Trek-Based Parameter Identification for Linear Causal Models With Arbitrarily Structured Latent Variables

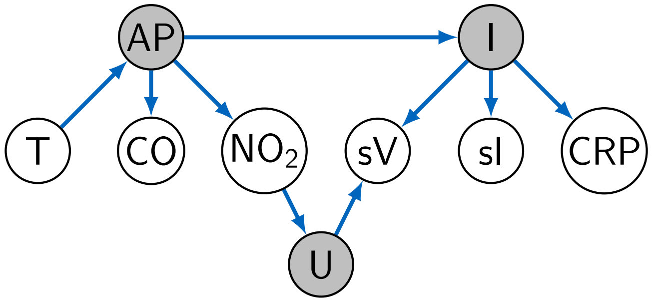
Nils Sturma

Insitute of Mathematics

École Polytechnique Fédérale de Lausanne (EPFL)

(joint work with Mathias Drton)

Example: Effect of Air Pollution on Inflammation



Goal:

Can we recover the causal effect of Air Pollution on Inflammation from the observable covariance matrix?

Topic of this talk: Are the semi-direct effects between the observed variables identifiable?

From the observable population covariance matrix?

Setup

Graph:

Directed graph $G = (V, D)$ with observed nodes $\mathcal{O}$ and latent nodes $\mathcal{L}$, and $V := \mathcal{O} \sqcup \mathcal{L}$.

Linear Causal Model:

Random vector $X = (X_v)_{v \in V}$ satisfies

$$X_v = \sum_{w \in \mathrm{pa}(v)} \lambda_{wv} X_w + \varepsilon_v,$$

where $\varepsilon = (\varepsilon_v)_{v \in V}$ are jointly independent with $\mathrm{Var}[\varepsilon_v] = \phi_v$.

Matrix Form:

$$X = \Lambda^\top X + \varepsilon,$$

where parameter matrix Λ is sparse and supported over edge set D, and $\Phi = \mathrm{Var}(\varepsilon) = \mathrm{diag}(\phi_v : v \in V)$.

Observable Population Covariance Matrix:

$$\Sigma := \mathrm{Cov}[X_{\mathcal{O}}] = [(I - \Lambda)^{-\top} \Phi (I - \Lambda)^{-1}]_{\mathcal{O},\mathcal{O}}.$$

Semi-direct Effects

Consider $v, w \in \mathcal{O}$. We say that there is a semi-direct effect $v \rightsquigarrow w$ in G if

$$v \to w \in G \qquad \text{or} \qquad v \to h_1 \to \cdots \to h_k \to w \in G$$

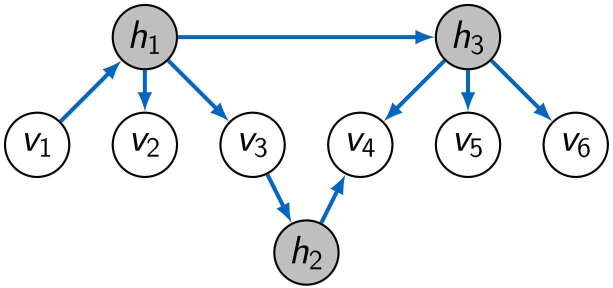
with $h_1, \ldots, h_k \in \mathcal{L}$.

Matrix of Interest:

$$\overline{\Lambda} := \Lambda_{\mathcal{O},\mathcal{O}} + \Lambda_{\mathcal{O},\mathcal{L}}(I - \Lambda_{\mathcal{L},\mathcal{L}})^{-1}\Lambda_{\mathcal{L},\mathcal{O}}.$$

Entries are sums over all semi-direct effects.

Example:



$v_1 \rightsquigarrow v_4 \in G$

$v_3 \rightsquigarrow v_4 \in G$

$\ldots$

Question: For which graphs can we recover $\overline{\Lambda}$ from $\Sigma = \mathrm{Cov}[X_{\mathcal{O}}]$?

Factorization of the Covariance Matrix

It holds that

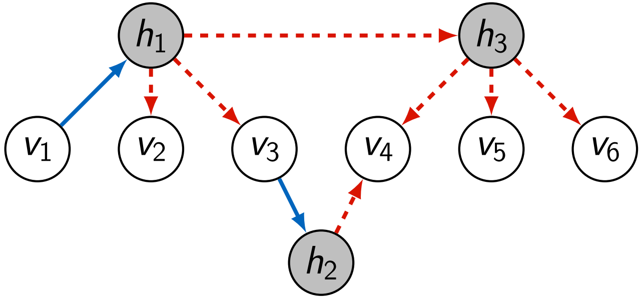
$$\Sigma = \mathrm{Cov}[X_{\mathcal{O}}] = (I - \overline{\Lambda})^{-\top}\Omega(I - \overline{\Lambda})^{-1}.$$

Observation:

Ω is cov. matrix in the model given by subgraph G_{lat} obtained from deleting all edges $v \to w$ with $v \in \mathcal{O}$.

Note: $\Omega = (I - \overline{\Lambda})^{\top}\Sigma(I - \overline{\Lambda})$.

Example:



Identifying $h_1 \to h_3$ in G_{lat} is well studied:

$$|\lambda_{h_1 h_3}| = \sqrt{\frac{\omega_{v_2 v_6}\omega_{v_3 v_5}}{\omega_{v_2 v_3}\omega_{v_5 v_6} - \omega_{v_2 v_6}\omega_{v_3 v_5}}}.$$

Identifying $\overline{\Lambda}$ allows to transform the given model into a simpler measurement model for which existing tools become applicable.

Key Idea

Exploit low rank in Ω:

Fix $v \in \mathcal{O}$ and consider $\overline{\Lambda}_{\overline{\text{pa}}(v),v}$.

- Find sets $Y, Z \subseteq \mathcal{O}$ such that $\Omega_{Y,Z\cup\{v\}}$ does not have full column rank, but $\text{rank}(\Omega_{Y,Z}) = |Z|$.
- Then, there is $\xi \in \mathbb{R}^{|Z|}$ s.t. $\Omega_{Y,Z} \cdot \xi = \Omega_{Y,v}$

$$\begin{aligned}
&\iff \Omega_{Y,Z} \cdot \xi = [(I-\overline{\Lambda})^\top \Sigma (I-\overline{\Lambda})]_{Y,v} \\
&\iff \Omega_{Y,Z} \cdot \xi = [(I-\overline{\Lambda})^\top \Sigma]_{Y,v} - [(I-\overline{\Lambda})^\top \Sigma \overline{\Lambda}]_{Y,v} \\
&\iff \left([(I-\overline{\Lambda})^\top \Sigma]_{Y,\overline{\text{pa}}(v)} \quad \Omega_{Y,Z} \right) \cdot \begin{pmatrix} \overline{\Lambda}_{\overline{\text{pa}}(v),v} \\ \xi \end{pmatrix} = [(I-\overline{\Lambda})^\top \Sigma]_{Y,v}.
\end{aligned}$$

- Ensure that the matrix on the left-hand side is invertible and that $\Omega_{Y,Z}$ and suitable entries in $\overline{\Lambda}$ are already certified to be identifiable from earlier calculations.

Trek Rule

$\Sigma_{vw} = [(I-\Lambda)^{-\top}\Phi(I-\Lambda)^{-1}]_{vw}$ is sum over 'treks' $v \leftarrow \cdots \leftarrow k \rightarrow \cdots \rightarrow w$, i.e,

$$\Sigma_{vw} = \sum_{\pi \in \mathcal{T}(v,w)} \phi_k \prod_{x \to y \in \pi} \lambda_{xy}.$$

$\implies$ Ω_{vw} is sum over treks in G_{lat}.

$\implies$ $[(I-\overline{\Lambda})^{\top}\Sigma]_{vw} = [\Omega(I-\overline{\Lambda})^{-1}]_{vw}$ is sum over treks with left part in G_{lat}.

Trek-Separation in Subgraphs:

If there exists a system of treks with no sided intersection from Y to $\overline{\mathrm{pa}}(v) \cup Z$ in G such that the left part of every trek only takes edges in G_{lat}, and the right part of every trek ending in Z only takes edges in G_{lat}, then we have generically

$$\det\left([(I-\overline{\Lambda})^{\top}\Sigma]_{Y,\overline{\mathrm{pa}}(v)} \quad \Omega_{Y,Z}\right) \neq 0.$$

Latent Subgraph Criterion

Definition

Given a node $v \in \mathcal{O}$, the 4-tuple $(Y, Z, H_1, H_2) \in 2^{\mathcal{O}\setminus\{v\}} \times 2^{\mathcal{O}\setminus\{v\}} \times 2^{\mathcal{L}} \times 2^{\mathcal{L}}$ satisfies the *latent-subgraph criterion* (LSC) for v if

(i) $|Y| = |\overline{\mathrm{pa}}(v)| + |Z|$ and $|Z| = |H_1| + |H_2|$ with $Z \cap \overline{\mathrm{pa}}(v) = \emptyset$,

(ii) $Y \cap (Z \cup \{v\}) = \emptyset$ and (H_1, H_2) trek separates Y and $Z \cup \{v\}$ in the latent subgraph G_{lat},

(iii) there exists a system of treks with no sided intersection from Y to $\overline{\mathrm{pa}}(v) \cup Z$ in G such that the left part of every trek only takes edges in G_{lat}, and the right part of every trek ending in Z only takes edges in G_{lat}.

Algorithm and Software

Theorem (LSC-identifiability)

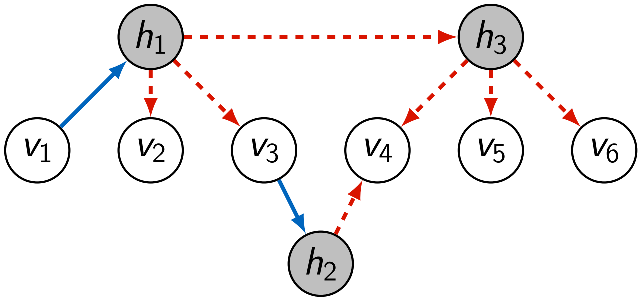
If the tuple $(Y, Z, H_1, H_2) \in 2^{\mathcal{O}\setminus\{v\}} \times 2^{\mathcal{O}\setminus\{v\}} \times 2^{\mathcal{L}} \times 2^{\mathcal{L}}$ *satisfies the LSC for* $v \in \mathcal{O}$*, then* $\overline{\Lambda}_{*,v}$ *is a a rational function of the observed covariance matrix* Σ*, the semi-direct effects* $(\overline{\Lambda}_{*,z})_{z\in Z}$*, and the semi-direct effects* $(\overline{\Lambda}_{*,y})_{y\in y}$ *for those* $y \in Y$ *that can be reached from* $Z \cup \{v\}$ *via a trek whose left part only takes edges in* G_{lat} *and avoids* H_2, H_1.

Algorithm:

Recursively cycle through nodes $v \in \mathcal{O}$ to find LSC-tuples solving for $\overline{\Lambda}_{*,v}$.

- Find LSC-tuples via integer linear program that generalizes classical max-flow problem.
- Sound and complete for determining LSC-identifiability.
- GitHub: `NilsSturma/LSC`; soon on `CRAN` within `SEMID` package.

Example



The graph is LSC-identifiable:

$\underline{v = v_1:}$ $(Y, Z, H_1, H_2) = (\emptyset, \emptyset, \emptyset, \emptyset)$.

$\underline{v = v_2, v_3, v_5, v_6:}$ $(Y, Z, H_1, H_2) = (\{v_1\}, \emptyset, \emptyset, \emptyset)$ and $\Pi = \{v_1\}$.

$\underline{v = v_4:}$ $(Y, Z, H_1, H_2) = (\{v_1, v_2, v_3\}, \{v_5\}, \emptyset, \{h_1\})$ and

$$\Pi = \{v_1, \ v_3, \ v_2 \leftarrow h_1 \rightarrow h_2 \rightarrow v_5\}.$$

Conclusion

EPFL

- Identifiability of semi-direct effects allows to transform the model into a simpler measurement model.
- Many applications require modeling effects of latent variables. Still lots to explore ...
- Formulas motivate easy-to-use "plug-in" estimators.

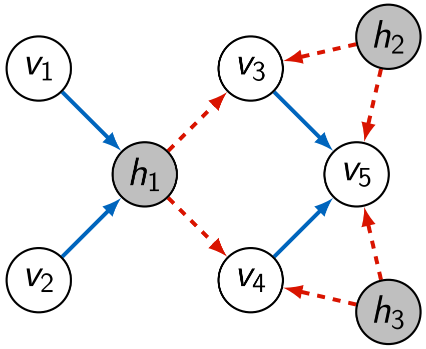
Articles:

Sturma, Drton (2025).
Trek-Based Parameter Identification for Linear Causal Models With Arbitrarily Structured Latent Variables.
arXiv:2507.18170

Barber, Drton, Sturma, Weihs (2022).
Half-Trek Criterion for Identifiability of Latent Variable Models.
Ann. Stat. 50, no. 6, 3174–3196.

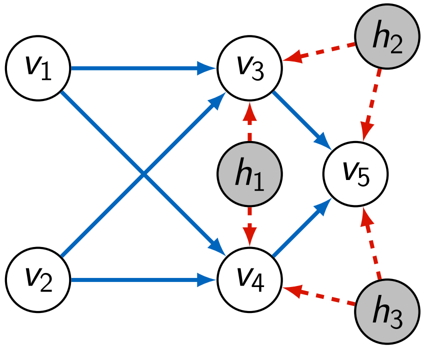
Appendix: Subtleties of Canonicalizations

Original graph:



LSC not applicable and not rationally identifiable.

Canonicalization:



LSC applicable $\Longrightarrow$ rationally identifiable.

Subtlety: Identifiability may be due to assumption that causal effects are not mediated by the same latent variable!