DIRECT LINEAR LATENT CAUSAL STRUCTURE LEARNING VIA TRIPLET DIFFERENCE CONSTRAINT

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning causal structures among unobserved latent variables is fundamental yet challenging, especially when latent variables are causally related. Existing non-Gaussian methods rely heavily on repeated, pairwise or subset-wise independence tests, suffering from error accumulation and heavy computational overheads. To tackle this challenge, we propose a direct framework for latent causal discovery based on higher-order statistical information. Specifically, we introduce the the *Triplet Difference* (TD), a novel algebraic criterion formulated in terms of second- and third-order cumulants, which enables the characterization of latent causal relationships through observable data. Building on this criterion, we establish identifiability conditions for latent causal clusters and causal structures among latent variables. Rather than relying on repeated independence testing, our framework exploits the constraint-satisfaction patterns of TD across variable triplets to distinguish latent causal topologies, thereby enabling latent cluster identification and causal ordering. Complexity analysis confirms that our algorithm avoids exponential evaluation costs over observed variables. Extensive experiments on both synthetic and real-world datasets demonstrate the effectiveness of the proposed approach in recovering latent causal structures, corroborating the theoretical findings and practical utility of the framework.

1 INTRODUCTION

Uncovering causal relationships among latent variables (unobserved representations) is a fundamental task across machine learning and statistical inference. In complex real-world systems, key functional factors are often not directly measurable but instead manifest through multi-dimensional observed indicators. Ignoring such latent factors may lead to spurious dependencies or incorrect causal interpretations among observed variables. Therefore, recovering both the latent structure and the causal relationships of these latent variables (as illustrated in Figure 1) is essential for transparent representation learning and causal reasoning.

Existing latent causal discovery frameworks primarily leverage non-Gaussianity to overcome the limitations of Markov equivalence. Methodologies such as Triad(Cai et al., 2019) and Generalized Independent Noise (GIN)(Xie et al., 2020; 2022) have achieved significant progress by identifying independent noise conditions. However, they suffer from two critical limitations: (1) Heavy Computational and Memory Overhead: They rely heavily on repeated, pairwise or subset-wise kernel-based independence tests (e.g., KCIT(Zhang et al., 2012) or HSIC(Gretton et al., 2007)). This results in combinatorial search spaces over observed variables and high memory complexity ($O(N^2)$ for Gram matrices), making them unscalable to large datasets. (2) Parameter Sensitivity and Error Propagation: Kernel independence tests are notoriously sensitive to hyper-parameters such as kernel bandwidth and significance thresholds, leading to unstable test decisions in finite samples.

Fortunately, non-Gaussian distributions contain rich higher-order statistical information beyond covariance(Cai et al., 2023; Chen et al., 2024). In this paper, we show that structural implications of latent variables can be directly converted into explicit algebraic constraints among different kinds of cross-cumulants, thereby bypassing independence testing entirely. Our key insight stems from an inherent algebraic symmetry induced by shared latent causes. Specifically, suppose two observed variables X_i and X_j belong to the same pure latent cluster generated by L. Let $M_{k,L}$ denote the total causal effect from L to X_k. When evaluated relative to an arbitrary correlated auxiliary variable X_k,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

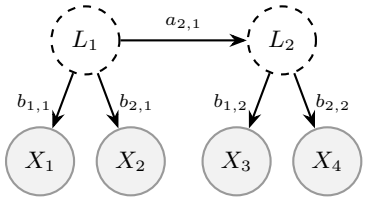


Figure 1: An example of the latent causal structure, where L_1 and L_2 are latent variables and $X_1, X_2, \cdots, X_4$ are observed variables.

the cross-cumulant ratio for X_i: $\frac{\mathcal{C}(X_i,X_k^2)\mathcal{C}(X_i,X_k)}{\mathcal{C}(X_i^2,X_k)} = M_{k,L}^2\mathcal{C}(L^2)$ is completely invariant to the specific choice of X_i, as the latent-to-observed causal coefficient cancels out naturally. Consequently, evaluating the same ratio for X_j yields an identical quantity, establishing a Triplet Difference $\tau_{(i,j|k)}$. Rather than relying on isolated cumulant equalities, we characterize latent causal topologies through the constraint-satisfaction support patterns of Triplet Difference across variable combinations. This algebraic primitive enables pure cluster identification and latent causal ordering without repeated independence testing.

Our main contributions are summarized as follows: (1) We propose the Triplet Difference (TD), a novel cumulant-based algebraic criterion that characterizes latent causal dependencies through explicit constraints on observed data, exploiting higher-order statistical information without directly observing latent variables. (2) We establish identifiability conditions for latent causal clusters and causal order among latent variables, theoretically characterizing how Triplet Difference constraint-satisfaction patterns uniquely encode latent causal topologies under the stated assumptions. (3) We develop TCD, a TD-based algorithm that directly recovers latent causal structures with polynomial complexity $\mathcal{O}(Np^2 + p^3)$.

2 Preliminary

2.1 Linear Non-Gaussian Latent Causal Model

Let $\mathbf{X}$ denote the set of observed variables and $\mathbf{L}$ the set of latent variables. Each variable is assumed to follow the Linear Non-Gaussian Latent Causal Model described as:

$$\begin{aligned} \mathbf{L} &= \mathbf{A}\mathbf{L} + \mathbf{E}_\mathbf{L}, \\ \mathbf{X} &= \mathbf{B}\mathbf{L} + \mathbf{E}_\mathbf{X}, \end{aligned} \tag{1}$$

where $\mathbf{A}$ denotes the causal strength matrix among latent variables, and $\mathbf{B}$ denotes the causal strength matrix from latent variables to observed variables. $\mathbf{E}_L$ and $\mathbf{E}_X$ represent the non-Gaussian noises of the latent and observed variables, respectively, which are assumed to be mutually independent. Without loss of generality, we assume all variables have zero mean; otherwise, they can be centered by a preprocessing step.To ensure the identification of the Linear Non-Gaussian Latent Causal Model, some assumptions are required.

Assumption 2.1 (Purity Assumption)**.** There are no direct directed edges between observed variables.

Assumption 2.2 (Two-Pure-Child Assumption)**.** Each latent variable has at least two pure observed variables as its direct children.

Remark 2.3. These assumptions are commonly used in linear latent causal representation learning. Assumption 2.1 is a standard identification condition widely adopted in latent variable modeling, measurement theory, and structural equation modeling Bollen (2014). In latent causal discovery, pure measurement models serve as the foundational bedrock for identifying underlying latent structures Silva et al. (2006); Kummerfeld & Ramsey (2016); Cai et al. (2019); Xie et al. (2020). It reflects many real-world domains—such as psychometrics, biology, and sensor monitoring—where measurement indicators are designed to selectively reflect specific target latent factors. For Assumption 2.2, the requirement of at least two pure children for each latent variable is a minimal structural condition to guarantee identifiability against causal transitivity Cai et al. (2019); Adams et al. (2021). Compared to classic constraint-based methods like BPC Silva et al. (2006) that strictly demand at

least three pure measurement indicators per latent component, the Two-Pure-Child condition relaxes the measurement requirement while maintaining identifiability under non-Gaussian regimes.

2.2 CUMULANTS

To fully exploit the non-Gaussianity inherent in the data, we introduce cumulants to capture higher-order distributional information from observed data for causal discovery.

Definition 2.4 (**Cumulant Tensor** (Brillinger, 2001))**.** Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote an n-dimensional random vector. Its k-th order cumulant tensor $\mathcal{C}^{(k)}$ forms an $n \times n \times \cdots \times n$ (k modes) array, where the entry indexed by $(i_1, \ldots, i_k)$ is defined as

$$\mathcal{C}^{(k)}_{i_1,\ldots,i_k} = \mathrm{Cum}\left(x_{i_1}, \ldots, x_{i_k}\right) = \sum_{(d_1,\ldots,d_h)} (-1)^{h-1}(h-1)!\,\mathbb{E}\left[\prod_{j\in d_1} x_j\right] \cdots \mathbb{E}\left[\prod_{j\in d_h} x_j\right]. \quad (2)$$

For clarity, the k-th order cumulant of X_i is denoted as

$$\mathcal{C}(X_i^k) = cum(\underbrace{X_i, X_i, \ldots, X_i}_{k\ times}). \quad (3)$$

The cumulant of m-th x_i and n-th x_j is denoted as

$$\mathcal{C}(X_i^m, X_j^n) = cum(\underbrace{X_i, X_i, \ldots, X_i}_{m\ times}, \underbrace{X_j, X_j, \ldots, X_j}_{n\ times}). \quad (4)$$

3 MOTIVATION: INDEPENDENCE BETWEEN SURROGATE VARIABLES AND PSEUDO-VARIABLE

To illustrate the intuition behind our approach, consider the latent-variable causal structure shown in Fig. 1. In this example, L_1 and L_2 are latent variables, while $X_1, \ldots, X_4$ are observed variables. We first review the intuition underlying the Triad constraint, GIN condition, and TIN condition, which share a common strategy of constructing appropriate quantities to eliminate the statistical information shared by observed variables and then exploiting the resulting independence to uncover latent causal structures.

Consider the observed variables X_1, X_2, and X_3. Under the Triad constraint Cai et al. (2019), a pseudo-residual is constructed as $\hat{R} = X_1 - \frac{\mathrm{Cov}(X_3,X_1)}{\mathrm{Cov}(X_3,X_2)} X_2$. Intuitively, this construction cancels the contribution of L_1 shared by X_1 and X_3, or equivalently, removes the component of X_1 that is statistically shared with X_3. Under the corresponding model assumptions, the resulting pseudo-residual satisfies $\hat{R} \perp X_3$. Thus, rather than directly observing L_1, the Triad constraint exploits the independence induced after the shared latent contribution has been eliminated.

The GIN condition Xie et al. (2020; 2022) follows a similar principle but provides a more general way to construct such a quantity. Specifically, it first finds a nonzero vector $\boldsymbol{\omega}$ satisfying $\boldsymbol{\omega}^\top \mathrm{Cov}(\boldsymbol{Y}, X_3) = 0, \quad \boldsymbol{Y} = (X_1, X_2)^\top$, and then constructs the surrogate variable $\hat{R} = \boldsymbol{\omega}^\top \boldsymbol{Y}$. The resulting $\hat{R}$ can be viewed as a surrogate variable or pseudo-residual whose component shared with X_3 has been eliminated. Under the corresponding LiNGAM assumptions, the remaining component becomes independent of X_3. GIN therefore infers latent causal structures not by directly recovering the latent variables, but by exploiting the independence of such constructed surrogate variables.

The TIN condition Dai et al. (2022) further generalizes this idea by seeking a transformation $T(\cdot)$ of $\boldsymbol{Y}$ that removes the shared latent effect in $\boldsymbol{Y}$, such that $T(\boldsymbol{Y}) \perp\!\!\!\perp X_3$.

Although these methods differ in how the transformed quantity is constructed, they share a common underlying principle: construct an appropriate quantity from observed variables so that the contribution shared with a reference set is canceled, and then use the resulting independence as evidence for the underlying latent structure. Importantly, this strategy does not require the latent variables to be directly observed. Instead, it exploits the statistical footprints that latent variables leave in the observed distribution to infer their existence and causal relationships.

4 TRIPLET DIFFERENCE: ALGEBRAIC CONSTRAINT FOR LATENT EFFECTS ELIMINATION

Inspired by this common principle, we ask a different but closely related question: can we construct another type of statistical quantity that removes the indicator-specific effects of a latent variable and yields an invariant signature shared by its observed children, while avoiding the need for repeatedly testing independence?

Our key idea is to exploit the multiplicative structure of higher-order joint cumulants induced by latent-variable models. Consider the data generated according to Figure 1. For the three observed variables X_1, X_2, X_3, the higher-order joint cumulants can be expressed as

$$\begin{aligned}\mathcal{C}(X_1, X_3) &= a_{2,1} b_{1,1} b_{3,2} \mathcal{C}(L_1, L_1),\\ \mathcal{C}(X_1, X_3^2) &= a_{2,1}^2 b_{1,1} b_{3,2}^2 \mathcal{C}(L_1^3),\\ \mathcal{C}(X_1^2, X_3) &= a_{2,1} b_{1,1}^2 b_{3,2} \mathcal{C}(L_1^3).\end{aligned} \tag{5}$$

Similarly, one has

$$\begin{aligned}\mathcal{C}(X_2, X_3) &= a_{2,1} b_{2,1} b_{3,2} \mathcal{C}(L_1, L_1),\\ \mathcal{C}(X_2, X_3^2) &= a_{2,1}^2 b_{2,1} b_{3,2}^2 \mathcal{C}(L_1^3),\\ \mathcal{C}(X_2^2, X_3) &= a_{2,1} b_{2,1}^2 b_{3,2} \mathcal{C}(L_1^3).\end{aligned} \tag{6}$$

Although these quantities involve different observed variables and different loading coefficients, they share a crucial algebraic structure: the contribution of the latent variable L_1 enters all of them in a highly structured multiplicative form. This structure allows us to construct a ratio that cancels the loading coefficients associated with the individual observed indicators. Specifically,

$$\begin{aligned}\frac{\mathcal{C}(X_1, X_3^2)\mathcal{C}(X_1, X_3)}{\mathcal{C}(X_1^2, X_3)} &= a_{2,1}^2 b_{3,2}^2 \mathcal{C}(L_1, L_1),\\ \frac{\mathcal{C}(X_2, X_3^2)\mathcal{C}(X_2, X_3)}{\mathcal{C}(X_2^2, X_3)} &= a_{2,1}^2 b_{3,2}^2 \mathcal{C}(L_1, L_1).\end{aligned} \tag{7}$$

Interestingly, the causal coefficients $b_{1,1}$ and $b_{1,2}$ are canceled out. Consequently, although X_1 and X_2 are different observed variables, they yield the same quantity when they share the same latent parent L_1. By taking the difference between the two ratio-based quantities, we obtain

$$\frac{\mathcal{C}(X_1, X_3^2)\mathcal{C}(X_1, X_3)}{\mathcal{C}(X_1^2, X_3)} - \frac{\mathcal{C}(X_2, X_3^2)\mathcal{C}(X_2, X_3)}{\mathcal{C}(X_2^2, X_3)} = 0. \tag{8}$$

Therefore, Eq. 8 provides an algebraic signature of shared latent ancestry. More importantly, this signature is obtained directly from higher-order cumulants, rather than by repeatedly testing conditional independence.

Based on this observation, we formally introduce our cumulant-based discriminative statistic, namely *Triplet Difference*.

Definition 4.1 (Triplet Difference)**.** Suppose that X_i, X_j and X_k are distinct and correlated variables and that all noise variables are non-Gaussian. Define the third-order cumulant Triplet Difference of $\{X_i, X_j\}$ relative to X_k, as

$$\tau_{(i,j|k)} := \frac{\mathcal{C}(X_i, X_k^2)\mathcal{C}(X_i, X_k)}{\mathcal{C}(X_i^2, X_k)} - \frac{\mathcal{C}(X_j, X_k^2)\mathcal{C}(X_j, X_k)}{\mathcal{C}(X_j^2, X_k)}. \tag{9}$$

We say that a *vanishing triplet difference* holds for $\{X_i, X_j\}$ and X_k whenever $\tau_{(i,j|k)} = 0$.

Using higher-order cumulants, it can be generalized to the n-th order cumulant-based measure as:

Definition 4.2 (N-th Order Cumulant-based Triplet Difference)**.** Suppose that X_i, X_j and X_k are distinct and correlated variables and that all noise variables are non-Gaussian. For $n \geq 3$ and $m \geq 1$, define the *N-th Order Cumulant-based Triplet Difference* of $\{X_i, X_j\}$ relative to X_k, as

$$\tau^{(n)}_{(i,j|k)} := \frac{\mathcal{C}(X_i^m, X_k^{n-m})\mathcal{C}(X_i, X_k)}{\mathcal{C}(X_i^{m+1}, X_k^{n-m-1})} - \frac{\mathcal{C}(X_j^m, X_k^{n-m})\mathcal{C}(X_j, X_k)}{\mathcal{C}(X_j^{m+1}, X_k^{n-m-1})}. \tag{10}$$

We say that a *vanishing n-th order cumulant-based measure* holds for $\{X_i, X_j\}$ and X_k if and only if $\tau^{(n)}_{(i,j|k)} = 0$ with $n \geq 3$.

5 TRIPLET DIFFERENCE BASED METHOD FOR LATENT CAUSAL STRUCTURE LEARNING

Triplet difference based latent Causal Discovery (TCD).

In this section, we develop a unified identifiability framework based on triplet difference constraints derived from higher-order cumulants. We first establish an equivalent characterization of shared latent variables, showing that the membership of observed variables in the same latent causal cluster can be determined through observable statistical constraint. We further investigate the identifiability of latent causal ordering. These results connect observable higher-order statistics to latent causal structural properties, providing the theoretical foundation for the proposed latent causal discovery algorithm.

5.1 IDENTIFICATION OF CAUSAL CLUSTERS

We first address the identification of groups of observed variables with a shared latent direct cause, which we term *causal cluster*. The key to identifying latent clusters is to determine whether the structural constraint of shared latent parenthood admits an equivalent characterization in the observable distribution. Under the assumed model, the higher-order cumulant structure induces specific triplet difference constraints among the observed descendants of a common latent parent. These constraints provide a distributional characterization of latent cluster membership, allowing the underlying grouping structure to be identified without directly observing the latent variables.

Theorem 5.1. *Under Assumptions 2.1 and 2.2, let* $\mathbf{X}$ *be the set of all observed variables in the model as Eq. (1). Suppose* X_i *and* X_j *are distinct and correlated variables. If and only if* $\forall X_k \in \mathbf{X} \setminus \{X_i, X_j\}$ *and vanishing triplet differences hold among* $\{X_i, X_j\}$ *and* X_k*, then* X_i *and* X_j *belong to the same cluster.*

Theorem 5.1 establishes a correspondence between latent structural properties and observable statistical criteria. Specifically, the theorem states that under the given assumptions, shared latent parent relationships can be characterized by specified triplet difference constraints. This equivalence provides a completeness guarantee for identifying valid cluster members and supplies a structural criterion to eliminate incompatible cluster partitions at the population level.

Based on Definition 4.2 and Theorem 5.1, we first determine whether two observed variables belong to the same causal cluster, and then merge the corresponding clusters accordingly. As for the merging criterion, we find that overlapping causal clusters can be consolidated into a single cluster. The validity of this merging procedure is guaranteed by Proposition 5.2.

Proposition 5.2. *Let* C_1 *and* C_2 *be two clusters. If* C_1 *and* C_2 *are overlapping,* C_1 *and* C_2 *share the same latent parent.*

This proposition holds true because of the equivalence of the pure variables in terms of triplet-difference constraints. In particular, as shown in Theorem 5.1, all variables in a cluster have the same triplet-difference constraints.

These results lead to a discovering causal clusters algorithm consisting of three main steps. First, we screen the observed variables according to their statistical dependence to construct candidate causal clusters. Second, we apply the criterion established in Theorem 5.1 to identify causal clusters among the candidates. Finally, following Proposition 5.2, we merge overlapping causal clusters to obtain the resulting cluster structure.

Example 5.3. *Consider the example in Figure 1. First, we identify* $\{X_1, X_2\}$ *and* $\{X_3, X_4\}$ *as causal clusters based on Theorem 5.1. Then, we introduce* L_1 *and* L_2 *as their respective latent parents. Finally, we return the partial graph* $\mathcal{G} = \{L_1 \to \{X_1, X_2\}, L_2 \to \{X_3, X_4\}\}$*, where the causal direction between* L_1 *and* L_2 *remains undetermined.*

5.2 IDENTIFICATION OF CAUSAL ORDER AMONG LATENT VARIABLES

Having identified the latent clusters, we next investigate whether their causal ordering can be recovered from observable higher-order statistics. The central difficulty is that the latent variables are not directly observed, and therefore conventional causal orientation procedures cannot be applied

Algorithm 1 Causal Clusters Discovering

Input: Observed variable set $\mathbf{X} = \{X_1, \ldots, X_d\}$
Output: Causal cluster set $\mathcal{S}$

Initialize the candidate cluster set $\mathcal{S} \leftarrow \varnothing$;
for all $(X_i, X_j) \in \mathbf{X}$ **do**
  if $\tau^{(n)}_{(i,j|k)} = 0$ for every $X_k \in \mathbf{X} \setminus \{X_i, X_j\}$ **then**
    $\mathcal{S} \leftarrow \mathcal{S} \cup \big\{\{X_i, X_j\}\big\}$;
  end if
end for
Merge all the overlapping sets in $\mathcal{S}$;
return $\mathcal{S}$

Algorithm 2 Latent Causal Structure Learning

Input: Observed data $\mathbf{X}$; latent clusters $\mathcal{C} = \{C_1, \ldots, C_r\}$;
Output: Latent DAG $\mathcal{G}_L$

Initialize latent variables $\mathcal{L} = \{L_1, \ldots, L_r\}$
Construct the proxy-correlation graph $\mathcal{G}_\rho = (\mathcal{L}, F)$ and obtain its connected components $\mathcal{H}$
$\mathcal{G}_L \leftarrow \emptyset$
for each $H \in \mathcal{H}$ **do**
  for each pair $L_a, L_b \in H$ with $a < b$ **do**
    $N_{a\to b} \leftarrow \Big|\Big\{(X_i, X_j, X_k) : X_i, X_j \in C_a,\ X_k \in C_b,\ |\tau^{(n)}_{(i,k|j)}| \le \delta\Big\}\Big|$
    Compute $N_{b\to a}$ analogously
    $D_{ab} \leftarrow \mathrm{sgn}(N_{a\to b} - N_{b\to a}), \quad D_{ba} \leftarrow -D_{ab}$
  end for
  $s_a \leftarrow \sum_{L_b \in H\setminus\{L_a\}} D_{ab}, \quad \forall L_a \in H$
  Sort latent variables in H by decreasing s_a to obtain $\pi_H = (L_{(1)}, \ldots, L_{(|H|)})$
  Add $L_{(i)} \to L_{(j)}$ to $\mathcal{G}_L$ for all $i < j$
end for
if $r > 2$ **then**
  for each candidate edge $L_a \to L_b$ **do**
    $\mathcal{P}_b \leftarrow \mathrm{Pred}(L_b) \setminus \{L_a\}$
    for each $Q \subseteq \mathcal{P}_b$ in increasing $|Q|$ **do**
      Construct $M_{a,b|Q}$
      if $\mathrm{rank}(M_{a,b|Q}) \le |Q|$ **then**
        Remove $L_a \to L_b$ from $\mathcal{G}_L$
        break
      end if
    end for
  end for
end if
return $\mathcal{G}_L$

to them directly. We show that triplet difference constraints involving the observed descendants of different latent clusters provide an identifiable characterization of latent ancestral relations under the assumed structural conditions.

Theorem 5.4. *Under Assumptions 2.1 and 2.2, suppose the data of observed variables* $\mathbf{X}$ *are generated by Eq. 1. Let* L_a *and* L_b *be two directed connected latent variables without confounders, and let* $\{X_i, X_j\}$ *be two children of* L_a*. Then if and only if there exist an observed variable* X_k *that is a child of* L_b*, such that* $\tau^{(n)}_{(i,k|j)} = 0$*, then,* L_a *is an ancestor of* L_b*.*

Once the latent causal order is identified, indirect ancestral relations can be removed using conditional independence among the latent variables by using rank-constraint characterization of latent conditional independence in Silva et al. (2006). Thus, we can derive Algorithm 2 to recover the latent causal structure.

6 COMPLEXITY ANALYSIS

In this section, we analyze the computational complexity of TCD and representative baselines in terms of the number of observed variables p and the sample size N. Let r denote the number of discovered latent variables.

TCD first estimates the required low-order cumulants and then performs triplet-wise algebraic constraints. With cached cumulants, causal-cluster discovery and latent-order recovery require $\mathcal{O}\big(Np^2 + p^3\big)$. Redundant latent edges are subsequently removed by conditioning on subsets of latent predecessors, yielding an additional worst-case cost of $\mathcal{O}(2^r r^5)$. Thus, the overall worst-case complexity is $\mathcal{O}\big(Np^2 + p^3 + 2^r r^5\big)$. Under our assumption that each latent variable has at least two observed children, $r \leq \lfloor p/2 \rfloor$, and in typical settings $r \ll p$. Therefore, the combinatorial search in TCD is performed only in the much smaller latent space rather than over all observed variables.

In terms of memory, TCD requires only the observed data and cached low-order cumulant statistics. In contrast, kernel-based independence tests such as HSIC typically construct $N \times N$ Gram matrices, leading to $\mathcal{O}(N^2)$ memory per test. This applies to Triad, LSTC, GIN, and LaHiCaSl when HSIC-based tests are employed, making repeated kernel tests increasingly costly as N grows. Therefore, TCD avoids the repeated kernel-based testing bottleneck while maintaining polynomial dependence on the observed sample size and variable dimension in its main computational stages. More detailed elaborations and complexity comparisons with state-of-the-art methods are provided in the appendix.

7 EXPERIMENTS

7.1 SYNTHETIC DATA

In our simulation study, we consider a relatively complex set of model mechanisms. Several causal graphs involving latent variables are considered. In all these model settings, the observed variables are randomly generated according to Equation 1, with sample sizes $n \in \{500, 1000, 2000\}$. The causal coefficients are sampled from a uniform distribution over $[-2, -1] \cup [1, 2]$. The noise terms e are drawn from various non-Gaussian distributions, specifically: (a) cubed Gamma distribution; (b) cubed Gumbel distribution; and (c) cubed Laplace distribution.

Case 1 : L_1 and L_2 each have two pure measurement variables, i.e., $L_1 \to \{X_1, X_2\}$ and $L_2 \to \{X_3, X_4\}$.

Case 2 : L_1 and L_2 have different numbers of measurement variables: $L_1 \to \{X_1, X_2, X_3\}$ and $L_2 \to \{X_4, X_5\}$.

Case 3 : Each latent variable has two measurement variables: $L_1 \to \{X_1, X_2\}$, $L_2 \to \{X_3, X_4\}$, $L_3 \to \{X_5, X_6\}$.

Case 4 : Each latent variable has multiple measurement variables: $L_1 \to \{X_1, X_2, X_3, X_4\}$, $L_2 \to \{X_5, X_6, X_7, X_8\}$, $L_3 \to \{X_9, X_{10}, X_{11}, X_{12}\}$.

Case 5 : Impure variables are added to Case 1: $L_1 \to \{X_1, X_2, X_3\}$, $L_2 \to \{X_4, X_5, X_6\}$, with extra edges $X_2 \to X_3$ and $X_5 \to X_6$.

Case 6 : Each latent variable has two measurement variables (same as Case 3), but the causal directions among the latents are different: $L_1 \to L_2 \leftarrow L_3$.

We compared the proposed algorithm with ETCD(Li & Liu, 2025), RLCD(Dong et al., 2024),LaHiCaSl(Xie et al., 2024),GIN(Xie et al., 2020), LSTC(Cai et al., 2019), BPC(Silva et al., 2006), and FOFC(Kummerfeld & Ramsey, 2016). The overall experimental procedure consisted of two stages. The first stage was causal cluster discovery, which aimed to identify groups of observed variables influenced by the same latent variable. The second stage focused on discovering causal relationships among the latent variables. Since BPC and FOFC are not capable of learning the structure among causal clusters, they were included only in the comparison for the first stage. All experimental results reported below were averaged over 10 independent runs of each algorithm on randomly generated datasets.

To evaluate the quality of the recovered measurement model and structural model against the ground truth, we adopt the following metrics. For the measurement model, we use *latent omission* $= \frac{OL}{TL}$,

Table 1: Measurement-model recovery results under the cubed Gumbel distribution.

Case	n	Latent Omission↓								Latent Commission↓								Mismeasurement↓							
		TCD	ETCD	RLCD	LaHiCaSl	GIN	LSTC	FOFC	BPC	**TCD**	ETCD	RLCD	LaHiCaSl	GIN	LSTC	FOFC	BPC	**TCD**	ETCD	RLCD	LaHiCaSl	GIN	LSTC	FOFC	BPC
Case 1	500	0.25	0.00	0.00	0.00	0.30	0.40	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.08	0.00	0.00	0.00
	1000	0.10	0.10	0.10	0.10	0.10	0.55	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00
	2000	0.10	0.15	0.00	0.00	0.00	0.60	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00
Case 2	500	0.15	0.00	0.00	0.00	0.30	0.35	1.00	0.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.02	0.00	0.06	0.00	0.00	0.16
	1000	0.05	0.05	0.00	0.00	0.25	0.45	1.00	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.18
	2000	0.05	0.10	0.00	0.00	0.40	0.65	1.00	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.18
Case 3	500	0.23	0.10	0.07	0.00	0.40	0.60	1.00	0.67	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.02	0.02	0.00	0.12	0.00	0.00	0.28
	1000	0.17	0.07	0.03	0.03	0.30	0.50	1.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.03	0.10	0.00	0.00	0.33
	2000	0.20	0.07	0.07	0.07	0.40	0.57	1.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.08	0.18	0.00	0.00	0.25
Case 4	500	0.07	0.00	0.00	0.00	0.00	0.53	1.00	0.33	0.03	0.00	0.00	0.03	0.20	0.00	0.00	0.00	0.07	0.00	0.01	0.00	0.08	0.00	0.00	0.06
	1000	0.07	0.00	0.00	0.00	0.07	0.60	1.00	0.23	0.07	0.00	0.03	0.07	0.06	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.11	0.00	0.00	0.04
	2000	0.03	0.00	0.00	0.00	0.03	0.40	1.00	0.23	0.07	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.04
Case 5	500	0.20	0.80	0.10	0.00	0.15	0.75	1.00	0.75	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.08	0.00	0.00	0.08
	1000	0.20	0.75	0.10	0.00	0.35	0.85	1.00	0.80	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.13	0.00	0.03	0.05	0.12	0.00	0.00	0.08
	2000	0.40	0.85	0.00	0.00	0.35	0.95	1.00	0.95	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.12	0.00	0.00	0.02
Case 6	500	0.07	0.37	0.37	0.10	0.40	0.53	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.02	0.03	0.18	0.00	0.00	0.28
	1000	0.17	0.37	0.20	0.07	0.40	0.57	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.03	0.03	0.28	0.00	0.00	0.33
	2000	0.10	0.27	0.00	0.00	0.43	0.60	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.18	0.00	0.00	0.25

Table 2: Direction recovery results under the cubed Gumbel noise.

Algorithm		direction_pre_f1↑						direction_true_f1↑					
		TCD	ETCD	RLCD	LaHiCaSl	GIN	LSTC	**TCD**	ETCD	RLCD	LaHiCaSl	GIN	LSTC
Case 1	500	0.50	**0.90**	0.10	**0.90**	0.30	0.30	**1.00**	0.90	0.10	0.90	0.70	**1.00**
	1000	**0.80**	**0.80**	0.50	0.50	0.10	0.30	**1.00**	0.70	0.50	0.60	0.40	**1.00**
	2000	**0.80**	0.60	0.00	**0.80**	0.00	0.20	**1.00**	0.80	0.00	0.80	0.60	**1.00**
Case 2	500	0.70	0.70	**0.80**	0.60	0.40	0.40	**1.00**	0.80	0.80	0.80	0.80	**1.00**
	1000	**0.90**	0.70	0.40	0.60	0.40	0.30	**1.00**	0.90	0.40	0.90	0.80	**1.00**
	2000	**0.90**	0.60	0.30	0.60	0.20	0.10	**1.00**	0.90	0.30	0.80	0.60	**1.00**
Case 3	500	0.70	0.70	0.45	**0.82**	0.21	0.18	**0.97**	0.70	0.45	0.82	0.58	0.80
	1000	0.73	0.80	0.45	**0.88**	0.07	0.21	**0.95**	0.80	0.45	0.92	0.30	0.80
	2000	**0.70**	0.60	0.52	**0.70**	0.35	0.24	**1.00**	0.80	0.52	0.93	0.42	0.80
Case 4	500	**0.77**	**0.77**	0.57	0.60	0.38	0.00	**1.00**	0.90	0.57	0.70	0.17	0.80
	1000	**0.74**	0.50	0.55	0.73	0.29	0.07	**0.90**	0.70	0.60	0.77	0.26	0.80
	2000	**0.82**	0.47	0.45	0.77	0.37	0.13	**1.00**	0.80	0.45	0.75	0.42	0.80
Case 5	500	**0.60**	0.00	0.00	0.00	0.00	0.00	**1.00**	**1.00**	0.00	0.90	0.00	**1.00**
	1000	**0.50**	0.00	0.20	0.00	0.00	0.00	**1.00**	0.80	0.20	0.70	0.00	**1.00**
	2000	**0.40**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.80	0.00	0.70	0.00	**1.00**
Case 6	500	**0.85**	0.20	0.05	0.61	0.21	0.13	**0.96**	0.90	0.05	0.76	0.53	0.68
	1000	0.70	0.30	0.20	**0.73**	0.36	0.16	**1.00**	0.90	0.20	0.85	0.69	0.76
	2000	**0.90**	0.37	0.30	0.66	0.20	0.00	**1.00**	0.50	0.30	0.66	0.31	0.80

latent commission $= \frac{FL}{TL}$, and *measurement error* $= \frac{MO}{TO}$, where OL is the number of omitted latent variables, FL the number of falsely claimed latent variables, TL the total number of latent variables in the true graph, MO the number of misspecified measurement edges, and T_O the total number of true measurement edges (See the details in (Silva et al., 2006)). For the recovered structural model, we use the F_1 score: $F_1 = 2 \times \frac{P \times R}{P+R}$, where P and R denote precision and recall, respectively.

In the main text, we only present the results for the noise distributions (a) cubed Gamma distribution; (b) cubed Gumbel distribution; (c) cubed Laplace distribution;. The complete results for all other noise settings are provided in the appendix.

As shown in Table 1, TCD achieves stable and competitive measurement-model recovery across the three non-Gaussian noise distributions. It generally yields lower latent omission than GIN, LSTC, FOFC, and BPC, while remaining competitive with ETCD, RLCD, and LaHiCaSl, including in the more challenging Cases 5 and 6. Across all three metrics, TCD maintains a balanced recovery profile, with near-zero latent commission in most settings and generally low mismeasurement. While RLCD and LaHiCaSl attain lower omission in some cases, RLCD is the strongest overall baseline, whereas LaHiCaSl can suffer from larger commission or mismeasurement errors in the more difficult settings.

As shown in Table 2, TCD exhibits strong and stable performance in latent causal direction recovery across all three non-Gaussian noise distributions. When the ground-truth latent clusters are provided, `direction_true_f1` remains close to 1.0 in nearly all settings, confirming that the proposed direction-identification criterion can accurately recover latent causal relations once the measurement structure is correctly specified.

More importantly, TCD achieves the best or near-best `direction_pre_f1` in the majority of settings. Since this metric evaluates the complete two-stage pipeline using the clusters recovered in the first stage, these results highlight the overall advantage of TCD in end-to-end latent causal structure recovery. This advantage is particularly clear in the more challenging Cases 5 and 6, where several baselines suffer substantial degradation after cluster-estimation errors propagate to the direction-recovery stage, whereas TCD remains comparatively stable. Overall, TCD combines accurate latent-direction identification with stronger robustness at the full-pipeline level, demonstrating a clear advantage in jointly recovering both the measurement structure and the latent causal structure.

7.2 Real-World Data

We evaluate the proposed method on the "12 Industry Portfolios Daily" dataset from the Kenneth R. French Data Library, covering 12 industry portfolios from July 1, 1926, to April 30, 2026, with 52,471 daily observations. To account for potential deviations from model assumptions in real-world data, we employ bootstrap confidence intervals based on 500 resamples to improve the robustness of maximal causal cluster identification.

The identified latent causal structure by our method is given in Fig. 2. Our method identifies three causal clusters: L_1, comprising non-durables, energy, and finance; L_2, comprising durables, manufacturing, business equipment, telecommunications, wholesale/retail, and others; and L_3, comprising chemicals and healthcare. This result reveals economically interpretable inter-industry relationships and is broadly consistent with the cascade propagation mechanism discussed by Acemoglu et al. (2012). Specifically, L_1 captures macro-financial and fundamental demand shocks, including basic consumption demand, energy prices, and inflation expectations, and occupies the upstream position in the inferred causal structure. This suggests that such shocks may propagate to downstream sectors. L_2 encompasses production, investment, distribution, and final demand, whose returns tend to co-move with market demand, investment sentiment, and business-cycle fluctuations. L_3, comprising chemicals and healthcare, represents specialized production sectors whose returns may respond to upstream economic shocks and input-cost fluctuations.

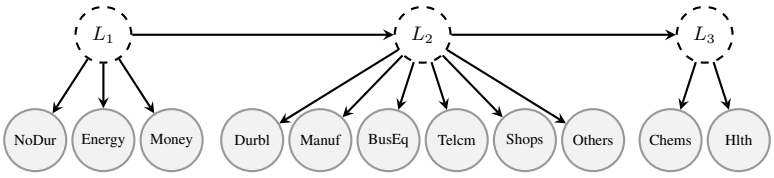


Figure 2: Latent causal structure identified by TCD on the 12 Industry Portfolios Daily dataset.

8 Conclusion

In this paper, we propose a new Triplet Difference Constraint, which constructs difference-based constraints among triplets of variables to effectively eliminate the influence of latent variables on the dependencies among observed variables. Based on this criterion, we establish identifiability theories for latent causal clusters and the causal structures among latent variables, and further propose Triplet difference based latent Causal Discovery (TCD). By exploiting higher-order statistical information captured by joint cumulants of observed variables, TCD enables the identification of latent causal clusters and the causal structures among latent variables without directly observing the latent variables. Experiments on both simulated and real-world datasets demonstrate the effectiveness of the proposed theory and method. Future work will investigate identifiability under more complex latent-variable structures, nonlinear and nonstationary settings, as well as explore applications of the proposed approach to scientific discovery and causal analysis of complex systems.

AI use statement

In preparing this paper, we used generative AI tools to polish the language of the manuscript. Generative AI was not used to develop the research ideas, design the methodology or experiments, analyze

the results, or draw conclusions. The authors carefully reviewed and revised all AI-assisted language edits. We take full responsibility for the final content of the paper.

ETHICS STATEMENT

Our work develops a new approach for latent causal discovery only from observed variables. All empirical evaluations rely on synthetic datasets and public financial data generated from specified causal models. No human data, private records, or proprietary sensitive data are used in the experiments. The method is designed for scientific causal discovery tasks. While causal learning techniques may have downstream applications, we are not aware of immediate harmful risks introduced by this theoretical contribution. We adhere to the ethical principles for machine learning research.

REPRODUCIBILITY STATEMENT

All theoretical derivations, proofs, and complexity comparisons against existing methods are provided in the appendix. The synthetic data generation pipeline, hyperparameter settings and code of our proposed method to reproduce our experimental results will be publicly released upon acceptance of this paper. All experiments are conducted with fixed random seeds, and we report the mean across multiple runs. The software environment and computational resources are documented in the appendix. No proprietary datasets are used in this work.

REFERENCES

Daron Acemoglu, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 2012.

Jeffrey Adams, Niels Hansen, and Kun Zhang. Identification of partially observed linear causal models: Graphical conditions for the non-gaussian and heterogeneous cases. *Advances in Neural Information Processing Systems*, 34:22822–22833, 2021.

Kenneth A Bollen. *Structural equations with latent variables*. John Wiley & Sons, 2014.

David R Brillinger. *Time series: data analysis and theory*. SIAM, 2001.

Ruichu Cai, Feng Xie, Clark Glymour, Zhifeng Hao, and Kun Zhang. Triad constraints for learning causal structure of latent variables. *Advances in neural information processing systems*, 32, 2019.

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*, pp. 3380–3407. PMLR, 2023.

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*, volume 38, pp. 20353–20361, 2024.

Wei Chen, Linjun Peng, Zhiyi Huang, Ruichu Cai, Zhifeng Hao, and Kun Zhang. Higher order cumulants-based method for direct and efficient causal discovery. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

Wei Chen, Wanyang Gu, Linjun Peng, Ting Yan, Ruichu Cai, Zhifeng Hao, and Kun Zhang. Horizontal and vertical federated causal structure learning via higher-order cumulants. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 20280–20288, 2026.

Zhitang Chen, Kun Zhang, and Laiwan Chan. Causal discovery with scale-mixture model for spatiotemporal variance dependencies. *Advances in Neural Information Processing Systems*, 25, 2012.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Nov):507–554, 2002.

Diego Colombo, Marloes H Maathuis, Markus Kalisch, and Thomas S Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pp. 294–321, 2012.

Haoyue Dai, Peter Spirtes, and Kun Zhang. Independence testing-based approach to causal discovery under measurement error and linear non-gaussian models. *Advances in Neural Information Processing Systems*, 35:27524–27536, 2022.

Xinshuai Dong, Biwei Huang, Ignavier Ng, Xiangchen Song, Yujia Zheng, Songyao Jin, Roberto Legaspi, Peter Spirtes, and Kun Zhang. A versatile causal discovery framework to allow causally-related hidden variables. In *International Conference on Learning Representations*, volume 2024, pp. 43084–43118, 2024.

Doris Entner and Patrik O Hoyer. Discovering unconfounded causal relationships using linear non-gaussian models. In *JSAI International Symposium on Artificial Intelligence*, pp. 181–195. Springer, 2010.

Arthur Gretton, Kenji Fukumizu, Choon Teo, Le Song, Bernhard Schölkopf, and Alex Smola. A kernel statistical test of independence. *Advances in neural information processing systems*, 20, 2007.

Patrik Hoyer, Dominik Janzing, Joris M Mooij, Jonas Peters, and Bernhard Schölkopf. Nonlinear causal discovery with additive noise models. *Advances in neural information processing systems*, 21, 2008a.

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, 2008b.

Erich Kummerfeld and Joseph Ramsey. Causal clustering for 1-factor measurement models. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1655–1664, 2016.

Xiuchuan Li and Tongliang Liu. Efficient and trustworthy causal discovery with latent variables and complex relations. In *International Conference on Learning Representations*, volume 2025, pp. 34349–34383, 2025.

Jie Qiao, Yu Xiang, Zhengming Chen, Ruichu Cai, and Zhifeng Hao. Causal discovery from poisson branching structural causal model using high-order cumulant with path analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 20524–20531, 2024.

Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006.

Shohei Shimizu, Patrik O Hoyer, and Aapo Hyvärinen. Estimation of linear non-gaussian acyclic models for latent factors. *Neurocomputing*, 72(7-9):2024–2027, 2009.

Ricardo Silva, Richard Scheines, Clark Glymour, and Peter Spirtes. Learning the structure of linear latent variable models. *Journal of Machine Learning Research*, 7(2), 2006.

Charles Spearman. Pearson's contribution to the theory of two factors. *British Journal of Psychology*, 19(1):95, 1928.

Peter Spirtes and Clark Glymour. An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, 9(1):62–72, 1991.

Peter L Spirtes. Calculation of entailed rank constraints in partially non-linear and cyclic models. *arXiv preprint arXiv:1309.7004*, 2013.

Peter L Spirtes, Christopher Meek, and Thomas S Richardson. Causal inference in the presence of latent variables and selection bias. *arXiv preprint arXiv:1302.4983*, 2013.

Tatsuya Tashiro, Shohei Shimizu, Aapo Hyvärinen, and Takashi Washio. Parcelingam: A causal ordering method robust against latent confounders. *Neural computation*, 26(1):57–83, 2014.

Feng Xie, Ruichu Cai, Biwei Huang, Clark Glymour, Zhifeng Hao, and Kun Zhang. Generalized independent noise condition for estimating latent variable causal graphs. *Advances in neural information processing systems*, 33:14891–14902, 2020.

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*, pp. 24370–24387. PMLR, 2022.

Feng Xie, Biwei Huang, Zhengming Chen, Ruichu Cai, Clark Glymour, Zhi Geng, and Kun Zhang. Generalized independent noise condition for estimating causal structure with latent variables. *Journal of Machine Learning Research*, 25(191):1–61, 2024.

Kun Zhang, Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. Kernel-based conditional independence test and application in causal discovery. *arXiv preprint arXiv:1202.3775*, 2012.

A PROOFS OF THEOREMS

Theorem 5.1. Under Assumptions 2.1 and 2.2, let $\mathbf{X}$ be the set of all observed variables in the model as Eq. (1). Suppose X_i and X_j are distinct and correlated variables. If and only if $\forall X_k \in \mathbf{X} \setminus \{X_i, X_j\}$ and vanishing triplet differences hold among $\{X_i, X_j\}$ and X_k, then X_i and X_j belong to the same cluster.

Proof. $\mathbf{X}$ generated by Eq.(1) can be written in terms of the mixing matrix form:

$$\mathbf{X} = [\mathbf{M}_{\mathbf{X},\mathbf{E_L}} \quad \mathbf{M}_{\mathbf{X},\mathbf{E_X}}] \begin{bmatrix} \mathbf{E_L} \\ \mathbf{E_X} \end{bmatrix}. \tag{11}$$

$\tau_{(i,j|k)}$ can be rewritten as follows:

$$\begin{aligned} \tau^{(n)}_{(i,j|k)} =& \frac{\mathcal{C}(X_i^m, X_k^{n-m})\mathcal{C}(X_i, X_k)}{\mathcal{C}(X_i^{m+1}, X_k^{n-m-1})} - \frac{\mathcal{C}(X_j^m, X_k^{n-m})\mathcal{C}(X_j, X_k)}{\mathcal{C}(X_j^{m+1}, X_k^{n-m-1})} \\ =& \frac{\sum_z M^m_{X_i,E_{L_z}} M^{n-m}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z}) \sum_z M_{X_i,E_{L_z}} M_{X_k,E_{L_z}} \mathcal{C}(E^2_{L_z})}{\sum_z M^{m+1}_{X_i,E_{L_z}} M^{n-m-1}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z})} \\ & - \frac{\sum_z M^m_{X_j,E_{L_z}} M^{n-m}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z}) \sum_z M_{X_j,E_{L_z}} M_{X_k,E_{L_z}} \mathcal{C}(E^2_{L_z})}{\sum_z M^{m+1}_{X_j,E_{L_z}} M^{n-m-1}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z})} \end{aligned} \tag{12}$$

where $\mathcal{C}_3(E_{L_z})$ is an acronym for $cum(E_{L_z}, E_{L_z}, E_{L_z})$.

There are two cases to consider.

Case 1) X_i and X_j belong to the same cluster, which means $\{X_i, X_j\}$ share the same one parental latent variable. Let L_a be the parental latent variable. We have $\forall E_{L_z} \in \mathbf{E_L}, M_{X_i,E_{L_z}} = B_{X_i,L_a} M_{L_a,E_{L_z}}$ and $\forall E_{L_z} \in \mathbf{E_L}, M_{X_j,E_{L_z}} = B_{X_j,L_a} M_{L_a,E_{L_z}}$. So the triplet difference can be rewritten as follows:

$$\begin{aligned} \tau^{(n)}_{(i,j|k)} =& \frac{\sum_z M^m_{X_i,E_{L_z}} M^{n-m}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z}) \sum_z M_{X_i,E_{L_z}} M_{X_k,E_{L_z}} \mathcal{C}(E^2_{L_z})}{\sum_z M^{m+1}_{X_i,E_{L_z}} M^{n-m-1}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z})} \\ & - \frac{\sum_z M^m_{X_j,E_{L_z}} M^{n-m}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z}) \sum_z M_{X_j,E_{L_z}} M_{X_k,E_{L_z}} \mathcal{C}(E^2_{L_z})}{\sum_z M^{m+1}_{X_j,E_{L_z}} M^{n-m-1}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z})} \\ =& \frac{\sum_z M^m_{X_i,E_{L_z}} M^{n-m}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z}) \sum_z M_{X_i,E_{L_z}} M_{X_k,E_{L_z}} \mathcal{C}(E^2_{L_z})}{\sum_z M^{m+1}_{X_i,E_{L_z}} M^{n-m-1}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z})} \\ & - \frac{\sum_z M^m_{X_i,E_{L_z}} M^{n-m}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z}) \sum_z M_{X_i,E_{L_z}} M_{X_k,E_{L_z}} \mathcal{C}(E^2_{L_z})}{\sum_z M^{m+1}_{X_i,E_{L_z}} M^{n-m-1}_{X_k,E_{L_z}} \mathcal{C}(E^n_{L_z})} \\ =& 0 \end{aligned} \tag{13}$$

Therefore, $\forall X_k \in \mathbf{X} \setminus \{X_i, X_j\}$ and vanishing triplet differences hold among $\{X_i, X_j\}$ and X_k.

Case 2) X_i and X_j doesn't belong to the same cluster, which means X_i and X_j have different parental latent variables. Let L_a and L_b be the two latent variables, and let the parent of x_i be L_a and the parent of X_j be L_b. Without loss of generality, let the causal order between latent variables $\{L_a, L_b\}$ be $L_a \prec L_b$. There must $\exists X_k \in \mathbf{X} \setminus \{X_i, X_j\}$ be the child of L_b and $\exists L_z \in \mathbf{L}, M_{X_i,L_z} = 0, M_{X_j,L_z} \neq 0, M_{X_k,L_z} \neq 0$, which leads to $\tau^{(n)}_{(i,j|k)} \neq 0$. Therefore, $\exists X_k \in \mathbf{X} \setminus \{X_i, X_j\}$ and vanishing triplet differences doesn't hold among $\{X_i, X_j\}$ and X_k. □

Theorem 5.4. Under Assumptions 2.1 and 2.2, suppose the data of observed variables $\mathbf{X}$ are generated by Eq. 1. Let L_a and L_b be two directed connected latent variables without confounders, and let $\{X_i, X_j\}$ be two children of L_a. Then if and only if there exist an observed variable X_k that is a child of L_b, such that $\tau^{(n)}_{(i,k|j)} = 0$, then, L_a is an ancestor of L_b.

Proof. Case 1) If L_a is an ancestor of L_b, then we have

$$\begin{aligned} L_a =& E_{L_a}, \\ L_b =& A_{b,a}L_a + E_{L_b} = A_{b,a}E_{L_a} + E_{L_b}, \\ X_i =& B_{X_i,L_a}L_a + E_{X_i} = B_{X_i,L_a}E_{L_a} + E_{X_i}, \\ X_j =& B_{X_j,L_a}L_a + E_{X_j} = B_{X_j,L_a}E_{L_a} + E_{X_j}, \\ X_k =& B_{X_k,L_b}L_b + E_{X_k} = B_{X_k,L_b}A_{b,a}E_{L_a} + B_{X_k,L_b}E_{L_b} + E_{X_k}. \end{aligned} \tag{14}$$

Then the Triplet Difference $\tau^{(n)}_{(i,k|j)}$ can be written as follows:

$$\begin{aligned} \tau^{(n)}_{(i,k|j)} =& \frac{\mathcal{C}(X_i^m, X_j^{n-m})\mathcal{C}(X_i, X_j)}{\mathcal{C}(X_i^{m+1}, X_j^{n-m-1})} - \frac{\mathcal{C}(X_k^m, X_j^{n-m})\mathcal{C}(X_k, X_j)}{\mathcal{C}(X_k^{m+1}, X_j^{n-m-1})}, \\ =& \frac{B^{m+1}_{X_i,L_a}B^{n-m+1}_{X_j,L_a}\mathcal{C}(E^n_{L_a})\mathcal{C}(E^2_{L_a})}{B^{(m+1)}_{X_i,L_a}B^{n-m-1}_{X_j,L_a}\mathcal{C}(E^n_{L_a})} - \frac{B^{m+1}_{X_k,L_b}A^{m+1}_{b,a}B^{n-m+1}_{X_j,L_a}\mathcal{C}(E^n_{L_a})\mathcal{C}(E^2_{L_a})}{B^{m+1}_{X_k,L_b}A^{m+1}_{b,a}B^{n-m-1}_{X_j,L_a}\mathcal{C}(E^n_{L_a})}, \\ =& B^2_{X_j,L_b}\mathcal{C}(E^2_{L_a}) - B^2_{X_j,L_b}\mathcal{C}(E^2_{L_a}), \\ =& 0. \end{aligned} \tag{15}$$

Therefore, $\forall X_k \in ch(L_b)$ and vanishing triplet differences hold among $\{X_i, X_j\}$ and X_k.

Case 2) If L_b is an ancestor of L_a, then we have

$$\begin{aligned} L_b =& E_{L_b}, \\ L_a =& A_{a,b}L_b + E_{L_a} = A_{a,b}E_{L_b} + E_{L_a}, \\ X_i =& B_{X_i,L_a}L_a + E_{X_i} = B_{X_i,L_a}A_{a,b}E_{L_b} + B_{X_i,L_a}E_{L_a} + E_{X_i}, \\ X_j =& B_{X_j,L_a}L_a + E_{X_j} = B_{X_j,L_a}A_{a,b}E_{L_b} + B_{X_j,L_a}E_{L_a} + E_{X_j}, \\ X_k =& B_{X_k,L_b}L_b + E_{X_k} = B_{X_k,L_b}E_{L_b} + E_{X_k}. \end{aligned} \tag{16}$$

Then the Triplet Difference $\tau^{(n)}_{(i,k|j)}$ can be written as follows:

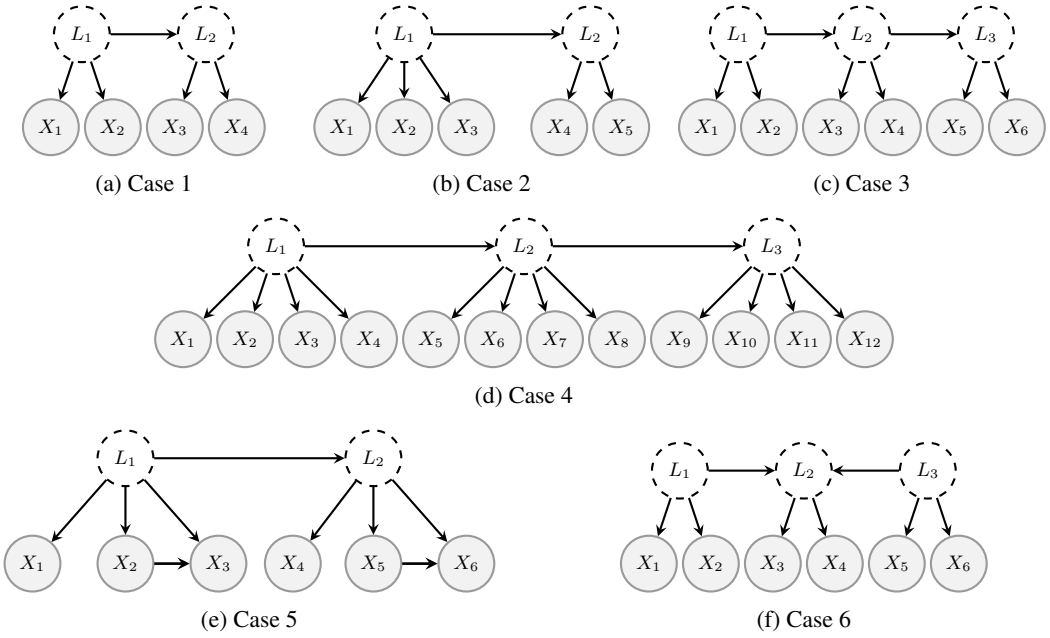
$$\begin{aligned} \tau^{(n)}_{(i,k|j)} =& \frac{\mathcal{C}(X_i^m, X_j^{n-m})\mathcal{C}(X_i, X_j)}{\mathcal{C}(X_i^{m+1}, X_j^{n-m-1})} - \frac{\mathcal{C}(X_k^m, X_j^{n-m})\mathcal{C}(X_k, X_j)}{\mathcal{C}(X_k^{m+1}, X_j^{n-m-1})} \\ =& \frac{B^{m+1}_{X_i,L_a}B^{n-m+1}_{X_j,L_a}\Big[\mathcal{C}(E^n_{L_a}) + A^n_{a,b}\mathcal{C}(E^n_{L_b})\Big]\Big[\mathcal{C}(E^2_{L_a}) + A^2_{a,b}\mathcal{C}(E^2_{L_b})\Big]}{B^{m+1}_{X_i,L_a}B^{n-m-1}_{X_j,L_a}\Big[\mathcal{C}(E^n_{L_a}) + A^n_{a,b}\mathcal{C}(E^n_{L_b})\Big]} \\ & - \frac{B^{m+1}_{X_k,L_b}B^{n-m+1}_{X_j,L_a}A^{n-m+1}_{a,b}\mathcal{C}(E^n_{L_b})\mathcal{C}(E^2_{L_b})}{B^{m+1}_{X_k,L_b}B^{n-m-1}_{X_j,L_a}A^{n-m-1}_{a,b}\mathcal{C}(E^n_{L_b})} \\ =& B^2_{X_j,L_a}\mathcal{C}(E^2_{L_a}) \neq 0. \end{aligned} \tag{17}$$

Therefore, $\exists X_k \in ch(L_b)$ and vanishing triplet differences doesn't hold among $\{X_i, X_j\}$ and X_k. □

B ADDITIONAL INFORMATION ON EXPERIMENTS

All experiments were conducted on a machine equipped with an Intel Xeon E5-2620 v4 CPU and an NVIDIA GeForce RTX 2080 Ti GPU. For each synthetic experimental setting, each method was run independently 10 times, and the results were averaged.

To complement the experiments in the main text, we further evaluate TCD under cubed Gamma and cubed Laplace noise across six graph structures and sample sizes $N \in \{500, 1000, 2000\}$. The measurement-model recovery results in Tables 3 and 5 show that TCD maintains low latent commission, with values no greater than 0.07 under cubed Gamma noise and zero throughout the cubed Laplace experiments, while its relative performance in latent omission and mismeasurement



(a) Case 1 (b) Case 2 (c) Case 3

(d) Case 4

(e) Case 5 (f) Case 6

Figure 3: Causal structures used in the simulation experiments. Solid nodes represent observed variables, while dashed nodes represent latent variables.

varies across settings. For direction recovery, Tables 4 and 6 show that TCD achieves the highest average scores on both reported F_1 metrics across the 18 graph–sample-size combinations for each noise distribution. Specifically, its `direction_true_f1` ranges from 0.90 to 1.00 under cubed Gamma noise and from 0.98 to 1.00 under cubed Laplace noise. Case 5 further evaluates robustness in a setting that violates the structural assumptions of our method. Although its absolute `direction_pre_f1` scores are lower in this setting, TCD outperforms all compared baselines on this metric at every sample size under both noise distributions. These results support the effectiveness of TCD across the evaluated non-Gaussian noise distributions and provide empirical evidence of robustness to the specific structural assumption violation considered in Case 5.

Table 3: Measurement-model recovery results under the cubed Gamma distribution.

Case	n	Latent Omission↓								Latent Commission ↓								Mismeasurement ↓							
		TCD	ETCD	RLCD	LaHiCaSl	GIN	LSTC	FOFC	BPC	**TCD**	ETCD	RLCD	LaHiCaSl	GIN	LSTC	FOFC	BPC	**TCD**	ETCD	RLCD	LaHiCaSl	GIN	LSTC	FOFC	BPC
Case 1	500	0.15	0.10	0.10	0.10	0.40	0.35	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.05	0.00	0.00	0.03	0.00	0.00	0.00
	1000	0.05	0.05	0.00	0.00	0.40	0.20	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00
	2000	0.15	0.00	0.00	0.00	0.30	0.15	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Case 2	500	0.10	0.05	0.10	0.00	0.35	0.35	1.00	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.02	0.02	0.00	0.00	0.00	0.18
	1000	0.05	0.00	0.00	0.00	0.05	0.20	1.00	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18
	2000	0.15	0.00	0.00	0.00	0.25	0.30	1.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.20
Case 3	500	0.17	0.27	0.20	0.07	0.20	0.23	1.00	0.67	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.03	0.15	0.05	0.07	0.10	0.02	0.00	0.32
	1000	0.20	0.00	0.00	0.00	0.30	0.50	1.00	0.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.08	0.00	0.00	0.28
	2000	0.10	0.00	0.00	0.00	0.23	0.40	1.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.03	0.00	0.00	0.32
Case 4	500	0.03	0.07	0.00	0.00	0.00	0.23	1.00	0.50	0.03	0.00	0.03	0.03	0.10	0.03	0.00	0.00	0.04	0.06	0.00	0.00	0.02	0.00	0.00	0.05
	1000	0.00	0.00	0.00	0.00	0.00	0.27	1.00	0.30	0.03	0.00	0.00	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05
	2000	0.00	0.00	0.00	0.00	0.00	0.27	1.00	0.23	0.07	0.00	0.00	0.00	0.13	0.03	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.05
Case 5	500	0.45	0.90	0.15	0.05	0.15	0.55	1.00	0.75	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.13	0.00	0.02	0.07	0.10	0.00	0.00	0.08
	1000	0.20	0.85	0.10	0.00	0.10	0.65	1.00	0.85	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.12	0.00	0.02	0.00	0.03	0.00	0.00	0.05
	2000	0.30	0.90	0.00	0.00	0.20	0.75	1.00	1.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.12	0.00	0.02	0.00	0.07	0.00	0.00	0.00
Case 6	500	0.13	0.33	0.17	0.03	0.30	0.37	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.22	0.00	0.00	0.03	0.00	0.00	0.00
	1000	0.20	0.27	0.00	0.03	0.23	0.43	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.08	0.00	0.00	0.05	0.00	0.00	0.00
	2000	0.07	0.30	0.00	0.00	0.33	0.43	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.07	0.00	0.00	0.00

C Complexity Analysis

In this section, we analyze and compare the computational characteristics of TCD and representative baseline methods for latent-variable causal discovery with respect to the number of observed variables p and the sample size N. Let $T_{\mathrm{CI}}(N)$ and $M_{\mathrm{CI}}(N)$ denote, respectively, the time and memory costs of one independence test, let R denote the number of iterations in Stage 1 of ETCD, and let r denote the number of discovered latent variables.

Table 4: Direction recovery results under the cubed Gamma noise.

Algorithm		direction_pre_f1↑						direction_true_f1↑					
		TCD	ETCD	RLCD	LaHiCaSl	GIN	LSTC	**TCD**	ETCD	RLCD	LaHiCaSl	GIN	LSTC
Case 1	500	0.70	**0.80**	0.20	**0.80**	0.40	0.50	**1.00**	0.90	0.20	0.90	0.90	0.90
	1000	**0.90**	0.80	0.40	**0.90**	0.40	0.60	**1.00**	0.90	0.40	0.90	0.90	0.90
	2000	0.70	**1.00**	0.30	**1.00**	0.30	0.70	**1.00**	**1.00**	0.30	**1.00**	0.90	0.90
Case 2	500	**0.80**	0.70	0.30	**0.80**	0.20	0.40	**1.00**	0.90	0.30	0.80	0.80	**1.00**
	1000	**0.90**	**0.90**	0.40	**0.90**	0.80	0.50	**1.00**	0.90	0.40	0.90	0.90	**1.00**
	2000	0.70	**1.00**	0.30	0.80	0.30	0.40	**1.00**	**1.00**	0.30	0.80	0.80	**1.00**
Case 3	500	**0.77**	0.63	0.37	0.63	0.41	0.44	**1.00**	0.78	0.37	0.87	0.53	0.72
	1000	0.70	**1.00**	0.40	0.92	0.43	0.37	**0.90**	0.88	0.40	**0.90**	0.42	0.76
	2000	0.73	**0.80**	0.45	0.73	0.29	0.36	**0.92**	0.80	0.45	0.73	0.44	0.76
Case 4	500	**0.85**	0.77	0.42	0.77	0.49	0.43	**1.00**	0.80	0.47	0.83	0.60	0.80
	1000	**1.00**	0.80	0.55	0.83	0.42	0.20	**1.00**	0.70	0.55	0.93	0.44	0.80
	2000	**0.90**	**0.90**	0.50	0.88	0.50	0.35	**1.00**	**1.00**	0.50	0.92	0.58	0.80
Case 5	500	**0.20**	0.00	0.10	0.10	0.00	0.00	**1.00**	**1.00**	0.10	0.90	0.00	**1.00**
	1000	**0.40**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.90	0.00	0.80	0.00	**1.00**
	2000	**0.30**	0.00	0.00	0.00	0.00	0.00	**1.00**	**1.00**	0.00	0.90	0.00	**1.00**
Case 6	500	**0.71**	0.43	0.27	**0.71**	0.33	0.26	0.98	**1.00**	0.27	0.78	0.58	0.68
	1000	0.70	0.32	0.20	**0.83**	0.41	0.15	0.97	**1.00**	0.20	0.84	0.50	0.72
	2000	**0.87**	0.25	0.25	0.85	0.33	0.39	**1.00**	**1.00**	0.25	0.85	0.50	0.76

Table 5: Measurement-model recovery results under the cubed Laplace distribution.

Case	n	Latent Omission↓								Latent Commission↓								Mismeasurement↓							
		TCD	ETCD	RLCD	LaHiCaSl	GIN	LSTC	FOFC	BPC	**TCD**	ETCD	RLCD	LaHiCaSl	GIN	LSTC	FOFC	BPC	**TCD**	ETCD	RLCD	LaHiCaSl	GIN	LSTC	FOFC	BPC
Case 1	500	0.15	0.10	0.00	0.10	0.40	0.30	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.08	0.00	0.00	0.00
	1000	0.05	0.05	0.10	0.20	0.20	0.40	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00
	2000	0.20	0.00	0.10	0.10	0.40	0.35	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00
Case 2	500	0.15	0.00	0.00	0.00	0.10	0.10	1.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.20
	1000	0.05	0.00	0.05	0.00	0.20	0.25	1.00	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.18
	2000	0.05	0.00	0.00	0.00	0.10	0.10	1.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.20
Case 3	500	0.17	0.03	0.13	0.03	0.30	0.20	1.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.03	0.03	0.05	0.00	0.00	0.33
	1000	0.13	0.07	0.07	0.07	0.30	0.37	1.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.03	0.05	0.00	0.00	0.27
	2000	0.17	0.03	0.00	0.00	0.17	0.23	1.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.28
Case 4	500	0.13	0.07	0.00	0.00	0.00	0.13	1.00	0.40	0.00	0.00	0.03	0.03	0.06	0.00	0.00	0.00	0.13	0.07	0.01	0.00	0.02	0.00	0.00	0.04
	1000	0.07	0.00	0.00	0.00	0.00	0.07	1.00	0.37	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03
	2000	0.03	0.00	0.00	0.00	0.00	0.27	1.00	0.23	0.00	0.00	0.00	0.07	0.13	0.10	0.00	0.00	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.03
Case 5	500	0.35	0.80	0.05	0.05	0.15	0.75	1.00	0.85	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.20	0.00	0.03	0.07	0.10	0.00	0.00	0.05
	1000	0.40	0.95	0.15	0.05	0.20	0.65	1.00	0.75	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.12	0.00	0.05	0.05	0.13	0.00	0.00	0.08
	2000	0.35	0.90	0.00	0.00	0.00	0.60	1.00	1.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00	0.00	0.03	0.00	0.02	0.02	0.08	0.00	0.00	0.00
Case 6	500	0.17	0.30	0.13	0.03	0.23	0.27	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.07	0.00	0.03	0.10	0.00	0.00	0.00
	1000	0.17	0.13	0.13	0.07	0.27	0.23	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.07	0.03	0.00	0.00	0.00
	2000	0.10	0.10	0.03	0.10	0.27	0.30	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.02	0.00	0.03	0.15	0.00	0.00	0.00

BPC (Silva et al., 2006) and FOFC (Kummerfeld & Ramsey, 2016) mainly rely on covariance-based Tetrad constraints, with worst-case structural search costs of $\mathcal{O}(p^5)$ and $\mathcal{O}(p^4)$, respectively. Triad and LSTC (Cai et al., 2019) examine triplet-wise conditions and require approximately $\mathcal{O}(p^3)$ candidate evaluations per major round, with the cost of each evaluation depending on the adopted independence test. GIN (Xie et al., 2020) searches over subsets of increasing sizes, and its worst-case number of candidate subsets is exponential in p, yielding a search complexity of $\mathcal{O}(2^p T_{\mathrm{CI}}(N))$.

LaHiCaSl (Xie et al., 2024) extends the GIN principle to latent hierarchical structures and repeatedly performs cluster identification, local structure discovery, and causal-order inference. Since these stages involve subset search together with repeated independence tests, its worst-case search cost remains combinatorial in the number of active variables. We therefore express its worst-case time complexity as $\mathcal{O}(2^p T_{\mathrm{CI}}(N))$, while noting that practical pruning can substantially reduce the number of tested subsets.

RLCD (Dong et al., 2024) relies primarily on covariance-rank constraints and proceeds through CI-skeleton discovery, causal-cluster identification, and cluster refinement. The rank statistics themselves can be computed from the empirical covariance matrix, whose construction requires $\mathcal{O}(Np^2)$ time. However, the subsequent structural search considers multiple candidate subsets and overlapping clique structures; hence its worst-case structural search is combinatorial. We summarize its cost as $\mathcal{O}(Np^2 + 2^p T_{\mathrm{rank}}(p))$, where $T_{\mathrm{rank}}(p)$ denotes the cost of one covariance-rank test.

ETCD (Li & Liu, 2025) adopts a sequential two-stage procedure. Its original analysis gives a complexity of $\mathcal{O}(R_E p^3)$ for latent-variable identification in Stage 1 and cubic complexity for causal-relation recovery in Stage 2.

Table 6: Direction recovery results under the cubed Laplace noise.

Algorithm		direction_pre_f1↑						direction_true_f1↑					
		TCD	ETCD	RLCD	LaHiCaSl	GIN	LSTC	**TCD**	ETCD	RLCD	LaHiCaSl	GIN	LSTC
Case 1	500	0.70	**0.90**	0.50	0.80	0.40	0.50	**1.00**	0.90	0.50	0.90	0.70	**1.00**
	1000	**0.90**	**0.90**	0.20	0.60	0.20	0.40	**1.00**	**1.00**	0.20	0.80	0.60	**1.00**
	2000	**0.70**	**0.70**	0.40	**0.70**	0.40	0.40	**1.00**	0.70	0.40	0.80	0.70	**1.00**
Case 2	500	0.70	**0.90**	0.30	**0.90**	0.40	0.70	**1.00**	0.90	0.30	0.90	0.90	**1.00**
	1000	**0.90**	**0.90**	0.40	0.80	0.40	0.50	**1.00**	**1.00**	0.40	0.90	0.80	**1.00**
	2000	**0.90**	0.80	0.40	0.70	0.20	0.80	**1.00**	0.90	0.40	0.70	0.80	**1.00**
Case 3	500	0.77	**0.97**	0.28	0.92	0.35	0.53	**1.00**	0.90	0.28	0.95	0.66	0.80
	1000	**0.80**	0.70	0.27	0.78	0.37	0.46	**1.00**	0.90	0.27	0.87	0.58	0.76
	2000	0.70	0.90	0.55	**0.95**	0.38	0.59	**1.00**	0.90	0.55	0.95	0.47	0.80
Case 4	500	**0.83**	0.77	0.50	0.67	0.53	0.51	**1.00**	0.88	0.50	0.77	0.65	0.80
	1000	**0.87**	0.70	0.39	0.76	0.50	0.29	**1.00**	0.70	0.45	0.88	0.53	0.76
	2000	0.97	**1.00**	0.50	0.86	0.57	0.40	**1.00**	**1.00**	0.50	0.87	0.64	0.80
Case 5	500	**0.30**	0.00	0.20	0.10	0.00	0.00	**1.00**	**1.00**	0.10	0.80	0.00	**1.00**
	1000	**0.30**	0.00	0.00	0.00	0.00	0.00	**1.00**	**1.00**	0.00	**1.00**	0.00	**1.00**
	2000	**0.50**	0.00	0.00	0.10	0.00	0.00	**1.00**	**1.00**	0.00	0.80	0.00	**1.00**
Case 6	500	**0.77**	0.35	0.10	0.61	0.44	0.33	**1.00**	**1.00**	0.20	0.71	0.69	0.68
	1000	0.67	0.65	0.25	**0.83**	0.36	0.36	**1.00**	0.90	0.25	0.86	0.61	0.72
	2000	**0.77**	0.43	0.30	0.71	0.51	0.43	**0.98**	0.90	0.30	0.89	0.77	0.80

TCD instead evaluates triplet-wise algebraic constraints after estimating the required low-order cumulants. With cached cumulants, causal-cluster discovery and latent-order recovery require

$$\mathcal{O}\big(Np^2 + p^3\big)\,.$$

After obtaining the latent order, TCD removes redundant latent edges by testing conditioning subsets over the predecessors of each latent variable. In the unrestricted worst case, this step incurs an additional $\mathcal{O}(2^r r^5)$ cost, giving an overall worst-case complexity of

$$\mathcal{O}\big(Np^2 + p^3 + 2^r r^5\big)\,.$$

Under our assumption that each latent variable has at least two observed children, $r \leq \lfloor p/2 \rfloor$, and in typical settings $r \ll p$. Therefore, the combinatorial search in TCD is performed only in the much smaller latent space rather than over all observed variables.

The considered methods also exhibit substantially different memory-scaling behaviors with respect to N. Covariance-, rank-, and cumulant-based statistics can be computed by scanning the samples and storing the data together with low-dimensional summary statistics. In contrast, methods whose implementations rely on kernel-based independence tests, such as HSIC, may require the construction of $N \times N$ Gram matrices for each test, resulting in $\mathcal{O}(N^2)$ memory usage. This is particularly relevant for Triad, LSTC, GIN, and LaHiCaSl when HSIC-based independence tests are used. Consequently, repeated kernel-based independence testing can become a major computational and memory bottleneck as the sample size increases. By contrast, TCD avoids repeated kernel-based independence tests and requires only the observed data and cached low-order cumulant statistics.

D Related Work

D.1 Causal Discovery with Latent Variables

Classical causal discovery methods mainly study causal relations among observed variables. Constraint-based methods, such as PC (Spirtes & Glymour, 1991), recover causal structures from conditional independence relations, while score-based methods, such as GES (Chickering, 2002), search over Markov equivalence classes. Functional causal models further exploit distributional asymmetries to improve identifiability, including ANM (Hoyer et al., 2008a) and LiNGAM (Shimizu et al., 2006).

In the presence of latent variables, FCI (Spirtes et al., 2013) and RFCI (Colombo et al., 2012) characterize identifiable causal relations among observed variables, but do not explicitly recover the latent variables or their internal causal structure. Another line of work exploits rank constraints induced by latent structures. Tetrad constraints (Spearman, 1928) and trek-separation (Spirtes, 2013)

Table 7: Computational complexity and memory characteristics of representative latent-variable causal discovery methods.

Method	Time complexity	Space complexity
BPC	$\mathcal{O}(Np^2 + p^5)$	$\mathcal{O}(Np + p^2)$
FOFC	$\mathcal{O}(Np^2 + p^4)$	$\mathcal{O}(Np + p^2)$
Triad	$\mathcal{O}\big(p^3 T_{\mathrm{CI}}(N)\big)$	$\mathcal{O}(Np + M_{\mathrm{CI}}(N))$
LSTC	$\mathcal{O}\big(R_L p^3 T_{\mathrm{CI}}(N)\big)$	$\mathcal{O}(Np + M_{\mathrm{CI}}(N))$
GIN	$\mathcal{O}(2^p T_{\mathrm{CI}}(N))$	$\mathcal{O}(Np + M_{\mathrm{CI}}(N))$
LaHiCaSl	$\mathcal{O}(2^p T_{\mathrm{CI}}(N))$	$\mathcal{O}(Np + M_{\mathrm{CI}}(N))$
RLCD	$\mathcal{O}\big(Np^2 + 2^p T_{\mathrm{rank}}(p)\big)$	$\mathcal{O}(Np + p^2)$
ETCD	$\mathcal{O}\big(R_E p^3 + m^3\big)$	$\mathcal{O}(Np + M_{\mathrm{CI}}(N))$
TCD	$\mathcal{O}\big(Np^2 + p^3 + 2^r r^5\big)$	$\mathcal{O}(Np + p^2)$

p denotes the number of observed variables and N the sample size. R_L denotes the number of recursive rounds in LSTC, R_E the number of iterations in Stage 1 of ETCD, and m the size of the active variable set entering Stage 2 of ETCD. r denotes the number of discovered latent variables; under the assumption that each latent variable has at least two observed children, $r \leq \lfloor p/2 \rfloor$. $T_{\mathrm{CI}}(N)$ and $M_{\mathrm{CI}}(N)$ denote the time and memory costs of one independence test, respectively, and $T_{\mathrm{rank}}(p)$ denotes the cost of one covariance-rank test. The exponential terms for GIN, LaHiCaSl, and RLCD represent worst-case subset-search behavior over observed variables, whereas the 2^r term of TCD arises only from latent-edge pruning over the considerably smaller latent space. Practical implementations may terminate substantially earlier due to pruning, sparsity, and early stopping.

motivate methods such as BPC (Silva et al., 2006) and FOFC (Kummerfeld & Ramsey, 2016), which recover groups of observed variables sharing latent causes under suitable measurement assumptions.

Recent methods consider substantially more general latent structures. RLCD (Dong et al., 2024) uses covariance-rank information to locate latent variables and recover causal relations involving both observed and causally related hidden variables. ETCD (Li & Liu, 2025) further allows complex relations among latent and observed variables and develops an efficient structure-learning procedure under a pure-children assumption. These methods substantially extend classical measurement-model settings, but primarily rely on second-order rank information and corresponding structural assumptions.

D.2 Non-Gaussian Independence-Based Methods

Non-Gaussianity provides additional information beyond covariance for latent causal discovery. Overcomplete ICA has been used to recover linear mixing structures in the presence of latent sources (Hoyer et al., 2008b; Shimizu et al., 2009), although the recovered mixing model may still correspond to multiple observationally equivalent causal structures (Entner & Hoyer, 2010; Tashiro et al., 2014).

The Triad condition (Cai et al., 2019) exploits independence between pseudo-residuals and observed variables to identify latent clusters and causal relations. The Generalized Independent Noise (GIN) condition (Xie et al., 2024) extends this idea to sets of variables by searching for linear combinations that are independent of another variable set, enabling the identification of latent variables and causal relations in hierarchical linear non-Gaussian models. Related independence-based approaches have substantially improved the identifiability of latent causal structures.

However, these methods often require independence testing over many candidate variables or variable subsets. In particular, kernel-based independence tests may be computationally expensive for large sample sizes, and their finite-sample decisions can be sensitive to test parameters such as kernel bandwidths and significance thresholds. Our work instead seeks observable algebraic constraints that can be evaluated directly without repeated independence testing.

D.3 Higher-Order Cumulants for Causal Discovery

Higher-order cumulants provide a complementary way to exploit non-Gaussianity. Unlike covariance, they preserve higher-order distributional information and can yield explicit algebraic constraints among observed variables. Cai et al. (2023) introduced the One-Latent-Component condition and derived closed-form cumulant relations for identifying latent structures and causal orders. Chen et al. (2024) further studied settings where observed variables may share a latent cause while also being directly causally related. More recently, Chen et al. (2025) proposed a cause-difference criterion based on products of joint cumulants for identifying causal relations between observed variables.

Higher-order cumulants have been applied to causal discovery in a variety of settings. For spatiotemporal variance dependencies, Chen et al. (2012) proposed a probabilistic causal scale-mixture model that employs a structural vector autoregressive (SVAR) model to characterize contemporaneous and lagged causal relationships among latent scale variables, while using fourth-order cross-cumulants to identify the contemporaneous causal ordering. In the presence of measurement error, Dai et al. (2022) leveraged non-Gaussianity and higher-order statistical information to formulate the Transformed Independent Noise (TIN) condition. Their method identifies an ordered group decomposition of the latent target variables through independence testing rather than by directly solving algebraic constraints involving cumulants. For Poisson branching structural causal models (PB-SCMs), Qiao et al. (2024) established a connection between higher-order cumulants and directed-path information, thereby identifying causal directions that cannot be distinguished using second-order statistics alone. In federated causal discovery, Chen et al. (2026) exploited the fact that cross-cumulants depend only on the joint distribution of the variables involved. They aggregated locally estimated cross-cumulant matrices across clients to recover the global causal ordering and causal strengths. Collectively, these studies demonstrate that higher-order statistical information captures non-Gaussian structural properties beyond covariance and facilitates causal discovery in settings involving spatiotemporal variance dependencies, measurement error, count data, and federated data.

Existing cumulant-based methods typically derive criteria for specific local causal configurations. A single cumulant equality, however, may be shared by multiple candidate structures when latent variables are causally related or when direct observed-to-observed effects are present. Our work therefore shifts the focus from an individual cumulant equality to the *pattern of satisfied constraints across multiple variable combinations*. We introduce the Triplet Difference as a common algebraic primitive and use its support pattern to distinguish latent structures. This yields a unified approach to latent-cluster discovery and subsequent causal-structure recovery without relying on repeated independence tests.