Causal Inference for Health Data

(STATS C160/C260 — Winter 2026)

Lecture 15:
Unobserved Confounding — Part II

Drago Plečko

UCLA College | Physical Sciences
Statistics & Data Science

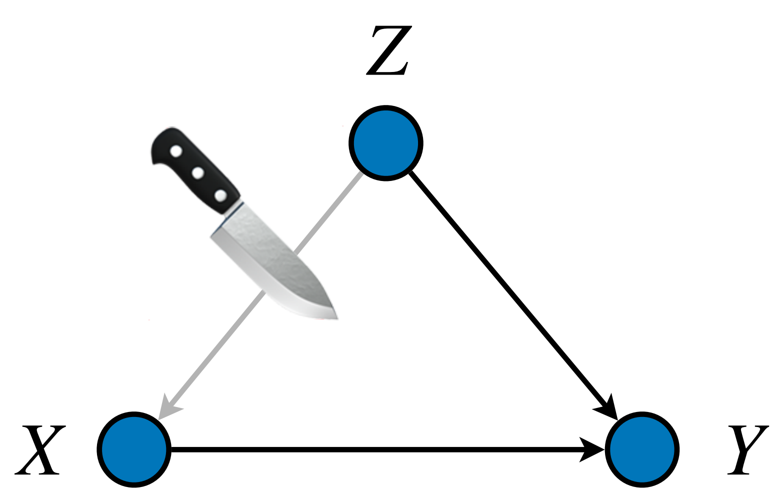
Recap: Unobserved Confounding Bias

- Common bias culprit: unobserved confounding

Our Analysis

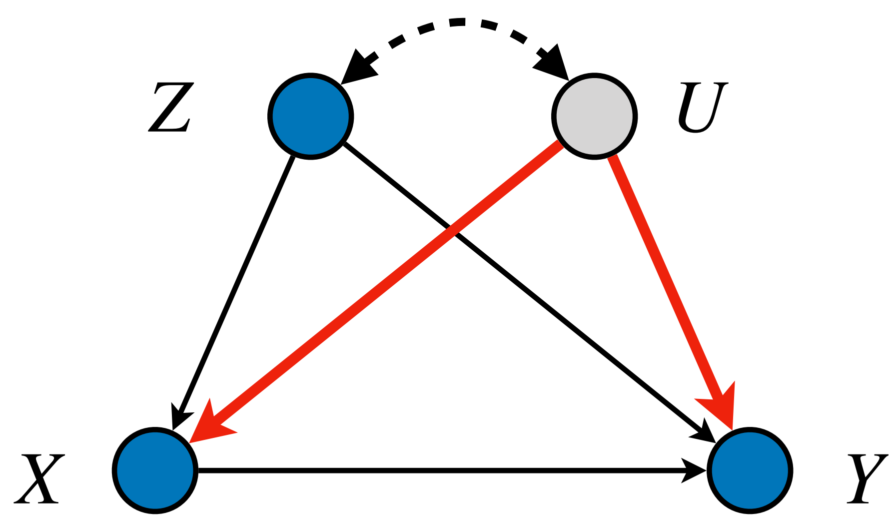
Reality

In this case, our findings will not translate to the RCT setting



Recap: E-values & Sensitivity Parameters

- Recall that we studied a sensitivity analysis for binary outcomes Y,
- The important sensitivity parameters were related to the strength of $U \to X, U \to Y$ effects,



- Is there an approach that can be used for the case of continuous outcomes?
- In this lecture, we study sensitivity in the linear case.

Non-Parametric Models in Causal Inference

SCM Knowledge

very detailed knowledge

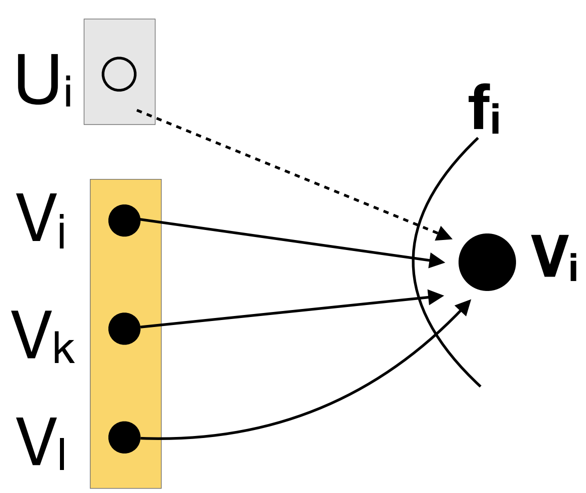
$$V_i \leftarrow \underbrace{\boxed{f_i}}(\mathrm{pa}(V_i), U_i)$$

exact functional form of how nature works

e.g., $V_i \leftarrow V_j^{2/3} \sin V_k + V_\ell \log U_i$

Diagram Knowledge

knowledge about f_i **arguments only**



Non-Parametric Inference

- with the non parametric approach, we are trying to learn conditionals $P(V_i \mid \mathrm{pa}(V_i))$,
- we put no restriction on the possible class of functions we fit, meaning that conditionals are fully determined by the data.

Parametric Models in Causal Inference

SCM Knowledge

very detailed knowledge

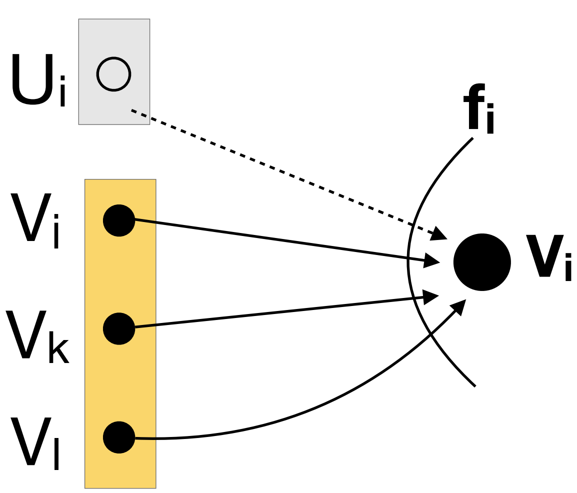
$$V_i \leftarrow \underbrace{\boxed{f_i}}(\mathrm{pa}(V_i), U_i)$$

exact functional form of how nature works

e.g., $V_i \leftarrow V_j^{2/3} \sin V_k + V_\ell \log U_i$

Diagram Knowledge

knowledge about f_i arguments only



Parametric Inference

- with the parametric approach, we are specifying a class for the conditionals $P(V_i \mid \mathrm{pa}(V_i))$,
- conditionals are determined by the data, but only within a function class,
- implicitly, we are making assumptions about f_i!

Parametric Models: Recovering Causal Effects?

- care needs to be taken when using parametric models, since they recover correct causal effects under stricter conditions,
- e.g.,

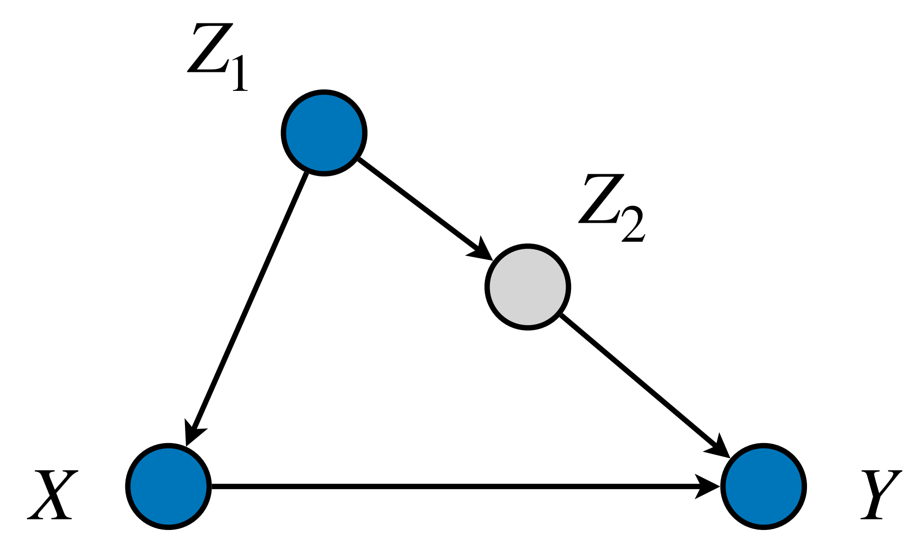
Proposition. Let Z be a back-door set for (X, Y), and let the mechanism f_Y be linear in its arguments. Then, linear regression $Y \sim X + Z$ recovers the true causal effect.

✔ ? ✘ ?

Parametric Models: Recovering Causal Effects?

- Consider the following example:

$$Z_1 \leftarrow \epsilon_{Z_1}$$
$$Z_2 \leftarrow Z_1^2 + \epsilon_{Z_2}$$
$$X \leftarrow Z_1 + \epsilon_X$$
$$Y \leftarrow \beta X + Z_2 + \epsilon_Y$$

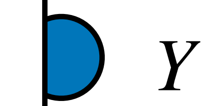


- Z_1 is back-door for (X, Y); f_Y linear,
- Is the correct causal effect recovered by linear regression $Y \sim X + Z_1$? **Simulation.**

Parametric Models: Recovering Causal Effects?

- Consider the following example:

Simulation Conclusion: the correct causal effect is not recovered.



- Z_1 is back-door for (X, Y); f_Y linear,
- Is the correct causal effect recovered by linear regression $Y \sim X + Z_1$? **Simulation.**

Parametric Models: Recovering Causal Effects?

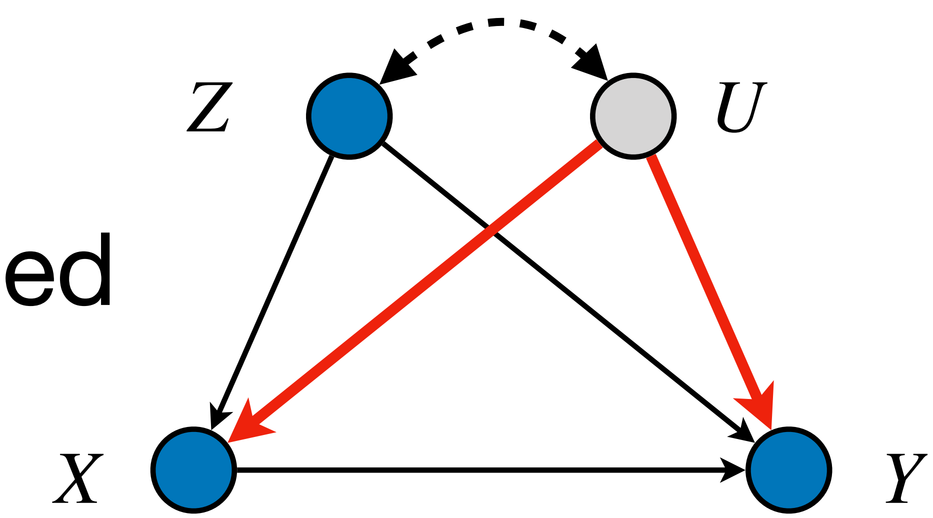
Proposition. Let Z be a back-door set for (X, Y), and let the mechanism f_Y be linear in Z, X. Then, linear regression $Y \sim X + Z$ recovers the true causal effect.

Parametric assumptions can be strong in practice; we use them when they are necessary.

Linear Sensitivity Analysis

- In this lecture, we are interested in the setting of an unobserved confounder U,

- The important sensitivity parameters, as before, are related to the strength of $U \to X, U \to Y$ effects,



- We assume that the f_Y mechanism is linear in X, Z, U; but we do not make additional parametric assumptions.

Omitted Variable Bias (OVB)

- Our first goal is to quantify the so-called *omitted variable bias* (OVB),
- Suppose that the regression model

$$Y \leftarrow X\beta + Z\gamma + U\delta + \epsilon$$

is true, but we only regress $Y \sim X + Z$ using OLS, obtaining the parameter $\hat{\beta}_{uc}$ for X

- How much bias does $\hat{\beta}_{uc}$ incurr compared to the true β?

Helper: Frisch-Waugh-Lovell Theorem

Theorem (FWL). Consider the regression model

$$Y = X\beta + Z\gamma + \epsilon,$$

where X is an $n \times p$ matrix, and Z is $n \times k$. Let $\hat{\beta}$ be the OLS estimate for β. Let $Y^{\perp Z}$, $X^{\perp Z}$ denote the residuals of Y, X after performing OLS regression onto Z. Denote by $\hat{\beta}_{ts}$ the coefficient for X when regressing

$$Y^{\perp Z} \sim X^{\perp Z}.$$

Then, we have $\hat{\beta} = \hat{\beta}_{ts}$.

FWL Theorem Proof

Proof (FWL). Consider the OLS optimization problem

$$\operatorname{argmin}_{\beta,\gamma} \|Y - X\beta - Z\gamma\|^2,$$

which can be re-written as

$$\operatorname{argmin}_{\beta} \operatorname{argmin}_{\gamma} \|Y - X\beta - Z\gamma\|^2.$$

Let $g(\beta) = \operatorname{argmin}_{\gamma} \|Y - X\beta - Z\gamma\|^2$. The OLS solution for γ in $g(\beta)$ is given by $(Z^T Z)^{-1} Z^T (Y - X\beta)$.

Therefore, the fitted values $Z\hat{\gamma}$ equals $P_Z(Y - X\beta)$, where $P_Z = Z(Z^T Z)^{-1} Z^T$ is the projection matrix on span(Z).

FWL Theorem Proof

Proof (FWL continued).

Then,

$$\begin{aligned}
\text{argmin}_\beta g(\beta) &= \text{argmin}_\beta \|Y - X\beta - P_Z(Y - X\beta)\|^2 \\
&= \text{argmin}_\beta \|(I - P_Z)Y - (I - P_Z)X\beta\|^2
\end{aligned}$$

which is exactly the OLS regression $Y^{\perp Z}$ onto $X^{\perp Z}$.

Omitted Variable Bias via FWL

- Consider now the setting $Y \leftarrow \beta X + Z\gamma + \delta U + \epsilon$, with X, U 1-dimensional. Using FWL, we have

$$\hat{\beta}_{uc} = \frac{Cov(X^{\perp Z}, Y^{\perp Z})}{Var(X^{\perp Z})}$$

$$Y = \hat{\beta} X + Z\hat{\gamma} + \hat{\delta} U + R$$

where $R \perp X, Z, U$

$$= \frac{Cov(X^{\perp Z}, \hat{\beta} X^{\perp Z} + \hat{\delta} U^{\perp Z})}{Var(X^{\perp Z})}$$

$$= \hat{\beta} + \hat{\delta} \times \frac{Cov(X^{\perp Z}, U^{\perp Z})}{Var(X^{\perp Z})}$$

Estimate of β in complete regression

Estimate of δ in complete regression

Coefficient of $X \rightarrow U$ after partialing out Z

OVB Interpretation & Drawbacks

- Therefore, we can compute the resulting bias:

$$\text{bias} \triangleq \hat{\beta}_{uc} - \hat{\beta} = \hat{\delta} \times \frac{Cov(X^{\perp Z}, U^{\perp Z})}{Var(X^{\perp Z})}$$

linear impact of $U \rightarrow Y$

linear imbalance caused by $X \rightarrow U$ (after partialing out Z)

- However, this parameterization has a drawback: for certain confounders, units may be unclear (e.g., “health-seeking behavior” -> how does one measure it?)

- It makes the δ parameter difficult to specify.

Rewriting the Bias via R^2

- We thus re-write the bias:

force corr instead of Cov:

$$\mathrm{corr}(A, B) = \frac{Cov(A, B)}{\sqrt{Var(A)Var(B)}}$$

$$\text{bias} = \frac{Cov(Y^{\perp Z,X}, U^{\perp Z,X})}{Var(U^{\perp Z,X})} \times \frac{Cov(X^{\perp Z}, U^{\perp Z})}{Var(X^{\perp Z})}$$

$$= \frac{\mathrm{corr}(Y^{\perp Z,X}, U^{\perp Z,X})sd(Y^{\perp Z,X})}{sd(U^{\perp Z,X})} \times \frac{\mathrm{corr}(X^{\perp Z}, U^{\perp Z})sd(U^{\perp Z})}{sd(X^{\perp Z})}$$

$$= \frac{\mathrm{corr}(Y^{\perp Z,X}, U^{\perp Z,X})\mathrm{corr}(X^{\perp Z}, U^{\perp Z})}{sd(U^{\perp Z,X})/sd(U^{\perp Z})} \times \frac{sd(Y^{\perp Z,X})}{sd(X^{\perp Z})}$$

Question: are these correlations something familiar?

Rewriting the Bias via R^2

- Note that:

$$\text{corr}(A,B)^2 = \frac{Cov(A,B)^2}{Var(A)Var(B)} = \frac{Cov(A,B)}{Var(A)}\frac{Cov(A,B)}{Var(B)}$$

$$= \hat{\beta}_{B\sim A} \times \frac{Cov(A,B)}{Var(B)} = \frac{Cov(\hat{\beta}_{B\sim A}A,B)}{Var(B)}$$

$$= \frac{Cov(\hat{B},B)}{Var(B)} = \frac{Var(\hat{B},\hat{B})}{Var(B)}$$

This is the coefficient of determination R^2!

Corr = ... = R^2

- Note that:

$$\boxed{\mathrm{corr}(A,B)^2} = \frac{Cov(A,B)^2}{Var(A)Var(B)} = \frac{Cov(A,B)}{Var(A)}\frac{Cov(A,B)}{Var(B)}$$

$$= \hat{\beta}_{B\sim A} \times \frac{Cov(A,B)}{Var(B)} = \frac{Cov(\hat{\beta}_{B\sim A}A,B)}{Var(B)}$$

$$= \frac{Cov(\hat{B},B)}{Var(B)} = \frac{Var(\hat{B},\hat{B})}{Var(B)}$$

$$\boxed{= R^2_{B\sim A}}$$

This is the coefficient of determination R^2 (proportion variance explained)

Rewriting the Bias via R^2

$$\text{bias} = \frac{\underbrace{\text{corr}(Y^{\perp Z,X}, U^{\perp Z,X})}_{\sqrt{R^2_{Y\sim U|Z,X}}}\,\underbrace{\text{corr}(X^{\perp Z}, U^{\perp Z})}_{\sqrt{R^2_{X\sim U|Z}}}}{\underbrace{sd(U^{\perp Z,X})/sd(U^{\perp Z})}_{\sqrt{1 - R^2_{X\sim U|Z}}}} \times \frac{sd(Y^{\perp Z,X})}{sd(X^{\perp Z})}$$

is this also familiar to us?

$$\Longrightarrow \qquad \text{bias} = \sqrt{\frac{R^2_{Y\sim U|Z} R^2_{X\sim U|Z}}{1 - R^2_{X\sim U|Z}}} \times \boxed{\frac{sd(Y^{\perp Z,X})}{sd(X^{\perp Z})}}$$

Rewriting the Bias — $se(\hat{\beta})$

$$\frac{sd(Y^{\perp Z,X})}{sd(X^{\perp Z})} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i^{\perp Z,X})^2}}{sd(X^{\perp Z})}$$

$$= \frac{\sqrt{\frac{1}{n-p}\sum_{i=1}^{n}(Y_i^{\perp Z,X})^2}}{\sqrt{n}\cdot sd(X^{\perp Z})}\sqrt{n-p}$$

$$= \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i^{\perp Z} - \bar{X}^{\perp Z})^2}}\sqrt{n-p}$$

this is the standard error of $\hat{\beta}_{uc}$

$$= se(\hat{\beta}_{uc})\sqrt{n-p}$$

Towards a Relative Bias

- Therefore, the bias can be written as:

$$\text{bias} = \sqrt{\frac{R^2_{Y\sim U|Z,X} R^2_{X\sim U|Z}}{1 - R^2_{X\sim U|Z}}(n-p)} \times se(\hat{\beta}_{uc})$$

$$\implies \text{rel-bias} \triangleq \frac{\text{bias}}{\hat{\beta}_{uc}} = \underbrace{\sqrt{\frac{R^2_{Y\sim U|Z,X} R^2_{X\sim U|Z}}{1 - R^2_{X\sim U|Z}}}}_{\text{comes from sensitivity parameters!}} \times \underbrace{\left(\frac{\hat{\beta}_{uc}}{se(\hat{\beta}_{uc})\sqrt{n-p}}\right)^{-1}}_{\text{comes from data: t-value divide by } \sqrt{\text{df}}\text{!}}$$

$R^2_{Y\sim U|Z,X} = R^2_{X\sim U|Z}$ Simplification

- Recall that for E-values, we assumed equal strength of the sensitivity parameters $RR_{XU} = RR_{UY}$,
- We can make a similar simplification in the linear case,
- What is a concerning bias level? Relative bias of 100%.

with $R^2_{Y\sim U|Z,X} = R^2_{X\sim U|Z} = r^2$, we want:

$$\sqrt{\frac{r^4}{1-r^2}} \times \boxed{\left(\frac{\hat{\beta}_{uc}}{se(\hat{\beta}_{uc})\sqrt{n-p}}\right)}^{-1} < 1$$

label this with f

Robustness Value (RV)

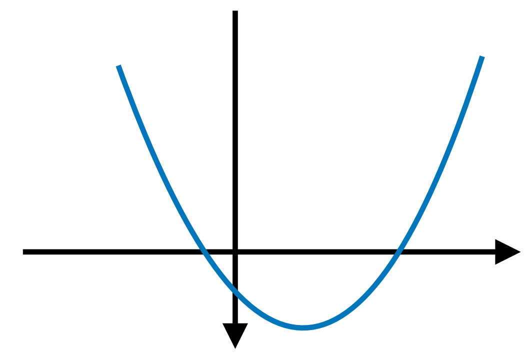
- Re-arranging, we obtain:

$$\sqrt{\frac{r^4}{1-r^2}} \times \frac{1}{f} < 1$$

$$\iff \quad r^4 + r^2 f^2 - f^2 < 0$$

$$\iff \quad r^2 \leq \frac{1}{2}(\underbrace{\sqrt{f^4 - 4f^2}}_{\text{this is the robustness value (RV)}} - f^2)$$

quadratic in r^2



- If $R^2_{Y\sim U|Z,X}, R^2_{X\sim U|Z} < RV$, or conclusions are robust.

Discussion: does $X \to U$ need to be linear?

- In this analysis, we assumed that the f_Y mechanism was linear in its arguments,
- But we also used the $R^2_{X \sim U|Z}$ from a linear model for U; does this mean some linearity in $X \to U$ is needed?
- It turns out the answer is no; even if $E[U \mid X, Z]$ takes a complex form, our sensitivity analysis still works; the only thing that matters is the linear imbalance in U caused by X, parameterized by $R^2_{X \sim U|Z}$.

Heuristic Approach: Leave-One-Confounder-Out

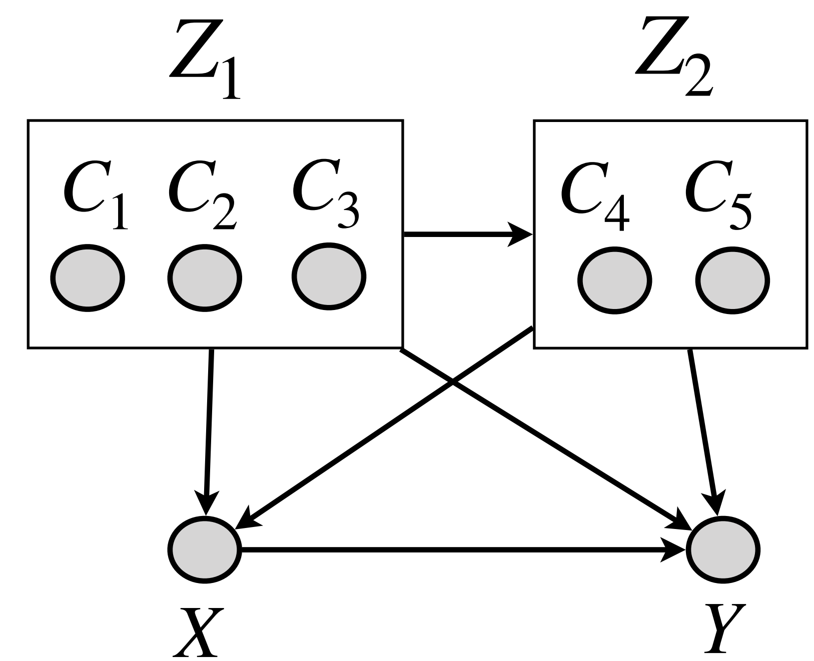
- So, how can one reason about possibly appropriate values of the parameters $R^2_{Y\sim U|Z,X}$, $R^2_{X\sim U|Z}$?

- We can proceed similarly as for the setting of E-values, by leaving out observed confounders and computing their R^2 values,

- This gives a sense of how large a confounder would have to be, to explain away the effect we are seeing.

VAHCS: Causal Diagram

- Variables:
  - parental education C_1,
  - parental separation C_2,
  - antisocial behavior in adolescence C_3,
  - adolescent depression/anxiety C_4,
  - alcohol use in adolescence C_5,
  - cannabis use in adolescence X,
  - adulthood mental health score Y.



Z_1, Z_2 **is back-door for** $X, Y \Longrightarrow E[y \mid do(x)] = \sum_{z_1, z_2} E[y \mid x, z_1, z_2] P(z_1, z_2)$