Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Conditional Independent Test in the Presence of Measurement Error with Causal Structure Learning

Hongbing Zhang[2], Kezhou Chen[2], Nankai Lin[3], Aimin Yang[2], Zhifeng Hao[1], Zhengming Chen[1,2]*

[1]College of Science, Shantou University, Shantou, Guangdong, China
[2]School of Computer Science, Guangdong University of Technology, Guangzhou, China
[3]School of Information Science and Technology, Guangdong University of Foreign Studies, China
zhbin@mail2.gdut.edu.cn, chenkz@mail2.gdut.edu.cn, neakail@outlook.com, amyang18@163.com, haozhifeng@stu.edu.cn, chenzhengming1103@gmail.com

Abstract

Testing conditional independence is a critical task, particularly in causal discovery and learning in Bayesian networks. However, in many real-world scenarios, variables are often measured with errors, such as those introduced by insufficient measurement accuracy, complicating the testing process. This paper focuses on testing conditional independence in the linear non-Gaussian measurement error model, under the condition that measurement error noise follows a Gaussian distribution. By leveraging high-order cumulants, we derive rank constraints on the cumulant matrix and establish their role in effectively assessing conditional independence, even in the presence of measurement errors. Based on these theoretical results, we leverage the rank constraints of the cumulant matrix as a tool for conditional independence testing and incorporate it into the PC algorithm, resulting in the PC-ME algorithm — a method designed to learn causal structures from observed data while accounting for measurement errors. Experimental results demonstrate that the proposed method outperforms existing approaches, particularly in cases other methods encounter difficulties.

1 Introduction

The conditional independence (CI) test aims to assess whether a correlation relation exists between two variables when conditioned on a set of other variables. It has been broadly applied across various fields like statistics, machine learning, and causal discovery [Zhou *et al.*, 2020; Watson and Wright, 2021; Bouezmarni and Taamouti, 2014; Lundborg *et al.*, 2022; Spirtes *et al.*, 2000].

When the data model is presumed to be linear, the partial correlation test [Baba *et al.*, 2004] is a well-known and effective tool for examining conditional independence relations among variables. Specifically, let X_i and X_j be two random variables and $\mathbf{X}_p$ be a vector, if the CI relation between X_i and X_j given $\mathbf{X}_p$ holds, i.e., $X_i \perp\!\!\!\perp X_k|\mathbf{X}_p$, the partial correlation coefficient would be zero, denoted by $\rho_{X_iX_j.\mathbf{X}_p} = 0$. Thus, one can examine whether the partial correlation coefficient is zero to discover the CI relations among observed variables. Obtaining these CI relations usually can be used to improve downstream tasks, such as constructing causal structures over observed variables, where the CI relations correspond to d-separation in the causal graphical model [Spirtes and Glymour, 1991; Spirtes *et al.*, 2000].

However, due to uncertainties in the real environment, the variable of interest is often difficult to measure directly and instead collected under the existence of measurement errors [Totton and White, 2011; Fuller, 2009; Kelly, 2007]. An example of measurement error is depicted in Fig. 1, where X_i, X_j, and X_k represent hidden variables of interest. When there exists some disturbing noise that directly causes the observed variable, one can only observe the variables $\tilde{X}_i$, $\tilde{X}_j$, and $\tilde{X}_k$ (known as *measured variables*). Since the CI relations only hold on $X_i \perp\!\!\!\perp X_j|X_k$, in this case, the partial correlation coefficient $\rho_{\tilde{X}_i\tilde{X}_j.\tilde{X}_k} \neq 0$, implying the CI relations of $X_i \perp\!\!\!\perp X_j|X_k$ can not be tested from $\tilde{X}_i$, $\tilde{X}_j$, and $\tilde{X}_k$ (further details can be found in Example 2). Such a problem will lead to uninformative relations discovery or incorrect causal structure learning. Therefore, it is necessary to investigate how to learn the CI relations among hidden variables only from their measured variables.

Measurement error has been a topic of increasing interest in causal discovery. For instance, [Zhang *et al.*, 2018] introduces a method for inferring causal structures in the presence of measurement error by leveraging non-Gaussianity, showing that the causal structure can be identified up to an ancestral ordered grouping (AOG) equivalence class. Similarly, [Yang *et al.*, 2022] exploits the sparsity of the mixing matrix to propose a method for identifying linear latent variable models affected by measurement error. These identifiability results mainly rely on overcomplete independent component analysis (ICA), which, despite its theoretical appeal, often encounters computational challenges and risks convergence to local optima. Recently, [Dai *et al.*, 2022] introduced the TIN condition to capture causal structure patterns in linear non-Gaussian models, proving that the partially causal or-

*Corresponding author.

der can be identified even in the presence of measurement error. While this approach effectively verifies d-separation relations (i.e., conditional independence) among hidden variables, it imposes structural constraints, such as the requirement that each hidden variable must have at least two observed variables as children. This assumption significantly restricts the applicability of the TIN condition in scenarios involving measurement error. Overall, existing methods struggle to efficiently identify conditional independence (CI) relations among latent variables solely from the observed variables without imposing additional assumptions. As shown in Fig. 1, for instance, existing methods may fail to effectively test the $X_i \perp\!\!\!\perp X_j|X_k$ by only using measured variables $\tilde{X}_i, \tilde{X}_j$ and $\tilde{X}_k$.

In this paper, we address the problem of testing conditional independence (CI) relations among hidden variables only from their measured variables, under the linear non-Gaussian measurement error model. We find out a sufficient condition—namely, the partial Gaussian noise assumption, where the noise term of the observed variables follows a Gaussian distribution—that enables the identification of d-separation relations from the high-order statistics of the hidden variables. Specifically, we first establish an equivalence between partial correlation and the rank deficiency of the correlation coefficient matrix. We then prove that rank constraints on the covariance matrix can be generalized to the higher-order cumulant matrix, enabling the identification of d-separation relations among latent variables. Based on this theoretical result, we propose a structure learning algorithm, the PC-ME algorithm, to infer causal structures among latent variables from their measured variables. Furthermore, we show that the PC-ME algorithm can identify the causal structure up to a Markov equivalence class.

Our contributions are summarised as follows:

- We generalize partial correlation to the case of measurement error and study the properties of rank deficiency of the higher-order cumulant matrix, under the linear non-Gaussian measurement model and the partial Gaussian noise assumption.
- We propose an efficient algorithm to infer causal structures among hidden variables up to a Markov equivalence class by leveraging the properties of rank deficiency in the higher-order cumulant matrix.
- We empirically demonstrate that the proposed algorithm can asymptotically recover the correct causal structure under mild assumptions.

2 Linear Non-Gaussian Measurement Error Model

In this paper, we focus on the problem of testing conditional independence in the presence of measurement error, specifically within the framework of linear non-Gaussian models. Our theoretical results are built upon causal graphical models. In a causal graph, we use $Pa(V_i) = \{V_j|V_j \rightarrow V_i\}$, $Ch(V_i) = \{V_j|V_i \rightarrow V_j\}$ to denote the set of parents and children of V_i, respectively. Due to space limitation, some graphical notations, such as d-separation and V-structure, are used here without explicit definitions, which can be found in the standard literature [Spirtes *et al.*, 2000].

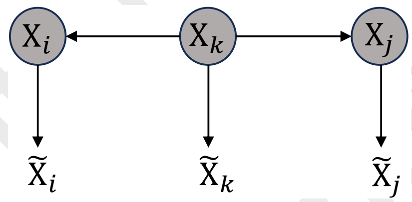

Figure 1: An example of the linear non-Gaussian measurement error model, where X_i, X_j and X_k are hidden variables, $\tilde{X}_i, \tilde{X}_j$ and $\tilde{X}_k$ are observed variables. This paper aims to test the conditional independent relations, e.g., $X_i \perp\!\!\!\perp X_j|X_k$, by only using the observed variables $\tilde{X}_i, \tilde{X}_j$ and $\tilde{X}_k$.

We start with the definition of the linear non-Gaussian model. Let $\mathbf{X} = \{X_1, \cdots, X_n\}$ denote the variable set of interest, in the *linear non-Gaussian model*, the data generation process for $X_i \in \mathbf{X}$ is given by the following equation:

$$X_i = \sum_{X_j \in Pa_{X_i}} b_{ij} X_j + n_i, \tag{1}$$

where Pa_{X_i} is the parent set of X_i, b_{ij} is the causal effect from X_j to X_i, and n_i represents noise terms of X_i that follows non-Gaussian distribution. Without loss of generality, we assume all noise terms are independent of each other and the mean of noise is zero.

In real-world scenarios, variables of interest are often difficult to observe directly (commonly referred to as unobserved or hidden variables) and are instead measured with additive measurement error. To formalize the existence of measurement error, we present the linear non-Gaussian Measurement error Model (LiNGMME) as follows.

Definition 1 (LiNGMME). *Let $\mathbf{X}$ be the hidden variable of interest that follows the linear non-Gaussian model, and $\tilde{\mathbf{X}}$ be the measured variable set of $\mathbf{X}$, we say $\tilde{\mathbf{X}}$ follows the LiNGMME model if for $\tilde{X}_i$, $\tilde{X}_i \in \tilde{\mathbf{X}}$, it is generated by:*

$$\tilde{X}_i = \beta_{X_i \rightarrow \tilde{X}_i} X_i + e_i, \tag{2}$$

where $X_i \in \mathbf{X}$ is hidden variable, $\beta_{X_i \rightarrow \tilde{X}_i}$ is disturbing strength from X_i to $\tilde{X}_i$, and e_i are measurement error term.

Recently, the linear non-Gaussian model with measurement error has been extensively studied, such as [Yang *et al.*, 2022; Zhang *et al.*, 2018; Dai *et al.*, 2022]. These works usually assume that all noise term, including those associated with hidden variables and measured variables, follow a non-Gaussian distribution. Under this assumption, the causal structure can be identified up to an Ancestral Ordered Grouping (AOG) equivalence class [Zhang *et al.*, 2018]. In this paper, we explore a different condition where the noise term of the measured variables follows a Gaussian distribution, which enables the identifiability of conditional independence relations among hidden variables.

Condition 1 (Partial Gaussian Condition). *In LiNGMME, the noise term of the measured variable e_i follows a Gaussian distribution.*

In general, Condition 1 is reasonable and holds in most cases, as the Gaussian distribution is ubiquitous and commonly observed in many real-world scenarios, as discussed in [Totton and White, 2011; Fuller, 2009; Kelly, 2007].

Objective. In this paper, assuming Condition 1 holds, our goal is to investigate how to identify the conditional independence (CI) relations among hidden variables $\mathbf{X}$ only using their measured variables $\tilde{\mathbf{X}}$, under the LiNGAMME.

3 Conditional Independence Test Under Measurement Error

To achieve our goal, in this section, we provide an algebraic solution for identifying the conditional independence (CI) relations among hidden variables. First, we revisit partial correlation testing, a well-known method for testing CI, and explain why it fails in the presence of measurement error (Sec. 3.1). Next, we derive an equivalent expression for partial correlation from an algebraic perspective, specifically through the rank deficiency of the covariance matrix, and demonstrate how this can be extended to rank constraints on the higher-order cumulant matrix. By exploring these properties, we prove that the CI relations among latent variables can be identified (Sec. 3.2).

3.1 Partial Correlation Test

In a linear model, such as the linear Gaussian model (i.e., noise term follow Gaussian distribution in Eq. (1)), partial correlation is a commonly used tool for testing conditional independence, and is specifically defined as follows:

Definition 2 (**P**artial **C**orrelation)**.** *The partial correlation coefficient of random variables X_i and X_j given $\mathbf{X}_p$, is denoted as*

$$\rho_{X_iX_j\cdot\mathbf{X}_p} = \frac{\rho_{X_iX_j} - \rho_{X_i\mathbf{X}_p}\cdot\rho_{X_j\mathbf{X}_p}}{\sqrt{(1-\rho_{X_i\mathbf{X}_p})\cdot(1-\rho_{X_j\mathbf{X}_p})}}, \tag{3}$$

where $\rho_{X_iX_j}$, $\rho_{X_i\mathbf{X}_p}$, and $\rho_{X_j\mathbf{X}_p}$ represent the correlation coefficients of X_i and X_j, X_i and $\mathbf{X}_p$, X_j and $\mathbf{X}_p$, respectively

Specifically, CI relations can be examined by checking whether the partial correlation coefficient is zero. For instance, consider the structure in Fig. 1, where the CI relation $X_i \perp\!\!\!\perp X_j|X_k$ holds. This can be tested by checking whether the partial correlation coefficient $\rho_{X_iX_j\cdot X_k}$ is zero. However, in the presence of measurement error, only $\tilde{X}_i, \tilde{X}_j$ and $\tilde{X}_k$ are observed. In this case, the partial correlation test cannot reliably identify the CI relations among the latent variables X_i and X_j. Next, we give an example to illustrate this.

Illustrative Example

Consider the graph in Fig. 1, where X_i, X_j and X_k are hidden variables and $\tilde{X}_i$, $\tilde{X}_j$, and $\tilde{X}_k$ are observed variables that are measured with measurement error. According to the definition of LiNGAMME, the data generation processes are

$$\begin{aligned}\tilde{X}_k &= \beta_{X_k\to\tilde{X}_k}n_k + e_k,\\ \tilde{X}_i &= \beta_{X_i\to\tilde{X}_i}(b_{ik}n_k + n_i) + e_i,\\ \tilde{X}_j &= \beta_{X_j\to\tilde{X}_j}(b_{jk}n_k + n_j) + e_j.\end{aligned} \tag{4}$$

To construct the CI test using partial correlation, it is sufficient to check whether the numerator is zero. Specifically, each term of the numerator of the partial correlation coefficient between $\tilde{X}_i$ and $\tilde{X}_j$ given $\tilde{X}_k$ is as follows.

$$\begin{aligned}\rho_{\tilde{X}_i,\tilde{X}_j} &= \beta_{X_i\to\tilde{X}_i}\beta_{X_j\to\tilde{X}_j}\cdot\frac{b_{ik}b_{jk}Var(n_k)}{Var(\tilde{X}_i)Var(\tilde{X}_j)},\\ \rho_{\tilde{X}_i,\tilde{X}_k}\rho_{\tilde{X}_j,\tilde{X}_k} &= \beta_{X_i\to\tilde{X}_i}\beta_{X_j\to\tilde{X}_j}\beta^2_{X_k\to\tilde{X}_k}\cdot\\ &\quad\frac{b_{ik}b_{jk}Var^2(n_k)}{Var(\tilde{X}_i)Var(\tilde{X}_j)Var^2(\tilde{X}_k)}.\end{aligned} \tag{5}$$

Without loss of generality, assume that all observed variables are standardized, i.e., have unit variance. If the numerator of the partial correlation coefficient is zero, then we have

$$1-\beta^2_{X_k\to\tilde{X}_k}Var(n_k) = 0, \tag{6}$$

which **does not** hold in most cases. Note that if the above holds, it implies that either no measurement error exists or the measurement error can be ignored. There are two factors that prevent the partial correlation test from working: one is the disturbing strength from X_k to $\tilde{X}_k$, and the other is the measurement error noise e_k (or, $\tilde{X}_k$). Generally speaking, without further constraints, the partial correlation test cannot be used in the presence of measurement error.

3.2 Rank Deficiency of High-order Cumulants Matrix

The above example illustrates why partial correlation does not work in the presence of measurement error: there exists disturbing strength $\beta_{X_k\to\tilde{X}_k}$ and an additional noise term e_k, which causes the vanishing test to fail. To tackle this issue, fortunately, we have found that these unexpected influences can be mitigated under Condition 1, by appropriately utilizing the properties of higher-order cross-cumulants.

Before introducing the main result, we first provide an equivalent representation of partial correlation from the algebraic perspective, used as the building block for later results. We begin with the following example.

Example 1 (Equivalent representation of partial correlation)**.** *Consider the graph in Fig. 1, where the CI relationship $X_i \perp\!\!\!\perp X_j|X_k$ holds. This CI relation can be tested by checking whether the partial correlation coefficient $\rho_{X_iX_j\cdot X_k}$ is zero, i.e., $\rho_{X_iX_j} - \rho_{X_iX_k}\cdot\rho_{X_kX_j} = 0$ holds. Without loss generality, assume X_i, X_j and X_k are all standardized, so that $\rho_{X_kX_k} = 1$. The vanishing of the partial correlation can then be rewritten as:*

$$\rho_{X_iX_j}\rho_{X_kX_k} - \rho_{X_iX_k}\cdot\rho_{X_kX_j} = 0. \tag{7}$$

Such an equality can be transferred to the determinant of the correlation coefficient matrix between $\{X_i, X_k\}$ and $\{X_j, X_k\}$, as shown below:

$$\begin{vmatrix}\rho_{X_iX_j} & \rho_{X_iX_k}\\ \rho_{X_jX_k} & \rho_{X_kX_k}\end{vmatrix} = \rho_{X_iX_j}\rho_{X_kX_k} - \rho_{X_iX_k}\cdot\rho_{X_kX_j} = 0, \tag{8}$$

which means that the correlation coefficient matrix is not full rank, i.e., rank deficiency.

In our work, we refer to this algebraic property as a "rank constraint," following the terminology used in the original literature [Sullivant *et al.*, 2010; Huang *et al.*, 2022]. We further present the following lemma to establish the connection between the CI relation and the rank constraints of the coefficient matrix. In other words, the relationship between partial correlation and the rank deficiency of the coefficient matrix.

Lemma 1 (Rank constraints of covariance matrix [Sullivant *et al.*, 2010])**.** *Assume X_i, X_j and $\mathbf{X}_p$ follow the linear Gaussian model, and let $\mathbf{Y} = (X_i, \mathbf{X}_p)$ and $\mathbf{Z} = (X_j, \mathbf{X}_p)$ represent two vectors, then $X_i \perp\!\!\!\perp X_j | \mathbf{X}_p$ if and only if the rank of correlation coefficient matrix between $\mathbf{Y}$ and $\mathbf{Z}$ is $|\mathbf{X}_p|$.*

Lemma 1 provides an equivalent formalization of the partial correlation test, where the rank constraint of the covariance matrix between two vectors holds, with the intersection set of the two vectors serving as the conditional set in the CI relation. It is important to note that, when all variables are standardized, the rank constraints of the coefficient matrix are equivalent to those of the covariance matrix.

Recall the illustrative example in Sec 3.1. Let $\mathbf{Y} = \{\tilde{X}_i, \tilde{X}_k\}$ and $\mathbf{Z} = \{\tilde{X}_j, \tilde{X}_k\}$, assuming all variables are standardized. In most cases, the correlation coefficient (covariance) matrix between $\mathbf{Y}$ and $\mathbf{Z}$ has full rank due to the influence of measurement error. Under Condition 1, the key idea to address this issue is the fact that the higher-order cross-cumulants of Gaussian noise are zero. Based on this observation, one may wonder whether the rank constraints of the covariance matrix can be extended to a higher-order cross-cumulants matrix, thereby removing the influence of the Gaussian noise term of the measured variables. Before delving into this fact, we first revisit the definition of cumulants and some fundamental properties.

Definition 3 (k-th order joint cumulant tensor)**.** *The k-th order joint cumulant tensor of a random vector $X = [X_1, ..., X_n]^T$ is the k-way tensor $\mathcal{T}_X^{(k)}$ in $R^{n\times\cdots\times n} \equiv (R^n)^k$ whose entry in $(i_1, ..., i_k)$ is the joint cumulant:*

$$\mathcal{T}_{X\ i_1,\ldots,i_k}^{(k)} = \mathrm{cum}(X_{i_1}, \ldots, X_{i_k}) := \sum_{(B_1,\ldots,B_L)} (-1)^{L-1}(L-1)!\mathbb{E}\Big[\prod_{j\in B_1} X_j\Big]\cdots\mathbb{E}\Big[\prod_{j\in B_L} X_j\Big], \tag{9}$$

where the sum is taken over all partitions $(B_1, \ldots, B_L)$ of the multiset $\{i_1, ..., i_k\}$.

Remark 1. *Consider $k = 4$, one can see that the four-order cumulant $\mathrm{cum}(X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4})$ is defined as*

$$\begin{aligned}\mathbb{E}[X_{i_1}X_{i_2}X_{i_3}X_{i_4}] &- \mathbb{E}[X_{i_1}X_{i_2}]\mathbb{E}[X_{i_3}X_{i_4}] \\ &- \mathbb{E}[X_{i_1}X_{i_3}]\mathbb{E}[X_{i_2}X_{i_4}] \\ &- \mathbb{E}[X_{i_1}X_{i_4}]\mathbb{E}[X_{i_2}X_{i_3}].\end{aligned}$$

That is, the fourth-order cumulant is not equal to the fourth-order moment. Specifically, all non-fourth-order interaction terms are absent in the fourth-order cumulant.

One can see that the second-order cumulant is the covariance, and higher-order cumulants can be viewed as an extension of covariance, where certain properties, such as additivity, are preserved. This motivates us to further investigate the matrix of cross-cumulants, which corresponds to the 2D slice of the joint cumulant tensor.

Definition 4 (2D slice of joint cumulant tensor)**.** *For a random vector $\mathbf{X}$ with k-th order joint cumulant tensor $\mathcal{T}_X^{(k)}$ where $k \geq 2$ and k is any even, denote its 2D matrix slice of k-th order joint cumulant tensor as $\mathcal{C}^{(k)}$, where*

$$\mathcal{C}_{i,j}^{(k)} := \mathrm{cum}(\underbrace{X_i, \cdots, X_j}_{\frac{k}{2}\text{ times}}, \underbrace{X_j, \cdots, X_j}_{\frac{k}{2}\text{ times}}) \tag{10}$$

In particular, when $k = 2$, for a random vector $\mathbf{X}$, $\mathcal{C}_{\mathbf{X}}^{(2)}$ is the covariance matrix $\Sigma_{\mathbf{X}}$. Similar to Lemma 1, one can construct the cumulant matrix between two vectors $\mathbf{Y}$ and $\mathbf{Z}$, as shown in the following example.

Example 2 (K-order cumulant matrix)**.** *For two vertices sets $\mathbf{Z} = \{X_1, X_2, \cdots, X_n\}, \mathbf{Y} = \{X_{n+1}, X_{n+2}, \cdots, X_{2n}\}$, let $k \geq 2$, the k-order cumulant matrix between $\mathbf{Z}$ and $\mathbf{Y}$ is*

$$\mathcal{C}_{\mathbf{Y};\mathbf{Z}}^{(k)} := \begin{pmatrix} \mathcal{C}_{X_1,X_{n+1}}^{(k)} & \cdots & \mathcal{C}_{X_1,X_{2n}}^{(k)} \\ \vdots & \mathcal{C}_{X_i,X_{n+j}}^{(k)} & \vdots \\ \mathcal{C}_{X_n,X_{n+1}}^{(k)} & \cdots & \mathcal{C}_{X_n,X_{2n}}^{(k)} \end{pmatrix}. \tag{11}$$

Next, we will show that in the LiNGAMME, the CI relations among hidden variables can be tested by checking the rank constraints over the cumulant matrix of observed variables, as shown in Theorem 1.

Theorem 1 (Rank constraints of cumulant matrix)**.** *For hidden variables X_i, X_j, and a vector $\mathbf{X}_p$, assume that their observed variables $\tilde{X}_i$, $\tilde{X}_j$, $\tilde{\mathbf{X}}_p$ follow the LiNGMME and any k-order cumulant among them exists. Let $\mathbf{Y} = \{\tilde{X}_i, \tilde{\mathbf{X}}_p\}$ and $\mathbf{Z} = \{\tilde{X}_j, \tilde{\mathbf{X}}_p\}$. Then, under Condition 1, $X_i \perp\!\!\!\perp X_j | \mathbf{X}_p$ generic hold if and only if the rank of k-order cumulant matrix between $\mathbf{Z}$ and $\mathbf{Y}$ is $|\mathbf{X}_p|$, i.e.,* Rank*($\mathcal{C}_{\mathbf{Y};\mathbf{Z}}^{k}$) = $|\mathbf{X}_p|$.

Remark 2. *One can see that Lemma 1 is a special case (when $k = 2$) of Theorem 1. The key difference is that, while CI relations cannot be identified using the rank of the covariance matrix, they can be tested by Theorem 1 under Condition 1.*

We further provide an example to show how CI relations can be tested using the rank constraints of the cumulant matrix, even when the partial correlation test fails.

Example 3 (Testing CI relations using Theorem 1)**.** *Continuous the illustrative example in Sec. 3.1. Let $\mathbf{Y} = \{\tilde{X}_i, \tilde{X}_k\}$ and $\mathbf{Z} = \{\tilde{X}_j, \tilde{X}_k\}$, assuming the 4-order cross-cumulant exists. Then the cumulants matrix between $\mathbf{Y}$ and $\mathbf{Z}$ is*

$$\mathcal{C}_{\mathbf{Y};\mathbf{Z}}^{(4)} := \begin{pmatrix} \mathcal{C}_{X_i,X_j}^{(4)} & \mathcal{C}_{X_i,X_k}^{(4)} \\ \mathcal{C}_{X_j,X_k}^{(4)} & \mathcal{C}_{X_k,X_k}^{(4)} \end{pmatrix}, \tag{12}$$

where

$$\begin{aligned} \mathcal{C}_{X_i,X_j}^{(4)} &= \beta_{X_i\to\tilde{X}_i}^2 \beta_{X_j\to\tilde{X}_j}^2 b_{ik}^2 b_{jk}^2 \mathbb{E}(n_k^4), \\ \mathcal{C}_{X_i,X_k}^{(4)} &= \beta_{X_i\to\tilde{X}_i}^2 \beta_{X_k\to\tilde{X}_k}^2 b_{ik}^2 \mathbb{E}(n_k^4), \\ \mathcal{C}_{X_j,X_k}^{(4)} &= \beta_{X_j\to\tilde{X}_j}^2 \beta_{X_k\to\tilde{X}_k}^2 b_{jk}^2 \mathbb{E}(n_k^4), \\ \mathcal{C}_{X_k,X_k}^{(4)} &= \beta_{X_k\to\tilde{X}_k}^4 \mathbb{E}(n_k^4). \end{aligned} \tag{13}$$

One can see that $\mathcal{C}^{(4)}_{X_i,X_j}\mathcal{C}^{(4)}_{X_k,X_k} - \mathcal{C}^{(4)}_{X_i,X_k}\mathcal{C}^{(4)}_{X_j,X_k} = 0$, *which implies that the cumulant matrix* $\mathcal{C}^{(4)}_{\mathbf{Y};\mathbf{Z}}$ *is rank deficiency, i.e.,* $\mathrm{Rank}(\mathcal{C}^{(4)}_{\mathbf{Y};\mathbf{Z}}) = 1$. *By Theorem 1, we can infer that* $X_i \perp\!\!\!\perp X_j | X_k$.

Intuitively, the reason why CI relations can be tested by checking the rank constraints of the cumulant matrix is that the cumulant matrix excludes any information from Gaussian measurement error noise due to the property that higher-order cumulants of Gaussian distributions are zero, e.g., $\mathbb{E}(e_k^4) = 0$ in $\mathcal{C}^{(4)}_{X_k,X_k}$. This result can be directly applied to the linear non-Gaussian causal model, providing an alternative method for CI testing.

Corollary 1 (Rank constraints in linear non-Gaussian model)**.** *Let* X_i, X_j *be two random variables and* $\mathbf{X}_p$ *be a vector in the linear non-Gaussian model. Suppose any k-order cross-cumulant exists. Then,* $X_i \perp\!\!\!\perp X_j | \mathbf{X}_p$ *generic holds if and only if the rank of any k-order cumulant matrix between vectors* $\mathbf{Y} = (X_i, \mathbf{X}_p)$ *and vectors* $\mathbf{Z} = (X_j, \mathbf{X}_p)$ *is* $|\mathbf{X}_p|$, *i.e.,* $\mathrm{Rank}(\mathcal{C}^{k}_{\mathbf{Y};\mathbf{Z}}) = |\mathbf{X}_p|$, *where* $|\mathbf{X}_p|$ *are the dimension size of the vector.*

In other words, Corollary 1 shows that if there is no measurement error, Lemma 1 generally holds and can be viewed as a special case of the rank constraints of the cumulant matrix, so that $\mathrm{Rank}(\mathcal{C}^{2}_{\mathbf{Y},\mathbf{Z}}) = \mathrm{Rank}(\mathcal{C}^{k}_{\mathbf{Y},\mathbf{Z}})$ with $k \geq 2$. Practically, in the linear non-Gaussian model, one can enhance the statistical robustness of the CI test by summarizing the results across multiple k-order cumulants matrices.

Example 4. *Let* $k = \{4, 6, 8\}$, $\mathbf{Y} = \{\tilde{X}_i, \tilde{X}_k\}$ *and* $\mathbf{Z} = \{\tilde{X}_j, \tilde{X}_k\}$. *If we obtain results from observational data such that* $\mathrm{Rank}(\mathcal{C}^{(4)}_{\mathbf{Y},\mathbf{Z}}) = 1$, $\mathrm{Rank}(\mathcal{C}^{(6)}_{\mathbf{Y},\mathbf{Z}}) = 1$, *and* $\mathrm{Rank}(\mathcal{C}^{(8)}_{\mathbf{Y},\mathbf{Z}}) = 2$, *we can infer that* $X_i \perp\!\!\!\perp X_j | X_k$, *because most of the results tend to support the rank deficiency of the cumulant matrix.*

Moreover, in practice, an important problem is how to test the rank constraints of the cumulant matrix using empirical data. To address this, we propose the following hypothesis test.

Test the rank of Cumulant Matrix We test the rank of cumulant by leveraging CR statistic [Schkoda and Drton, 2023; Robin and Smith, 2000], which is formed to test the null hypothesis

$$\mathcal{H}_0 : \mathrm{Rank}(\mathcal{C}) = r \quad \text{vs.} \quad \mathcal{H}_1 : \mathrm{Rank}(\mathcal{C}) \neq r,$$

where $\hat{\mathcal{C}}$ be an asymptotically normal estimator of $\mathcal{C}$ and the null distribution of the CR statistic may be asymptotically approximated by a weighted sum of independent χ^2_1 random variables, the weights being determined by the asymptotic covariance matrix of $\hat{\mathcal{C}}$.

4 Application to Causal Structure Learning

Conditional independence (CI) tests are widely used in causal discovery, where that the true causal structure of n random variables can be represented by a directed acyclic graph (DAG) $\mathcal{G}$. In this section, we apply the proposed CI testing tool to the causal discovery task, presenting an algorithm that can infer causal structures from observational data in the presence of measurement error. To ensure the asymptotic correctness of the causal discovery methods, some necessary causal assumptions are required.

Assumption 1 (Causal Markov condition)**.** *The joint distribution* $\mathbb{P}$ *satisfies all CIs that are imposed by the true causal graph (generating process of the data).*

Assumption 2 (Rank Faithfulness [Spirtes, 2013])**.** *Let a distribution* P *be (linearly) rank-faithful to a directed acyclic graph* $\mathcal{G}$ *if every rank-constraint on a sub-cumulants matrix that holds in* P *is entailed by every free-parameter linear structural model with path diagram equal to* $\mathcal{G}$.

Note that the causal Markov condition and faithfulness assumption are commonly used in the constraint-based causal discovery algorithm, such as PC algorithm [Spirtes *et al.*, 2000]. Traditionally, the PC algorithm recovers the graph structure by leveraging the (conditional) independence relations identified in the data. However, it is well known that small mistakes early in the algorithm (e.g., failing to detect an independence relation) can propagate and lead to significant errors in the resulting DAG. Consequently, the performance of such methods heavily depends on the accuracy of the (conditional) independence testing procedures.

In the LiNGAMME framework, as shown in the illustrative example in Sec. 3.1, the conditional independence test becomes unreliable in the presence of measurement error, which can lead to incorrect results produced by the PC algorithm. To address this issue, by replacing the conditional independence test in the PC algorithm with our proposed tool, we present the PC algorithm with measurement error (abbreviated as **PC-ME**), as shown in Algorithm 1.

As shown in Algorithm 1, we adopt the standard framework of the PC algorithm to learn the causal structure from observed variables in the LiNGAMME framework. Specifically, we initialize the complete undirected graph $\mathcal{G}$ on the set of hidden variables $\mathbf{X}$. Then, we iteratively check whether each pair of hidden variables, X_i and X_j, are conditionally independent given a set of hidden variables $\mathbf{X}_p$. This is achieved by testing the rank constraint of the cumulant matrix among the observed variables. Formally, $X_i \perp\!\!\!\perp X_j \mid \mathbf{X}_p$ holds *if and only if* $\mathrm{Rank}(\mathcal{C}^{k}_{\mathbf{Y};\mathbf{Z}}) = |\mathbf{X}_p|$, where $\mathbf{Y} = \{\tilde{X}_i, \tilde{\mathbf{X}}_p\}$ and $\mathbf{Z} = \{\tilde{X}_j, \tilde{\mathbf{X}}_p\}$. If X_i and X_j are conditionally independent given a conditional set, the edge between X_i and X_j in $\mathcal{G}$ is removed. By repeating this procedure, we obtain the skeleton of the hidden variables. Simultaneously, we record the corresponding CI test results, including the conditional set of X_i and X_j in the d-separated set $\mathrm{Sepset}(X_i, X_j)$, which is useful for inferring causal directions and reducing redundant tests in subsequent steps. Next, we orient the causal directions by detecting V-structures, i.e., checking whether a local structure $X_i - X_k - X_j$ forms a V-structure. If it does, we orient it as $X_i \rightarrow X_k \leftarrow X_j$. Finally, we apply Meek's rules [Meek, 1995] to infer additional causal directions and output a partial DAG (PDAG) of the hidden variables with respect to the observational data.

Algorithm 1 PC algorithm with measurement error (PC-ME)

Require: Observational dataset $\tilde{\mathbf{X}} = \{\tilde{X_1}, \ldots, \tilde{X_m}\}$
Ensure: A partial DAG $\mathcal{G}$.

Let $\tilde{X}_i$ denote as its corresponding hidden variable X_i, $X_i \in \mathbf{X}$;
Initialize the complete undirected graph G over the hidden variable set $\mathbf{X}$;
for $\forall X_i, X_j \in \mathbf{X}$ and adjacent in G **do**
    if $\exists \mathbf{X}_p \subseteq \mathbf{X} \setminus \{X_i, X)j\}$ and $(|\mathbf{X}_p| < k)$ such that $\text{Rank}(\mathcal{C}^k_{\mathbf{Y};\mathbf{Z}}) = |\mathbf{X}_p|$ with $\mathbf{Y} = \{\tilde{X_i}, \tilde{\mathbf{X}_p}\}$, $\mathbf{Z} = \{\tilde{X_j}, \tilde{\mathbf{X}_p}\}$ **then**
        delete edge $X_i - X_j$ from $\mathcal{G}$;
        record $\mathbf{X}_p$ in Sepset(X_i, X_j);
        record $\mathbf{X}_p$ in Sepset(X_j, X_i);
    end if
end for
for $\forall X_i, X_j, X_k \in \mathbf{X}$ such that the pair X_i, X_k and the pair $X_j, X)k$ are adjacent in $\mathcal{G}$ but the pair X_i, X_j are not adjacent in $\mathcal{G}$ **do**
    if $X_k \notin$ Sepset$(X_i, X_j) \cup$ Sepset(X_j, X_i) **then**
        orient $X_i - X_k - X_j$ as $X_i \to X_k \leftarrow X_j$;
    end if
end for
while at least one edge can be oriented **do**
    if $\exists X_i \to X_j$, X_j and X_k are adjacent, X_i and X_k are not adjacent and there is no arrowhead at X_j **then**
        orient $X_j - X_k$ as $X_j \to X_k$
    end if
    if $\exists X_i$ and X_j, there is a directed path from X_i to X_j and an edge between X_i and X_j **then**
        orient $X_i - X_j$ as $X_i \to X_j$
    end if
end while
Return a partial DAG of hidden variables;

Complexity analysis. The complexity of the PC-ME algorithm for a hidden variable causal structure graph G is bounded by the largest degree in G. Let k be the maximal degree of any vertex and let n be the number of vertices. Then in the worst case, the number of conditional independence tests among hidden variables required by the algorithm is bounded by

$$\frac{n^2(n-1)^{k-1}}{(k-1)!} \tag{14}$$

Since the PC-ME algorithm relies only on the constraints of conditional independence (CI), it generally identifies the causal structure only up to Markov equivalence classes, which are sets of graphs that represent exactly the same independences and CIs. Thus, we next give the identifiability results of the PC-ME algorithm.

Theorem 2 (The correctness of algorithm)**.** *Suppose the data generation process follows the LiNGAMME, and the causal Markov condition, rank faithfulness assumption, and Condition 1 hold. Given a sufficiently large sample size, the PC-ME algorithm asymptotically outputs the Markov equivalence classes of the hidden variables.*

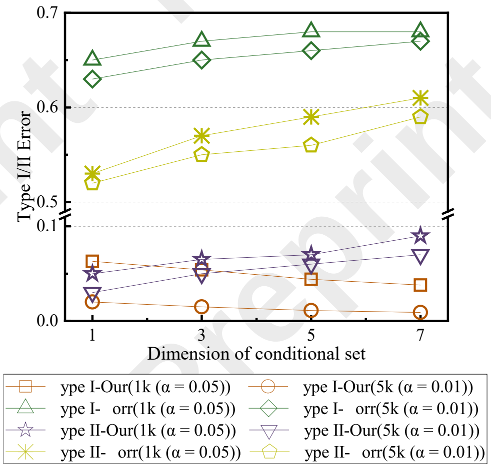

Figure 2: The average of Type I and Type II errors occurred in simulated case of $X_i \perp\!\!\!\perp X_j | \mathbf{X}_p$.

5 Simulation Experiments

In this section, we evaluate the practical performance of the proposed algorithm using synthetic data. First, we examine the performance in the CI test based on the rank constraints of cumulant matrices. Then, we verify the correctness of the PC-ME algorithm in the task of causal structure learning.

5.1 Performance of Conditional Independence Test

We conducted an experiment to verify the performance of the conditional independence test under the LiNGAMME. We consider the case whether X_i and X_j are conditional independence given by $\mathbf{X}_p$, i.e., $X_i \perp\!\!\!\perp X_j | \mathbf{X}_p$, where $\mathbf{X}_p = \{X_1, ..., X_m\}$. We adjust the dimension m of $\mathbf{X}_p$ to show that the proposed method is able to deal with the complexity case of conditional independent, ranging from 1 to 7. Specifically, we first generated the independent vector $\mathbf{X}_p$, then X_i and X_i were generated as $\beta^\intercal \mathbf{X}_p + \varepsilon$, where β is causal strength vector, which is generated uniformly in $[-2, -0.5] \cup [0.5, 2]$, ε is the noise term that follow the uniform distribution with $\mathcal{U}(-1, 1)$. Finally, we generate the observed variable according to the definition of LiNGAMME.

We compare our method with the partial correlation [Baba *et al.*, 2004]- the well-known CI test method in linear models. We use Type I and Type II errors as evaluation metrics (for more details, refer to the appendix) to investigate how the size of the conditional set impacts performance. To see the performance of Type II, not all elements in $\mathbf{X}_p$ are the conditional set of X_i and X_j. We randomly select the independent variable into $\mathbf{X}_p$ with ranging 1 to 3. In our simulations, the sample size is fixed at $[1000, 5000]$, and the significance level is set to $[0.05, 0.01]$ for all methods.

The results are shown in Fig. 2, one can see that the probabilities of Type I errors closely align with the significance levels for our method, and meanwhile, the Type I error is close to significance level. These results show that our CI test tool can tackle the CI test under the existence of measurement error. Although with the increase in the dimensionality of $\mathbf{X}_p$, the performance of our method noticeably decreases. This

		Precision ↑			**F1 ↑**		
Algorithm		Ours	PC	LPC	Ours	PC	LPC
D=5	3k	0.95	0.39	0.52	0.95	0.54	0.67
	5k	0.96	0.45	0.50	0.96	0.61	0.66
	10k	0.96	0.43	0.51	0.97	0.59	0.66
D=10	3k	0.92	0.42	0.51	0.93	0.58	0.66
	5k	0.95	0.43	0.46	0.95	0.59	0.62
	10k	0.95	0.45	0.52	0.96	0.61	0.67
D=15	3k	0.90	0.41	0.48	0.91	0.57	0.64
	5k	0.92	0.42	0.48	0.93	0.58	0.64
	10k	0.93	0.47	0.51	0.94	0.63	0.67

Table 1: Performance on structure learning (uniform distribution).

		Precision ↑			**F1 ↑**		
Algorithm		Ours	PC	LPC	Ours	PC	LPC
D=5	3k	0.95	0.48	0.49	0.95	0.65	0.65
	5k	0.96	0.59	0.52	0.96	0.65	0.67
	10k	0.97	0.55	0.54	0.97	0.70	0.68
D=10	3k	0.93	0.50	0.51	0.94	0.66	0.66
	5k	0.95	0.48	0.47	0.95	0.65	0.63
	10k	0.96	0.49	0.52	0.96	0.65	0.67
D=15	3k	0.92	0.47	0.51	0.92	0.63	0.66
	5k	0.94	0.49	0.45	0.94	0.65	0.61
	10k	0.95	0.52	0.53	0.95	0.67	0.68

Table 2: Performance on structure learning (standard exponential).

decrease is reasonable because the accuracy of the rank test decreases with the higher dimension matrix [Schkoda and Drton, 2023]. Furthermore, all of baseline methods exhibit poor performance in terms of Type I&II error, indicating the existing methods are not capable of handling the CI test in the presence of measurement error.

5.2 Performance of Causal Structure Learning

To evaluate the effectiveness of the proposed algorithm in causal structure learning, we employ the PC-ME algorithm on the random graph under the LiNGAMME model. We then compare its performance with that of the original PC algorithm, where CI relations are tested by partial correlation [Spirtes *et al.*, 2000], and latent PC algorithm (abbreviated as LPC) [Chen *et al.*, 2024], which allow to test the CI relations among hidden variables form measured variables.

Firstly, we construct the latent structure by randomly generating a Directed Acyclic Graph (DAG) [Kalisch and Bühlman, 2007], which can be represented as an $n \times n$ adjacency matrix, where n denotes the dimensionality of the latent variables. This matrix encodes the causal relationships among the hidden variables, with each element representing the causal strength between them. The hidden variable data is then generated from this DAG, with causal strength constrained to the interval $[-2, -0.5] \cup [0.5, 2]$, while the error term n_i of each hidden variable X_i is sampled from either (i) a uniform distribution $n_i \sim U(0.2, 1)$ (here four-order cumulant matrix is used in our method), or (ii) a standard exponential distribution. Subsequently, we generate observed variables corresponding to each hidden variable based on the definition of the LiNGAMEE. For instance, $\tilde{X}_i = \beta_{X_i \to \tilde{X}_i} X_i + e_i$, where $\tilde{X}_i$ is the observed variable of hidden variable X_i, $\beta_{X_i \to \tilde{X}_i}$ is disturbing strength, and e_i is the corresponding noise term that follows the Gaussian distribution.

We simulate observational data under different settings and apply our algorithm along with baseline methods to these datasets. This allows us to assess how the performance varies across different scenarios. We control the following parameters: (1) varying the number of latent variables $D = \{5, 10, 15\}$, (2) adjusting the sample sizes $\{3k, 5k, 10k\}$. Additionally, we set the sparseness parameter when generating the DAG to $s = 2/(k-1)$, where k represents the number of hidden variables, ensuring an average of two neighbors for each hidden variable [Cui *et al.*, 2018]. Moreover, since the output of the PC algorithm is a partial DAG, we utilize precision, recall, and F1 score as metrics. In our analysis, the significance level of all methods is set to $\alpha = 0.01$.

The experiment result of learning the causal skeleton is shown in Table 1 and Table 2. Our method outperforms all of the baseline methods, particularly in terms of precision and F1 score. This implies that the causal structure of the hidden variable estimated by our method is correct in most cases. Besides, one can see that all of the baseline methods exhibit poor F1 scores, indicating that existing methods are not able to handle scenarios involving measurement errors effectively. Due to space limitations, additional details on the experimental results are provided in the appendix.

6 Related Works

Numerous methods exist for conditional independence (CI) testing and causal discovery with latent variables. For CI testing, there are the residual-based methods [Zhang *et al.*, 2022; Zhang *et al.*, 2019; Chen *et al.*, 2024] and kernel-based approaches [Zhang *et al.*, 2011; Doran *et al.*, 2014]. For the task of causal discovery with latent variables, there are second-order-based ones [Silva *et al.*, 2006; Chen *et al.*, 2022; Huang *et al.*, 2022; Xie *et al.*, 2023], higher-order statistics-based approaches [Xie *et al.*, 2022; Xie *et al.*, 2024; Jin *et al.*, 2024; Chen *et al.*, 2023]. While effective for CI testing and latent causal discovery, these methods do not account for measurement error, leaving a key gap in existing approaches.

7 Conclusion

We study the problem of testing conditional independence among observed variables in the presence of measurement error. In the LiNGAMEE model, by incorporating the higher-order cumulant, we present the rank constraint of the cumulant matrix, a novel CI tool capable of accommodating measurement errors. We further apply the proposed tool to the task of causal structure learning and present a practical algorithm PC-ME. The theoretical tool elegantly extends the application scope of partial correlation to account for measurement errors, and the proposed algorithm offers a solution for investigating and addressing measurement errors in the analysis. How to relax the model assumptions, e.g., allowing non-linear relations, would be interesting future directions.

Acknowledgements

This research was supported in part by the National Natural Science Foundation of China under grants No. 62476163 and U24A20233, and the Guangdong Basic and Applied Basic Research Foundation under grant number 2023B1515120020. We appreciate the comments from anonymous reviewers, which greatly helped to improve the paper.

References

[Baba *et al.*, 2004] Kunihiro Baba, Ritei Shibata, and Masaaki Sibuya. Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics*, 46(4):657–664, 2004.

[Bouezmarni and Taamouti, 2014] Taoufik Bouezmarni and Abderrahim Taamouti. Nonparametric tests for conditional independence using conditional distributions. *Journal of Nonparametric Statistics*, 26(4):697–719, 2014.

[Chen *et al.*, 2022] Zhengming Chen, Feng Xie, Jie Qiao, Zhifeng Hao, Kun Zhang, and Ruichu Cai. Identification of linear latent variable model with arbitrary distribution. In *Proceedings 36th AAAI Conference on Artificial Intelligence (AAAI)*, 2022.

[Chen *et al.*, 2023] Zhengming Chen, Feng Xie, Jie Qiao, Zhifeng Hao, and Ruichu Cai. Some general identification results for linear latent hierarchical causal structure. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 3568–3576, 2023.

[Chen *et al.*, 2024] Zhengming Chen, Jie Qiao, Feng Xie, Ruichu Cai, Zhifeng Hao, and Keli Zhang. Testing conditional independence between latent variables by independence residuals. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[Cui *et al.*, 2018] Ruifei Cui, Perry Groot, Moritz Schauer, and Tom Heskes. Learning the causal structure of copula models with latent variables. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018*, pages 188–197. AUAI Press, 2018.

[Dai *et al.*, 2022] 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.

[Doran *et al.*, 2014] Gary Doran, Krikamol Muandet, Kun Zhang, and Bernhard Schölkopf. A permutation-based kernel conditional independence test. In *UAI*, pages 132–141, 2014.

[Fuller, 2009] Wayne A Fuller. *Measurement error models*. John Wiley & Sons, 2009.

[Huang *et al.*, 2022] Biwei Huang, Charles Jia Han Low, Feng Xie, Clark Glymour, and Kun Zhang. Latent hierarchical causal structure discovery with rank constraints. In *Advances in Neural Information Processing Systems*, 2022.

[Jin *et al.*, 2024] Songyao Jin, Feng Xie, Guangyi Chen, Biwei Huang, Zhengming Chen, Xinshuai Dong, and Kun Zhang. Structural estimation of partially observed linear non-gaussian acyclic model: A practical approach with identifiability. In *ICLR*, 2024.

[Kalisch and Bühlman, 2007] Markus Kalisch and Peter Bühlman. Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *Journal of Machine Learning Research*, 8(3), 2007.

[Kelly, 2007] Brandon C Kelly. Some aspects of measurement error in linear regression of astronomical data. *The Astrophysical Journal*, 665(2):1489, 2007.

[Lundborg *et al.*, 2022] Anton Rask Lundborg, Rajen D Shah, and Jonas Peters. Conditional independence testing in hilbert spaces with applications to functional data analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5):1821–1850, 2022.

[Meek, 1995] Christopher Meek. Causal inference and causal explanation with background knowledge. In *UAI*, pages 403–410, 1995.

[Robin and Smith, 2000] Jean-Marc Robin and Richard J Smith. Tests of rank. *Econometric Theory*, 16(2):151–175, 2000.

[Schkoda and Drton, 2023] Daniela Schkoda and Mathias Drton. Goodness-of-fit tests for linear non-gaussian structural equation models. *arXiv preprint arXiv:2311.04585*, 2023.

[Silva *et al.*, 2006] Ricardo Silva, Richard Scheine, Clark Glymour, and Peter Spirtes. Learning the structure of linear latent variable models. *JMLR*, 7(Feb):191–246, 2006.

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

[Spirtes *et al.*, 2000] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT press, 2000.

[Spirtes, 2013] Peter Spirtes. Calculation of entailed rank constraints in partially non-linear and cyclic models. In *UAI*, pages 606–615. AUAI Press, 2013.

[Sullivant *et al.*, 2010] Seth Sullivant, Kelli Talaska, Jan Draisma, et al. Trek separation for gaussian graphical models. *The Annals of Statistics*, 38(3):1665–1685, 2010.

[Totton and White, 2011] Nicola Totton and Paul White. The ubiquitous mythical normal distribution. *Research and Innovation*, 1:1–23, 2011.

[Watson and Wright, 2021] David S Watson and Marvin N Wright. Testing conditional independence in supervised learning algorithms. *Machine Learning*, 110(8):2107–2129, 2021.

[Xie *et al.*, 2022] 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. PMLR, 2022.

[Xie *et al.*, 2023] Feng Xie, Yan Zeng, Zhengming Chen, Yangbo He, Zhi Geng, and Kun Zhang. Causal discovery of 1-factor measurement models in linear latent variable models with arbitrary noise distributions. *Neurocomputing*, 2023.

[Xie *et al.*, 2024] 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.

[Yang *et al.*, 2022] Yuqin Yang, AmirEmad Ghassami, Mohamed Nafea, Negar Kiyavash, Kun Zhang, and Ilya Shpitser. Causal discovery in linear latent variable models subject to measurement error. *Advances in Neural Information Processing Systems*, 35:874–886, 2022.

[Zhang *et al.*, 2011] Kun Zhang, Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 804–813, 2011.

[Zhang *et al.*, 2018] Kun Zhang, Mingming Gong, Joseph D Ramsey, Kayhan Batmanghelich, Peter Spirtes, and Clark Glymour. Causal discovery with linear non-gaussian models under measurement error: Structural identifiability results. In *UAI*, pages 1063–1072, 2018.

[Zhang *et al.*, 2019] Hao Zhang, Shuigeng Zhou, Jihong Guan, and Jun Huan. Measuring conditional independence by independent residuals for causal discovery. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(5):1–19, 2019.

[Zhang *et al.*, 2022] Hao Zhang, Shuigeng Zhou, Kun Zhang, and Jihong Guan. Residual similarity based conditional independence test and its application in causal discovery. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 5942–5949, 2022.

[Zhou *et al.*, 2020] Yeqing Zhou, Jingyuan Liu, and Liping Zhu. Test for conditional independence with application to conditional screening. *Journal of Multivariate Analysis*, 175:104557, 2020.