Graphical criteria for efficient total effect estimation

Leonard Henckel, Emilija Perković, and Marloes H. Maathuis

Seminar for Statistics, ETH Zurich

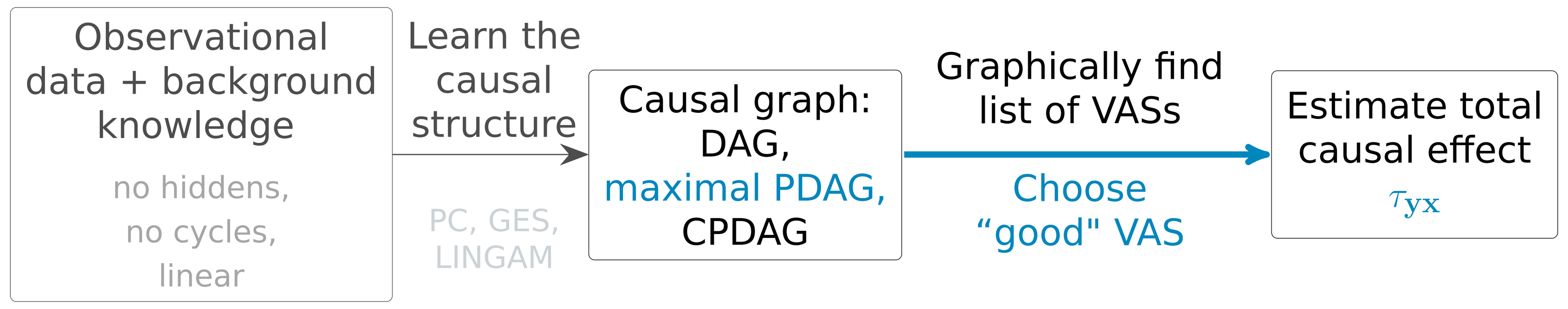
ETH zürich

Problem To consistently estimate a total causal effect from observational data an appropriate adjustment set needs to be chosen. The adjustment criterion [3] is a necessary and sufficient graphical criterion to identify those that provide consistent estimates, but typically multiple such valid adjustment sets (VAS) exist. Which of them should one ideally use?

Idea Develop graphical criteria to identify efficient valid adjustment sets in the linear causal model setting.

Results Graphical criteria for efficient total effect estimation:
1. Graphical criterion for qualitative asymptotic variance comparisons.
2. Identification of an asymptotically optimal valid adjustment set.
3. Asymptotic variance reducing, stable pruning procedure.

Framework



Criteria for adjustment set selection:

- Unbiased total effect estimate, i.e. a valid adjustment set
- Asymptotic variance of total effect estimate
- Amount of graphical information required
- Ease of data collection and size
- Stability under graph estimation errors

We consider the asymptotic variance of the total effect estimates.

Background

Theorem [2] $\mathbf{Z}$ is a **valid adjustment set** relative to (X, Y) in $\mathcal{G}$ iff

(Forbidden set) $\mathbf{Z}$ does not contain nodes in $\text{Forb}(X, Y, \mathcal{G})$.

(Blocking) $\mathbf{Z}$ blocks all **non-causal** paths from X to Y in $\mathcal{G}$.

Definition $\mathbf{V} = (V_1, \dots, V_p)^T$, $p \geq 1$ follows a **linear causal model** iff

(Causality) $\mathbf{V}$ is compatible with a causal DAG $\mathcal{G}$,

(Linearity) such that for all $V_i \in \mathbf{V}$

$$V_i = \sum_{V_j \in pa(V_i, \mathcal{G})} \alpha_{ij} V_j + \epsilon_i, \ \alpha_{ij} \in \mathbb{R},$$

with $\epsilon_1, \dots, \epsilon_p$ jointly independent, mean 0 with finite variance.

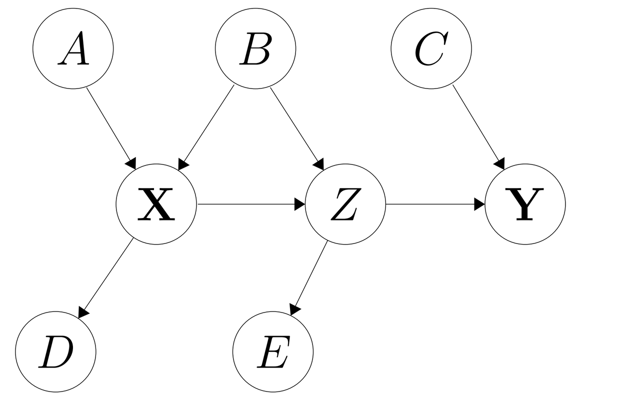
Given a VAS the total causal effect of X on Y in a CLM setting is the linear regression coefficient $\beta_{yx.\mathbf{z}}$ of X in the regression $Y \sim X + \mathbf{Z}$ [1].

Lemma $\mathbf{Z}$ a VAS in a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, such that $\mathbf{V}$ follows a causal linear model, then

$$\text{a.var}(\hat{\tau}^{\mathbf{z}}_{\mathbf{yx}}) = \text{a.var}(\hat{\beta}_{\mathbf{yx.z}}) = \frac{\sigma_{\mathbf{yy.xz}}}{\sigma_{\mathbf{xx.z}}}.$$

- $\hat{\beta}_{yx.\mathbf{z}}$ the least squares estimator of $\beta_{yx.\mathbf{z}}$
- $\sigma_{xx.\mathbf{z}} = var(X - \beta_{x\mathbf{z}}\mathbf{Z})$ and $\sigma_{yy.x\mathbf{z}} = var(Y - \beta_{yx.\mathbf{z}}X - \beta_{y\mathbf{z}.x}\mathbf{Z})$ the variance of the population error term of the regression specified by their respective subscripts

Example



$\tau^{\mathbf{z}}_{\mathbf{yx}} = \alpha_{\mathbf{yz}}\alpha_{\mathbf{zx}} = \mathbf{1}$

$\text{Forb}(X, Y, \mathcal{G}) = \{X, Z, Y, E\}$

$\text{cn}(X, Y, \mathcal{G}) = \{Z, Y\}$

```
n <- 100000
a <- runif(n)
b <- rnorm(n)
c <- rnorm(n)
x <- a + b + rnorm(n)
z <- x + b + rnorm(n)
y <- z + c + rnorm(n)
d <- x + rnorm(n)
e <- z + rnorm(n)

fit <- lm(y~x + S -1)
```

$\mathbf{S}$	$\hat{\beta}_{yx.\mathbf{s}}$	$\widehat{a.var}(\hat{\beta}_{yx.\mathbf{s}})$	$a.var(\hat{\beta}_{yx.\mathbf{s}})$
$\emptyset$	1.33	1.21	1.22
$\{a\}$	1.50	1.73	1.75
$\{b\}$	1.00	1.49	1.50
$\{c\}$	1.33	0.89	0.89
$\{z\}$	0.00	2.77	2.80
$\{a, b\}$	1.01	2.98	3.00
$\{b, c\}$	1.00	1.00	1.00
$\{b, d\}$	1.00	4.48	4.50
$\{b, e\}$	0.50	2.48	2.50
$\{a, b, c\}$	1.00	1.99	2.00

Results

Asymptotic variance comparison criterion: $\mathbf{Z_1}$ and $\mathbf{Z_2}$ two valid adjustment sets in a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, such that $\mathbf{V}$ follows a causal linear model. If

$$i)\ \mathbf{Z_1} \setminus \mathbf{Z_2} \perp_{\mathcal{G}} Y | \mathbf{Z_2} \cup X \textit{ and } ii)\ \mathbf{Z_2} \setminus \mathbf{Z_1} \perp_{\mathcal{G}} X | \mathbf{Z_1},$$

then $\text{a.var}(\hat{\tau}^{\mathbf{z_2}}_{\mathbf{yx}}) \leq \text{a.var}(\hat{\tau}^{\mathbf{z_1}}_{\mathbf{yx}})$.

- $\perp_{\mathcal{G}}$ indicates **d-separation**; the graphical criterion through which the encoded conditional independences can be read of a causal graph [1].

The optimal VAS: Let $\mathbf{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \mathbf{pa}(\text{cn}(\mathbf{X}, \mathbf{Y}, \mathcal{G}), \mathcal{G}) \setminus \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ and assume that $Y \in de(X, \mathcal{G})$. Then

(VAS) $\mathbf{O}$ is a **VAS** relative to (X, Y) in $\mathcal{G}$.

(Asymptotic optimality) If $\mathbf{V}$ follows a linear causal model compatible with $\mathcal{G}$, then for any VAS $\mathbf{Z}$

$$\text{a.var}(\hat{\tau}^{\mathbf{o}}_{\mathbf{yx}}) \leq \text{a.var}(\hat{\tau}^{\mathbf{z}}_{\mathbf{yx}}).$$

(Minimality) Let $\mathbf{Z}$ be a VAS, such that $\text{a.var}(\hat{\tau}^{\mathbf{o}}_{\mathbf{yx}}) = \text{a.var}(\hat{\tau}^{\mathbf{z}}_{\mathbf{yx}})$. If $\mathbf{V}$ follows a linear causal model compatible with $\mathcal{G}$ and f is faithful to $\mathcal{G}$ then $\mathbf{O} \subseteq \mathbf{Z}$.

- Total effects on non-descendants are 0 and hence assuming $\{Y\} \subseteq de(X, \mathcal{G})$ only limits us from superfluously estimating some zero values.

VAS pruning procedure:

input : Causal DAG $\mathcal{G}$, disjoint node sets $\mathbf{X}$ and $\mathbf{Y}$ and a VAS $\mathbf{Z}$
output: **VAS** $\mathbf{Z}' \subseteq \mathbf{Z}$, such that $\text{a.var}(\hat{\tau}^{\mathbf{z}'}_{\mathbf{yx}}) \leq \text{a.var}(\hat{\tau}^{\mathbf{z}}_{\mathbf{yx}})$

```
begin
  Z' = Z;
  foreach  Z ∈ Z' do
    if  Y ⊥_G Z | Z'_{-z} ∪ X then
      Z' = Z'_{-z};
  return Z';
```

- No other subset of $\mathbf{Z}$ is guaranteed by our variance comparison criterion to have a better asymptotic variance than the output set $\mathbf{Z}' \subseteq \mathbf{Z}$.
- $\mathbf{Z}'$ will be the same regardless of the order in which the nodes in $\mathbf{Z}$ are considered.

Application of results to example

I) Identifying the **valid adjustment sets** $\mathbf{Z}$ in the example DAG:
- $\mathbf{Z}$ may not contain any nodes in $\text{Forb}(X, Y, \mathcal{G}) = \{X, Z, Y, E\}$.
- $\mathbf{Z}$ must contain B to block the one open non-causal path.

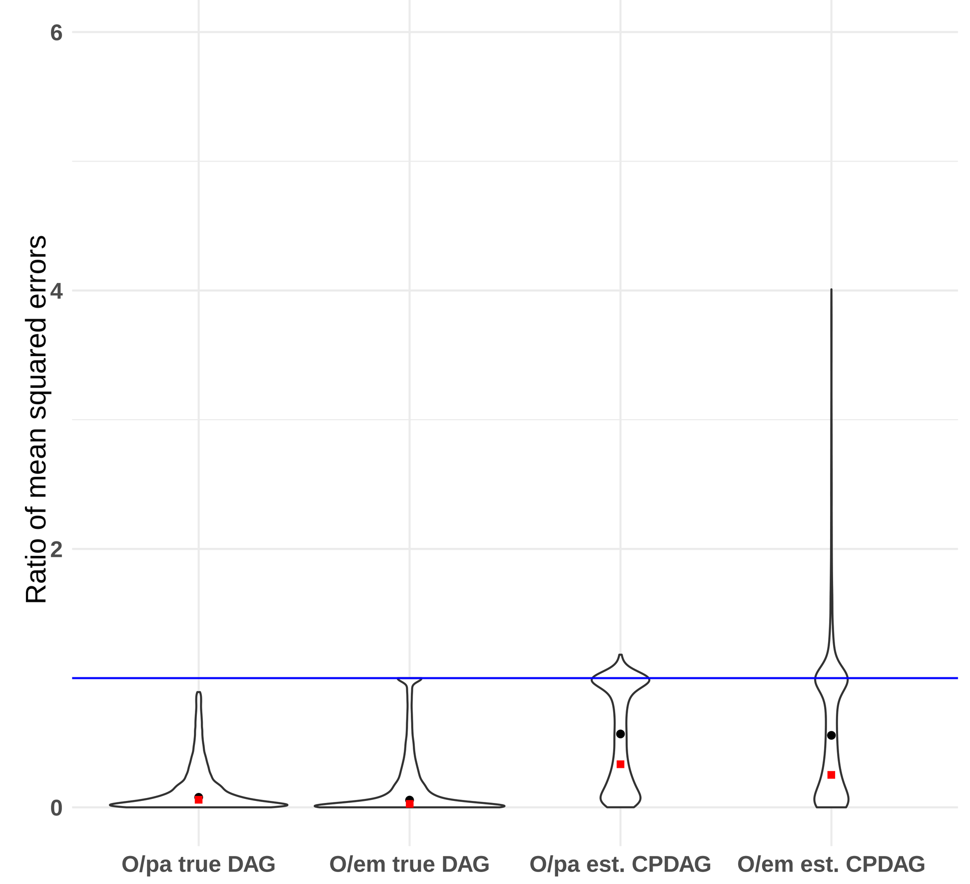
Therefore, $\mathbf{Z} = \{B\} \cup \mathbf{R}$, where $\mathbf{R} = \{A, C, D\}$.

II) Applying the **variance comparison criterion** and the **pruning procedure**:
- i) $A \perp_{\mathcal{G}} Y | X \cup \mathbf{Z} \setminus \{A\}$, ii) $D \perp_{\mathcal{G}} Y | X \cup \mathbf{Z} \setminus \{D\}$ and iii) $XC \perp_{\mathcal{G}} X | \mathbf{Z} \setminus \{C\}$.
- Including A and D is detrimental and C beneficial for the asymptotic variance.
- Output $\mathbf{Z}'$ is $\{B\}$ or $\{B, C\}$ depending on whether $C \in \mathbf{Z}$.

III) $\mathbf{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = pa(\{Z, Y\}, \mathcal{G}) \setminus \{X, Z, Y, E\} = \{B, C\}$, the VAS which in fact provides the smallest asymptotic variance among all VAS

Quantifying possible efficiency gains

We compared $\mathbf{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$with the typically used VAS $\mathbf{pa}(\mathbf{X}, \mathcal{G})$ and the non-causal/graphical **empty set** as a baseline, first using the true causal DAG and then estimating it.



10 000 sampled DAGs:
$p \in \{10, 20, 50, 100\}$,
$E[N] \in \{2, 3, 4, 5\}$,
$n \in \{125, 500, 2000, 8000\}$

X - randomly chosen,
$Y \in de(X, \mathcal{G})$

Figure showing mean squared error ratios. Values below 1 indicate improvement by using $\mathbf{O}$, the black dot is the median, the red one the geometric average of the ratios.

Remark

For simplicity we present simplfified results in this poster. Our complete results in the working paper hold for
1. **joint interventions**, i.e. sets $\mathbf{X}$ and $\mathbf{Y}$
2. larger classes of causal graphs, namely **maximally oriented PDAGs** [2] and
3. we only require the errors in the CLM to have **finite variance** and *not* that they be Gaussian.

References

[1] J. Pearl. *Causality*. Cambridge University Press, 2009.

[2] E. Perković, M. Kalisch, and M. H. Maathuis. Interpreting and using CPDAGs with background knowledge. In *UAI*, 2017.

[3] E. Perković, J. Textor, M. Kalisch, and M. H. Maathuis. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *Journal of Machine Learning Research*, 18(220):1–62, 2018.