A More Robust Optimal Adjustment Set

Chase Mathis

December 2023

Goals and New Research

- Similar to how a trek is associated with covariance, can one find a graphical structure or function of one that defines the bias term.
- In sparse data situations, should researchers include forbidden terms if sample covariance meet certain constraints.

Previous Causal DAGs

- The optimal adjustment set is the unbiased adjustment set that has the *lowest asymptotic variance*.
- Henckel et al. [**?**] that for Structural Equation Models (SEMs), an optimal adjustment set for finding the causal effect of $T \to Y$ is:

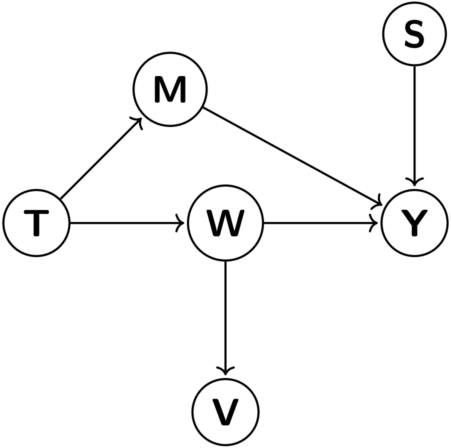
$$O_g(T \to Y) = pa_g(\mathrm{cn}_g(T \to Y)) \setminus (\mathrm{cn}_g(T \to Y) \cup \{T\}).$$

- It was also shown for the non-parametric case [**?**]

Example

So for instance in the graph below the

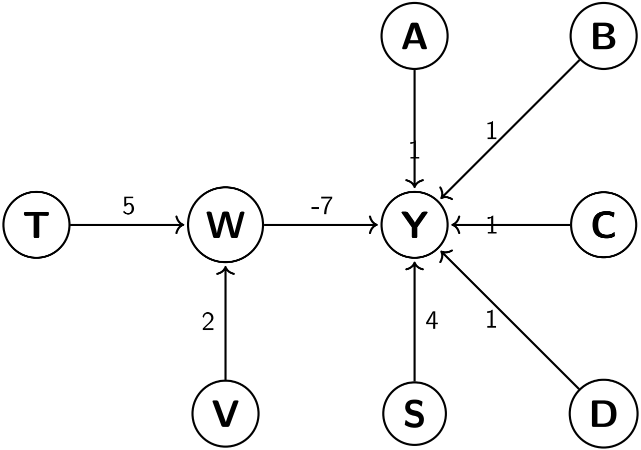
- $cn_G(T \to Y) = \{W, M, Y\}$
- $pa_g(cn_g(T \to Y)) = \{T, W, M, S\}$
- $\implies O_g(T \to Y) = \{S\}$



Graph Example

Consider the Structural Equation Model

- In the graph below, the $O_g(T \to Y) = \{A, B, C, D, V, S\}$



In R

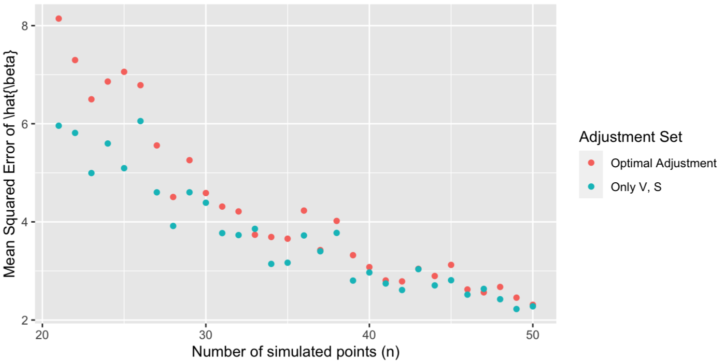
In R, this can be seen as

```
T <- rnorm(n)
V <- rnorm(n)
S <- rnorm(n)
A <- rnorm(n)
B <- rnorm(n)
C <- rnorm(n)
D <- rnorm(n)
E <- rnorm(n)

W <- 5 * T + 2 * V  +  + rnorm(n)
Y <- -7 * W  + A + B + 4 *S + C + D + rnorm(n)
```

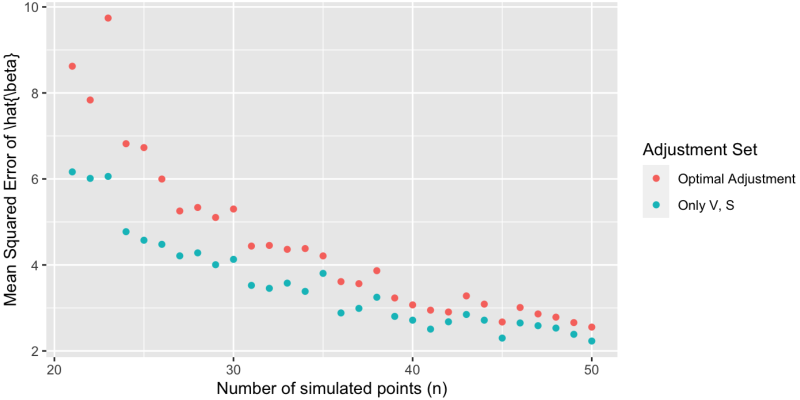
Results

I calculate the MSE of the $\hat{\beta_T} \mid \beta = -35$ over the optimal adjustment set and instead only using the strongly causal nodes V, S, I see the following:



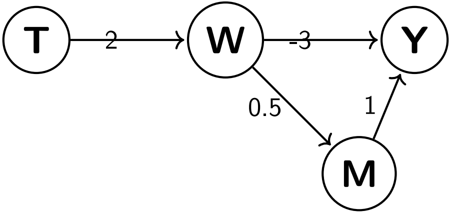
Results 2: Changing Parameters

If I make the variance of A, B, C, D, E small or V, S large, the difference is more pronounced



Forbidden Nodes

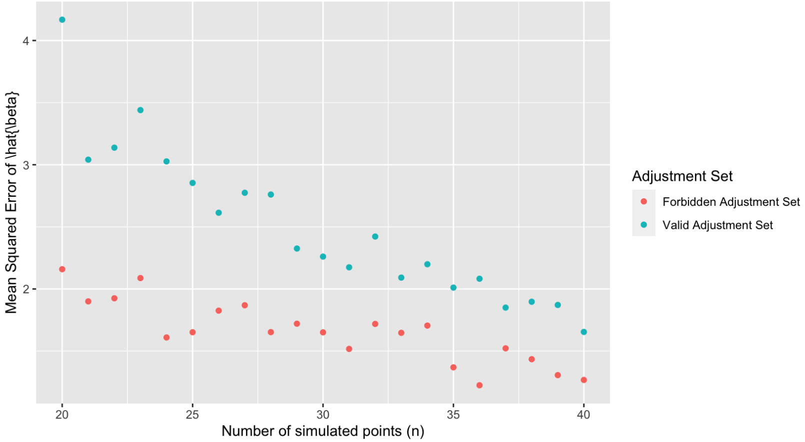
I've also been considering Forbidden Nodes and including them in adjustment sets.



Here M is a forbidden node and including it will give a biased estimate for $T \rightarrow Y$

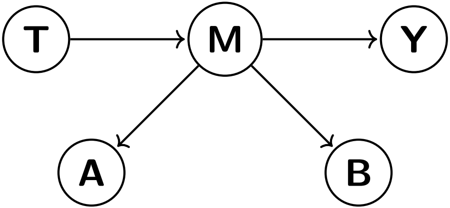
Variable Mediator

If data is sparse and M is quite a bit more variable than the other variables (5x), I found that the MSE of $\hat{\beta_T}$ is smaller when adjusting for M then for not adjusting for M



Adjusting for Two Forbidden Nodes

In the graph shown below, nodes A, B are forbidden as they are descendants of a causal node.



If we adjust for them, however, the bias is a ratio of determinants.

Proof

$$\beta_{ty\cdot ab} = \frac{\sigma_{ty\cdot ab}}{\sigma_{tt\cdot ab}} \tag{1}$$

$$= \frac{\sigma_{ty} - \Sigma_{t,AB}\Sigma_{AB}^{-1}\Sigma_{AB,y}}{|\Sigma_{AB\cdot t}|\sigma_{tt}/|\Sigma_{AB}|} \tag{2}$$

Focusing on the numerator, $\sigma_{ty} = \beta_{ty} * \sigma_{tt} = \beta_{tm}\beta_{my} * \sigma_{tt}$

$$= \frac{\beta_{tm}\beta_{my}\sigma_{tt} - \Sigma_{t,AB}\Sigma_{AB}^{-1}\Sigma_{AB,y}}{|\Sigma_{AB\cdot t}|\sigma_{tt}/|\Sigma_{AB}|} \tag{3}$$

Proof Cont.

Focusing on $\Sigma_{t,AB}\Sigma_{AB}^{-1}\Sigma_{AB,y}$, we get the following:

$$\Sigma_{t,AB}\Sigma_{AB}^{-1}\Sigma_{AB,y} = tr(\Sigma_{t,AB}\Sigma_{AB}^{-1}\Sigma_{AB,y}) \tag{4}$$

$$= tr(\Sigma_{t,AB}^{T}\Sigma_{AB,y}\Sigma_{AB}^{-1}) \tag{5}$$

$$= tr(\begin{bmatrix} \sigma_{ta} & \sigma_{tb} \end{bmatrix} \begin{bmatrix} \sigma_{ay} \\ \sigma_{by} \end{bmatrix} \Sigma_{AB}^{-1}) \tag{6}$$

We can use the trek rule to find that $\sigma_{ay} = \sigma_{mm}\beta_{ma}\beta_{my}$ and $\sigma_{by} = \sigma_{mm}\beta_{mb}\beta_{my}$

$$= \sigma_{mm}\beta_{my} tr(\begin{bmatrix} \sigma_{ta} & \sigma_{tb} \end{bmatrix} \begin{bmatrix} \beta_{ma} \\ \beta_{mb} \end{bmatrix} \Sigma_{AB}^{-1}) \tag{7}$$

$\sigma_{ta} = \beta_{ta}\sigma_{tt}, \sigma_{tb} = \beta_{tb}\sigma_{tt}$

Proof Cont.

$$= \sigma_{mm}\sigma_{tt}\beta_{my} tr(\begin{bmatrix} \beta_{ta} & \beta_{tb} \end{bmatrix} \begin{bmatrix} \beta_{ma} \\ \beta_{mb} \end{bmatrix} \Sigma_{AB}^{-1}) \tag{8}$$

$$\tag{9}$$

Further, $\beta_{ta} = \beta_{tm}\beta_{ma}, \beta_{tb} = \beta_{tm}\beta_{mb}$

$$= \sigma_{mm}\sigma_{tt}\beta_{my}\beta_{tm} tr(\begin{bmatrix} \beta_{ma} & \beta_{mb} \end{bmatrix} \begin{bmatrix} \beta_{ma} \\ \beta_{mb} \end{bmatrix} \Sigma_{AB}^{-1}) \tag{10}$$

$$= \sigma_{mm}\sigma_{tt}\beta_{my}\beta_{tm}(\beta_{ma}^2 + \beta_{mb}^2) tr(\Sigma_{AB}^{-1}) \tag{11}$$

$$= \sigma_{tt}\beta_{my}\beta_{tm}(\sigma_{ma}^2 + \sigma_{mb}^2) tr(\Sigma_{AB}^{-1})/\sigma_{mm} \tag{12}$$

$$= \sigma_{tt}\beta_{my}\beta_{tm}(\sigma_{ma}^2 + \sigma_{mb}^2)(\frac{1}{\sigma_{aa\cdot b}} + \frac{1}{\sigma_{bb\cdot a}})/\sigma_{mm} \tag{13}$$

$$\tag{14}$$

Proof Cont.

Returning to the original equation

$$= \frac{\beta_{tm}\beta_{my}\sigma_{tt} - \sigma_{tt}\beta_{my}\beta_{tm}(\sigma_{ma}^2 + \sigma_{mb}^2)(\frac{1}{\sigma_{aa\cdot b}} + \frac{1}{\sigma_{bb\cdot a}})/\sigma_{mm}}{|\Sigma_{AB\cdot t}|\sigma_{tt}/|\Sigma_{AB}|} \quad (15)$$

$$= \frac{\beta_{tm}\beta_{my}|\Sigma_{AB}|\left(1 - (\sigma_{ma}^2 + \sigma_{mb}^2)(\frac{1}{\sigma_{aa\cdot b}} + \frac{1}{\sigma_{bb\cdot a}})/\sigma_{mm}\right)}{|\Sigma_{AB\cdot t}|} \quad (16)$$

$$= \frac{\beta_{tm}\beta_{my}|\Sigma_{AB}|\frac{1}{\sigma_{mm}}\left(\sigma_{mm} - (\sigma_{ma}^2 + \sigma_{mb}^2)(\frac{1}{\sigma_{aa\cdot b}} + \frac{1}{\sigma_{bb\cdot a}})\right)}{|\Sigma_{AB\cdot t}|} \quad (17)$$

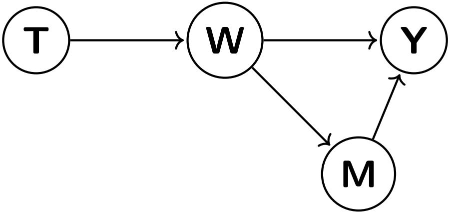
$$= \frac{\beta_{tm}\beta_{my}|\Sigma_{AB}|\frac{1}{\sigma_{mm}}\sigma_{mm\cdot AB}}{|\Sigma_{AB\cdot t}|} \quad (18)$$

$$\sigma_{mm\cdot AB}|\Sigma_{AB}| = |\Sigma_{AB\cdot m}||\sigma_{mm}|$$

$$= \beta_{tm}\beta_{my}\frac{|\Sigma_{AB\cdot m}|}{|\Sigma_{AB\cdot t}|}$$

Adjusting for A Mediator

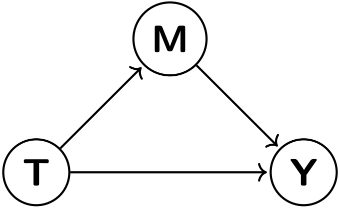
In the graph shown below, nodes M is a mediator, and the bias term is a ratio of schur complements minus a path from $W \to Y$ through only M.



$$\beta_{ty \cdot M} = \beta_{tw}\beta_{wy}\left(\frac{\sigma_{mm \cdot w}}{\sigma_{mm \cdot t}} - \beta_{wm}\sigma_{mm}\beta_{my}\right)$$

Future Work

After talking with Dr. Evans, I plan to look into the simple linear scenario to come up with constraints about whether to include M as a confounding factor.



Perhaps, if one is uncertain if M is a mediator or not, one should still include it if it meets a certain criterion.

Linear Case Constraint

We will compare two estimators MSE: a biased $\hat{\beta}_{ty \cdot m}$ and unbiased $\hat{\beta}_{ty}$
$MSE(\hat{\beta}_{ty}) = \sigma^2(X^T X)^{-1}$, where X is the design matrix and σ^2 is the residual variance. Note $X = \mathbf{t}$, so we will not include an intercept and mean center the variables. Therefore

$$MSE(\hat{\beta}_{ty}) = \frac{\sigma_{yy \cdot t}}{\sigma_{tt} * n} \tag{19}$$

The residual variance for $\hat{\beta}_{ty \cdot m} = d_{yy} = \sigma_{yy \cdot tm}$. However $(X^T X)^{-1}$ is now a matrix. We can find the variance still by using schur-complements.$(X^T X)^{-1}[1, 1] = (\sigma_{tt \cdot m} * n)^{-1} = (\sigma_{tt} * n)^{-1} = (d_{tt} * n)^{-1}$
Therefore,

$$Var(\hat{\beta}_{ty \cdot m}) = \frac{\sigma_{yy \cdot t,m}}{\sigma_{tt \cdot m} * n} \tag{20}$$

Bias for $\hat{\beta}_{ty \cdot m}$

Via the trek rule,

$$\sigma_{tm} = d_{tt}\beta_{tm}$$

and

$$\sigma_{my} = d_{mm}\beta_{my} + d_{tt}\beta_{tm}\beta_{ty} + d_{tt}\beta_{tm}^2\beta_{my}$$

Therefore,

$$\begin{aligned}
\sigma_{my} * \sigma_{tm} &= d_{tt}d_{mm}\beta_{my}\beta_{tm} + d_{tt}^2\beta_{tm}^2\beta_{ty} + d_{tt}^2\beta_{tm}^3\beta_{my} && (21)\\
&= d_{tt}d_{mm}\beta_{my}\beta_{tm} + d_{tt}^2\beta_{tm}^2\left(\beta_{tm}\beta_{my} + \beta_{ty}\right) && (22)\\
&= d_{tt}d_{mm}\beta_{my}\beta_{tm} + \frac{\sigma_{tt}^2\sigma_{tm}^2}{\sigma_{tt}^2}\left(\beta_{tm}\beta_{my} + \beta_{ty}\right) && (23)\\
&= d_{tt}d_{mm}\beta_{my}\beta_{tm} + \sigma_{tm}^2\beta_{ty}^T && (24)
\end{aligned}$$

Note the true causal effect of T on Y, which I will call β_{ty}^T is $\beta_{tm}\beta_{my} + \beta_{ty}$

Bias for $\hat{\beta}_{ty\cdot m}$ cont.

$$
\begin{aligned}
\hat{\beta}_{ty\cdot m} &= \frac{\sigma_{ty\cdot m}}{\sigma_{tt\cdot m}} && (25)\\
&= \frac{\sigma_{ty} - \sigma_{tm}\sigma_{my}/\sigma_{mm}}{\sigma_{tt} * \sigma_{mm\cdot t}/\sigma_{mm}} && (26)\\
&= \frac{\beta_{ty}^{T} * \sigma_{tt} - \sigma_{tm}\sigma_{my}/\sigma_{mm}}{\sigma_{tt} * \sigma_{mm\cdot t}/\sigma_{mm}} && (27)\\
&= \frac{\beta_{ty}^{T} * \sigma_{mm} - \sigma_{tm}\sigma_{my}/\sigma_{tt}}{\sigma_{mm\cdot t}} && (28)\\
&= \frac{\beta_{ty}^{T} * \sigma_{mm} - \left(d_{tt}d_{mm}\beta_{my}\beta_{tm} + \sigma_{tm}^{2}\beta_{ty}^{T}\right)/\sigma_{tt}}{\sigma_{mm\cdot t}} && (29)\\
&= \frac{\beta_{ty}^{T}\left(\sigma_{mm} - \sigma_{tm}^{2}/\sigma_{tt}\right) - d_{tt}d_{mm}\beta_{my}\beta_{tm}/\sigma_{tt}}{\sigma_{mm\cdot t}} && (30)
\end{aligned}
$$

Final simplifications for Bias for $\hat{\beta}_{ty \cdot m}$

$$= \frac{\beta_{ty}^T (\sigma_{mm \cdot t}) - d_{mm} \beta_{my} \beta_{tm} / \sigma_{tt}}{\sigma_{mm \cdot t}} \tag{31}$$

$$= \beta_{ty}^T - \frac{d_{mm} \beta_{my} \beta_{tm}}{\sigma_{mm \cdot t}} \tag{32}$$

$$= \beta_{ty}^T - \beta_{my} \beta_{tm} \tag{33}$$

Thus, $bias\left(\hat{\beta}_{ty \cdot m}\right) = -\beta_{my} \beta_{tm}$

Comparing MSE

$$MSE(\hat{\beta}_{ty\cdot m}) = \beta_{my}^2 \beta_{tm}^2 + \frac{d_{yy}}{\sigma_{tt\cdot m} n} \tag{34}$$

$$MSE(\hat{\beta_{ty}}) = \frac{\sigma_{yy\cdot t}}{d_{tt} n} \tag{35}$$

$$\sigma_{yy\cdot tm} = \sigma_{yy\cdot t} - \frac{\sigma_{ym\cdot t}^2}{\sigma_{mm\cdot t}} \tag{36}$$

$$\implies \sigma_{yy\cdot tm} - \sigma_{yy\cdot t} = -\frac{\sigma_{ym\cdot t}^2}{\sigma_{mm\cdot t}} \tag{37}$$

$$= \beta_{my}^2 * d_{mm} \tag{38}$$

Subtracting Conditional Variances

$$\frac{d_{yy}}{\sigma_{tt\cdot m}} - \frac{\sigma_{yy\cdot t}}{d_{tt}} \tag{39}$$

$$\sigma_{tt\cdot m} = \sigma_{mm\cdot t} d_{tt} / \sigma_{mm} \tag{40}$$

$$\sigma_{mm} = d_{tt}\beta_{tm}^2 + d_{mm} \tag{41}$$

$$= \frac{d_{yy}\left(\frac{d_{tt}\beta_{tm}^2 + d_{mm}}{d_{mm}}\right)}{dtt} - \frac{\sigma_{yy\cdot t}}{d_{tt}} \tag{42}$$

$$= \frac{d_{yy}\beta_{tm}^2}{d_{mm}} + \frac{d_{yy}}{d_{tt}} - \frac{\sigma_{yy\cdot t}}{d_{tt}} \tag{43}$$

$$= \frac{d_{yy}\beta_{tm}^2}{d_{mm}} - \frac{d_{mm} * \beta_{my}^2}{d_{tt}} \quad (20) \tag{44}$$

Putting a Constraint on MSE

$$MSE(\hat{\beta}_{ty\cdot m}) - MSE(\hat{\beta}_{ty}) \tag{45}$$

$$= \beta_{my}^2 \beta_{tm}^2 + \frac{d_{yy}}{\sigma_{tt\cdot m} n} - \frac{\sigma_{yy\cdot t}}{d_{tt} n} \tag{46}$$

$$= \beta_{my}^2 \beta_{tm}^2 + \frac{d_{yy} \beta_{tm}^2}{n * d_{mm}} - \frac{d_{mm} * \beta_{my}^2}{d_{tt} * n} \tag{47}$$

$$\implies \beta_{my}^2 \beta_{tm}^2 + \frac{d_{yy} \beta_{tm}^2}{n * d_{mm}} - \frac{d_{mm} * \beta_{my}^2}{d_{tt} * n} < 0 \tag{48}$$

$$\implies n * \beta_{my}^2 \beta_{tm}^2 + \frac{d_{yy} \beta_{tm}^2}{d_{mm}} < \frac{d_{mm} \beta_{my}^2}{d_{tt}} \tag{49}$$

This equation makes sense. If d_{mm} is large, we should include it. Further if d_{tt} is small, adding in d_{mm} will help estimate $T \rightarrow Y$.

Adjusting for the Bias

I ran some simulations below on how we can achieve a better estimate if we correct for the bias we induce.
Mathematically the adjustment estimate is

$$\hat{\beta}_{ty\cdot a} = \hat{\beta}_{ty\cdot m} + \hat{\beta}_{tm} * \hat{\beta}_{my\cdot t} \tag{50}$$

I also looked at the *R* mediator package, which employs a simply total effect regression for their estimate.