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

Experience of supervisor

Other PhD Students

Conference attendance

Field of research

# of meetings per month

# of published papers

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

Emilija Perković, University of Washington
What is an adjustment set?

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

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

  
  $Z$ is a valid adjustment set if relative to $(\mathcal{X}, Y)$ and any $f$ compatible with $\mathcal{D}$:
  
  $$f(y|\text{do}(x)) = \int_Z f(y|x, z) f(z) dz.$$
What is an adjustment set?

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

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

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

• **(causal) DAG**: (causal) Directed Acyclic Graph.

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

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

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

Emilija Perković, University of Washington

Valid adjustment set selection 8 / 26
Adjustment Sets

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

Emilija Perković, University of Washington

Valid adjustment set selection 9 / 26
Observational data (+ bg knowledge, or h. confounders) → Learn the causal structure (PC, GES, PC LINGAM, GIES, IGSP, AGES FCI, RFCI...) → Causal graph → Graphically find adjustment sets → Efficiency estimate total causal effect

**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 \textbf{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 \( \mathbf{X} \) on \( \mathbf{Y} \) is the coefficient \( \beta_{yx.z} \) of \( \mathbf{X} \) in the regression \( \mathbf{Y} \sim \mathbf{X} + \mathbf{Z} \).
Theorem (Perković et al., 2015, 2017, 2018): \( Z \) is a valid adjustment set relative to \((X, Y)\) and \( G \) if:

- **Amenability** \( G \) is **amenable** relative to \((X, Y)\).
- **Forbidden Set** \( Z \) does not contain nodes in **Forbidden** \((X, Y, G)\).
- **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 \( \beta_{yx.z} \) of \( X \) in the regression \( Y \sim X + Z \).

- We have algorithms to list all valid adjustment sets (see \texttt{adjustment()} in R package pcalg.)
Theorem (Perković et al., 2015, 2017, 2018):
\( Z \) is a valid adjustment set relative to \((X, Y)\) and \( \mathcal{G} \) if:

**Amenability** \( \mathcal{G} \) is \textit{amenable} relative to \((X, Y)\).

**Forbidden Set** \( Z \) does not contain nodes in \textit{Forbidden}(\( X, Y, \mathcal{G} \)).

**Blocking** \( Z \) blocks all \textit{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 \( \beta_{yx,z} \) of \( X \) in the regression \( Y \sim X + Z \).

- We have algorithms to list all valid adjustment sets (see \texttt{adjustment()} in R package \texttt{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:

\[ \tau_{yx} = 0.8 \cdot 2 = 1.6 \]

**Z VAS:**
- \( B \in Z \)
- \( E \notin Z \)
- \( A, C, D \) 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

<table>
<thead>
<tr>
<th>AS</th>
<th>mean</th>
<th>var</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \emptyset )</td>
<td>2.27</td>
<td>3.57</td>
</tr>
<tr>
<td>A</td>
<td>2.60</td>
<td>4.92</td>
</tr>
<tr>
<td>B</td>
<td>1.60</td>
<td>4.53</td>
</tr>
<tr>
<td>C</td>
<td>2.27</td>
<td>2.21</td>
</tr>
<tr>
<td>D</td>
<td>2.27</td>
<td>2.89</td>
</tr>
<tr>
<td>E</td>
<td>0.00</td>
<td>0.82</td>
</tr>
<tr>
<td>A+B</td>
<td>1.60</td>
<td>8.96</td>
</tr>
<tr>
<td>B+C</td>
<td>1.60</td>
<td>2.52</td>
</tr>
<tr>
<td>B+D</td>
<td>1.60</td>
<td>3.53</td>
</tr>
<tr>
<td>B+E</td>
<td>0.00</td>
<td>0.83</td>
</tr>
<tr>
<td>A+B+C</td>
<td>1.60</td>
<td>5.04</td>
</tr>
</tbody>
</table>
Asymptotic variance formula

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

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

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

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

**Remark:** This is not trivial in the non-Gaussian case.
**Goal:** Minimize \( a \cdot \text{var}(\hat{\tau}_{yx}^z) = \frac{\sigma_{yy.xz}}{\sigma_{xx.z}} \):

- minimize \( \sigma_{yy.xz} = \text{Var}(Y - \beta_{yx.z}X - \beta_{yz.x}^T Z) \)
- maximize \( \sigma_{xx.z} = \text{Var}(X - \beta_{xz}^T Z) \)
**Goal:** Minimize $a \cdot \text{var}(\hat{t}_{yx}^z) = \frac{\sigma_{yy \cdot xz}}{\sigma_{xx \cdot z}}$:

- minimize $\sigma_{yy \cdot xz} = \text{Var}(Y - \beta_{yx \cdot z}X - \beta_{yz \cdot x}^T Z)$
- maximize $\sigma_{xx \cdot z} = \text{Var}(X - \beta_{xz}^T Z)$

**Z as VAS:**
- $B \in Z$, $E \notin Z$
- $A \perp_g Y | (Z \setminus A) \cup X$
- $D \perp_g Y | (Z \setminus D) \cup X$
- $C \perp_g X | Z \setminus C$
### Variance with random edge coefficients

<table>
<thead>
<tr>
<th>AS</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A, B}</td>
<td>5.38</td>
<td>5.47</td>
<td>0.85</td>
<td>(A, D) increase variance</td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>1.44</td>
<td>4.44</td>
<td>0.51</td>
<td>(C) decreases variance</td>
</tr>
<tr>
<td>{B}</td>
<td>3.49</td>
<td>4.40</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>0.94</td>
<td>3.58</td>
<td>0.32</td>
<td>{A, B, D} is worst set</td>
</tr>
<tr>
<td>{A, B, D}</td>
<td>7.20</td>
<td>7.39</td>
<td>12.65</td>
<td>{B, C} is best set</td>
</tr>
<tr>
<td>{A, B, C, D}</td>
<td>1.93</td>
<td>6.01</td>
<td>7.59</td>
<td></td>
</tr>
<tr>
<td>{B, D}</td>
<td>5.31</td>
<td>6.33</td>
<td>12.34</td>
<td></td>
</tr>
<tr>
<td>{B, C, D}</td>
<td>1.42</td>
<td>5.15</td>
<td>7.41</td>
<td>not all comparisons are consistent</td>
</tr>
</tbody>
</table>

![Diagram](image.png)
### Variance with random edge coefficients

#### Graph Representation

![Graph](attachment:graph.png)

<table>
<thead>
<tr>
<th>AS</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A, B}</td>
<td>5.38</td>
<td>5.47</td>
<td>0.85</td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>1.44</td>
<td>4.44</td>
<td>0.51</td>
</tr>
<tr>
<td>{B}</td>
<td>3.49</td>
<td>4.40</td>
<td>0.54</td>
</tr>
<tr>
<td>{B, C}</td>
<td>0.94</td>
<td>3.58</td>
<td>0.32</td>
</tr>
<tr>
<td>{A, B, D}</td>
<td>7.20</td>
<td>7.39</td>
<td>12.65</td>
</tr>
<tr>
<td>{A, B, C, D}</td>
<td>1.93</td>
<td>6.01</td>
<td>7.59</td>
</tr>
<tr>
<td>{B, D}</td>
<td>5.31</td>
<td>6.33</td>
<td>12.34</td>
</tr>
</tbody>
</table>

- \{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

**Diagram:**

![Graph with nodes A, B, C, X, D, E, Y and edges between them.]

<table>
<thead>
<tr>
<th>AS</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A, B}</td>
<td>5.38</td>
<td>5.47</td>
<td>0.85</td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>1.44</td>
<td>4.44</td>
<td>0.51</td>
</tr>
<tr>
<td>{B}</td>
<td>3.49</td>
<td>4.40</td>
<td>0.54</td>
</tr>
<tr>
<td>{B, C}</td>
<td>0.94</td>
<td>3.58</td>
<td>0.32</td>
</tr>
<tr>
<td>{A, B, D}</td>
<td>7.20</td>
<td>7.39</td>
<td>12.65</td>
</tr>
<tr>
<td>{A, B, C, D}</td>
<td>1.93</td>
<td>6.01</td>
<td>7.59</td>
</tr>
<tr>
<td>{B, D}</td>
<td>5.31</td>
<td>6.33</td>
<td>12.34</td>
</tr>
<tr>
<td>{B, C, D}</td>
<td>1.42</td>
<td>5.15</td>
<td>7.41</td>
</tr>
</tbody>
</table>

- 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

\[
\begin{align*}
A & \quad B & \quad C \\
D & \quad X & \quad E & \quad Y
\end{align*}
\]

<table>
<thead>
<tr>
<th>AS</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A, B}</td>
<td>5.38</td>
<td>5.47</td>
<td>0.85</td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>1.44</td>
<td>4.44</td>
<td>0.51</td>
</tr>
<tr>
<td>{B}</td>
<td>3.49</td>
<td>4.40</td>
<td>0.54</td>
</tr>
<tr>
<td>{B, C}</td>
<td>0.94</td>
<td>3.58</td>
<td>0.32</td>
</tr>
<tr>
<td>{A, B, D}</td>
<td>7.20</td>
<td>7.39</td>
<td>12.65</td>
</tr>
<tr>
<td>{A, B, C, D}</td>
<td>1.93</td>
<td>6.01</td>
<td>7.59</td>
</tr>
<tr>
<td>{B, D}</td>
<td>5.31</td>
<td>6.33</td>
<td>12.34</td>
</tr>
<tr>
<td>{B, C, D}</td>
<td>1.42</td>
<td>5.15</td>
<td>7.41</td>
</tr>
</tbody>
</table>

- \(A, D\) 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

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).
Main results

- Graphical criterion for qualitative asymptotic variance comparison
- Variance reducing pruning procedure

Remark: The results are 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).
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 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).
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).
Asymptotic variance comparison criterion: \( Z_1 \) and \( Z_2 \) VAS wrt \((X, Y)\) in a DAG \( G = (V, E) \), such that

- \( Z_1 \setminus Z_2 \perp_G Y|Z_2 \cup X \)
- \( Z_2 \setminus Z_1 \perp_G X|Z_1 \)

then \( a.var(\hat{\tau}_{Z_2}^{z_2}) \leq a.var(\hat{\tau}_{Z_1}^{z_1}) \).

- \( \perp_G \) 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 $G$, disjoint node sets $X$ and $Y$ and a VAS $Z$

**Output**: VAS $Z' \subseteq Z$, such that $a.\text{var}(\hat{\tau}_{yx}') \leq a.\text{var}(\hat{\tau}_{yx})$

```markdown
begin
1. $Z' = Z$;
2. foreach $Z \in Z'$ do
3.   if $Y \perp_{G} Z|Z'_{-z} \cup X$ and $Z'_{-z}$ is a VAS then
4.     $Z' = Z'_{-z}$;
5. return $Z'$;
```

i) order independent
ii) no other VAS $Z'' \subseteq Z$ is assured a better asymptotic variance
The optimal VAS

**Definition:** $O(X, Y, G) = pa(cn(X, Y, G), G) \setminus forb(X, Y, G)$
The optimal VAS

**Definition:** \( O(X, Y, G) = pa(cn(X, Y, G), G) \setminus forb(X, Y, G) \)

\( X, Y \) two nodes in causal DAG \( G = (V, E) \), such that \( Y \in de(X, G) \).

Then

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

**(Optimality)** For any VAS \( Z \)

\[
a \cdot var(\hat{\tau}_{yx}^O) \leq a \cdot var(\hat{\tau}_{yx}^Z).
\]
The optimal VAS

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

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

*(Validity)* If a VAS exists, \( \mathbf{O} \) is one.

*(Optimality)* For any VAS \( \mathbf{Z} \)

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

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

**Remark:** If \( Y \notin de(X, G) \), then \( \tau_{yx} = 0 \).
Example: the optimal VAS

- $cn(X, Y, G) = \{Y\}$
- $forb(X, Y, G) = \{X, Y, F\}$
- $pa(cn(X, Y, G), G) = \{X, A_2, B_2, R\}$
- $O(X, Y, G) = \{A_2, B_2, R\}$
Quantifying the practical efficiency gain

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

$X$ - randomly chosen, $Y$ - descendant of $X$

$\text{pa} : \text{pa}(X, G)$,
$\text{em} : \emptyset$,
$\text{O} : \text{O}(X, Y, G)$,
$\text{adj} : \text{adjust}(X, Y, G)$
Joint interventions

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

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

\[ \text{pa} : \text{pa}(\mathbf{X}, G), \]
\[ \text{em} : \emptyset, \]
\[ \text{O} : \mathbf{O}(\mathbf{X}, Y, G), \]
\[ \text{adj} : \text{adjust}(\mathbf{X}, Y, G) \]
• 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!