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. [1] that for Structural Equation Models (SEMs), an optimal adjustment set for finding the causal effect of $T \to Y$ is:

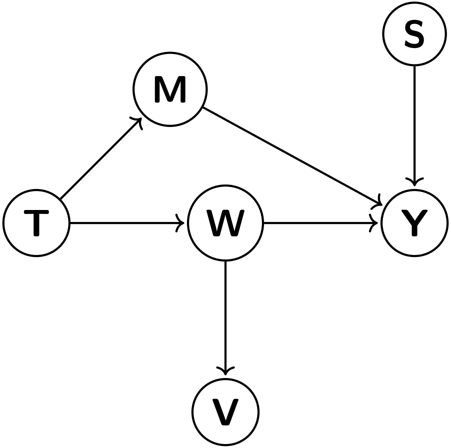
$$O_g(T \to Y) = pa_g(\mathrm{cn}_g(T \to Y)) \setminus (\mathrm{cn}_g(T \to Y) \cup \{T\}).$$

- It was also shown for the non-parametric case [2]

Example

So for instance in the graph below the

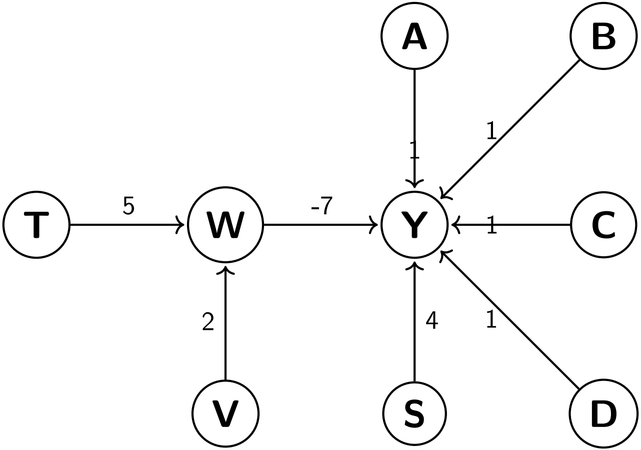
- $cn_G(T \to Y) = \{W, M, Y\}$
- $pa_g(cn_g(T \to Y)) = \{T, W, M, S\}$
- $\implies O_g(T \to Y) = \{S\}$



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

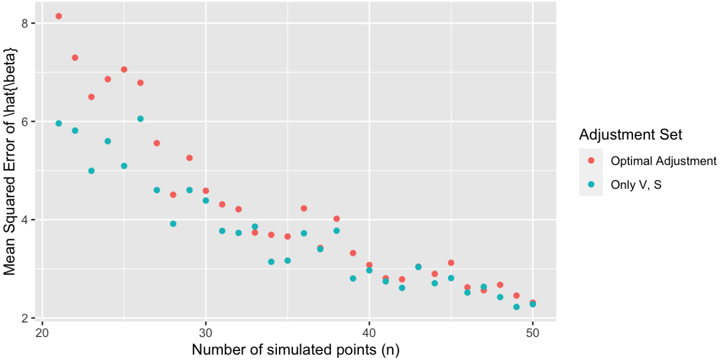
In R, this can be seen as

```
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)
```

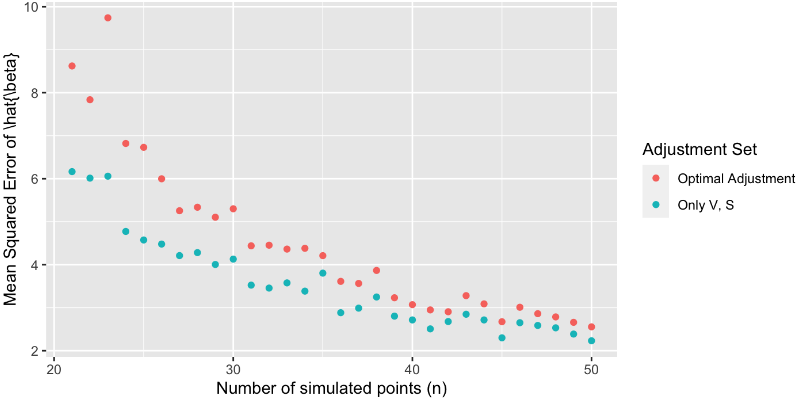
Results

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:



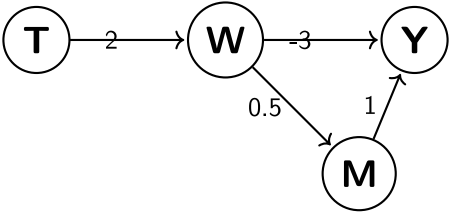
Results 2: Changing Parameters

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



Forbidden Nodes

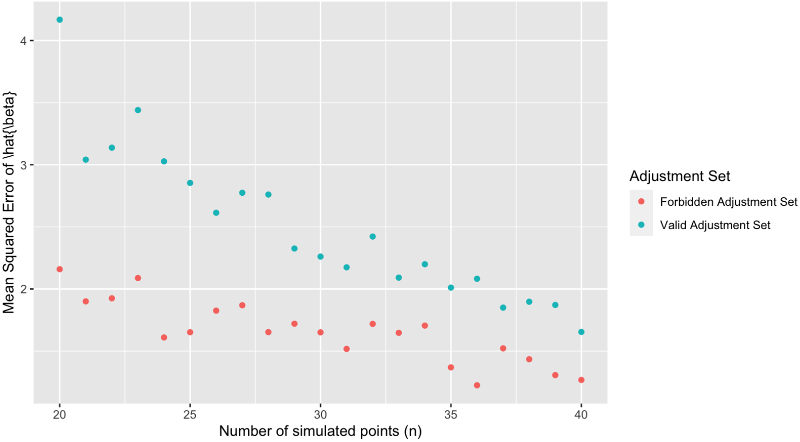
I've also been considering Forbidden Nodes and including them in adjustment sets.



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

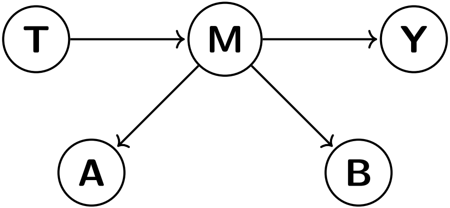
Variable Mediator

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:

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

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

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

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

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

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

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

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

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

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

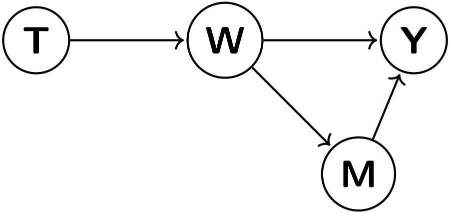
$$= \frac{\beta_{tm}\beta_{my}|\Sigma_{AB}|\frac{1}{\sigma_{mm}}\sigma_{mm\cdot AB}}{|\Sigma_{AB\cdot t}|} \tag{18}$$

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

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

Adjusting for A Mediator

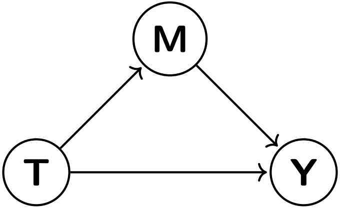
In the graph shown below, nodes M is a mediator, and the bias term is a ratio of schur complements minus a path from $W \to Y$ through only M.



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

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.

References

Leonard Henckel, Emilija Perković, and Marloes H. Maathuis.
Graphical criteria for efficient total effect estimation via adjustment in causal linear models.
Journal of the Royal Statistical Society Series B: Statistical Methodology, 84(2):579–599, March 2022.

Andrea Rotnitzky and Ezequiel Smucler.
Efficient adjustment sets for population average causal treatment effect estimation in graphical models.
Journal of Machine Learning Research, 21(188):1–86, 2020.