Seminar:
Learning distribution-free anchored linear structural equation models in the presence of measurement error.

Junhyoung Chung

July 10, 2024

Seoul National University
Department of Statistics

· Chung, J., Ahn, Y., Shin, D., & Park, G. (2024)
Learning distribution-free anchored linear structural equation models in the presence of measurement error.

Main contributions and outline

Main contributions

- Establish an identifiability of distribution-free anchored linear SEMs based on the geometry-faithfulness assumption.
- Propose a consistent algorithm to discover a latent structure in the presence of measurement error.
- Provide various numerical experiments and analysis of real galaxy data.

Introduction

Introduction

- Identifiability of directed acyclic graphical models (DAG) is usually achieved by posing additional assumptions. For example,
  - **Causal minimality**: True graph is a minimal structure that is Markov to its distribution.
  - **Faithfulness**: Conditional independence implies d-separation.
  - **Distributional constraints**: Gaussian errors with equal variance (Peters and Bühlmann, 2014), non-Gaussian errors (Shimizu et al., 2006), etc.
- The aforementioned identifiability results work under **causal sufficiency** regime, excluding the presence of latent variables.
- However, in many real-world setting, observed variables are **imperfect measures** of corresponding true variables.

Motivating example

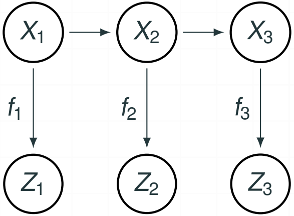


Figure 1: Anchored DAG between the latent variables X and the observed variables Z.

- Figure 1 visualizes the relationship between the latent variables X and the observed variables Z.
- One can observe that X_1 and X_3 are d-separated(blocked) by X_2, while the statement becomes false if we replace X to Z.

Previous research

	Latent structure	Contamination	Distributional constraints
Halpern et al., 2015	$P(X) = \prod_j P(X_j \mid X_{\mathrm{Pa}(j)})$	$Z_j = f_j(X_j)$	X,Z are binary
Zhang et al., 2017	$X = BX + \epsilon$	$Z = X + \Psi$	ϵ, Ψ are Gaussian
Zhang et al., 2018	$X = BX + \epsilon$	$Z = X + \Psi$	ϵ is non-Gaussian
Saeed et al., 2020	$X = BX + \epsilon$	$Z_j = f_j(X_j)$	ϵ is Gaussian Known moment relationship
Liu et al., 2022	$P(X) = \prod_j P(X_j \mid X_{\mathrm{Pa}(j)})$	$Z_j = f_j(X_j)$	Likelihood is given
Ours	$X = BX + \epsilon$	$Z_j = f_j(X_j)$	**Known moment relationship**

Preliminaries

Directed acyclic graph

Figure 2: 5-node DAG example.

- A DAG $\mathcal{G} = (V, E)$ consists of a set of nodes $V = \{1, ..., p\}$ and a set of directed edges $E \subset V \times V$ with no directed cycles. Its **skeleton** is an undirected graph obtained by removing directions in the edges.
- A set of **parents** of node k, denoted by $\mathrm{Pa}(k)$, consists of all nodes j such that $(j, k) \in E$.
- If there is a directed path $j \to \cdots \to k$, then k is a **descendant** of j, and j is called an **ancestor** of k.
- A node k is a **collider** if there exists a triple (j, k, ℓ) such that $j \to k \leftarrow \ell$, and we say such triple generates a **v-structure**.

D-separation and d-connection

Figure 3: 5-node DAG example.

- Two nodes j and k in DAG $\mathcal{G}$ are **d-connected by a node set** $S \subset V$ if there exists a path $\mathcal{P}$ between j and k such that for every node ℓ on the path $\mathcal{P}$
  1. if ℓ is a collider, either ℓ or its descendant is in S,
  2. otherwise ℓ is not in S.
- If j and k are not d-connected by S, we say j and k are **d-separated by** S.
  - 1 and 5 are d-separated by $\emptyset, \{2\}, \{4\}, \{2, 4\}$.
  - 1 and 5 are d-connected by $\{3\}$.

Markov equivalence class

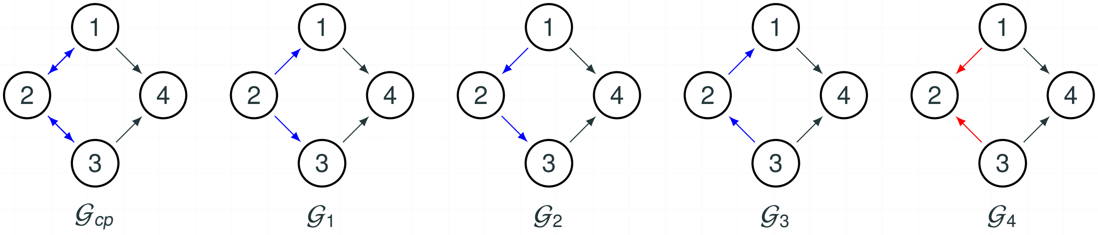


Figure 4: Markov equivalence class of $\mathcal{G}_1$ ($\mathcal{G}_2$ and $\mathcal{G}_3$) and its CPDAG.

- A **Markov equivalence class** (MEC) is a set of DAGs that encode the same set of d-separations.
- It is known that all DAGs in the MEC **have the same skeleton and the same v-structures**.
- A **complete partially directed acyclic graph** (CPDAG) is a unique representation of MEC.

Linear SEM

- The joint distribution generated by a **DAG model** $(\mathcal{G}, P)$ can be factorized as follows:

$$P(X) = P(X_1, ..., X_p) = \prod_{j=1}^{p} P(X_j \mid X_{\text{Pa}(j)}). \tag{1}$$

- A **linear SEM** is a special DAG model of (1) where the joint distribution of a linear SEM is defined by the following linear equations: For all $j \in V$,

$$X_j = \sum_{k \in \text{Pa}(j)} \beta_{kj} X_k + \epsilon_j, \tag{2}$$

where $(\epsilon_j)_{j \in V}$ are independent, but possibly not identical errors with mean 0 and variance $(\sigma_j^2)_{j \in V}$.

- The linear SEM in (2) can be restated as a matrix form:

$$X = BX + \epsilon. \tag{3}$$

- We denote $\mathcal{L}(\mathcal{G}, B, F)$ as the linear SEM in (3) where B is the edge weight matrix, $\mathcal{G}$ is the underlying true DAG, and $\epsilon \sim F$.

Anchored linear SEM

- An **anchored DAG model** considers a DAG model with latent variables.
- In our framework, we consider an **anchored linear SEM**, special case of an anchored DAG model, as follows: For all $j \in V$,

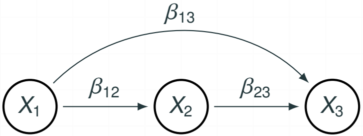
$$Z_j = f_j(X_j) \quad \text{and} \quad X \sim \mathcal{L}\left(\mathcal{G}, B, (0, \Sigma_\epsilon)\right), \tag{4}$$

where $\Sigma_\epsilon = \text{diag}(\sigma_1^2, ..., \sigma_p^2)$, and $f_j : \mathbb{R} \to \mathbb{R}$ can be linear, non-linear, or even non-deterministic function.

  - (Additive measurement error model) $f_j(X_j) = X_j + \psi_j$, where $\psi_j \sim (0, \eta_j^2)$.
  - (Dropout model) $f_j(X_j) = X_j \psi_j$, where $\psi_j \sim \text{Bernoulli}(p)$.
  - (Poisson transformation) $f_j(X_j) = \text{Poisson}(|X_j|)$.

Identifiability for distribution-free anchored linear SEMs

Geometry-faithfulness



- To provide an intuition, consider a 3-node fully connected linear SEM:

$$X_1 = \epsilon_1, \quad X_2 = \beta_{12} X_1 + \epsilon_2, \quad X_3 = \beta_{13} X_1 + \beta_{23} X_2 + \epsilon_3,$$

  where $\epsilon_j \sim (0, 1)$ for all $j \in \{1, 2, 3\}$.
- Then, the inverse covariance matrix is

$$\Sigma^{-1} = \begin{pmatrix} 1 + \beta_{12}^2 + \beta_{13}^2 & -\beta_{12} + \beta_{13}\beta_{23} & -\beta_{13} \\ - & 1 + \beta_{23}^2 & -\beta_{23} \\ - & - & 1 \end{pmatrix}.$$

- Observe that X_1 and X_3 are d-separated by X_2 iff $[\Sigma^{-1}]_{13} = 0$. In addition, X_2 and X_3 are d-separated by X_1 iff $[\Sigma^{-1}]_{23} = 0$.

Geometry-faithfulness

Assumption 1. Geometry-faithfulness

Consider a linear SEM $\mathcal{L}(\mathcal{G}, B, (0, \Sigma_\epsilon))$ that generates $P(X)$, i.e., $X \sim \mathcal{L}(\mathcal{G}, B, (0, \Sigma_\epsilon))$. Then, for any pair of nodes $j, k \in V$, and for any subset $S \subset V \setminus \{j, k\}$,

$$j \text{ and } k \text{ are d-separated by } S \text{ in } \mathcal{G} \iff \rho_{j,k,S} \propto [(\Sigma_{L,L})^{-1}]_{j,k} = 0,$$

where $\Sigma = (I_p - B)^{-1}\Sigma_\epsilon(I_p - B)^{-\top}$, $L = S \cup \{j, k\}$, and $\rho_{j,k,S}$ is the partial correlation coefficient of X_j and X_k given X_S.

- Geometry-faithfulness ensures that partial correlations directly reflect d-separations and connections within the graph.
- Under the geometry-faithfulness assumption, j and k are d-separated by S if and only if the residuals obtained by projecting X_j and X_k onto X_S are orthogonal.

Identifiability for distribution-free anchored linear SEMs

Theorem 1. Identifiability for distribution-free anchored linear SEMs

Consider a distribution-free anchored linear SEM with $\mathcal{L}(G, B, (0, \Sigma_\epsilon))$. Then, model is identifiable up to the MEC if the followings are satisfied.

(A1). The latent distribution $P(X)$ is geometry-faithful to $\mathcal{G}$.

(A2). The observed random variables satisfy the following condition: For all $j \in V$,

$$Z_j \perp\!\!\!\perp \{Z_1, ..., Z_p, X_1, ..., X_p\} \setminus \{Z_j, X_j\} \mid X_j.$$

(A3). For all $j, k \in V$, there exists a finite-dimensional vector δ_j of monomials in Z_j and a finite-dimensional vector δ_{jk} of monomials in Z_j and Z_k, such that their means can be mapped to the moments of the latent variables by continuously differentiable functions g_{jj} and g_{jk}, such that $\mathbb{E}[X_j] = g_j(\mathbb{E}[\delta_j])$, $\mathbb{E}[X_j^2] = g_{jj}(\mathbb{E}[\delta_{jj}])$, and $\mathbb{E}[X_j X_k] = g_{jk}(\mathbb{E}[\delta_{jk}])$, and their covariance satisfies $\mathrm{Cov}(\delta_j, \delta_{jk}) < \infty$.

Algorithm

Algorithm

Algorithm 1. CPDAG learning algorithm for distribution-free anchored linear SEMs

Input: n i.i.d. observations from an anchored linear SEM, $Z^{1:n}$, transformation $\mathcal{T}$ and g such that $\mathrm{Cov}(X) = \mathcal{T}(\mathbb{E}[g(Z)])$, and significance level α.

Output: Complete Partial DAG (CPDAG), $\hat{\mathcal{G}}_{cp}$.

1. Estimate the mean of $g(Z)$ from $Z^{1:n}$.
2. Estimate the covariance matrix $\hat{\Sigma}$ for latent variables using $\mathcal{T}$ and g.
3. Estimate the partial correlations of X using $\hat{\Sigma}$.
4. Estimate a CPDAG using a constraint-based algorithm (e.g., the PC algorithm) by conducting a consistent partial correlation test with α.

Return: Estimated CPDAG, $\hat{\mathcal{G}}_{cp}$.

- Under appropriate conditions, **Algorithm 1 is guaranteed to be consistent**.

Consistency

Assumption 2. Strong geometry-faithfulness

Consider an anchored linear SEM (4). For any $j, k \in V$ and $S \subset V \setminus \{j, k\}$, if j and k are d-separated by S, then the corresponding partial correlation is zero; otherwise, it is bounded below by a constant. Precisely, there exists a constant $\rho_{\min} > 0$ such that

$$0 < \rho_{\min} \leq \inf \left\{ |\rho_{j,k,S}| : j \text{ and } k \text{ are d-connected by } S \right\}.$$

Theorem 2. Consistency

Consider an anchored linear SEM (4) with the true CPDAG $\mathcal{G}_{cp}$. Suppose that the strong geometry-faithfulness assumption specified in Assumption 2 and Assumptions (A2)-(A3) in Theorem 1 are satisfied. Additionally, suppose that the sequence of significance level $\{\alpha_n : n \in \mathbb{N}\}$ satisfies $2\left(1 - \Phi\left(0.5\rho_{\min}\sqrt{n-p-1}\right)\right) < \alpha_n < 1$ for all $n \in \mathbb{N}$ where $\Phi(\cdot)$ denotes the CDF of $N(0, 1)$. Then, Algorithm 1 is consistent, i.e.,

$$\hat{\mathcal{G}}_{cp} \to G_{cp} \quad \text{as} \quad n \to \infty,$$

where $\hat{\mathcal{G}}_{cp}$ is the estimated CPDAG by Algorithm 1.

Numerical experiments

Experiment setting

- The number of its parents was uniformly selected from the set $\{1, ..., d_{in}\}$.

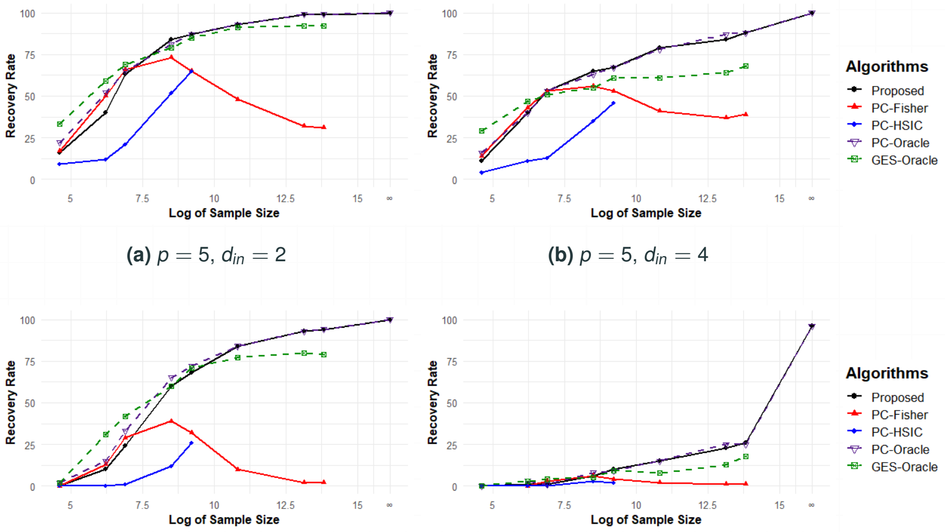
- β_{jk} in Equation (2) were randomly chosen to fall within the intervals $(-0.75, -0.25) \cup (0.25, 0.75)$.

- Models were constructed with various noise distributions: (i) Gaussian, (ii) Uniform, (iii) Student's t, and (iv) Discrete uniform.

- For the contamination process, (i) dropout models, (ii) additive measurement error models were considered.
  - ▹ In dropout models, the dropout probability was set to $\gamma = 0.1$.
  - ▹ In additive measurement error models, the variance of the measurement error was set at $\eta^2 = 0.25$.

- Finally, the significance level was set at $\alpha_n = 1 - \Phi(n^{1/5}/2)$, respecting the theoretical result from Theorem 2.

Dropout models



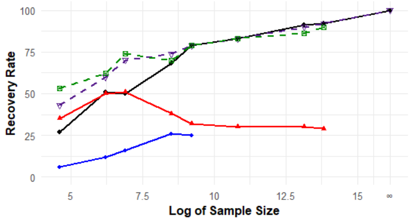
(a) $p = 5$, $d_{in} = 2$

(b) $p = 5$, $d_{in} = 4$

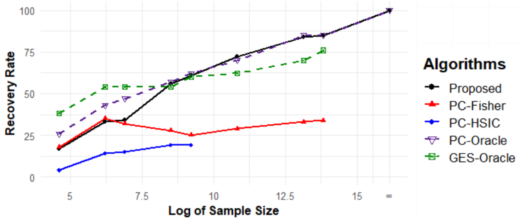
(c) $p = 10$, $d_{in} = 2$

(d) $p = 10$, $d_{in} = 4$

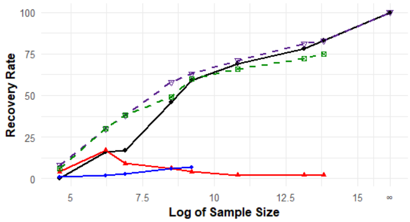
Additive measurement error models



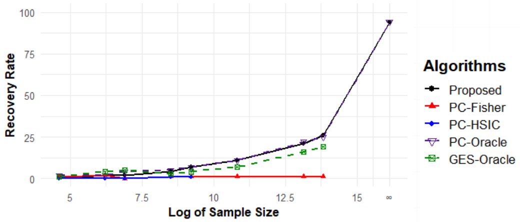
(a) $p = 5$, $d_{in} = 2$



(b) $p = 5$, $d_{in} = 4$



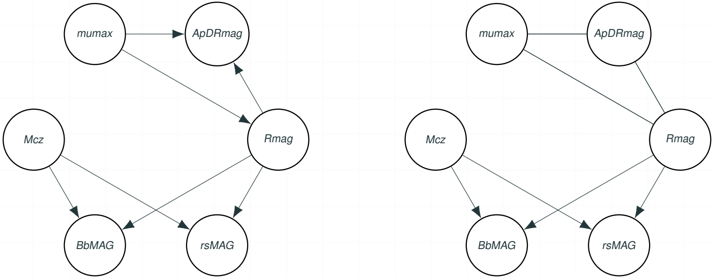
(c) $p = 10$, $d_{in} = 2$



(d) $p = 10$, $d_{in} = 4$

Real data analysis

Data description



(a) True DAG **(b)** True CPDAG

Figure 7: The true DAG and CPDAG of variables within galaxy data.

- The data ($p = 6, n = 3462$) comprises galaxy brightness measurements, which also includes measurement errors for each variable.
  - Rmag (0.0069), mumax (0), ApDRmag (0), Mcz (0.0038), BbMAG (1.5233) and rsMAG (1.6508).

Results

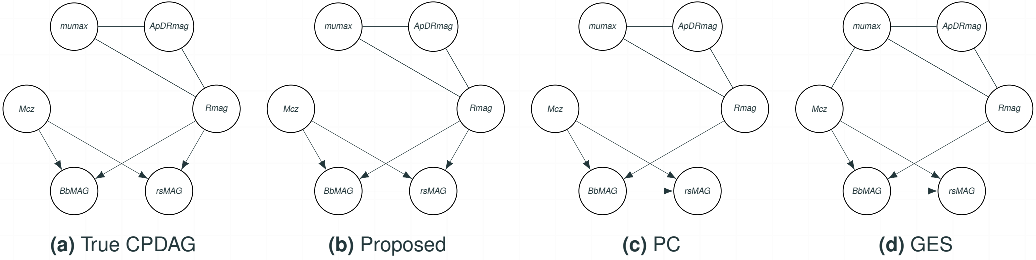


Figure 8: CPDAGs estimated by the proposed, PC, and GES algorithms.

- The proposed algorithm successfully detects all true edges, while falsely detects an undirected edge (BbMAG, rsMAG).
- Contrary to the proposed method, the PC and GES algorithms fail to recover the true edges.

Discussion

Discussion

- Introduce the **geometry-faithfulness assumption**.
- Present a **consistent learning algorithm based on the PC-algorithm**.
- **Known moment relationships** between observed and latent variables is required.
- Develop **high-dimensional models** and **DAG recovery algorithms**.

Reference

- Halpern, Y., Horng, S., & Sontag, D. (2015). Anchored discrete factor analysis. *arXiv preprint arXiv:1511.03299*.
- Liu, Y., Constantinou, A. C., & Guo, Z. (2022). Improving Bayesian network structure learning in the presence of measurement error. *Journal of Machine Learning Research, 23*(324), 1-28.
- Peters, J., & Bühlmann, P. (2014). Identifiability of Gaussian structural equation models with equal error variances. *Biometrika, 101*(1), 219-228.
- Saeed, B., Belyaeva, A., Wang, Y., & Uhler, C. (2020). Anchored causal inference in the presence of measurement error. In *Conference on uncertainty in artificial intelligence* (pp. 619-628). PMLR.
- Shimizu, S., Hoyer, P. O., Hyvärinen, A., Kerminen, A., & Jordan, M. (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research, 7*(10).
- Zhang, K., Gong, M., Ramsey, J., Batmanghelich, K., Spirtes, P., & Glymour, C. (2017). Causal discovery in the presence of measurement error: Identifiability conditions. *arXiv preprint arXiv:1706.03768.*
- Zhang, K., Gong, M., Ramsey, J. D., Batmanghelich, K., Spirtes, P., & Glymour, C. (2018). Causal Discovery with Linear Non-Gaussian Models under Measurement Error: Structural Identifiability Results. In *UAI* (pp. 1063-1072).