)} > 0$, we have $-\infty \leq \bar{\mu}^{(3)} \leq \infty$. If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available, then this bound is sharp.

(ii). If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$, we have $\bar{\mu}^{(3)} = 0$. Otherwise, if $\bar{\sigma}_1^{(2)} = 0$, we have $\bar{\mu}^{(3)} = -\bar{\sigma}_0^{(3)}$; if $\bar{\sigma}_0^{(2)} = 0$, we have $\bar{\mu}^{(3)} = \bar{\sigma}_1^{(3)}$.

(2) Assume that $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available, and that at least one of $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available. Then we have

$$\begin{aligned} -\frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^3 &\leq \bar{\mu}^{(3)} \\ &\leq \frac{\sqrt{2}}{\sqrt[4]{27}} \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^3. \end{aligned} \quad (24)$$

Proof. (1)-(ii): If $\bar{\sigma}_x^{(2)} = 0$ for some $x \in \{0, 1\}$, then $\tilde{Y}_x = 0$ almost surely and thus $\bar{\sigma}_x^{(3)} = 0$. If $x = 1$, then $\tilde{\Delta} = -\tilde{Y}_0$ and $\bar{\mu}^{(3)} = -\bar{\sigma}_0^{(3)}$. If $x = 0$, then $\tilde{\Delta} = \tilde{Y}_1$ and $\bar{\mu}^{(3)} = \bar{\sigma}_1^{(3)}$.

(1)-(i): Assume $\bar{\sigma}_0^{(2)} \bar{\sigma}_1^{(2)} > 0$ and fix $M > 0$. Choose $\eta \in (0, \min\{\bar{\sigma}_0^{(2)}, \bar{\sigma}_1^{(2)}\})$.

Sharpness (upper bound). Set $a > 0$ with $p := 2\eta/(3a^2) < 1/3$ and $a > (M - (\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)}))/3\eta$. Take independent random variables Z_1, Z_0 with

$$\mathbb{E}[Z_1] = \mathbb{E}[Z_0] = 0, \quad \mathbb{E}[Z_x^2] = \frac{\bar{\sigma}_x^{(2)} - \eta}{1 - 3p} \quad (x \in \{0, 1\}), \quad \mathbb{E}[Z_1^3] = \frac{\bar{\sigma}_1^{(3)} - \eta a/2}{1 - 3p}, \quad \mathbb{E}[Z_0^3] = \frac{\bar{\sigma}_0^{(3)} + \eta a/2}{1 - 3p}, \quad (95)$$

which is possible by Lemma 5. Define $(\tilde{Y}_1, \tilde{Y}_0)$ by

$$(\tilde{Y}_1, \tilde{Y}_0) = \begin{cases} (Z_1, Z_0) & \text{with probability } 1 - 3p, \\ (a, -a) & \text{with probability } p, \\ (-a/2, a/2) & \text{with probability } 2p. \end{cases} \quad (96)$$

Then $\mathbb{E}[\tilde{Y}_x^2] = \bar{\sigma}_x^{(2)}$ and $\mathbb{E}[\tilde{Y}_x^3] = \bar{\sigma}_x^{(3)}$ for $x \in \{0, 1\}$. Moreover, since Z_1, Z_0 are independent and $\mathbb{E}[Z_1] = \mathbb{E}[Z_0] = 0$,

$$\bar{\mu}^{(3)} = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^3] = (1 - 3p)(\mathbb{E}[Z_1^3] - \mathbb{E}[Z_0^3]) + p(2a)^3 + 2p(-a)^3 = \bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)} + 3\eta a > M. \quad (97)$$

Lemma 2 implies the existence of an SCM realizing this.

Sharpness (lower bound). Set $a > 0$ with $p := \eta/(3a^2) < 2/3$ and $a > (M + (\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)}))/6\eta$. Take independent random variables Z_1, Z_0 with

$$\mathbb{E}[Z_1] = \mathbb{E}[Z_0] = 0, \quad \mathbb{E}[Z_x^2] = \frac{\bar{\sigma}_x^{(2)} - \eta}{1 - 3p/2} \quad (x \in \{0, 1\}), \quad \mathbb{E}[Z_1^3] = \frac{\bar{\sigma}_1^{(3)} + \eta a}{1 - 3p/2}, \quad \mathbb{E}[Z_0^3] = \frac{\bar{\sigma}_0^{(3)} - \eta a}{1 - 3p/2}, \quad (98)$$

which is possible by Lemma 5. Define $(\tilde{Y}_1, \tilde{Y}_0)$ by

$$(\tilde{Y}_1, \tilde{Y}_0) = \begin{cases} (Z_1, Z_0) & \text{with probability } 1 - 3p/2, \\ (a, -a) & \text{with probability } p, \\ (-2a, 2a) & \text{with probability } p/2. \end{cases} \quad (99)$$

Then $\mathbb{E}[\tilde{Y}_x^2] = \bar{\sigma}_x^{(2)}$ and $\mathbb{E}[\tilde{Y}_x^3] = \bar{\sigma}_x^{(3)}$ for $x \in \{0, 1\}$. Moreover, since Z_1, Z_0 are independent and $\mathbb{E}[Z_1] = \mathbb{E}[Z_0] = 0$,

$$\bar{\mu}^{(3)} = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^3] = (1 - 3p/2)(\mathbb{E}[Z_1^3] - \mathbb{E}[Z_0^3]) + p(2a)^3 + (p/2)(-4a)^3 = \bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)} - 6\eta a < -M. \quad (100)$$

Lemma 2 implies the existence of an SCM realizing this.

(2) follows from Theorem 10. □

Theorem 12. Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ exist and are available.

(1). If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$, we have $\bar{\mu}^{(4)} = 0$.

(2). If $\bar{\sigma}_1^{(2)} > 0$ or $\bar{\sigma}_0^{(2)} > 0$, we have

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^4 \leq \bar{\mu}^{(4)} \leq \infty. \quad (25)$$

If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ are available, then this bound is sharp.

Proof. (1): If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$, we have $\tilde{Y}_1 = 0$ almost surely and $\tilde{Y}_0 = 0$ almost surely. Therefore we have $\bar{\mu}^{(4)} = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^4] = 0$.

(2): By the Cauchy–Schwarz inequality, $\bar{\mu}^{(4)} = \mathbb{E}[\tilde{\Delta}^4] \geq (\mathbb{E}[\tilde{\Delta}^2])^2$. Let $\|Z\|_2 := (\mathbb{E}[Z^2])^{1/2}$. By the triangle inequality in L^2 , $\|\tilde{\Delta}\|_2 \geq \|\tilde{Y}_1\|_2 - \|\tilde{Y}_0\|_2$. Since $\|\tilde{Y}_x\|_2 = \sqrt{\bar{\sigma}_x^{(2)}}$, we have $\bar{\mu}^{(4)} \geq \left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}} \right)^4$.

Sharpness (lower bound). Take a random variable $S \in \{-1, 1\}$ with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$ and set $\tilde{Y}_x := \sqrt{\bar{\sigma}_x^{(2)}} S$. Then $\mathbb{E}[\tilde{Y}_x] = 0$, $\mathbb{E}[\tilde{Y}_x^2] = \bar{\sigma}_x^{(2)}$ and $\tilde{\Delta} = (\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}})S$, so the lower bound is attained.

Sharpness (upper bound). Fix $M > 0$. We assume $\bar{\sigma}_1^{(2)} > 0$, by exchanging the roles of 0 and 1 if necessary. Choose a constant $A > \sqrt{\bar{\sigma}_1^{(2)}}$ such that $(\bar{\sigma}_1^{(2)}/A^2)(A - \sqrt{\bar{\sigma}_0^{(2)}})^4 > M$.

Let I be a random variable with $\mathbb{P}(I = 1) = \bar{\sigma}_1^{(2)}/A^2$ and $\mathbb{P}(I = 0) = 1 - (\bar{\sigma}_1^{(2)}/A^2)$, and let S be a random variable with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$, independent of I . Define $\tilde{Y}_1 := IAS$ and $\tilde{Y}_0 := \sqrt{\bar{\sigma}_0^{(2)}}S$. Then $\mathbb{E}[\tilde{Y}_x^2] = \bar{\sigma}_x^{(2)}$ ($x \in \{0, 1\}$) and $\bar{\mu}^{(4)} = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^4] \geq \mathbb{P}(I = 1)(A - \sqrt{\bar{\sigma}_0^{(2)}})^4 > M$. By Lemma 2, these can be realized by an SCM. $\square$

Theorem 13. Assume that $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available. Then we have $0 \leq \bar{\mu}^{(4)} \leq \infty$. If only $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available, then this bound is sharp.

Proof. The inequality is trivial.

Sharpness (lower bound). Fix $\varepsilon > 0$. Choose a random variable B with $\mathbb{E}[B] = 0$ and $\mathbb{E}[B^3] = 2\bar{\sigma}_0^{(3)}$. Take a constant $M_0 > 0$ such that $M_0^2 > \mathbb{E}[B^2]$ and $|\delta| < \sqrt[4]{\varepsilon}$, where $\delta := \frac{2(\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)})}{3(M_0^2 - \mathbb{E}[B^2])}$. Then define $\tilde{Y}_0$ by

$$\tilde{Y}_0 := \begin{cases} B & \text{with probability } 1/2, \\ M_0 & \text{with probability } 1/4, \\ -M_0 & \text{with probability } 1/4 \end{cases} \quad (101)$$

and define $\tilde{Y}_1$ by

$$\tilde{Y}_1 := \begin{cases} \tilde{Y}_0 - \delta & \text{if } \tilde{Y}_0 = B, \\ \tilde{Y}_0 + \delta & \text{if } \tilde{Y}_0 \in \{M_0, -M_0\}. \end{cases} \quad (102)$$

Then $\mathbb{E}[\tilde{Y}_0] = \mathbb{E}[\tilde{\Delta}] = 0$, hence $\mathbb{E}[\tilde{Y}_1] = 0$. Moreover $\tilde{\Delta}^2 \equiv \delta^2$ and $\mathbb{E}[\tilde{\Delta}^3] = 0$, so

$$\mathbb{E}[\tilde{Y}_1^3] = \mathbb{E}[\tilde{Y}_0^3] + 3\mathbb{E}[\tilde{Y}_0^2\tilde{\Delta}]. \quad (103)$$

Since $\mathbb{E}[\tilde{Y}_0^3] = \bar{\sigma}_0^{(3)}$ and $\mathbb{E}[\tilde{Y}_0^2\tilde{\Delta}] = -\frac{1}{2}\mathbb{E}[B^2]\delta + \frac{1}{2}M_0^2\delta = \frac{\delta}{2}(M_0^2 - \mathbb{E}[B^2])$, we have $\mathbb{E}[\tilde{Y}_1^3] = \bar{\sigma}_1^{(3)}$. We also see $\bar{\mu}^{(4)} = \mathbb{E}[\tilde{\Delta}^4] = \delta^4 < \varepsilon$.

Sharpness (upper bound). Fix $M > 0$. Let B_0, B_1 be random variables with $\mathbb{E}[B_x] = 0$, $\mathbb{E}[B_x^3] = 2\bar{\sigma}_x^{(3)}$ ($x \in \{0, 1\}$), independent of each other. Take a constant $A > 0$ and define $(\tilde{Y}_1, \tilde{Y}_0)$ by

$$(\tilde{Y}_1, \tilde{Y}_0) := \begin{cases} (B_1, B_0) & \text{with probability } 1/2, \\ (A, -A) & \text{with probability } 1/4, \\ (-A, A) & \text{with probability } 1/4. \end{cases} \quad (104)$$

Then $\mathbb{E}[\tilde{Y}_1^3] = \bar{\sigma}_1^{(3)}$ and $\mathbb{E}[\tilde{Y}_0^3] = \bar{\sigma}_0^{(3)}$. We have $\tilde{\Delta} = \tilde{Y}_1 - \tilde{Y}_0 = \pm 2A$ when $(\tilde{Y}_1, \tilde{Y}_0) = (\pm A, \mp A)$, and we have $\bar{\mu}^{(4)} = \mathbb{E}[\tilde{\Delta}^4] \geq \frac{1}{2}(2A)^4 = 8A^4$. We obtain $\bar{\mu}^{(4)} > M$ by taking $A > \sqrt[4]{M/8}$.

By Lemma 2, these distributions can be realized by SCMs. Therefore we have the sharpness. $\square$

Theorem 14. Assume that $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\left(\sqrt[4]{\bar{\sigma}_1^{(4)}} - \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^4 \leq \bar{\mu}^{(4)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}} \right)^4. \quad (26)$$

If only $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ are available, then this bound is sharp.

Proof. Recall that $\tilde{\Delta} = \tilde{Y}_1 - \tilde{Y}_0$. Since 4 is even, we have

$$\|\tilde{Y}_x\|_4 = (\mathbb{E}[|\tilde{Y}_x|^4])^{1/4} = \sqrt[4]{\bar{\sigma}_x^{(4)}} \quad (x \in \{0, 1\}), \quad \|\tilde{\Delta}\|_4^4 = \mathbb{E}[|\tilde{\Delta}|^4] = \bar{\mu}^{(4)}. \quad (105)$$

By Minkowski's inequality in L^4 -space,

$$\|\tilde{\Delta}\|_4 \leq \|\tilde{Y}_1\|_4 + \|\tilde{Y}_0\|_4, \quad \|\tilde{Y}_1\|_4 \leq \|\tilde{\Delta}\|_4 + \|\tilde{Y}_0\|_4, \quad \text{and} \quad \|\tilde{Y}_0\|_4 \leq \|\tilde{Y}_1\|_4 + \|\tilde{\Delta}\|_4 \quad (106)$$

hold. Thus we have $|\|\tilde{Y}_1\|_4 - \|\tilde{Y}_0\|_4| \leq \|\tilde{\Delta}\|_4 \leq \|\tilde{Y}_1\|_4 + \|\tilde{Y}_0\|_4$. Raising both sides to the fourth power yields (26).

Sharpness. Take a random variable $S \in \{-1, 1\}$ with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$. Setting $\tilde{Y}_1 := \sqrt[4]{\bar{\sigma}_1^{(4)}}S$ and $\tilde{Y}_0 := -\sqrt[4]{\bar{\sigma}_0^{(4)}}S$ attains the upper bound. Setting $\tilde{Y}_1 := \sqrt[4]{\bar{\sigma}_1^{(4)}}S$ and $\tilde{Y}_0 := \sqrt[4]{\bar{\sigma}_0^{(4)}}S$ attains the lower bound. By Lemma 2, these can be realized by an SCM. $\square$

Theorem 15. (1). Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ exist and are available.

(i). If $\bar{\sigma}_1^{(2)} = \bar{\sigma}_0^{(2)} = 0$, then $\bar{\mu}^{(4)} = 0$.

(ii). If $\bar{\sigma}_1^{(2)} > 0$ and $\bar{\sigma}_0^{(2)} > 0$, then

$$\left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}\right)^4 \leq \bar{\mu}^{(4)} \leq \infty. \quad (27)$$

(iii). If $\bar{\sigma}_1^{(2)} > 0$ and $\bar{\sigma}_0^{(2)} = 0$, then $\bar{\sigma}_0^{(3)} = 0$, and

$$(\bar{\sigma}_1^{(2)})^2 + \frac{(\bar{\sigma}_1^{(3)})^2}{\bar{\sigma}_1^{(2)}} \leq \bar{\mu}^{(4)} \leq \infty. \quad (28)$$

(iv). If $\bar{\sigma}_1^{(2)} = 0$ and $\bar{\sigma}_0^{(2)} > 0$, then $\bar{\sigma}_1^{(3)} = 0$, and

$$(\bar{\sigma}_0^{(2)})^2 + \frac{(\bar{\sigma}_0^{(3)})^2}{\bar{\sigma}_0^{(2)}} \leq \bar{\mu}^{(4)} \leq \infty. \quad (29)$$

If only $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ are available, then these bounds are sharp.

(2). Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\begin{aligned} \max \left\{ \left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}\right)^4, \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} - \sqrt[4]{\bar{\sigma}_0^{(4)}}\right)^4 \right\} \\ \leq \bar{\mu}^{(4)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}}\right)^4. \end{aligned} \quad (30)$$

(3). Assume $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\left(\sqrt[4]{\bar{\sigma}_1^{(4)}} - \sqrt[4]{\bar{\sigma}_0^{(4)}}\right)^4 \leq \bar{\mu}^{(4)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}}\right)^4. \quad (31)$$

(4). Assume $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$, $(\bar{\sigma}_1^{(3)}, \bar{\sigma}_0^{(3)})$, and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ exist and are available. Then we have

$$\begin{aligned} \max \left\{ \left(\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}}\right)^4, \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} - \sqrt[4]{\bar{\sigma}_0^{(4)}}\right)^4 \right\} \\ \leq \bar{\mu}^{(4)} \leq \left(\sqrt[4]{\bar{\sigma}_1^{(4)}} + \sqrt[4]{\bar{\sigma}_0^{(4)}}\right)^4. \end{aligned} \quad (32)$$

Proof. (1)-(i) follows from Theorem 12(1).

(1)-(ii): By the Cauchy-Schwarz inequality, $\mathbb{E}[\tilde{\Delta}^4] \geq \mathbb{E}[\tilde{\Delta}^2]^2$, hence $\bar{\mu}^{(4)} \geq (\bar{\mu}^{(2)})^2$. Then Theorem 3 gives $\bar{\mu}^{(2)} \geq (\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}})^2$.

Sharpness (lower bound). Fix $\varepsilon > 0$. Choose $\eta \in (0, \frac{1}{4})$ and let S be a random variable with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$.

If $\bar{\sigma}_0^{(3)} \neq 0$, let U be a random variable $\mathbb{E}[U] = 0$, $\mathbb{E}[U^2] = \eta^{-2/3}$, and $\mathbb{E}[U^3] = \bar{\sigma}_0^{(3)}/\eta$ (Lemma 5). Otherwise set $U := 0$. Let W be a random variable with

$$\mathbb{P}(W = \eta^{-7/16}) = \mathbb{P}(W = -\eta^{-7/16}) = \mathbb{P}\left(W = \frac{\eta^{-7/16}}{2}\right) = \mathbb{P}\left(W = -\frac{\eta^{-7/16}}{2}\right) = \frac{1}{4} \quad (107)$$

and define a random variable H by

$$H = \begin{cases} \frac{8(\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)})}{9\eta a^2} & \text{if } |W| = \eta^{-7/16}, \\ -\frac{8(\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)})}{9\eta a^2} & \text{if } |W| = \frac{\eta^{-7/16}}{2}, \end{cases} \quad (108)$$

where $a := \eta^{-7/16}$. Then we have $\mathbb{E}[W] = \mathbb{E}[H] = 0$ and, since $\mathbb{E}[W^3] = \mathbb{E}[H^3] = \mathbb{E}[WH^2] = 0$,

$$\mathbb{E}[(W + H)^3] = 3\mathbb{E}[W^2H] = 3\left(\frac{1}{2}a^2 \cdot \frac{8(\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)})}{9\eta a^2} - \frac{1}{2}\left(\frac{a}{2}\right)^2 \cdot \frac{8(\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)})}{9\eta a^2}\right) = \frac{\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)}}{\eta}. \quad (109)$$

Define

$$c_0^2 := \frac{\bar{\sigma}_0^{(2)} - \eta\mathbb{E}[U^2] - \eta\mathbb{E}[W^2]}{1 - 2\eta}, \quad c_1^2 := \frac{\bar{\sigma}_1^{(2)} - \eta\mathbb{E}[U^2] - \eta\mathbb{E}[(W + H)^2]}{1 - 2\eta}. \quad (110)$$

For η small enough, $c_0^2 > 0$ and $c_1^2 > 0$. Assume (U, W, S) are independent and define $(\tilde{Y}_1, \tilde{Y}_0)$ by

$$(\tilde{Y}_1, \tilde{Y}_0) = \begin{cases} (U, U) & \text{with probability } \eta, \\ (W + H, W) & \text{with probability } \eta, \\ (c_1 S, c_0 S) & \text{with probability } 1 - 2\eta. \end{cases} \quad (111)$$

Then $\mathbb{E}[\tilde{Y}_x^2] = \bar{\sigma}_x^{(2)}$ for $x \in \{0, 1\}$ by the definition of c_0^2, c_1^2 . Moreover, $\mathbb{E}[\tilde{Y}_0^3] = \eta\mathbb{E}[U^3] = \bar{\sigma}_0^{(3)}$ and $\mathbb{E}[\tilde{Y}_1^3] = \eta\mathbb{E}[U^3] + \eta\mathbb{E}[(W + H)^3] = \bar{\sigma}_1^{(3)}$.

In addition, we have

$$\bar{\mu}^{(4)} = \mathbb{E}[\tilde{\Delta}^4] = (1 - 2\eta)(c_1 - c_0)^4 + \eta\mathbb{E}[H^4]. \quad (112)$$

If $\bar{\sigma}_0^{(3)} \neq 0$, then $\eta\mathbb{E}[U^2] = \eta^{1/3}$, while if $\bar{\sigma}_0^{(3)} = 0$, then $\eta\mathbb{E}[U^2] = 0$. Moreover, we have

$$\eta\mathbb{E}[W^2] = \frac{5}{8}\eta^{1/8}, \quad \eta\mathbb{E}[H^2] = \frac{64}{81}(\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)})^2\eta^{3/4}, \quad \eta\mathbb{E}[H^4] = \frac{4096}{6561}(\bar{\sigma}_1^{(3)} - \bar{\sigma}_0^{(3)})^4\eta^{1/2}. \quad (113)$$

Since $\eta\mathbb{E}[(W + H)^2] \leq 2\eta\mathbb{E}[W^2] + 2\eta\mathbb{E}[H^2]$, we have $c_0^2 \rightarrow \bar{\sigma}_0^{(2)}$, $c_1^2 \rightarrow \bar{\sigma}_1^{(2)}$, and $\bar{\mu}^{(4)} \rightarrow (\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}})^4$ as $\eta \rightarrow 0$.

Therefore, by choosing η small enough, we have $\bar{\mu}^{(4)} < (\sqrt{\bar{\sigma}_1^{(2)}} - \sqrt{\bar{\sigma}_0^{(2)}})^4 + \varepsilon$.

Sharpness (upper bound). Fix $M > 0$. Choose $b > 0$ such that $b > M/16$ and $b^{-1} < \min\{\bar{\sigma}_0^{(2)}, \bar{\sigma}_1^{(2)}\}$, and set $p := b^{-3}$. Let Z_x be a random variable with

$$\mathbb{E}[Z_x] = 0, \quad \mathbb{E}[Z_x^2] = \frac{\bar{\sigma}_x^{(2)} - pb^2}{1 - p}, \quad \mathbb{E}[Z_x^3] = \frac{\bar{\sigma}_x^{(3)}}{1 - p}, \quad (114)$$

for each $x \in \{0, 1\}$ (Lemma 5), and assume Z_1, Z_0 are independent. Define $(\tilde{Y}_1, \tilde{Y}_0)$ by

$$(\tilde{Y}_1, \tilde{Y}_0) = \begin{cases} (Z_1, Z_0) & \text{with probability } 1 - p, \\ (b, -b) & \text{with probability } p/2, \\ (-b, b) & \text{with probability } p/2. \end{cases} \quad (115)$$

Then $\mathbb{E}[\tilde{Y}_x^2] = \bar{\sigma}_x^{(2)}$ and $\mathbb{E}[\tilde{Y}_x^3] = \bar{\sigma}_x^{(3)}$ for $x \in \{0, 1\}$. Moreover, we obtain $\bar{\mu}^{(4)} \geq p(2b)^4 = 16b > M$.

(1)-(iii): Since $\bar{\sigma}_0^{(2)} = 0$, $\tilde{Y}_0 = 0$ almost surely and $\bar{\sigma}_0^{(3)} = 0$. Since $\bar{\sigma}_1^{(3)} = \mathbb{E}[\tilde{Y}_1(\tilde{Y}_1^2 - \bar{\sigma}_1^{(2)})]$, we have

$$(\bar{\sigma}_1^{(3)})^2 \leq \bar{\sigma}_1^{(2)}(\bar{\mu}^{(4)} - (\bar{\sigma}_1^{(2)})^2) \quad (116)$$

by the Cauchy–Schwarz inequality. Thus $\bar{\mu}^{(4)} \geq (\bar{\sigma}_1^{(2)})^2 + (\bar{\sigma}_1^{(3)})^2/\bar{\sigma}_1^{(2)}$.

Sharpness (lower bound). Let Z be a random variable with $\mathbb{E}[Z] = 0$, $\mathbb{E}[Z^2] = \bar{\sigma}_1^{(2)}$ and $\mathbb{E}[Z^3] = \bar{\sigma}_1^{(3)}$ (Lemma 5). Then $Z^2 = (\bar{\sigma}_1^{(3)}/\bar{\sigma}_1^{(2)})Z + \bar{\sigma}_1^{(2)}$ almost surely, and hence $\mathbb{E}[Z^4] = (\bar{\sigma}_1^{(2)})^2 + (\bar{\sigma}_1^{(3)})^2/\bar{\sigma}_1^{(2)}$. Taking $\tilde{Y}_1 = Z$ and $\tilde{Y}_0 = 0$ attains the lower bound.

Sharpness (upper bound). Fix $M > 0$. Choose $b > 0$ such that $b > M$ and $b^{-1} < \bar{\sigma}_1^{(2)}$, and set $p := b^{-3}$. Take a random variable Z with

$$\mathbb{E}[Z] = 0, \quad \mathbb{E}[Z^2] = \frac{\bar{\sigma}_1^{(2)} - pb^2}{1-p}, \quad \mathbb{E}[Z^3] = \frac{\bar{\sigma}_1^{(3)}}{1-p} \quad (117)$$

(Lemma 5). Define $\tilde{Y}_0 = 0$ almost surely and let

$$\tilde{Y}_1 = \begin{cases} Z & \text{with probability } 1-p, \\ b & \text{with probability } p/2, \\ -b & \text{with probability } p/2. \end{cases} \quad (118)$$

Then $\mathbb{E}[\tilde{Y}_1^2] = \bar{\sigma}_1^{(2)}$, $\mathbb{E}[\tilde{Y}_1^3] = \bar{\sigma}_1^{(3)}$, and $\bar{\mu}^{(4)} \geq pb^4 = b > M$.

(1)-(iv): The proof is identical after interchanging the roles of 1 and 0.

By Lemma 2, these are realized by some SCMs. Thus the bounds in (1) are sharp.

(2), (3), and (4) follow from Theorems 12–14. Lemma 2 yields SCMs. □

Theorem 16. Let $m \geq 2$ be an even integer. Assume $(\bar{\sigma}_1^{(m)}, \bar{\sigma}_0^{(m)})$ exist and are available. Then we have

$$\begin{aligned} \left(\sqrt[m]{\bar{\sigma}_1^{(m)}} - \sqrt[m]{\bar{\sigma}_0^{(m)}} \right)^m &\leq \bar{\mu}^{(m)} \\ &\leq \left(\sqrt[m]{\bar{\sigma}_1^{(m)}} + \sqrt[m]{\bar{\sigma}_0^{(m)}} \right)^m. \end{aligned} \quad (33)$$

If only $(\bar{\sigma}_1^{(m)}, \bar{\sigma}_0^{(m)})$ are available, then this bound is sharp.

Proof. Recall that $\tilde{\Delta} = \tilde{Y}_1 - \tilde{Y}_0$. Since m is even, we have

$$\|\tilde{Y}_x\|_m = (\mathbb{E}[|\tilde{Y}_x|^m])^{1/m} = \sqrt[m]{\bar{\sigma}_x^{(m)}} \quad (x \in \{0, 1\}), \quad \|\tilde{\Delta}\|_m^m = \mathbb{E}[|\tilde{\Delta}|^m] = \bar{\mu}^{(m)}. \quad (119)$$

By Minkowski's inequality in L^m -space,

$$\|\tilde{\Delta}\|_m \leq \|\tilde{Y}_1\|_m + \|\tilde{Y}_0\|_m, \quad \|\tilde{Y}_1\|_m \leq \|\tilde{\Delta}\|_m + \|\tilde{Y}_0\|_m, \quad \text{and} \quad \|\tilde{Y}_0\|_m \leq \|\tilde{Y}_1\|_m + \|\tilde{\Delta}\|_m \quad (120)$$

hold. Thus we have $|\|\tilde{Y}_1\|_m - \|\tilde{Y}_0\|_m| \leq \|\tilde{\Delta}\|_m \leq \|\tilde{Y}_1\|_m + \|\tilde{Y}_0\|_m$. Raising both sides to the m -th power yields (33).

Sharpness. Take a random variable $S \in \{-1, 1\}$ with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$. Setting $\tilde{Y}_1 := \sqrt[m]{\bar{\sigma}_1^{(m)}}S$ and $\tilde{Y}_0 := -\sqrt[m]{\bar{\sigma}_0^{(m)}}S$ attains the upper bound. Setting $\tilde{Y}_1 := \sqrt[m]{\bar{\sigma}_1^{(m)}}S$ and $\tilde{Y}_0 := \sqrt[m]{\bar{\sigma}_0^{(m)}}S$ attains the lower bound. By Lemma 2, these can be realized by an SCM. □

Theorem 17. Let $m \geq 2$ be an even integer. Assume that $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ exist and are available for a single $k < m$.

(1). If k is odd, we have $0 \leq \bar{\mu}^{(m)} \leq \infty$.

(2). If k is even, we have the following:

(i). If $\bar{\sigma}_1^{(k)} = \bar{\sigma}_0^{(k)} = 0$, we have $\bar{\mu}^{(m)} = 0$.

(ii). If $\bar{\sigma}_1^{(k)} > 0$ or $\bar{\sigma}_0^{(k)} > 0$, we have

$$\left(\sqrt[k]{\bar{\sigma}_1^{(k)}} - \sqrt[k]{\bar{\sigma}_0^{(k)}} \right)^m \leq \bar{\mu}^{(m)} \leq \infty. \quad (34)$$

If only $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ are available, then these bounds are sharp.

Proof. (1): The inequality $0 \leq \bar{\mu}^{(m)} \leq \infty$ follows from the definition.

Sharpness (lower bound). Fix $\varepsilon > 0$. Choose a random variable B with $\mathbb{E}[B] = 0$ and $\mathbb{E}[B^k] = 2\bar{\sigma}_0^{(k)}$ (Lemma 7). For $a, \delta \in \mathbb{R}$, define a random variable W_a by

$$W_a := \begin{cases} B & \text{with probability } 1/2, \\ a & \text{with probability } 1/4, \\ -a & \text{with probability } 1/4 \end{cases} \quad (121)$$

and define $V_{a,\delta}$ by

$$V_{a,\delta} := \begin{cases} W_a - \delta & \text{if } W_a = B, \\ W_a + \delta & \text{if } W_a \in \{a, -a\}. \end{cases} \quad (122)$$

Then $\mathbb{E}[W_a] = \mathbb{E}[V_{a,\delta}] = 0$ and, since k is odd, $\mathbb{E}[W_a^k] = \bar{\sigma}_0^{(k)}$. Moreover $V_{a,\delta} - W_a = \pm\delta$, hence $\mathbb{E}[(V_{a,\delta} - W_a)^m] = |\delta|^m$. Let $F_a(\delta) := \mathbb{E}[V_{a,\delta}^k]$. Then

$$F_a(\delta) = \frac{1}{2}\mathbb{E}[(B - \delta)^k] + \frac{1}{4}(a + \delta)^k + \frac{1}{4}(-a + \delta)^k = \frac{1}{2}\mathbb{E}[(B - \delta)^k] + \frac{1}{4}(a + \delta)^k - \frac{1}{4}(a - \delta)^k, \quad (123)$$

so $F_a(\delta) \rightarrow \infty$ and $F_a(-\delta) \rightarrow -\infty$ as $a \rightarrow \infty$.

Hence, for some $a_0 > 0$ we have $F_{a_0}(-\sqrt[m]{\varepsilon}) \leq \bar{\sigma}_1^{(k)} \leq F_{a_0}(\sqrt[m]{\varepsilon})$. Since F_{a_0} is continuous, there exists $\delta_0 \in [-\sqrt[m]{\varepsilon}, \sqrt[m]{\varepsilon}]$ such that $F_{a_0}(\delta_0) = \bar{\sigma}_1^{(k)}$. With $(\tilde{Y}_1, \tilde{Y}_0) := (V_{a_0,\delta_0}, W_{a_0})$, we have $\mathbb{E}[\tilde{Y}_x^k] = \bar{\sigma}_x^{(k)}$ and $\bar{\mu}^{(m)} = |\delta_0|^m \leq \varepsilon$.

Sharpness (upper bound). Fix $M > 0$. Let B_0, B_1 be random variables with $\mathbb{E}[B_x] = 0$, $\mathbb{E}[B_x^k] = 2\bar{\sigma}_x^{(k)}$ ($x \in \{0, 1\}$) (Lemma 7). Take a constant $A > 0$ with $\frac{1}{2}(2A)^m > M$, and define $(\tilde{Y}_1, \tilde{Y}_0)$ by

$$(\tilde{Y}_1, \tilde{Y}_0) := \begin{cases} (B_1, B_0) & \text{with probability } 1/2, \\ (A, -A) & \text{with probability } 1/4, \\ (-A, A) & \text{with probability } 1/4. \end{cases} \quad (124)$$

Then $\mathbb{E}[\tilde{Y}_1^k] = \bar{\sigma}_1^{(k)}$ and $\mathbb{E}[\tilde{Y}_0^k] = \bar{\sigma}_0^{(k)}$. We have $\tilde{\Delta} = \tilde{Y}_1 - \tilde{Y}_0 = \pm 2A$ when $(\tilde{Y}_1, \tilde{Y}_0) = (\pm A, \mp A)$, and we have $\bar{\mu}^{(m)} = \mathbb{E}[\tilde{\Delta}^m] \geq \frac{1}{2}(2A)^m > M$. Taking A large yields $\bar{\mu}^{(m)} > M$.

By Lemma 2, these distributions can be realized by SCMs.

(2)-(i): If $\bar{\sigma}_1^{(k)} = \bar{\sigma}_0^{(k)} = 0$, we have $\tilde{Y}_1 = 0$ almost surely and $\tilde{Y}_0 = 0$ almost surely. Hence $\bar{\mu}^{(m)} = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^m] = 0$.

(2)-(ii): Since $m > k$, $\|\tilde{\Delta}\|_m \geq \|\tilde{\Delta}\|_k$. By the triangle inequality in L^k -space, $\|\tilde{\Delta}\|_k \geq |\|\tilde{Y}_1\|_k - \|\tilde{Y}_0\|_k| = |\sqrt[k]{\bar{\sigma}_1^{(k)}} - \sqrt[k]{\bar{\sigma}_0^{(k)}}|$. Raising to the m -th power yields the lower bound.

Sharpness (lower bound). Take a random variable S with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$ and set $\tilde{Y}_x := \sqrt[k]{\bar{\sigma}_x^{(k)}} S$. Then $\tilde{\Delta} = (\sqrt[k]{\bar{\sigma}_1^{(k)}} - \sqrt[k]{\bar{\sigma}_0^{(k)}})S$, so the lower bound is attained.

Sharpness (upper bound). If $\bar{\sigma}_1^{(k)} = \bar{\sigma}_0^{(k)} = 0$, then $\bar{\mu}^{(m)} = 0$. Assume $\bar{\sigma}_1^{(k)} + \bar{\sigma}_0^{(k)} > 0$, and fix $M > 0$. We may assume $\bar{\sigma}_1^{(k)} > 0$, by exchanging the roles of 0 and 1 if necessary. Choose a constant $A > \sqrt[k]{\bar{\sigma}_1^{(k)}}$ such that $(\bar{\sigma}_1^{(k)}/A^k)(A - \sqrt[k]{\bar{\sigma}_0^{(k)}})^m > M$.

Let I be a random variable with $\mathbb{P}(I = 1) = \bar{\sigma}_1^{(k)}/A^k$, and S be a random variable with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$, independent of I . Define $\tilde{Y}_1 := IAS$ and $\tilde{Y}_0 := \sqrt[k]{\bar{\sigma}_0^{(k)}}S$. Then $\mathbb{E}[\tilde{Y}_x^k] = \bar{\sigma}_x^{(k)}$ ($x \in \{0, 1\}$) and $\bar{\mu}^{(m)} = \mathbb{E}[\tilde{\Delta}^m] \geq \mathbb{P}(I = 1)(A - \sqrt[k]{\bar{\sigma}_0^{(k)}})^m > M$.

By Lemma 2, these can be realized by SCMs. Therefore we have the sharpness. $\square$

Theorem 18. Let $m \geq 3$ be an odd number. Assume that $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ exist and are available for a single $k \leq m$.

(1). If k is even and $\bar{\sigma}_1^{(k)} = \bar{\sigma}_0^{(k)} = 0$, we have $\bar{\mu}^{(m)} = 0$.

(2). Otherwise, we have $-\infty \leq \bar{\mu}^{(m)} \leq \infty$. If only $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$ are available, then this bound is sharp.

Proof. (1): If $\bar{\sigma}_1^{(k)} = \bar{\sigma}_0^{(k)} = 0$ for an even k , we have $\tilde{Y}_1 = 0$ almost surely and $\tilde{Y}_0 = 0$ almost surely. Hence $\bar{\mu}^{(m)} = \mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_0)^m] = 0$.

(2): Fix $M > 0$.

The case k is odd. For each $x \in \{0, 1\}$, take a random variable B_x with $\mathbb{E}[B_x] = 0$ and $\mathbb{E}[B_x^k] = 2\bar{\sigma}_x^{(k)}$ (Lemma 7). For $A > 0$, define (U_A, V_A) by

$$(U_A, V_A) = \begin{cases} (A, -A) & \text{with probability } 1/3, \\ (-A, 0) & \text{with probability } 1/3, \\ (0, A) & \text{with probability } 1/3. \end{cases} \quad (125)$$

Then $\mathbb{E}[U_A^k] = \mathbb{E}[V_A^k] = 0$ and $\mathbb{E}[(U_A - V_A)^m] = \frac{2^m - 2}{3}A^m$. Define $(\tilde{Y}_1, \tilde{Y}_0)$ by

$$(\tilde{Y}_1, \tilde{Y}_0) = \begin{cases} (B_1, B_0) & \text{with probability } 1/2, \\ (U_A, V_A) & \text{with probability } 1/2. \end{cases} \quad (126)$$

Then $\mathbb{E}[\tilde{Y}_x^k] = \bar{\sigma}_x^{(k)}$ and

$$\bar{\mu}^{(m)} = \mathbb{E}[\tilde{\Delta}^m] = \frac{1}{2}\mathbb{E}[(B_1 - B_0)^m] + \frac{2^m - 2}{6}A^m. \quad (127)$$

Taking A large yields $\bar{\mu}^{(m)} > M$. Replacing (U_A, V_A) by (V_A, U_A) yields $\bar{\mu}^{(m)} < -M$.

The case k is even. We assume $\bar{\sigma}_1^{(k)} > 0$, by exchanging the roles of 0 and 1 if necessary. Take a random variable S with $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$ and put $W_0 := \sqrt[k]{2\bar{\sigma}_0^{(k)}}S$. Choose a constant A such that $p_A := \frac{2\bar{\sigma}_1^{(k)}}{A^k(1/2 + 1/2^k)} \leq \frac{2}{3}$. Let Z_A be a random variable with

$$\mathbb{P}(Z_A = A) = \frac{p_A}{2}, \quad \mathbb{P}\left(Z_A = -\frac{A}{2}\right) = p_A \quad \text{and} \quad \mathbb{P}(Z_A = 0) = 1 - \frac{3p_A}{2}. \quad (128)$$

Then $\mathbb{E}[Z_A] = 0$, $\mathbb{E}[Z_A^k] = 2\bar{\sigma}_1^{(k)}$, and $\mathbb{E}[Z_A^m] = 2\bar{\sigma}_1^{(k)} \frac{1/2 - 1/2^m}{1/2 + 1/2^k} A^{m-k}$. Define $(\tilde{Y}_1, \tilde{Y}_0)$ by

$$(\tilde{Y}_1, \tilde{Y}_0) = \begin{cases} (0, W_0) & \text{with probability } 1/2, \\ (Z_A, 0) & \text{with probability } 1/2. \end{cases} \quad (129)$$

Then $\mathbb{E}[\tilde{Y}_x^k] = \bar{\sigma}_x^{(k)}$ and $\bar{\mu}^{(m)} = \mathbb{E}[Z_A^m]/2$. Taking A large yields $\bar{\mu}^{(m)} > M$. Replacing Z_A by $-Z_A$ yields $\bar{\mu}^{(m)} < -M$. By Lemma 2, these can be realized by SCMs. $\square$

D NUMERICAL EXPERIMENTS

In this appendix, we conduct numerical experiments to illustrate the properties of our results in a finite-sample setting.

Table 5: Estimates of bounds of central moments of ICE under SCM (A). MCI denotes the estimates obtained by [Kawakami and Tian, 2025]. LB denotes the lower bound, and UB denotes the upper bound.

Target m	Theorem	Available $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$	$N = 20$		$N = 1000$		True
			LB	UB	LB	UB	
2	3	2	0.092 [0.000,0.456]	1.128 [0.317,2.088]	0.005 [0.001,0.013]	1.492 [1.386,1.600]	0.083
	5	3	0.000 [0.000,0.000]	∞	0.000 [0.000,0.000]	∞	0.083
	6	4	0.000 [0.000,0.000]	1.603 [0.523,2.737]	0.000 [0.000,0.000]	2.157 [2.051,2.270]	0.083
	MCI	-	0.004 [0.000,0.044]	1.020 [0.296,1.904]	0.000 [0.000,0.000]	1.410 [1.243,1.576]	0.083
3	8	2	$-\infty$	∞	$-\infty$	∞	0.000
	9	3	$-\infty$	∞	$-\infty$	∞	0.000
	10	4	-1.328 [-2.810,-0.235]	1.328 [0.235,2.810]	-1.966 [-2.122,-1.822]	1.966 [1.822,2.122]	0.000
	MCI	-	-0.981 [-2.157,-0.201]	0.450 [0.021,1.338]	-1.613 [-1.840,-1.395]	0.629 [0.475,0.814]	0.000
4	12	2	0.023 [0.000,0.208]	∞	0.000 [0.000,0.000]	∞	0.013
	13	3	0.000 [0.000,0.000]	∞	0.000 [0.000,0.000]	∞	0.013
	14	4	0.058 [0.000,0.455]	2.937 [0.274,7.494]	0.000 [0.000,0.001]	4.657 [4.207,5.152]	0.013
	MCI	-	0.000 [0.000,0.000]	2.240 [0.233,5.663]	0.000 [0.000,0.000]	3.980 [3.355,4.728]	0.013

Table 6: Estimates of bounds of central moments of ICE under SCM (B). MCI denotes the estimates obtained by [Kawakami and Tian, 2025]. LB denotes the lower bound, and UB denotes the upper bound.

Target m	Theorem	Available $(\bar{\sigma}_1^{(k)}, \bar{\sigma}_0^{(k)})$	$N = 20$		$N = 1000$		True
			LB	UB	LB	UB	
2	3	2	0.858 [0.604,0.978]	1.021 [0.746,1.142]	0.887 [0.879,0.895]	1.117 [1.107,1.125]	0.903
	5	3	0.000 [0.000,0.000]	∞	0.000 [0.000,0.000]	∞	0.903
	6	4	0.000 [0.000,0.000]	1.144 [1.024,1.286]	0.000 [0.000,0.000]	1.139 [1.131,1.148]	0.903
	MCI	-	0.086 [0.000,0.147]	0.843 [0.412,1.117]	0.021 [0.001,0.055]	1.071 [0.990,1.140]	0.903
3	8	2	$-\infty$	∞	$-\infty$	∞	0.000
	9	3	$-\infty$	∞	$-\infty$	∞	0.000
	10	4	-0.760 [-0.904,-0.643]	0.760 [0.643,0.904]	-0.754 [-0.763,-0.746]	0.754 [0.746,0.763]	0.000
	MCI	-	-0.403 [-0.606,-0.101]	0.405 [0.123,0.601]	-0.555 [-0.650,-0.458]	0.556 [0.465,0.646]	0.000
4	12	2	0.745 [0.365,0.956]	∞	0.787 [0.772,0.801]	∞	0.819
	13	3	0.000 [0.000,0.000]	∞	0.000 [0.000,0.000]	∞	0.819
	14	4	0.904 [0.730,1.153]	1.313 [1.048,1.653]	0.761 [0.750,0.774]	1.298 [1.280,1.318]	0.819
	MCI	-	0.000 [0.000,0.001]	0.792 [0.286,1.248]	0.000 [0.000,0.000]	1.158 [0.960,1.351]	0.819

Table 7: Estimates of central moments of ICE under IED condition for SCM (C) based on Theorem 1.

Target	$N = 20$	$N = 1000$	True
2	0.412 [0.000,0.903]	0.335 [0.266,0.406]	0.333
3	-0.010 [-0.500,0.455]	0.000 [-0.072,0.072]	0.000
4	0.461 [0.000,1.596]	0.205 [0.141,0.278]	0.200

Baseline. We compare our bounds with those given by the Monte Carlo integration-based estimator (MCI) [Kawakami and Tian, 2025]. Note that only the upper bounds of MCI are sharp when m is even. [Fan and Park, 2010, Post, 2023, Post and Van Den Heuvel, 2025] do not provide results for moments of causal effects; thus, there is no baseline other than [Kawakami and Tian, 2025]. Our estimator uses only the marginal central moments of each PO, whereas MCI requires the full marginal distribution of each PO.

Setting. We work with the following SCMs:

$$\text{SCM(A)} : \begin{cases} Y := -(X + 1)U \mathbb{1}(XU \geq 0), \\ X \sim \text{Bern}(0.8), U \sim \text{Unif}(-1, 1) \end{cases} \quad (130)$$

$$\text{SCM(B)} : \begin{cases} Y := X \text{sign}(U) + 0.1(1 - X)U, \\ X \sim \text{Bern}(0.8), U \sim \text{Unif}(-1, 1) \end{cases} \quad (131)$$

$$\text{SCM(C)} : \begin{cases} Y = U_1 + XU_2, \\ U_1 \sim \text{Unif}(-1, 1), U_2 \sim \text{Unif}(-1, 1), \\ X \sim \text{Bern}(0.8), U_1 \perp\!\!\!\perp U_2 \end{cases} \quad (132)$$

where $\text{Bern}(p)$ is a Bernoulli distribution with probability p , and $\text{Unif}(-1, 1)$ is a uniform distribution over $[-1, 1]$. SCM (A) is the setting used in the experiments of Kawakami and Tian [2025]. SCM(A) and (B) do not satisfy the IED condition, and the SCM(C) satisfies the IED condition. We estimate the bounds by plugging in the empirical central moments of each potential outcome. For even-order moments, if the plug-in estimate is negative, we set it to zero. We conduct 1,000 simulation runs for sample sizes $N = 20, 1000$. We report the means of the estimates and their 95% confidence intervals (CIs).

Results. From Table 5, in setting (A), when the sample size is relatively large ($N = 1000$), the tightest bounds for the second and fourth central moments of ICE obtained by our method—using $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$, respectively—are similar to those of MCI, with comparable CIs. In contrast, the tightest bound for the third central moment of ICE using $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$ is slightly wider than those of MCI. When the sample size is relatively small ($N = 20$), the estimated bounds from our method have relatively wide CIs compared to those of MCI.

In setting (B), from Table 6, the tightest bounds for the second and fourth central moments of ICE obtained by our method—using $(\bar{\sigma}_1^{(2)}, \bar{\sigma}_0^{(2)})$ and $(\bar{\sigma}_1^{(4)}, \bar{\sigma}_0^{(4)})$, respectively—are narrower than those of MCI.

In setting (C), as shown in Table 7, when the sample size is relatively large ($N = 1000$), we obtain reliable estimates of the second, third, and fourth central moments of ICE, as indicated by the narrow CIs. In contrast, when the sample size is relatively small ($N = 20$), the estimates of the second, third, and fourth central moments of ICE are unstable, with wide CIs.

Remark (Confidence intervals for plug-in estimates.). *The empirical case studies report plug-in estimates because the original articles provide only point summaries. This appendix reports confidence intervals in simulated finite-sample settings. If confidence intervals for marginal moments are available in empirical applications, they can be propagated through the formulas in this paper. For example, under IED, $\bar{\mu}^{(2)} = \bar{\sigma}_1^{(2)} - \bar{\sigma}_0^{(2)}$. If $L^1 \leq \bar{\sigma}_1^{(2)} \leq U^1$ and $L^0 \leq \bar{\sigma}_0^{(2)} \leq U^0$ hold simultaneously, then*

$$L^1 - U^0 \leq \bar{\mu}^{(2)} \leq U^1 - L^0. \quad (133)$$

If the two marginal intervals each have coverage at least α , then the simultaneous coverage is at least $2\alpha - 1$, since $\mathbb{P}(A \cap B) \geq 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c)$. The same argument applies to the bounds derived in this paper.