Causal Effect Identification with a Single Agnostic Proxy

Xiu-Chuan Li
Sydney AI Centre
The University of Sydney
lixiuchuan2020@gmail.com

Tongliang Liu*
Sydney AI Centre
The University of Sydney
tongliang.liu@sydney.edu.au

Abstract

Latent confounding poses a significant obstacle to identifying the causal effect of a treatment on an outcome. To address this, many existing studies leverage proxies of the latent confounder to indirectly adjust for the confounding bias. However, they impose specific structural constraints on the proxies and typically require multiple proxies. In this paper, focusing on the latent variable linear non-Gaussian acyclic model (lvLiNGAM), we propose a causal effect identification algorithm requiring only a single agnostic proxy. Crucially, the term "agnostic" means that the causal connections between the proxy and the treatment-outcome pair can be arbitrary and are not required to be known a priori. This structural complexity precludes identifying the causal effect via a simple closed-form formula. Consequently, our identification algorithm is designed to first derive candidate solutions from cumulants and then isolate the valid solution by examining certain independence relationships. We present a series of new theoretical results, which collectively establish the soundness of our identification algorithm: given the observational population distribution, it correctly identifies the true causal effect when identifiable, and correctly reports unidentifiability otherwise. Finally, we empirically validate our theoretical results. Our code is available at https://github.com/XiuchuanLi/NeurIPS2026-agnostic-proxy.

1 Introduction

Ascertaining the causal effect of a treatment on an outcome is a crucial challenge in various fields such as social sciences [1, 2], public health [3, 4], and agriculture [5]. The gold standard is conducting randomized experiments, but this is usually prohibitively expensive and even infeasible due to ethical or legal concerns. Consequently, researchers increasingly study causal effect identification, that is, whether and how the causal effect can be uniquely determined from the observational (non-experimental) population distribution. The backdoor formula [6] is a widely-used identification strategy, but it is unsuitable for scenarios involving latent confounding. To tackle latent confounding, many studies utilize proxies of the latent confounder that are often its observed children other than the treatment and the outcome. Proxies serve as imperfect measurements of the latent confounder, enabling us to adjust for the confounding bias indirectly.

Most early studies on proximal causal effect identification focus on non-parametric structural causal models (SCMs) [7, 8, 9, 10, 11]. Generally, they require multiple proxies subject to specific structural constraints, a requirement that is often practically infeasible, unless certain more stringent conditions are satisfied. For instance, when the latent confounder L has only a single proxy Z, Kuroki et al. [7] require that $P(Z|L)$ is known a priori. Even with this prior knowledge, they still impose a structural constraint that Z shares no causal edge with either the treatment T or the outcome O. Hereafter,

*Corresponding author

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

we refer to such a proxy as an isolated proxy. Recently, some research investigates parametric SCMs, particularly linear SCMs [12, 13, 14, 15, 16]. Although a single proxy still cannot guarantee identifiability in all linear SCMs [12], this is achievable in latent variable linear non-Gaussian acyclic models (lvLiNGAMs) where all exogenous noises are non-Gaussian [17]. Within lvLiNGAM, Kivva et al. [12] propose an identification algorithm that can work with a single isolated proxy, while the algorithm of Tramontano et al. [16] slightly relaxes the requirement for an isolated proxy by allowing a causal edge from Z to T. Moreover, they both assume that the underlying causal structure is known a priori. In essence, they still impose restrictive structural constraints on the proxies.

In this paper, we reduce the number of required proxies and eliminate the structural constraints imposed on proxies simultaneously. Focusing on lvLiNGAM, we propose a novel causal effect identification algorithm requiring only a single agnostic proxy. Crucially, the term "agnostic" encompasses two distinct structural properties: flexibility and uncertainty. Flexibility means that the causal connections between the proxy and the treatment-outcome pair can be arbitrary, while uncertainty means that the causal connections are not required to be known a priori. A more concrete characterization of the agnostic proxy is provided at the beginning of Sec. 3. Agnostic proxies are common in the real world. Consider the causal structure shown in Fig. 1, if it is unknown whether academic club scheduling conflicts with classroom sessions, we cannot determine whether classroom attendance rate impacts academic club attendance rate; if it is also unknown whether club curriculum and exam scope overlaps, we cannot determine whether academic club attendance rate affects exam score. Therefore, academic club attendance rate is an agnostic proxy.

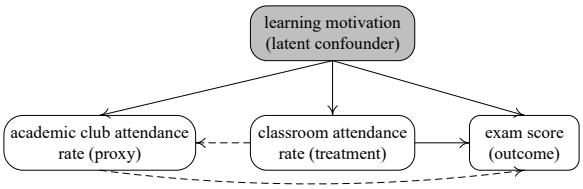


Figure 1: Academic club attendance rate is an agnostic proxy of learning motivation. The dashed lines represent potential causal edges whose existence is uncertain.

The structural complexity of the agnostic proxy precludes identifying the causal effect via a simple closed-form formula. Therefore, our identification algorithm is designed to first generate multiple candidate solutions by solving quadratic equations derived from cumulants and then isolate the valid solution from these candidates by examining certain independence relationships. We present a series of theoretical results, providing a formal guarantee for the soundness of our identification algorithm: given the observational population distribution, it correctly identifies the true causal effect when identifiable, and correctly reports unidentifiability otherwise. Also, we develop a plug-in estimation method based on our identification algorithm and show that its error converges to zero as the sample size increases, empirically validating our theoretical results.

2 Preliminary

§ lvLiNGAM

Throughout this paper, we focus on causal effect identification in latent variable linear non-Gaussian acyclic model (lvLiNGAM). Within a lvLiNGAM whose causal structure is a directed acyclic graph (DAG) $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, each $V_i \in \mathbf{V}$ follows

$$V_i = \sum_{V_j \in \mathbf{V}} a_{V_i V_j} V_j + \epsilon_{V_i}, \tag{1}$$

where $a_{V_i V_j}$ refers to the direct causal strength from V_j to V_i and ϵ_{V_i} refers to the exogenous noise of V_i. $a_{V_i V_j} \neq 0$ iff $V_j \in \mathrm{Pa}(V_i)$ where $\mathrm{Pa}(V_i)$ denotes the parents of V_i in $\mathcal{G}$. All exogenous noises have non-Gaussian distributions and are mutually independent. Without loss of generality, we further assume they all have a mean of zero.

Assumption 1. For any polynomial in $\{a_{V_i V_j} | V_j \in \mathrm{Pa}(V_i)\}$ with rational coefficients, it equals 0 only if it is identically 0.

Remark 1. This assumption holds generically. Specifically, if a polynomial is not identically zero, its zeros set has Lebesgue measure zero in the parameter space. Moreover, there are only countably many polynomials with rational coefficients, so the set of coefficients violating this assumption also has Lebesgue measure zero.

Based on Eq. (1), each $V_i \in \mathbf{V}$ can also be written as a linear combination of exogenous noises:

$$V_i = \sum_{V_j \in \mathbf{V}} m_{V_i V_j} \epsilon_{V_j}, \tag{2}$$

where $M = (I - A)^{-1} = I + \sum_{i=1}^{l} A^i$ with M being the matrix whose (i, j)-entry is $m_{V_i V_j}$, A being the matrix whose (i, j)-entry is $a_{V_i V_j}$, and l being the length of the longest directed path in $\mathcal{G}$. Under Assumption 1, $m_{V_i V_j} = 0$ iff $V_j \notin \mathrm{An}(V_i)$, where $\mathrm{An}(V_i)$ denotes the ancestor of V_i (including V_i itself) in $\mathcal{G}$. The causal effect of V_j on V_i is defined as $\mathbb{E}[V_i|do(V_j = 1)] - \mathbb{E}[V_i|do(V_j = 0)]$, which is exactly $m_{V_i V_j}$.

Remark 2. Our focus on lvLiNGAM is motivated by two primary considerations. First, as discussed in Sec. 5, it is generally impossible to identify the causal effect with a single proxy outside lvLiNGAM framework, let alone with a single agnostic proxy whose causal connections with the treatment-outcome pair are both flexible and uncertain, which further elevates the identification challenge. Second, real-world systems across diverse domains [18, 19, 20, 21, 22] can be approximated by LiNGAM, a simplified version of lvLiNGAM without latent confounders, so lvLiNGAM enjoys widespread popularity in causal inference literature [23, 24, 25, 12, 26, 27, 28, 16, 29].

§ Cumulant

Our identification algorithm first derives candidate solutions from cumulants, which is a basic concept in statistics. We present the definition of cumulant along with some properties below.

Definition 1. (Cumulant) Given k random variables $X_1, ..., X_k$, the cumulant $\mathrm{cum}(X_1, ..., X_k)$ is defined as

$$\mathrm{Cum}(X_1, ..., X_k) = \sum_{(A_1, ..., A_l)} (-1)^{l-1}(l-1)! \prod_{i=1}^{l} \mathbb{E}\left[\prod_{j \in A_i} X_j\right],$$

where the sum is taken over all partitions $(A_1, ..., A_l)$ of $\{1, ..., k\}$.

Cumulants are known to manifest the following properties [30]:

1. (Exchangeability) $\mathrm{Cum}(V_1, ..., V_i, ..., V_j, ..., V_n) = \mathrm{Cum}(V_1, ..., V_j, ..., V_i, ..., V_n)$.
2. (Linearity) If k is a constant, then $\mathrm{Cum}(kV_1, ..., V_n) = k\mathrm{Cum}(V_1, ..., V_n)$; if V_1' is a random variable, then $\mathrm{Cum}(V_1 + V_1', ..., V_n) = \mathrm{Cum}(V_1, ..., V_n) + \mathrm{Cum}(V_1', ..., V_n)$.
3. (Independence) If $(U_1, ..., U_n) \perp\!\!\!\perp (V_1, ..., V_n)$, then $\mathrm{Cum}(U_1+V_1, ..., U_n+V_n) = \mathrm{Cum}(U_1, ..., U_n)+\mathrm{Cum}(V_1, ..., V_n)$.

For ease of exposition, we denote $\mathrm{Cum}(\underbrace{X, \ldots}_{i \text{ times}})$ by $C_i(X)$ and $\mathrm{Cum}(\underbrace{X, \ldots}_{i \text{ times}}, \underbrace{Y, \ldots}_{j \text{ times}})$ by $C_{i,j}(X, Y)$ in the following. Using this notation, the sum of the subscripts corresponds to the order of the cumulant.

Assumption 2. For each variable V, $|C_4(\epsilon_V)| < \infty$ and $C_4(\epsilon_V) \neq 0$.

Remark 3. First, $|C_4(\epsilon_V)| < \infty$ is a technical assumption that primarily serves to exclude those pathological distributions characterized by extremely heavy tails (e.g., Cauchy distribution). Second, note that ϵ_V is non-Gaussian, $C_4(\epsilon_V) \neq 0$ simply specifies this non-Gaussianity in terms of a non-vanishing fourth-order cumulant. Intuitively, this implies that the distribution has tail behavior that is not identical to a Gaussian distribution. Finally, Asmp. 2 logically implies that $|C_2(\epsilon_V)| < \infty$ and $C_2(\epsilon_V) \neq 0$. Notably, Tramontano et al. [16] implicitly assume that for any $k \geq 2$, $|C_k(\epsilon_V)| < \infty$ and $C_k(\epsilon_V) \neq 0$, which is strictly stronger than our assumption.

Assumption 3. For every $V_i \in \mathbf{V}$ and $\mathbf{V}_i \subseteq \mathrm{An}(V_i)$ such that $V_i \in \mathbf{V}_i$, $\sum_{V \in \mathbf{V}_i} m_{V_i V}^4 C_4(\epsilon_V) \neq 0$.

Remark 4. $\sum_{V \in \mathbf{V}_i} m_{V_i V}^4 C_4(\epsilon_V)$ can be regarded as a polynomial in $\{a_{V_i V_j} | V_j \in \mathrm{Pa}(V_i)\}$ with $C_4(\epsilon_V)$ entering the coefficients. Trivially, for every $V_i \in \mathbf{V}$ and $\mathbf{V}_i \subseteq \mathrm{An}(V_i)$, this polynomial is not identically zero. Following Remark 1, Assumption 3 also holds generically.

§ Independence

After deriving candidate solutions from cumulants, our identification algorithm isolates the valid solution from candidates by examining certain independence relationships. Darmois-Skitovitch Theorem [31], which establishes the connection between independence and non-Gaussianity, constitutes a theoretical cornerstone.

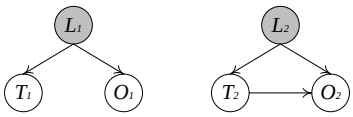


Figure 2: Two causal structures showing different behaviors on cumulants.

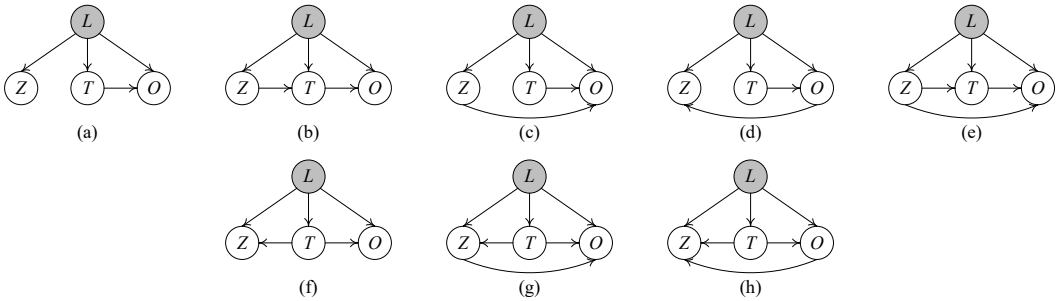


Figure 3: Causal structures considered in Sec. 3. Notably, they encompass all possible patterns of causal connections between Z and (T, O).

Darmois-Skitovitch Theorem. *Suppose two random variables V_1 and V_2 are both linear combinations of independent non-constant random variables $\{e_i\}_i$:*

$$V_1 = \sum_i \alpha_i e_i, \quad V_2 = \sum_i \beta_i e_i.$$

If $V_1 \perp\!\!\!\perp V_2$, then for every non-Gaussian e_j, $\alpha_j \beta_j = 0$, that is, at least one of V_1 and V_2 does not contain e_j.

3 Main Results

We investigate such a problem: There are a treatment variable T and an outcome variable O where T is a parent of O. These two variables have a common parent L which is a latent variable, and L has another observed child Z serving as its agnostic proxy. Specifically, the causal connections between Z and (T, O) can be arbitrary and are not known a priori, which means that the underlying causal structure can be any one in Fig. 3 and it is unknown which one corresponds to the ground truth. Our goal is to identify the causal effect of T on O (i.e. m_{OT}) from the joint distribution of (T, O, Z).

3.1 High-level Intuition

In this section, we intuitively illustrate how cumulants can be used to generate candidate solutions and explain why independence relationships can be used to isolate the valid solution. Specifically, consider the two causal structures shown as Fig. 2. First, according to Lems. 1 in App. A,

$$C_{3,1}(T_1, O_1)C_{1,3}(T_1, O_1) = C_{2,2}^2(T_1, O_1).$$

Second, if removing the causal edge $T_2 \to O_2$, that is, subtracting $m_{O_2T_2}T_2$ from O_2, the two causal structures shown as Fig. 2 are equivalent to each other. That is, one root of the equation in λ:

$$C_{3,1}(T_2, O_2 - \lambda T_2)C_{1,3}(T_2, O_2 - \lambda T_2) = C_{2,2}^2(T_2, O_2 - \lambda T_2), \tag{3}$$

is exactly $m_{O_2T_2}$. In fact, Eq. (3) can be simplified as a quadratic equation (see Eq. (7) in App. A), two roots of which are $m_{O_2L_2}/m_{T_2L_2}$ and $m_{O_2T_2}$ (see Lem. 2 in App. A), so cumulants can be used to generate candidate solutions.

Moreover, $O_2 - \lambda T_2$ contains different exogenous noises depending on the value of λ: it contains ϵ_{T_2} but not ϵ_{L_2} when $\lambda = m_{O_2L_2}/m_{T_2L_2}$, whereas it contains ϵ_{L_2} but not ϵ_{T_2} when $\lambda = m_{O_2T_2}$. According to the Darmois-Skitovitch Theorem, these different underlying noise structures imply that the independence relationship between $O_2 - \lambda T_2$ and other specific variables might vary according to the value of λ, so we can distinguish between $m_{O_2T_2}$ and $m_{O_2L_2}/m_{T_2L_2}$ by examining certain independence relationships.

For ease of exposition, we define cumulant constraint and independence constraint as follows.

Definition 2. (Cumulant constraint) Given two random variables V_1, V_2, we say (V_1, V_2) satisfies the cumulant constraint if $C_{3,1}(V_1, V_2)C_{1,3}(V_1, V_2) = C_{2,2}^2(V_1, V_2)$.

Definition 3. (Independence constraint) Given three random variables V_1, V_2, V_3, we say (V_1, V_2, V_3) satisfies the independence constraint if $\mathrm{Cov}(V_1, V_3)\mathrm{Cov}(V_2, V_3) \neq 0$ and $\mathrm{Cov}(V_2, V_3)V_1 - \mathrm{Cov}(V_1, V_3)V_2 \perp\!\!\!\perp V_3$.

Remark 5. While we synergistically combines the cumulant constraint and the independence constraint within a coherent paradigm for causal effect identification, previous studies use them separately for different purposes. For instance, Cai et al. [32] utilize the triad constraint, which is analogue to the independence constraint, to infer the causal structure among latent variables, while Chen et al. [26] employ the cumulant constraint to determine whether there exists a causal edge between two observed variables.

3.2 Detailed Analysis

§ Flexibility

The first property of the agnostic proxy is flexibility, which means that the underlying causal structure may be any one of those depicted in Fig. 3. In the following, we investigate whether and how, given any causal structure in Fig. 3, the causal effect of interest can be uniquely determined from the joint distribution of (T, O, Z).

In Fig. 3(a), Z is an isolated proxy. In this case, m_{OT} can be identified via an closed-form formula [12]. Specifically, Let

$$g(X, Y) := \mathrm{sgn}(\mathrm{Cov}(X, Y))\sqrt{\frac{C_{3,1}(X, Y)}{C_{1,3}(X, Y)}}, \tag{4}$$

$$f(T, O, Z) := \frac{\mathrm{Cov}(T, O) - g(T, Z)\mathrm{Cov}(O, Z)}{\mathrm{Var}(T) - g(T, Z)\mathrm{Cov}(T, Z)}, \tag{5}$$

where $\mathrm{sgn}(\cdot)$ is the sign function.

Theorem 1. *In Fig. 3(a), $m_{OT} = f(T, O, Z)$.*

In Fig. 3(b), following the discussion in Sec. 3.1, $(T, O - \lambda T)$ satisfy the cumulant constraint iff $\lambda = m_{OT}$ or m_{OL}/m_{TL}. Furthermore, $(Z, T, O - \lambda T)$ satisfies the independence constraint iff $\lambda = m_{OT}$. Formally,

Theorem 2. *In Fig. 3(b), $m_{OT} = \lambda$ iff $(T, O - \lambda T)$ satisfies the cumulant constraint and $(Z, T, O - \lambda T)$ satisfies the independence constraint.*

In Fig. 3(c), following the discussion in Sec. 3.1, $(T, O - \lambda T)$ satisfy the cumulant constraint iff $\lambda = m_{OT}$ or m_{OL}/m_{TL}. Furthermore, $(O - \lambda T, Z, T)$ satisfies the independence constraint iff $\lambda = m_{OT}$. Formally,

Theorem 3. *In Fig. 3(c), $m_{OT} = \lambda$ iff $(T, O - \lambda T)$ satisfies the cumulant constraint and $(O - \lambda T, Z, T)$ satisfies the independence constraint.*

In Fig. 3(d), m_{ZO} can be identified in the same way that m_{OT} in Fig 3(b) is identified. Next, note that Z becomes an isolated proxy after subtracting $m_{ZO}O$ from Z, so m_{OT} can be identified via Equation (5). Formally,

Theorem 4. *In Fig. 3(d), $m_{OT} = f(T, O, Z - \lambda O)$ where $(O, Z - \lambda O)$ satisfies the cumulant constraint and $(T, O, Z - \lambda O)$ satisfies the independence constraint.*

In Fig. 3(e), following the discussion in Sec. 3.1, $(Z, T - \lambda_1 Z)$ satisfies the cumulant constraint iff $\lambda_1 = m_{TZ}$ or m_{TL}/m_{ZL} and $(Z, O - \lambda_2 Z)$ satisfies the cumulant constraint iff $\lambda_2 = m_{OZ}$ or m_{OL}/m_{ZL}. Furthermore, $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint iff $(\lambda_1, \lambda_2) = (m_{TZ}, m_{OZ})$ or $(m_{TL}/m_{ZL}, m_{OL}/m_{ZL})$. In both cases, after subtracting $\lambda_1 Z, \lambda_2 Z$ from T, O respectively, Z can be regarded as an isolated proxy, so m_{OT} can be identified via Eq. (5). Formally,

Theorem 5. *In Fig. 3(e), $m_{OT} = f(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ where both $(Z, T - \lambda_1 Z)$ and $(Z, O - \lambda_2 Z)$ satisfy the cumulant constraint and $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint.*

Algorithm 1: Causal effect identification.

Input: Joint distribution of (T, O, Z)
Output: m_{OT}

if (T, O, Z) *satisfies the independence constraint* **then**
  return $f(T, O, Z)$
else if *there exists one and only one* λ *s.t. the constraints in Thm. 2 are satisfied* **then**
  return λ
else if *there exists* λ *s.t. the constraints in Thm. 3 are satisfied* **then**
  return λ
else if *there exists* λ *s.t. the constraints in Thm. 4 are satisfied* **then**
  return $f(T, O, Z - \lambda O)$
else if *there exist* λ_1, λ_2 *s.t. the constraints in Thm. 5 are satisfied* **then**
  return $f(T - \lambda_1 Z, O - \lambda_2 Z, Z)$
else
  raise Error("m_{OT} is unidentifiable.")

We can construct two SCMs $\mathcal{M}$ and $\tilde{\mathcal{M}}$ that share the same causal structure shown as Fig. 3(f):

$$\mathcal{M} := \begin{cases} L = \epsilon_L, \\ T = a_{TL}L + \epsilon_T, \\ O = a_{OT}T + a_{OL}L + \epsilon_O, \\ Z = a_{ZT}T + a_{ZL}L + \epsilon_Z, \end{cases} \quad , \quad \tilde{\mathcal{M}} := \begin{cases} \tilde{L} = \frac{1}{a_{\tilde{T}\tilde{L}}}\epsilon_T, \\ \tilde{T} = a_{\tilde{T}\tilde{L}}\tilde{L} + a_{TL}\epsilon_L, \\ \tilde{O} = \frac{a_{OT}a_{TL}+a_{OL}}{a_{TL}}\tilde{T} - \frac{a_{OL}a_{\tilde{T}\tilde{L}}}{a_{TL}}\tilde{L} + \epsilon_O, \\ \tilde{Z} = \frac{a_{ZT}a_{TL}+a_{ZL}}{a_{TL}}\tilde{T} - \frac{a_{ZL}a_{\tilde{T}\tilde{L}}}{a_{TL}}\tilde{L} + \epsilon_Z, \end{cases} . \quad (6)$$

It is easy to verify that the joint distribution of (T, O, Z) is identical to that of $(\tilde{T}, \tilde{O}, \tilde{Z})$. However, $m_{OT} = a_{OT}$ while $m_{\tilde{O}\tilde{T}} = \frac{a_{OT}a_{TL}+a_{OL}}{a_{TL}}$. Therefore, m_{OT} in Fig. 3(f) is unidentifiable. Analogously, we can also construct such two SCMs for Fig. 3(g) and Fig. 3(h). Specifically, for Fig. 3(g), we add $a_{OZ}Z$ to O, add $a_{OZ}\tilde{Z}$ to $\tilde{O}$, and keep all other assignments fixed; for Fig. 3(h), we add $a_{ZO}O$ to Z, add $a_{ZO}\tilde{O}$ to $\tilde{Z}$, and keep all other assignments fixed.

Theorem 6. *In Fig. 3(f)–(h),* m_{OT} *is unidentifiable.*

§ Uncertainty

The second property of the agnostic proxy is uncertainty, which means that it is unknown which causal structure in Fig. 3 corresponds to the ground truth. In the following, we investigate how to determine the underlying causal structure from the joint distribution of (T, O, Z).

First, since it is only in Fig. 3(a) that Z shares no causal edge with (T, O), there is an independence relationship that holds only in Fig. 3(a). Specifically,

Corollary 1. *The underlying causal structure is Fig. 3(a) iff* (T, O, Z) *satisfies the independence constraint.*

Fortunately, we find that Thms. 2–5, which are developed to identify causal effects, can also be used to identify causal structures actually. Specifically,

Corollary 2. *If the underlying causal structure is not Fig. 3(a), it is Fig. 3(b) iff there exists one and only one* λ *s.t. the constraints in Thm. 2 are satisfied.*

Corollary 3. *If the underlying causal structure is not one of Fig. 3(a)–(b), it is Fig. 3(c) iff there exists* λ *s.t. the constraints in Thm. 3 are satisfied.*

Corollary 4. *If the underlying causal structure is not one of Fig. 3(a)–(c), it is Fig. 3(d) iff there exists* λ *s.t. the constraints in Thm. 4 are satisfied.*

Corollary 5. *If the underlying causal structure is not one of Fig. 3(a)–(d), it is Fig. 3(e) iff there exist* λ_1 *and* λ_2 *s.t. the constraints in Thm. 5 are satisfied.*

Finally, it is unnecessary to distinguish between Fig. 3(f)–(h), as m_{OT} is unidentifiable in each case.

§ Identification algorithm

We summarize our identification algorithm in Alg. 1. Based on Thms. 1–6 and Cors. 1–5, our identification algorithm is sound. Formally,

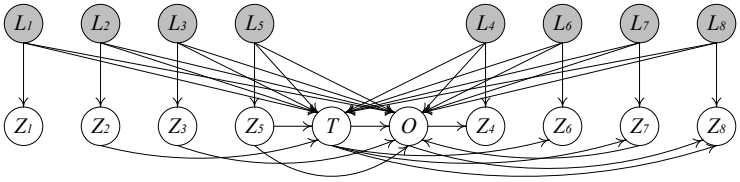


Figure 4: A typical example investigated in Sec. 4.1. The causal connections between Z_1 and (T, O) are identical to those in Fig. 3(a), those between Z_2 and (T, O) are identical to those in Fig. 3(b), ..., and those between Z_8 and (T, O) are identical to those in Fig. 3(h).

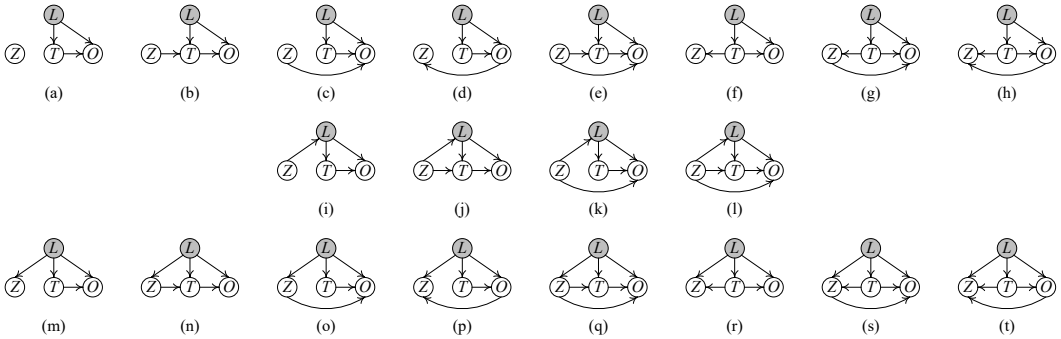


Figure 5: Causal structures considered in Sec. 4.2. Notably, they encompass all possible patterns of causal connections between Z and (L, T, O).

Theorem 7. *If Asmps. 1, 2, and 3 hold, Alg. 1 correctly identifies the true causal effect when identifiable, and correctly reports unidentifiability otherwise.*

Particularly, if the underlying causal structure is known a priori, the identification algorithm can be substantially simplified: without the need for if-else logic, it can directly identifies the causal effect of interest or reports unidentifiability based on one of Thms. 1–6.

4 Extensions

In this section, we generalize the problem setting investigated in Sec. 3 in two distinct directions.

4.1 Multiple Latent Confounders

In Sec. 3, we focus on the setting with a single latent confounder. In this section, we investigate a more complex setting with multiple latent confounders: T is still a parent of O, but T and O have multiple common parents $\mathbf{L} = \{L_1, ..., L_m\}$ that are all latent variables and also root variables. Each L_i has an observed child Z_i which is not a child of any other $L_j \neq L_i$ serving as its agnostic proxy and there exists no causal edge between any two $Z_i, Z_j \in \mathbf{Z} = \{Z_1, ..., Z_m\}$. A typical example is shown as Fig. 4. Our goal is to identify the m_{OT} from the joint distribution of $(T, O, Z_1, ..., Z_m)$.

For this more complex setting, we present and prove a series of theoretical results (App. C.1), which collectively establish the soundness of our proposed identification algorithm (App. C.2). Intuitively, although multiple latent confounders exist between T and O, there is only a single latent confounder L_i between each Z_i and (T, O). Following the theoretical results in Sec. 3, the existence, direction, and strength of the causal connections between Z_i and (T, O) can be determined. With this knowledge, the causal connections between Z_i and (T, O) can be removed, thereby rendering Z_i an isolated proxy. With isolated proxies, m_{OT} can be identified via an closed-form formula similar to Equation (5) [12].

4.2 Generalized Agnostic Proxy

While Z is always a child of L in Sec. 3, we let Z be a generalized agnostic proxy in this section: T is still a parent of O and L is still a common parent of both T and O, but the causal connections between Z and (L, T, O) can be arbitrary and are not known a priori. That is, the underlying causal

structure may be any one of those depicted in Fig. 5 (note that (m)–(t) in Fig. 5 are exactly (a)–(h) in Fig. 3) and it is unknown which one corresponds to the ground truth. Our goal is to identify m_{OT} from the joint distribution of (T, O, Z).

For this more complex setting, we present and prove a series of theoretical results (App. D.1), which collectively establish the soundness of our identification algorithm (App. D.2). In brief, our identification algorithm first determines whether the underlying causal structure is one of Fig. 5(m)–(t) by examining certain independence relationships. If it is, the rest of the problem can be solved by Alg. 1; if it is not, our identification algorithm reports unidentifiability because the causal effect of interest is unidentifiable without prior knowledge about the underlying causal structure.

5 Related Work

A substantial body of work in machine learning has focused on studying and exploiting causality [33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48]. Within this broader literature, our work focuses on causal effect identification in the presence of latent confounding [49, 50, 51, 7, 52, 53]. One of the most frequently used strategies to handle latent confounding is through proxies. An early study [7] investigates two different settings. First, in non-parametric SCMs, the causal effect of the treatment T on the outcome O can be identified if (i) the latent confounder L has an isolated proxy Z and $P(Z|L)$ is known a priori; or (ii) L has two proxies Z and W where $W \perp\!\!\!\perp (T, O, Z)|L$ and $(Z, W) \perp\!\!\!\perp O|\{L, T\}$. Second, in linear SCMs, the causal effect can be identified if L has two proxies Z and W where $Z \perp\!\!\!\perp O|\{L, T\}$ and $W \perp\!\!\!\perp (T, Z)|L$.

Following this seminal work, many subsequent studies delve deeper into causal effect identification in non-parametric SCMs [8, 9, 54, 55, 10, 11]. On the one hand, it is known that unless certain stringent conditions (e.g., $P(Z|L)$ is known a priori and Z is an isolated proxy) are satisfied, the causal effect is generally unidentifiable with a single proxy [56, 7, 57], so all of these studies require multiple proxies subject to specific structural constraints. On the other hand, a few recent works establish various other conditions for single-proxy identifiability. For instance, Xu et al. [57] rely on two assumptions: (i) the outcome O is deterministically generated given the treatment T and the latent confounder L, and (ii) the single proxy Z is an isolated proxy. Iglesias-Alonso et al. [58] need multi-domain data exhibiting sufficient variability where the domain shifts are driven exclusively by changes of the latent confounder. In addition, they require that the single proxy Z has no causal connection with T and it is not a child of O. Evidently, these conditions are all highly restrictive.

Recently, many works investigate parametric SCMs, particularly linear SCMs [12, 13, 14, 15, 16]. The major contributions of [13, 14] lie in data-driven methods for selecting valid proxies from all observed variables, they still require multiple proxies. Kivva et al. [12] prove that in linear SCMs where exogenous noises are Gaussian, the causal effect is unidentifiable with a single isolated proxy. Within lvLiNGAM, they propose an identification algorithm requiring only a single isolated proxy, while the identification algorithm of Tramontano et al. [16] slightly relaxed the requirement for an isolated proxy by allowing a causal edge from the proxy to the treatment. The former is based on the moment while the latter is based on the cumulant, which are two informationally equivalent characterizations of a probability distribution. Tramontano et al. [15] regard lvLiNGAM as an overcomplete independent components analysis (OICA) model, then causal effect identification essentially boils down to identifying an OICA model, which is an intractable problem. Their identification algorithm is based on graphical reconstruction independent components analysis (GRICA), which lacks a theoretical guarantee for correctly identifying the causal effect. Notably, all of [12, 15, 16] require the underlying causal structure to be known a priori. In contrast, we require only a single agnostic proxy and provide rigorous theoretical guarantees.

6 Experiments

Based on our proposed identification strategy in Alg. 1, it is straightforward to construct a plug-in estimation method that operates on finite samples. For instance, we can use sample cumulants for the cumulant constraints in Alg. 1 and independence tests based on the Hilbert-Schmidt Independence Criterion (HSIC) for the independence constraints in Alg. 1. The resulting plug-in method can be made asymptotically consistent under suitable regularity conditions. First, sample cumulants and the quantities derived from them are consistent by the law of large numbers and the continuous mapping

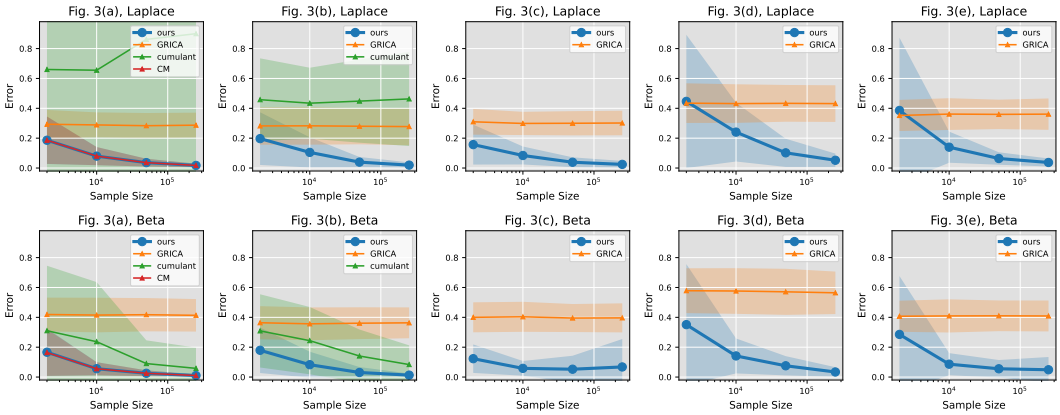


Figure 6: The underlying causal structure is known a priori and the causal effect is identifiable.

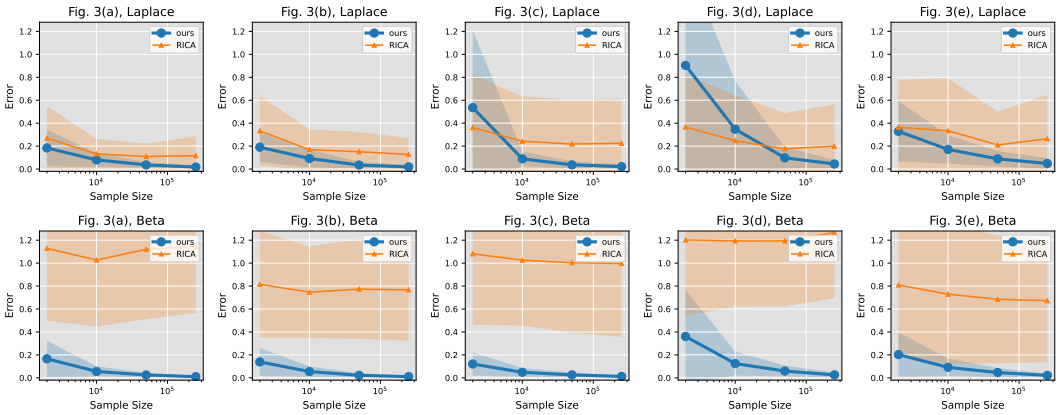


Figure 7: The underlying causal structure is not known a priori and the causal effect is identifiable.

theorem. Second, with a bounded characteristic kernel such as the Gaussian kernel, HSIC vanishes if and only if the two variables are independent [59, 60], and its empirical estimate converges to the corresponding population value [61]. By letting the significance level tend to zero sufficiently slowly, both type-I and type-II error probabilities can be made to vanish asymptotically. Finally, since only finitely many cumulant estimations and independence tests are performed, a union-bound argument establishes asymptotic consistency: the estimated causal effect converges in probability to its true value when identifiable, and unidentifiability is reported with probability tending to one otherwise. In practice, we employ large-scale independence tests based on the Random Fourier Feature (RFF) HSIC [62], a computationally efficient approximation to HSIC whose computational complexity scales linearly with the sample size. In addition, we fix the significance level at 0.05 in all experiments for simplicity.

Data. For each causal structure in Fig. 3(a)–(h), we consider sample sizes of 2k, 10k, 50k, 250k; for each sample size, we produce 500 datasets. Across all 500 datasets, each direct causal strength is sampled from a uniform distribution over $[-1.0, -0.5] \cup [0.5, 1.0]$. In 250 of the 500 datasets, exogenous noises are drawn from a Laplace distribution with $\mu = 0, b = 1$, while in the remaining 250 datasets they are drawn from a Beta distribution with $\alpha = 1/3, \beta = 2/3$.

Results in Fig. 6. When the underlying causal structure is known a priori and the causal effect of interest is identifiable, we use CM [12], cumulant [16], and GRICA [15] as baselines, and report the performances in Fig. 6, where the error is given by |Estimated Value-True Value| / |True Value| following [12, 16, 15] and the shadowed band denotes standard deviation. As CM is only applicable to isolated proxies, we report its performance only on Fig. 3(a). For the same reason, we report that of cumulants only on Fig. 3(a) and (b). As the sample size increases, the errors of our method gradually approach zero and eventually become the lowest among all baselines, providing empirical support for our theoretical results. In contrast, the errors of GRICA remain consistently at a substantial level because it lacks a theoretical guarantee of correctness as stated in Sec. 5.

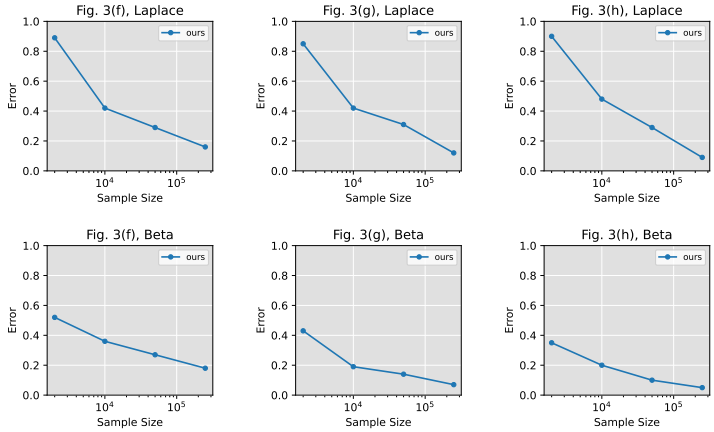


Figure 8: The underlying causal structure is not known a priori and the causal effect is unidentifiable.

Results in Fig. 7. When the underlying causal structure is not known a priori and the causal effect of interest is identifiable, we use RICA [23] as the baseline, which is designed to recover the entire causal model rather than a specific causal effect, and report the performances in Fig. 7. While the errors of our method gradually approach zero as the sample size increases, the errors of RICA remain consistently substantial and do not decrease as the sample size increases. This is because RICA also lacks a theoretical guarantee of correctness, similar to GRICA.

Results in Fig. 8. When the underlying causal structure is not known a priori and the causal effect of interest is unidentifiable, we report the performances in Fig. 8, where the error is defined as the proportion of unidentifiable cases that are incorrectly determined as identifiable by our estimation method. As the sample size increases, the errors of our method gradually approach zero.

7 Conclusion

This paper investigates causal effect identification in the presence of latent confounding with a single agnostic proxy. Within lvLiNGAM, we propose a novel identification algorithm that synergistically combines the cumulant constraint and the independence constraint within a coherent paradigm, which first derives candidate solutions via the former and then isolates the valid solution via the latter. We provide both theoretical guarantees and empirical validation. Our work not only broadens the applicability of causal effect identification but also may benefit research in natural and social sciences.

Limitations. First, since we assume that T is a parent of O, our theoretical guarantees do not cover the case where the causal effect is zero. Second, this paper primarily focuses on causal effect identification (whether and how the causal effect can be uniquely determined from the observational population distribution) rather than estimation (how to compute a reliable estimate for the causal effect given a finite sample). Accordingly, we construct our estimation method as a straightforward finite-sample version of the identification algorithm, without optimizing for its small-sample performance or evaluating its robustness in the main text. Nevertheless, we provide some preliminary discussions on these aspects in Apps. E.2 and E.3, leaving a deeper investigation for future work.

Acknowledgment

Tongliang Liu is partially supported by the following Australian Research Council projects: DP260102466, FT220100318, DP220102121, LP220100527, LP220200949. Xiu-Chuan Li is partially supported by ARC FT220100318 and Google PhD Fellowship.

References

[1] Michael E Sobel. Causal inference in the social sciences. *Journal of the American Statistical Association*, 95(450):647–651, 2000.

[2] Elizabeth L Ogburn, Oleg Sofrygin, Ivan Diaz, and Mark J Van der Laan. Causal inference for social network data. *Journal of the American Statistical Association*, 119(545):597–611, 2024.

[3] Miguel A Hernán and James M Robins. Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health*, 60(7):578–586, 2006.

[4] Michele Hilton Boon, Peter Craig, Hilary Thomson, Mhairi Campbell, and Laurence Moore. Regression discontinuity designs in health: a systematic review. *Epidemiology*, 32(1):87–93, 2021.

[5] David Wuepper and Robert Finger. Regression discontinuity designs in agricultural and environmental economics. *European Review of Agricultural Economics*, 50(1):1–28, 2023.

[6] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

[7] Manabu Kuroki and Judea Pearl. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2):423–437, 2014.

[8] Wang Miao, Zhi Geng, and Eric J Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018.

[9] Xu Shi, Wang Miao, Jennifer C Nelson, and Eric J Tchetgen Tchetgen. Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(2):521–540, 2020.

[10] Ilya Shpitser, Zach Wood-Doughty, and Eric J Tchetgen Tchetgen. The proximal id algorithm. *Journal of Machine Learning Research*, 24(188):1–46, 2023.

[11] Yifan Cui, Hongming Pu, Xu Shi, Wang Miao, and Eric Tchetgen Tchetgen. Semiparametric proximal causal inference. *Journal of the American Statistical Association*, 119(546):1348–1359, 2024.

[12] Yaroslav Kivva, Saber Salehkaleybar, and Negar Kiyavash. A cross-moment approach for causal effect estimation. In *Conference on Neural Information Processing Systems*, pages 9944–9955, 2023.

[13] Erich Kummerfeld, Jaewon Lim, and Xu Shi. Data-driven automated negative control estimation (dance): search for, validation of, and causal inference with negative controls. *Journal of Machine Learning Research*, 25(229):1–35, 2024.

[14] Feng Xie, Zhengming Chen, Shanshan Luo, Wang Miao, Ruichu Cai, and Zhi Geng. Automating the selection of proxy variables of unmeasured confounders. In *International Conference on Machine Learning*, pages 54430–54459, 2024.

[15] Daniele Tramontano, Yaroslav Kivva, Saber Salehkaleybar, Mathias Drton, and Negar Kiyavash. Causal effect identification in lingam models with latent confounders. In *International Conference on Machine Learning*, pages 48468–48493, 2024.

[16] Daniele Tramontano, Yaroslav Kivva, Saber Salehkaleybar, Negar Kiyavash, and Mathias Drton. Causal effect identification in lvlingam from higher-order cumulants. In *International Conference on Machine Learning*, pages 59924–59944, 2025.

[17] Patrik O Hoyer, Shohei Shimizu, Antti J Kerminen, and Markus Palviainen. Estimation of causal effects using linear non-gaussian causal models with hidden variables. *International Journal of Approximate Reasoning*, 49(2):362–378, 2008.

[18] Alessio Moneta, Doris Entner, Patrik O Hoyer, and Alex Coad. Causal inference by independent component analysis: Theory and applications. *Oxford Bulletin of Economics and Statistics*, 75(5):705–730, 2013.

[19] Pablo Campomanes, Marilisa Neri, Bruno AC Horta, Ute F Röhrig, Stefano Vanni, Ivano Tavernelli, and Ursula Rothlisberger. Origin of the spectral shifts among the early intermediates of the rhodopsin photocycle. *Journal of the American Chemical Society*, 136(10):3842–3851, 2014.

[20] Takeshi Ogawa, Hideki Shimobayashi, Jun-Ichiro Hirayama, and Motoaki Kawanabe. Asymmetric directed functional connectivity within the frontoparietal motor network during motor imagery and execution. *Neuroimage*, 247:118794, 2022.

[21] Taku Moriyama and Masashi Kuwano. Causal inference for contemporaneous effects and its application to tourism product sales data. *Journal of Marketing Analytics*, 10(3):250–260, 2022.

[22] Atsushi Kurotani, Hirokuni Miyamoto, and Jun Kikuchi. Validation of causal inference data using directlingam in an environmental small-scale model and calculation settings. *MethodsX*, 12:102528, 2024.

[23] Saber Salehkaleybar, AmirEmad Ghassami, Negar Kiyavash, and Kun Zhang. Learning linear non-gaussian causal models in the presence of latent variables. *Journal of Machine Learning Research*, 21(39):1–24, 2020.

[24] Feng Xie, Biwei Huang, Zhengming Chen, Yangbo He, Zhi Geng, and Kun Zhang. Identification of linear non-gaussian latent hierarchical structure. In *International Conference on Machine Learning*, pages 24370–24387, 2022.

[25] Ruichu Cai, Zhiyi Huang, Wei Chen, Zhifeng Hao, and Kun Zhang. Causal discovery with latent confounders based on higher-order cumulants. In *International Conference on Machine Learning*, pages 3380–3407, 2023.

[26] Wei Chen, Zhiyi Huang, Ruichu Cai, Zhifeng Hao, and Kun Zhang. Identification of causal structure with latent variables based on higher order cumulants. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 20353–20361, 2024.

[27] Xiu-Chuan Li, Jun Wang, and Tongliang Liu. Recovery of causal graph involving latent variables via homologous surrogates. In *International Conference on Learning Representations*, 2025.

[28] Xiu-Chuan Li and Tongliang Liu. Efficient and trustworthy causal discovery with latent variables and complex relations. In *International Conference on Learning Representations*, 2025.

[29] Yewei Xia, Zhengming Chen, Haoyue Dai, Fuhong Wang, Yixin Ren, Yiqing Li, Kun Zhang, and Shuigeng Zhou. Conditional independent component analysis for estimating causal structure with latent variables. In *International Conference on Learning Representations*, 2026.

[30] Mikhail Belkin, Luis Rademacher, and James Voss. Blind signal separation in the presence of gaussian noise. In *Conference on Learning Theory*, pages 270–287, 2013.

[31] George Darmois. Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire. *Revue de l'Institut international de statistique*, pages 2–8, 1953.

[32] Ruichu Cai, Feng Xie, Clark Glymour, Zhifeng Hao, and Kun Zhang. Triad constraints for learning causal structure of latent variables. In *Conference on Neural Information Processing Systems*, pages 12883–12892, 2019.

[33] Yongqiang Chen, Yonggang Zhang, Yatao Bian, Han Yang, MA Kaili, Binghui Xie, Tongliang Liu, Bo Han, and James Cheng. Learning causally invariant representations for out-of-distribution generalization on graphs. In *Conference on Neural Information Processing Systems*, pages 22131–22148, 2022.

[34] Zhuo Huang, Xiaobo Xia, Li Shen, Bo Han, Mingming Gong, Chen Gong, and Tongliang Liu. Harnessing out-of-distribution examples via augmenting content and style. In *International Conference on Learning Representations*, 2023.

[35] Yuanyuan Wang, Xi Geng, Wei Huang, Biwei Huang, and Mingming Gong. Generator identification for linear sdes with additive and multiplicative noise. In *Conference on Neural Information Processing Systems*, pages 64103–64138, 2023.

[36] Yexiong Lin, Yu Yao, Xiaolong Shi, Mingming Gong, Xu Shen, Dong Xu, and Tongliang Liu. Cs-isolate: Extracting hard confident examples by content and style isolation. In *Conference on Neural Information Processing Systems*, pages 58556–58576, 2023.

[37] Xiu-Chuan Li, Kun Zhang, and Tongliang Liu. Causal structure recovery with latent variables under milder distributional and graphical assumptions. In *International Conference on Learning Representations*, 2024.

[38] Yuhang Liu, Zhen Zhang, Dong Gong, Erdun Gao, Biwei Huang, Mingming Gong, Anton van den Hengel, Kun Zhang, and Javen Qinfeng Shi. Towards identifiable latent additive noise models. *arXiv preprint arXiv:2403.15711*, 2024.

[39] Yexiong Lin, Yu Yao, and Tongliang Liu. Learning the latent causal structure for modeling label noise. In *Conference on Neural Information Processing Systems*, pages 120549–120577, 2024.

[40] Yuanyuan Wang, Biwei Huang, Wei Huang, Xi Geng, and Mingming Gong. Identifiability analysis of linear ode systems with hidden confounders. In *Conference on Neural Information Processing Systems*, pages 59054–59092, 2024.

[41] Yuanyuan Wang, Wei Huang, Mingming Gong, Xi Geng, Tongliang Liu, Kun Zhang, and Dacheng Tao. Identifiability and asymptotics in learning homogeneous linear ode systems from discrete observations. *Journal of Machine Learning Research*, 25(154):1–50, 2024.

[42] Yexiong Lin, Yu Yao, Zhaoqing Wang, Xu Shen, Jun Yu, Bo Han, and Tongliang Liu. Improving the instance-dependent transition matrix estimation by exploiting self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[43] Yuhang Liu, Zhen Zhang, Dong Gong, Mingming Gong, Biwei Huang, Anton van den Hengel, Kun Zhang, and Javen Qinfeng Shi. Identifying weight-variant latent causal models. *Journal of Machine Learning Research*, 27(4):1–49, 2026.

[44] Erdun Gao and Dino Sejdinovic. Activecq: Active estimation of causal quantities. In *International Conference on Learning Representations*, 2026.

[45] Yuhang Liu, Zhen Zhang, Dong Gong, Erdun Gao, Biwei Huang, Mingming Gong, Anton Hengel, Kun Zhang, and Javen Qinfeng Shi. Beyond dags: A latent partial causal model for multimodal learning. In *International Conference on Learning Representations*, 2026.

[46] Heejeong Nam, Quentin Le Lidec, Lucas Maes, Yann LeCun, and Randall Balestriero. Causal-jepa: Learning world models through object-level latent masking. In *International Conference on Machine Learning*, 2026.

[47] Erdun Gao, Jake Fawkes, and Dino Sejdinovic. Causal-epig: Causally aligned active cate estimation. In *International Conference on Machine Learning*, 2026.

[48] Ziming Xu, Shuang Liang, Ruobing Han, Ziqiao Xi, Mingxing Rao, Kun Zhou, Zijun Zhang, Yuchen Yan, Yufan Wei, Junbo Huang, et al. Causalwm: Causal chain-of-thought reasoning for embodied world model. *arXiv preprint arXiv:2609.23184*, 2026.

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

[50] Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

[51] Carlos Brito and Judea Pearl. Generalized instrumental variables. In *Conference on Uncertainty in Artificial Intelligence*, pages 85–93, 2002.

[52] Dominik Rothenhäusler, Peter Bühlmann, and Nicolai Meinshausen. Causal dantzig. *The Annals of Statistics*, 47(3):1688–1722, 2019.

[53] Xiu-Chuan Li, James Kwok, Jiaxian Guo, and Tongliang Liu. Causal effect identifiability in the presence of latent confounders without auxiliary variables. In *International Conference on Machine Learning*, 2026.

[54] Afsaneh Mastouri, Yuchen Zhu, Limor Gultchin, Anna Korba, Ricardo Silva, Matt J Kusner, Arthur Gretton, and Krikamol Muandet. Proximal causal learning with kernels: Two-stage estimation and moment restriction. In *International Conference on Machine Learning*, pages 7512–7523, 2021.

[55] Liyuan Xu, Heishiro Kanagawa, and Arthur Gretton. Deep proxy causal learning and its application to confounded bandit policy evaluation. In *Conference on Neural Information Processing Systems*, pages 26264–26275, 2021.

[56] Judea Pearl. On measurement bias in causal inference. In *Conference on Uncertainty in Artificial Intelligence*, pages 425–432, 2010.

[57] Liyuan Xu and Arthur Gretton. Kernel single proxy control for deterministic confounding. In *International Conference on Artificial Intelligence and Statistics*, pages 3736–3744, 2025.

[58] Manuel Iglesias-Alonso, Felix Schur, Julius von Kügelgen, and Jonas Peters. Transferring causal effects using proxies. In *Conference on Neural Information Processing Systems*, 2025.

[59] Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International Conference on Algorithmic Learning Theory*, pages 63–77, 2005.

[60] Bharath K Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert RG Lanckriet. Hilbert space embeddings and metrics on probability measures. *The Journal of Machine Learning Research*, 11:1517–1561, 2010.

[61] Arthur Gretton, Kenji Fukumizu, Choon Teo, Le Song, Bernhard Schölkopf, and Alex Smola. A kernel statistical test of independence. In *Conference on Neural Information Processing Systems*, 2007.

[62] Qinyi Zhang, Sarah Filippi, Arthur Gretton, and Dino Sejdinovic. Large-scale kernel methods for independence testing. *Statistics and Computing*, 28(1):113–130, 2018.

[63] Susanna Wang, Nadir Yehya, Eric E Schadt, Hui Wang, Thomas A Drake, and Aldons J Lusis. Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS genetics*, 2(2):148–159, 2006.

[64] Wang Miao, Wenjie Hu, Elizabeth L Ogburn, and Xiao-Hua Zhou. Identifying effects of multiple treatments in the presence of unmeasured confounding. *Journal of the American Statistical Association*, 118(543):1953–1967, 2023.

[65] Stephen B Wheatcroft, Mark T Kearney, Ajay M Shah, Vivienne A Ezzat, John R Miell, Michael Modo, Stephen CR Williams, Will P Cawthorn, Gema Medina-Gomez, Antonio Vidal-Puig, et al. Igf-binding protein-2 protects against the development of obesity and insulin resistance. *Diabetes*, 56(2):285–294, 2007.

[66] Marc Schneeberger. Irx3, a new leader on obesity genetics. *Ebiomedicine*, 39:19–20, 2019.

Table of Contents

A Important Lemmas **15**

B Proof of Theoretical Results in Sec. 3.2 **17**

C Extension 1: Multiple Latent Confounders **22**

C.1 Theoretical Results 22

C.2 Identification Algorithm 27

D Extension 2: Generalized Agnostic Proxy **28**

D.1 Theoretical Results 28

D.2 Identification Algorithm 31

E Causal Effect Estimation **32**

E.1 Real-world Data 32

E.2 Small-Sample Performance 32

E.3 Robustness 32

A Important Lemmas

Under Asmp. 1, we have

1. $m_{V_2V_1} \neq 0$ iff $V_1 \in \text{An}(V_2)$.
2. $m_{V_iV_j} = a_{V_iV_j} \neq 0$ iff $V_j \in \text{Pa}(V_i)$ and there is no other directed path from V_j to V_i.
3. If $\{V_1, V_2\} \subseteq \text{An}(V_3) \cap \text{An}(V_4)$ and there exist two non-intersecting directed paths from $\{V_1, V_2\}$ to $\{V_3, V_4\}$, then $m_{V_3V_1}/m_{V_3V_2} \neq m_{V_4V_1}/m_{V_4V_2}$.

§ Cumulants

Lemma 1. *Suppose three random variables V, V_1, V_2 can be expressed as*

$$V = e, \quad V_1 = \alpha_1 e + e_1, \quad V_2 = \alpha_2 e + e_2,$$

where e, e_1, e_2 are mutually independent, $\alpha_1\alpha_2 \neq 0$, and $C_4(e) \neq 0$, then (V_1, V_2) satisfies the cumulant constraint.

Proof. Based on the independence property of cumulants, we have

$$\begin{aligned} C_{3,1}(V_1, V_2) &= \alpha_1^3\alpha_2 C_4(e), \\ C_{1,3}(V_1, V_2) &= \alpha_1\alpha_2^3 C_4(e), \\ C_{2,2}(V_1, V_2) &= \alpha_1^2\alpha_2^2 C_4(e). \end{aligned}$$

Therefore, $C_{3,1}(V_1, V_2)C_{1,3}(V_1, V_2) = C_{2,2}^2(V_1, V_2)$. □

Lemma 2. *Suppose three random variables V, V_1, V_2 can be expressed as*

$$V = e, \quad V_1 = \alpha_1 e + e_1, \quad V_2 = \alpha_2 e + \beta e_1 + e_2,$$

where e, e_1, e_2 are mutually independent, $\alpha_1\alpha_2\beta \neq 0$, $2\alpha_1\alpha_2\beta - \alpha_2^2 - \alpha_1^2\beta^2 \neq 0$, and $C_4(e)C_4(e_1) \neq 0$, then $(V_1, V_2 - \lambda V_1)$ satisfies the cumulant constraint iff $\lambda = \beta$ or $\frac{\alpha_2}{\alpha_1}$.

Proof. Based on the exchangeability and linearity of cumulants, we have

$$C_{3,1}(V_1, V_2 - \lambda V_1) = C_{3,1}(V_1, V_2) - \lambda C_4(V_1),$$
$$C_{1,3}(V_1, V_2 - \lambda V_1) = C_{1,3}(V_1, V_2) - 3\lambda C_{2,2}(V_1, V_2) + 3\lambda^2 C_{3,1}(V_1, V_2) - \lambda^3 C_4(V_1),$$
$$C_{2,2}(V_1, V_2 - \lambda V_1) = C_{2,2}(V_1, V_2) - 2\lambda C_{3,1}(V_1, V_2) + \lambda^2 C_4(V_1).$$

Let $C_{3,1}(V_1, V_2 - \lambda V_1)C_{1,3}(V_1, V_2 - \lambda V_1) = C_{2,2}(V_1, V_2 - \lambda V_1)^2$, we have

$$(C_{3,1}^2(V_1, V_2) - C_{2,2}(V_1, V_2)C_4(V_1))\lambda^2 + (C_{1,3}(V_1, V_2)C_4(V_1) - C_{3,1}(V_1, V_2)C_{2,2}(V_1, V_2))\lambda + C_{2,2}^2(V_1, V_2) - C_{3,1}(V_1, V_2)C_{1,3}(V_1, V_2) = 0. \quad (7)$$

Based on the independence property of cumulants, we have

$$C_{3,1}(V_1, V_2) = \alpha_1^3\alpha_2 C_4(e) + \beta C_4(e_1),$$
$$C_{1,3}(V_1, V_2) = \alpha_1\alpha_2^3 C_4(e) + \beta^3 C_4(e_1),$$
$$C_{2,2}(V_1, V_2) = \alpha_1^2\alpha_2^2 C_4(e) + \beta^2 C_4(e_1),$$
$$C_4(V_1) = \alpha_1^4 C_4(e) + C_4(e_1).$$

We can write Eq. (7) as

$$(2\alpha_1^2\alpha_2\beta - \alpha_1\alpha_2^2 - \alpha_1^3\beta^2)\lambda^2 + (\alpha_2^3 + \alpha_1^3\beta^3 - \alpha_1\alpha_2^2\beta - \alpha_1^2\alpha_2\beta^2)\lambda + (2\alpha_1\alpha_2^2\beta^2 - \alpha_1^2\alpha_2\beta^3 - \alpha_2^3\beta) = 0. \quad (8)$$

Because $\alpha_1\alpha_2\beta \neq 0$ and $2\alpha_1\alpha_2\beta - \alpha_2^2 - \alpha_1^2\beta^2 \neq 0$, Eq. (8) is a quadratic equation in λ. Trivially, its two roots are β and $\frac{\alpha_2}{\alpha_1}$. □

§ Independence

In the following, Lem. 3 gives a sufficient condition for the independence constraint to be satisfied, whereas Lems. 4 and 5 give two sufficient conditions for the independence constraint to be violated.

Lemma 3. *Suppose three random variables V_1, V_2, V_3 can be expressed as*

$$V_1 = c_1 e + e_1, \quad V_2 = c_2 e + e_2, \quad V_3 = c_3 e + e_3,$$

where $e, (e_1, e_2), e_3$ are mutually independent[2]*, $c_1c_2c_3 \neq 0$, and $\mathrm{Var}(e) \neq 0$, then (V_1, V_2, V_3) satisfies the independence constraint.*

Proof.

$$\mathrm{Cov}(V_1, V_3)\mathrm{Cov}(V_2, V_3) = c_1c_2c_3^2(\mathrm{Var}(e))^2 \neq 0,$$
$$\begin{aligned}\mathrm{Cov}(V_2, V_3)V_1 - \mathrm{Cov}(V_1, V_3)V_2 &= c_2c_3\mathrm{Var}(e)(c_1e + e_1) - c_1c_3\mathrm{Var}(e)(c_2e + e_2)\\ &= c_2c_3\mathrm{Var}(e)e_1 - c_1c_3\mathrm{Var}(e)e_2 \perp\!\!\!\perp V_3.\end{aligned}$$

Therefore, (V_1, V_2, V_3) satisfies the independence constraint. □

Lemma 4. *Given three variables V_1, V_2, V_3, if $\exists V \in \mathbf{V}$ s.t. only one of m_{V_1V} and m_{V_2V} is non-zero and m_{V_3V} is non-zero, then (V_1, V_2, V_3) does not satisfy the independence constraint.*

Proof. If $\mathrm{Cov}(V_1, V_3)\mathrm{Cov}(V_2, V_3) = 0$, (V_1, V_2, V_3) does not satisfy the independence constraint. Otherwise, both $\mathrm{Cov}(V_2, V_3)V_1 - \mathrm{Cov}(V_1, V_3)V_2$ and V_3 contains ϵ_V, so $\mathrm{Cov}(V_2, V_3)V_1 - \mathrm{Cov}(V_1, V_3)V_2 \not\perp\!\!\!\perp V_3$ based on Darmois-Skitovitch Theorem, that is, (V_1, V_2, V_3) also does not satisfy the independence constraint. □

Lemma 5. *Given three variables V_1, V_2, V_3, if (V_1, V_2, V_3) satisfies the independence constraint and $m_{V_3V_i}m_{V_3V_j} \neq 0$, then $m_{V_1V_i}m_{V_2V_j} = m_{V_1V_j}m_{V_2V_i}$.*

Proof. We prove this lemma by contradiction. Suppose $m_{V_1V_i}m_{V_2V_j} \neq m_{V_1V_j}m_{V_2V_i}$,

$$\begin{aligned}&\mathrm{Cov}(V_2, V_3)V_1 - \mathrm{Cov}(V_1, V_3)V_2\\ =&(\mathrm{Cov}(V_2, V_3)m_{V_1V_i} - \mathrm{Cov}(V_1, V_3)m_{V_2V_i})\epsilon_{V_i} + (\mathrm{Cov}(V_2, V_3)m_{V_1V_j} - \mathrm{Cov}(V_1, V_3)m_{V_2V_j})\epsilon_{V_j} + \ldots\end{aligned}$$

Since (V_1, V_2, V_3) satisfies the independence constraint, $\mathrm{Cov}(V_1, V_3)\mathrm{Cov}(V_2, V_3) \neq 0$, so $\mathrm{Cov}(V_2, V_3)V_1 - \mathrm{Cov}(V_1, V_3)V_2$ contains at least one of ϵ_{V_i} and ϵ_{V_j}. Based on Darmois-Skitovitch Theorem, $\mathrm{Cov}(V_2, V_3)V_1 - \mathrm{Cov}(V_1, V_3)V_2 \not\perp\!\!\!\perp V_3$, which leads to contradiction. □

[2] e_1 is not necessarily independent of e_2

B Proof of Theoretical Results in Sec. 3.2

Theorem 1. In Fig. 3(a), $m_{OT} = f(T, O, Z)$.

Proof. Note that

$$C_{3,1}(T, Z) = m_{TL}^3 m_{ZL} C_4(\epsilon_L) \neq 0,$$
$$C_{1,3}(T, Z) = m_{TL} m_{ZL}^3 C_4(\epsilon_L) \neq 0,$$
$$\text{Cov}(T, Z) = m_{TL} m_{ZL} \text{Var}(\epsilon_L) \neq 0,$$

so

$$g(T, Z) = \frac{m_{TL}}{m_{ZL}}.$$

Based on Eq. (4) and Thm. 1 in [12], this theorem holds. □

Theorem 2. In Fig. 3(b), $\lambda = m_{OT}$ iff $(T, O - \lambda T)$ satisfies the cumulant constraint and $(Z, T, O - \lambda T)$ satisfies the independence constraint.

Proof. In Fig. 3(b),

$$T = m_{TL}\epsilon_L + m_{TZ}\epsilon_Z + \epsilon_T,$$
$$O = m_{OL}\epsilon_L + m_{OT}(m_{TZ}\epsilon_Z + \epsilon_T) + \epsilon_O,$$

let ϵ_L and $m_{TZ}\epsilon_Z + \epsilon_T$ be e and e_1 in Lem. 2, $(T, O - \lambda T)$ satisfies the cumulant constraint iff $\lambda = m_{OT}$ or $\frac{m_{OL}}{m_{TL}}$.

"Only if": If $\lambda = m_{OT}$,

$$Z = m_{ZL}\epsilon_L + \epsilon_Z,$$
$$T = m_{TL}\epsilon_L + m_{TZ}\epsilon_Z + \epsilon_T,$$
$$O - \lambda T = a_{OL}\epsilon_L + \epsilon_O.$$

Let ϵ_L be e in Lem. 3, $(Z, T, O - \lambda T)$ satisfies the independence constraint.

"If": If $\lambda = m_{OL}/m_{TL}$,

$$Z = m_{ZL}\epsilon_L + \epsilon_Z,$$
$$T = m_{TL}\epsilon_L + m_{TZ}\epsilon_Z + \epsilon_T,$$
$$O - \lambda T = (m_{OZ} - \frac{m_{OL}}{m_{TL}} m_{TZ})\epsilon_Z + (m_{OT} - \frac{m_{OL}}{m_{TL}})\epsilon_T + \epsilon_O.$$

Note that $m_{OT} - \frac{m_{OL}}{m_{TL}} \neq 0$, $O - \lambda T$ contain ϵ_T. Besides, $m_{TT} \neq 0$ and $m_{ZT} = 0$, so $(Z, T, O - \lambda T)$ does not satisfy the independence constraint based on Lem. 4. □

Theorem 3. In Fig. 3(c), $m_{OT} = \lambda$ iff $(T, O - \lambda T)$ satisfies the cumulant constraint and $(O - \lambda T, Z, T)$ satisfies the independence constraint.

Proof. In Fig. 3(c),

$$T = m_{TL}\epsilon_L + \epsilon_T,$$
$$O = m_{OL}\epsilon_L + m_{OT}\epsilon_T + m_{OZ}\epsilon_Z + \epsilon_O,$$

let ϵ_L and ϵ_T be e and e_1 in Lem. 2, $(T, O - \lambda T)$ satisfies the cumulant constraint iff $\lambda = m_{OT}$ or $\frac{m_{OL}}{m_{TL}}$.

"Only if": If $\lambda = m_{OT}$, $(T, O - \lambda T)$ satisfies the cumulant constraint based on Lem. 2. Besides,

$$O - \lambda T = (m_{OL} - m_{OT} m_{TL})\epsilon_L + m_{OZ}\epsilon_Z + \epsilon_O,$$
$$Z = m_{ZL}\epsilon_L + \epsilon_Z,$$
$$T = m_{TL}\epsilon_L + \epsilon_T.$$

Note that $m_{OL} - m_{OT}m_{TL} \neq 0$, let ϵ_L be e in Lem. 3, $(O - \lambda T, Z, T)$ satisfies the independence constraint.

"If": If $\lambda = m_{OL}/m_{TL}$,

$$O - \lambda T = m_{OZ}\epsilon_Z + (m_{OT} - \frac{m_{OL}}{m_{TL}})\epsilon_T + \epsilon_O,$$
$$Z = m_{ZL}\epsilon_L + \epsilon_Z,$$
$$T = m_{TL}\epsilon_L + \epsilon_T.$$

Note that $m_{OT} - \frac{m_{OL}}{m_{TL}} \neq 0$, both $O - \lambda T$ and T contain ϵ_T while Z does not contain ϵ_T, so $(O - \lambda T, Z, T)$ does not satisfy the independence constraint based on Lem. 4. □

Theorem 4. In Fig. 3(d), $m_{OT} = f(T, O, Z - \lambda O)$ where $(O, Z - \lambda O)$ satisfies the cumulant constraint and $(T, O, Z - \lambda O)$ satisfies the independence constraint.

Proof. Based on Thm. 2, $\lambda = m_{ZO}$ iff $(O, Z - \lambda O)$ satisfies the cumulant constraint and $(T, O, Z - \lambda O)$ satisfies the independence constraint. Based on Thm. 1, $m_{OT} = f(T, O, Z - \lambda O)$. □

Theorem 5. In Fig. 3(e), $m_{OT} = f(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ where both $(Z, T - \lambda_1 Z)$ and $(Z, O - \lambda_2 Z)$ satisfy the cumulant constraint and $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint.

Proof. In Fig. 3(e),

$$Z = m_{ZL}\epsilon_L + \epsilon_Z,$$
$$T = m_{TL}\epsilon_L + m_{TZ}\epsilon_Z + \epsilon_T,$$
$$O = m_{OL}\epsilon_L + m_{OZ}\epsilon_Z + m_{OT}\epsilon_T + \epsilon_O$$

Let ϵ_L and ϵ_Z be e and e_1 in Lem. 2, $(Z, T - \lambda_1 Z)$ satisfies the cumulant constraint iff $\lambda_1 = m_{TZ}$ or $\frac{m_{TL}}{m_{ZL}}$, $(Z, O - \lambda_2 Z)$ satisfies the cumulant constraint iff $\lambda_2 = m_{OZ}$ or $\frac{m_{OL}}{m_{ZL}}$.

We first prove $(\lambda_1, \lambda_2) = (m_{TZ}, m_{OZ})$ or $(m_{TL}/m_{ZL}, m_{OL}/m_{ZL})$ iff both $(Z, T - \lambda_1 Z)$ and $(Z, O - \lambda_2 Z)$ satisfy the cumulant constraint and $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint.

"Only if": If $(\lambda_1, \lambda_2) = (m_{TZ}, m_{OZ})$, both $(Z, T - \lambda_1 Z)$ and $(Z, O - \lambda_2 Z)$ satisfy the cumulant constraint based on Lem. 2. Besides,

$$T - \lambda_1 Z = a_{TL}\epsilon_L + \epsilon_T, \tag{9}$$
$$O - \lambda_2 Z = (m_{OL} - m_{OZ}m_{ZL})\epsilon_L + m_{OT}\epsilon_T + \epsilon_O, \tag{10}$$
$$Z = m_{ZL}\epsilon_L + \epsilon_Z. \tag{11}$$

Note that $m_{OL} - m_{OZ}m_{ZL} \neq 0$, let ϵ_L be e in Lem. 3, $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint.

If $(\lambda_1, \lambda_2) = (m_{TL}/m_{ZL}, m_{OL}/m_{ZL})$,

$$T - \lambda_1 Z = (m_{TZ} - \frac{m_{TL}}{m_{ZL}})\epsilon_Z + \epsilon_T, \tag{12}$$
$$O - \lambda_2 Z = (m_{OZ} - \frac{m_{OL}}{m_{ZL}})\epsilon_Z + m_{OT}\epsilon_T + \epsilon_O, \tag{13}$$
$$Z = m_{ZL}\epsilon_L + \epsilon_Z. \tag{14}$$

Note that $(m_{TZ} - \frac{m_{TL}}{m_{ZL}})(m_{OZ} - \frac{m_{OL}}{m_{ZL}}) \neq 0$, let ϵ_Z be e in Lem. 3, $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint.

"If": If $(\lambda_1, \lambda_2) = (m_{TZ}, m_{OL}/m_{ZL})$, based on Eqs. (9) and (13), $T - \lambda_1 Z$ contains ϵ_L while $O - \lambda_2 Z$ does not contain ϵ_L, note that $m_{ZL} \neq 0$, so $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ does not satisfy the independence constraint based on Lem. 4.

If $(\lambda_1, \lambda_2) = (m_{TL}/m_{ZL}, m_{OZ})$, based on Eqs. (12) and (10), $T - \lambda_1 Z$ does not contains ϵ_L while $O - \lambda_2 Z$ contains ϵ_L, note that $m_{ZL} \neq 0$, so $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ does not satisfy the independence constraint based on Lem. 4. Therefore, $(\lambda_1, \lambda_2) = (m_{TZ}, m_{OZ})$ or $(m_{TL}/m_{ZL}, m_{OL}/m_{ZL})$.

In the following, we prove $m_{OT} = f(T - \lambda_1 Z, O - \lambda_2 Z, Z)$. Denote $T - \lambda_1 Z, O - \lambda_2 Z$ by T', O'. If $(\lambda_1, \lambda_2) = (m_{TZ}, m_{OZ})$, based on Eqs. (9) and (10), L, Z, T', O' can be expressed as

$$\begin{aligned} L &= \epsilon_L, \\ Z &= a_{ZL} L + \epsilon_Z, \\ T' &= a_{TL} L + \epsilon_T, \\ O' &= a_{OL} L + m_{OT} T' + \epsilon_O. \end{aligned}$$

If $(m_{TL}/m_{ZL}, m_{OL}/m_{ZL})$, based on Eqs. (12) and (13), L, Z, T', O' can be expressed as

$$\begin{aligned} L &= \epsilon_Z, \\ Z &= L + a_{ZL} \epsilon_L, \\ T' &= -\frac{a_{TL}}{m_{ZL}} L + \epsilon_T, \\ O' &= -\frac{a_{OL}}{m_{ZL}} L + m_{OT} T' + \epsilon_O. \end{aligned}$$

In both cases, $Z \perp\!\!\!\perp (T', O') | L$, that is, Z is an isolated proxy. Also, it is obvious that $m_{OT} = m_{O'T'}$, so $m_{OT} = f(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ based on Thm. 1. □

Corollary 1. The underlying causal structure is Fig. 3(a) iff (T, O, Z) satisfies the independence constraint.

Proof. "Only if": In Fig. 3(a),

$$\begin{aligned} Z &= m_{ZL} \epsilon_L + \epsilon_Z, \\ T &= m_{TL} \epsilon_L + \epsilon_T, \\ O &= m_{OL} \epsilon_L + m_{OT} \epsilon_T + \epsilon_O. \end{aligned}$$

Let ϵ_L be e in Lem. 3, (T, O, Z) satisfies the independence constraint.

"If": We prove this part by contradiction.

- In Fig. 3(b) and (e), $\{L, Z\} \subseteq \mathrm{An}(T) \cap \mathrm{An}(O)$ and there exist two non-intersecting directed paths from $\{L, Z\}$ to $\{T, O\}$ ($Z \to T$ and $L \to O$), so $m_{TL}/m_{TZ} \neq m_{OL}/m_{OZ}$. Besides, $m_{ZL} m_{ZZ} \neq 0$. Based on Lem. 5, (T, O, Z) does not satisfy the independence constraint, which leads to contradiction.
- In Fig. 3(c) and (g), $m_{TZ} = 0$, $m_{OZ} \neq 0$, and $m_{ZZ} \neq 0$. Based on Lem. 4, (T, O, Z) does not satisfy the independence constraint, which leads to contradiction.
- In Fig. 3(d) and (h), $m_{TO} = 0$, $m_{OO} \neq 0$, and $m_{ZO} \neq 0$. Based on Lem. 4, (T, O, Z) does not satisfy the independence constraint, which leads to contradiction.
- In Fig. 3(f), $\{L, T\} \subseteq \mathrm{An}(T) \cap \mathrm{An}(O)$ and there exist two non-intersecting directed paths from $\{L, T\}$ to $\{T, O\}$ (T and $L \to O$), so $m_{TL}/m_{TT} \neq m_{OL}/m_{OT}$. Besides, $m_{ZL} m_{ZT} \neq 0$. Based on Lem. 5, (T, O, Z) does not satisfy the independence constraint, which leads to contradiction.

□

Corollary 2. If the underlying causal structure is not Fig. 3(a), it is Fig. 3(b) iff there exists one and only one λ s.t. the constraints in Thm. 2 are satisfied.

Proof. "Only if": This part can be proven by Thm. 2.

"If": We prove this part by contradiction. Suppose there exists λ s.t. $(T, O - \lambda T)$ satisfies the cumulant constraint and $(Z, T, O - \lambda T)$ satisfies the independence constraint.

- In Fig. 3(c) and (g), $m_{TZ} = 0$ and $m_{ZZ} \neq 0$. In particular, for any λ, $O - \lambda T$ contains ϵ_Z. Based on Lem. 4, $(Z, T, O - \lambda T)$ does not satisfy the independence constraint, which leads to contradiction.
- In Fig. 3(d) and (h), $m_{TO} = 0$ and $m_{ZO} \neq 0$. In particular, for any λ, $O - \lambda T$ contains ϵ_O. Based on Lem. 4, $(Z, T, O - \lambda T)$ does not satisfy the independence constraint, which leads to contradiction.

- In Fig. 3(e), if $\lambda \neq m_{OT}$, $O - \lambda T$ contains ϵ_T, note that $m_{ZT} = 0$ and $m_{TT} \neq 0$, so $(Z, T, O - \lambda T)$ does not satisfy the independence constraint based on Lem. 4, which leads to contradiction. Otherwise, $O - \lambda T$ contain both ϵ_L and ϵ_Z, since $\{L, Z\} \subseteq \text{An}(Z) \cap \text{An}(T)$ and there exist two non-intersecting directed paths from $\{L, Z\}$ to $\{Z, T\}$ (Z and $L \to T$), so $m_{ZL}/m_{TL} \neq m_{ZZ}/m_{TZ}$. Based on Lem. 5, $(Z, T, O - \lambda T)$ does not satisfy the independence constraint, which leads to contradiction.
- In Fig. 3(f), if $\lambda = m_{OT}$,

$$\begin{aligned} Z &= m_{ZL}\epsilon_L + m_{ZT}\epsilon_T + \epsilon_Z, \\ T &= m_{TL}\epsilon_L + \epsilon_T, \\ O - \lambda T &= a_{OL}\epsilon_L + \epsilon_O. \end{aligned}$$

  Let ϵ_L be e in Lem. 1, $(T, O - \lambda T)$ satisfies the cumulant constraint. Let ϵ_L be e in Lem. 3, $(Z, T, O - \lambda T)$ satisfies the independence constraint. If $\lambda = m_{OL}/m_{TL}$,

$$\begin{aligned} Z &= m_{ZL}\epsilon_L + m_{ZT}\epsilon_T + \epsilon_Z, \\ T &= m_{TL}\epsilon_L + \epsilon_T, \\ O - \lambda T &= (m_{OT} - \frac{m_{OL}}{m_{TL}})\epsilon_T + \epsilon_O. \end{aligned}$$

  Let ϵ_T be e in Lem. 1, $(T, O - \lambda T)$ satisfies the cumulant constraint. Let ϵ_T be e in Lem. 3, $(Z, T, O - \lambda T)$ satisfies the independence constraint. Therefore, there exist two λ s.t. $(T, O - \lambda T)$ satisfies the cumulant constraint and $(Z, T, O - \lambda T)$ satisfies the independence constraint, which leads to contradiction.

□

Corollary 3. If the underlying causal structure is not one of Fig. 3(a)–(b), it is Fig. 3(c) iff there exists λ s.t. the constraints in Thm. 3 are satisfied.

Proof. "Only if": This part can be proven by Thm. 3.

"If": We prove this part by contradiction. Suppose there exists λ s.t. $(T, O - \lambda T)$ satisfies the cumulant constraint and $(O - \lambda T, Z, T)$ satisfies the independence constraint.

- In Fig. 3(d), (f), (g), and (h), based on Lem. 2, $(T, O - \lambda T)$ satisfies the cumulant constraint iff $\lambda = m_{OT}$ or m_{OL}/m_{TL}. If $\lambda = m_{OT}$, $O - \lambda T$ does not contain ϵ_T, note that $m_{TT}m_{ZT} \neq 0$, so $(O - \lambda T, Z, T)$ does not satisfy the independence constraint based on Lem. 4, which leads to contradiction. If $\lambda = m_{OL}/m_{TL}$, $O - \lambda T$ does not contain ϵ_L, note that $m_{TL}m_{ZL} \neq 0$, so $(O - \lambda T, Z, T)$ does not satisfy the independence constraint based on Lem. 4, which leads to contradiction.
- In Fig. 3(e), if $\lambda \neq m_{OT}$, $O - \lambda T$ contains ϵ_T, note that $m_{ZT} = 0$ and $m_{TT} \neq 0$, so $(O - \lambda T, Z, T)$ does not satisfy the independence constraint based on Lem. 4, which leads to contradiction. Otherwise, $O - \lambda T = a_{OL}L + a_{OZ}Z + \epsilon_O$, denote $O - \lambda T$ by O', since $\{L, Z\} \subseteq \text{An}(Z) \cap \text{An}(O')$ and there exist two non-intersecting directed paths from $\{L, Z\}$ to $\{Z, O'\}$ (Z and $L \to O'$), so $m_{ZL}/m_{O'L} \neq m_{ZZ}/m_{O'Z}$. Besides, $m_{TL}m_{TZ} \neq 0$. Based on Lem. 5, $(O - \lambda T, Z, T)$ does not satisfy the independence constraint, which leads to contradiction.

□

Corollary 4. If the underlying causal structure is not one of Fig. 3(a)–(c), it is Fig. 3(d) iff there exists λ s.t. the constraints in Thm. 4 are satisfied.

Proof. "Only if": This part can be proven by Thm. 4.

"If": First, based on Cor. 1, there does not exist $\lambda = 0$ s.t. $(T, O, Z - \lambda O)$ satisfies the independence constraint. Next, suppose there exists $\lambda \neq 0$ s.t. $(O, Z - \lambda O)$ satisfies the cumulant constraint and $(T, O, Z - \lambda O)$ satisfies the independence constraint.

- In Fig. 3(e), (f), and (g), $m_{TO} = 0$ and $m_{OO} \neq 0$. In particular, for any $\lambda \neq 0$, $Z - \lambda O$ contains ϵ_O. Based on Lem. 4, $(T, O, Z - \lambda O)$ does not satisfy the independence constraint, which leads to contradiction.

- In Fig. 3(h), if $\lambda \neq m_{ZO}$, $Z - \lambda O$ contains ϵ_O, note that $m_{TO} = 0$ and $m_{OO} \neq 0$, so $(T, O, Z - \lambda O)$ does not satisfy the independence constraint based on Lem. 4, which leads to contradiction. Otherwise, $Z - \lambda O$ contains both ϵ_L and ϵ_T, since $\{L, T\} \subseteq \mathrm{An}(T) \cap \mathrm{An}(O)$ and there exist two non-intersecting directed paths from $\{L, T\}$ to $\{T, O\}$ (T and $L \to O$), so $m_{TL}/m_{OL} \neq m_{TT}/m_{OT}$. Based on Lem. 5, $(T, O, Z - \lambda O)$ does not satisfy the independence constraint, which leads to contradiction.

□

Corollary 5. If the underlying causal structure is not one of Fig. 3(a)–(d), it is Fig. 3(e) iff there exist λ_1 and λ_2 s.t. the constraints in Thm. 5 are satisfied.

Proof. "Only if": This part can be proven by Thm. 5.

"If": First, based on Cor. 1, there does not exist $\lambda_1 = 0$ and $\lambda_2 = 0$ s.t. $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint. Next, suppose there exists λ_1 and λ_2 s.t. $\lambda_1^2 + \lambda_2^2 \neq 0$, $(Z, T - \lambda_1 Z)$ and $(Z, O - \lambda_2 Z)$ satisfy the cumulant constraint, and $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint.

- In Fig. 3(f), Z contains $\epsilon_L, \epsilon_T, \epsilon_Z$, and

$$T - \lambda_1 Z = (m_{TL} - \lambda_1 m_{ZL})\epsilon_L + (1 - \lambda_1 m_{ZT})\epsilon_T - \lambda_1 \epsilon_Z,$$
$$O - \lambda_2 Z = (m_{OL} - \lambda_2 m_{ZL})\epsilon_L + (m_{OT} - \lambda_2 m_{ZT})\epsilon_T - \lambda_2 \epsilon_Z + \epsilon_O.$$

  Suppose $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint, note that $\lambda_1^2 + \lambda_2^2 \neq 0$, based on Lem. 4, $\lambda_1 \lambda_2 \neq 0$. Moreover, based on Lem. 5,

$$(1 - \lambda_1 m_{ZT})(-\lambda_2) = (m_{OT} - \lambda_2 m_{ZT})(-\lambda_1), \tag{15}$$
$$(m_{TL} - \lambda_1 m_{ZL})(-\lambda_2) = (m_{OL} - \lambda_2 m_{ZL})(-\lambda_1). \tag{16}$$

  From Eqs. (15) and (16), we can derive $m_{OL} = m_{OT} m_{TL}$. Note that $m_{OL} = m_{OT} m_{TL} + a_{OL}$ in Fig. 3(f), this leads to contradiction.

- In Fig. 3(g), Z contains $\epsilon_L, \epsilon_T, \epsilon_Z$, and

$$T - \lambda_1 Z = (m_{TL} - \lambda_1 m_{ZL})\epsilon_L + (1 - \lambda_1 m_{ZT})\epsilon_T - \lambda_1 \epsilon_Z,$$
$$O - \lambda_2 Z = (m_{OL} - \lambda_2 m_{ZL})\epsilon_L + (m_{OT} - \lambda_2 m_{ZT})\epsilon_T + (m_{OZ} - \lambda_2)\epsilon_Z + \epsilon_O.$$

  In addition,

$$m_{OL} - m_{OZ} m_{ZL} = a_{OL} + a_{OT} m_{TL}, \quad m_{OT} - m_{OZ} m_{ZT} = a_{OT}. \tag{17}$$

  Suppose $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint, based on Lem. 4, either $\lambda_1 = 0$ and $m_{OZ} - \lambda_2 = 0$ or $\lambda_1 \neq 0$ and $m_{OZ} - \lambda_2 \neq 0$. Moreover, based on Lem. 5,

$$(m_{TL} - \lambda_1 m_{ZL})(m_{OT} - \lambda_2 m_{ZT}) = (1 - \lambda_1 m_{ZT})(m_{OL} - \lambda_2 m_{ZL}), \tag{18}$$
$$(1 - \lambda_1 m_{ZT})(m_{OZ} - \lambda_2) = (m_{OT} - \lambda_2 m_{ZT})(-\lambda_1), \tag{19}$$
$$(m_{TL} - \lambda_1 m_{ZL})(m_{OZ} - \lambda_2) = (m_{OL} - \lambda_2 m_{ZL})(-\lambda_1). \tag{20}$$

  If $\lambda_1 = 0$ and $m_{OZ} - \lambda_2 = 0$, Eqs. (17) and (18) imply that $m_{TL} a_{OT} = a_{OL} + a_{OT} m_{TL}$, which leads to contradiction. If $\lambda_1 \neq 0$ and $m_{OZ} - \lambda_2 \neq 0$, Eqs. (19) and (20) imply that $\frac{m_{OL} - m_{OZ} m_{ZL}}{m_{OT} - m_{OZ} m_{ZT}} = m_{TL}$, which contradicts Eq. (17).

- In Fig. 3(h), Z contains $\epsilon_L, \epsilon_T, \epsilon_O, \epsilon_Z$, and

$$T - \lambda_1 Z = (m_{TL} - \lambda_1 m_{ZL})\epsilon_L + (1 - \lambda_1 m_{ZT})\epsilon_T - \lambda_1 m_{ZO}\epsilon_O - \lambda_1 \epsilon_Z,$$
$$O - \lambda_2 Z = (m_{OL} - \lambda_2 m_{ZL})\epsilon_L + (m_{OT} - \lambda_2 m_{ZT})\epsilon_T + (1 - \lambda_2 m_{ZO})\epsilon_O - \lambda_2 \epsilon_Z.$$

  Suppose $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint, note that $\lambda_1^2 + \lambda_2^2 \neq 0$, based on Lem. 4, $\lambda_1 \lambda_2 \neq 0$. Furthermore, based on Lem. 5,

$$(-\lambda_1 m_{ZO})(-\lambda_2) = (-\lambda_1)(1 - \lambda_2 m_{ZO}),$$

  that is, $\lambda_1 = 0$, which leads to contradiction.

□

C Extension 1: Multiple Latent Confounders

C.1 Theoretical Results

For ease of exposition, we denote the set of Z_i whose causal relationship with (T, O) follows Fig. 3(a), Fig. 3(b), ..., Fig. 3(h) by $\mathbf{Z}_a, \mathbf{Z}_b, ..., \mathbf{Z}_h$. Using this notation, in Fig. 4, we have $Z_1 \in \mathbf{Z}_a, Z_2 \in \mathbf{Z}_b, ..., Z_8 \in \mathbf{Z}_h$. In addition, we abbreviate $\text{An}(T) \cap \mathbf{Z}, \text{An}(O) \cap \mathbf{Z}, \text{Pa}(T) \cap \mathbf{Z}, \text{Pa}(O) \cap \mathbf{Z}$ by $\text{An}_{\mathbf{Z}}(T), \text{An}_{\mathbf{Z}}(O), \text{Pa}_{\mathbf{Z}}(T), \text{Pa}_{\mathbf{Z}}(O)$ respectively.

Thm. 8 establishes the necessary and sufficient condition for $Z_i \in \mathbf{Z}_a$.

Theorem 8. *$Z_i \in \mathbf{Z}_a$ iff (T, O, Z_i) satisfies the independence constraint.*

Proof. "Only if": if $Z_i \in \mathbf{Z}_a$,

$$T = m_{TL_i}\epsilon_{L_i} + \sum_{L \in \mathbf{L} \setminus \{L_i\}} m_{TL}\epsilon_L + \sum_{Z \in \text{An}_{\mathbf{Z}}(T)} m_{TZ}\epsilon_Z + \epsilon_T, \tag{21}$$

$$O = m_{OL_i}\epsilon_{L_i} + \sum_{L \in \mathbf{L} \setminus \{L_i\}} m_{OL}\epsilon_L + \sum_{Z \in \text{An}_{\mathbf{Z}}(O)} m_{OZ}\epsilon_Z + m_{OT}\epsilon_T + \epsilon_O, \tag{22}$$

$$Z_i = m_{Z_iL_i}\epsilon_{L_i} + \epsilon_{Z_i}. \tag{23}$$

Note that $Z_i \notin \text{An}_{\mathbf{Z}}(T) \cup \text{An}_{\mathbf{Z}}(O)$, let ϵ_{L_i} be e in Lem. 3, (T, O, Z_i) satisfies the independence constraint.

"If": We prove this part by contradiction.

- If $Z_i \in \mathbf{Z}_b \cup \mathbf{Z}_e$, $\{L_i, Z_i\} \subseteq \text{An}(T) \cap \text{An}(O)$ and there exist two non-intersecting directed paths from $\{L_i, Z_i\}$ to $\{T, O\}$ ($Z_i \to T$ and $L_i \to O$), so $m_{TL_i}/m_{TZ_i} \neq m_{OL_i}/m_{OZ_i}$. Besides, $m_{Z_iL_i}m_{Z_iZ_i} \neq 0$. Based on Lem. 5, (T, O, Z_i) does not satisfy the independence constraint, which leads to contradiction.
- If $Z_i \in \mathbf{Z}_c \cup \mathbf{Z}_g$, $m_{TZ_i} = 0$ and $m_{Z_iZ_i}m_{OZ_i} \neq 0$. Based on Lem. 4, (T, O, Z_i) does not satisfy the independence constraint, which leads to contradiction.
- If $Z_i \in \mathbf{Z}_d \cup \mathbf{Z}_h$, $m_{TO} = 0$ and $m_{OO}m_{Z_iO} \neq 0$. Based on Lem. 4, (T, O, Z_i) does not satisfy the independence constraint, which leads to contradiction.
- If $Z_i \in \mathbf{Z}_f$, $\{L_i, T\} \subseteq \text{An}(T) \cap \text{An}(O)$ and there exist two non-intersecting directed paths from $\{L_i, T\}$ to $\{T, O\}$ (T and $L_i \to O$), so $m_{TL_i}/m_{TT} \neq m_{OL_i}/m_{OT}$. Besides, $m_{Z_iL}m_{Z_iT} \neq 0$. Based on Lem. 5, (T, O, Z_i) does not satisfy the independence constraint, which leads to contradiction.

□

Thm. 9 establishes the necessary and sufficient condition for $Z_i \in \mathbf{Z}_c$. Furthermore, it gives the value of m_{OZ_i} when $Z_i \in \mathbf{Z}_c$.

Theorem 9. *Suppose $Z_i \notin \mathbf{Z}_a$, $Z_i \in \mathbf{Z}_c$ iff there exists λ s.t. $(Z_i, O - \lambda Z_i)$ satisfies the cumulant constraint and $(O - \lambda Z_i, T, Z_i)$ satisfies the independence constraint. Furthermore, if $Z_i \in \mathbf{Z}_c$, $\lambda = m_{OZ_i}$ iff the above constraints are satisfied.*

Proof. "Only if": if $Z_i \in \mathbf{Z}_c$, let $\lambda = m_{OZ_i}$, then

$$O - \lambda Z_i = (m_{OL_i} - m_{OZ_i}m_{Z_iL_i})\epsilon_{L_i} + \sum_{L \in \mathbf{L} \setminus \{L_i\}} m_{OL}\epsilon_L + \sum_{Z \in \text{An}_{\mathbf{Z}}(O) \setminus \{Z_i\}} m_{OZ}\epsilon_Z + m_{OT}\epsilon_T + \epsilon_O, \tag{24}$$

$$T = m_{TL_i}\epsilon_{L_i} + \sum_{L \in \mathbf{L} \setminus \{L_i\}} m_{TL}\epsilon_L + \sum_{Z \in \text{An}_{\mathbf{Z}}(T)} m_{TZ}\epsilon_Z + \epsilon_T, \tag{25}$$

$$Z_i = m_{Z_iL_i}\epsilon_{L_i} + \epsilon_{Z_i}. \tag{26}$$

Let ϵ_{L_i} be e in Lem. 1, $(Z_i, O - \lambda Z_i)$ satisfies the cumulant constraint. Let ϵ_{L_i} be e in Lem. 3, $(O - \lambda Z_i, T, Z_i)$ satisfies the independence constraint.

"If": We prove this part by contradiction. Suppose $Z_i \notin \mathbf{Z}_a \cup \mathbf{Z}_c$, based on Thm. 8, there does not exist $\lambda = 0$ s.t. $(Z_i, O - \lambda Z_i)$ satisfies the cumulant constraint and $(O - \lambda Z_i, T, Z_i)$ satisfies the

independence constraint. Suppose there exists $\lambda \neq 0$ s.t. $(O - \lambda Z_i, T, Z_i)$ satisfies the independence constraint.

- If $Z_i \in \mathbf{Z}_d \cup \mathbf{Z}_f \cup \mathbf{Z}_h$, $m_{TZ_i} = 0$ and $m_{Z_iZ_i} \neq 0$. In particular, for any $\lambda \neq 0$, $O - \lambda Z_i$ contains ϵ_{Z_i}. Based on Lem. 4, $(O - \lambda Z_i, T, Z_i)$ does not satisfy the independence constraint, which leads to contradiction.
- If $Z_i \in \mathbf{Z}_b \cup \mathbf{Z}_e$, based on Lem. 2, $(Z_i, O - \lambda Z_i)$ satisfies the cumulant constraint iff $\lambda = m_{OZ_i}$ or $m_{OL_i}/m_{Z_iL_i}$. If $\lambda = m_{OZ_i}$, $O - \lambda Z_i$ does not contain ϵ_{Z_i}, note that $m_{TZ_i}m_{Z_iZ_i} \neq 0$, so $(O - \lambda Z_i, T, Z_i)$ does not satisfy the independence constraint based on Lem. 4, which leads to contradiction. If $\lambda = m_{OL_i}/m_{Z_iL_i}$, $O - \lambda Z_i$ does not contain ϵ_{L_i}, note that $m_{TL_i}m_{Z_iL_i} \neq 0$, so $(O - \lambda Z_i, T, Z_i)$ does not satisfy the independence constraint based on Lem. 4, which leads to contradiction..
- If $Z_i \in \mathbf{Z}_g$, we discuss two cases. If $\lambda \neq m_{OZ_i}$, $O - \lambda Z_i$ contains ϵ_{Z_i}, note that $m_{TZ_i} = 0$ and $m_{Z_iZ_i} \neq 0$, so $(O - \lambda Z_i, T, Z_i)$ does not satisfy the independence constraint based on Lem. 4, which leads to contradiction. If $\lambda = m_{OZ_i}$, $O - \lambda Z_i = a_{OL_i}L_i + a_{OT}T + \sum_{L\in\mathbf{L}\setminus\{L_i\}} a_{OL}L + \sum_{Z\in \mathrm{Pa}_{\mathbf{Z}}(O)\setminus\{Z_i\}} a_{OZ}Z + \epsilon_O$, denote $O - \lambda Z_i$ by O', since $\{L_i, T\} \subseteq \mathrm{An}(O') \cap \mathrm{An}(T)$ and there exist two non-intersecting directed paths from $\{L_i, T\}$ to $\{T, O'\}$ (T and $L_i \to O'$), so $m_{TL_i}/m_{O'L_i} \neq m_{TT}/m_{O'T}$. Besides, $m_{Z_iL_i}m_{Z_iT} \neq 0$. Based on Lem. 5, $(O - \lambda Z_i, T, Z_i)$ does not satisfy the independence constraint, which leads to contradiction.

Furthermore, if $Z_i \in \mathbf{Z}_c$ but $\lambda \neq m_{OZ_i}$, $O - \lambda Z_i$ contains ϵ_{Z_i}, note that $m_{TZ_i} = 0$ and $m_{Z_iZ_i} \neq 0$, $(O - \lambda Z_i, T, Z_i)$ does not satisfy the independence constraint based on Lem. 4. □

Thm. 10 establishes the necessary and sufficient condition for $Z_i \in \mathbf{Z}_d$. Furthermore, it gives the value of m_{Z_iO} when $Z_i \in \mathbf{Z}_d$.

Theorem 10. *Suppose $Z_i \notin \mathbf{Z}_a \cup \mathbf{Z}_c$, $Z_i \in \mathbf{Z}_d$ iff there exists λ s.t. $(O, Z_i - \lambda O)$ satisfies the cumulant constraint and $(T, O, Z_i - \lambda O)$ satisfies the independence constraint. Furthermore, if $Z_i \in \mathbf{Z}_d$, $\lambda = m_{Z_iO}$ iff the above constraints are satisfied.*

Proof. "Only if": If $Z_i \in \mathbf{Z}_d$, let $\lambda = m_{Z_iO}$, then

$$T = m_{TL_i}\epsilon_{L_i} + \sum_{L\in\mathbf{L}\setminus\{L_i\}} m_{TL}\epsilon_L + \sum_{Z\in\mathrm{An}_{\mathbf{Z}}(T)} m_{TZ}\epsilon_Z + \epsilon_T, \tag{27}$$

$$O = m_{OL_i}\epsilon_{L_i} + \sum_{L\in\mathbf{L}\setminus\{L_i\}} m_{OL}\epsilon_L + \sum_{Z\in\mathrm{An}_{\mathbf{Z}}(O)} m_{OZ}\epsilon_Z + m_{OT}\epsilon_T + \epsilon_O, \tag{28}$$

$$Z_i - \lambda O = a_{Z_iL_i}\epsilon_{L_i} + \epsilon_{Z_i} \tag{29}$$

Let ϵ_{L_i} be e in Lem. 1, $(O, Z_i - \lambda O)$ satisfies the cumulant constraint. Let ϵ_{L_i} be e in Lem. 3, $(T, O, Z_i - \lambda O)$ satisfies the independence constraint.

"If": We prove this part by contradiction. Suppose $Z_i \notin \mathbf{Z}_a \cup \mathbf{Z}_c \cup \mathbf{Z}_d$, based on Thm. 8, there does not exist $\lambda = 0$ s.t. $(T, O, Z_i - \lambda O)$ satisfies the independence constraint. Suppose there exists $\lambda \neq 0$ s.t. $(O, Z_i - \lambda O)$ satisfies the cumulant constraint and $(T, O, Z_i - \lambda O)$ satisfies the independence constraint.

- If $Z_i \in \mathbf{Z}_b \cup \mathbf{Z}_e \cup \mathbf{Z}_f \cup \mathbf{Z}_g$, $m_{TO} = 0$ and $m_{OO} \neq 0$. In particular, for any $\lambda \neq 0$, $Z_i - \lambda O$ contains ϵ_O. Based on Lem. 4, $(T, O, Z_i - \lambda O)$ does not satisfy the independence constraint, which leads to contradiction.
- If $Z_i \in \mathbf{Z}_h$, we discuss two cases. If $\lambda \neq m_{Z_iO}$, $Z_i - \lambda O$ contains ϵ_O, note that $m_{TO} = 0$ and $m_{OO} \neq 0$, so $(T, O, Z_i - \lambda O)$ does not satisfy the independence constraint based on Lem. 4, which leads to contradiction. If $\lambda = m_{Z_iO}$, $Z_i - \lambda O = a_{Z_iT}T + a_{Z_iL_i}L_i + \epsilon_{Z_i}$, which contains both ϵ_{L_i} and ϵ_T. Since $\{L_i, T\} \subseteq \mathrm{An}(T) \cap \mathrm{An}(O)$ and there exist two non-intersecting directed paths from $\{L_i, T\}$ to $\{T, O\}$ (T and $L_i \to O$), $m_{TL_i}/m_{OL_i} \neq m_{TT}/m_{OT}$. Based on Lem. 5, $(T, O, Z_i - \lambda O)$ does not satisfy the independence constraint, which leads to contradiction.

Furthermore, if $Z_i \in \mathbf{Z}_d$ but $\lambda \neq m_{Z_iO}$, $Z_i - \lambda O$ contains ϵ_O, note that $m_{TO} = 0$ and $m_{OO} \neq 0$, $(T, O, Z_i - \lambda O)$ does not satisfy the independence constraint based on Lem. 4. □

Thm. 11 establishes the necessary and sufficient condition for $Z_i \in \mathbf{Z}_f$. Furthermore, it gives the value of m_{Z_iT} when $Z_i \in \mathbf{Z}_f$.

Theorem 11. *Suppose $Z_i \notin \mathbf{Z}_a \cup \mathbf{Z}_c \cup \mathbf{Z}_d$, $Z_i \in \mathbf{Z}_f$ iff there exists λ s.t. $(T, Z_i - \lambda T)$ satisfies the cumulant constraint and $(T, O, Z_i - \lambda T)$ satisfies the independence constraint. Furthermore, if $Z_i \in \mathbf{Z}_f$, $\lambda = m_{Z_i T}$ iff the above constraints are satisfied.*

Proof. "Only if": If $Z_i \in \mathbf{Z}_f$, let $\lambda = m_{Z_i T}$, then

$$T = m_{TL_i}\epsilon_{L_i} + \sum_{L \in \mathbf{L}\backslash\{L_i\}} m_{TL}\epsilon_L + \sum_{Z \in \mathrm{An}_{\mathbf{Z}}(T)} m_{TZ}\epsilon_Z + \epsilon_T, \tag{30}$$

$$O = m_{OL_i}\epsilon_{L_i} + \sum_{L \in \mathbf{L}\backslash\{L_i\}} m_{OL}\epsilon_L + \sum_{Z \in \mathrm{An}_{\mathbf{Z}}(O)} m_{OZ}\epsilon_Z + m_{OT}\epsilon_T + \epsilon_O, \tag{31}$$

$$Z_i - \lambda T = a_{Z_i L_i}\epsilon_{L_i} + \epsilon_{Z_i} \tag{32}$$

Let ϵ_{L_i} be e in Lem. 1, $(T, Z_i - \lambda T)$ satisfies the cumulant constraint. Let ϵ_{L_i} be e in Lem. 3, $(T, O, Z_i - \lambda T)$ satisfies the independence constraint.

"If": We prove this part by contradiction. Suppose $Z_i \notin \mathbf{Z}_a \cup \mathbf{Z}_c \cup \mathbf{Z}_d \cup \mathbf{Z}_f$, based on Thm. 8, there does not exist $\lambda = 0$ s.t. $(T, O, Z_i - \lambda T)$ satisfies the independence constraint. Suppose there exists $\lambda \neq 0$ s.t. $(T, Z_i - \lambda T)$ satisfies the cumulant constraint and $(T, O, Z_i - \lambda T)$ satisfies the independence constraint.

- If $Z_i \in \mathbf{Z}_b \cup \mathbf{Z}_e$, let $L_j \in \mathbf{L}\backslash\{L_i\}$, since $\{T, L_j\} \subseteq \mathrm{An}(T) \cap \mathrm{An}(O)$ and there exist two non-intersecting directed paths from $\{T, L_j\}$ to $\{T, O\}$ (T and $L_j \to O$), so $m_{TT}/m_{TL_j} \neq m_{OT}/m_{OL_j}$. Besides, for any $\lambda \neq 0$, $Z_i - \lambda T$ contains ϵ_T and ϵ_{L_j}, based on Lem. 5, $(T, O, Z_i - \lambda T)$ does not satisfy the independence constraint, which leads to contradiction.
- If $Z_i \in \mathbf{Z}_g$, $m_{TZ_i} = 0$ and $m_{OZ_i} \neq 0$. In particular, for any $\lambda \neq 0$, $Z_i - \lambda T$ contains ϵ_{Z_i}. Based on Lem. 4, $(T, O, Z_i - \lambda T)$ does not satisfy the independence constraint, which leads to contradiction.
- If $Z_i \in \mathbf{Z}_h$, $m_{TO} = 0$ and $m_{OO} \neq 0$. In particular, for any $\lambda \neq 0$, $Z_i - \lambda T$ contains ϵ_O. Based on Lem. 4, $(T, O, Z_i - \lambda T)$ does not satisfy the independence constraint, which leads to contradiction.

Furthermore, If $Z_i \in \mathbf{Z}_f$ but $\lambda \neq m_{Z_i T}$, $Z_i - \lambda T$ contains ϵ_T. Besides, let $L_j \in \mathbf{L}\backslash\{L_i\}$, $Z_i - \lambda T$ contains ϵ_{L_j}. Since $\{L_j, T\} \subseteq \mathrm{An}(T) \cap \mathrm{An}(O)$ and there exist two non-intersecting directed paths from $\{L_j, T\}$ to $\{T, O\}$ (T and $L_j \to O$), $m_{TL_j}/m_{OL_j} \neq m_{TT}/m_{OT}$. Based on Lem. 5, $(T, O, Z_i - \lambda T)$ does not satisfy the independence constraint. □

Thm. 12 establishes the necessary and sufficient condition for $Z_i \in \mathbf{Z}_b \cup \mathbf{Z}_e$. Furthermore, it gives the values of (m_{TZ_i}, m_{OZ_i}) and $(m_{TL_i}/m_{Z_i L_i}, m_{OL_i}/m_{Z_i L_i})$ when $Z_i \in \mathbf{Z}_b \cup \mathbf{Z}_e$.

Theorem 12. *Suppose $Z_i \notin \mathbf{Z}_a \cup \mathbf{Z}_c \cup \mathbf{Z}_d \cup \mathbf{Z}_f$, $Z_i \in \mathbf{Z}_b \cup \mathbf{Z}_e$ iff there exist λ_1, λ_2 s.t. both $(Z_i, T - \lambda_1 Z_i)$ and $(Z_i, O - \lambda_2 Z_i)$ satisfy the cumulant constraint and $(T - \lambda_1 Z_i, O - \lambda_2 Z_i, Z_i)$ satisfies the independence constraint. Furthermore, if $Z_i \in \mathbf{Z}_b \cup \mathbf{Z}_e$, $(\lambda_1, \lambda_2) = (m_{TZ_i}, m_{OZ_i})$ or $(m_{TL_i}/m_{Z_i L_i}, m_{OL_i}/m_{Z_i L_i})$ iff the above constraints are satisfied.*

Proof. "Only if": If $Z_i \in \mathbf{Z}_b \cup \mathbf{Z}_e$, let $(\lambda_1, \lambda_2) = (m_{TZ_i}, m_{OZ_i})$, then

$$T - \lambda_1 Z_i = a_{TL_i}\epsilon_{L_i} + \sum_{L \in \mathbf{L}\backslash\{L_i\}} m_{TL}\epsilon_L + \sum_{Z \in \mathrm{An}_{\mathbf{Z}}(T)\backslash\{Z_i\}} m_{TZ}\epsilon_Z + \epsilon_T, \tag{33}$$

$$O - \lambda_2 Z_i = (m_{OL_i} - m_{OZ_i} m_{Z_i L_i})\epsilon_{L_i} + \sum_{L \in \mathbf{L}\backslash\{L_i\}} m_{OL}\epsilon_L + \sum_{Z \in \mathrm{An}_{\mathbf{Z}}(O)\backslash\{Z_i\}} m_{OZ}\epsilon_Z + m_{OT}\epsilon_T + \epsilon_O, \tag{34}$$

$$Z_i = m_{Z_i L_i}\epsilon_{L_i} + \epsilon_{Z_i}. \tag{35}$$

Let ϵ_{L_i} be e in Lem. 1, both $(Z_i, T - \lambda_1 Z_i)$ and $(Z_i, O - \lambda_2 Z_i)$ satisfy the cumulant constraint. Let ϵ_{L_i} be e in Lem. 3, $(T - \lambda_1 Z_i, O - \lambda_2 Z_i, Z_i)$ satisfies the independence constraint.

"If": We prove this part by contradiction. Suppose $Z_i \notin \mathbf{Z}_a \cup \mathbf{Z}_b \cup \mathbf{Z}_c \cup \mathbf{Z}_d \cup \mathbf{Z}_e \cup \mathbf{Z}_f$, based on Thm. 9, there does not exist $\lambda_1 = 0, \lambda_2$ s.t. $(Z_i, O - \lambda_2 Z_i)$ satisfies the cumulant constraint and $(T - \lambda_1 Z_i, O - \lambda_2 Z_i, Z_i)$ satisfies the independence constraint. Suppose there exist $\lambda_1 \neq 0, \lambda_2$ s.t. both $(Z_i, T - \lambda_1 Z_i)$ and $(Z_i, O - \lambda_2 Z_i)$ satisfy the cumulant constraint and $(T - \lambda_1 Z_i, O - \lambda_2 Z_i, Z_i)$ satisfies the independence constraint.

- If $Z_i \in \mathbf{Z}_g$, Z_i contains $\epsilon_{L_i}, \epsilon_T, \epsilon_{Z_i}$, and

$$T - \lambda_1 Z_i = (m_{TL_i} - \lambda_1 m_{Z_i L_i})\epsilon_{L_i} + (1 - \lambda_1 m_{Z_i T})\epsilon_T - \lambda_1 \epsilon_{Z_i} + ...,$$
$$O - \lambda_2 Z_i = (m_{OL_i} - \lambda_2 m_{Z_i L_i})\epsilon_{L_i} + (m_{OT} - \lambda_2 m_{Z_i T})\epsilon_T + (m_{OZ_i} - \lambda_2)\epsilon_{Z_i} + ...$$

  In addition,

$$m_{OL_i} - m_{OZ_i} m_{Z_i L_i} = a_{OL_i} + (m_{OT} - m_{OZ_i} m_{Z_i T}) m_{TL_i}. \tag{36}$$

  Suppose $(T - \lambda_1 Z_i, O - \lambda_2 Z_i, Z_i)$ satisfies the independence constraint, note that $\lambda_1 \neq 0$, based on Lem. 4, $m_{OZ_i} - \lambda_2 \neq 0$. Moreover, based on Lem. 5,

$$(1 - \lambda_1 m_{Z_i T})(m_{OZ_i} - \lambda_2) = (m_{OT} - \lambda_2 m_{Z_i T})(-\lambda_1), \tag{37}$$
$$(m_{TL_i} - \lambda_1 m_{Z_i L_i})(m_{OZ_i} - \lambda_2) = (m_{OL_i} - \lambda_2 m_{Z_i L_i})(-\lambda_1). \tag{38}$$

  From Eqs. (37) and (38), we can derive $\frac{m_{OL_i} - m_{OZ_i} m_{Z_i L_i}}{m_{OT} - m_{OZ_i} m_{Z_i T}} = m_{TL_i}$, which contradicts Eq. (36).

- If $Z_i \in \mathbf{Z}_h$, Z_i contains $\epsilon_O, \epsilon_{Z_i}$, and

$$T - \lambda_1 Z_i = -\lambda_1 m_{Z_i O}\epsilon_O - \lambda_1 \epsilon_{Z_i} + ...,$$
$$O - \lambda_2 Z_i = (1 - \lambda_2 m_{Z_i O})\epsilon_O - \lambda_2 \epsilon_{Z_i} + ...$$

  Suppose $(T - \lambda_1 Z, O - \lambda_2 Z, Z)$ satisfies the independence constraint, based on Lem. 5,

$$(-\lambda_1 m_{Z_i O})(-\lambda_2) = (-\lambda_1)(1 - \lambda_2 m_{Z_i O}),$$

  that is, $\lambda_1 = 0$, this leads to contradiction.

Furthermore, if $Z_i \in \mathbf{Z}_b \cup \mathbf{Z}_e$, based on Lem. 2, $(Z_i, T - \lambda_1 Z_i)$ satisfy the cumulant constraint iff $\lambda_1 = m_{TZ_i}$ or $m_{TL_i}/m_{Z_i L_i}$, $(Z_i, O - \lambda_2 Z_i)$ satisfy the cumulant constraint iff $\lambda_2 = m_{OZ_i}$ or $m_{OL_i}/m_{Z_i L_i}$.

- If $(\lambda_1, \lambda_2) = (m_{TL_i}/m_{Z_i L_i}, m_{OL_i}/m_{Z_i L_i})$,

$$T - \lambda_1 Z_i = (m_{TZ_i} - \frac{m_{TL_i}}{m_{Z_i L_i}})\epsilon_{Z_i} + \sum_{L \in \mathbf{L} \setminus \{L_i\}} m_{TL}\epsilon_L + \sum_{Z \in \mathrm{An}_{\mathbf{Z}}(T) \setminus \{Z_i\}} m_{TZ}\epsilon_Z + \epsilon_T, \tag{39}$$

$$O - \lambda_2 Z_i = (m_{OZ_i} - \frac{m_{OL_i}}{m_{Z_i L_i}})\epsilon_{Z_i} + \sum_{L \in \mathbf{L} \setminus \{L_i\}} m_{OL}\epsilon_L + \sum_{Z \in \mathrm{An}_{\mathbf{Z}}(O) \setminus \{Z_i\}} m_{OZ}\epsilon_Z + m_{OT}\epsilon_T + \epsilon_O, \tag{40}$$

$$Z_i = m_{Z_i L_i}\epsilon_{L_i} + \epsilon_{Z_i}. \tag{41}$$

  Note that $(m_{TZ_i} - \frac{m_{TL_i}}{m_{Z_i L_i}})(m_{OZ_i} - \frac{m_{OL_i}}{m_{Z_i L_i}}) \neq 0$, let ϵ_{Z_i} be e in Lem. 3, $(T - \lambda_1 Z_i, O - \lambda_2 Z_i, Z_i)$ satisfies the independence constraint.

- If $(\lambda_1, \lambda_2) = (m_{TZ_i}, m_{OL_i}/m_{Z_i L_i})$, based on Eqs. (33) and (40), $T - \lambda_1 Z_i$ contains ϵ_{L_i} while $O - \lambda_2 Z_i$ does not contain ϵ_{L_i}, note that $m_{Z_i L_i} \neq 0$, $(T - \lambda_1 Z_i, O - \lambda_2 Z_i, Z_i)$ does not satisfy the independence constraint based on Lem. 4.

- If $(\lambda_1, \lambda_2) = (m_{TL_i}/m_{Z_i L_i}, m_{OZ_i})$, based on Eqs. (39) and (34), $T - \lambda_1 Z_i$ does not contain ϵ_{L_i} while $O - \lambda_2 Z_i$ contains ϵ_{L_i}, note that $m_{Z_i L_i} \neq 0$, so $(T - \lambda_1 Z_i, O - \lambda_2 Z_i, Z_i)$ does not satisfy the independence constraint based on Lem. 4.

□

Thm. 13 establishes the necessary and sufficient condition for $Z_i \in \mathbf{Z}_g$. Furthermore, it gives the value of $(m_{Z_i T}, m_{OZ_i})$ when $Z_i \in \mathbf{Z}_g$.

Theorem 13. *Suppose $Z_i \notin \mathbf{Z}_a \cup \mathbf{Z}_b \cup \mathbf{Z}_c \cup \mathbf{Z}_d \cup \mathbf{Z}_e \cup \mathbf{Z}_f$, $Z_i \in \mathbf{Z}_g$ iff there exist λ_1, λ_2 s.t. both $(T, Z_i - \lambda_1 T)$ and $(Z_i - \lambda_1 T, O - \lambda_2(Z_i - \lambda_1 T))$ satisfy the cumulant constraint and $(T, O - \lambda_2(Z_i - \lambda_1 T), Z_i - \lambda_1 T)$ satisfies the independence constraint. Furthermore, if $Z_i \in \mathbf{Z}_g$, $(\lambda_1, \lambda_2) = (m_{Z_i T}, m_{OZ_i})$ iff the above constraints are satisfied.*

Proof. "Only if": If $Z_i \in \mathbf{Z}_g$, let $(\lambda_1, \lambda_2) = (m_{Z_iT}, m_{OZ_i})$,

$$T = m_{TL_i}\epsilon_{L_i} + \sum_{L\in\mathbf{L}\setminus\{L_i\}} m_{TL}\epsilon_L + \sum_{Z\in \mathrm{An}_{\mathbf{Z}}(T)} m_{TZ}\epsilon_Z + \epsilon_T, \tag{42}$$

$$Z_i - \lambda_1 T = a_{Z_iL_i}\epsilon_{L_i} + \epsilon_{Z_i}, \tag{43}$$

$$O - \lambda_2(Z_i - \lambda_1 T) = (m_{OL_i} - m_{OZ_i}a_{Z_iL_i})\epsilon_{L_i} + \sum_{L\in\mathbf{L}\setminus\{L_i\}} m_{OL}\epsilon_L + \sum_{Z\in \mathrm{An}_{\mathbf{Z}}(O)\setminus\{Z_i\}} m_{OZ}\epsilon_Z + m_{OT}\epsilon_T + \epsilon_O. \tag{44}$$

Let ϵ_{L_i} be e in Lem. 1, both $(T, Z_i - \lambda_1 T)$ and $(Z_i - \lambda_1 T, O - \lambda_2(Z_i - \lambda_1 T))$ satisfy the cumulant constraint. Let ϵ_{L_i} be e in Lem. 3, $(T, O - \lambda_2(Z_i - \lambda_1 T), Z_i - \lambda_1 T)$ satisfies the independence constraint.

"If": We prove this part by contradiction. Suppose $Z_i \notin \mathbf{Z}_a \cup \mathbf{Z}_b \cup \mathbf{Z}_c \cup \mathbf{Z}_d \cup \mathbf{Z}_e \cup \mathbf{Z}_f \cup \mathbf{Z}_g$, that is, $Z_i \in \mathbf{Z}_h$, based on Thm. 9, there does not exist $\lambda_1 = 0$ and λ_2 s.t. $(Z_i - \lambda_1 T, O - \lambda_2(Z_i - \lambda_1 T))$ satisfies the cumulant constraint and $(T, O - \lambda_2(Z_i - \lambda_1 T), Z_i - \lambda_1 T)$ satisfies the independence constraint; based on Thm. 11, there does not exist λ_1 and $\lambda_2 = 0$ s.t. $(T, Z_i - \lambda_1 T)$ satisfies the cumulant constraint and $(T, O - \lambda_2(Z_i - \lambda_1 T), Z_i - \lambda_1 T)$ satisfies the independence constraint. Suppose there exist $\lambda_1 \neq 0, \lambda_2 \neq 0$ s.t. both $(T, Z_i - \lambda_1 T)$ and $(Z_i - \lambda_1 T, O - \lambda_2(Z_i - \lambda_1 T))$ satisfy the cumulant constraint and $(T, O - \lambda_2(Z_i - \lambda_1 T), Z_i - \lambda_1 T)$ satisfies the independence constraint. Because $m_{TZ_i} = 0$ while both $O - \lambda_2(Z_i - \lambda_1 T)$ and $Z_i - \lambda_1 T$ contain ϵ_{Z_i} for any $\lambda_1 \neq 0, \lambda_2 \neq 0$. Based on Lem. 4, $(T, O - \lambda_2(Z_i - \lambda_1 T), Z_i - \lambda_1 T)$ does not satisfy the independence constraint.

Furthermore, if $Z_i \in \mathbf{Z}_g$, then $\lambda_1 \neq 0$ based on Thm. 9 and $\lambda_2 \neq 0$ based on Thm. 11.

- If $\lambda_2 \neq m_{OZ_i}$, $O - \lambda_2(Z_i - \lambda_1 T)$ contains ϵ_{Z_i} for any $\lambda_1 \neq 0$. Note that $m_{TZ_i} = 0$ and $Z_i - \lambda_1 T$ contains ϵ_{Z_i} for any $\lambda_1 \neq 0$, $(T, O - \lambda_2(Z_i - \lambda_1 T), Z_i - \lambda_1 T)$ does not satisfy the independence constraint based on Lem. 4.
- If $\lambda_2 = m_{OZ_i}$ and $\lambda_1 \neq m_{Z_iT}$, because $(T, Z_i - \lambda_1 T)$ satisfies the cumulant constraint, we have $\lambda_1 = \frac{m_{Z_iL_i}}{m_{TL_i}}$ based on Lem. 2. Let $L_j \in \mathbf{L}\setminus\{L_i\}$, $Z_i - \lambda_1 T$ contains $\epsilon_T, \epsilon_{L_j}$, and

  $$T = \epsilon_T + m_{TL_j}\epsilon_{L_j} + ...,$$
  $$O - \lambda_2(Z_i - \lambda_1 T) = (m_{OT} - m_{OZ_i}m_{Z_iT} + \frac{m_{OZ_i}m_{Z_iL_i}}{m_{TL_i}})\epsilon_T + (m_{OL_j} - m_{OZ_i}m_{Z_iL_j} + \frac{m_{OZ_i}m_{Z_iL_i}m_{TL_j}}{m_{TL_i}})\epsilon_{L_j} + ...$$

  Note that $m_{OL_j} \neq m_{OT}m_{TL_j}$ and $m_{Z_iL_j} = m_{Z_iT}m_{TL_j}$, we have

  $$m_{TL_j}(m_{OT} - m_{OZ_i}m_{Z_iT} + \frac{m_{OZ_i}m_{Z_iL_i}}{m_{TL_i}}) \neq m_{OL_j} - m_{OZ_i}m_{Z_iL_j} + \frac{m_{OZ_i}m_{Z_iL_i}m_{TL_j}}{m_{TL_i}},$$

  so $(T, O - \lambda_2(Z_i - \lambda_1 T), Z_i - \lambda_1 T)$ does not satisfy the independence constraint based on Lem. 5.

□

Thm. 14 gives the value of (a_{Z_iT}, a_{Z_iO}) when $Z_i \in \mathbf{Z}_h$.

Theorem 14. *If $Z_i \in \mathbf{Z}_h$, $(\lambda_1, \lambda_2) = (a_{Z_iT}, a_{Z_iO})$ iff both $(T, Z_i - \lambda_1 T - \lambda_2 O)$ and $(O, Z_i - \lambda_1 T - \lambda_2 O)$ satisfy the cumulant constraint and $(T, O, Z_i - \lambda_1 T - \lambda_2 O)$ satisfies the independence constraint.*

Proof. "Only if": If $(\lambda_1, \lambda_2) = (a_{Z_iT}, a_{Z_iO})$,

$$T = m_{TL_i}\epsilon_{L_i} + \sum_{L\in\mathbf{L}\setminus\{L_i\}} m_{TL}\epsilon_L + \sum_{Z\in \mathrm{An}_{\mathbf{Z}}(T)} m_{TZ}\epsilon_Z + \epsilon_T, \tag{45}$$

$$O = m_{OL_i}\epsilon_{L_i} + \sum_{L\in\mathbf{L}\setminus\{L_i\}} m_{OL}\epsilon_L + \sum_{Z\in \mathrm{An}_{\mathbf{Z}}(O)} m_{OZ}\epsilon_Z + m_{OT}\epsilon_T + \epsilon_O, \tag{46}$$

$$Z_i - \lambda_1 T - \lambda_2 O = a_{Z_iL_i}\epsilon_{L_i} + \epsilon_{Z_i}. \tag{47}$$

Note that $Z_i \notin \mathrm{An}_{\mathbf{Z}}(T) \cup \mathrm{An}_{\mathbf{Z}}(O)$, let ϵ_{L_i} be e in Lem. 1, both $(T, Z_i - \lambda_1 T - \lambda_2 O)$ and $(O, Z_i - \lambda_1 T - \lambda_2 O)$ satisfy the cumulant constraint. Let ϵ_{L_i} be e in Lem. 3, $(T, O, Z_i - \lambda_1 T - \lambda_2 O)$ satisfies the independence constraint.

"If": We prove this part by contradiction. If $\lambda_2 \neq a_{Z_iO}$, $Z_i - \lambda_1 T - \lambda_2 O$ contains ϵ_O. Note that $m_{OO} \neq 0$ and $m_{TO} = 0$, $(T, O, Z_i - \lambda_1 T - \lambda_2 O)$ does not satisfy the independence constraint based on Lem. 4. Moreover, if $\lambda_2 = a_{Z_iO}$ and $\lambda_1 \neq a_{Z_iT}$, similar to the proof of Thm. 11, we can prove $(T, O, Z_i - \lambda_1 T - \lambda_2 O)$ does not satisfy the independence constraint. Therefore, (λ_1, λ_2) has a unique value (a_{Z_iT}, a_{Z_iO}). □

C.2 Identification Algorithm

Let

$$f(T, O, \mathbf{Z}) := \frac{\mathrm{Cov}(T, O) - \sum_{Z_i \in \mathbf{Z}} g(T, Z_i)\mathrm{Cov}(O, Z_i)}{\mathrm{Var}(T) - \sum_{Z_i \in \mathbf{Z}} g(T, Z_i)\mathrm{Cov}(T, Z_i)}, \tag{48}$$

our identification algorithm is summarized in Alg. 2.

Algorithm 2: Causal effect identification (multiple latent confounders).

Input: Joint distribution of $(T, O, \mathbf{Z})$
Output: m_{OT}

1 $\mathbf{X}_a, \mathbf{X}_{be}, ..., \mathbf{X}_h := \emptyset, \emptyset, ..., \emptyset$.
2 **for** $Z_i \in \mathbf{Z}$ **do**
3 **if** *the constraint in Thm. 8 is satisfied* **then**
4 $\mathbf{X}_a := \mathbf{X}_a \cup \{Z_i\}$.
5 **else if** *there exists λ s.t. the constraints in Thm. 9 are satisfied* **then**
6 $\mathbf{X}_c := \mathbf{X}_c \cup \{(Z_i, \lambda)\}$.
7 **else if** *there exists λ s.t. the constraints in Thm. 10 are satisfied* **then**
8 $\mathbf{X}_d := \mathbf{X}_d \cup \{(Z_i, \lambda)\}$.
9 **else if** *there exists λ s.t. the constraints in Thm. 11 are satisfied* **then**
10 $\mathbf{X}_f := \mathbf{X}_f \cup \{(Z_i, \lambda)\}$.
11 **else if** *there exist λ_1, λ_2 s.t. the constraints in Thm. 12 are satisfied* **then**
12 $\mathbf{X}_{be} := \mathbf{X}_{be} \cup \{(Z_i, \lambda_1, \lambda_2)\}$.
13 **else if** *there exist λ_1, λ_2 s.t. the constraints in Thm. 13 are satisfied* **then**
14 $\mathbf{X}_g := \mathbf{X}_g \cup \{(Z_i, \lambda_1, \lambda_2)\}$.
15 **else**
16 $\mathbf{X}_h := \mathbf{X}_h \cup \{(Z_i, \lambda_1, \lambda_2)\}$ where λ_1, λ_2 satisfy Thm. 14.
17 **for** $(Z_i, \lambda) \in \mathbf{X}_d$ **do**
18 $Z_i := Z_i - \lambda O$.
19 **for** $(Z_i, \lambda) \in \mathbf{X}_f$ **do**
20 $Z_i := Z_i - \lambda T$.
21 **for** $(Z_i, \lambda_1, \lambda_2) \in \mathbf{X}_h$ **do**
22 $Z_i := Z_i - \lambda_1 T - \lambda_2 O$.
23 **for** $(Z_i, \lambda_1, \lambda_2) \in \mathbf{X}_g$ **do**
24 $O, Z_i := O - \lambda_2(Z_i - \lambda_1 T), Z_i - \lambda_1 T$.
25 **for** $(Z_i, \lambda_1, \lambda_2) \in \mathbf{X}_{be}$ **do**
26 $T, O := T - \lambda_1 Z_i, O - \lambda_2 Z_i$.
27 **for** $(Z_i, \lambda) \in \mathbf{X}_c$ **do**
28 $O := O - \lambda Z_i$.
29 **return** $f(T, O, \mathbf{Z})$

Theorem 15. *If Asmps. 1, 2, and 3 hold, Alg. 2 correctly identifies the true causal effect.*

Proof. Let T', O', Z_i' denote the variables T, O, Z_i after the update depicted in line 17–28 of Alg. 2. Define $L_i' = \epsilon_{Z_i}$ if $(Z_i, \lambda_1, \lambda_2) \in \mathbf{X}_{be}$ and $(\lambda_1, \lambda_2) = (m_{TL_i}/m_{Z_iL_i}, m_{OL_i}/m_{Z_iL_i})$, and define $L_i' = \epsilon_{L_i}$ otherwise. According to Eqs. (21)–(23), (24)–(26), (27)–(29), (30)–(32), (33)–(35), (39)–(41), (42)–(44), and (45)–(47), $Z_i' \perp\!\!\!\perp (T', O') | L_i'$ for each Z_i' and $m_{OT} = m_{O'T'}$, so $m_{OT} = f(T', O', \mathbf{Z}')$ based on Thm. 2 in [12]. □

D Extension 2: Generalized Agnostic Proxy

D.1 Theoretical Results

Assumption 4. For every $\{X_i, X_j\} \subseteq \{Z, T, O\}$, if $\text{An}(X_i) \cap \text{An}(X_j) \neq \emptyset$, then $\text{Cov}(X_i, X_j) \neq 0$.

Remark 6. Note that $\text{Cov}(X_i, X_j) = \sum_{V \in \text{An}(X_i) \cap \text{An}(X_j)} m_{X_i V} m_{X_j V} \text{Var}(\epsilon_V)$, which can be regarded as a polynomial in $\{a_{V_i V_j} | V_j \in \text{Pa}(V_i)\}$ with $\text{Var}(\epsilon_V)$ entering the coefficients. Trivially, for every $\{X_i, X_j\} \subseteq \{Z, T, O\}$, if $\text{An}(X_i) \cap \text{An}(X_j) \neq \emptyset$, this polynomial is not identically zero. Following Remark 1, Assumption 4 also holds generically.

Thms. 16 and 17 demonstrate that if the underlying causal structure is one of Fig. 5(a)–(l) and it is unknown which specific one the underlying causal structure is, m_{OT} is unidentifiable.

Theorem 16. *For every SCM $\mathcal{M}$ whose causal structure is Fig. 5(i), there exists a SCM $\tilde{\mathcal{M}}$ whose causal structure is Fig. 5(b) such that these two SCMs entail the same observational distribution but $m_{OT} \neq m_{\tilde{O}\tilde{T}}$, and vice versa.*

Proof. Let

$$\mathcal{M} := \begin{cases} Z = \epsilon_Z, \\ L = a_{LZ} Z + \epsilon_L, \\ T = a_{TL} L + \epsilon_T, \\ O = a_{OT} T + a_{OL} L + \epsilon_O, \end{cases} \quad , \quad \tilde{\mathcal{M}} := \begin{cases} \tilde{Z} = \epsilon_Z, \\ \tilde{L} = \frac{1}{a_{\tilde{T}\tilde{L}}} \epsilon_T, \\ \tilde{T} = a_{TL} a_{LZ} \tilde{Z} + a_{\tilde{T}\tilde{L}} \tilde{L} + a_{TL} \epsilon_L, \\ \tilde{O} = \frac{a_{OT} a_{TL} + a_{OL}}{a_{TL}} \tilde{T} - \frac{a_{OL} a_{\tilde{T}\tilde{L}}}{a_{TL}} \tilde{L} + \epsilon_O, \end{cases} .$$

$T = \tilde{T} = a_{TL} a_{LZ} \epsilon_Z + a_{TL} \epsilon_L + \epsilon_T$, $O = \tilde{O} = (a_{OT} a_{TL} a_{LZ} + a_{OL} a_{LZ}) \epsilon_Z + (a_{OT} a_{TL} + a_{OL}) \epsilon_L + a_{OT} \epsilon_T + \epsilon_O$, $Z = \tilde{Z} = \epsilon_Z$, that is, $\mathcal{M}$ and $\tilde{\mathcal{M}}$ entail the same observational distribution. However, $m_{OT} = a_{OT}$ while $m_{\tilde{O}\tilde{T}} = \frac{a_{OT} a_{TL} + a_{OL}}{a_{TL}}$.

Let

$$\mathcal{M} := \begin{cases} Z = \epsilon_Z, \\ L = \epsilon_L, \\ T = a_{TZ} Z + a_{TL} L + \epsilon_T, \\ O = a_{OT} T + a_{OL} L + \epsilon_O, \end{cases} \quad , \quad \tilde{\mathcal{M}} := \begin{cases} \tilde{Z} = \epsilon_Z, \\ \tilde{L} = \frac{a_{TZ}}{a_{\tilde{T}\tilde{L}}} \tilde{Z} + \frac{1}{a_{\tilde{T}\tilde{L}}} \epsilon_T, \\ \tilde{T} = a_{\tilde{T}\tilde{L}} \tilde{L} + a_{TL} \epsilon_L, \\ \tilde{O} = \frac{a_{OT} a_{TL} + a_{OL}}{a_{TL}} \tilde{T} - \frac{a_{OL} a_{\tilde{T}\tilde{L}}}{a_{TL}} \tilde{L} + \epsilon_O, \end{cases} .$$

$T = \tilde{T} = a_{TZ} \epsilon_Z + a_{TL} \epsilon_L + \epsilon_T$, $O = \tilde{O} = a_{OT} a_{TZ} \epsilon_Z + (a_{OT} a_{TL} + a_{OL}) \epsilon_L + a_{OT} \epsilon_T + \epsilon_O$, $Z = \tilde{Z} = \epsilon_Z$, that is, $\mathcal{M}$ and $\tilde{\mathcal{M}}$ entail the same observational distribution. However, $m_{OT} = a_{OT}$ while $m_{\tilde{O}\tilde{T}} = \frac{a_{OT} a_{TL} + a_{OL}}{a_{TL}}$. □

Theorem 17. *In Fig. 5(a), (c)–(h), and (j-l), m_{OT} is unidentifiable.*

Proof. We construct two SCMs $\mathcal{M}$ and $\tilde{\mathcal{M}}$ that share the same causal structure shown as Fig. 5(a):

$$\mathcal{M} := \begin{cases} L = \epsilon_L, \\ Z = \epsilon_Z, \\ T = a_{TL} L + \epsilon_T, \\ O = a_{OT} T + a_{OL} L + \epsilon_O, \end{cases} \quad , \quad \tilde{\mathcal{M}} := \begin{cases} \tilde{L} = \frac{1}{a_{TL}} \epsilon_T, \\ \tilde{Z} = \epsilon_Z, \\ \tilde{T} = a_{\tilde{T}\tilde{L}} \tilde{L} + a_{TL} \epsilon_L, \\ \tilde{O} = \frac{a_{OT} a_{TL} + a_{OL}}{a_{TL}} \tilde{T} - \frac{a_{OL} a_{\tilde{T}\tilde{L}}}{a_{TL}} \tilde{L} + \epsilon_O, \end{cases} . \tag{49}$$

$T = \tilde{T} = a_{TL} \epsilon_L + \epsilon_T$, $O = \tilde{O} = (a_{OT} a_{TL} + a_{OL}) \epsilon_L + a_{OT} \epsilon_T + \epsilon_O$, $Z = \tilde{Z} = \epsilon_Z$, that is, $\mathcal{M}$ and $\tilde{\mathcal{M}}$ entail the same observational distribution. However, $m_{OT} = a_{OT}$ while $m_{\tilde{O}\tilde{T}} = \frac{a_{OT} a_{TL} + a_{OL}}{a_{TL}}$. Analogously, we can also construct such two SCMs for Fig. 5(c)–(d) and (f)–(h).

- For Fig. 5(c), we add $a_{OZ} Z$ to O, add $a_{OZ} \tilde{Z}$ to $\tilde{O}$, and keep all other assignments fixed in Eq. (49).
- For Fig. 5(d), we add $a_{ZO} O$ to Z, add $a_{ZO} \tilde{O}$ to $\tilde{Z}$, and keep all other assignments fixed in Eq. (49).
- For Fig. 5(f), we add $a_{ZT} T$ to Z, add $a_{ZT} \tilde{T}$ to $\tilde{Z}$, and keep all other assignments fixed in Eq. (49).

- For Fig. 5(g), we add $a_{ZT}T, a_{OZ}Z$ to Z, O respectively, add $a_{ZT}\tilde{T}, a_{OZ}\tilde{Z}$ to $\tilde{Z}, \tilde{O}$ respectively, and keep all other assignments fixed in Eq. (49).
- For Fig. 5(h), we add $a_{ZT}T + a_{ZO}O$ to Z, add $a_{ZT}\tilde{T} + a_{ZO}\tilde{O}$ to $\tilde{Z}$, and keep all other assignments fixed in Eq. (49).

We construct two SCMs $\mathcal{M}$ and $\tilde{\mathcal{M}}$ that share the same causal structure shown as Fig. 5(e):

$$\mathcal{M} := \begin{cases} L = \epsilon_L, \\ Z = \epsilon_Z, \\ T = a_{TZ}Z + a_{TL}L + \epsilon_T, \\ O = a_{OZ}Z + a_{OT}T + a_{OL}L + \epsilon_O, \end{cases},$$

$$\tilde{\mathcal{M}} := \begin{cases} \tilde{L} = \frac{1}{a_{\tilde{T}\tilde{L}}}\epsilon_T, \\ \tilde{Z} = \epsilon_Z, \\ \tilde{T} = a_{TZ}\tilde{Z} + a_{\tilde{T}\tilde{L}}\tilde{L} + a_{TL}\epsilon_L, \\ \tilde{O} = \frac{a_{OZ}a_{TL} - a_{TZ}a_{OL}}{a_{TL}}\tilde{Z} + \frac{a_{OT}a_{TL} + a_{OL}}{a_{TL}}\tilde{T} - \frac{a_{OL}a_{\tilde{T}\tilde{L}}}{a_{TL}}\tilde{L} + \epsilon_O, \end{cases}.$$

$T = \tilde{T} = a_{TL}\epsilon_L + a_{TZ}\epsilon_Z + \epsilon_T$, $O = \tilde{O} = (a_{OT}a_{TL} + a_{OL})\epsilon_L + (a_{OT}a_{TZ} + a_{OZ})\epsilon_Z + a_{OT}\epsilon_T + \epsilon_O$, $Z = \tilde{Z} = \epsilon_Z$, that is, $\mathcal{M}$ and $\tilde{\mathcal{M}}$ entail the same observational distribution. However, $m_{OT} = a_{OT}$ while $m_{\tilde{O}\tilde{T}} = \frac{a_{OT}a_{TL} + a_{OL}}{a_{TL}}$.

We construct two SCMs $\mathcal{M}$ and $\tilde{\mathcal{M}}$ that share the same causal structure shown as Fig. 5(j):

$$\mathcal{M} := \begin{cases} Z = \epsilon_Z, \\ L = a_{LZ}Z + \epsilon_L, \\ T = a_{TZ}Z + a_{TL}L + \epsilon_T, \\ O = a_{OT}T + a_{OL}L + \epsilon_O, \end{cases}, \quad \tilde{\mathcal{M}} := \begin{cases} \tilde{Z} = \epsilon_Z, \\ \tilde{L} = \frac{a_{TZ}}{a_{\tilde{T}\tilde{L}}}\tilde{Z} + \frac{1}{a_{\tilde{T}\tilde{L}}}\epsilon_T, \\ \tilde{T} = a_{TL}a_{LZ}\tilde{Z} + a_{\tilde{T}\tilde{L}}\tilde{L} + a_{TL}\epsilon_L, \\ \tilde{O} = \frac{a_{OT}a_{TL} + a_{OL}}{a_{TL}}\tilde{T} - \frac{a_{OL}a_{\tilde{T}\tilde{L}}}{a_{TL}}\tilde{L} + \epsilon_O, \end{cases}.$$

$T = \tilde{T} = (a_{TL}a_{LZ} + a_{TZ})\epsilon_Z + a_{TL}\epsilon_L + \epsilon_T$, $O = \tilde{O} = (a_{OT}a_{TL}a_{LZ} + a_{OT}a_{TZ} + a_{OL}a_{LZ})\epsilon_Z + (a_{OT}a_{TL} + a_{OL})\epsilon_L + a_{OT}\epsilon_T + \epsilon_O$, $Z = \tilde{Z} = \epsilon_Z$, that is, $\mathcal{M}$ and $\tilde{\mathcal{M}}$ entail the same observational distribution. However, $m_{OT} = a_{OT}$ while $m_{\tilde{O}\tilde{T}} = \frac{a_{OT}a_{TL} + a_{OL}}{a_{TL}}$. Analogously, we can also construct such two SCMs for Fig. 5(l). Specifically, we add $a_{OZ}Z$ to O, add $a_{OZ}\tilde{Z}$ to $\tilde{O}$, and keep all other assignments fixed.

We construct two SCMs $\mathcal{M}$ and $\tilde{\mathcal{M}}$ that share the same causal structure shown as Fig. 5(k):

$$\mathcal{M} := \begin{cases} Z = \epsilon_Z, \\ L = a_{LZ}Z + \epsilon_L, \\ T = a_{TL}L + \epsilon_T, \\ O = a_{OZ}Z + a_{OT}T + a_{OL}L + \epsilon_O, \end{cases},$$

$$\tilde{\mathcal{M}} := \begin{cases} \tilde{Z} = \epsilon_Z, \\ \tilde{L} = \frac{a_{TL}a_{LZ}}{a_{\tilde{T}\tilde{L}}}\tilde{Z} + \frac{1}{a_{\tilde{T}\tilde{L}}}\epsilon_T, \\ \tilde{T} = a_{\tilde{T}\tilde{L}}\tilde{L} + a_{TL}\epsilon_L, \\ \tilde{O} = \frac{a_{OT}a_{TL} + a_{OL}}{a_{TL}}\tilde{T} - \frac{a_{OL}a_{\tilde{T}\tilde{L}}}{a_{TL}}\tilde{L} + (a_{OL}a_{LZ} + a_{OZ})\tilde{Z} + \epsilon_O, \end{cases}.$$

$T = \tilde{T} = a_{TL}a_{LZ}\epsilon_Z + a_{TL}\epsilon_L + \epsilon_T$, $O = \tilde{O} = (a_{OT}a_{TL}a_{LZ} + a_{OL}a_{LZ} + a_{OZ})\epsilon_Z + (a_{OT}a_{TL} + a_{OL})\epsilon_L + a_{OT}\epsilon_T + \epsilon_O$, $Z = \tilde{Z} = \epsilon_Z$, that is, $\mathcal{M}$ and $\tilde{\mathcal{M}}$ entail the same observational distribution. However, $m_{OT} = a_{OT}$ while $m_{\tilde{O}\tilde{T}} = \frac{a_{OT}a_{TL} + a_{OL}}{a_{TL}}$. □

Thms. 18–21 collectively establish the necessary and sufficient condition for the underlying causal structure to be one of Fig. 5(a)–(l).

Theorem 18. *The ground truth is one of Fig. 5(a)–(c), (e), and (i)–(l) iff* $T - \frac{\mathrm{Cov}(T,Z)}{\mathrm{Var}(Z)}Z \perp\!\!\!\perp Z$ *and* $O - \frac{\mathrm{Cov}(O,Z)}{\mathrm{Var}(Z)}Z \perp\!\!\!\perp Z$.

Proof. "Only if": In Fig. 5(a), $T - \frac{\mathrm{Cov}(T,Z)}{\mathrm{Var}(Z)}Z = T \perp\!\!\!\perp Z$ and $O - \frac{\mathrm{Cov}(O,Z)}{\mathrm{Var}(Z)}Z = O \perp\!\!\!\perp Z$.

In Fig. 5(b)–(c), (e), and (i)–(l), Z, T, O can be expressed as

$$Z = \epsilon_Z, \quad T = m_{TZ}\epsilon_Z + e_T, \quad O = m_{OZ}\epsilon_Z + e_O,$$

where $e_T \perp\!\!\!\perp \epsilon_Z$ and $e_O \perp\!\!\!\perp \epsilon_Z$. So $T - \frac{\mathrm{Cov}(T,Z)}{\mathrm{Var}(Z)}Z = e_T \perp\!\!\!\perp Z$ and $O - \frac{\mathrm{Cov}(O,Z)}{\mathrm{Var}(Z)}Z = e_O \perp\!\!\!\perp Z$.

"If": We prove this part by contradiction.

- In Fig. 5(d), (f)–(h), (m), (o)–(p), and (r)–(t), $m_{TZ} = 0$, $m_{ZZ} \neq 0$, and $\mathrm{Cov}(T, Z) \neq 0$, so both $T - \frac{\mathrm{Cov}(T,Z)}{\mathrm{Var}(Z)}Z$ and Z contains ϵ_Z, thus $T - \frac{\mathrm{Cov}(T,Z)}{\mathrm{Var}(Z)}Z \not\perp\!\!\!\perp Z$.
- In Fig. 5(n) and (q), $\{L, Z\} \subseteq \mathrm{An}(Z) \cap \mathrm{An}(T)$ and there exist two non-intersecting paths from $\{L, Z\}$ to $\{Z, T\}$ (e.g., Z and $L \to T$), so $m_{ZL}/m_{ZZ} \neq m_{TL}/m_{TZ}$, that is, $T - \frac{\mathrm{Cov}(T,Z)}{\mathrm{Var}(Z)}Z$ contains at least one of ϵ_L and ϵ_Z. Note that Z contains both ϵ_L and ϵ_Z, so $T - \frac{\mathrm{Cov}(T,Z)}{\mathrm{Var}(Z)}Z \not\perp\!\!\!\perp Z$.

□

Theorem 19. *Suppose the ground truth is not one of Fig. 5(a)–(c), (e), and (i)–(l), it is Fig. 5(f) or (g) iff* $Z - \frac{\mathrm{Cov}(Z,T)}{\mathrm{Var}(T)}T \perp\!\!\!\perp T$.

Proof. "Only if": In Fig. 5(f) and (g), $Z = a_{ZT}T + \epsilon_Z$ where $T \perp\!\!\!\perp \epsilon_Z$, $Z - \frac{\mathrm{Cov}(Z,T)}{\mathrm{Var}(T)}T = \epsilon_Z \perp\!\!\!\perp T$.

"If": We prove this part by contradiction.

- In Fig. 5(d), (h), (p), and (r)–(t), $\{L, T\} \subseteq \mathrm{An}(Z) \cap \mathrm{An}(T)$ and there exist two non-intersecting paths from $\{L, T\}$ to $\{Z, T\}$ (e.g., T and $L \to O \to Z$ in Fig. 5(d) and (h), T and $L \to Z$ in Fig. 5(p) and (r)–(t)), so $m_{ZL}/m_{ZT} \neq m_{TL}/m_{TT}$, that is, $Z - \frac{\mathrm{Cov}(Z,T)}{\mathrm{Var}(T)}T$ contains at least one of ϵ_L and ϵ_T. Note that T contains both ϵ_L and ϵ_T, so $Z - \frac{\mathrm{Cov}(Z,T)}{\mathrm{Var}(T)}T \not\perp\!\!\!\perp T$.
- In Fig. 5(m)–(o) and (q), $m_{ZT} = 0$, $m_{TT} \neq 0$, and $\mathrm{Cov}(Z, T) \neq 0$, so both $Z - \frac{\mathrm{Cov}(Z,T)}{\mathrm{Var}(T)}T$ and T contains ϵ_T, thus $Z - \frac{\mathrm{Cov}(Z,T)}{\mathrm{Var}(T)}T \not\perp\!\!\!\perp T$.

□

Theorem 20. *Suppose the ground truth is not one of Fig. 5(a)–(c), (e)–(g), and (i)–(l), it is Fig. 5(d) iff* $Z - \frac{\mathrm{Cov}(Z,O)}{\mathrm{Var}(O)}O \perp\!\!\!\perp O$.

Proof. "Only if": In Fig. 5(d), $Z = a_{ZO}O + \epsilon_Z$ where $O \perp\!\!\!\perp \epsilon_Z$, $Z - \frac{\mathrm{Cov}(Z,O)}{\mathrm{Var}(O)}O = \epsilon_Z \perp\!\!\!\perp O$.

"If": We prove this part by contradiction.

- In Fig. 5(h), (p), and (t), $\{L, O\} \subseteq \mathrm{An}(Z) \cap \mathrm{An}(O)$ and there exist two non-intersecting paths from $\{L, O\}$ to $\{Z, O\}$ (e.g., O and $L \to T \to Z$ in Fig. 5(h), O and $L \to Z$ in Fig. 5(p) and (t)), so $m_{ZL}/m_{ZO} \neq m_{OL}/m_{OO}$, that is, $Z - \frac{\mathrm{Cov}(Z,O)}{\mathrm{Var}(O)}O$ contains at least one of ϵ_L and ϵ_O. Note that O contains both ϵ_L and ϵ_O, so $Z - \frac{\mathrm{Cov}(Z,O)}{\mathrm{Var}(O)}O \not\perp\!\!\!\perp O$.
- In Fig. 5(m)–(o) and (q)–(s), $m_{ZO} = 0$, $m_{OO} \neq 0$, and $\mathrm{Cov}(Z, O) \neq 0$, so both $Z - \frac{\mathrm{Cov}(Z,O)}{\mathrm{Var}(O)}O$ and O contains ϵ_O, thus $Z - \frac{\mathrm{Cov}(Z,O)}{\mathrm{Var}(O)}O \not\perp\!\!\!\perp O$.

□

Theorem 21. *Suppose the ground truth is not one of Fig. 5(a)–(g) and (i)–(l), it is Fig. 5(h) iff*

$$Z - \begin{bmatrix} T & O \end{bmatrix} \begin{bmatrix} \mathrm{Var}(T) & \mathrm{Cov}(T,O) \\ \mathrm{Cov}(T,O) & \mathrm{Var}(O) \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{Cov}(T,Z) \\ \mathrm{Cov}(O,Z) \end{bmatrix} \perp\!\!\!\perp \{T, O\}$$

Proof. "Only if": In Fig. 5(h), $Z = a_{ZT}T + z_{ZO}O + \epsilon_Z$, there is

$$\begin{bmatrix} \mathrm{Cov}(T,Z) \\ \mathrm{Cov}(O,Z) \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(T) & \mathrm{Cov}(T,O) \\ \mathrm{Cov}(T,O) & \mathrm{Var}(O) \end{bmatrix} \begin{bmatrix} a_{ZT} \\ a_{ZO} \end{bmatrix},$$

so

$$Z - \begin{bmatrix} T & O \end{bmatrix} \begin{bmatrix} \mathrm{Var}(T) & \mathrm{Cov}(T,O) \\ \mathrm{Cov}(T,O) & \mathrm{Var}(O) \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{Cov}(T,Z) \\ \mathrm{Cov}(O,Z) \end{bmatrix} = \epsilon_Z \perp\!\!\!\perp \{T, O\}.$$

"If": We prove this part by contradiction. Specifically, we prove for any λ_1, λ_2, $Z - \lambda_1 T - \lambda_2 O \not\perp\!\!\!\perp \{T, O\}$.

- In Fig. 5(m)–(o) and (q), if $\lambda_2 \neq 0$ or $\lambda_1 \neq \frac{m_{ZL}}{m_{TL}}$, $Z - \lambda_1 T - \lambda_2 O$ contains ϵ_O or ϵ_L, so $Z - \lambda_1 T - \lambda_2 O \not\perp\!\!\!\perp O$. If $\lambda_2 = 0$ and $\lambda_1 = \frac{m_{ZL}}{m_{TL}} \neq 0$, note that $m_{ZT} = 0$ and $m_{TT} \neq 0$, $Z - \lambda_1 T - \lambda_2 O$ contains ϵ_T, so $Z - \lambda_1 T - \lambda_2 O \not\perp\!\!\!\perp T$.
- In Fig. 5(r) and (s), if $\lambda_2 \neq 0$ or $\lambda_1 \neq \frac{m_{ZL}}{m_{TL}}$, $Z - \lambda_1 T - \lambda_2 O$ contains ϵ_O or ϵ_L, so $Z - \lambda_1 T - \lambda_2 O \not\perp\!\!\!\perp O$. If $\lambda_2 = 0$ and $\lambda_1 = \frac{m_{ZL}}{m_{TL}}$, note that

$$\begin{aligned} Z &= (a_{ZT} a_{TL} + a_{ZL})\epsilon_L + a_{ZT}\epsilon_T + \epsilon_Z, \\ T &= a_{TL}\epsilon_L + \epsilon_T, \end{aligned}$$

  then $Z - \lambda_1 T - \lambda_2 O = -\frac{a_{ZL}}{a_{TL}}\epsilon_T + \epsilon_Z$ contains ϵ_T, so $Z - \lambda_1 T - \lambda_2 O \not\perp\!\!\!\perp T$.
- In Fig. 5(p), if $\lambda_2 \neq m_{ZO}$ or $\lambda_1 \neq \frac{m_{ZL} - m_{ZO} m_{OL}}{m_{TL}}$, $Z - \lambda_1 T - \lambda_2 O$ contains ϵ_O or ϵ_L, so $Z - \lambda_1 T - \lambda_2 O \not\perp\!\!\!\perp O$. If $\lambda_2 = m_{ZO}$ and $\lambda_1 = \frac{m_{ZL} - m_{ZO} m_{OL}}{m_{TL}}$, note that

$$\begin{aligned} T &= a_{TL}\epsilon_L + \epsilon_T, \\ O &= (a_{OL} + a_{OT} a_{TL})\epsilon_L + a_{OT}\epsilon_T + \epsilon_O, \\ Z &= (a_{ZL} + a_{ZO} a_{OL} + a_{ZO} a_{OT} a_{TL})\epsilon_L + a_{ZO} a_{OT}\epsilon_T + a_{ZO}\epsilon_O + \epsilon_Z, \end{aligned}$$

  then $Z - \lambda_1 T - \lambda_2 O = \epsilon_Z - \frac{a_{ZL}}{a_{TL}}\epsilon_T$ contains ϵ_T, so $Z - \lambda_1 T - \lambda_2 O \not\perp\!\!\!\perp T$.
- In Fig. 5(t), if $\lambda_2 \neq m_{ZO}$ or $\lambda_1 \neq \frac{m_{ZL} - m_{ZO} m_{OL}}{m_{TL}}$, $Z - \lambda_1 T - \lambda_2 O$ contains ϵ_O or ϵ_L, so $Z - \lambda_1 T - \lambda_2 O \not\perp\!\!\!\perp O$. If $\lambda_2 = m_{ZO}$ and $\lambda_1 = \frac{m_{ZL} - m_{ZO} m_{OL}}{m_{TL}}$, note that

$$\begin{aligned} T &= a_{TL}\epsilon_L + \epsilon_T, \\ O &= (a_{OL} + a_{OT} a_{TL})\epsilon_L + a_{OT}\epsilon_T + \epsilon_O, \\ Z &= (a_{ZL} + a_{ZO} a_{OL} + a_{ZT} a_{TL} + a_{ZO} a_{OT} a_{TL})\epsilon_L + (a_{ZO} a_{OT} + a_{ZT})\epsilon_T + a_{ZO}\epsilon_O + \epsilon_Z, \end{aligned}$$

  then $Z - \lambda_1 T - \lambda_2 O = \epsilon_Z - \frac{a_{ZL}}{a_{TL}}\epsilon_T$ contains ϵ_T, so $Z - \lambda_1 T - \lambda_2 O \not\perp\!\!\!\perp T$.

□

D.2 Identification Algorithm

Our identification algorithm is summarized in Alg. 3.

Algorithm 3: Causal effect identification (generalized agnostic proxy).

Input: Joint distribution of (T, O, Z)
Output: m_{OT}

if *the condition in Thm. 18 is satisfied* **then**
  raise Error("m_{OT} is unidentifiable.")
else if *the condition in Thm. 19 is satisfied* **then**
  raise Error("m_{OT} is unidentifiable.")
else if *the condition in Thm. 20 is satisfied* **then**
  raise Error("m_{OT} is unidentifiable.")
else if *the condition in Thm. 21 is satisfied* **then**
  raise Error("m_{OT} is unidentifiable.")
else
  Run Alg. 1

Theorem 22. *If Asmps. 1, 2, 3, and 4 hold, Alg. 3 correctly identifies the true causal effect when identifiable, and correctly reports unidentifiability otherwise.*

E Causal Effect Estimation

E.1 Real-world Data

We apply our estimation method to a real-world mouse obesity dataset[3] [63], which comprises mouse weight (as the outcome), 17 gene expressions that may influence mouse weight, and 5 potential instrumental variables. Following [64, 14], we assume the data generation adheres to a linear causal model and there is one latent confounder. Treating each gene expression as the treatment in turn, we use each remaining gene expression as an agnostic proxy to obtain an estimate of the causal effect, and use the median of all such estimates as the final result. In particular, we do not require 5 potential instrumental variables as auxiliary information. The majority of our findings align with those presented by [64, 14]. For instance, the gene expressions Gstm2, Sirpa, and 2010002N04Rik exhibit positive effects on body weight, whereas the gene expression Dscam, Igfbp2, Irx3 show negative impacts. In particular, the negative impacts of Igfbp2 and Irx3 are also supported by [65, 66].

E.2 Small-Sample Performance

The bottleneck in small-sample performance stems primarily from the large variance of estimating high-order cumulants. We find that a simple sub-sampling technique can reduce the variance, thereby improving small-sample performance substantially. Specifically, we partition the sample into 5 equal subsets, compute the estimates independently on each, and take the median as the final result. With the underlying causal structure not known a priori, we summarize the errors of our estimation method with and without this sub-sampling technique when the sample size is 2k in Tab. 1 below.

Table 1: Errors of our estimation method w/ and w/o sub-sampling when the dataset size is 2k.

Distribution	Sub-sampling	Fig. 3(a)	Fig. 3(b)	Fig. 3(c)	Fig. 3(d)	Fig. 3(e)
Laplace	w/o	0.18±0.2	0.19±0.1	0.54±0.7	0.90±0.9	0.33±0.3
	w/	0.17±0.1	0.19±0.1	0.44±0.4	0.72±0.7	0.35±0.3
Beta	w/o	0.17±0.2	0.14±0.1	0.12±0.1	0.36±0.4	0.20±0.2
	w/	0.14±0.1	0.14±0.1	0.12±0.1	0.29±0.3	0.19±0.1

E.3 Robustness

In this section, we evaluate the robustness of our estimation method. In this section, we always assume the underlying causal structure is not known a priori.

§ Gaussianity

We examine the robustness of our estimation method to Gaussianity by generating the exogenous noises from a generalized Gaussian distribution. Specifically, the probability density function of a generalized Gaussian X is given by $f_X(x;\beta) = \frac{\beta}{2\Gamma(1/\beta)} \exp(-|x|^\beta)$. Note that X follows a Gaussian distribution when $\beta = 2$. We test the performance for $\beta = 0.75, 1.0, 1.25, 1.5$. The experimental results summarized in Fig. 9 demonstrate that our estimation method performs better when the distribution exhibits stronger non-Gaussianity. Even with weak non-Gaussianity, it remains effective in the large-sample regime.

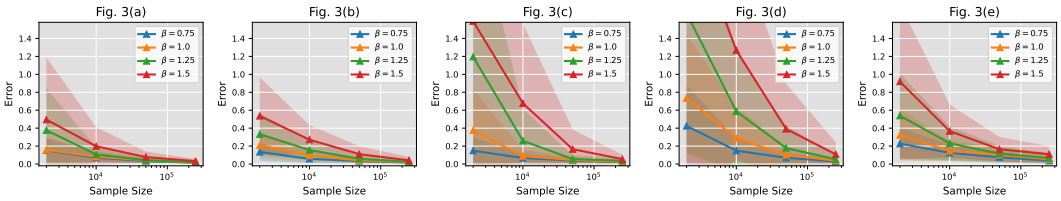


Figure 9: Robustness to Gaussianity.

[3]The dataset can be accessed through this link: https://ndownloader.figstatic.com/articles/17839318/versions/2

§ Nonlinearity

We examine the robustness of our estimation method to nonlinearity by letting the latent confounder causally influence the treatment and outcome in a nonlinear manner. Specifically, we replace the linear term $a_{TL}L$ in the structural assignment of T with a nonlinear term $a_{TL}L^{0.8}$ and the linear term $a_{OL}L$ in the structural assignment of O with a nonlinear term $a_{OL}L^{1.2}$. The results are summarized in Fig. 10, which show that in most cases, our estimation method's performance does not significantly deteriorate when the nonlinearity is not excessively strong.

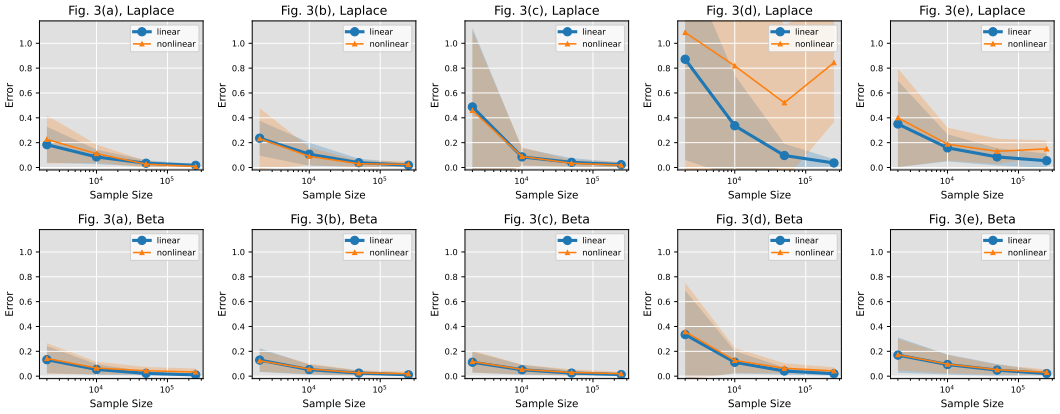


Figure 10: Robustness to nonlinearity.

§ Noisy Proxy

We examine the robustness of our estimation method to a noisy proxy by subjecting the proxy to measurement error. Specifically, we cannot access the true proxy Z but only a noisy version $\tilde{Z}$, where $\tilde{Z} = Z + \tilde{\epsilon}$. Here, $\tilde{\epsilon}$ is a Gaussian noise independent of Z, and the standard deviation of $\tilde{\epsilon}$ is set to 0.25 times that of Z. The experimental results shown in Fig. 11 indicate that the performance of our estimation method does not significantly degrade even with a noisy proxy.

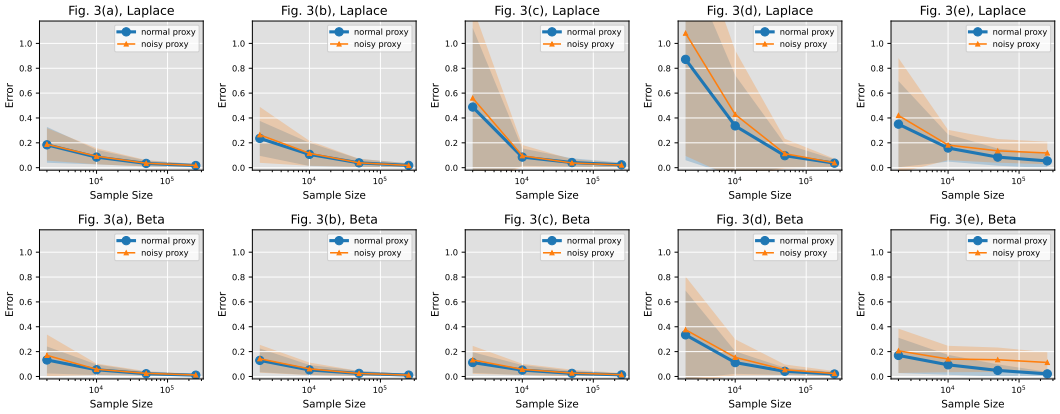


Figure 11: Robustness to noisy proxy.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: We have clearly stated the paper's contributions and scope in the abstract and introduction.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: We have discussed the limitations at the end of Sec. 7.

   Guidelines:

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

3. **Theory assumptions and proofs**

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

   Answer: [Yes]

Justification: We have provided the assumptions (Asmps. 1, 2, and 3) and proofs (see Apps. A, B, C.1, and D.1) for our theoretical results.

Guidelines:

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

4. **Experimental result reproducibility**

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

   Answer: [Yes]

   Justification: We have provided details in Sec. 6.

   Guidelines:

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

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

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

Answer: [Yes]

Justification: We have provided the link of the mouse obesity dataset in App. E.1. Our source code is available at https://github.com/XiuchuanLi/NeurIPS2026-agnostic-proxy.

Guidelines:

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

6. **Experimental setting/details**

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

Answer: [Yes]

Justification: We have provided details in Sec. 6 and App. E.

Guidelines:

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

7. **Experiment statistical significance**

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

Answer: [Yes]

Justification: The shadowed bands in Figs. 6 and 7 denote standard deviations.

Guidelines:

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

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

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: All experiments are performed on an Intel Core i9-13900F Processor.

   Guidelines:

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

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: The research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics

   Guidelines:

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

10. **Broader impacts**

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

    Answer: [Yes]

    Justification: As mentioned in Sec. 7: Our work not only broadens the applicability of causal effect identification but also may benefit research in natural and social sciences. Besides, we foresee no negative societal impacts that warrant specific highlighting.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

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

11. **Safeguards**

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

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

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

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: The creators or original owners of assets used in the paper have been properly credited.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

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

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

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

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

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

    Answer: [N/A]

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

    Guidelines:

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

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

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

    Answer: [N/A]

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

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

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

    Answer: [N/A]

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

    Guidelines:

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