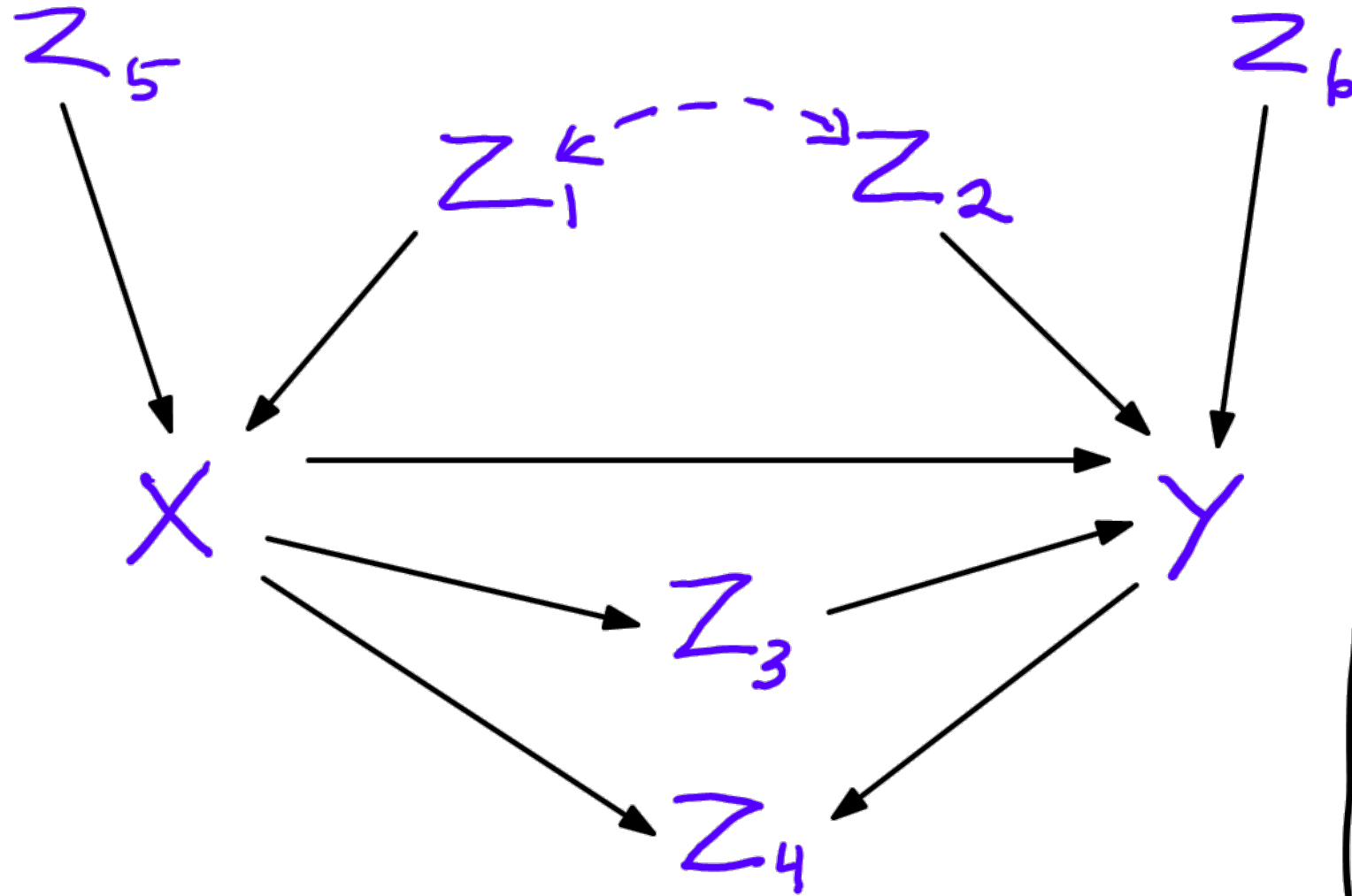


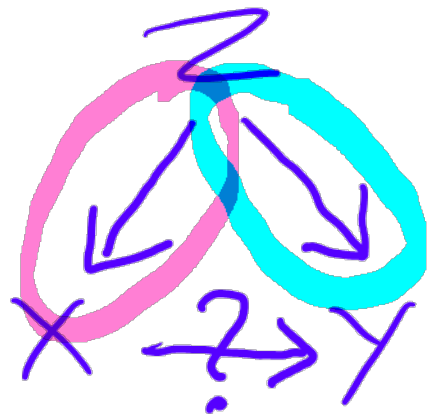
Does X cause Y?

- A bird's eye view of methods with observational data
- Lord's Paradox and the role of longitudinal data

Causal Graph Pearl & Mackenzie (2019)

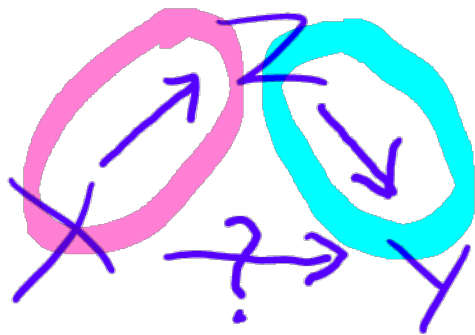


DAG =
Directed
Acyclic
graph



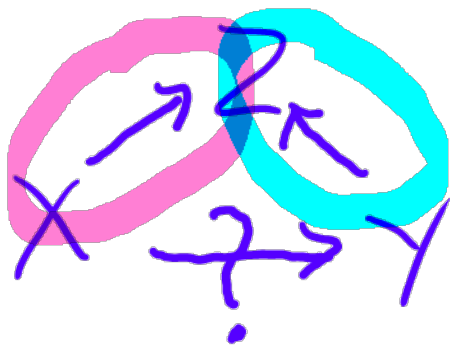
confounder

must include to see
see causal effect
of X on Y



mediator

must exclude
including may
wipe out a true effect



collider
e.g.
selection

must exclude
including may
create the impression
of an effect although there is
none

Moderators?

Can have - Confounder - moderators
- mediator - moderators
- collider - moderators

Also mediator-colliders, etc

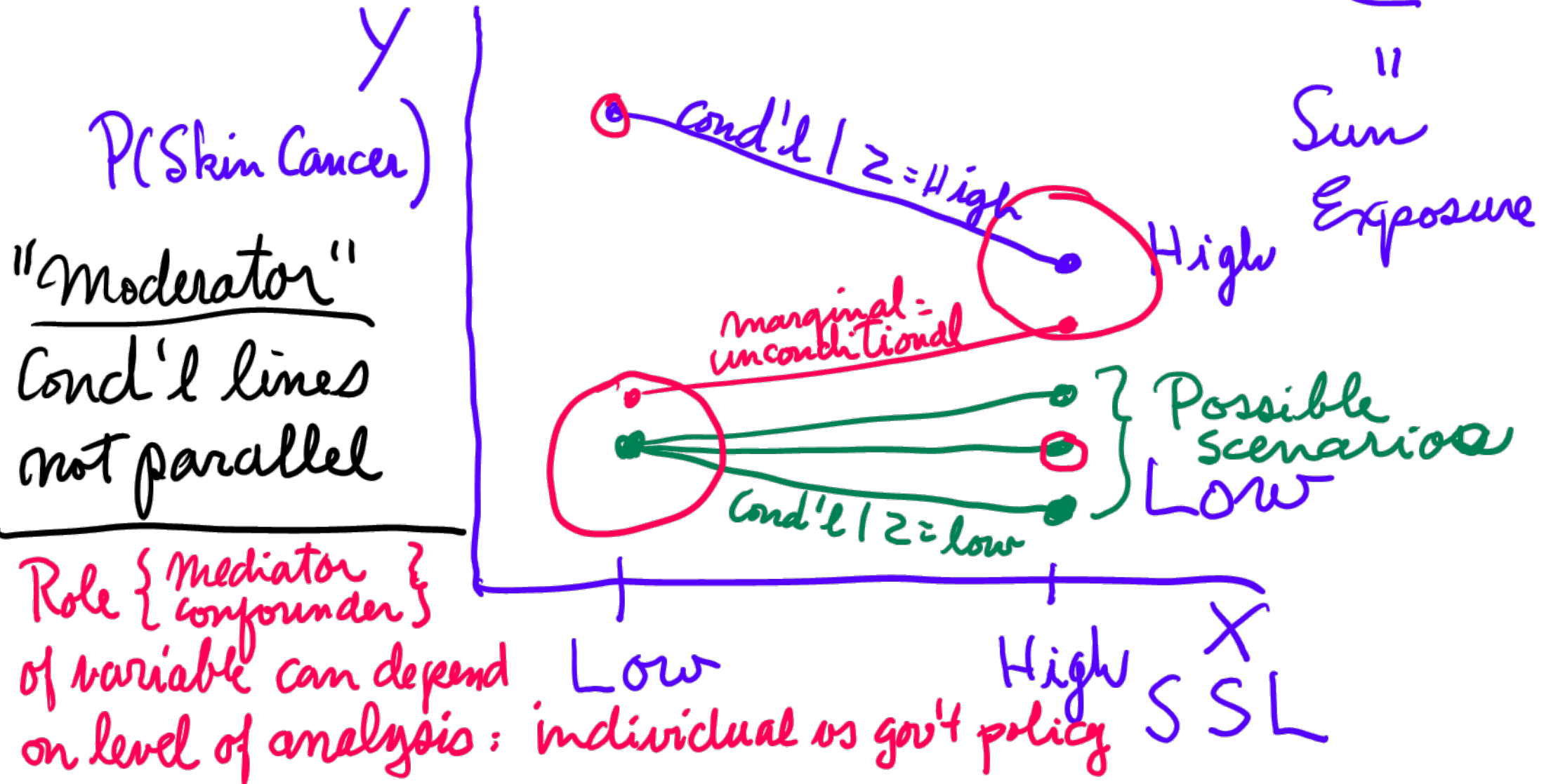
BUT not represented by DAGs

— So convenient visualizations of DAG are useful abstraction but limited in practice

— Avoiding inclusion of mediators more critical than avoiding colliders since inclusion of other confounders can correct for inclusion of a collider.

Moderator (= interaction) in data space

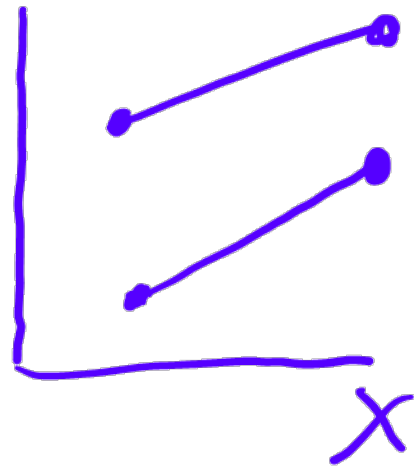
SSL example



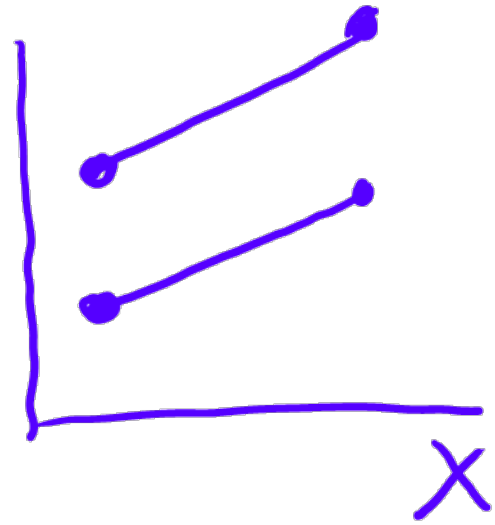
Moderation may be removable by ^{monotonically} transforming Y if

- 1) Cond'l effects in same direction
- 2) No crossings of lines

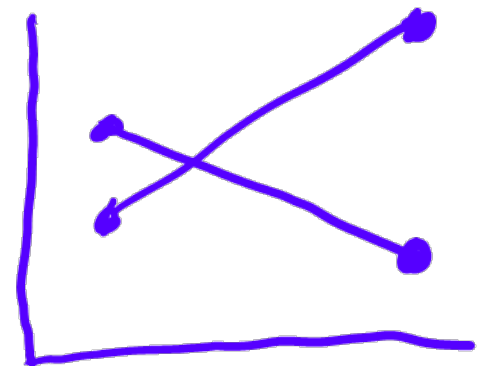
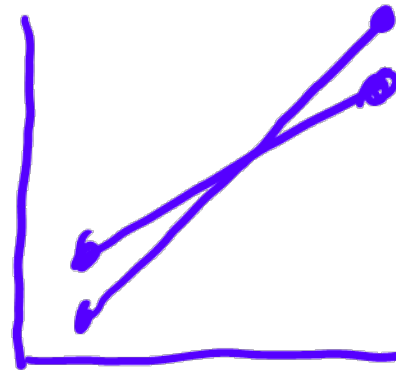
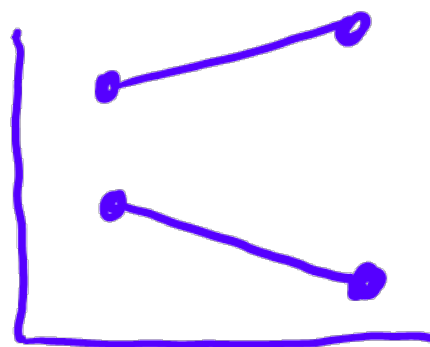
Removable $P(Y)$



log odds
"
 $\log\left(\frac{P(Y)}{1-P(Y)}\right)$

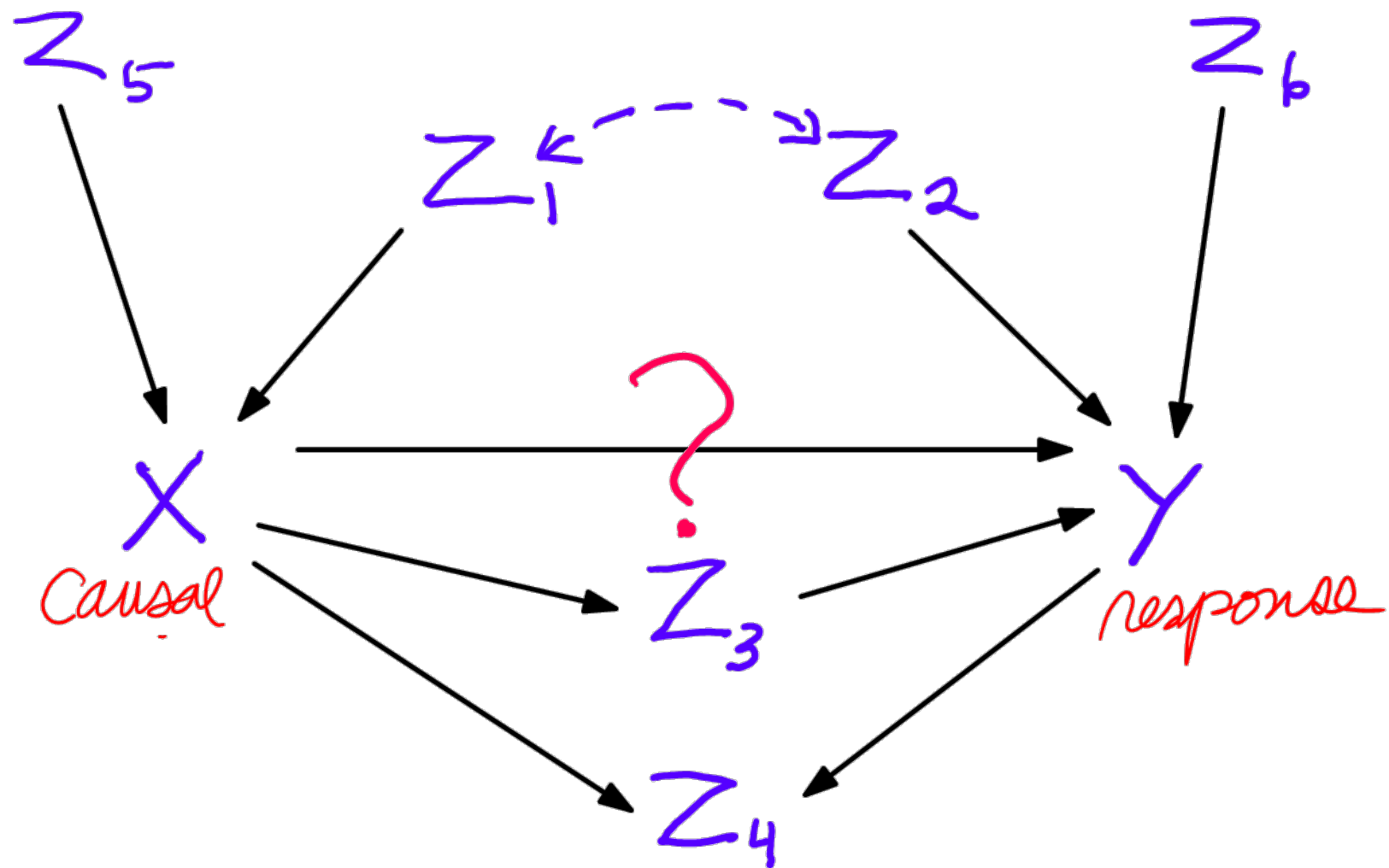


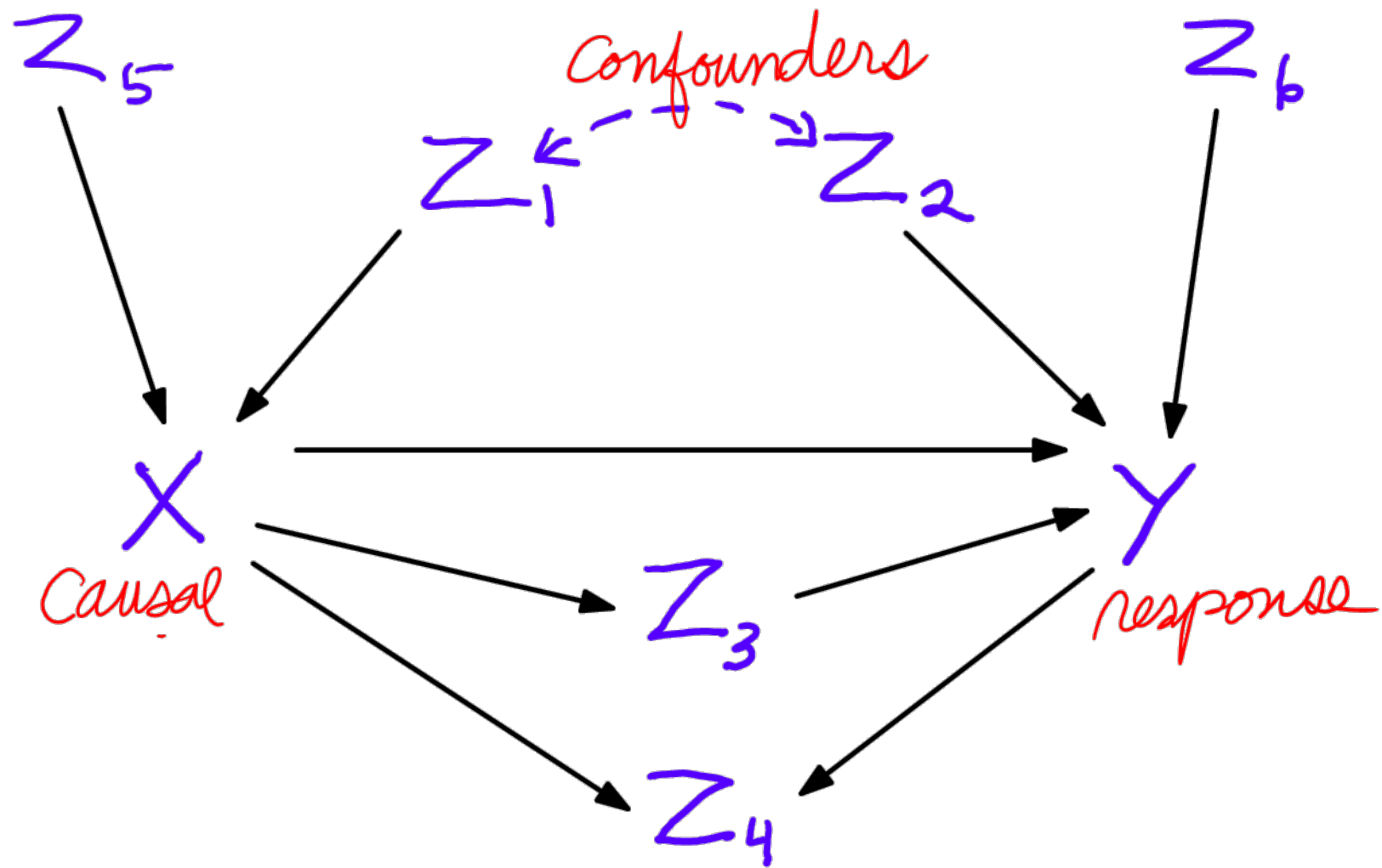
Nonremovable

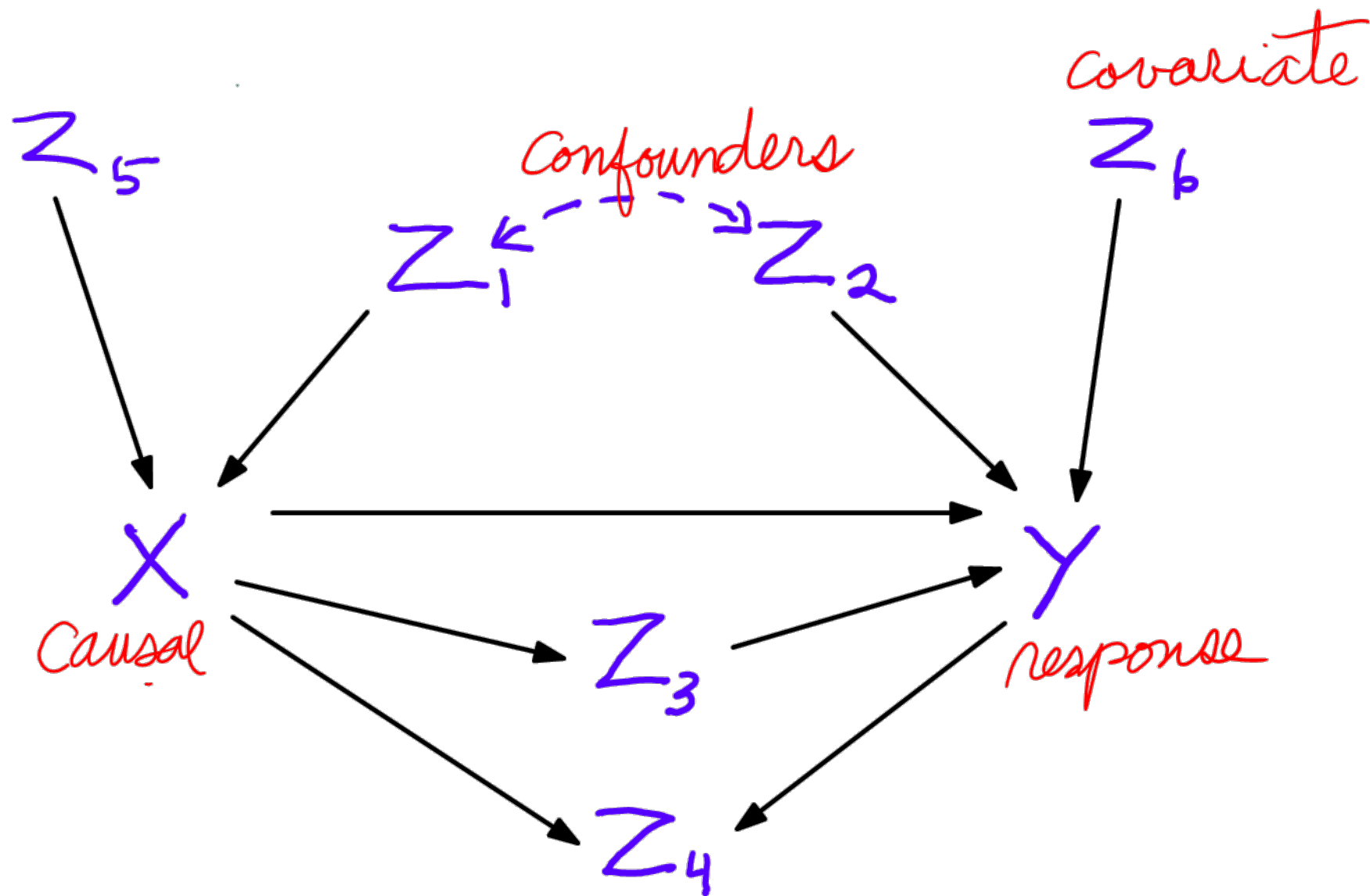


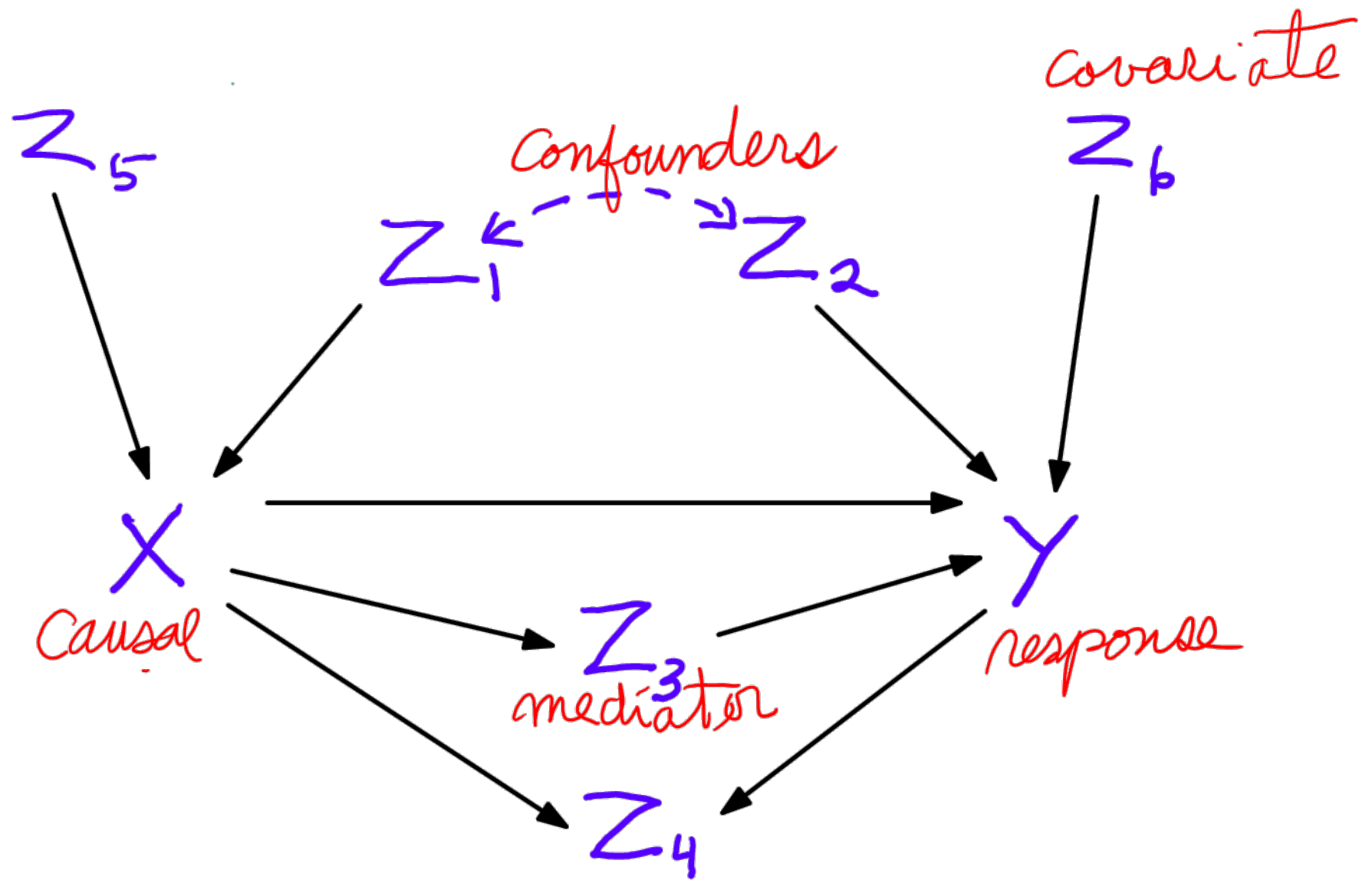
Note:

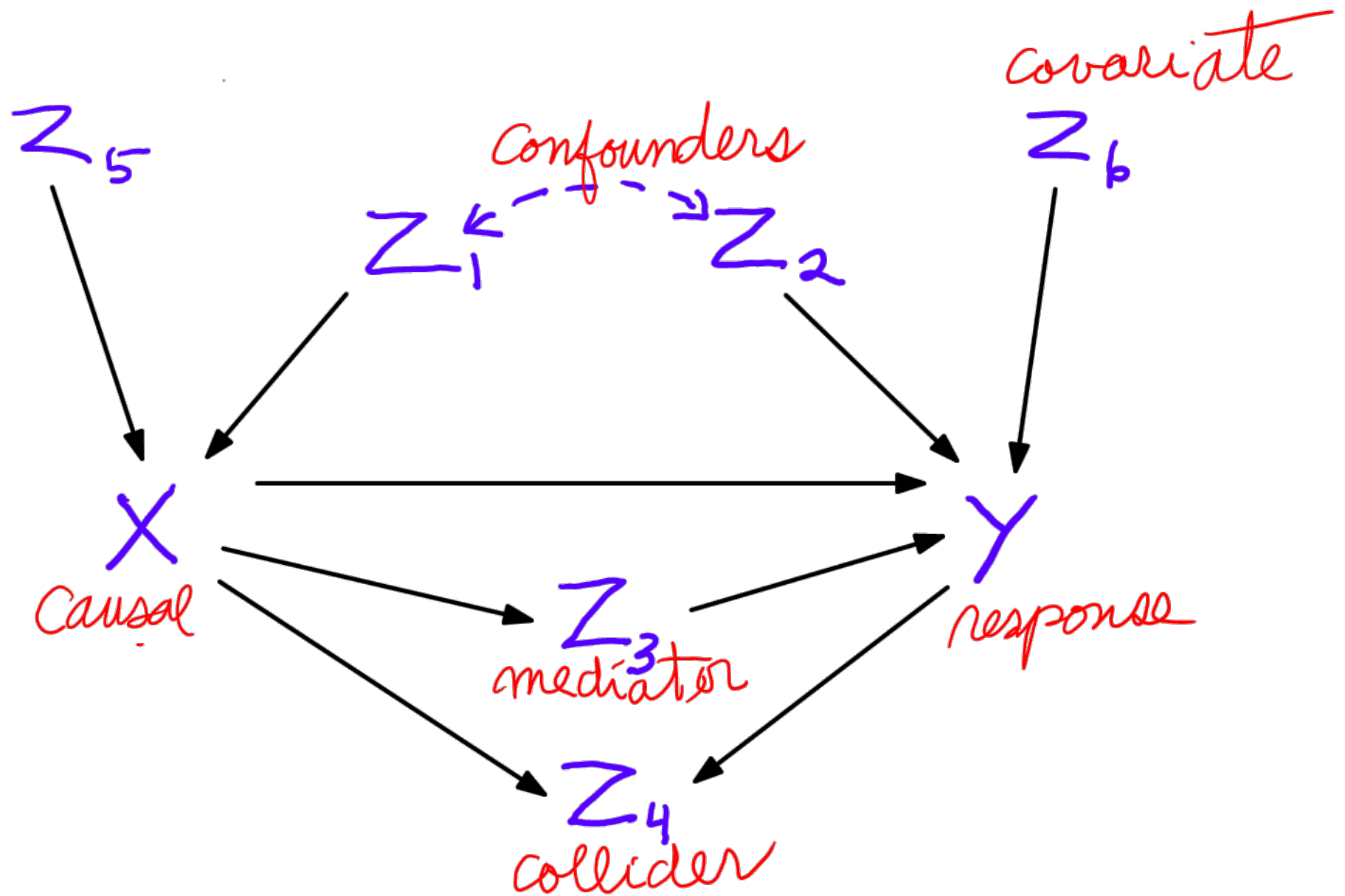
- Simpson effect: reversal of cond'l vs. marginal effects
 - Moderation (= interaction in relation of X and Z with Y)
 - Association (predictive)
 - Causality
- are distinct but not unrelated concepts

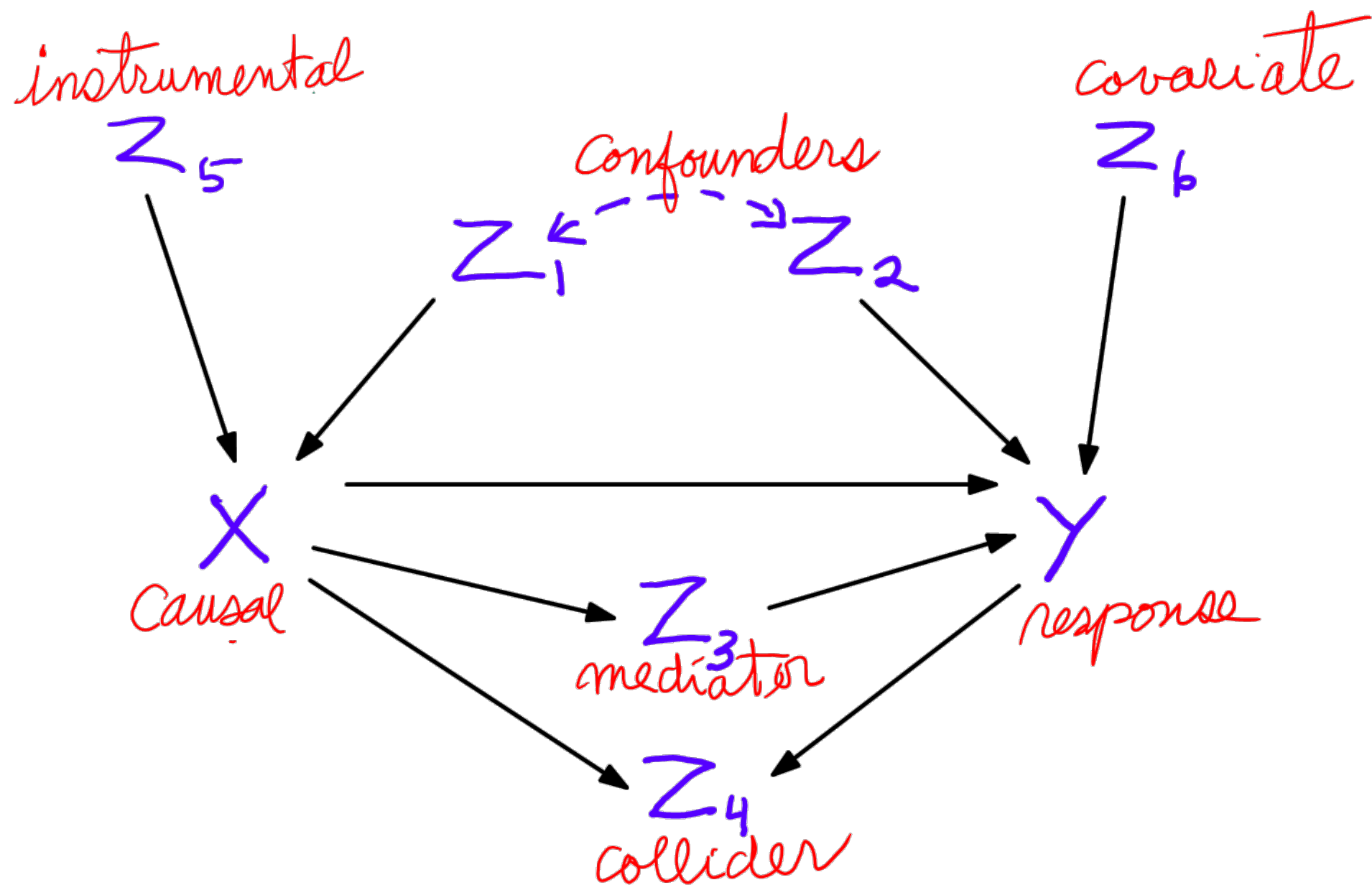


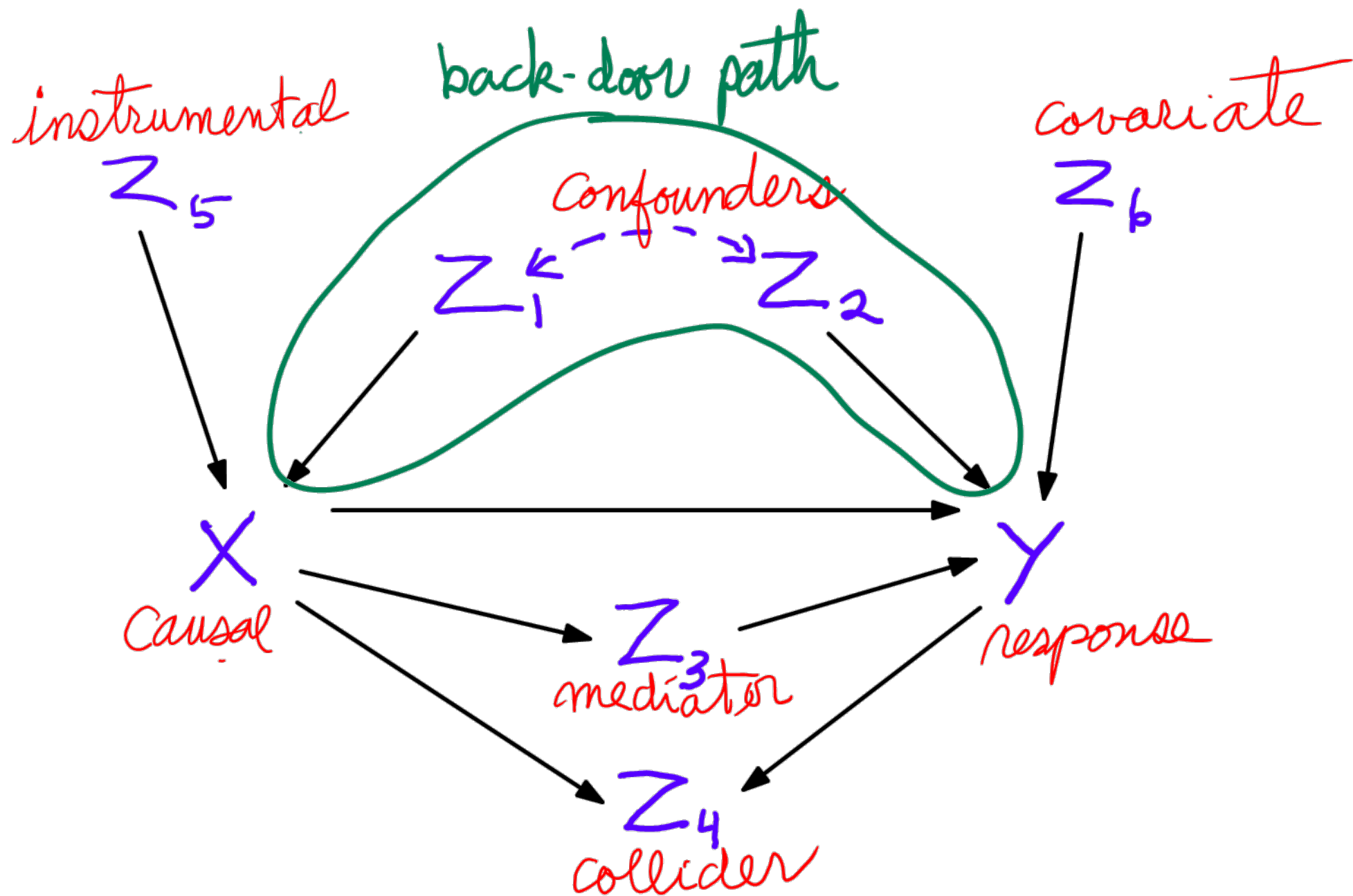


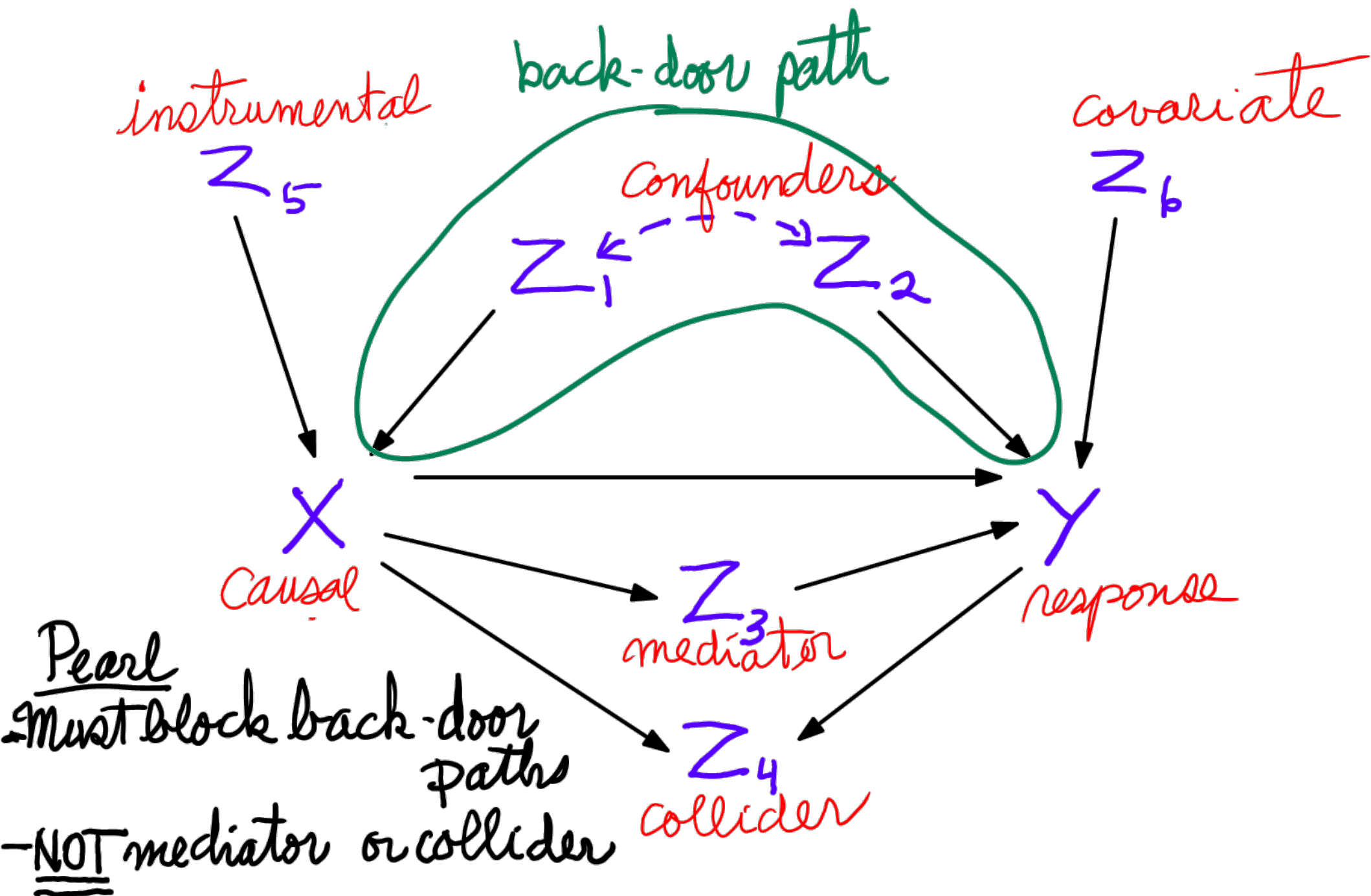


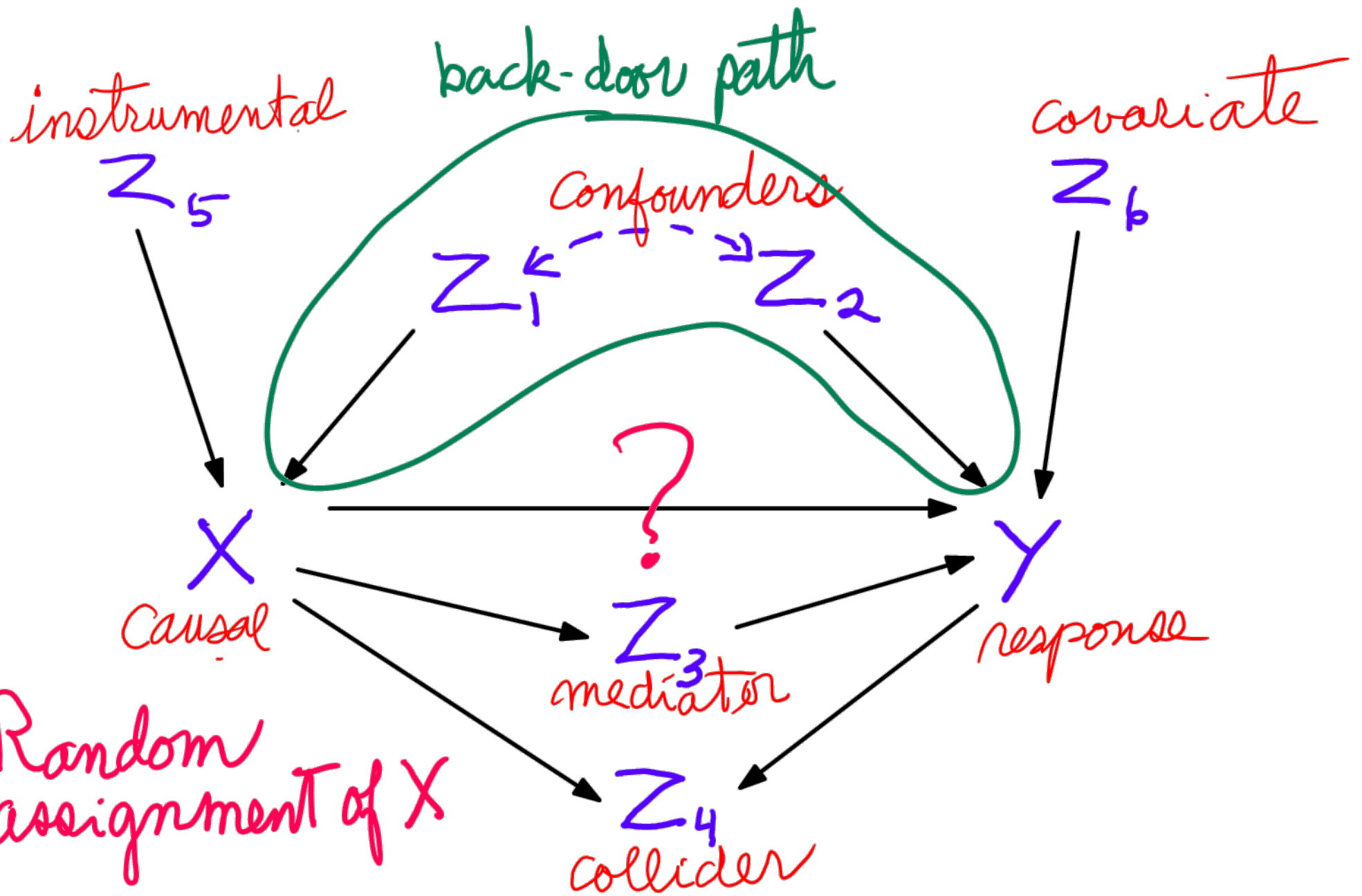




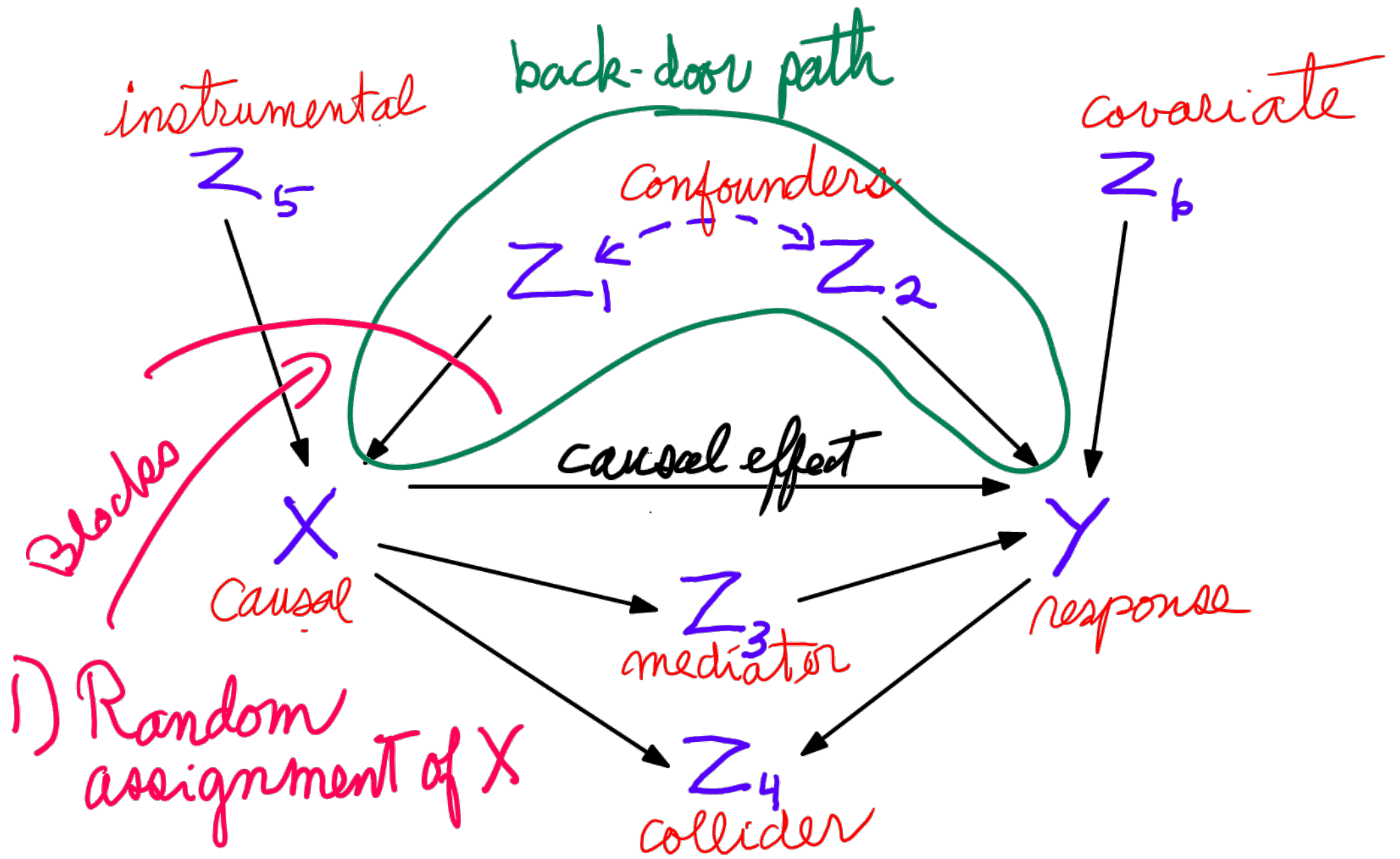


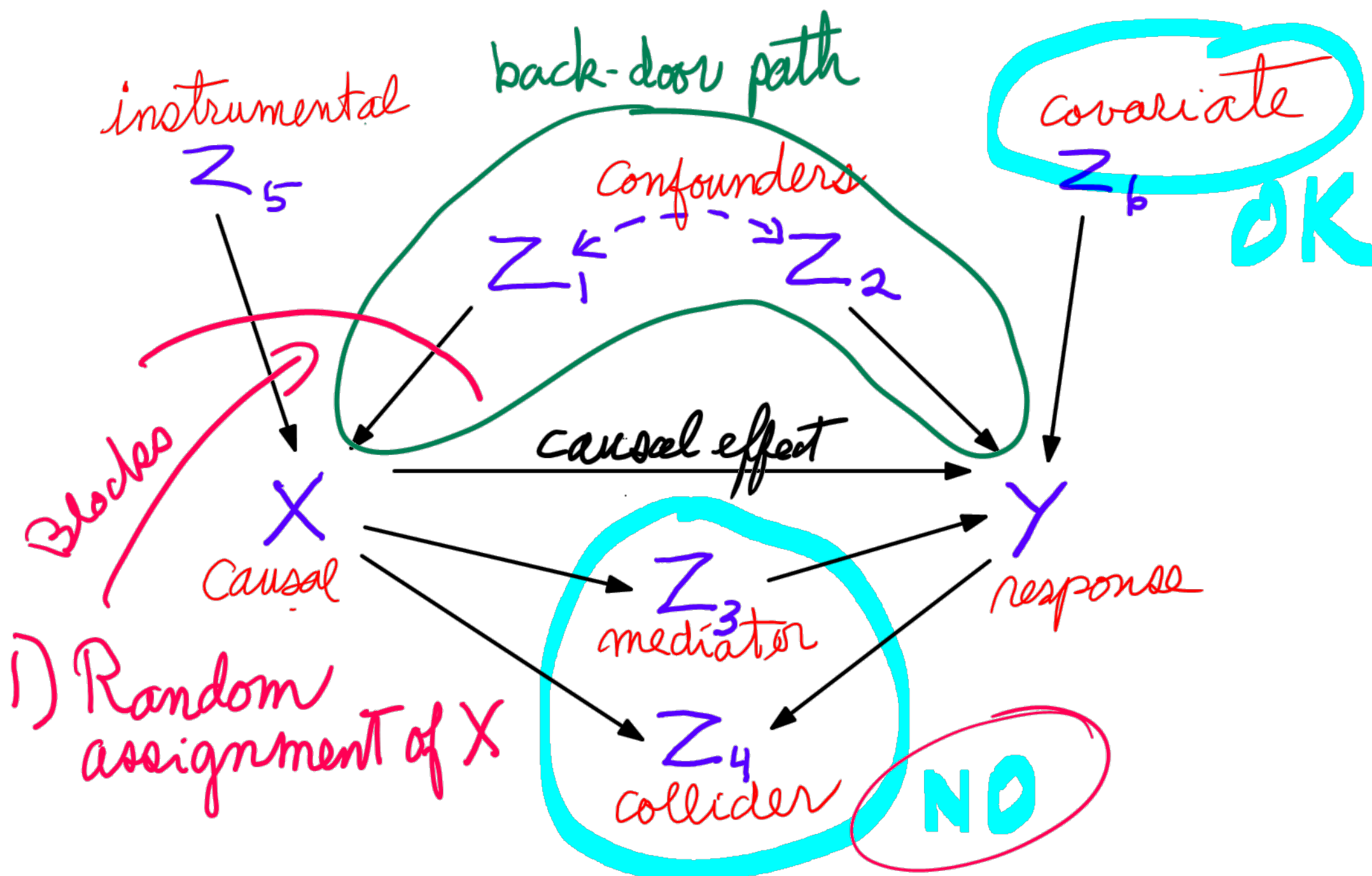


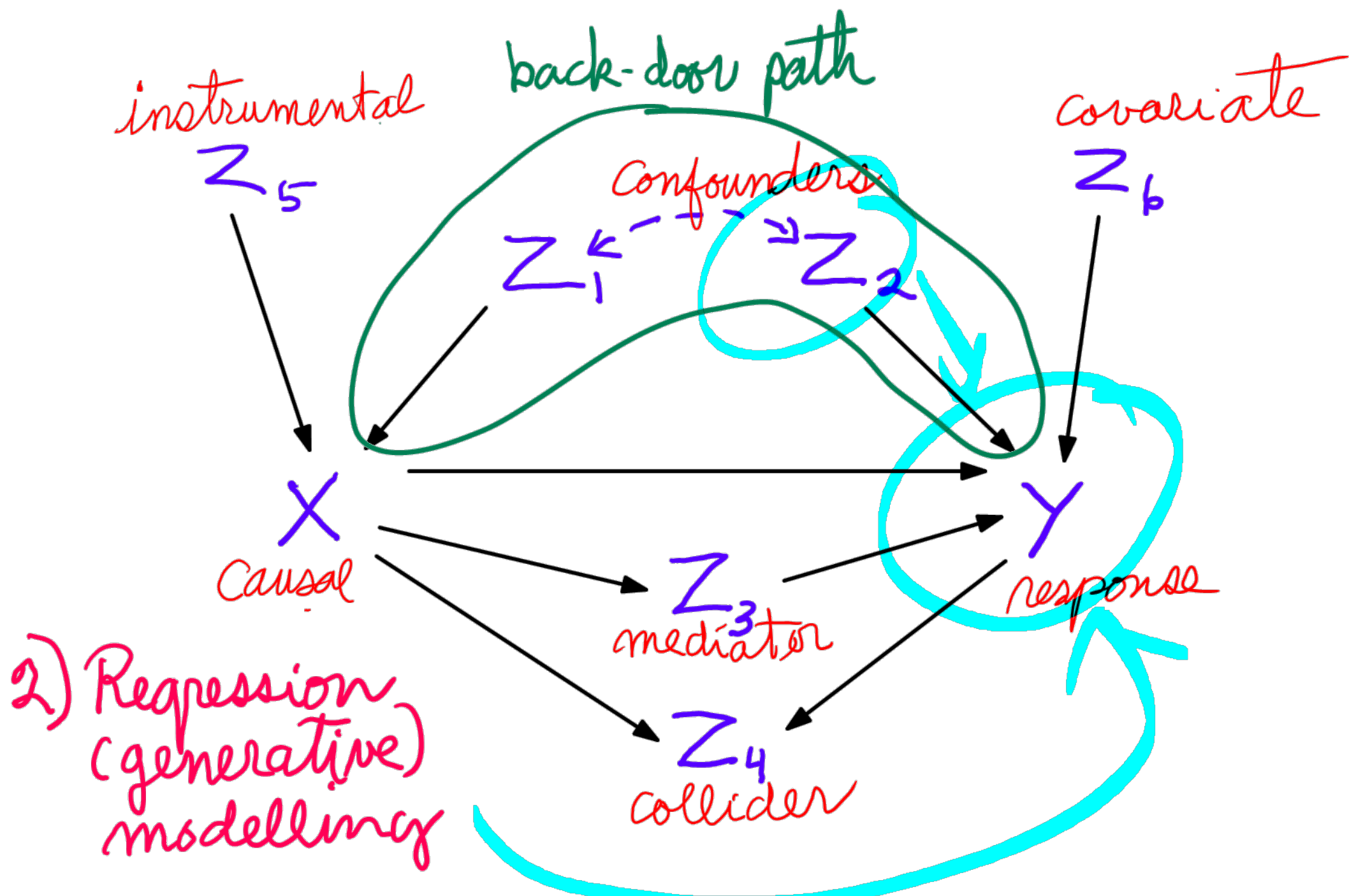


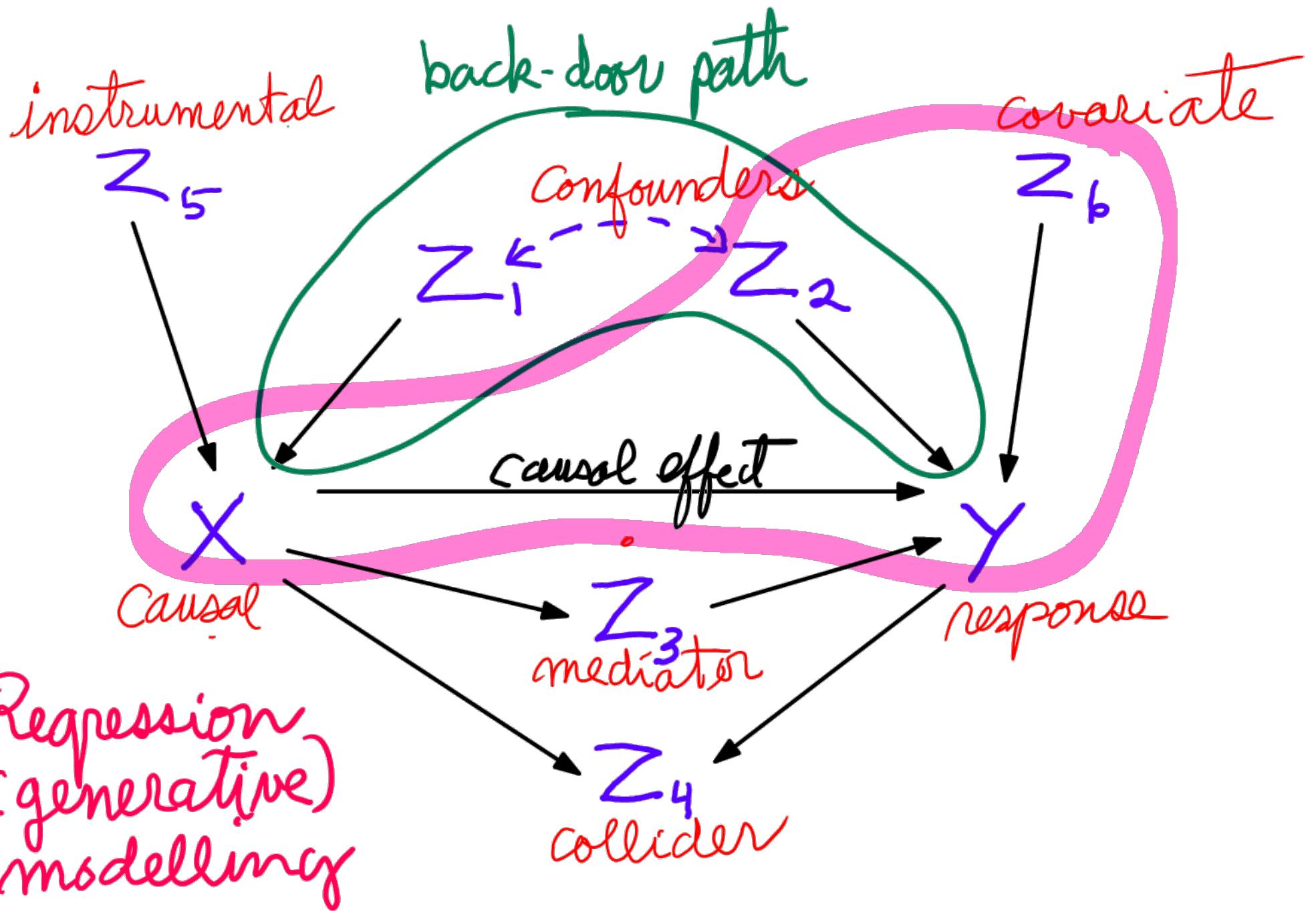


1) Random assignment of X

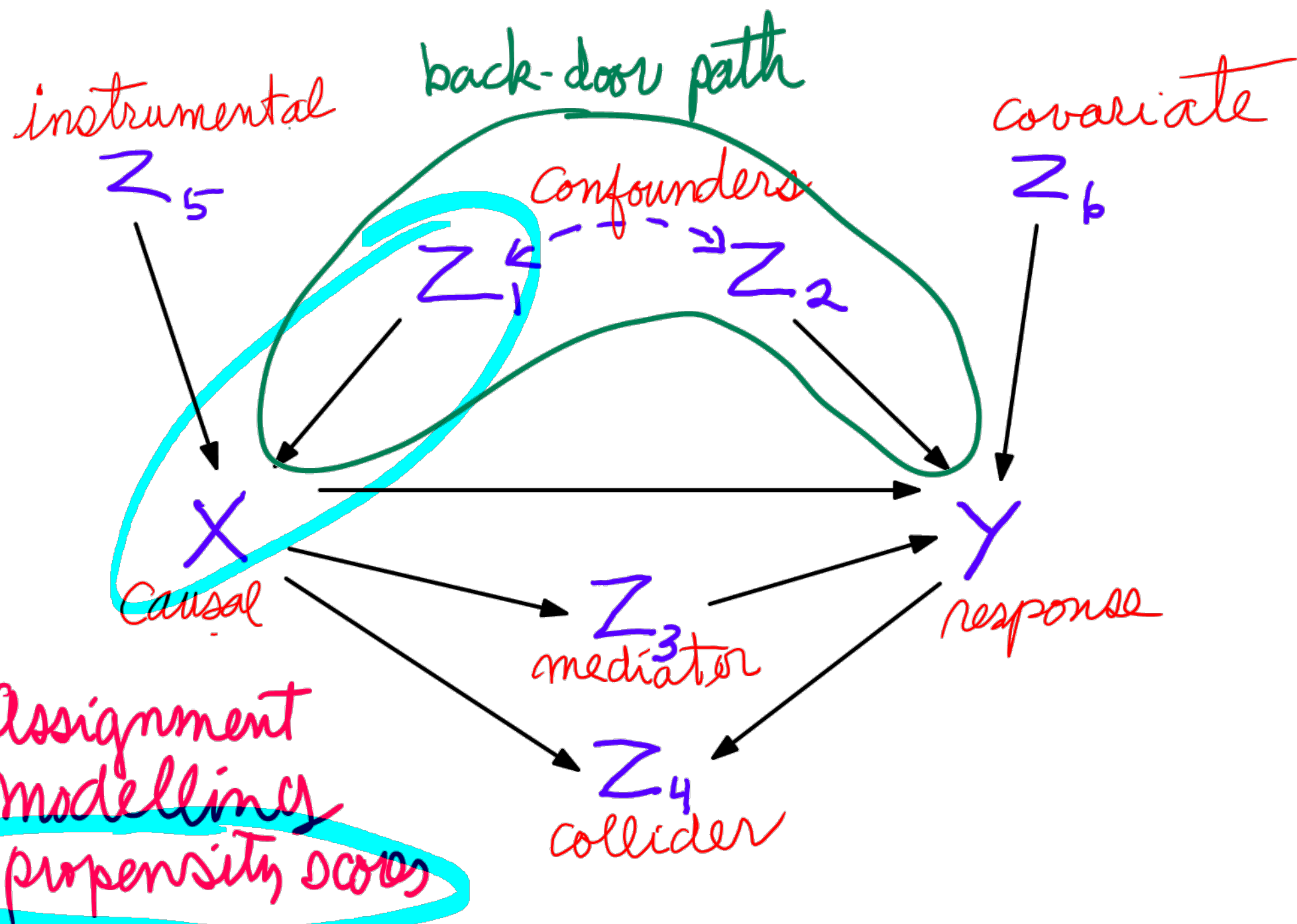




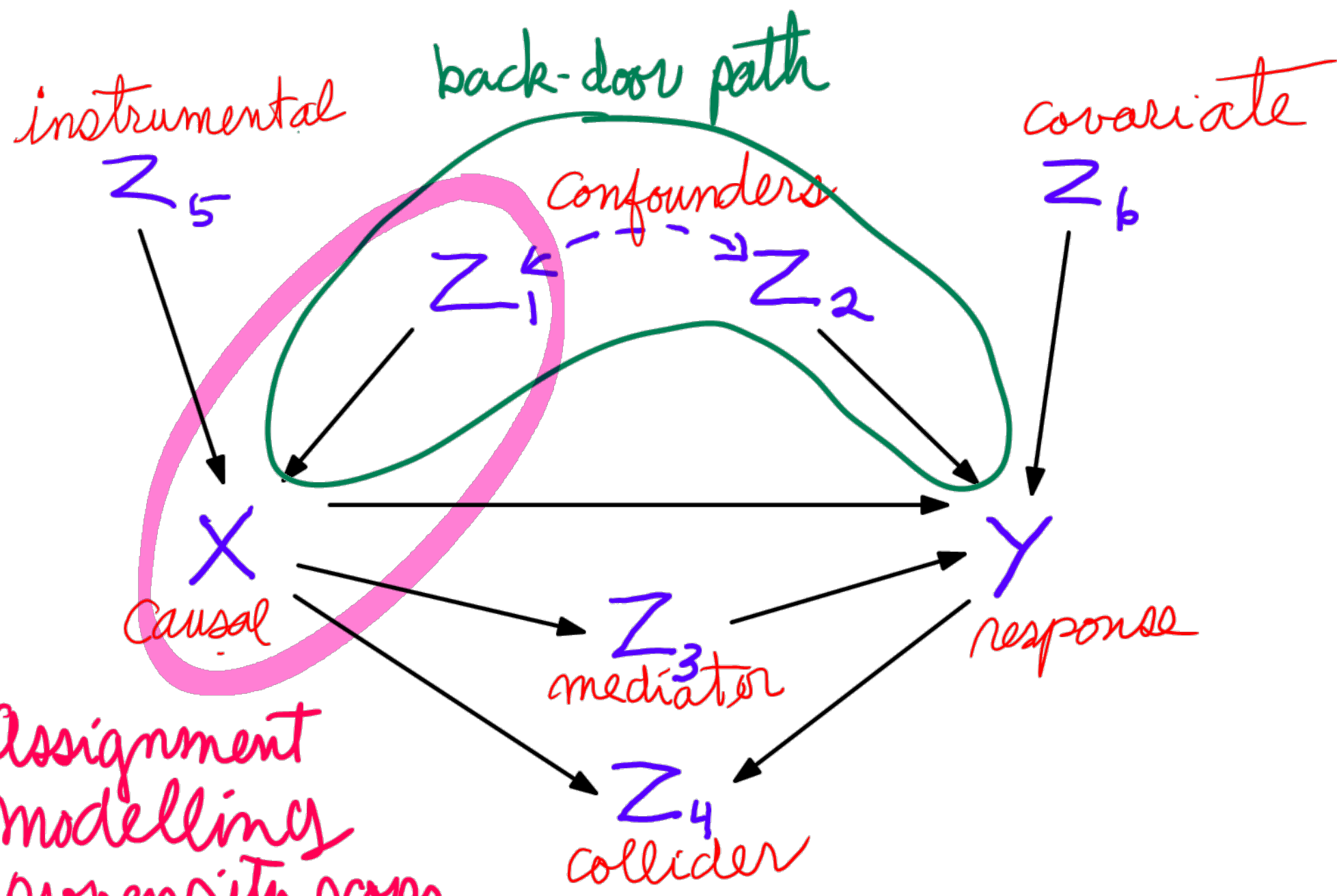




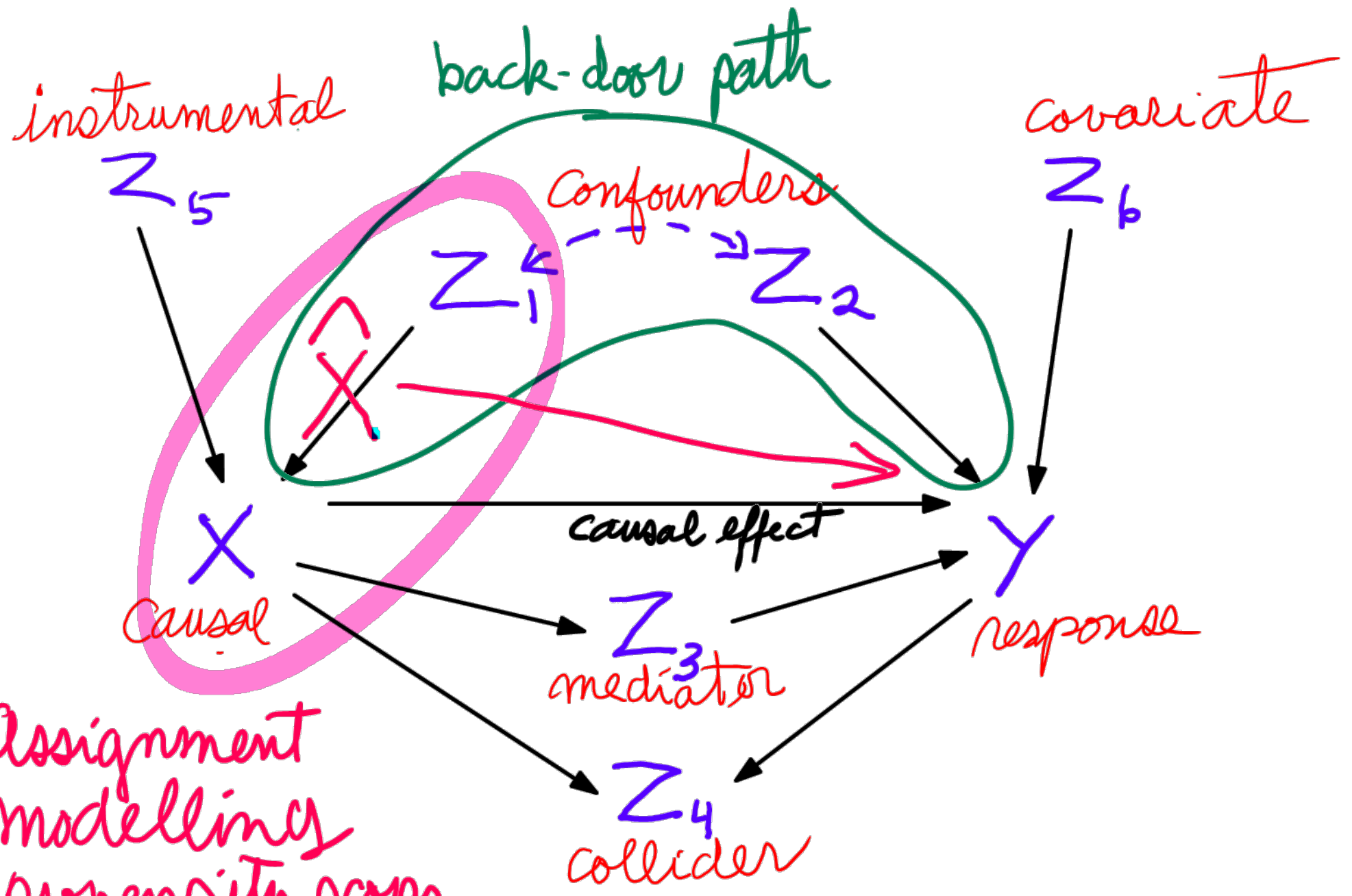
2) Regression (generative) modelling



3) Assignment modelling
- propensity scores

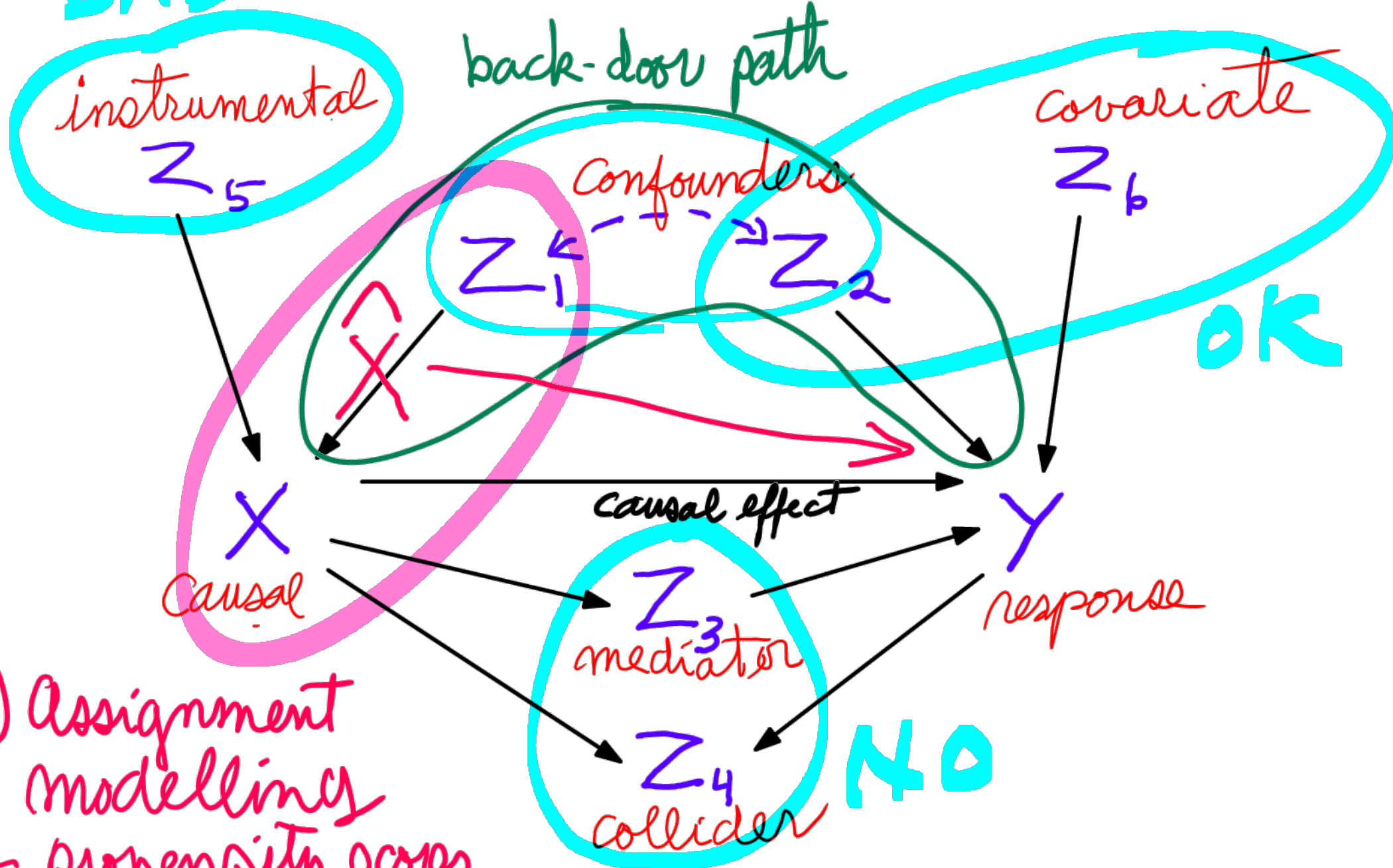


3) Assignment modelling
- propensity scores



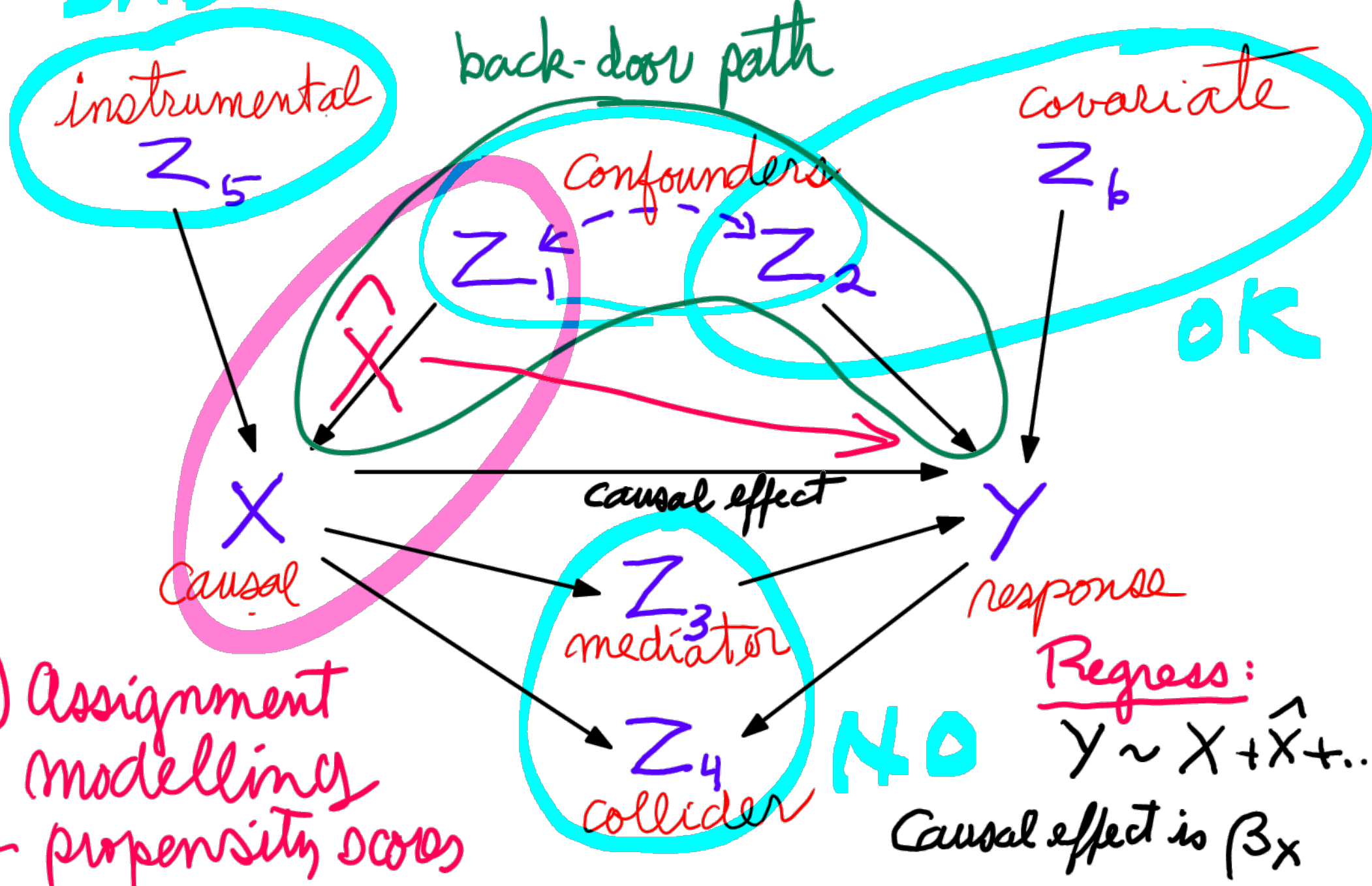
3) Assignment modelling
- propensity scores

BAD

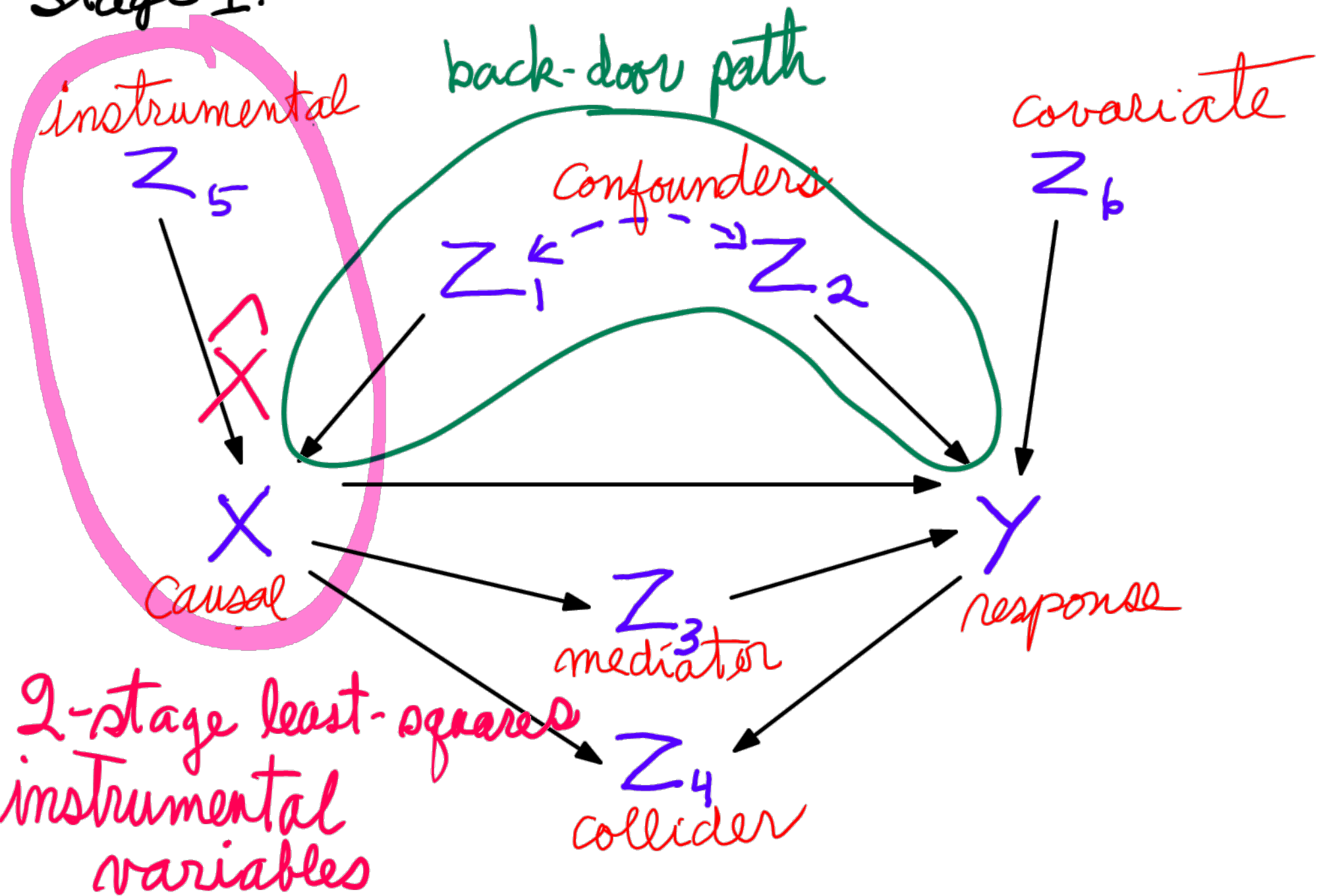


3) Assignment modelling
- propensity scores

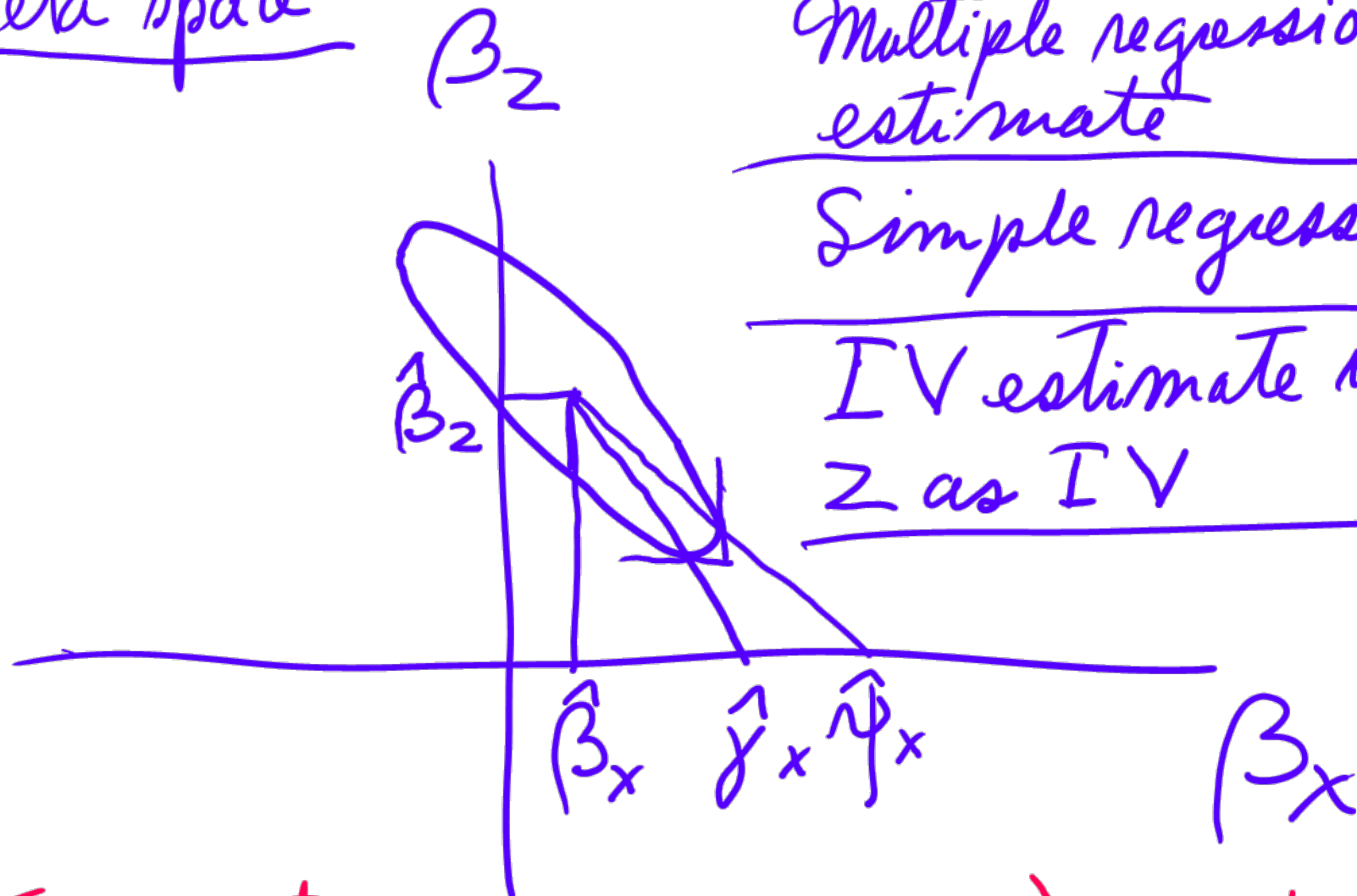
BAD



Stage 1:



Beta space



Multiple regression
estimate

$\hat{\beta}_x$

Simple regression

$\hat{\delta}_x$

IV estimate with
Z as IV

$\hat{\psi}_x$

For a strong IV, $\text{Corr}(X, Z)$ close to 1
and exclusion restriction $\Rightarrow \beta_z$ close to 0

So
 $\hat{\psi}_x$
close to
 $\hat{\delta}_x$

Note that we can't test the assumption of "exclusion restriction" by looking at the coefficient of the instrumental variable I in the regression

$$Y \sim I + X$$

since X is a collider if there is an omitted confounder C



and $\hat{\beta}_I$ should not be 0 even if I is a good instrument.

Lord's Paradox (Wainer version)

Y_2
post

2 Cafeterias

I - normal

II - weight
loss

Y_1 pre

y_2
post

pre = post

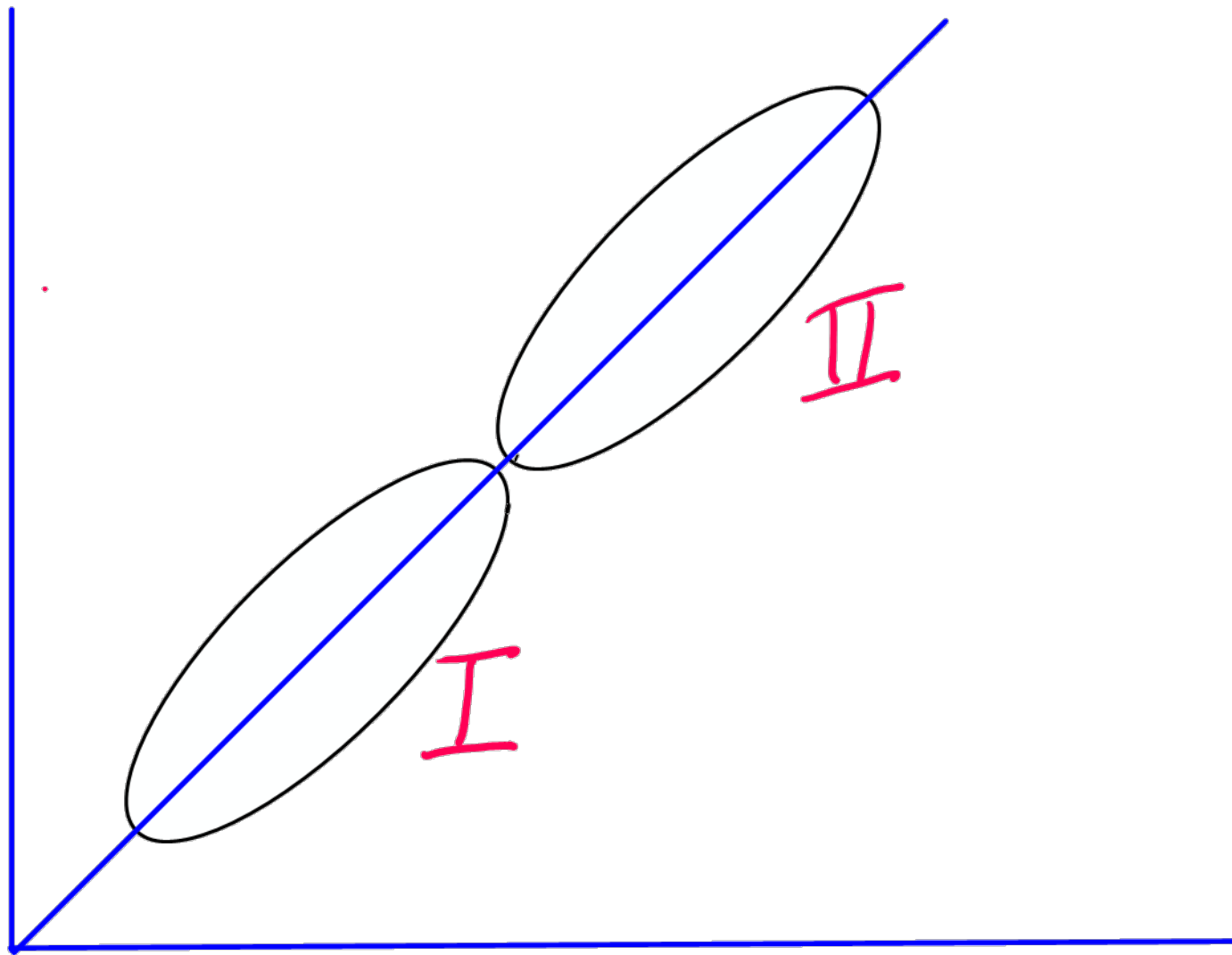
2 Cafeterias

I - normal

II - weight
loss

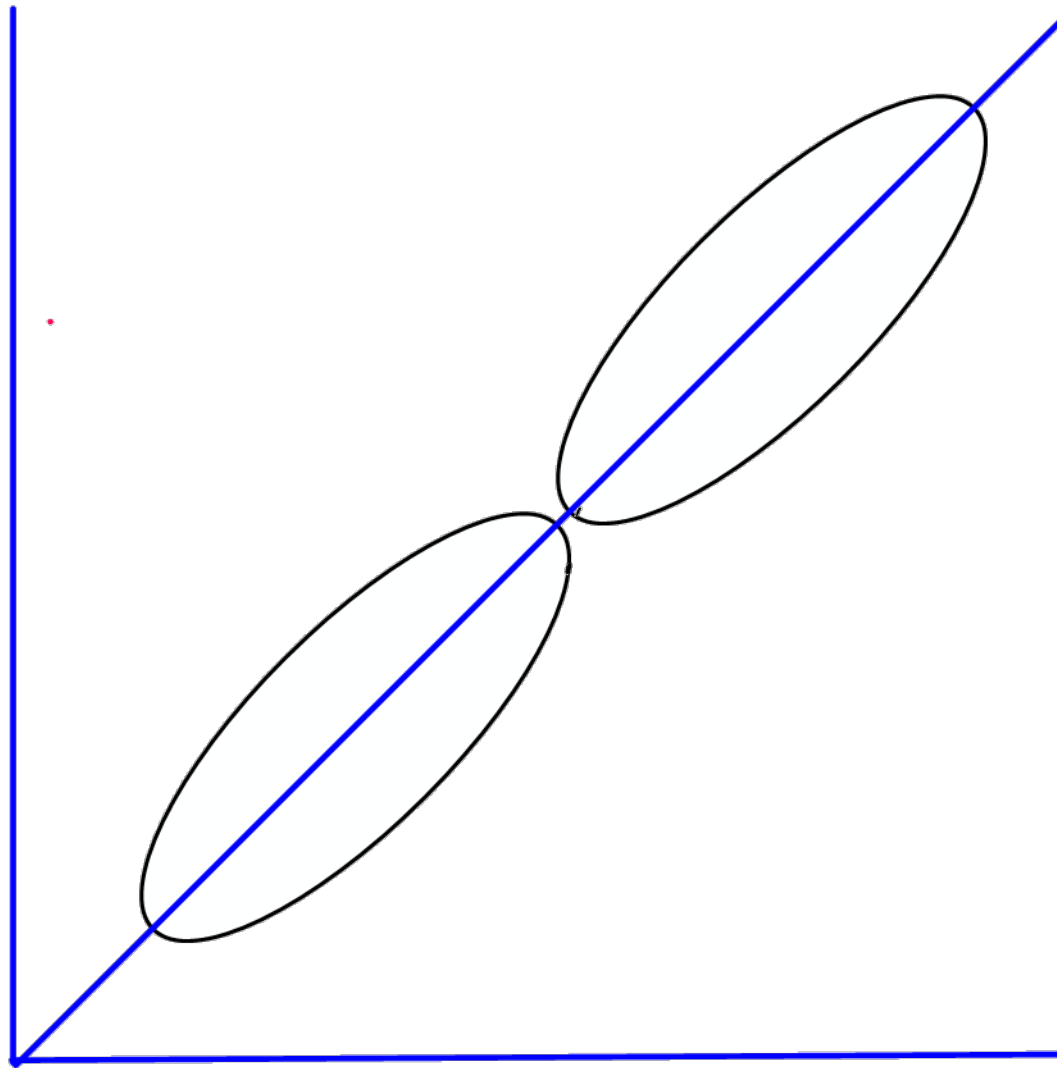
y_1 pre

y_2
post



y_1 pre

Y_2
post

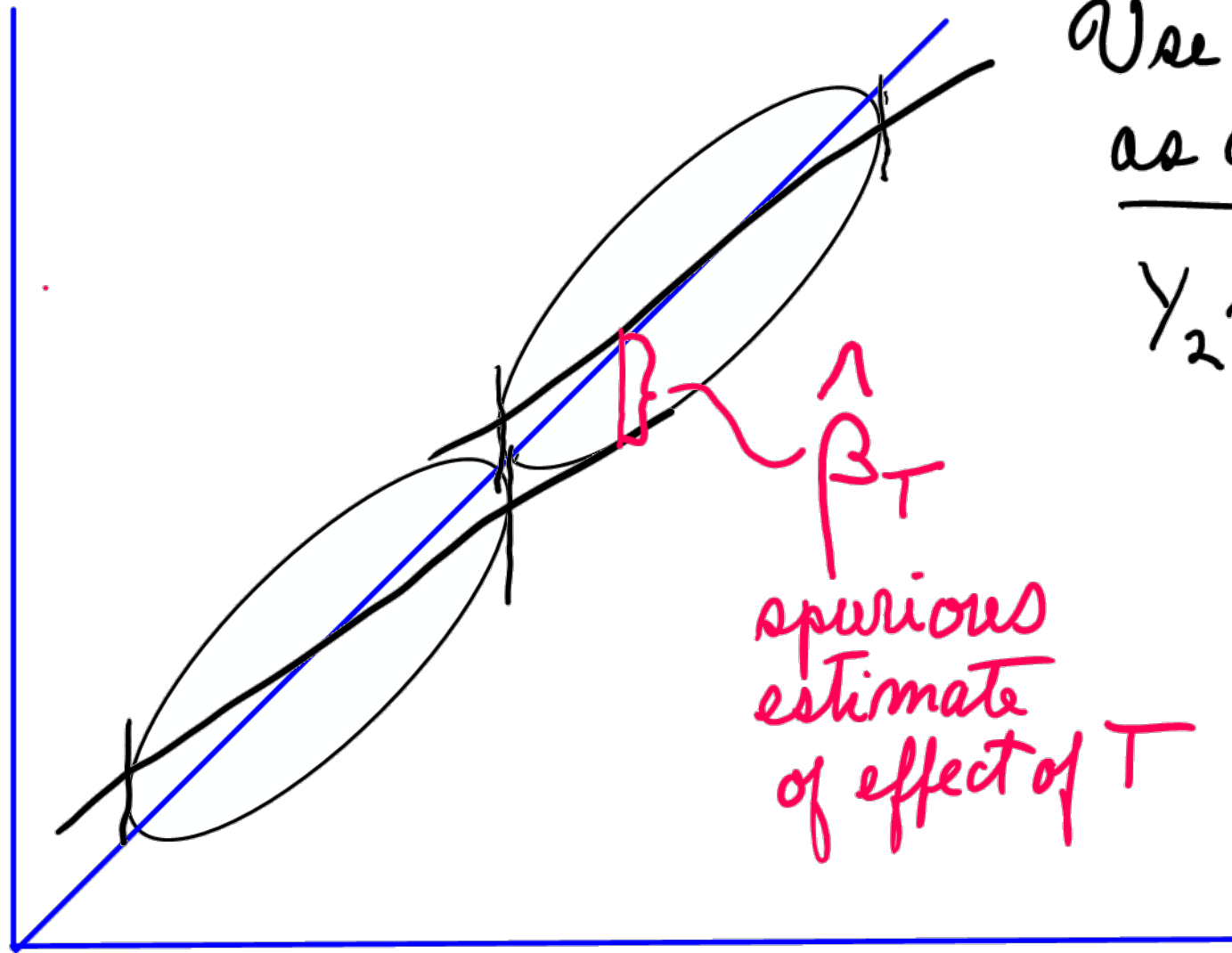


Use pretest
as covariate

$$Y_2 \sim T + Y_1$$

Y_1 pre

Y_2
post



Use pretest
as covariate

$$Y_2 \sim T + Y_1$$

$\hat{\beta}_T$
spurious
estimate
of effect of T

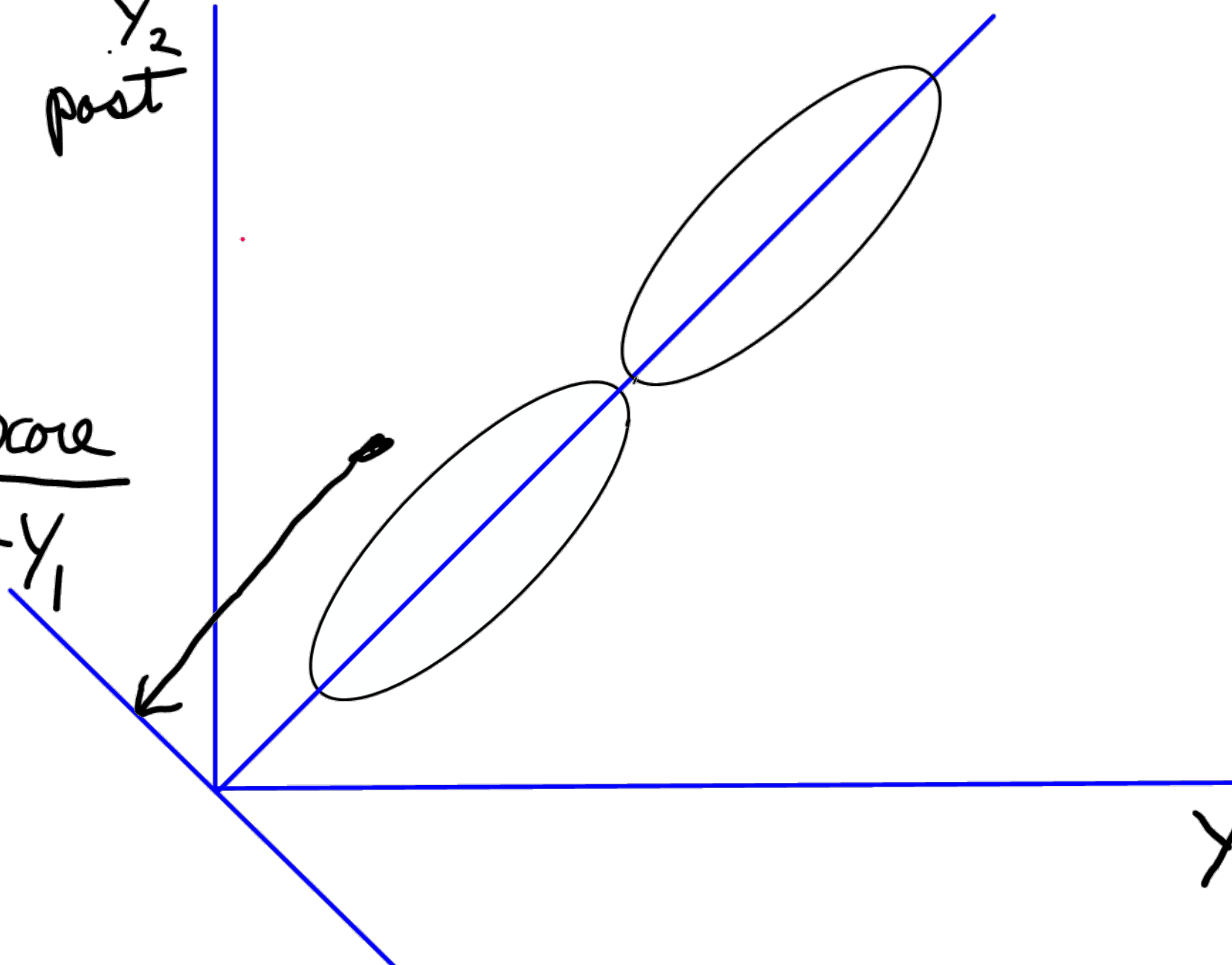
Y_1 pre

y_2
post

gain score

$$G = y_2 - y_1$$

y_1 pre



y_2
post

gain score

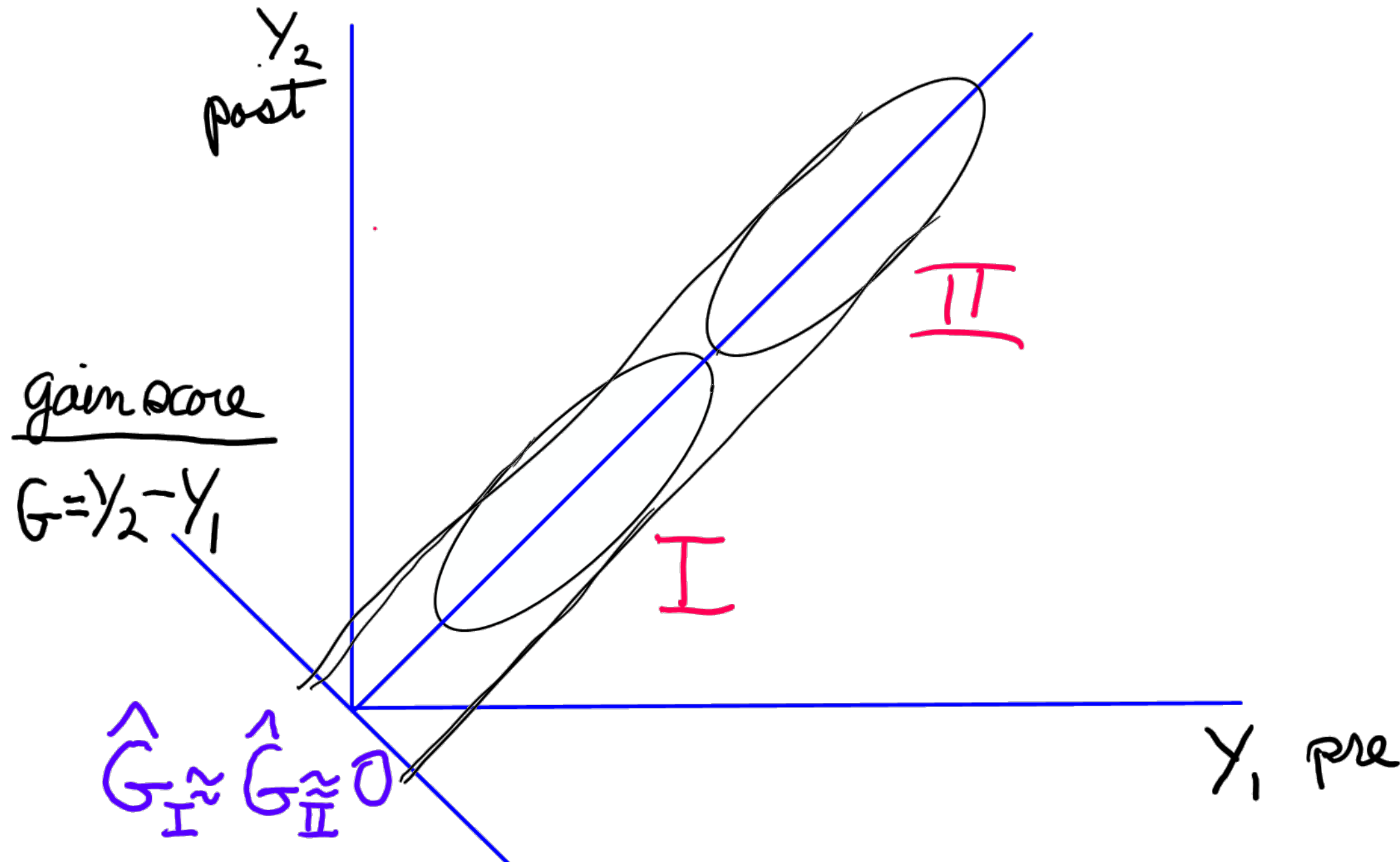
$$G = y_2 - y_1$$

II

I

$$\hat{G}_I \approx \hat{G}_{II} \approx 0$$

y_1 pre



y_2
post

gain score

$$G = y_2 - y_1$$

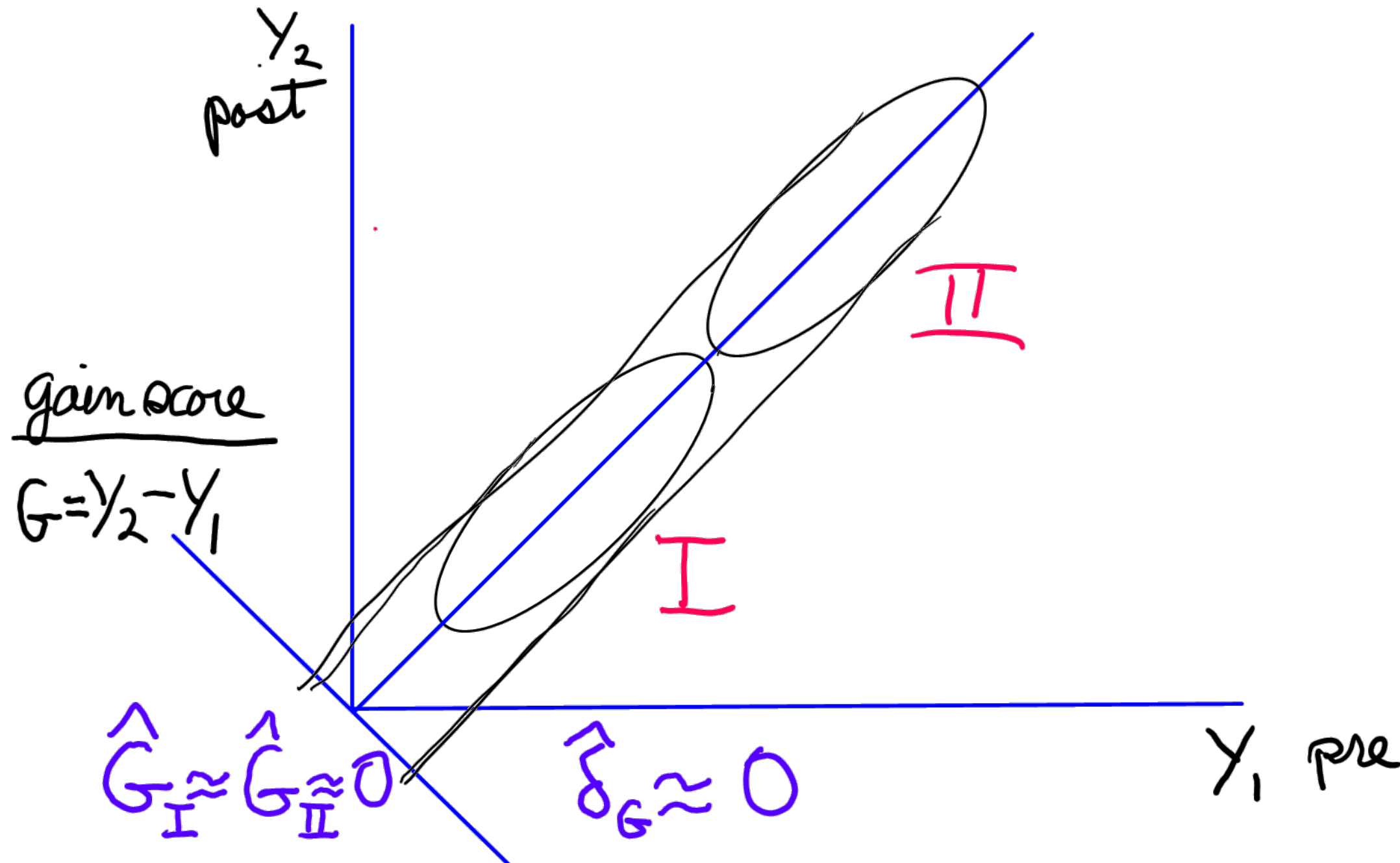
II

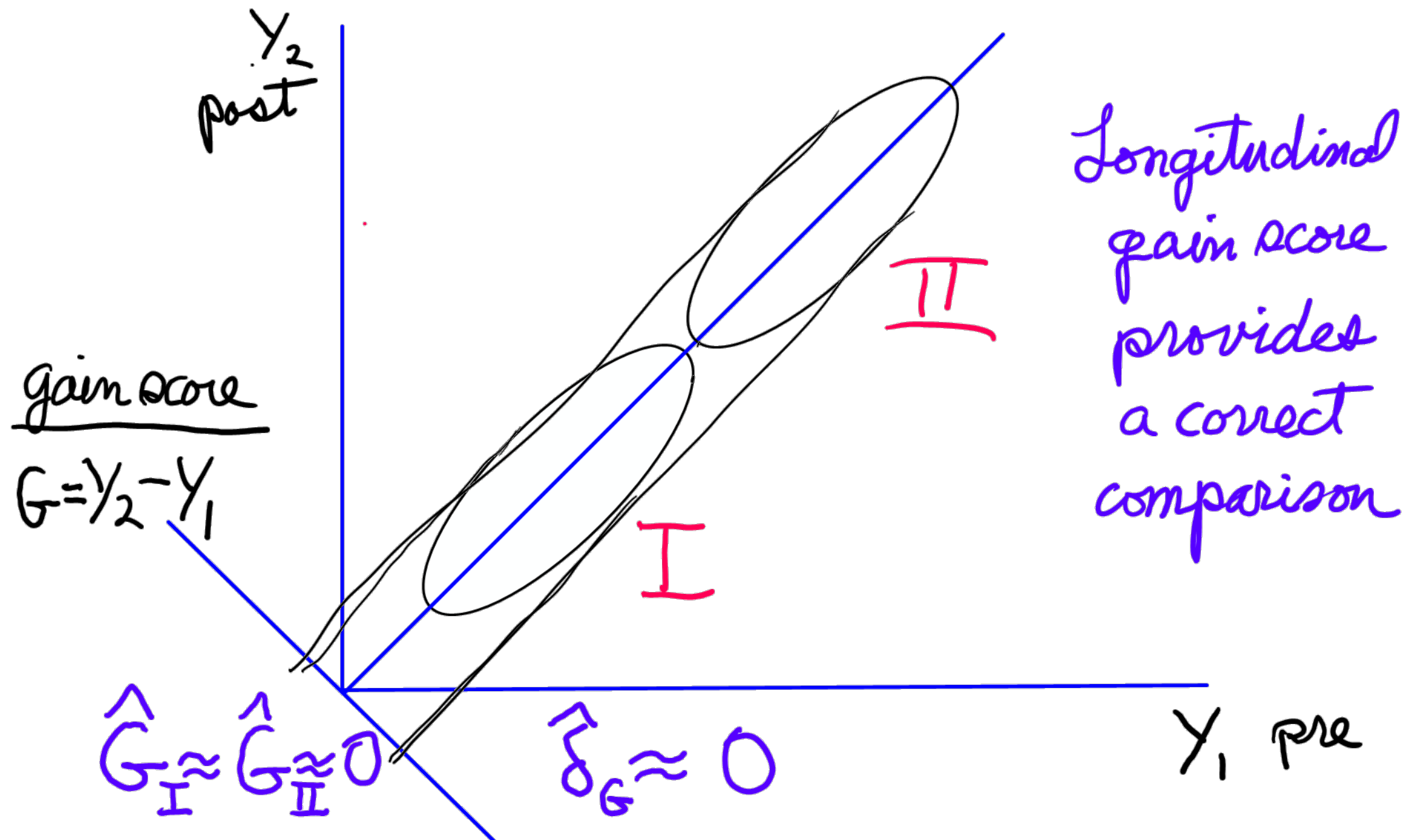
I

$$\hat{G}_I \approx \hat{G}_{II} \approx 0$$

$$\bar{\delta}_G \approx 0$$

y_1 pre





Conditions

- Same scale for Y_{pre} & Y_{post}
- No time-varying confounders

Within-subject effect adjusts for between-subject confounders whether measured or not.

Good model? $Y \sim X + Z_i + Z_j$

Want:

1) Unbiased - consistent

Block back doors - NOT mediators & colliders

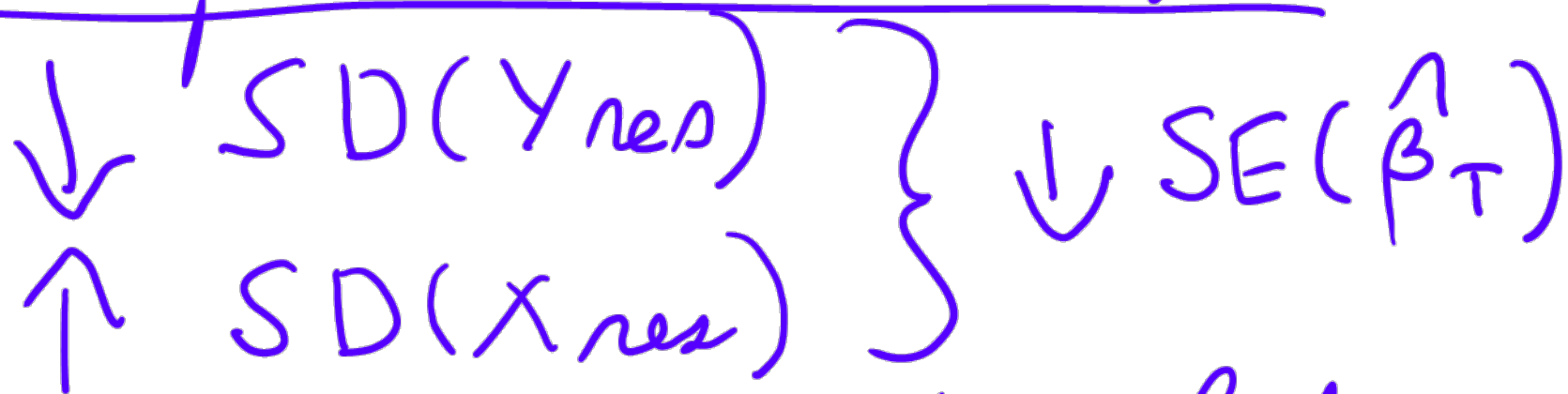
2) Low SE = $SD(Y_{res}) / SD(X_{res})$
Small $SD(Y_{res})$, Large $SD(X_{res})$

3) Honest SE

4) Robust Propensity scores - focus on X

Use the AVP to compare models.

Using confounders close to Y



But may not have knowledge about structure of model for Y

Using confounders close to X



But may have better understanding of assignment model.

Propensity score methods focus on predicting X with $\hat{X}$

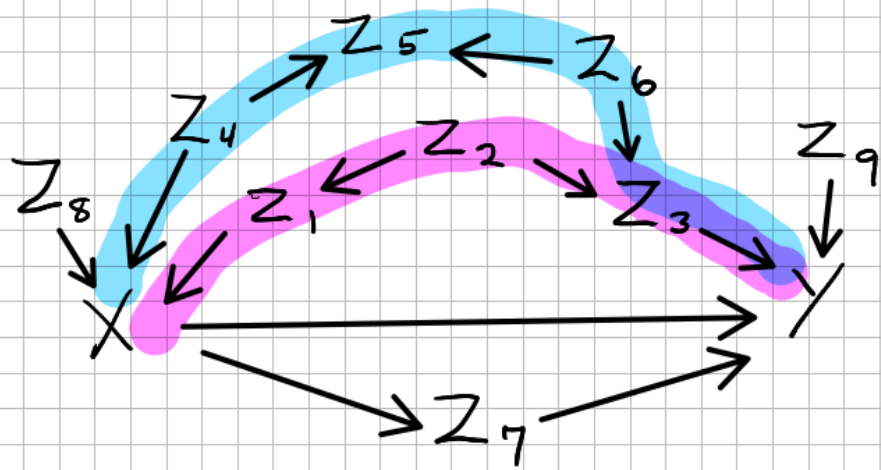
- no need to understand model for Y

- except to avoid mediators & colliders

Then regress Y on X and $\hat{X}$ (often grouped into intervals)

"Doubly robust:" throw in some Z 's close to Y and covariates.

Summary for linear models:



Back door path #1
 " " " #2

- Not including Z_5 blocks #1
- Any of Z_1, Z_2, Z_3 blocks # Z_9

$$SD(\hat{\beta}_x) = ? = \frac{1}{\sqrt{n}} \frac{se}{S_{x|others}}$$

Comparing models, consider impact on se & $S_{x|others}$

Will $Y \sim X + X_i + X_j$
 estimate the causal effect of X ?

2 requirements that are sufficient

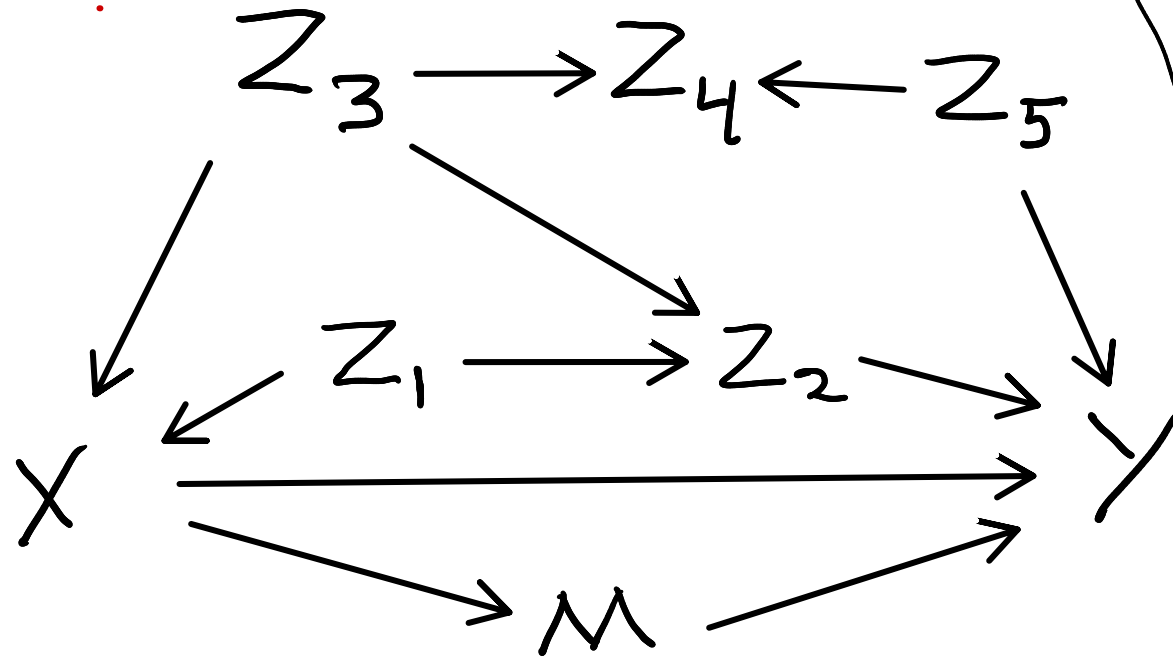
1) Block back-door paths
 How?

- a) Presence of a collider NOT
 in the model
 OR
 b) Including one or more
 non-colliders

2) Do not include descendants of X

When is $\hat{\beta}_X$ an estimate of the causal effect of X

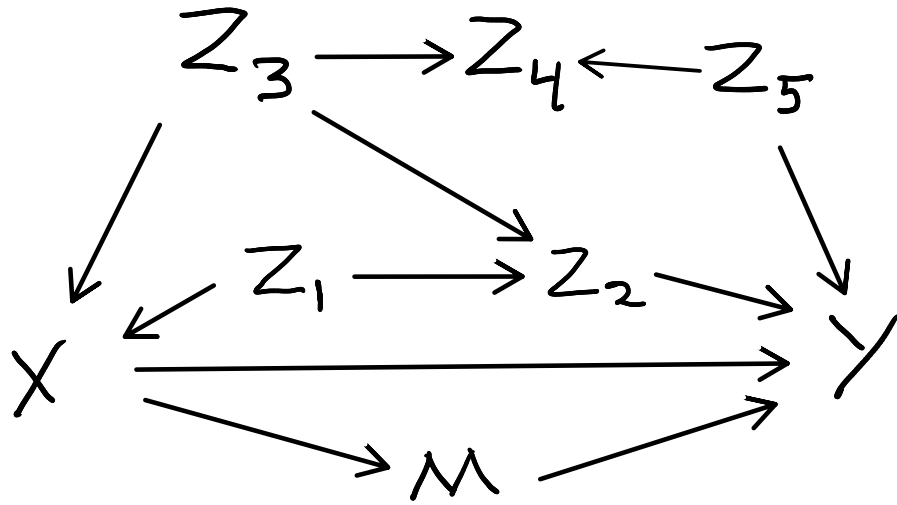
Backdoor criterion Judea Pearl
given a linear* causal graph



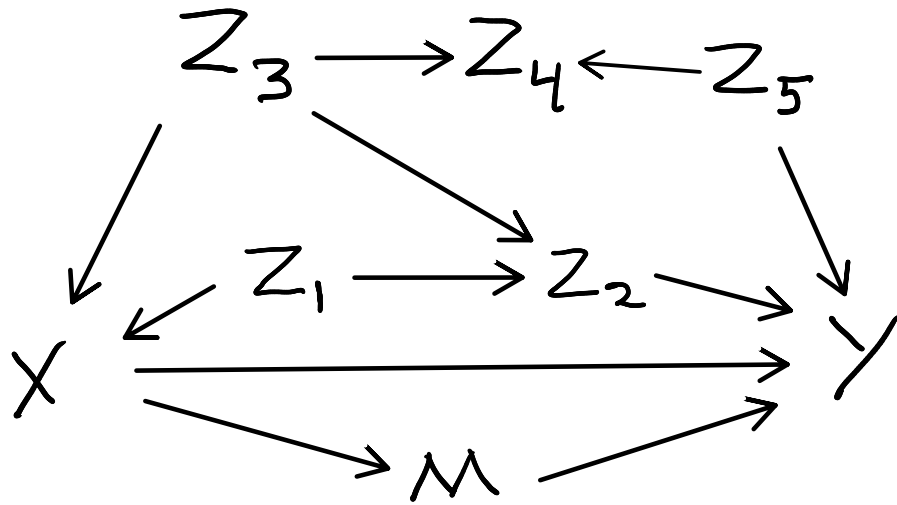
* A cyclic graph for "true" relationships that are linear in predictors and β 's

What model(s) would give us an estimate of the causal effect of X ?

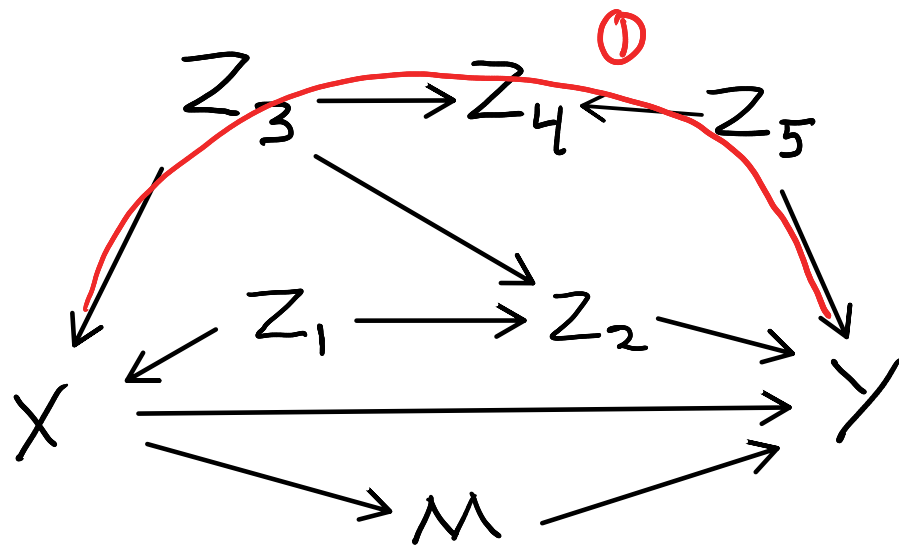
Step 1) Find all "backdoor" paths
i.e. path between X and Y
that has an arrow
pointing to X and an arrow
pointing to Y without
retracing.



Step 1) Find all "backdoor" paths
i.e. path between X and Y
that has an arrow
pointing to X and an arrow
pointing to Y without
retracing.

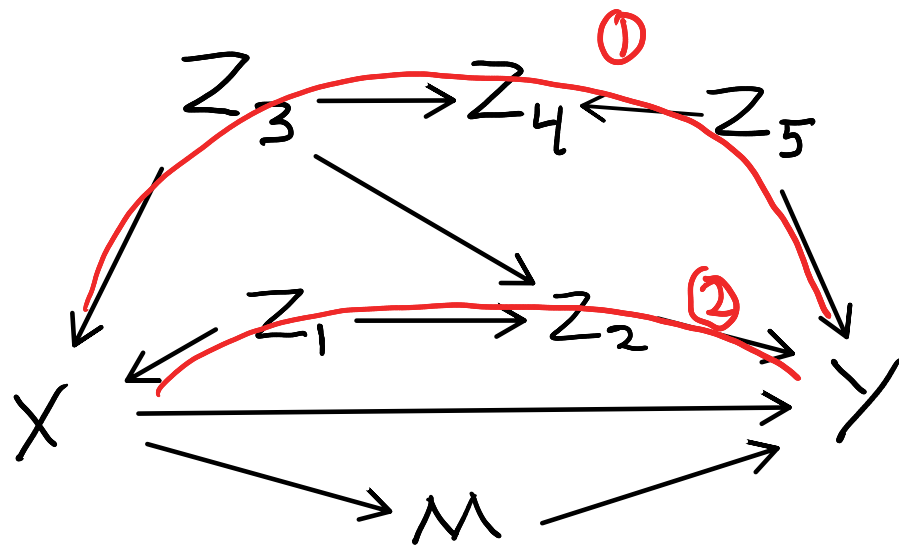


Here we have 3 paths



Step 1) Find all "backdoor" paths
i.e. path between X and Y
that has an arrow
pointing to X and an arrow
pointing to Y without
retracing.

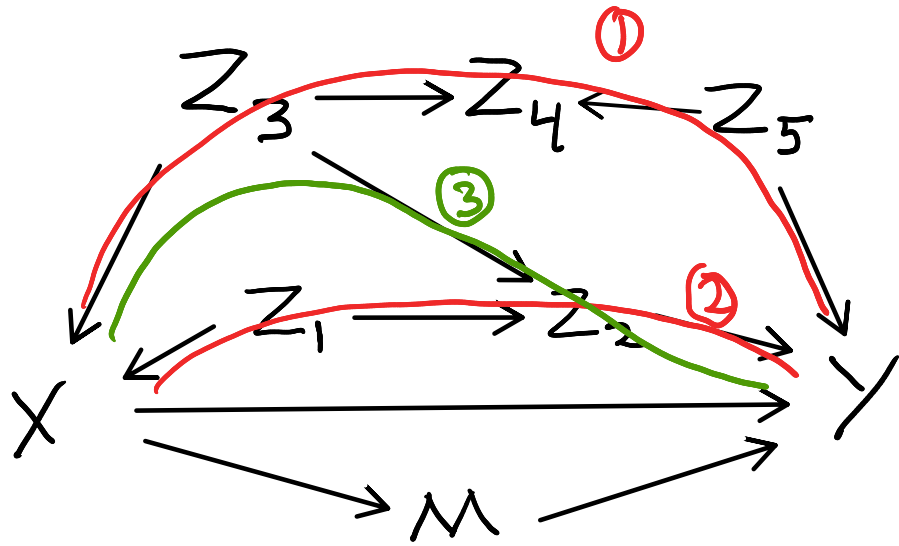
Here we have 3 paths



Step 1) Find all "backdoor" paths
i.e. path between X and Y
that has an arrow
pointing to X and an arrow
pointing to Y without
retracing.

Here we have 3 paths

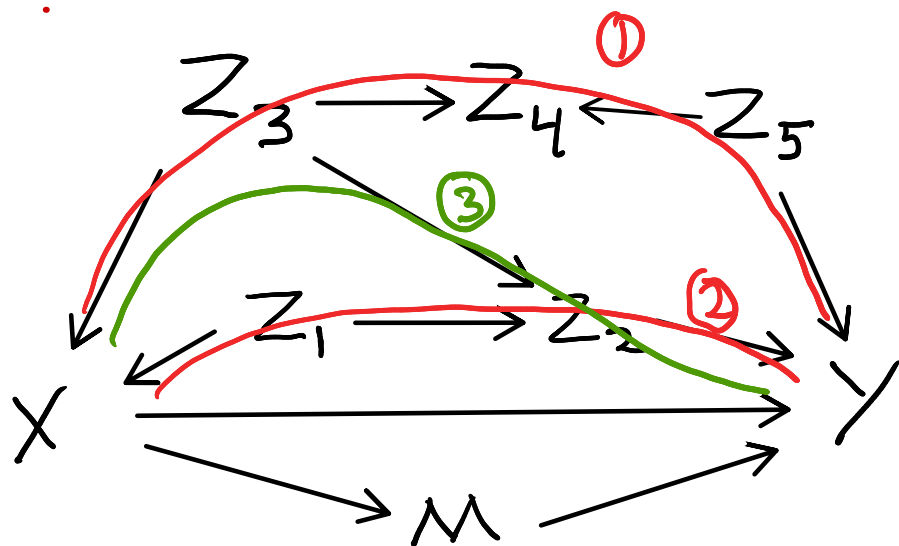
Step 1) Find all "backdoor" paths
i.e. path between X and Y
that has an arrow pointing to X and an arrow
pointing to Y without
retracing.



Here we have 3 paths

Step 2) Make sure all backdoor paths are closed "blocked"
How? a) presence of an uncontrolled collider
or b) one or more non-colliders that are controlled

If a collider is controlled then the path is open
and can be closed by controlling another non-collider

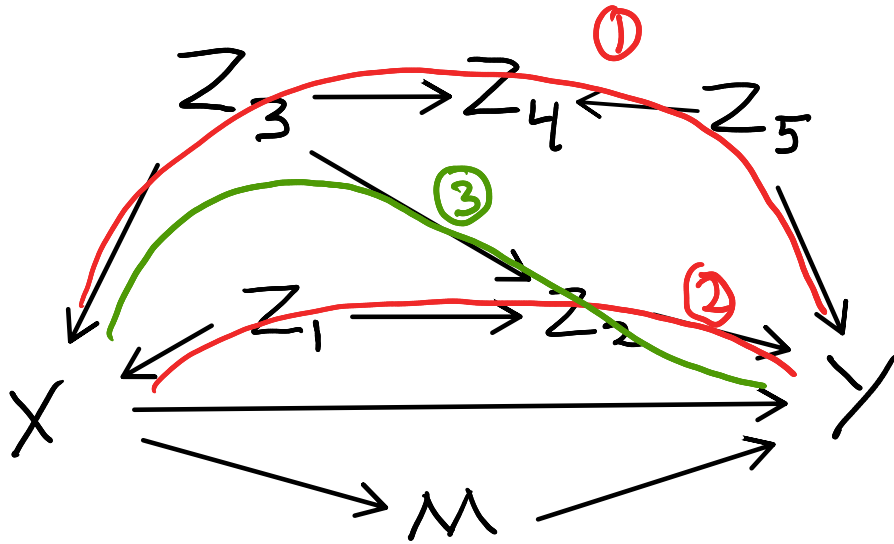


Checking the backdoor paths

①

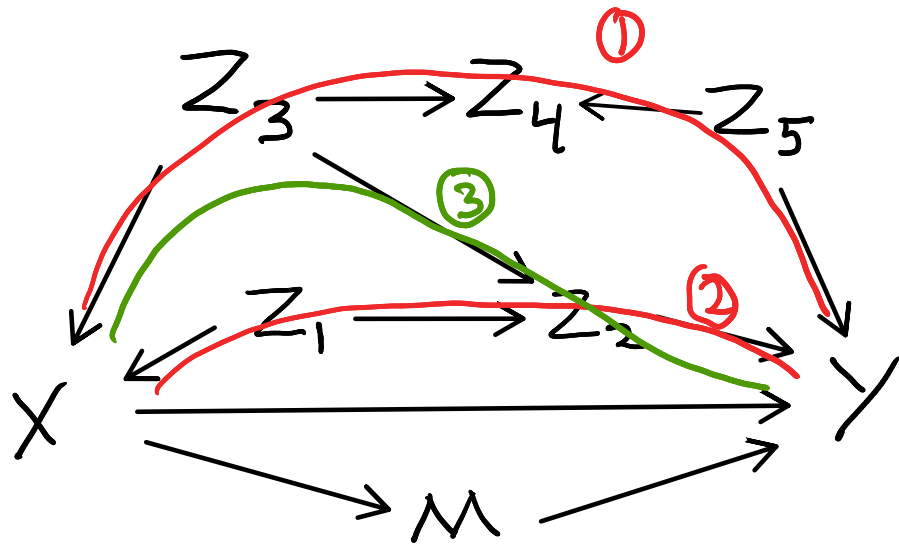
②

③



Checking the backdoor paths

- ① Closed by the collider Z_4
- ② Could be closed by controlling for Z_1 or Z_2
- ③ Could be closed by controlling for Z_1 or Z_2



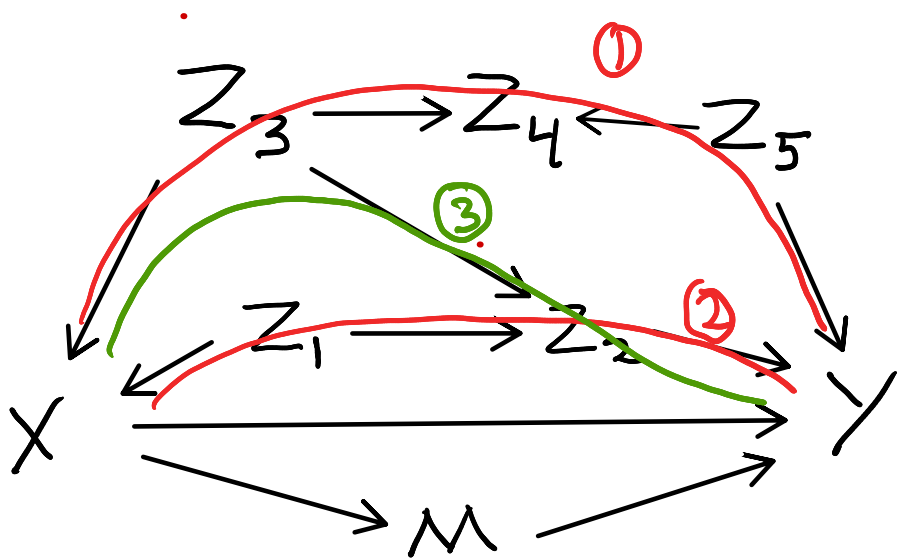
Checking the backdoor paths

- ① Closed by the collider Z_4
- ② Could be closed by controlling for Z_1 or Z_2
- ③ Could be closed by controlling for Z_1 or Z_2

Backdoor criterion

a model provides an estimate of the causal effect of X if (not iff)

- 1) all backdoor paths are closed
- and 2) No descendant of X is controlled.



Which models are causal?

1) $Y \sim X + Z_1 + Z_3$

2) $Y \sim X + Z_2$

3) $Y \sim X + Z_2 + Z_4$

4) $Y \sim X + Z_2 + Z_4 + Z_3$

5) $Y \sim X + Z_2 + Z_5$

Note: "Controlling" a variable can happen in two main ways:

- 1) By including it in a regression model with the correct functional form.
- 2) By selection, that is, the way in which the data were obtained. If the data has only one level of a multilevel character variable, then that variable is "controlled"

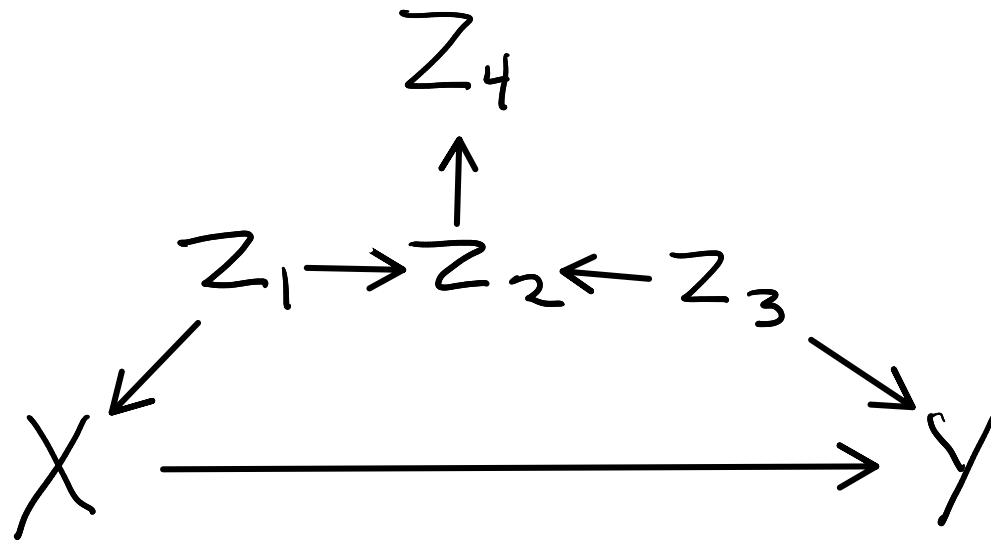
So it's vital to know how the data were generated:

- Was there random assignment^{*}
- Was there deliberate or implicit selection from a target population, which could create selection on a collider.

Non-statistician will often say 'yes' because they can't know how sacred the word "random" is to a statistician. It means^{*} *pristine* random, e.g. with a coin or a random number generator not just "haphazard" or disorganized.

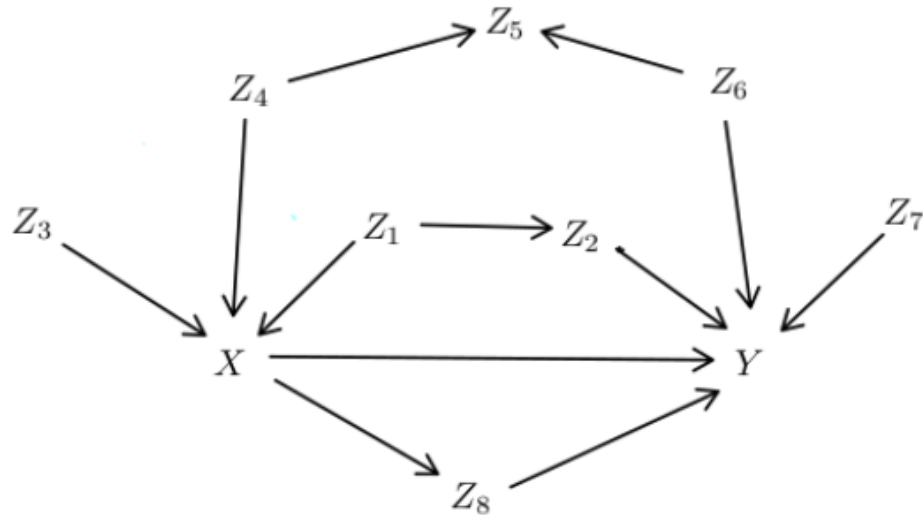
Caution re colliders : Colliders must not be even "partially" controlled, for example by controlling a descendant of the collider.

e.g. in



controlling for Z_4 would open the backdoor path:

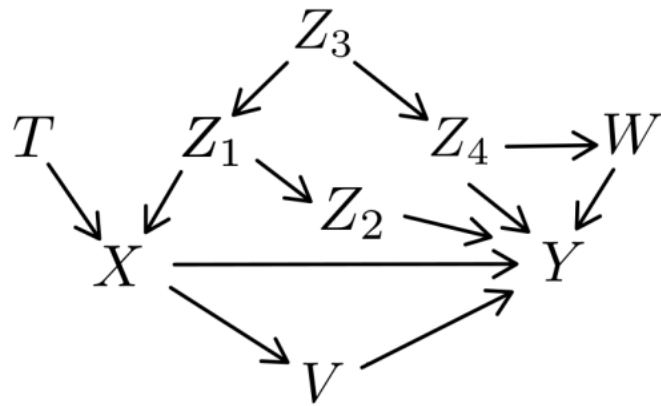
Question 12: (20 marks in 2 parts)



Consider the linear causal DAG above and the following models:

1. $Y \sim X$
2. $Y \sim X + Z1 + Z5$
3. $Y \sim X + Z1$
4. $Y \sim X + Z8$
5. $Y \sim X + Z2$
6. $Y \sim X + Z2 + Z6$
7. $Y \sim X + Z2 + Z8$
8. $Y \sim X + Z2 + Z7$
9. $Y \sim X + Z1 + Z5 + Z6$

- a) [10 marks] For each of these models discuss briefly **whether and why** fitting the model would produce, or not, an unbiased estimate of the causal effect of X .
- b) [10 marks] Choose any two models that produce unbiased estimates of the causal effect of X and discuss in detail whether one can be determined to have a lower standard error for its estimate of β_X and why.



a) [10 marks] Consider the linear causal DAG above and the following models. For each model state briefly whether and why the model would, or would not, produce an unbiased estimate of the causal effect of X .

1. $Y \sim X$

2. $Y \sim X + Z_1 + W$

3. $Y \sim X + Z_1$

4. $Y \sim X + Z_2 + W$

5. $Y \sim X + Z_1 + T$

6. $Y \sim X + Z_2 + Z_4$

7. $Y \sim X + Z_3 + Z_3$

8. $Y \sim X + Z_1 + V$

9. $Y \sim X + Z_2 + W$