Half-Trek Criterion for Identifiability of Latent Variable Models

at 17. Doktorand:innentreffen der Stochastik

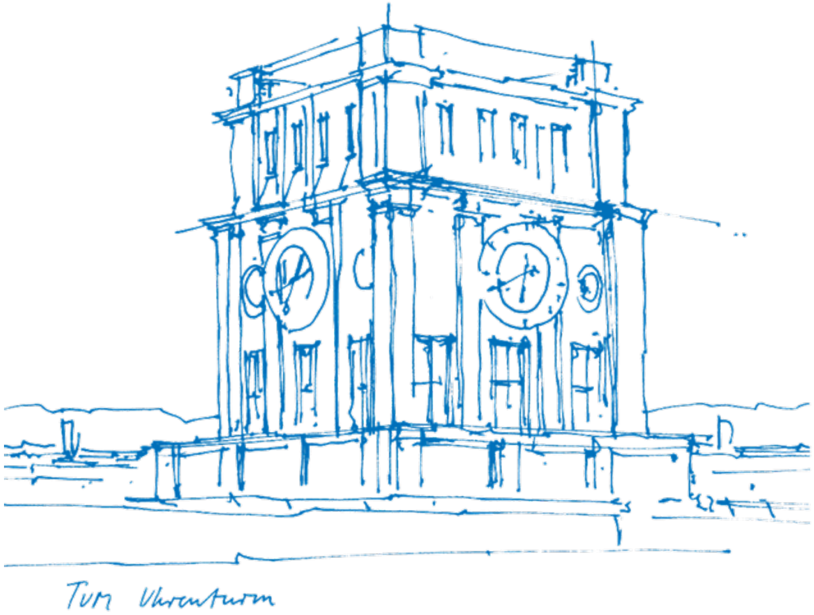
Nils Sturma

Research group Mathematical Statistics

Department of Mathematics

Technical University of Munich (TUM)

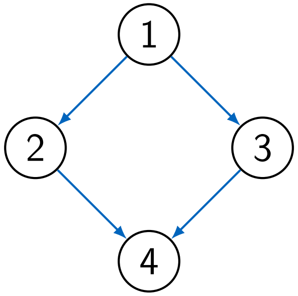
(joint work with Rina Foygel Barber, Mathias Drton, Luca Weihs)



Mathematical Statistics at TUM

About our Research Group

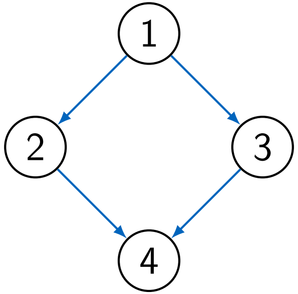
- Chair of Mathias Drton
- Multivariate and high-dimensional Statistics
- Graphical Models and Causality
- Algebraic Statistics



Mathematical Statistics at TUM

About our Research Group

- Chair of Mathias Drton
- Multivariate and high-dimensional Statistics
- Graphical Models and Causality
- Algebraic Statistics



In Klagenfurt...

Philipp Dettling

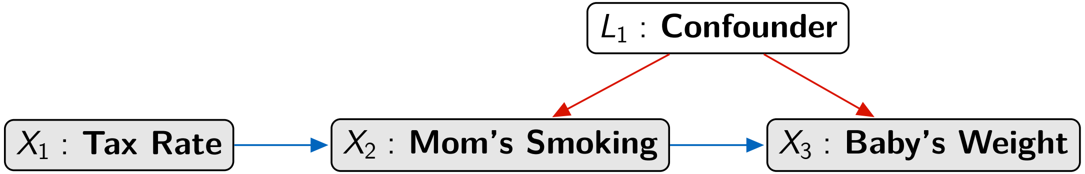
Leopold Mareis

David Strieder

Nils Sturma

Linear Structural Equation/ Causal Models

Each model is induced by a directed graph:



Linear structural equations:

$$
\begin{aligned}
X_1 &= \varepsilon_1, \\
X_2 &= \lambda_{12} X_1 + \gamma_2 L_1 + \varepsilon_2, \\
X_3 &= \lambda_{23} X_2 + \gamma_3 L_1 + \varepsilon_3, \\
L_1 &= \varepsilon_\ell.
\end{aligned}
$$

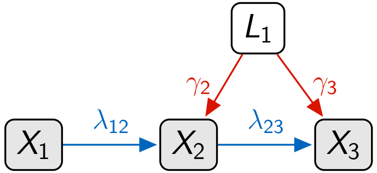
Independent errors:

$$\varepsilon_1 \perp\!\!\!\perp \varepsilon_2 \perp\!\!\!\perp \varepsilon_3 \perp\!\!\!\perp \varepsilon_\ell$$

$$\mathrm{Var}[\varepsilon_v] = \omega_v < \infty$$

Topic of the talk: If L_1 is latent, can we recover the direct effects $(\lambda_{12}, \lambda_{23})$ from $\Sigma = \mathrm{Var}[X]$?

Example: Instrumental Variable Model



$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} 0 & \lambda_{12} & 0 \\ 0 & 0 & \lambda_{23} \\ 0 & 0 & 0 \end{pmatrix}^T \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} + \begin{pmatrix} 0 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} L_1 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

Observed covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \cdot & \sigma_{22} & \sigma_{23} \\ \cdot & \cdot & \sigma_{33} \end{pmatrix} = \begin{pmatrix} \omega_1 & \boxed{\omega_1\lambda_{12}} & \boxed{\omega_1\lambda_{12}}\lambda_{23} \\ \cdot & \omega_2 + \gamma_2^2 + \omega_1\lambda_{12}^2 & \gamma_2\gamma_3 + \lambda_{23}\sigma_{22} \\ \cdot & \cdot & \omega_3 + \gamma_3^2 + 2\gamma_2\gamma_3\lambda_{23} + \lambda_{23}^2\sigma_{22} \end{pmatrix}$$

We see that

$$\lambda_{12} = \frac{\sigma_{12}}{\sigma_{11}} \qquad \text{with } \sigma_{11} > 0,$$

$$\lambda_{23} = \frac{\sigma_{13}}{\sigma_{12}} \qquad \text{with } \sigma_{12} = \omega_1\lambda_{12} \neq 0 \text{ ‘almost surely’}.$$

Setup

Variables:

Observed: $X = (X_v)_{v \in V}$ Latent: $L = (L_h)_{h \in \mathcal{L}}$

Graph:

Directed graph $G = (V \dot{\cup} \mathcal{L}, D)$ with directed cycles allowed.

Latent-factor assumption:

All latent variables are latent factors $\equiv$ all nodes in $\mathcal{L}$ are source nodes of G.

Structural equation model:

$$X = \Lambda^\top X + \Gamma^\top L + \varepsilon$$

- all latent factors and error terms in (L, ε) are mutually independent, so $\Omega_{\text{diag}} = \text{Var}[\varepsilon] = \text{diag}(\omega_v : v \in V)$ diagonal, and $\text{Var}[L] = I$ without loss of generality.
- parameter matrices Λ and Γ are sparse and supported over edge set D.

Identifiability

- Every latent-factor graph G yields a parametrization of the observed covariance matrix:

$$\phi_G : (\Lambda, \Gamma, \Omega_{\text{diag}}) \longmapsto \underbrace{(I - \Lambda)^{-\top}(\Omega_{\mathrm{diag}} + \Gamma^\top \Gamma)(I - \Lambda)^{-1}}_{=\Sigma=\text{Var}[X]}.$$

- The model given by G is rationally identifiable if

$$\exists \text{ rational map } \psi_G \ : \quad \psi_G \circ \phi_G(\Lambda, \Gamma, \Omega_{\text{diag}}) = \Lambda \quad \text{for 'almost all'} \ (\Lambda, \Gamma, \Omega_{\text{diag}}).$$

Identifiability

- Every latent-factor graph G yields a parametrization of the observed covariance matrix:

$$\phi_G : (\Lambda, \Gamma, \Omega_{\text{diag}}) \longmapsto \underbrace{(I - \Lambda)^{-\top}(\Omega_{\text{diag}} + \Gamma^{\top}\Gamma)(I - \Lambda)^{-1}}_{=\Sigma=\text{Var}[X]}.$$

- The model given by G is rationally identifiable if

$$\exists \text{ rational map } \psi_G \ : \quad \psi_G \circ \phi_G(\Lambda, \Gamma, \Omega_{\text{diag}}) = \Lambda \quad \text{for 'almost all'} \ (\Lambda, \Gamma, \Omega_{\text{diag}}).$$

- The problem may be solved via a Gröbner basis computation. . . on small scale.

- Main Contribution:
  - Sufficient graphical condition for rational identifiability.
  - Recursive polynomial time algorithm.

    (caveat: polynomial time when bounding a matrix rank in a search step)
  - Condition is not necessary but 'effective'; see simulations in paper.

Using Algebraic Relations in Latent Covariance Matrix

- Latent covariance matrix

$$\Omega \;\equiv\; \mathrm{Var}[\Gamma^\top L + \varepsilon] \;=\; \mathrm{Var}[\varepsilon] + \Gamma^\top \mathrm{Var}[L]\Gamma \;=\; \Omega_{\mathrm{diag}} + \Gamma^\top \Gamma.$$

- Observe that

$$\Sigma = (I - \Lambda)^{-\top}\Omega(I - \Lambda)^{-1} \iff \boxed{\Omega \;=\; (I - \Lambda)^T \Sigma (I - \Lambda)}$$

Using Algebraic Relations in Latent Covariance Matrix

- Latent covariance matrix

$$\Omega \;\equiv\; \mathrm{Var}[\Gamma^\top L + \varepsilon] \;=\; \mathrm{Var}[\varepsilon] + \Gamma^\top \mathrm{Var}[L]\Gamma \;=\; \Omega_{\mathrm{diag}} + \Gamma^\top\Gamma.$$

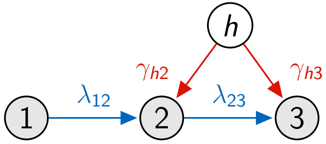
- Observe that

$$\Sigma = (I - \Lambda)^{-\top}\Omega(I-\Lambda)^{-1} \iff \boxed{\Omega \;=\; (I-\Lambda)^T\Sigma(I-\Lambda)}$$

- Algebraic relations between entries of $\Omega = \Omega_{\mathrm{diag}} + \Gamma^\top\Gamma$ yield relations between entries of Λ and Σ:

$$f(\Omega) \;=\; 0 \iff f((I-\Lambda)^T\Sigma(I-\Lambda)) = 0.$$

- Example:

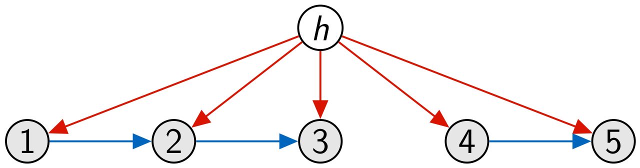


$$\Omega = \begin{pmatrix} \omega_1 & 0 & \mathbf{0} \\ 0 & \omega_2 + \gamma_{h2}^2 & \gamma_{h2}\gamma_{h3} \\ \mathbf{0} & \gamma_{h2}\gamma_{h3} & \omega_3 + \gamma_{h3}^2 \end{pmatrix}$$

$$\left[(I-\Lambda)^T\Sigma(I-\Lambda)\right]_{13} = \sigma_{13} - \lambda_{23}\,\sigma_{12} = 0$$

Latent Low Rank Structure

- Lots of existing work is based on using zero entries in latent covariance matrix.
- However, the resulting methods cannot cover situations such as



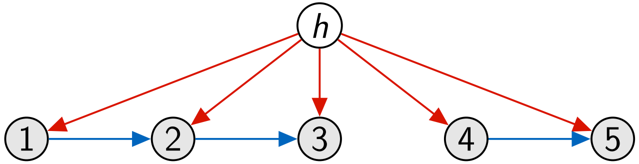
where the latent covariance matrix is dense:

$$\Omega = \Omega_{\text{diag}} + \gamma_h \gamma_h^\top = \text{diagonal} + \text{dense rank 1}.$$

- New paper: Generalize beyond zeros by exploiting

latent low rank structure.

Example: Latent Low Rank Structure



$$\Omega = \begin{pmatrix} \omega_1 & \gamma_{h1}\gamma_{h2} & \gamma_{h1}\gamma_{h3} & \gamma_{h1}\gamma_{h4} & \gamma_{h1}\gamma_{h5} \\ \gamma_{h1}\gamma_{h2} & \omega_2 + \gamma_{h2}^2 & \gamma_{h2}\gamma_{h3} & \gamma_{h2}\gamma_{h4} & \gamma_{h2}\gamma_{h5} \\ \gamma_{h1}\gamma_{h3} & \gamma_{h2}\gamma_{h3} & \omega_3 + \gamma_{h3}^2 & \gamma_{h3}\gamma_{h4} & \gamma_{h3}\gamma_{h5} \\ \gamma_{h1}\gamma_{h4} & \gamma_{h2}\gamma_{h4} & \gamma_{h3}\gamma_{h4} & \omega_4 + \gamma_{h4}^2 & \gamma_{h4}\gamma_{h5} \\ \gamma_{h1}\gamma_{h5} & \gamma_{h2}\gamma_{h5} & \gamma_{h3}\gamma_{h5} & \gamma_{h4}\gamma_{h5} & \omega_5 + \gamma_{h5}^2 \end{pmatrix}$$

Rank-deficient off-diagonal submatrix:

$$\Omega_{\{1,2\},\{3,4\}} = \begin{pmatrix} \gamma_{h1}\gamma_{h3} & \gamma_{h1}\gamma_{h4} \\ \gamma_{h2}\gamma_{h3} & \gamma_{h2}\gamma_{h4} \end{pmatrix} = \begin{pmatrix} \gamma_{h1} \\ \gamma_{h2} \end{pmatrix} \cdot \begin{pmatrix} \gamma_{h3} & \gamma_{h4} \end{pmatrix} \quad \Longrightarrow \quad \det(\Omega_{\{1,2\},\{3,4\}}) = 0.$$

Relations between Λ and Σ:

$$\det\left(\left[(I - \Lambda)^T \Sigma (I - \Lambda)\right]_{\{1,2\},\{3,4\}}\right) = \lambda_{23}\sigma_{12}\sigma_{24} - \lambda_{23}\sigma_{14}\sigma_{22} - \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23} = 0.$$

We see that

$$\lambda_{23} = \frac{\sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23}}{\sigma_{12}\sigma_{24} - \sigma_{14}\sigma_{22}} \qquad \text{with } \sigma_{12}\sigma_{24} - \sigma_{14}\sigma_{22} \neq 0 \text{ 'almost surely'}.$$

New Latent-Factor Half-Trek Criterion: Main Idea

- Digraph $(V \dot\cup \mathcal{L}, D)$ with observed variables in V and latent variables in $\mathcal{L}$.
- Recursive search for linear equation systems that determine columns $\Lambda_{\mathrm{pa}(v),v}$, $v \in V$.

New Latent-Factor Half-Trek Criterion: Main Idea

- Digraph $(V \dot{\cup} \mathcal{L}, D)$ with observed variables in V and latent variables in $\mathcal{L}$.

- Recursive search for linear equation systems that determine columns $\Lambda_{\text{pa}(v),v}$, $v \in V$.

- To this end, we find a rank-deficient off-diagonal submatrix

$$\Omega_{Y,Z\cup\{v\}} = \left[(I-\Lambda)^T \Sigma (I-\Lambda)\right]_{Y,Z\cup\{v\}} \quad \text{with } |Y| = |Z| + |\text{pa}(v)|.$$

- Our combinatorial conditions ensure a generically unique solution. In particular, we can write $Y = Y_Z \dot{\cup} Y_{\text{pa}(v)}$ such that $\det(\Omega_{Y_Z,Z} \neq 0)$ but

$$\det(\Omega_{Y_Z\cup\{w\},Z\cup\{v\}}) = 0 \quad \text{for all } w \in Y_{\text{pa}(v)}.$$

New Latent-Factor Half-Trek Criterion: Main Idea

- Digraph $(V \dot\cup \mathcal{L}, D)$ with observed variables in V and latent variables in $\mathcal{L}$.
- Recursive search for linear equation systems that determine columns $\Lambda_{\mathrm{pa}(v),v}$, $v \in V$.
- To this end, we find a rank-deficient off-diagonal submatrix

$$\Omega_{Y,Z\cup\{v\}} = \left[(I-\Lambda)^T \Sigma (I-\Lambda)\right]_{Y,Z\cup\{v\}} \quad \text{with } |Y| = |Z| + |\mathrm{pa}(v)|.$$

- Our combinatorial conditions ensure a generically unique solution. In particular, we can write $Y = Y_Z \dot\cup Y_{\mathrm{pa}(v)}$ such that $\det(\Omega_{Y_Z,Z} \neq 0)$ but

$$\det(\Omega_{Y_Z\cup\{w\},Z\cup\{v\}}) = 0 \quad \text{for all } w \in Y_{\mathrm{pa}(v)}.$$

- A half-trek from node v to node w is a path of the form:

$$v \to x_1 \to \ldots \to x_\ell \to w \quad \text{or} \quad v \leftarrow \ell \to x_1 \to \ldots \to x_\ell \to w.$$

Relevance: Entries of $(I-\Lambda)^T\Sigma$ are sums over half-treks.

Latent-Factor Half-Trek Criterion (LF-HTC)

Definition

Let $v \in V$ and $Y, Z \subseteq V \setminus \{v\}$ and $H \subseteq \mathcal{L}$. Triple (Y, Z, H) satisfies latent-factor half-trek criterion for v if

(i) $|Y| = |\text{pa}(v)| + |H|$ and $|Z| = |H|$;

(ii) $Y \cap (Z \cup \{v\}) = \emptyset$ and $[\text{pa}_{\mathcal{L}}(Y) \cap \text{pa}_{\mathcal{L}}(Z \cup \{v\})] \subseteq H$;

(iii) There is a system of half-treks from Y to $\text{pa}(v) \cup Z$ without sided intersection and all half-treks ending in Z have form $y \leftarrow \ell \rightarrow z$ for $\ell \in H$.

Theorem

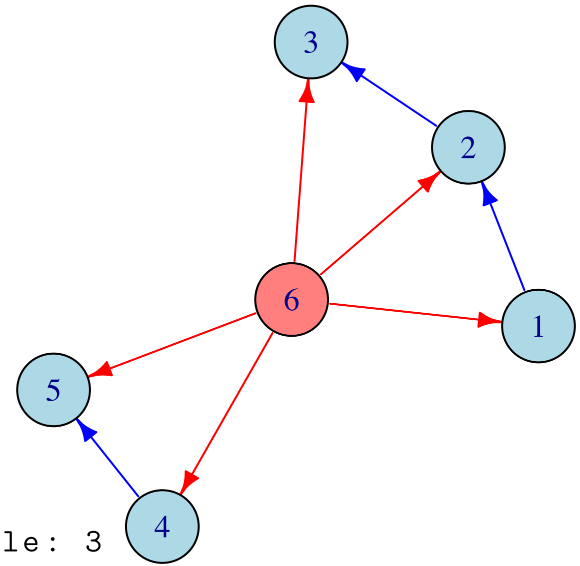
"If the triple (Y, Z, H) satisfies the LF-HTC for $v \in V$ then column $\Lambda_{*,v}$ is a rational function of Σ and certain other columns of Λ."

Our Software: SEMID (R Package)

Algorithm: Recursive Solving.

- Cycle through nodes v and search for LF-HTC triples (Y,Z,H) that allow solving for $\Lambda_{*,v}$.
- Network-flow setup finds LF-HTC triples in polynomial time under a bound on $|Z| = |H|$.

```
> # Define graph
> L = matrix(c(0, 1, 0, 0, 0, 0,
+              0, 0, 1, 0, 0, 0,
+              0, 0, 0, 0, 0, 0,
+              0, 0, 0, 0, 1, 0,
+              0, 0, 0, 0, 0, 0,
+              1, 1, 1, 1, 1, 0), 6, 6, byrow=TRUE)
> observedNodes = seq(1,5)
> latentNodes = c(6)
> g = LatentDigraph(L, observedNodes, latentNodes)
>
> # Check identifiability
> lfhtcID(g)
[1] nr. of edges between observed nodes shown rat. identifiable: 3
[2] rat. identifiable edges: 1->2, 2->3, 4->5
```



Conclusion

- Many applications require modeling effects of latent variables.
- Latent variable models may feature complicated parametrizations and geometry.
- Lots to explore still, in identification and for other problems...

Preprint:

Barber, Drton, Sturma, Weihs (2022).
Half-Trek Criterion for Identifiability of Latent Variable Models. arXiv:2201.04457.