Causal Discovery in Linear Models with Unobserved Variables and Measurement Error

arXiv

Prev. work

Yuqin Yang[1] Mohamed Nafea[2] Negar Kiyavash[3] Kun Zhang[4] AmirEmad Ghassami[5]

[1]Georgia Tech [2]Missouri S&T [3]EPFL [4]CMU & MBZUAI [5]Boston University

Motivation

The presence of **unobserved common causes** and **measurement error** are two of the most limiting challenges in causal discovery.

Existing work:

- Latent variable (LV): Unable to orient most edges, or provide non-trivial conditions for uniq. identifiability
- Measurement error (ME): Assume that each latent variable has at least two measurements
- Ours – NeurIPS'22: Considers LV or ME separately

This work studies the extent of identifiability from observational data when both challenges co-exist by leveraging special properties of measurement variables.

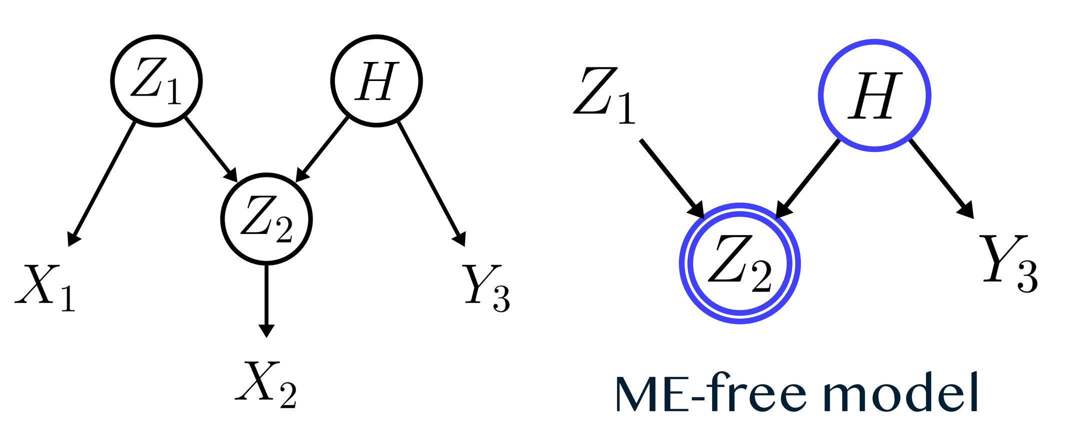
Linear LV-SEM-ME

Two sets of variables $V = [H; Z; Y]$ (underlying), X

- V follow a linear SCM: $V = \mathbf{A}V + N_V$
- Y : Observed (without error)
- Z : Measured variables, with measurements X
  - $X_i = Z_i + N_{X_i}$
- H : Unobserved (neither observed nor measured)

Canonical form: Measured leaf (mleaf) variables do not have exogenous noises; unobserved variables are roots.

- Define the set of cogent variables V^C as variables in V that are neither H nor mleaf.



We know whether each variable is observed ("Y") or measured ("X").

Identification Assumptions

Separability: Mixing matrix $\mathbf{W}$ transforming exog. noises to observed variables ($[X, Y]$) can be recovered from observational distribution.

- $[X; Y] = \mathbf{W} \cdot N$; $\mathbf{W}$ can be derived from $\mathbf{A}$
- Satisfied when all noises are non-Gaussian
- $\mathbf{W}^*$ corresp. to ME-free model can be deduced

Two-fold faithfulness assumption:

(a) Conventional (total causal effect not zero);

(b) Prevents measure-zero parameter cancellation or proportionality among specific edges

Minimality: Of the number of H with the same $\mathbf{W}$

- Provide equivalent graphical condition

Identification results & Algorithm

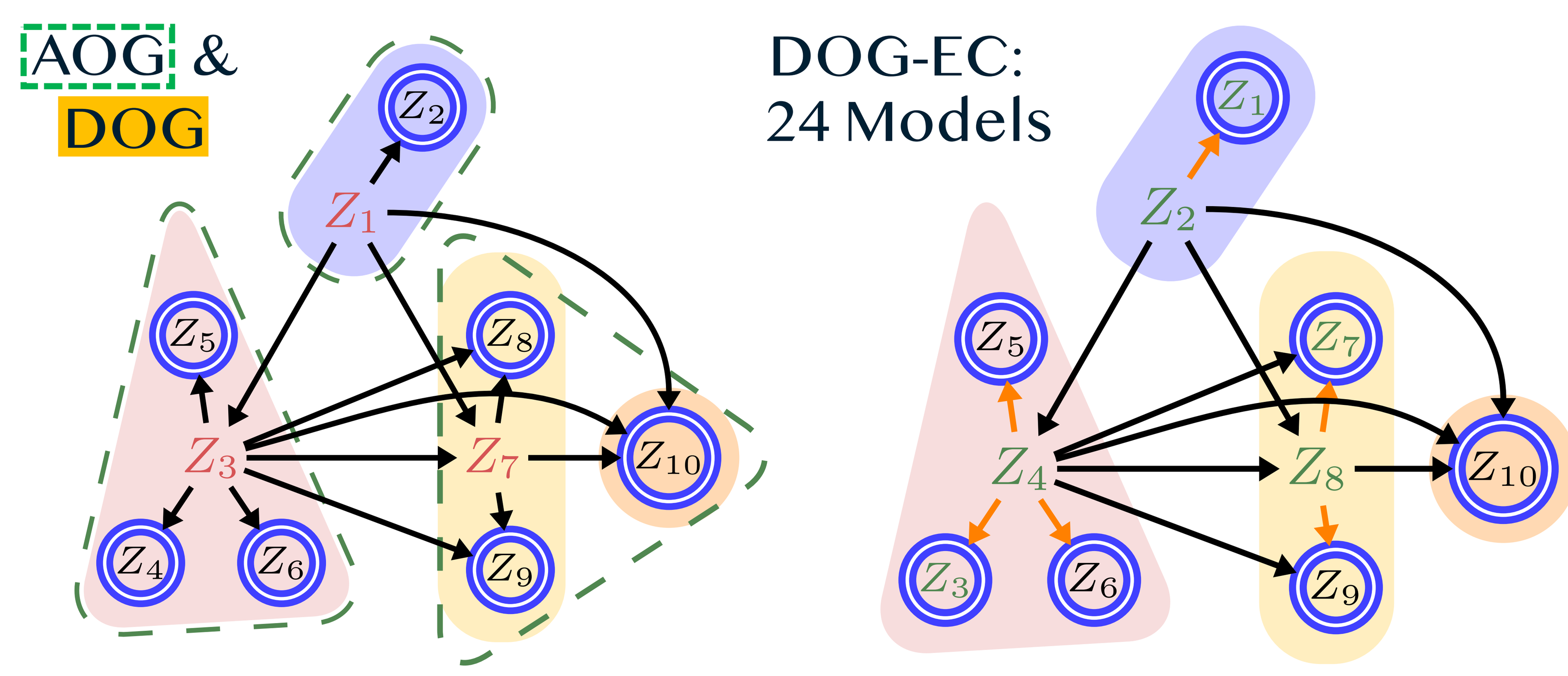
Theorem: Under conventional / LV-SEM-ME faithfulness assumption, an LV-SEM-ME can be recovered up to its **AOG / DOG equiv. class** (EC), where models in the same group have the same $\mathbf{W}$ and AOG / DOG.

Ancestral / Direct Ordered Grouping (AOG / DOG)

Partition variables in V into distinct groups:

1. Assign each cogent variable to a separate group.
2. Assign each unobserved / mleaf variable either to one of its children's / measured parents' group, or a separate group based on different graphical conditions for AOG / DOG (see the paper).

- DOG is a finer partition than AOG, hence DOG-EC is a subset of AOG-EC.

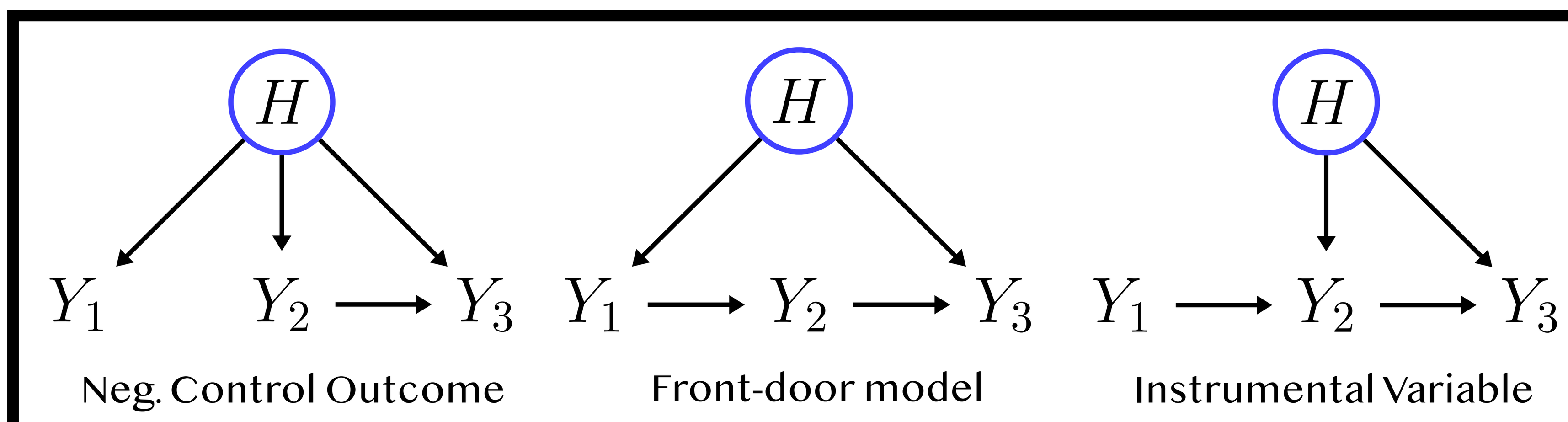


- The induced structure on each group is a star graph
  - Define the center of the "star" as the cogent var.
- Each model corresponds to a distinct choice of the centers of the stars (or their corresp. exogenous noises).
- Models in the same DOG-EC of an LV-SEM-ME have the same unlabeled graph structure.

Algorithm 1: Recover all models in the AOG / DOG Equivalence Class

1. Recover the AOG of the true model (see the paper). Initialize $\mathcal{M}_{AOG} = \emptyset$.
2. **for** *all possible selections* `row`, `col` *of the centers and the corresponding exog. noises in the groups* **do**
3. Recover $\mathbf{A}$ using sub-matrices of $\mathbf{W}^*$ partitioned by `row` and `col`.
4. Add $\mathbf{A}$ to $\mathcal{M}_{AOG}$.
5. Select $\mathcal{M}_{DOG}$ as the set of models in $\mathcal{M}_{AOG}$ that have the fewest total number of edges (non-zero entries) in $\mathbf{A}$.

Simulations



Relative Error	DOGEC (ours)			GRICA			lvLiNGAM			Cross-Moment		
	Mean	20%	80%	Mean	20%	80%	Mean	20%	80%	Mean	20%	80%
NCO	**0.08**	**0.06**	**0.11**	0.38	0.30	0.43	1.15	0.88	1.49	**0.08**	**0.04**	**0.10**
Front	**0.02**	**0.01**	**0.02**	0.46	0.35	0.57	0.84	0.60	0.93	0.23	0.03	0.23
IV	**0.19**	**0.02**	**0.36**	0.50	0.37	0.59	0.84	0.69	0.99	1.33	0.49	1.21
Union	**0.11**	**0.08**	**0.14**	0.39	0.30	0.42	0.73	0.28	0.89	0.30	0.23	0.38

- Estimate the edge weight (direct effect) of $Y_2 \rightarrow Y_3$
- Baselines: Require non-Gaussianity and known graph