Graphical criteria for efficient total effect estimation in causal linear models

Emilija Perković, University of Washington

Joint work with Leonard Henckel, Marloes H. Maathuis
ETH Zurich

Adjustment Sets

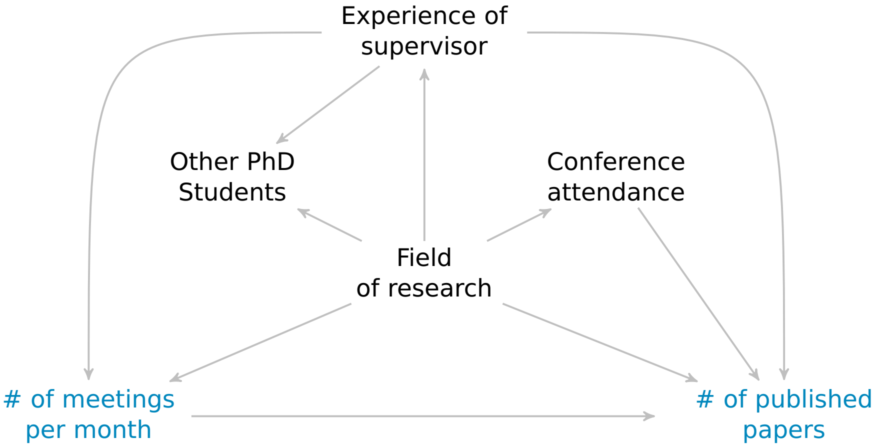


Figure: DAG $\mathcal{D}$.

Adjustment Sets

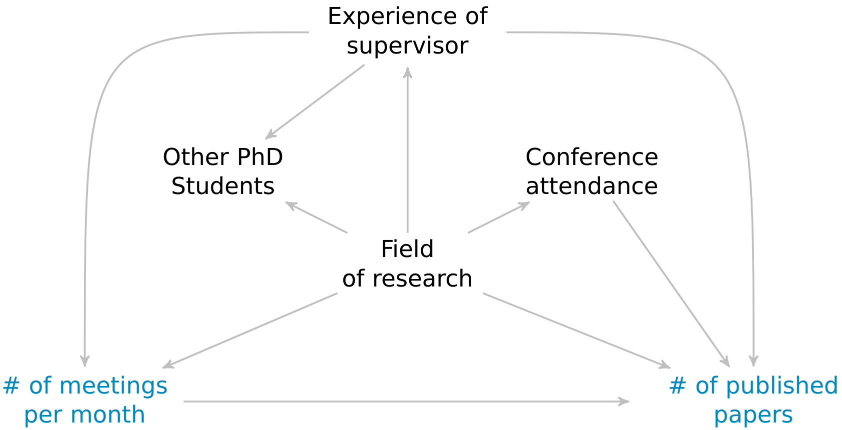


Figure: Use the back-door criterion (Pearl, 1993) or the adjustment criterion (Shpitser et. al, 2012) on DAG $\mathcal{D}$.

Adjustment Sets

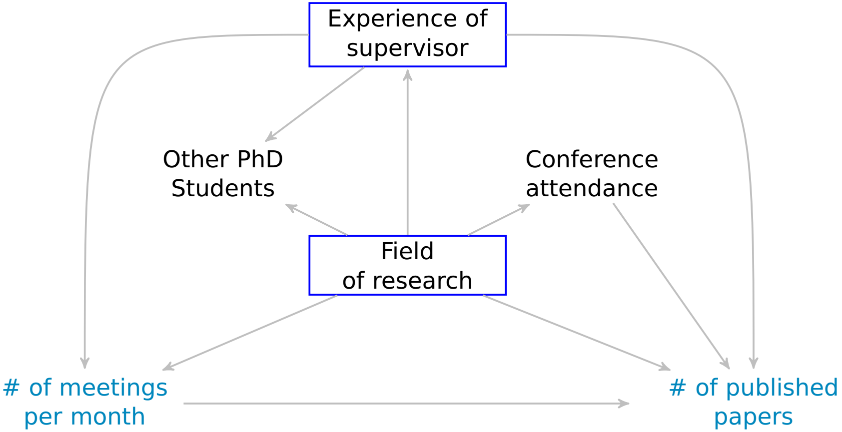


Figure: Use the back-door criterion (Pearl, 1993) or the adjustment criterion (Shpitser et. al, 2012) on DAG $\mathcal{D}$.

What is an adjustment set?

- (causal) DAG: (causal) Directed Acyclic Graph.

- A probability density f is compatible with the causal DAG $\mathcal{D}$ if: $f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j | pa(x_j, \mathcal{D}))$ and $f(\mathbf{v} | do(\mathbf{x})) = \prod_{X_j \in \mathbf{V} \setminus \mathbf{X}} f(x_j | pa(x_j, \mathcal{D}))$.

What is an adjustment set?

- (causal) DAG: (causal) Directed Acyclic Graph.

- A probability density f is compatible with the causal DAG $\mathcal{D}$ if: $f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j | pa(x_j, \mathcal{D}))$ and $f(\mathbf{v} | do(\mathbf{x})) = \prod_{X_j \in \mathbf{V} \setminus \mathbf{X}} f(x_j | pa(x_j, \mathcal{D}))$.

- $\mathbf{Z}$ is a valid adjustment set if relative to $(\mathbf{X}, \mathbf{Y})$ and any f compatible with $\mathcal{D}$:

  $f(\mathbf{y} | do(\mathbf{x})) = \int_{\mathbf{Z}} f(\mathbf{y} | \mathbf{x}, \mathbf{z}) f(\mathbf{z}) d\mathbf{z}$.

What is an adjustment set?

- (causal) DAG: (causal) Directed Acyclic Graph.

- A probability density f is compatible with the causal DAG $\mathcal{D}$ if: $f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j|pa(x_j, \mathcal{D}))$ and $f(\mathbf{v}|do(\mathbf{x})) = \prod_{X_j \in \mathbf{V} \setminus \mathbf{X}} f(x_j|pa(x_j, \mathcal{D}))$.

- $\mathbf{Z}$ is a valid adjustment set if relative to $(\mathbf{X}, \mathbf{Y})$ and any f compatible with $\mathcal{D}$:
  $f(\mathbf{y}|do(\mathbf{x})) = \int_{\mathbf{Z}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}) f(\mathbf{z}) d\mathbf{z}$.

- In a causal linear model, if $\mathbf{Z}$ a valid adjustment set then the total effect of X on Y is the coefficient $\beta_{yx.\mathbf{z}}$ of X in the regression $Y \sim X + \mathbf{Z}$.

Adjustment Sets

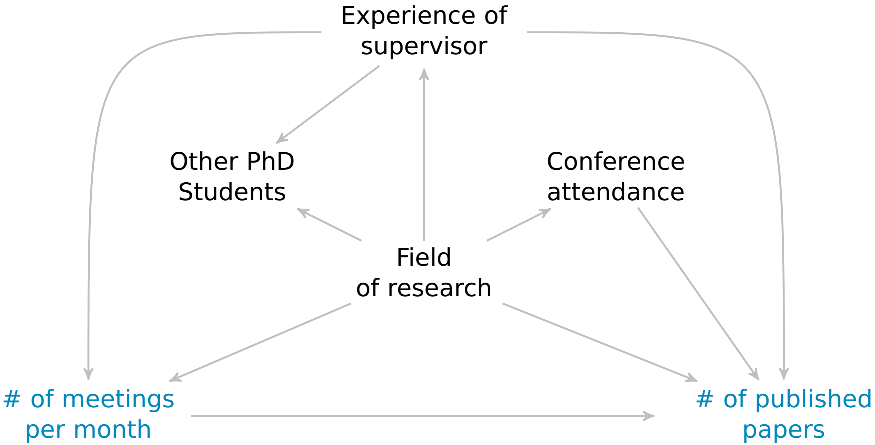


Figure: DAG $\mathcal{D}$.

Adjustment Sets

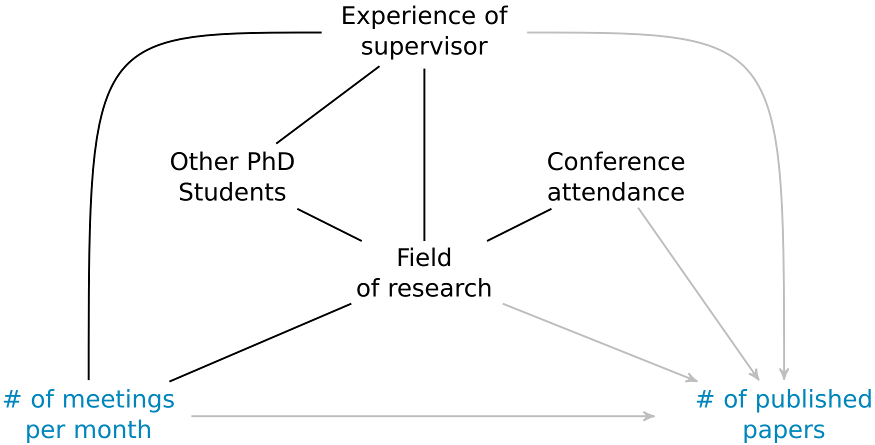


Figure: CPDAG $\mathcal{C}$ of DAG $\mathcal{D}$.

Adjustment Sets

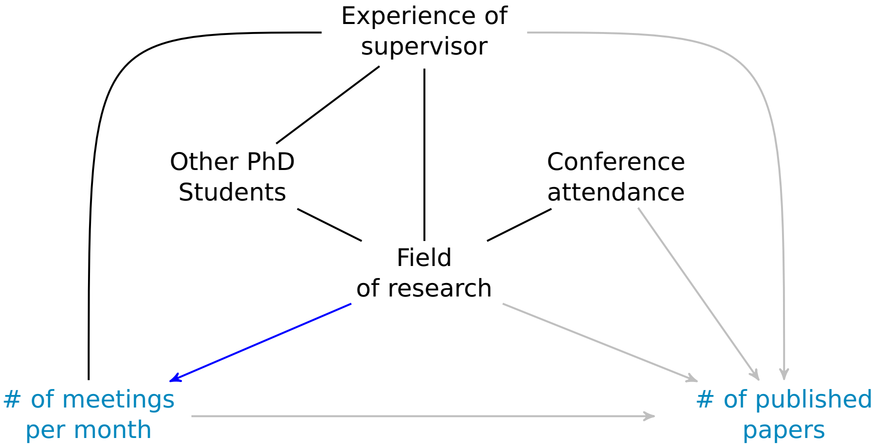


Figure: PDAG of DAG $\mathcal{D}$.

Adjustment Sets

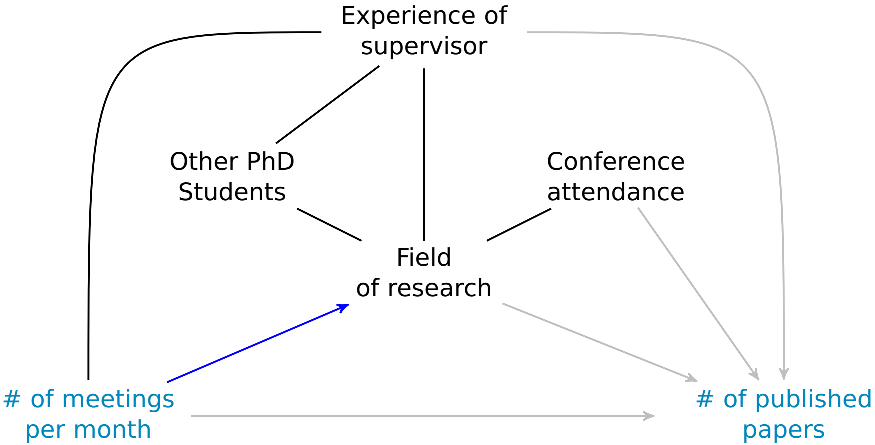


Figure: PDAG of DAG $\mathcal{D}$.

Adjustment Sets

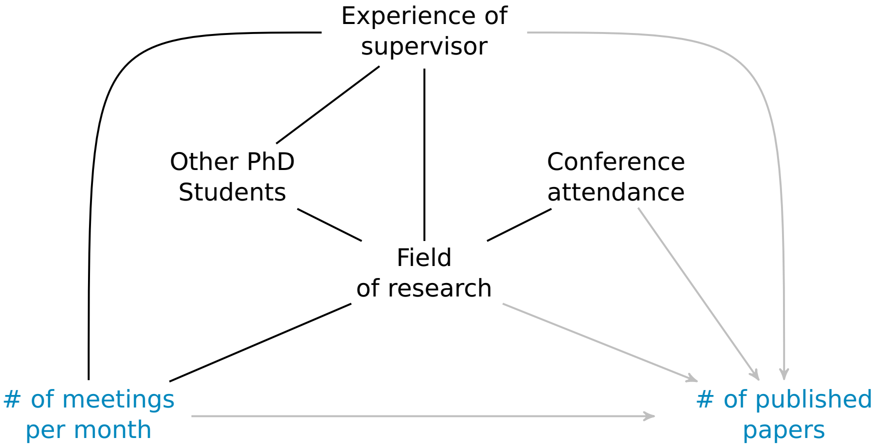
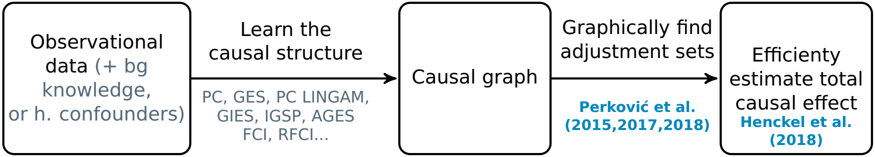


Figure: CPDAG $\mathcal{C}$ of DAG $\mathcal{D}$.

Framework



- Perković, Textor, Kalisch and Maathuis (2015). A Complete Generalized Adjustment Criterion. *UAI 2015*.
- Perković, Textor, Kalisch and Maathuis (2018). Complete Graphical Characterization and Construction of Adjustment Sets in Markov Equivalence Classes of Ancestral Graphs. *Journal of Machine Learning Research*, to appear.
- Perković, Kalisch and Maathuis (2017). Interpreting and Using CPDAGs with Background Knowledge. *UAI 2017*.
- Henckel, Perković, and Maathuis (2018). Graphical Criteria for Efficient Total Effect Estimation via Adjustment in Causal Linear Structural Equation Models. Working paper.

Generalized adjustment criterion

Theorem (Perković et al., 2015, 2017, 2018):
$\mathbf{Z}$ is a valid adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ and $\mathcal{G}$ if:

Amenability	$\mathcal{G}$ is **amenable** relative to $(\mathbf{X}, \mathbf{Y})$.
Forbidden Set	$\mathbf{Z}$ does not contain nodes in **Forbidden**$(\mathbf{X}, \mathbf{Y}, \mathcal{G})$.
Blocking	$\mathbf{Z}$ blocks all **proper non-causal definite status** paths from $\mathbf{X}$ to $\mathbf{Y}$.

In a causal linear model, if $\mathbf{Z}$ a valid adjustment set then the total effect of X on Y is the coefficient $\beta_{yx.\mathbf{z}}$ of X in the regression $Y \sim X + \mathbf{Z}$.

Generalized adjustment criterion

Theorem (Perković et al., 2015, 2017, 2018):
Z is a valid adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ and $\mathcal{G}$ if:

Amenability	$\mathcal{G}$ is **amenable** relative to $(\mathbf{X}, \mathbf{Y})$.
Forbidden Set	**Z** does not contain nodes in **Forbidden**$(\mathbf{X}, \mathbf{Y}, \mathcal{G})$.
Blocking	**Z** blocks all **proper non-causal definite status** paths from **X** to **Y**.

In a causal linear model, if **Z** a valid adjustment set then the total effect of X on Y is the coefficient $\beta_{yx.\mathbf{z}}$ of X in the regression $Y \sim X + \mathbf{Z}$.

- We have algorithms to list all valid adjustment sets (see `adjustment()` in R package pcalg.)

Generalized adjustment criterion

Theorem (Perković et al., 2015, 2017, 2018):
$\mathbf{Z}$ is a valid adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ and $\mathcal{G}$ if:

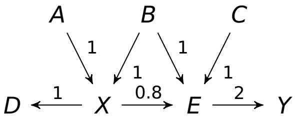
Amenability	$\mathcal{G}$ is **amenable** relative to $(\mathbf{X}, \mathbf{Y})$.
Forbidden Set	$\mathbf{Z}$ does not contain nodes in **Forbidden**$(\mathbf{X}, \mathbf{Y}, \mathcal{G})$.
Blocking	$\mathbf{Z}$ blocks all **proper non-causal definite status** paths from $\mathbf{X}$ to $\mathbf{Y}$.

In a causal linear model, if $\mathbf{Z}$ a valid adjustment set then the total effect of X on Y is the coefficient $\beta_{yx.\mathbf{z}}$ of X in the regression $Y \sim X + \mathbf{Z}$.

- We have algorithms to list all valid adjustment sets (see `adjustment()` in R package pcalg.)
- All of them will provide consistent estimators of the total effect, but which one will be asymptotically most efficient?

Example: efficient estimates

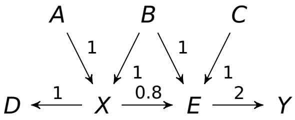
Directed acyclic graph (DAG) with weighted edges:



$\tau_{yx} = 0.8 \cdot 2 = 1.6$

Example: efficient estimates

Directed acyclic graph (DAG) with weighted edges:



$$\tau_{yx} = 0.8 \cdot 2 = 1.6$$

AS	mean	var
$\emptyset$	2.27	3.57
A	2.60	4.92
B	1.60	4.53
C	2.27	2.21
D	2.27	2.89
E	0.00	0.82
A+B	1.60	8.96
B+C	1.60	2.52
B+D	1.60	3.53
B+E	0.00	0.83
A+B+C	1.60	5.04

Z VAS:

- $B \in \mathbf{Z}$
- $E \notin \mathbf{Z}$
- A, C, D may be in **Z**

So total of 8 VAS here!

Variance varies significantly:

- $pa(X, \mathcal{G}) = \{A, B\}$ bad
- minimal set $\{B\}$ average
- $\{B, C\}$ best

Asymptotic variance formula

- $(X, Y, \mathbf{Z})$ joint normal, then $\sqrt{n}(\hat{\beta}_{yx.\mathbf{z}} - \beta_{yx.\mathbf{z}}) \xrightarrow{D} \mathcal{N}(0, \frac{\sigma_{yy.x\mathbf{z}}}{\sigma_{xx.\mathbf{z}}})$.

Asymptotic variance formula

- $(X, Y, \mathbf{Z})$ joint normal, then $\sqrt{n}(\hat{\beta}_{yx.\mathbf{z}} - \beta_{yx.\mathbf{z}}) \xrightarrow{D} \mathcal{N}(0, \frac{\sigma_{yy.x\mathbf{z}}}{\sigma_{xx.\mathbf{z}}})$.
- If $\mathbf{Z}$ a VAS wrt (X, Y), then $\sqrt{n}(\hat{\beta}_{yx.\mathbf{z}} - \tau_{yx}) \xrightarrow{D} \mathcal{N}(0, \frac{\sigma_{yy.x\mathbf{z}}}{\sigma_{xx.\mathbf{z}}})$,

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

Remark: This is not trivial in the non-Gaussian case.

Intuition

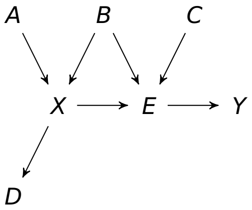
Goal: Minimize $a.var(\hat{\tau}_{yx}^{\mathbf{z}}) = \frac{\sigma_{yy.x\mathbf{z}}}{\sigma_{xx.\mathbf{z}}}$:

- minimize $\sigma_{yy.x\mathbf{z}} = Var(Y - \beta_{yx.\mathbf{z}}X - \beta_{y\mathbf{z}.x}^{T}\mathbf{Z})$
- maximize $\sigma_{xx.\mathbf{z}} = Var(X - \beta_{x\mathbf{z}}^{T}\mathbf{Z})$

Intuition

Goal: Minimize $a.var(\hat{\tau}_{yx}^{\mathbf{z}}) = \frac{\sigma_{yy.x\mathbf{z}}}{\sigma_{xx.\mathbf{z}}}$:

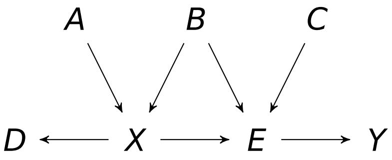
- minimize $\sigma_{yy.x\mathbf{z}} = Var(Y - \beta_{yx.\mathbf{z}} X - \beta_{y\mathbf{z}.x}^{T} \mathbf{Z})$
- maximize $\sigma_{xx.\mathbf{z}} = Var(X - \beta_{x\mathbf{z}}^{T} \mathbf{Z})$



Z a VAS:

- $B \in \mathbf{Z}$, $E \notin \mathbf{Z}$
- $A \perp_{\mathcal{G}} Y | (\mathbf{Z} \setminus A) \cup X$
- $D \perp_{\mathcal{G}} Y | (\mathbf{Z} \setminus D) \cup X$
- $C \perp_{\mathcal{G}} X | \mathbf{Z} \setminus C$

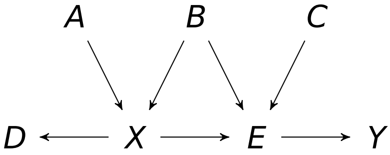
Variance with random edge coefficients



AS	Case 1	Case 2	Case 3
$\{A, B\}$	5.38	5.47	0.85
$\{A, B, C\}$	1.44	4.44	0.51
$\{B\}$	3.49	4.40	0.54
$\{B, C\}$	0.94	3.58	0.32
$\{A, B, D\}$	7.20	7.39	12.65
$\{A, B, C, D\}$	1.93	6.01	7.59
$\{B, D\}$	5.31	6.33	12.34
$\{B, C, D\}$	1.42	5.15	7.41

- A, D increase variance
- C decreases variance
- $\{A, B, D\}$ is worst set
- $\{B, C\}$ is best set
- not all comparisons are consistent

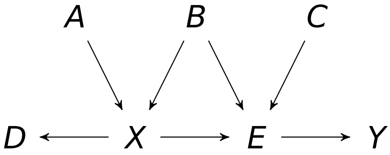
Variance with random edge coefficients



AS	Case 1	Case 2	Case 3
$\{A, B\}$	5.38	5.47	0.85
$\{A, B, C\}$	1.44	4.44	0.51
$\{B\}$	3.49	4.40	0.54
$\{B, C\}$	0.94	3.58	0.32
$\{A, B, D\}$	7.20	7.39	12.65
$\{A, B, C, D\}$	1.93	6.01	7.59
$\{B, D\}$	5.31	6.33	12.34
$\{B, C, D\}$	1.42	5.15	7.41

- A, D increase variance
- C decreases variance
- $\{A, B, D\}$ is worst set
- $\{B, C\}$ is best set
- not all comparisons are consistent

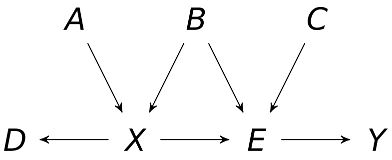
Variance with random edge coefficients



AS	Case 1	Case 2	Case 3
$\{A, B\}$	5.38	5.47	0.85
$\{A, B, C\}$	1.44	4.44	0.51
$\{B\}$	3.49	4.40	0.54
$\{B, C\}$	0.94	3.58	0.32
$\{A, B, D\}$	7.20	7.39	12.65
$\{A, B, C, D\}$	1.93	6.01	7.59
$\{B, D\}$	5.31	6.33	12.34
$\{B, C, D\}$	1.42	5.15	7.41

- A, D increase variance
- C decreases variance
- $\{A, B, D\}$ is worst set
- $\{B, C\}$ is best set
- not all comparisons are consistent

Variance with random edge coefficients



AS	Case 1	Case 2	Case 3
$\{A,B\}$	5.38	5.47	0.85
$\{A,B,C\}$	1.44	4.44	0.51
$\{B\}$	3.49	4.40	0.54
$\{B,C\}$	0.94	3.58	0.32
$\{A,B,D\}$	7.20	7.39	12.65
$\{A,B,C,D\}$	1.93	6.01	7.59
$\{B,D\}$	5.31	6.33	12.34
$\{B,C,D\}$	1.42	5.15	7.41

- A, D increase variance
- C decreases variance
- $\{A,B,D\}$ is worst set
- $\{B,C\}$ is best set
- not all comparisons are consistent

Main results

- Graphical criterion for qualitative asymptotic variance comparison

Main results

- Graphical criterion for qualitative asymptotic variance comparison

- Variance reducing pruning procedure

Main results

- Graphical criterion for qualitative asymptotic variance comparison
- Variance reducing pruning procedure
- Asymptotically optimal valid adjustment set (does not hold in the hidden variable setting)

Main results

- Graphical criterion for qualitative asymptotic variance comparison
- Variance reducing pruning procedure
- Asymptotically optimal valid adjustment set (does not hold in the hidden variable setting)

Remark: The results are in presented in the simplified form for singleton X and Y and DAGs, but also hold for joint interventions and more general graphs (CPDAGs, maximally oriented PDAGs, MAGs PAGs).

Variance comparison criterion

Asymptotic variance comparison criterion:
$\mathbf{Z_1}$ and $\mathbf{Z_2}$ VAS wrt (X, Y) in a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, such that

- $\mathbf{Z_1} \setminus \mathbf{Z_2} \perp_{\mathcal{G}} Y | \mathbf{Z_2} \cup X$
- $\mathbf{Z_2} \setminus \mathbf{Z_1} \perp_{\mathcal{G}} X | \mathbf{Z_1}$,

then $a.var(\hat{\tau}_{yx}^{\mathbf{Z_2}}) \leq a.var(\hat{\tau}_{yx}^{\mathbf{Z_1}})$.

- $\perp_{\mathcal{G}}$ indicates d-separation

Remark: This is an extension to non-disjoint sets (Kuroki and Cai, 2004) of size larger than 2 (Kuroki and Miyakawa, 2003) and to arbitrary error types.

Pruning procedure

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

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

i) order independent
ii) no other VAS $\mathbf{Z}'' \subseteq \mathbf{Z}$ is assured a better asymptotic variance

The optimal VAS

Definition: $\mathbf{O}(X, Y, \mathcal{G}) = pa(cn(X, Y, \mathcal{G}), \mathcal{G}) \setminus forb(X, Y, \mathcal{G})$

The optimal VAS

Definition: $\mathbf{O}(X, Y, \mathcal{G}) = pa(cn(X, Y, \mathcal{G}), \mathcal{G}) \setminus forb(X, Y, \mathcal{G})$

X, Y two nodes in causal DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, such that $Y \in de(X, \mathcal{G})$. Then

(Validity) If a VAS exists, **O** is one.

(Optimality) For any VAS **Z**

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

The optimal VAS

Definition: $\mathbf{O}(X, Y, \mathcal{G}) = pa(cn(X, Y, \mathcal{G}), \mathcal{G}) \setminus forb(X, Y, \mathcal{G})$

X, Y two nodes in causal DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, such that $Y \in de(X, \mathcal{G})$. Then

(Validity) If a VAS exists, **O** is one.

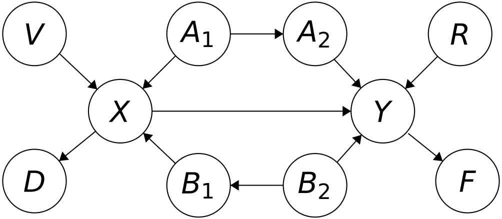
(Optimality) For any VAS **Z**

$$a.var(\hat{\tau}_{yx}^{\mathbf{O}}) \leq a.var(\hat{\tau}_{yx}^{\mathbf{Z}}).$$

(Minimality) If $a.var(\hat{\tau}_{yx}^{\mathbf{O}}) = a.var(\hat{\tau}_{yx}^{\mathbf{Z}})$ and we assume faithfulness, then $\mathbf{O} \subseteq \mathbf{Z}$.

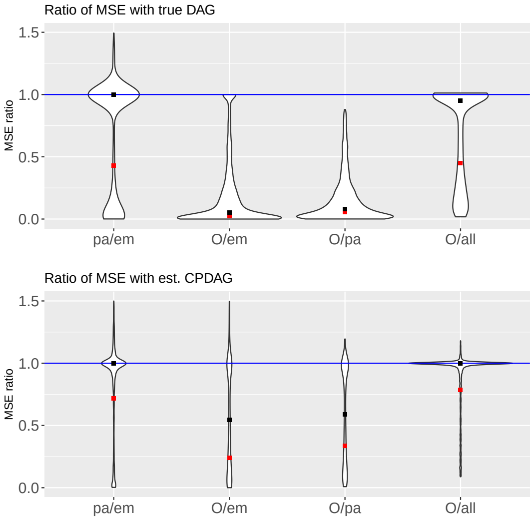
Remark: If $Y \notin de(X, \mathcal{G})$, then $\tau_{yx} = 0$.

Example: the optimal VAS



- $cn(X, Y, \mathcal{G}) = \{Y\}$
- $forb(X, Y, \mathcal{G}) = \{X, Y, F\}$
- $pa(cn(X, Y, \mathcal{G}), \mathcal{G}) = \{X, A_2, B_2, R\}$
- $\mathbf{O}(X, Y, \mathcal{G}) = \{A_2, B_2, R\}$

Quantifying the practical efficiency gain

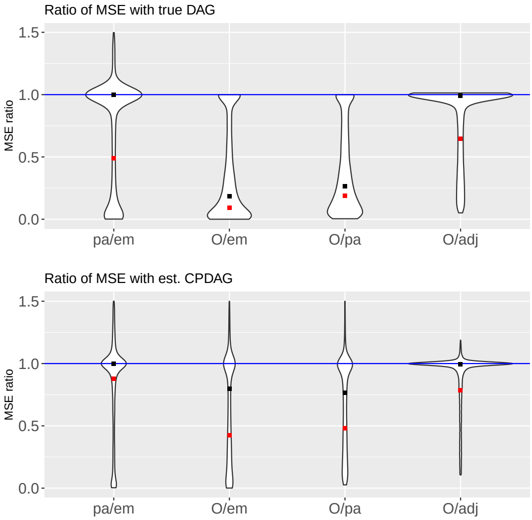


5000 random settings:
100 data sets sampled,
empirical MSE computed

X - randomly chosen,
Y - descendant of X

pa : $pa(X, \mathcal{G})$,
em : $\emptyset$,
O : $\mathbf{O}(X, Y, \mathcal{G})$,
adj : $adjust(X, Y, \mathcal{G})$

Joint interventions



5000 random settings:
100 data sets sampled,
empirical MSE computed

X - randomly chosen,
Y - descendant of
each $X_i \in \mathbf{X}$

pa : $pa(\mathbf{X}, \mathcal{G})$,
em : $\emptyset$,
O : $\mathbf{O}(\mathbf{X}, Y, \mathcal{G})$,
adj : $adjust(\mathbf{X}, Y, \mathcal{G})$

Summary

- Graphical criterion for qualitative asymptotic variance comparisons

- Variance decreasing pruning procedure

- Asymptotically optimal VAS

Summary

- Graphical criterion for qualitative asymptotic variance comparisons
- Variance decreasing pruning procedure
- Asymptotically optimal VAS

Thanks!