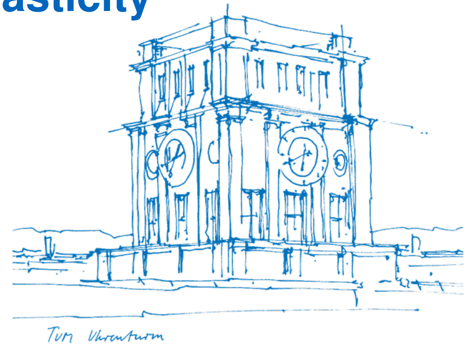
Munich Center for Machine Learning (MCML)
TUM School of Computation, Information and Technology
Technical University of Munich

Identifying Total Causal Effects in Linear Models under Partial Homoscedasticity

David Strieder, Mathias Drton

Probabilistic Graphical Models (PGM)
Nijmegen, September 11-13, 2024



Starting point

Munich Center for Machine Learning

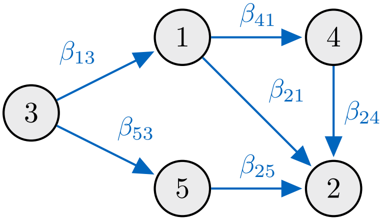
Research question: What is the total causal effect of X_i on X_j? Confidence?

- Only observational data available.
- Underlying causal structure is unknown.

Starting point
Underlying Linear SCM with Gaussian errors

Example:

$$
\begin{aligned}
X_1 &= \beta_{13} X_3 + \varepsilon_1 \\
X_2 &= \beta_{21} X_1 + \beta_{24} X_4 + \beta_{25} X_5 + \varepsilon_2 \\
X_3 &= \varepsilon_3 \\
X_4 &= \beta_{41} X_1 + \varepsilon_4 \\
X_5 &= \beta_{25} X_3 + \varepsilon_5
\end{aligned}
$$



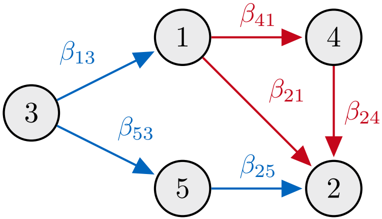
where $\varepsilon_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_j^2)$

Starting point

Underlying Linear SCM with Gaussian errors

Example: Causal effect $\mathcal{C}(1 \to 2) := \frac{\mathrm{d}}{\mathrm{d}x_1}\mathbb{E}[X_2 | \,\mathrm{do}(X_1 = x_1)] = \varphi(G, \Sigma)$.

$$
\begin{aligned}
X_1 &= \beta_{13} X_3 + \varepsilon_1 \\
X_2 &= \beta_{21} X_1 + \beta_{24} X_4 + \beta_{25} X_5 + \varepsilon_2 \\
X_3 &= \varepsilon_3 \\
X_4 &= \beta_{41} X_1 + \varepsilon_4 \\
X_5 &= \beta_{25} X_3 + \varepsilon_5
\end{aligned}
$$



where $\varepsilon_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_j^2)$

Identifiabilty

Question: Given an observational distribution P_{Σ_0} from a set of possible distributions $\{P_\Sigma : \Sigma \in \mathcal{M}\}$, can we recover the (causal) target of interest $\mathcal{C}(i \to j)$?

Identifiabilty

Munich Center for Machine Learning

Question: Given an observational distribution P_{Σ_0} from a set of possible distributions $\{P_\Sigma : \Sigma \in \mathcal{M}\}$, can we recover the (causal) target of interest $\mathcal{C}(i \to j)$?

- Each DAG generates (sub-)set $\mathcal{M}(G) \subseteq \mathcal{M}$.
- What happens if $\Sigma_0 \in \mathcal{M}(G^1) \cap \mathcal{M}(G^2)$? $\varphi(G^1, \Sigma_0) = \varphi(G^2, \Sigma_0)$?

Identifiabilty

Question: Given an observational distribution P_{Σ_0} from a set of possible distributions $\{P_\Sigma : \Sigma \in \mathcal{M}\}$, can we recover the (causal) target of interest $\mathcal{C}(i \to j)$?

- Each DAG generates (sub-)set $\mathcal{M}(G) \subseteq \mathcal{M}$.
- What happens if $\Sigma_0 \in \mathcal{M}(G^1) \cap \mathcal{M}(G^2)$? $\varphi(G^1, \Sigma_0) = \varphi(G^2, \Sigma_0)$?
- Only need to consider all complete DAGs $\mathcal{G}(d)$.

Identifiability
Arbitrary error variances

Linear Gaussian SCM with arbitrary error variances: $\mathcal{M} = \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}(G)$, where

$$\mathcal{M}(G) = \left\{ \Sigma \in \text{PD}(d) : \exists B \in \mathbb{R}^G, \omega \in \mathbb{R}^d \text{ with } \Sigma = (I_d - B)^{-1} \text{diag}(\omega) (I_d - B)^{-T} \right\}$$

Theorem (see, e.g., Pearl (2009)). *The total causal effect $\mathcal{C}(i \to j)$ is not generically identifiable under the assumption of an underlying linear Gaussian SCM.*

Identifiability
Equal error variances

Linear Gaussian SCM with partially equal error variances: $\mathcal{M}_{EV} = \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}_{EV}(G)$, where

$$\mathcal{M}_{EV}(G) = \Big\{\Sigma \in \text{PD}(d) : \exists \omega > 0 \text{ with } \omega = \Sigma_{k,k|p(k)} \quad \forall\, k = 1, \dots, d\Big\}.$$

Theorem (Peters and Bühlmann (2014))**.** *The total causal effect $\mathcal{C}(i \to j)$ is globally identifiable under the assumption of an underlying linear Gaussian SCM with equal error variances.*

Identifiability
Partially equal error variances

Linear Gaussian SCM with partially equal error variances: $\mathcal{M}_{PEV} = \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}_{PEV}(G)$, where

$$\mathcal{M}_{PEV}(G) = \left\{ \Sigma \in \mathsf{PD}(d) : \Sigma_{i,i|p(i)} = \Sigma_{j,j|p(j)} \right\}.$$

Identifiability

Partially equal error variances

Linear Gaussian SCM with partially equal error variances: $\mathcal{M}_{PEV} = \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}_{PEV}(G)$, where

$$\mathcal{M}_{PEV}(G) = \left\{ \Sigma \in \mathsf{PD}(d) : \Sigma_{i,i|p(i)} = \Sigma_{j,j|p(j)} \right\}.$$

Theorem. *The total causal effect $\mathcal{C}(i \to j)$ is generically identifiable under the assumption of an underlying linear Gaussian SCM with equal error variances among i and j.*

Causal Inference under Structure Uncertainty

Question: Given observational data from P_{Σ_0} out of a set of possible distributions $\{P_\Sigma : \Sigma \in \mathcal{M}\}$, can we construct a confidence region for $\mathcal{C}(i \to j)$?

Causal Inference under Structure Uncertainty

Question: Given observational data from P_{Σ_0} out of a set of possible distributions $\{P_\Sigma : \Sigma \in \mathcal{M}\}$, can we construct a confidence region for $\mathcal{C}(i \to j)$?

- Statistical uncertainty in the numerical size of the effect.
- Structural uncertainty in the underlying causal graph.
- Causal uncertainty due to equivalence of multiple models.

Causal Inference under Structure Uncertainty

Question: Given observational data from P_{Σ_0} out of a set of possible distributions $\{P_\Sigma : \Sigma \in \mathcal{M}\}$, can we construct a confidence region for $\mathcal{C}(i \to j)$?

- **Main Idea:** Use test inversion.
- Employ Dual-LRT for the arising testing problem, that is, for all $\psi \in \mathbb{R}$ and $G \in \mathcal{G}(d)$ invert joint test of structure and effect size:

$$\mathrm{H}_0^{(\psi)}(G) : \Sigma \in \mathcal{M}_\psi(G) \quad \text{against} \quad \mathrm{H}_1 : \Sigma \in \mathcal{M} \backslash \mathcal{M}_\psi(G).$$

Causal Inference under Structure Uncertainty
Arbitrary error variances

Theorem. *An asymptotic $(1-\alpha)$-confidence set for the total causal effect $\mathcal{C}(i \to j)$ under the assumption of an underlying linear Gaussian SCM is given by*

$$\bigcup_{G \in \mathcal{G}(d) \,:\, D(G) \geq 0} [L(G), U(G)] \bigcup \{0\},$$

where

$$L(G) := \frac{-(\widehat{\Sigma}^{-1})_{j,i|d(i)\setminus\{j\}} - \sqrt{D(G)}}{(\widehat{\Sigma}^{-1})_{j,j|d(i)\setminus\{j\}}}, \qquad U(G) := \frac{-(\widehat{\Sigma}^{-1})_{i,j|d(i)\setminus\{j\}} + \sqrt{D(G)}}{(\widehat{\Sigma}^{-1})_{j,j|d(i)\setminus\{j\}}}.$$

Causal Inference under Structure Uncertainty

Partially equal error variances

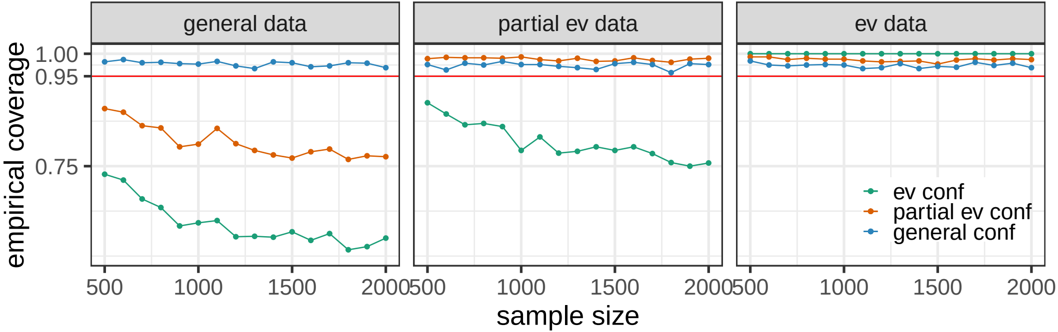
Theorem. *An asymptotic* $(1-\alpha)$*-confidence set for the total causal effect* $\mathcal{C}(i \to j)$ *under the assumption of an underlying linear Gaussian SCM with (partially) equal error variance among* i *and* j *is given by*

$$\bigcup_{G \in \mathcal{G}(d)\,:\,D(G) \geq 0} [L(G), U(G)] \bigcup \{0 : Z \leq K \exp\left(\frac{\chi^2_{1,1-\alpha}}{2n}\right)\},$$

where

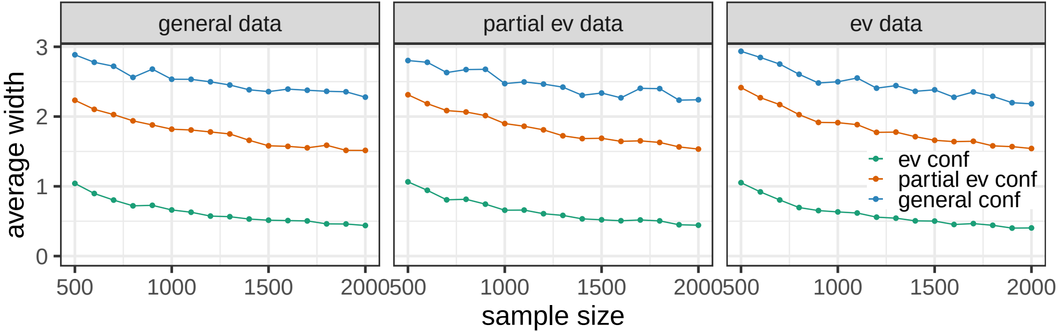
$$L(G) := \frac{-(\widehat{\Sigma}^{-1})_{j,i|d(i)\setminus\{j\}} - \sqrt{D(G)}}{(\widehat{\Sigma}^{-1})_{j,j|d(i)\setminus\{j\}}}, \qquad U(G) := \frac{-(\widehat{\Sigma}^{-1})_{i,j|d(i)\setminus\{j\}} + \sqrt{D(G)}}{(\widehat{\Sigma}^{-1})_{j,j|d(i)\setminus\{j\}}}.$$

Simulations



Empirical coverage of 95%-confidence intervals for the total causal effect.

Simulations



Mean width of 95%-confidence intervals for the total causal effect.

Conclusion

- The total causal effect is generically identifiable under partial homoscedasticity.
- Closed-form solution for constructing confidence regions for total causal effects that:
  - account for **causal structure uncertainty**
  - as well as **statistical uncertainty** about the numerical size of the effect.