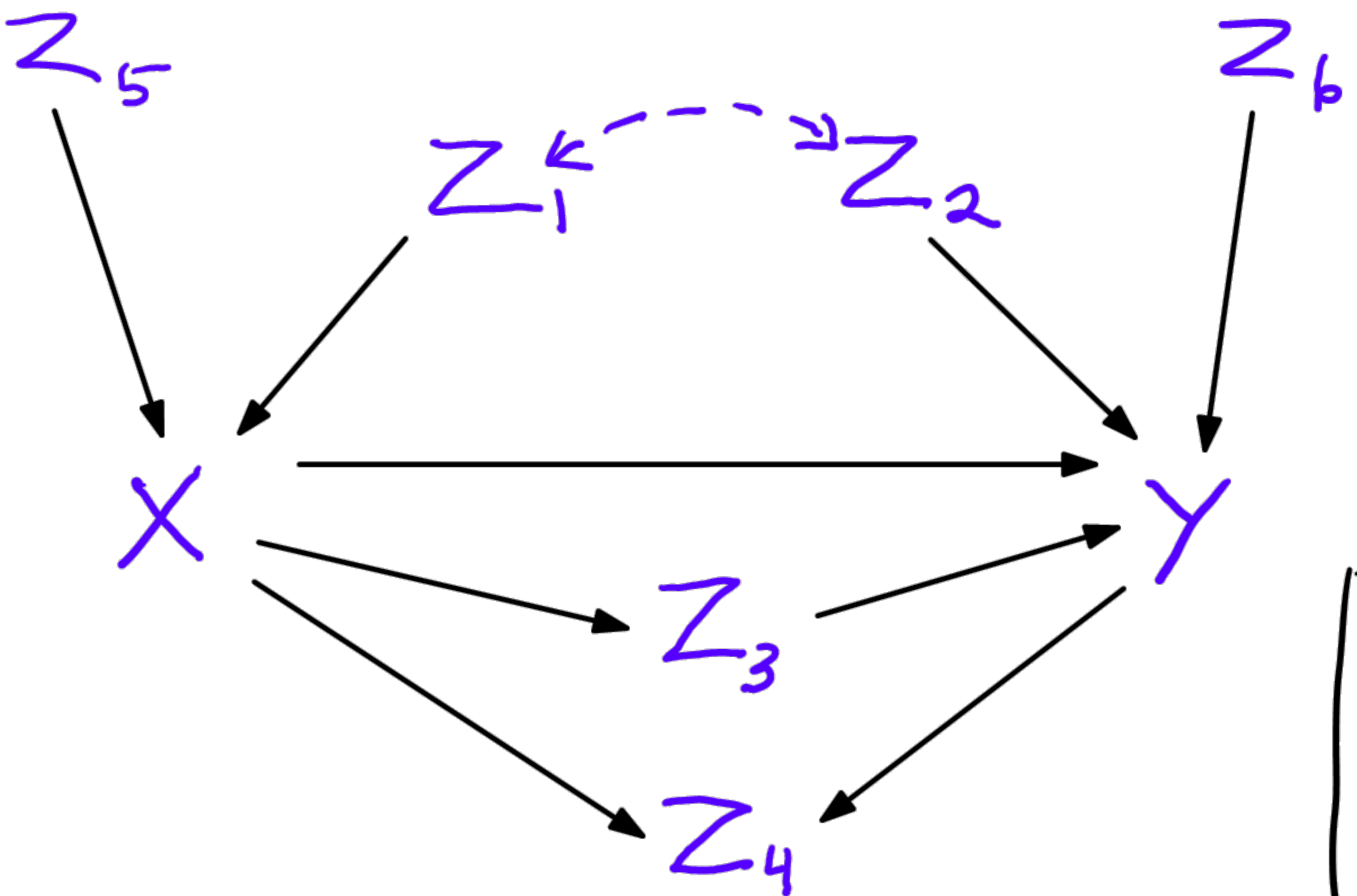
Does X cause Y?

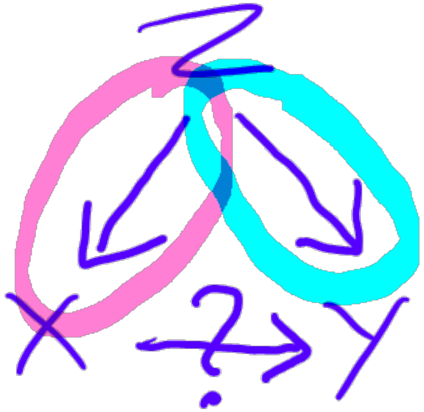
- A bird's eye view of methods with observational data
- Lord's Paradox and the role of longitudinal data

Causal Graph

Pearl & Mackenzie (2019)

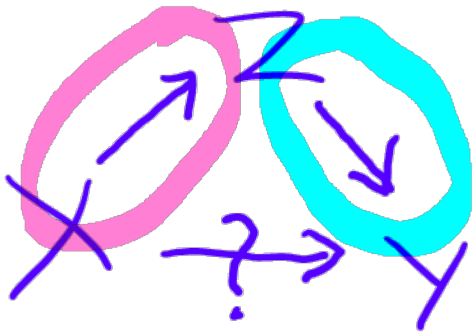


DAG = Directed acyclic graph



confounder

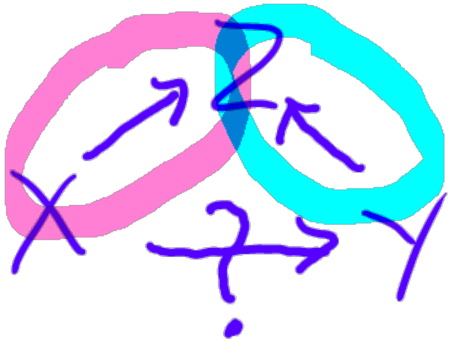
must include to see
see causal effect
of X on Y



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

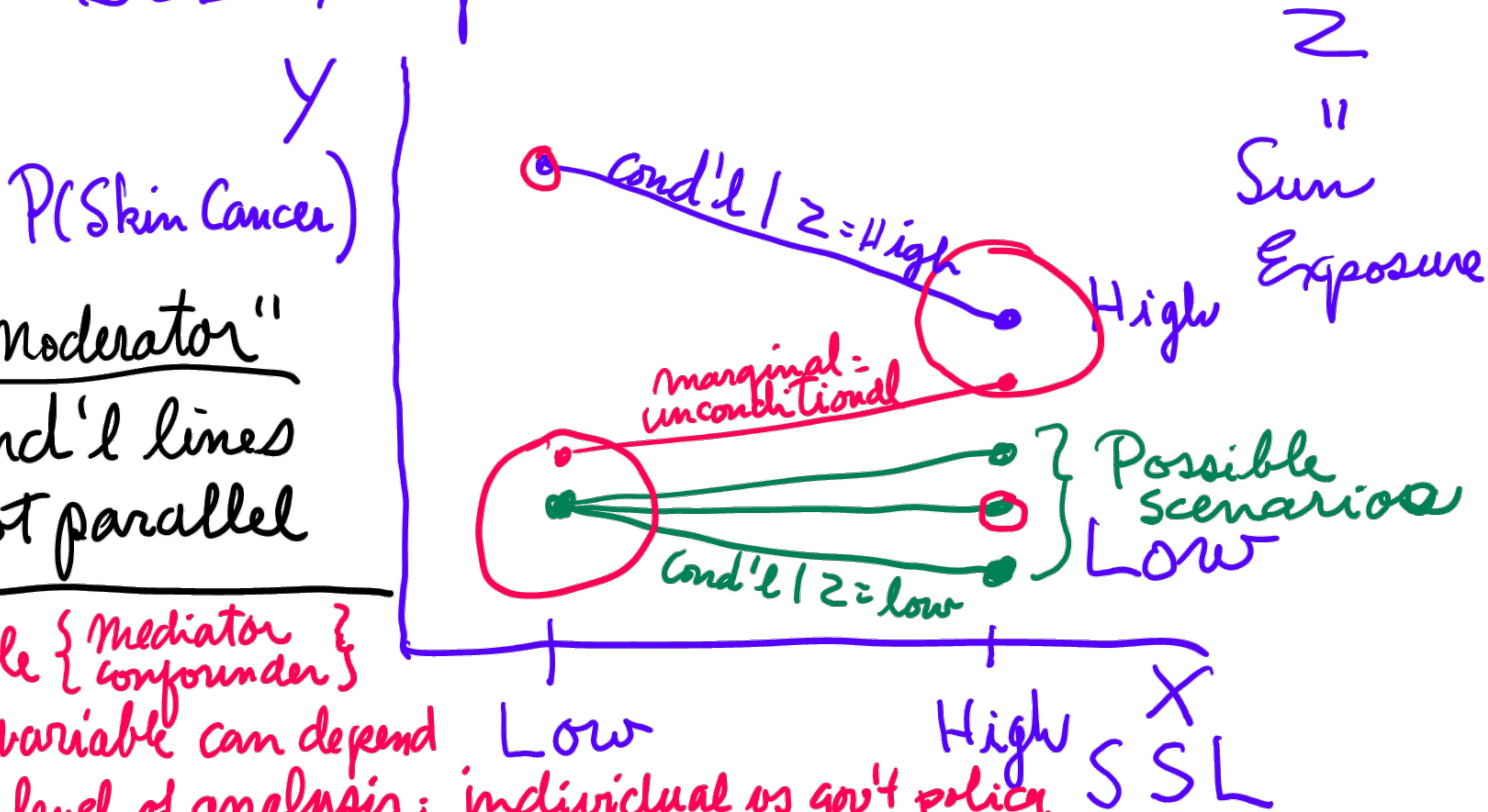
Also mediator-colliders, etc
BUT not represented by DAGs

- 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

"Moderator"
Cond'l lines not parallel



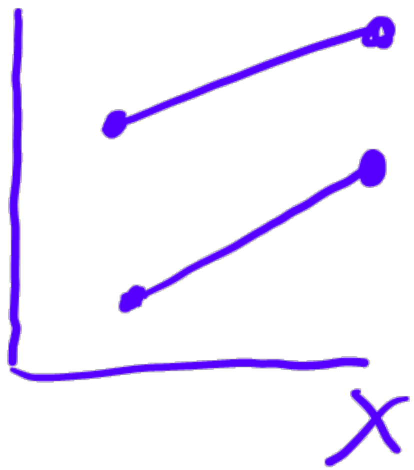
Role {mediator / confounder} of variable can depend on level of analysis: individual vs gov't policy

Moderation may be removable by monotonically transforming Y if

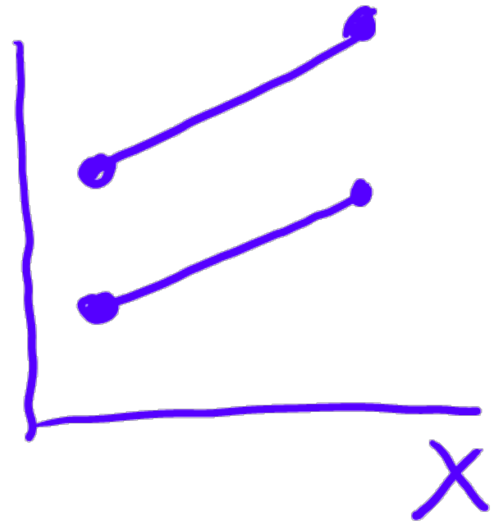
1) Cond'l effects in same direction
2) No crossings of lines

Removable

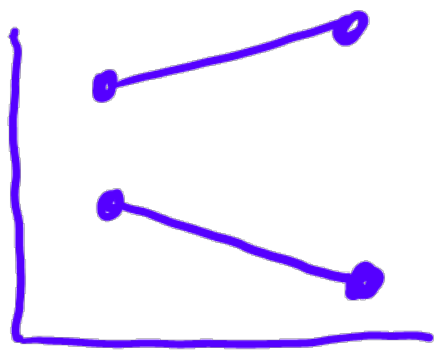
P(Y)

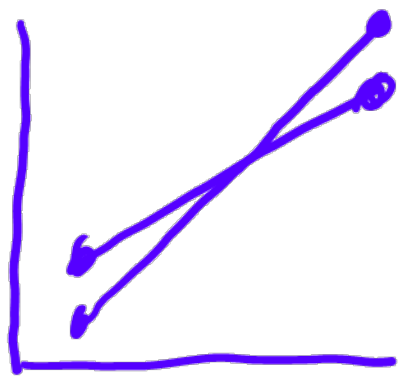


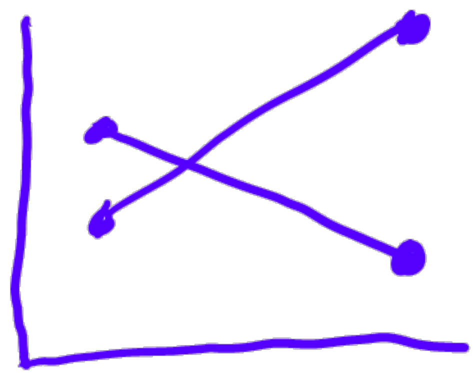
log odds = $\log\left(\frac{P(Y)}{1-P(Y)}\right)$



Nonremovable



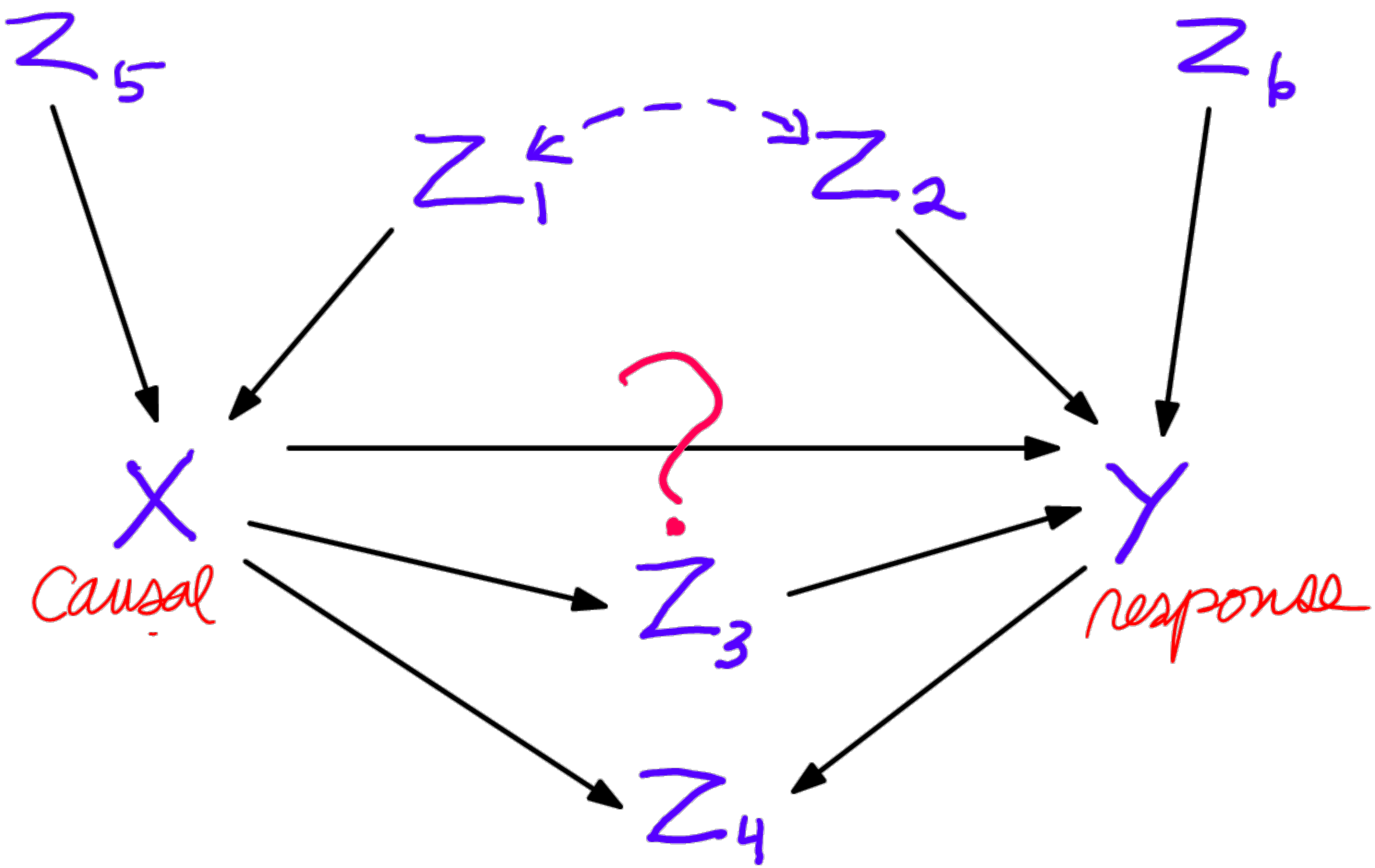




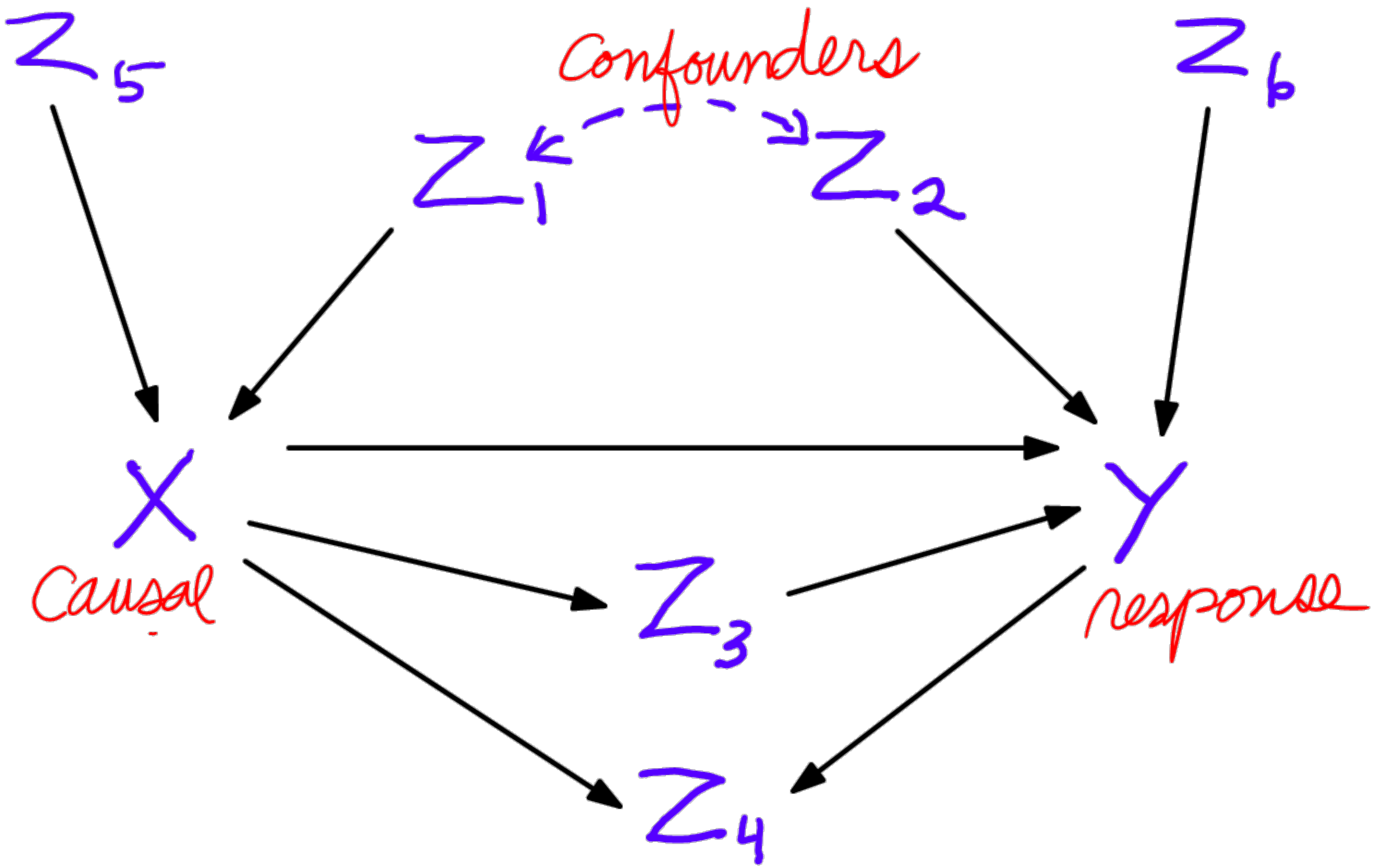
Note:

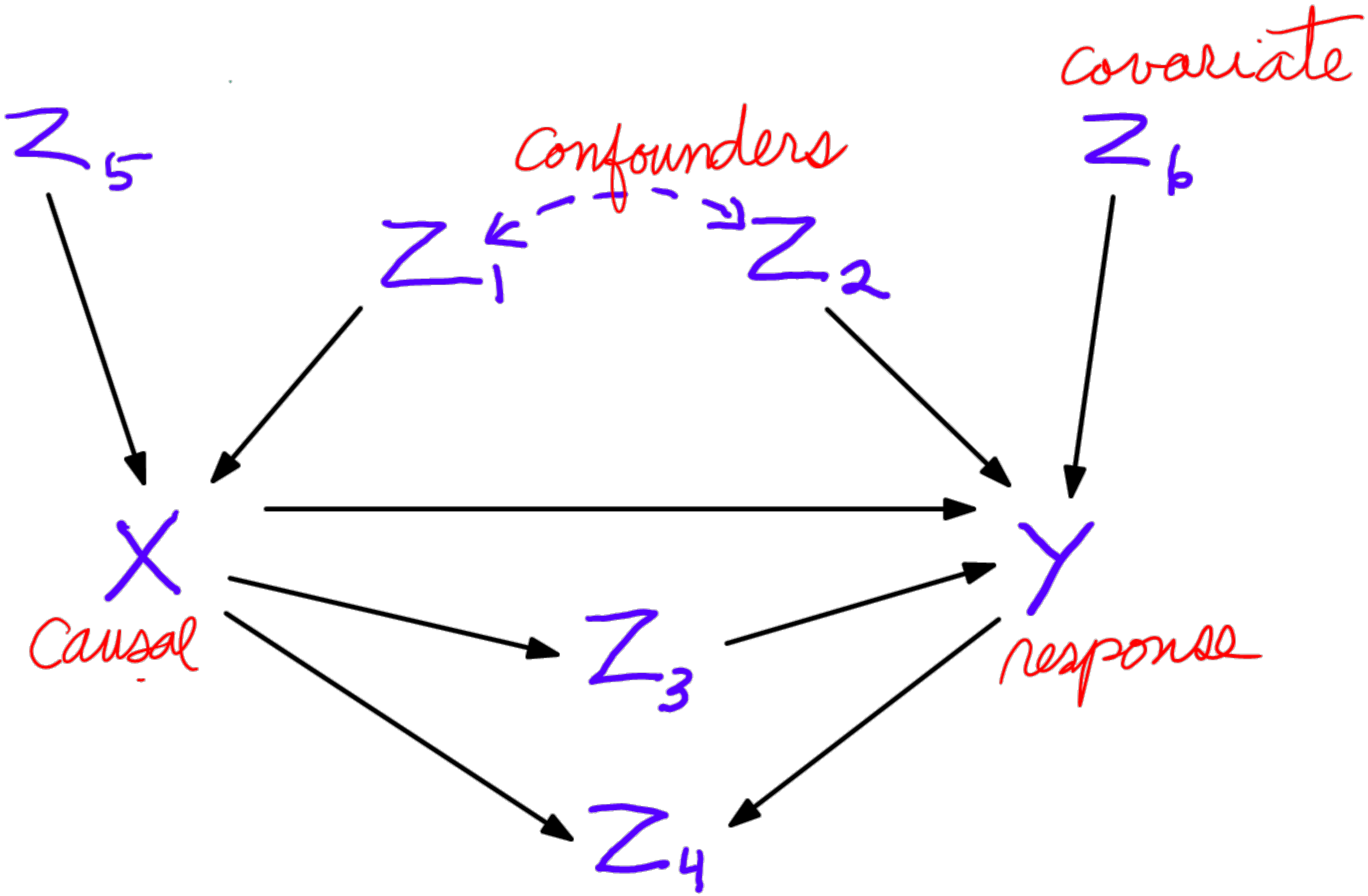
- Simpson effect: reversal of cond'l vs. marginal effects
- Moderation (= interaction in relation of X and Z with Y)
- Association (predictive)
- Causality

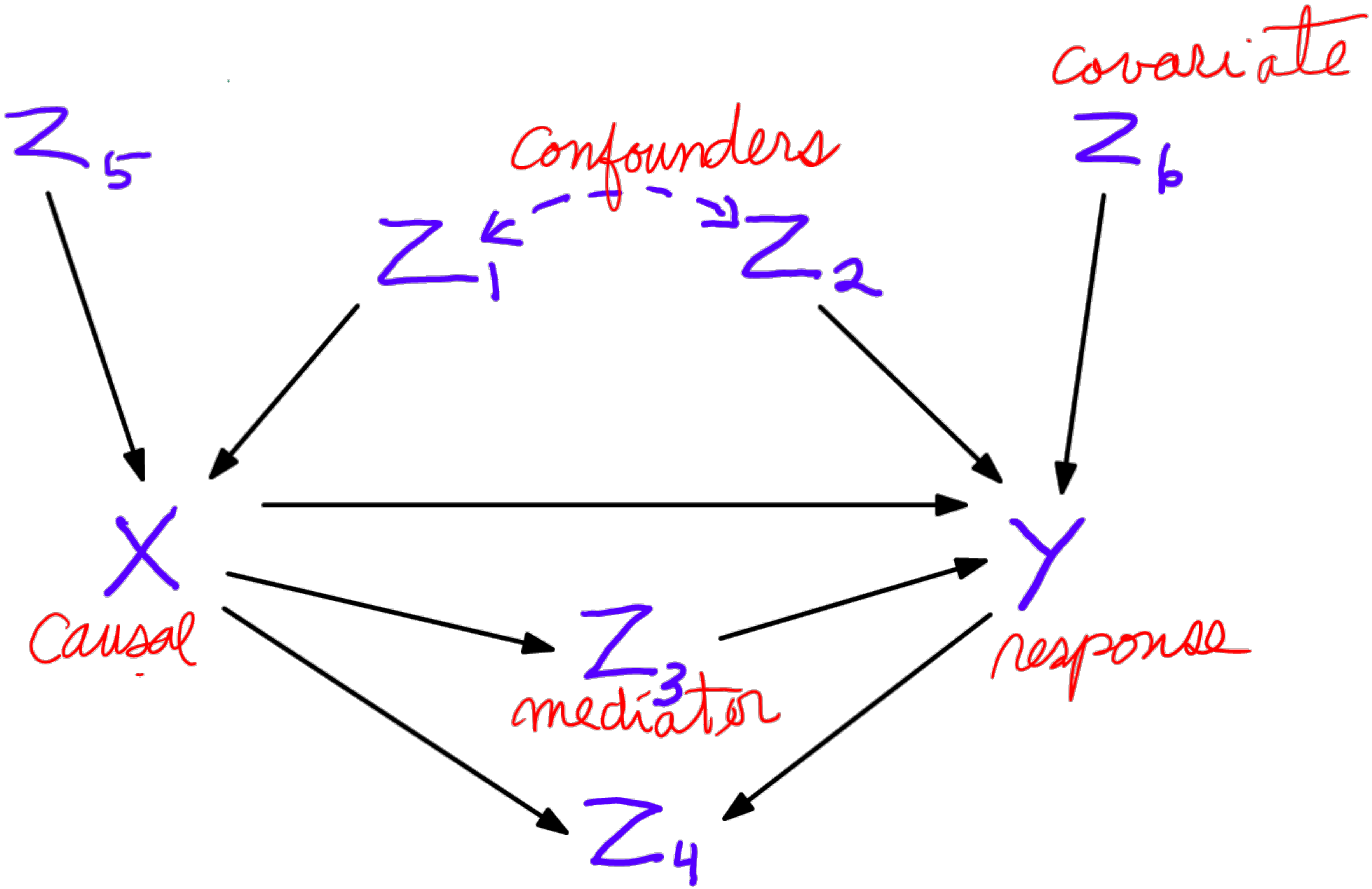
are distinct but not unrelated concepts

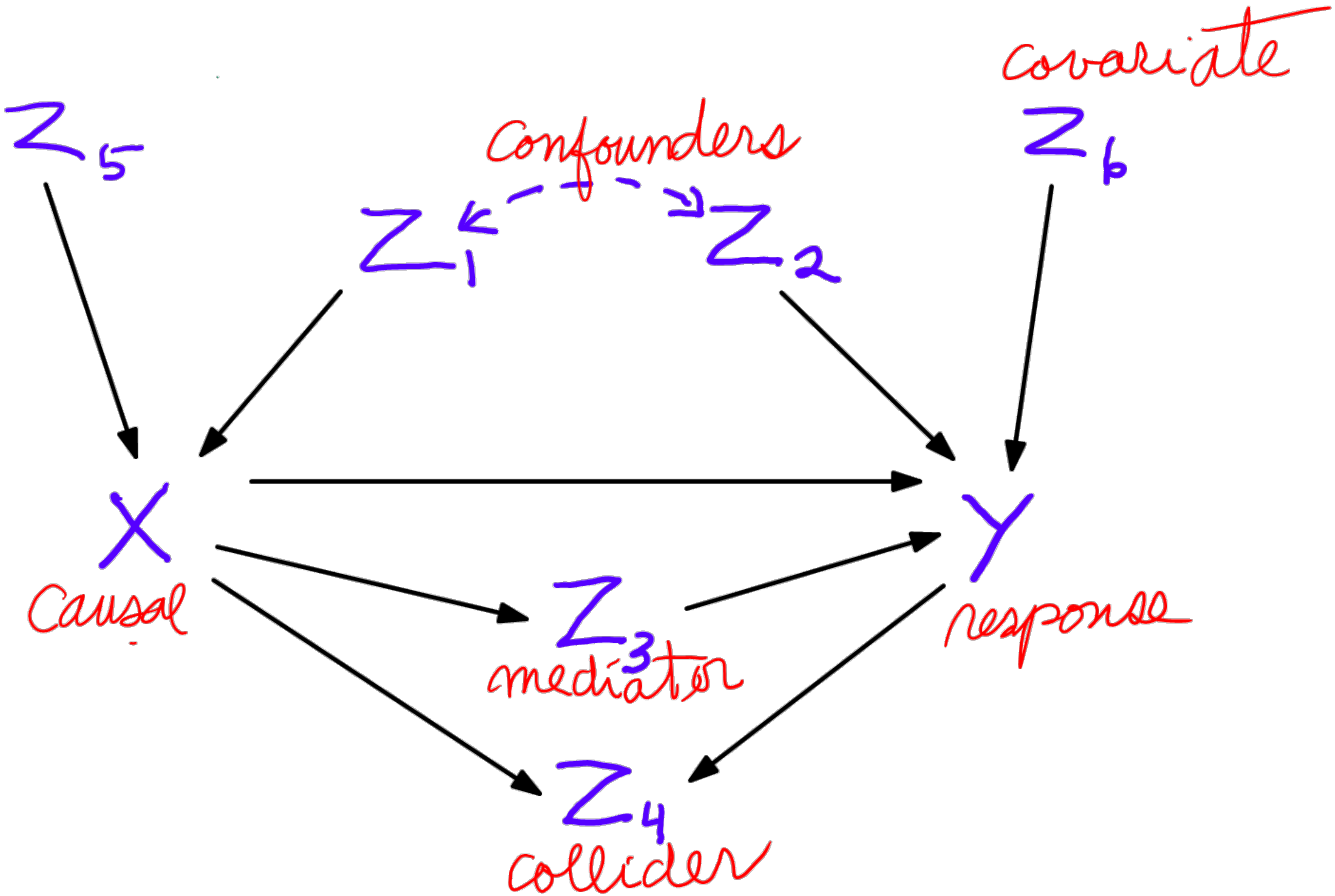


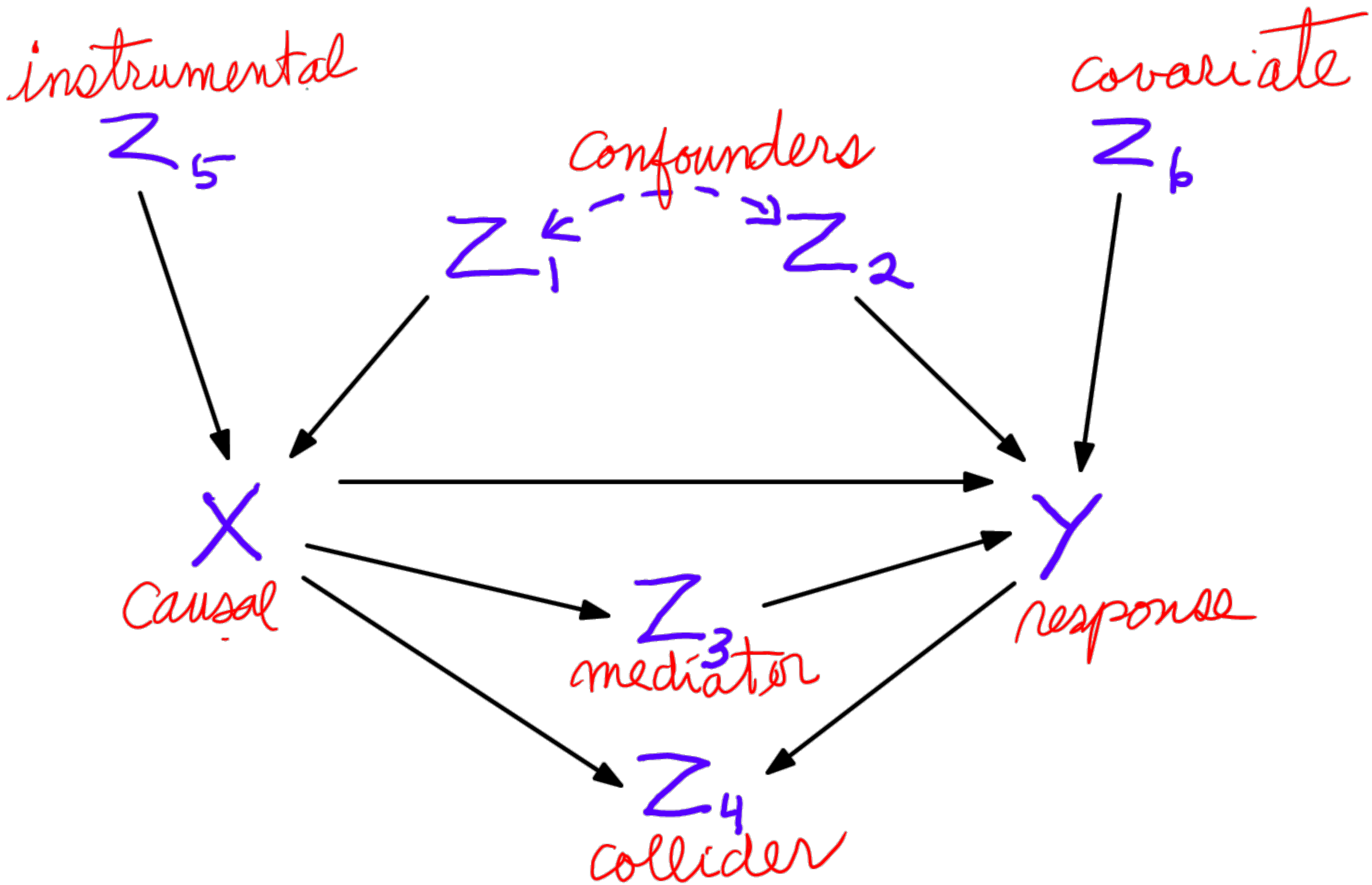
Z_5
Z_6
Z_1
Z_2
?
X
Y
Z_3
Z_4
causal
response


Z5
Confounders
Z6
Z1
Z2
X
Y
Causal
response
Z3
Z4

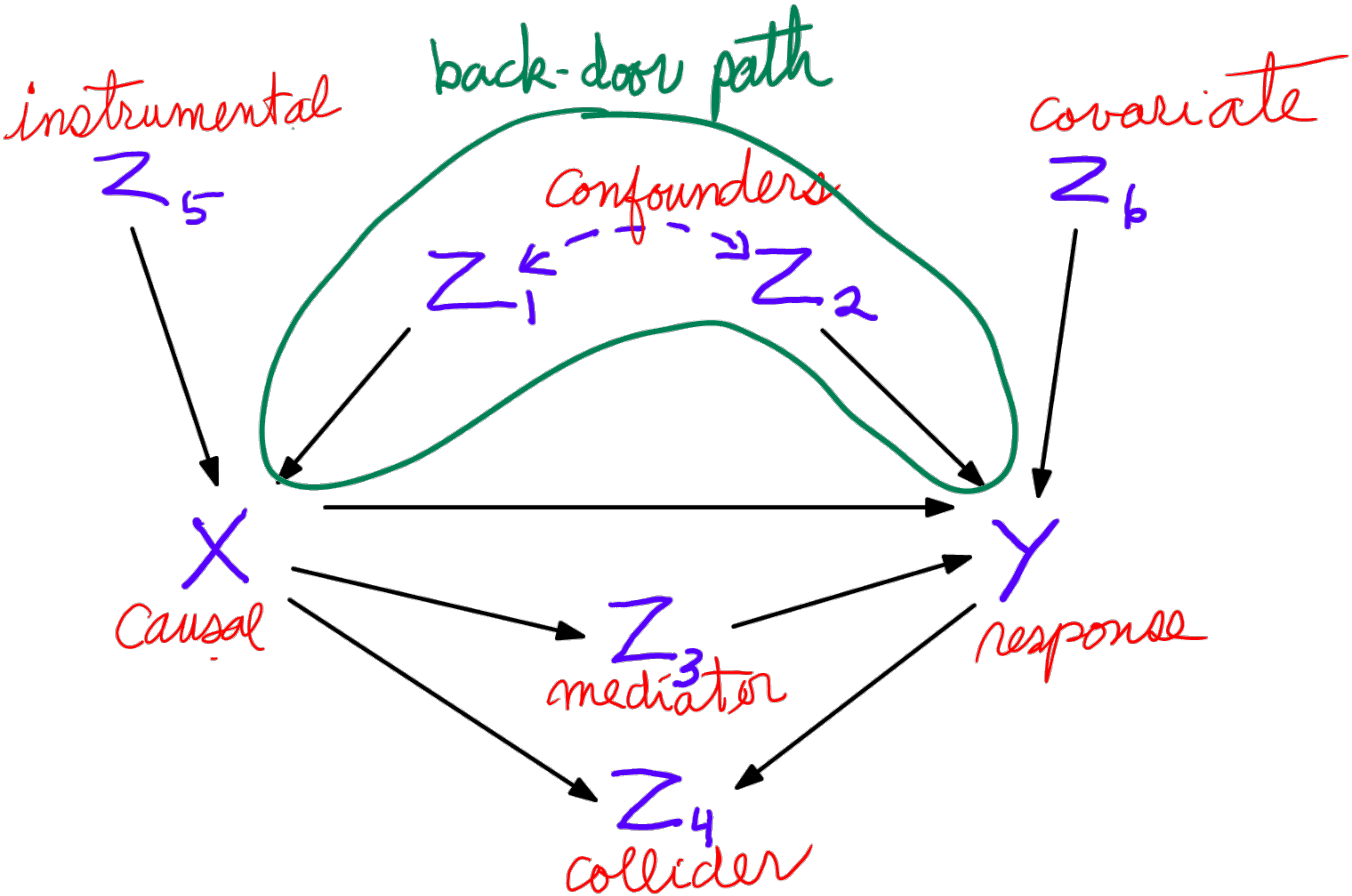

covariate
Confounders
Z_5
Z_1
Z_2
Z_6
X
Y
Causal
response
Z_3
Z_4

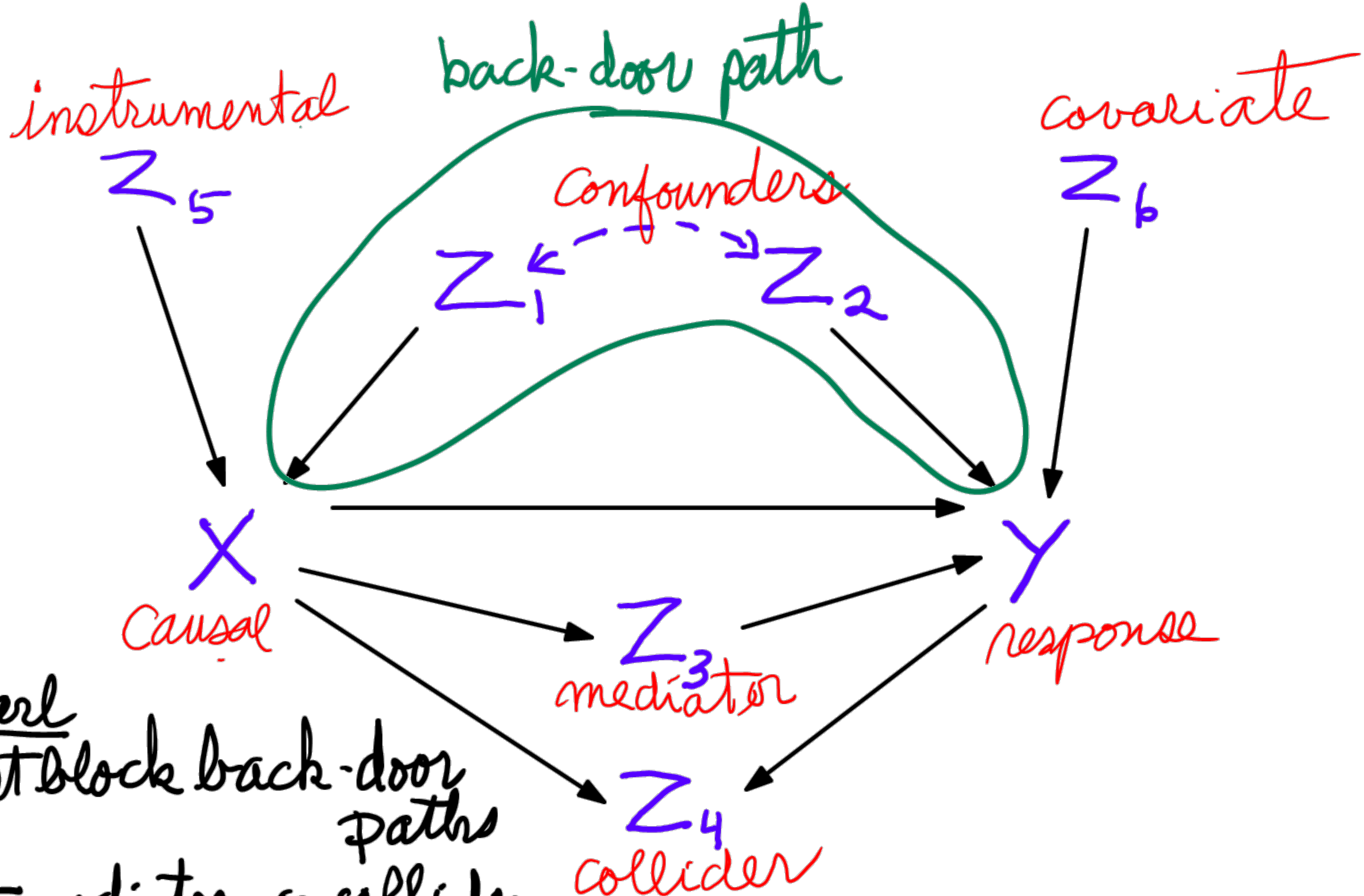

covariate
Z_5
Confounders
Z_1
Z_2
Z_6
X
Causal
Y
response
Z_3
mediator
Z_4


covariate
Z_5
confounders
Z_1
Z_2
Z_6
X
causal
Y
response
Z_3
mediator
Z_4
collider



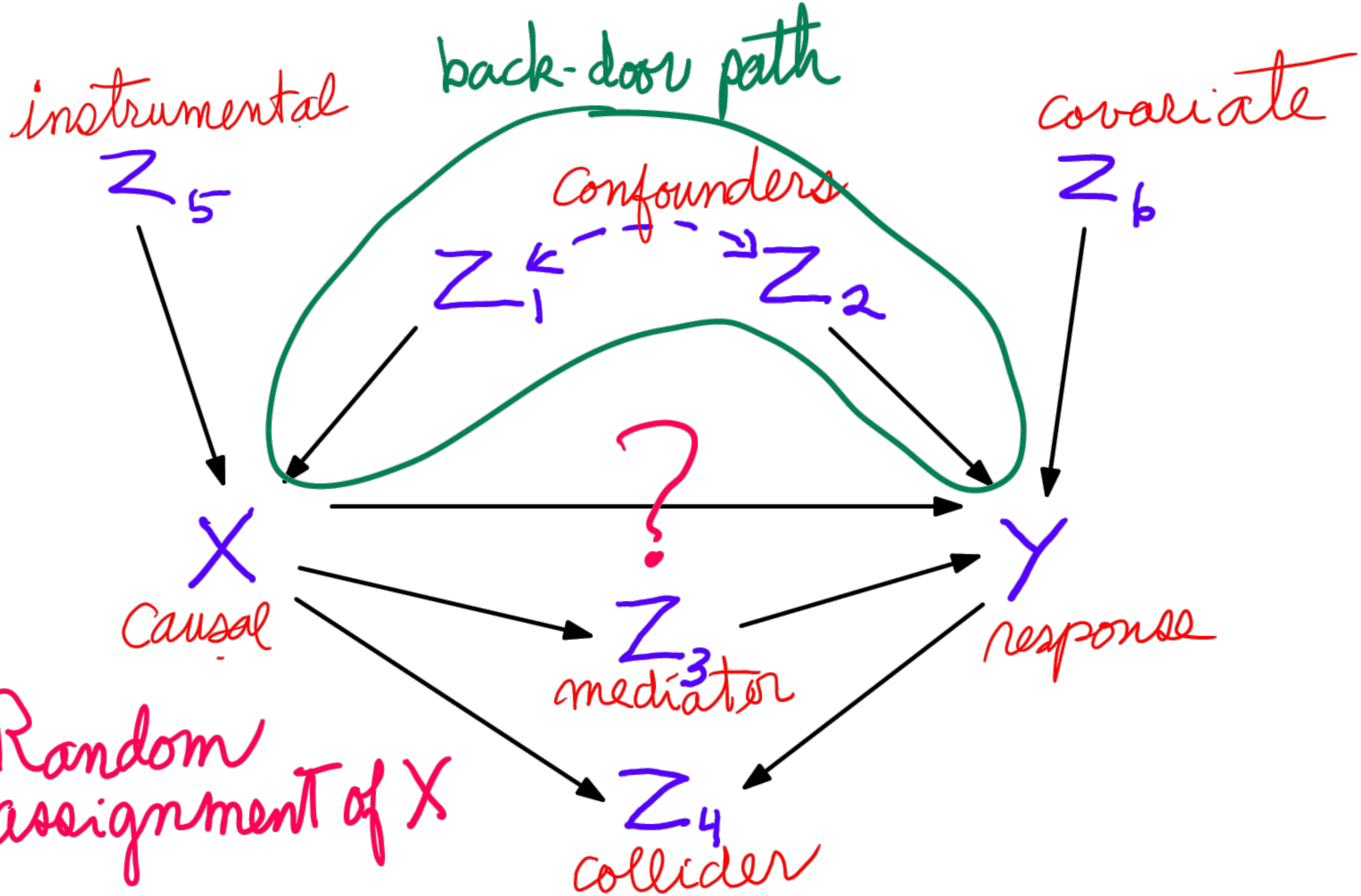
instrumental
Z5
confounders
Z1
Z2
covariate
Z6
X
causal
Y
response
Z3
mediator
Z4
collider


back-door path
instrumental
Z_5
confounders
Z_1
Z_2
covariate
Z_6
X
causal
Y
response
Z_3
mediator
Z_4
collider

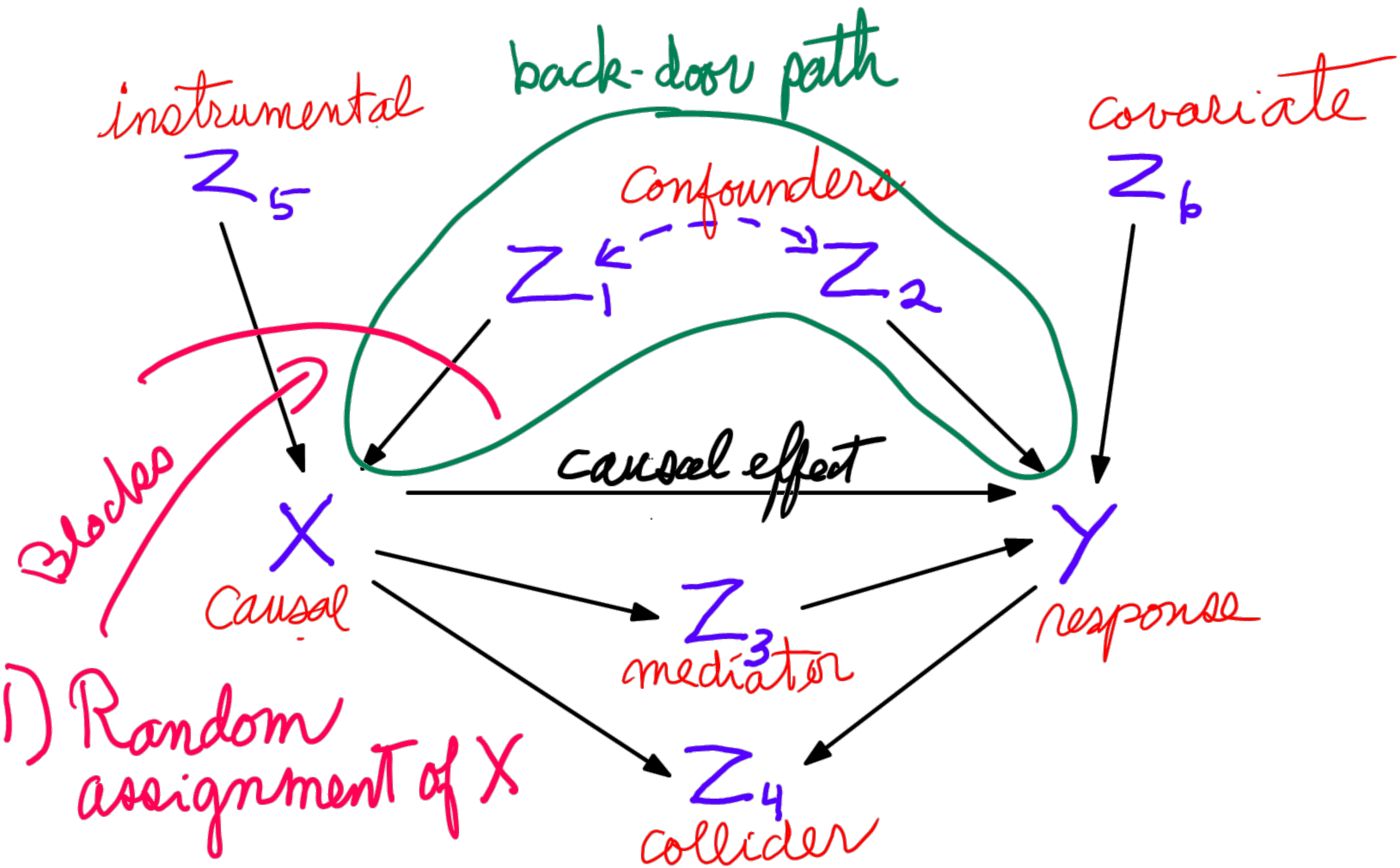


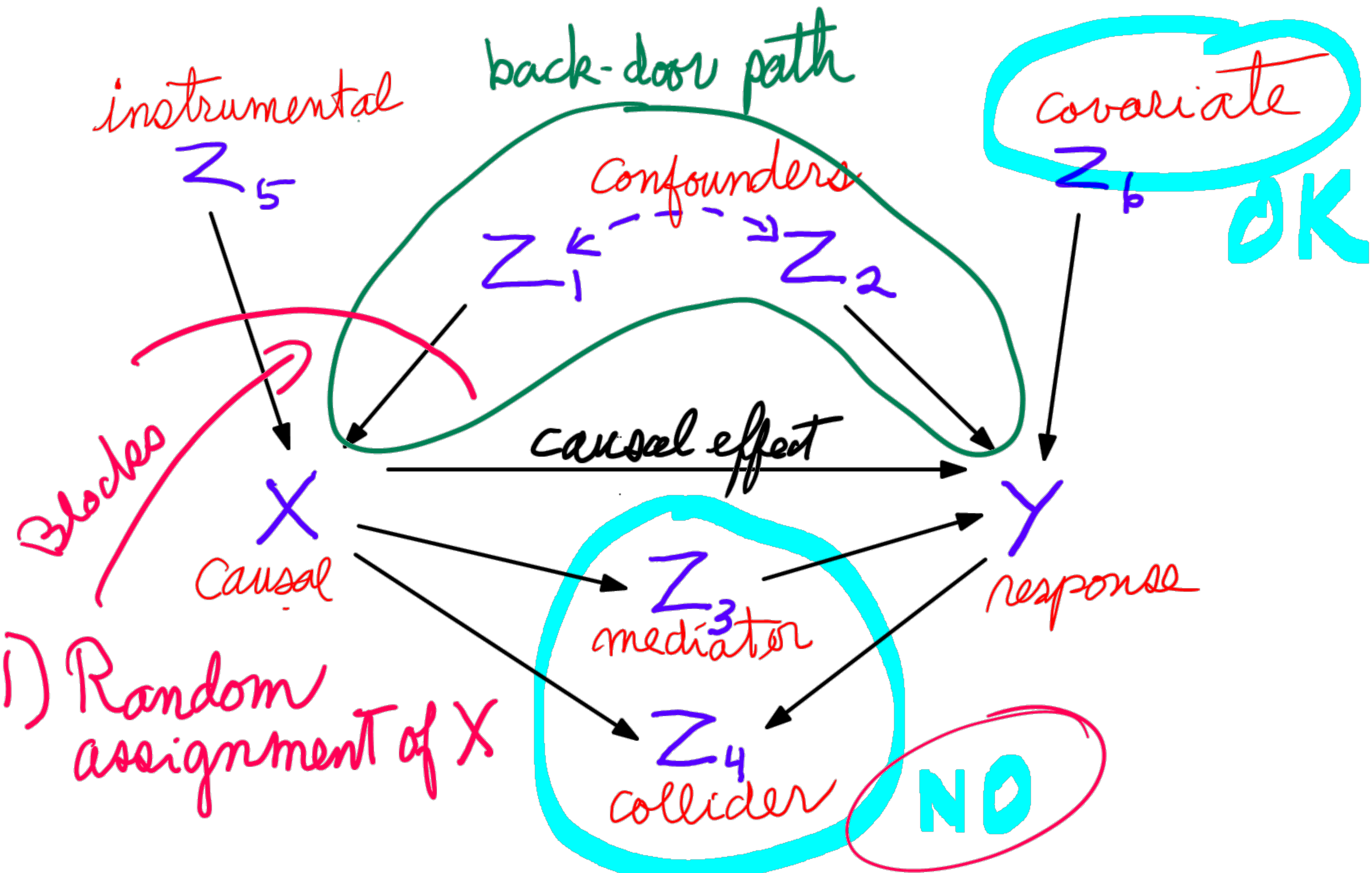
Pearl

- Must block back-door paths
- **NOT** mediator or collider

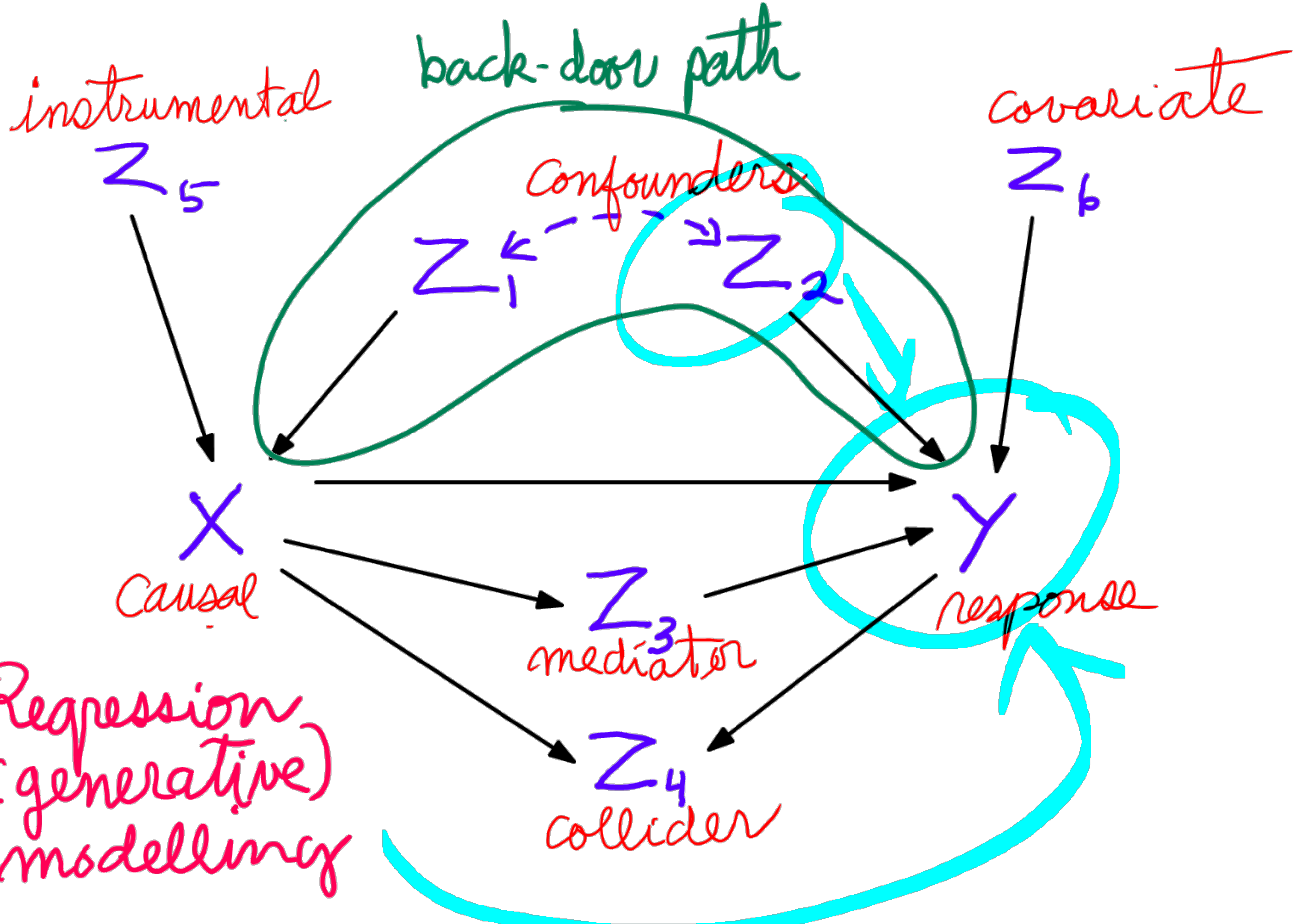


1) Random assignment of X

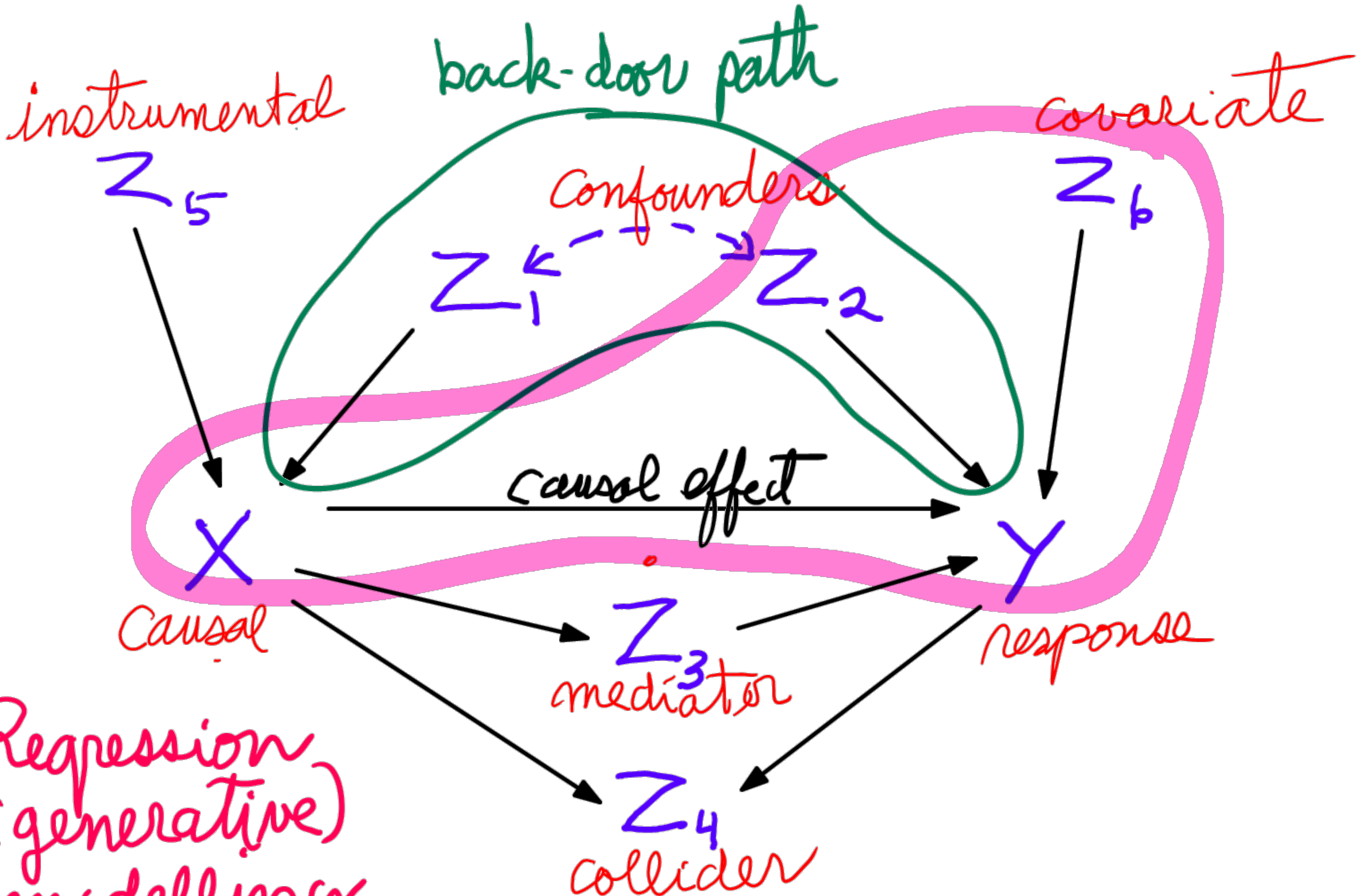

back-door path
instrumental
Z_5
covariate
Z_6
Confounders
Z_1
Z_2
causal effect
Blocks
X
causal
Y
response
Z_3
mediator
Z_4
collider
1) Random assignment of X



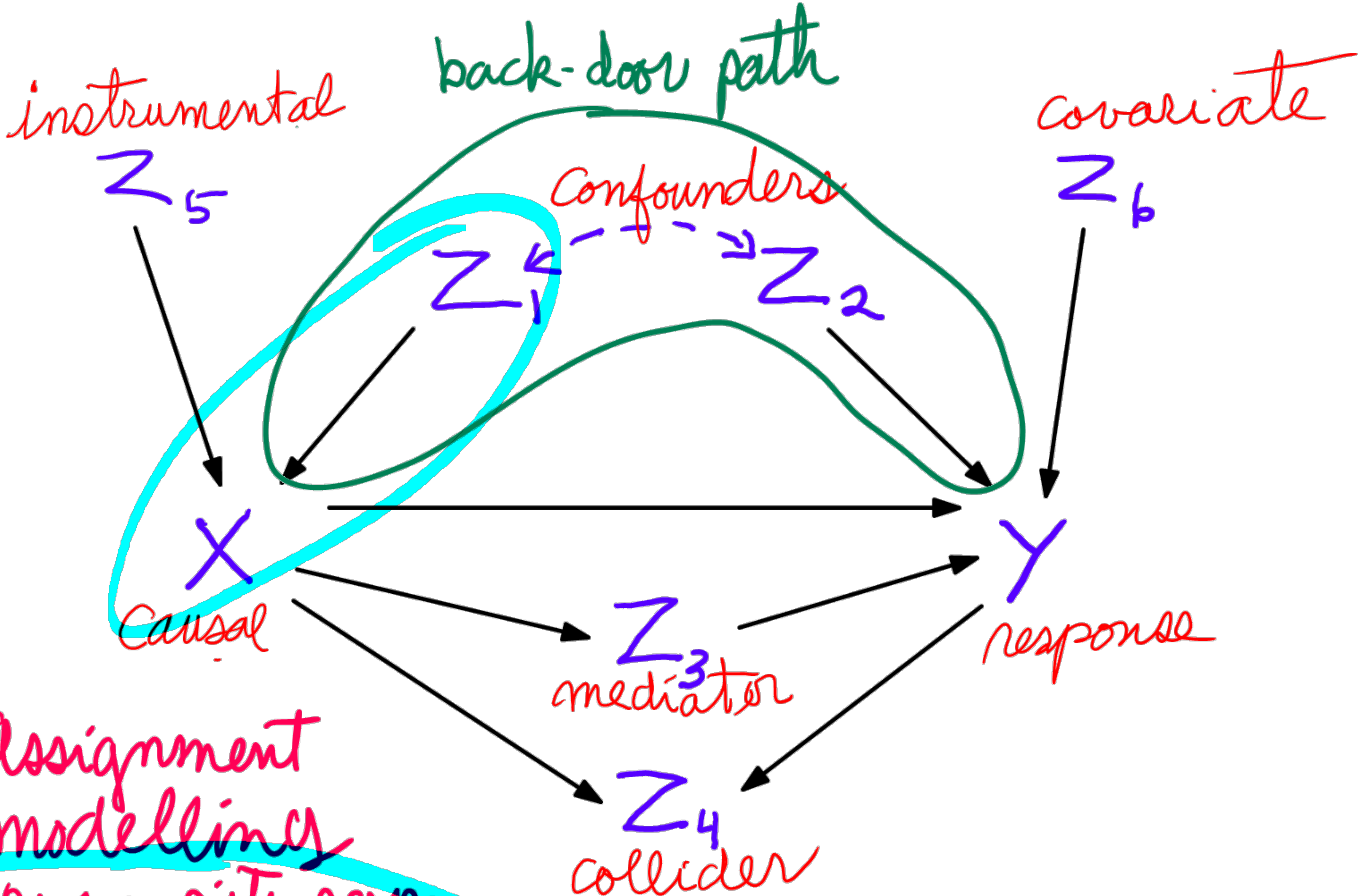
instrumental
Z5
back-door path
confounders
Z1
Z2
covariate
Z6
OK
Blocks
X
causal effect
Y
causal
Z3
mediator
response
1) Random assignment of X
Z4
collider
NO



2) Regression (generative) modelling

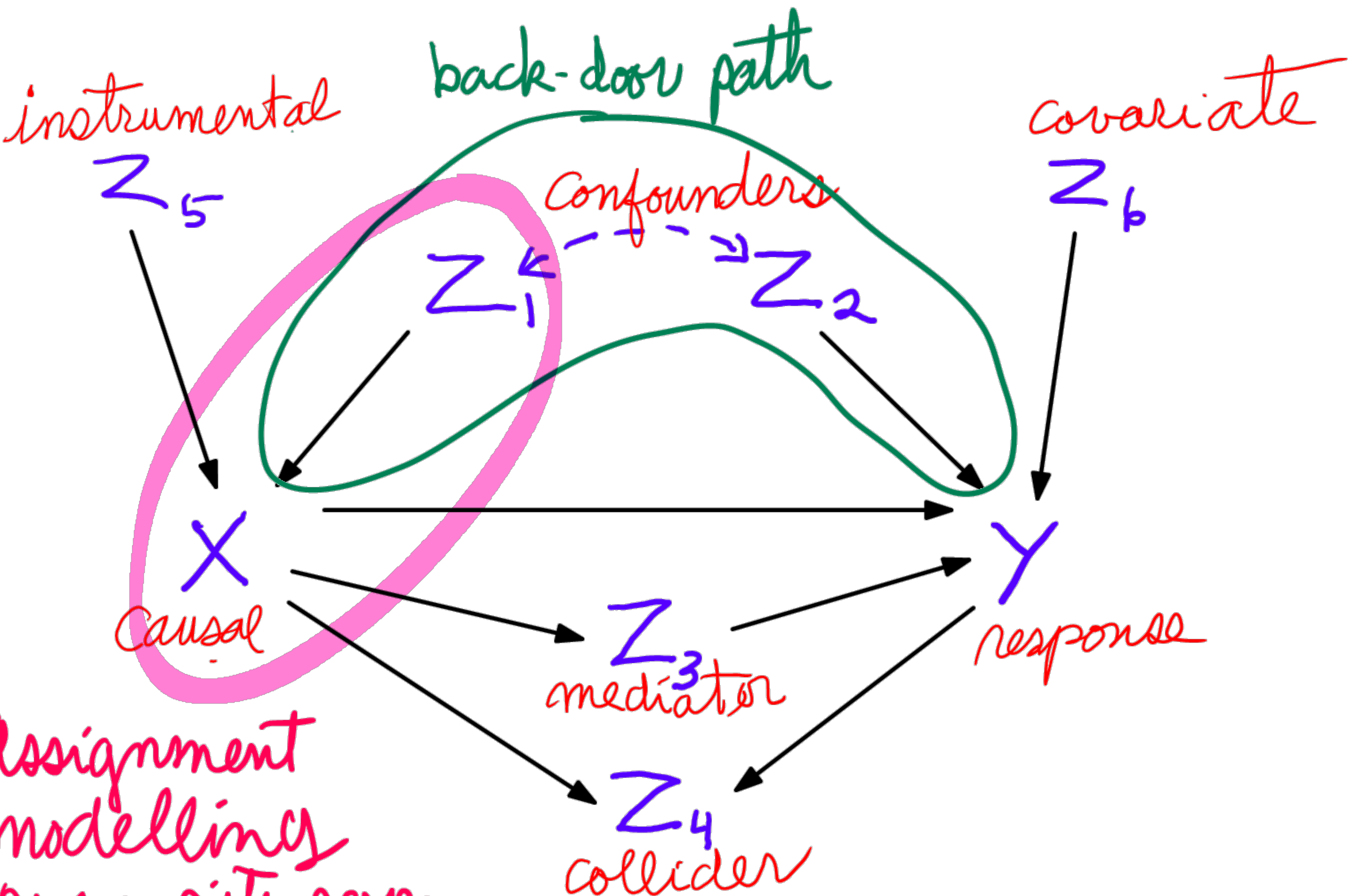


2) Regression (generative) modelling

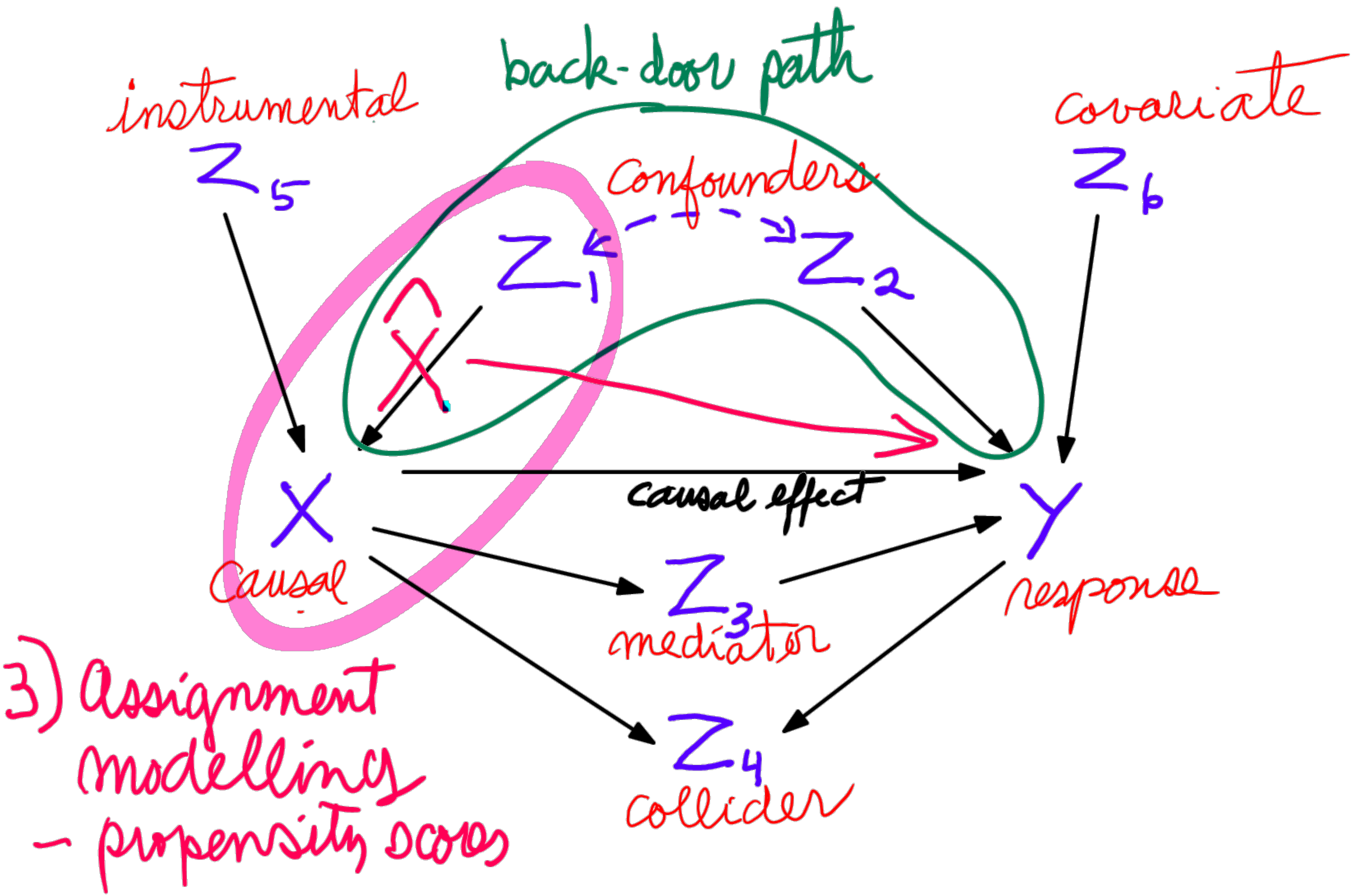


3) Assignment modelling

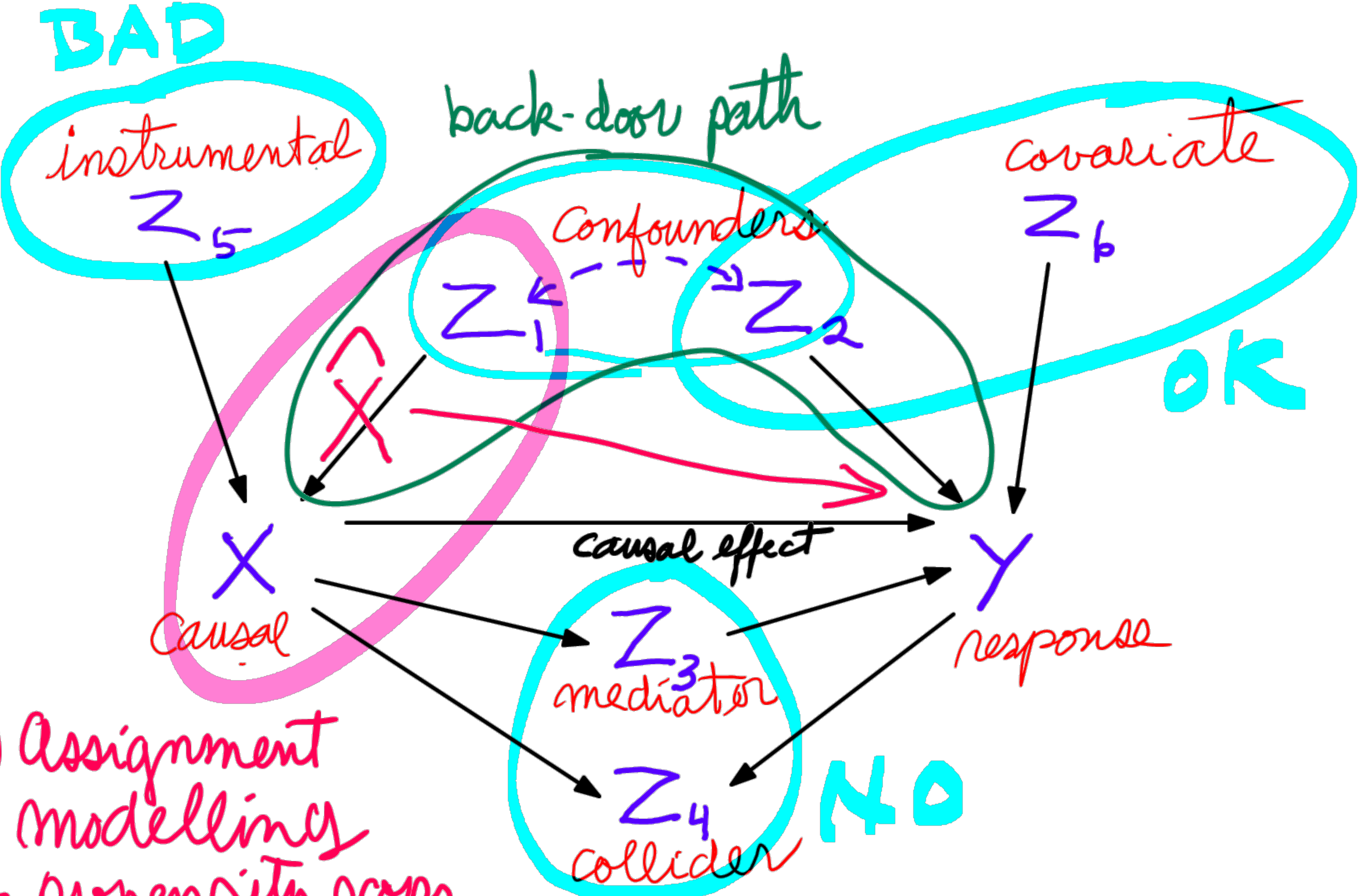
– propensity scores



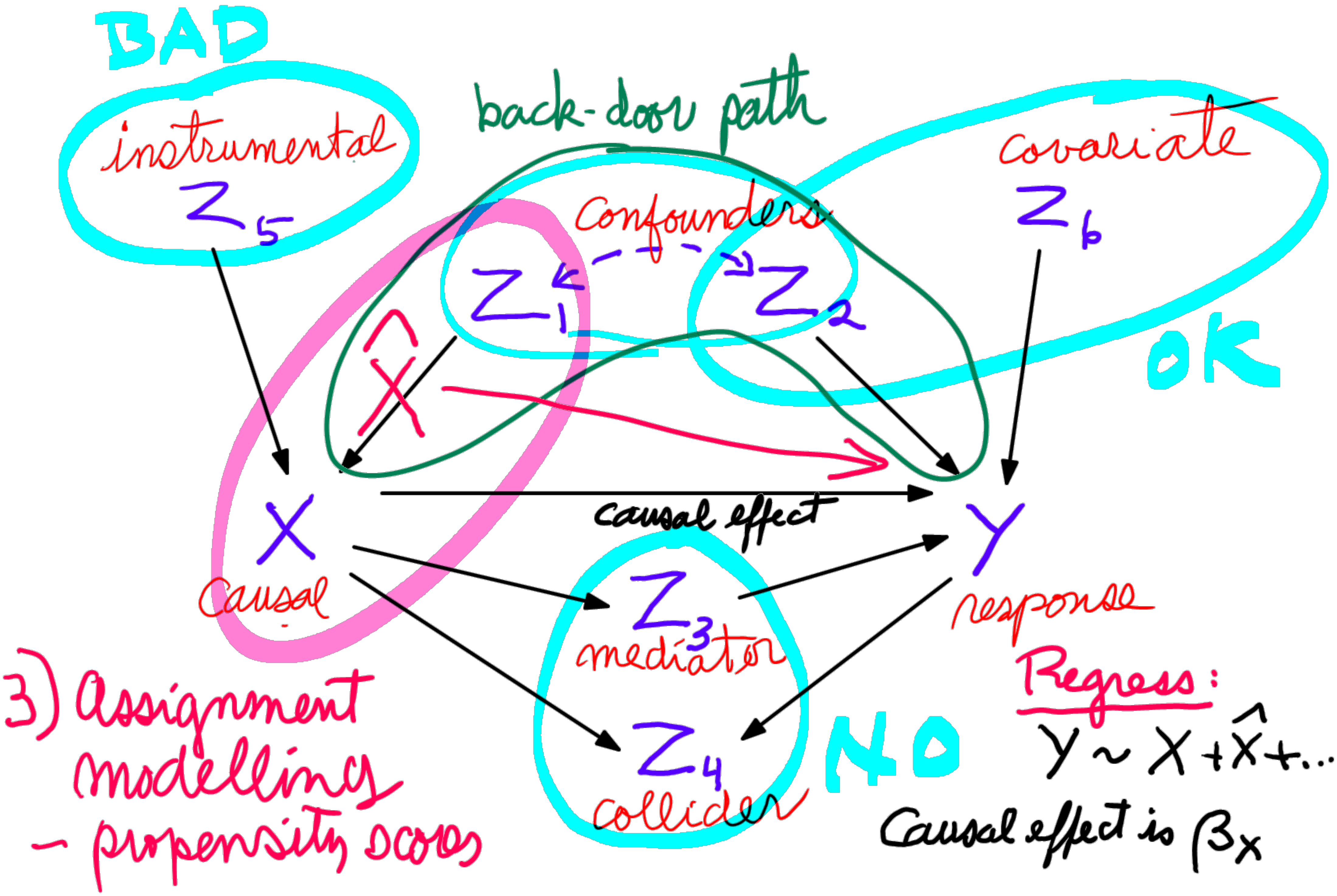
3) Assignment modelling
- propensity scores



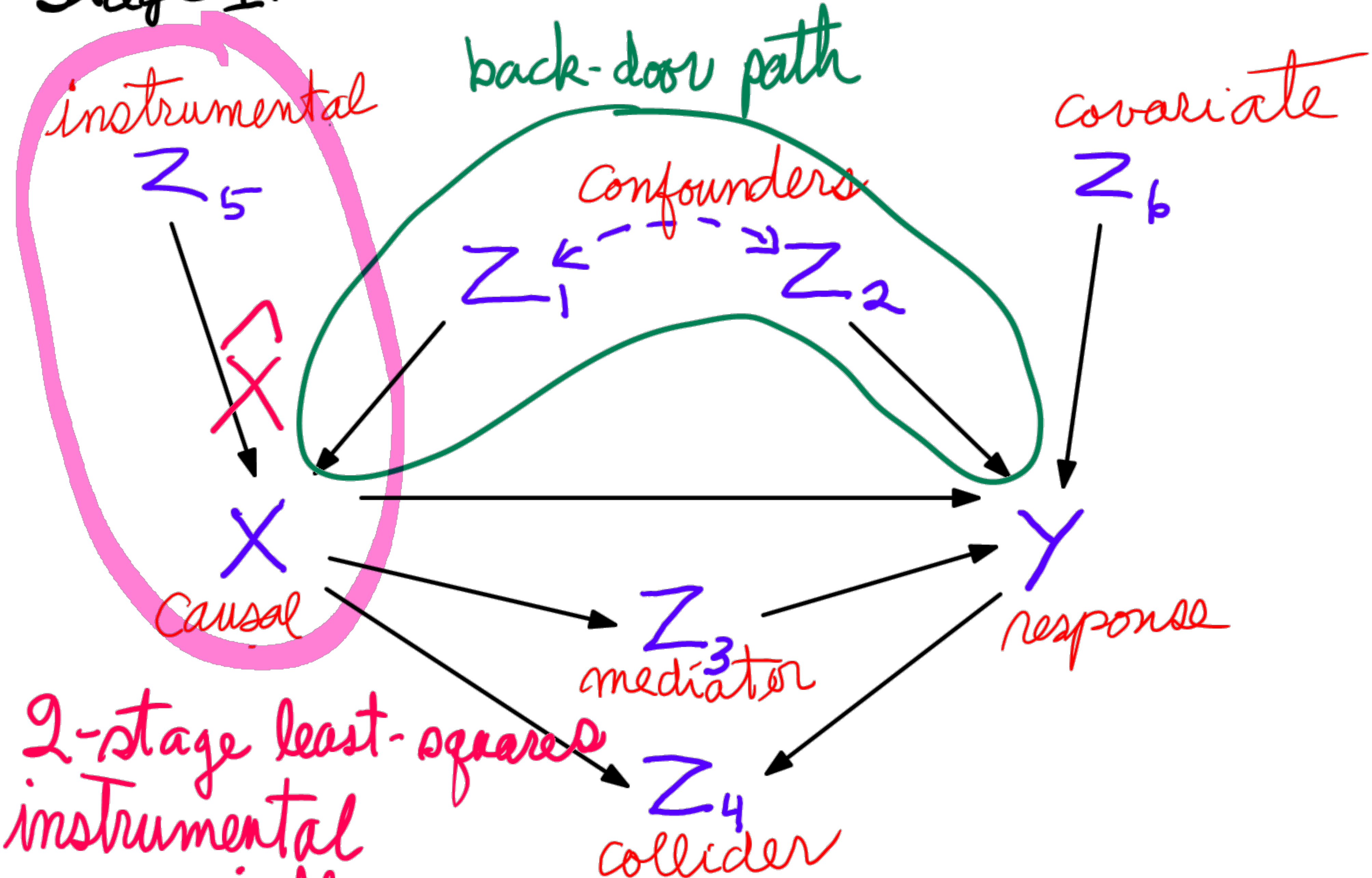
back-door path
instrumental
Z_5
covariate
Z_6
Confounders
Z_1
Z_2
X
causal effect
Y
Causal
response
Z_3
mediator
Z_4
collider
3) Assignment modelling
- propensity scores



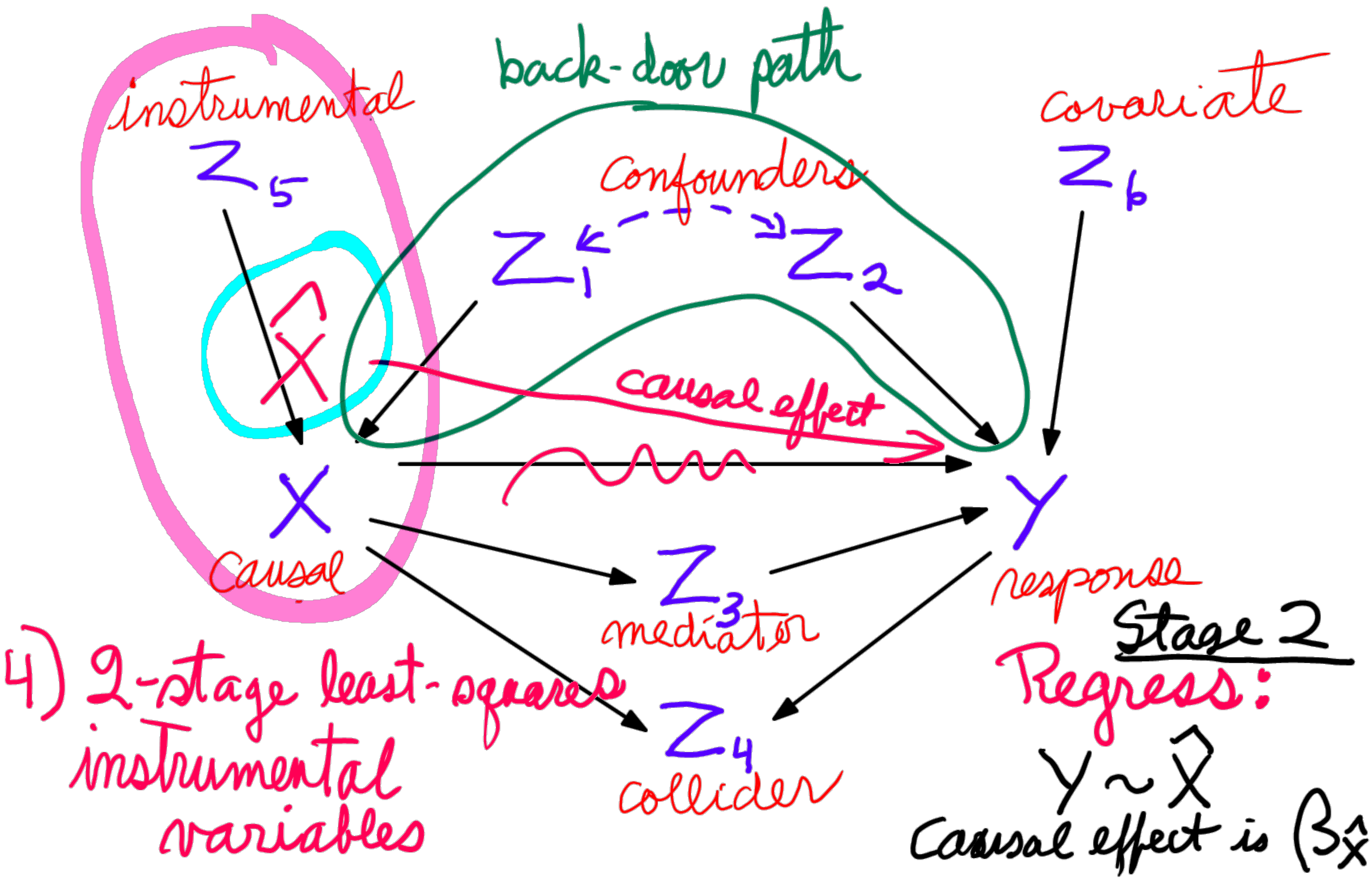
3) Assignment modelling
– propensity scores


BAD
instrumental
Z_5
back-door path
confounders
Z_1
Z_2
$\hat{X}$
covariate
Z_6
OK
X
Causal
causal effect
Y
response
Z_3
mediator
Z_4
collider
NO
3) Assignment modelling
- propensity scores
Regress:
$Y \sim X + \hat{X} + \ldots$
Causal effect is β_X

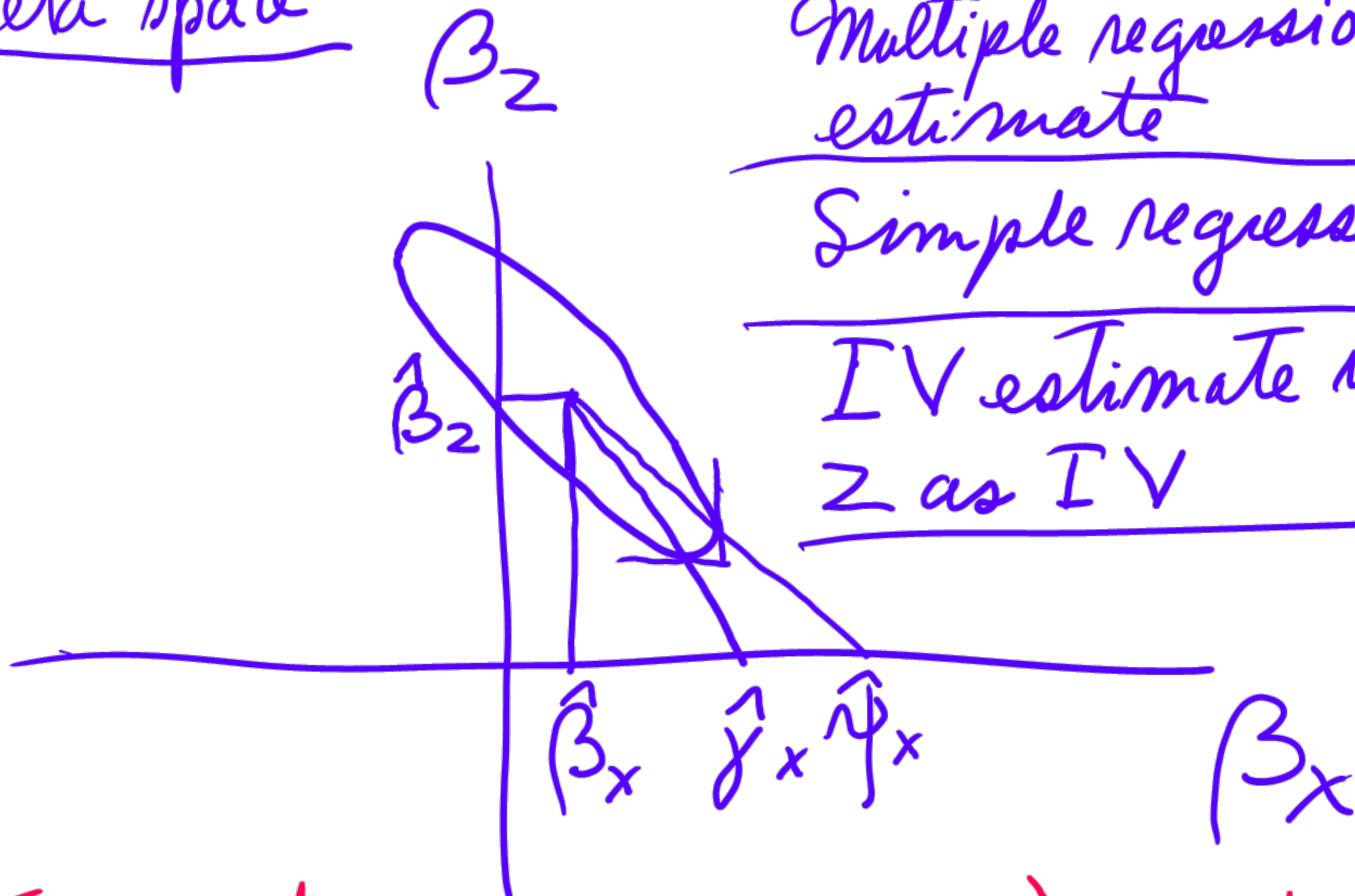
Stage 1:



4) 2-stage least-squares instrumental variables


back-door path
instrumental
Z_5
covariate
Z_6
confounders
Z_1
Z_2
$\hat{X}$
causal effect
X
Y
causal
Z_3
mediator
response
Z_4
collider
4) 2-stage least-squares
instrumental
variables
Stage 2
Regress:
$Y \sim \hat{X}$
Causal effect is $\beta_{\hat{X}}$

Beta space



Multiple regression estimate	$\hat{\beta}_x$
Simple regression	$\hat{\gamma}_x$
IV estimate with Z as IV	$\hat{\psi}_x$

For a strong IV, Corr(X,Z) close to 1 and exclusion restriction ⇒ β_2 close to 0

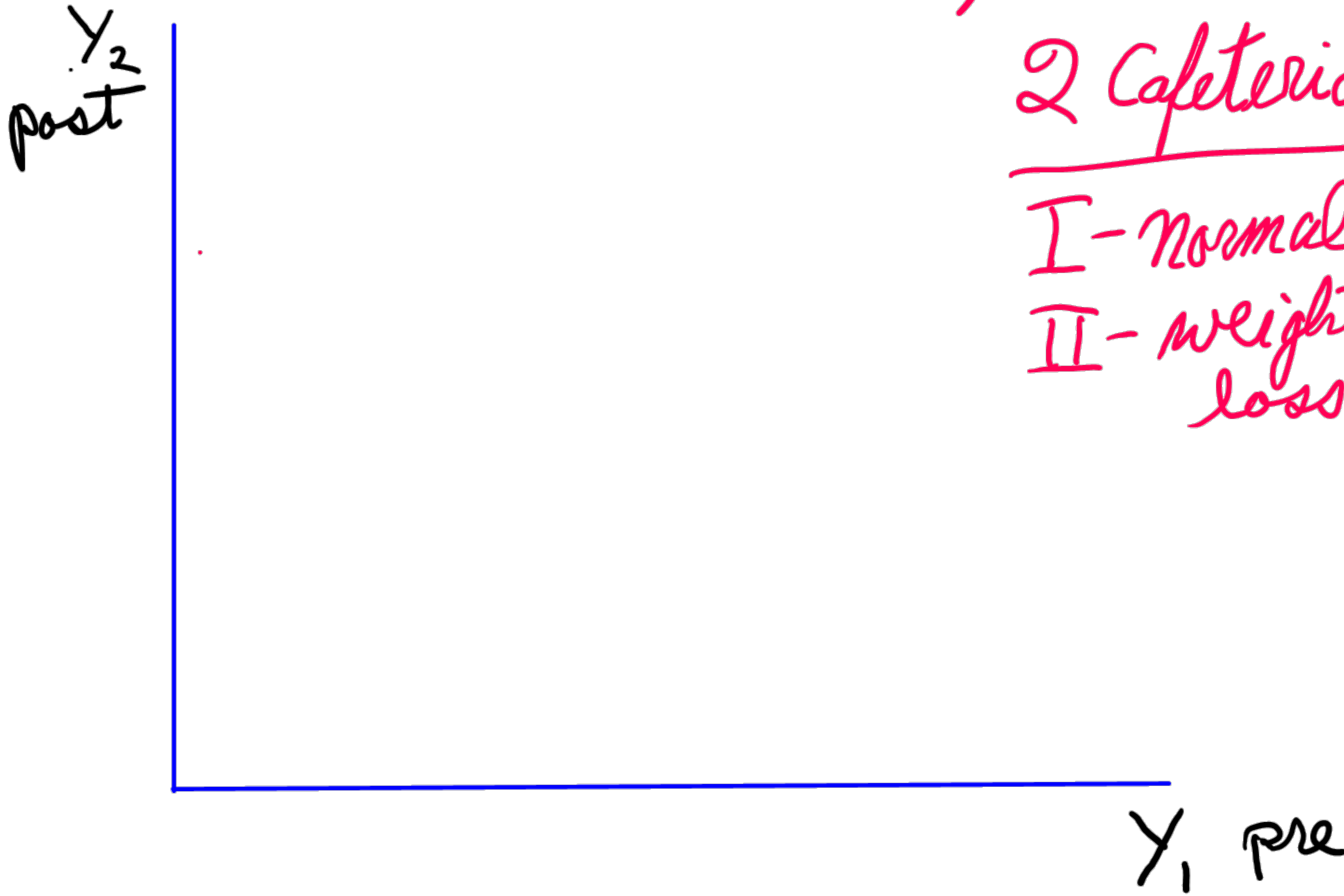
So $\hat{\psi}$ close to $\hat{\gamma}_x$

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

$$Y \sim I + X$$

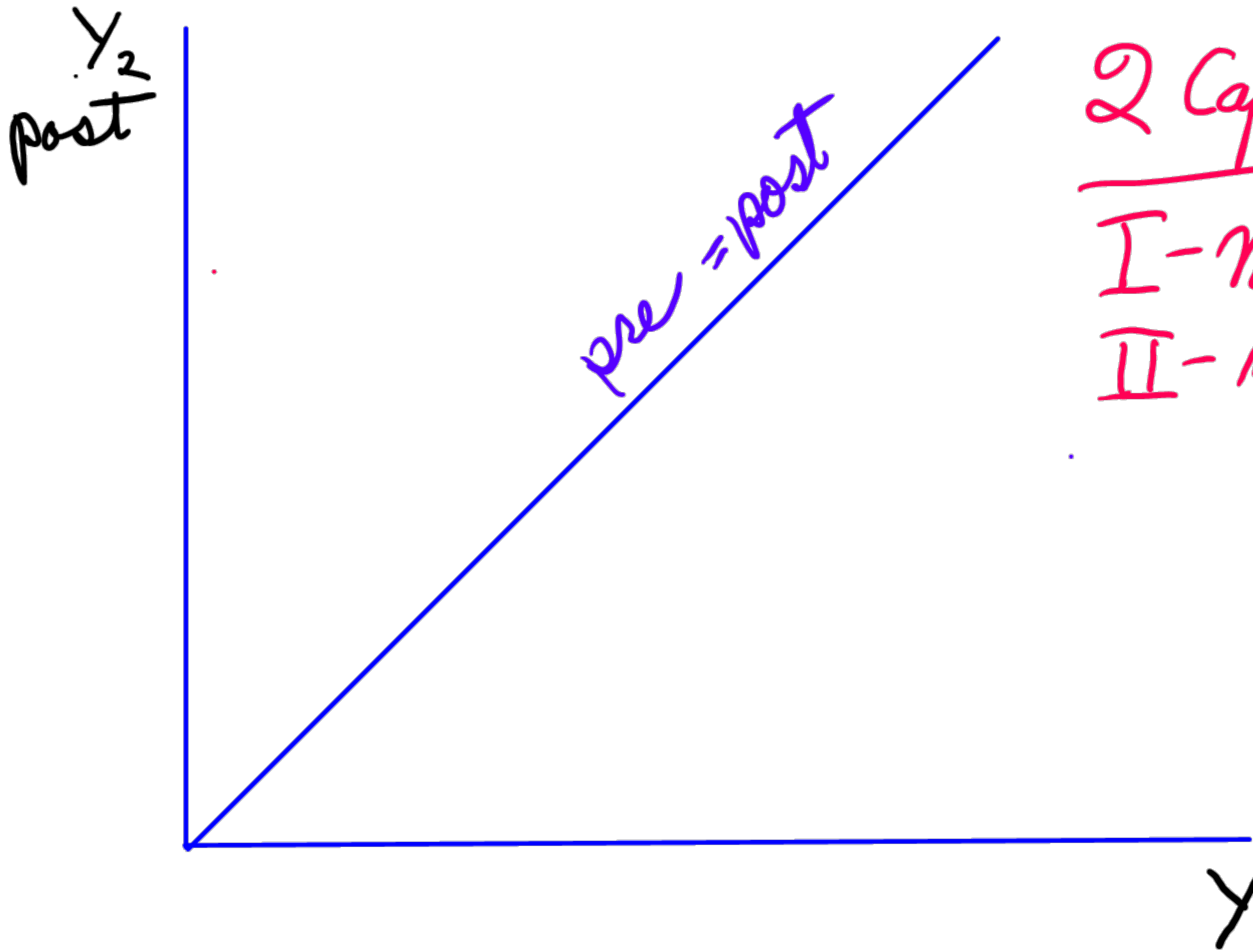
since X is a collider if there is an omitted confounder C and $\hat{\beta}_I$ should not be 0 even if I is a good instrument.

Lord's Paradox (Wainer version)



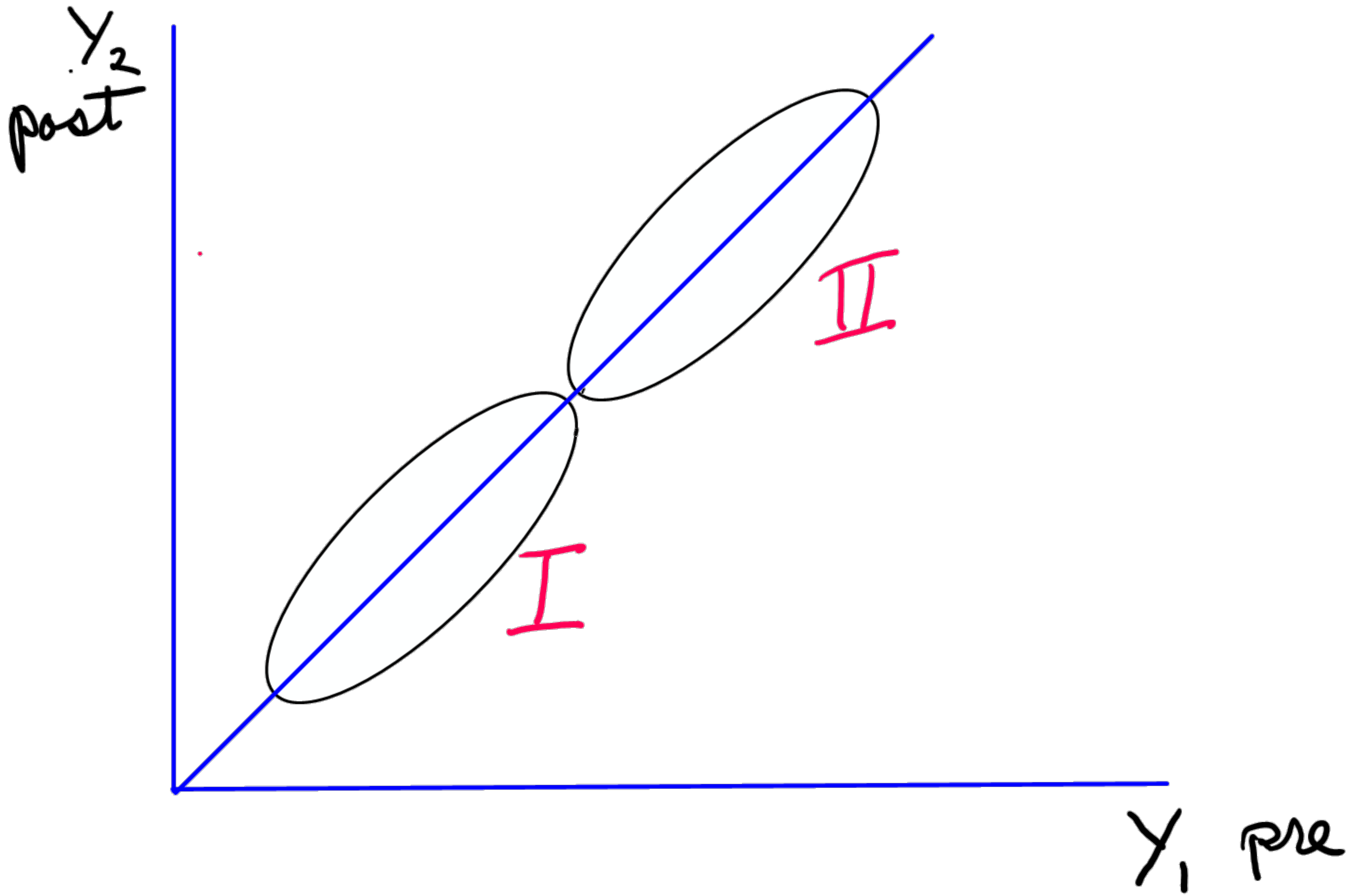
2 Cafeterias

I - normal
II - weight loss

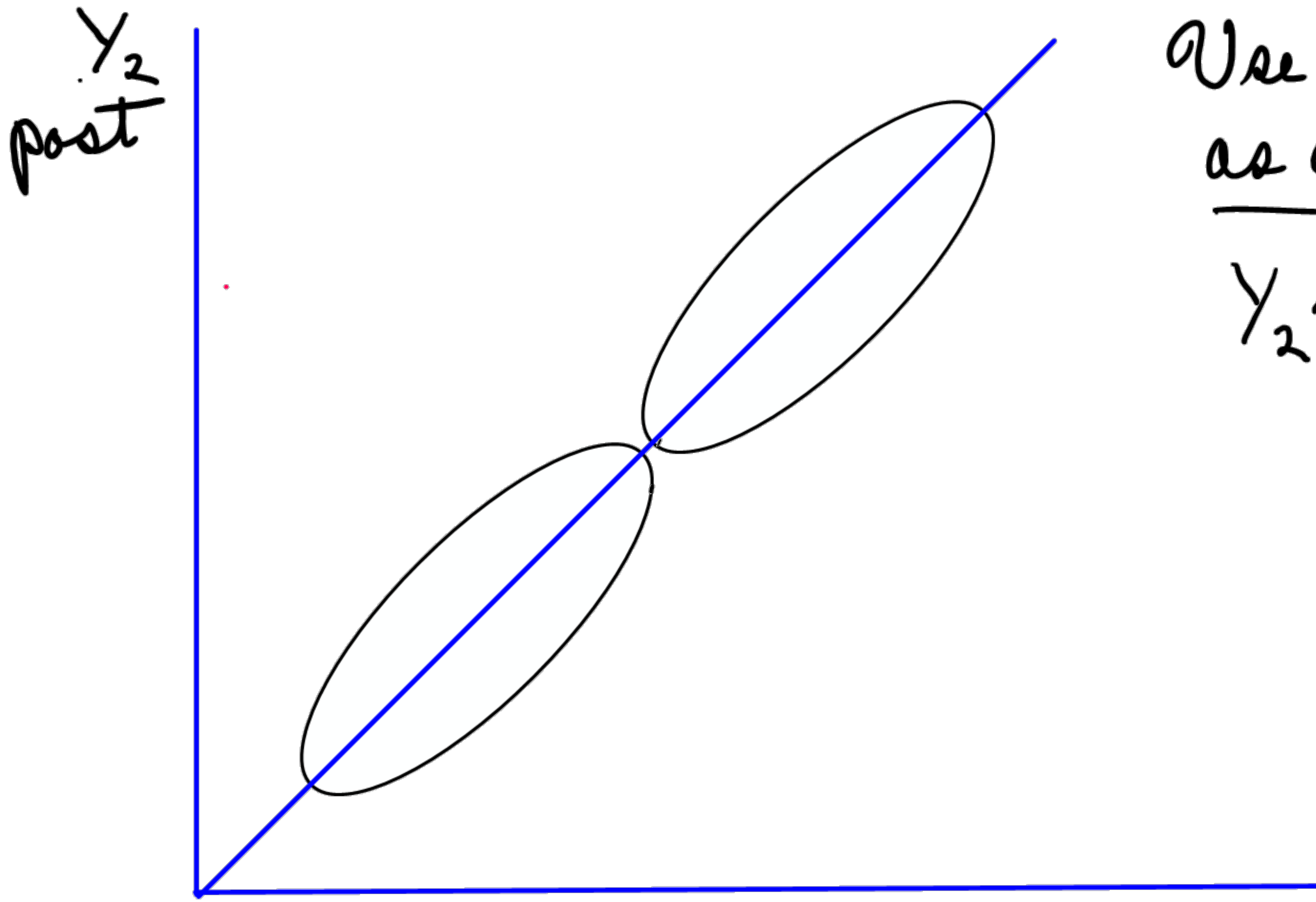


2 Cafeterias

I - normal
II - weight loss

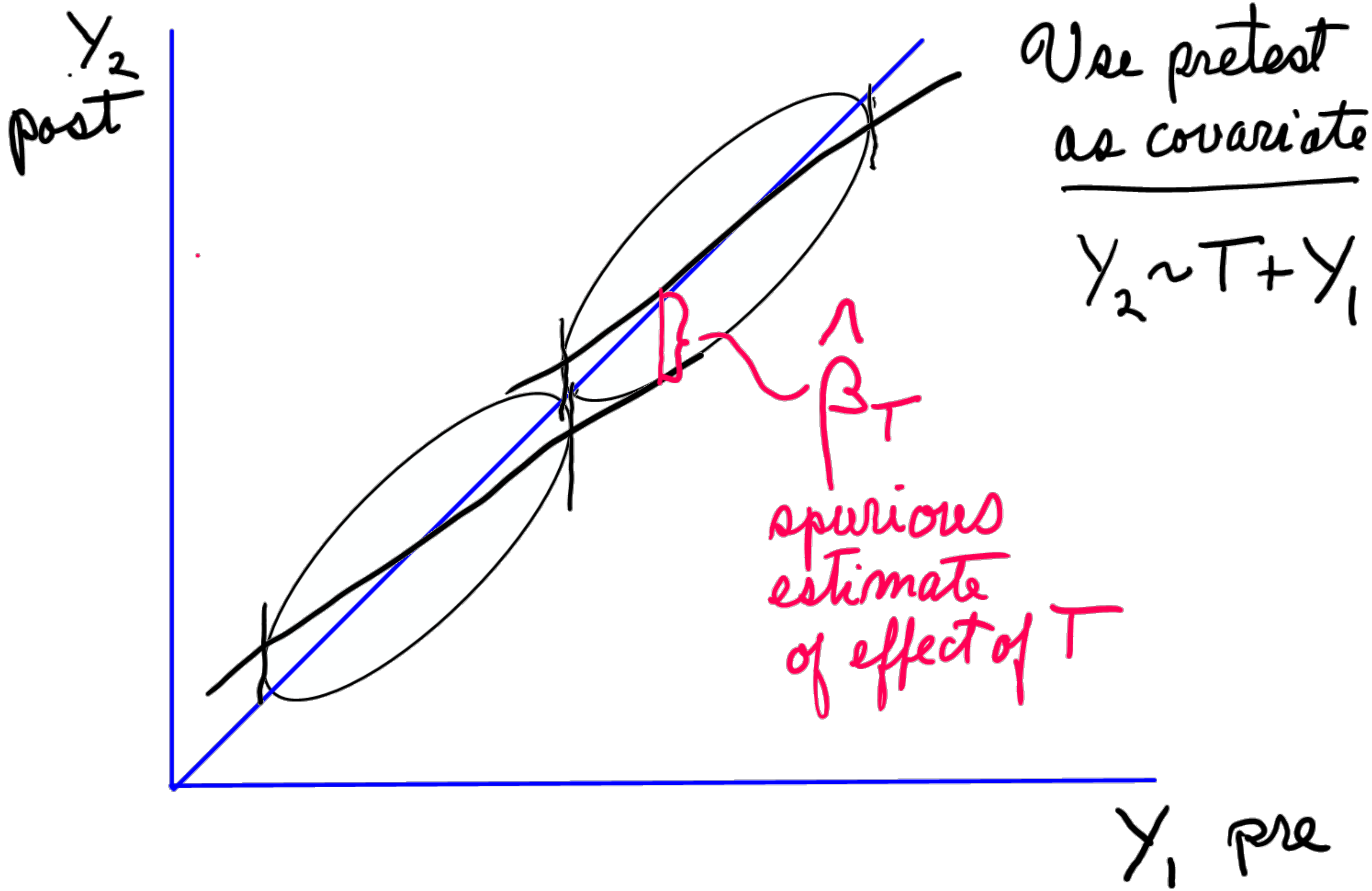


Y_2
post
I
II
Y_1 pre

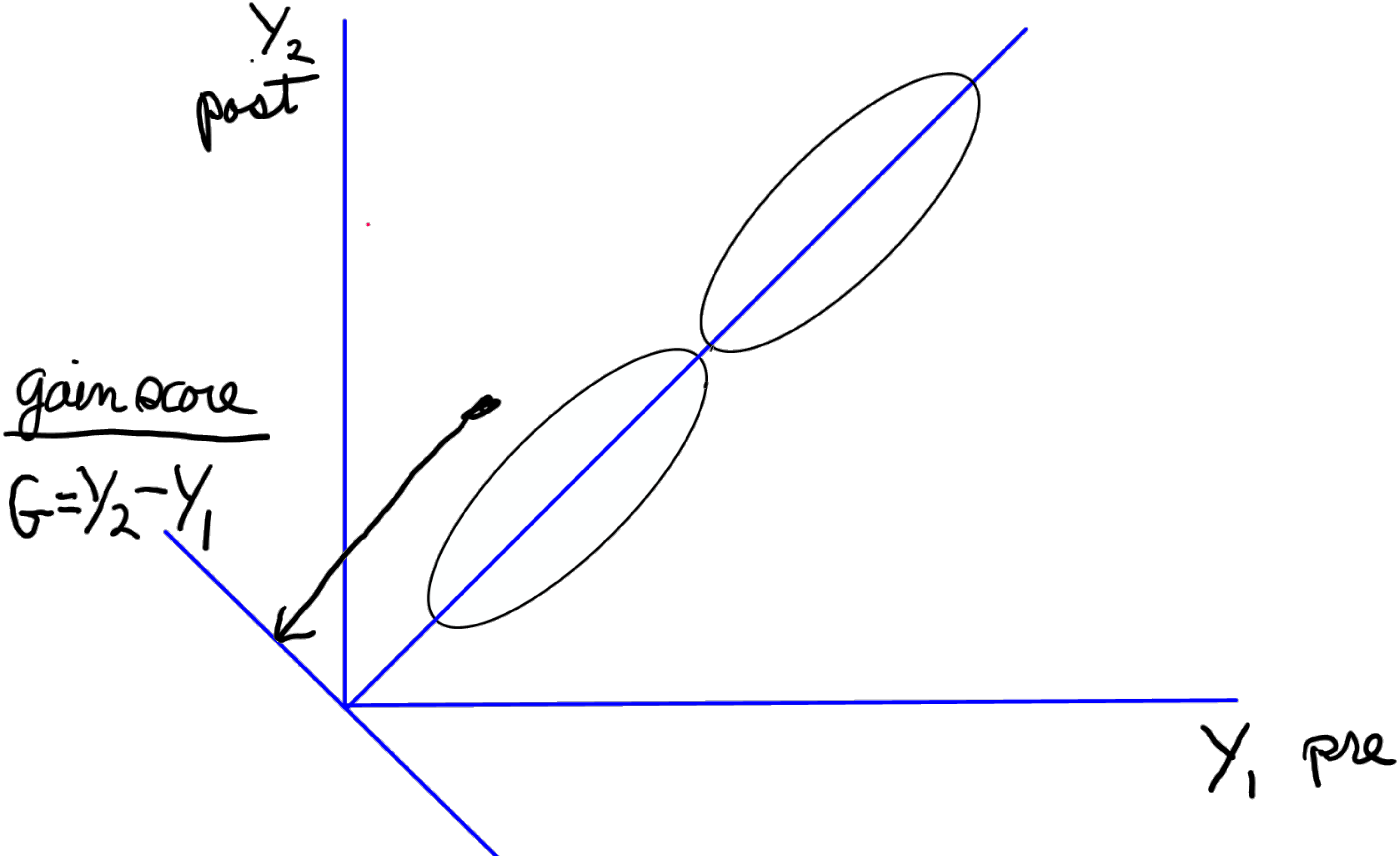


Use pretest as covariate

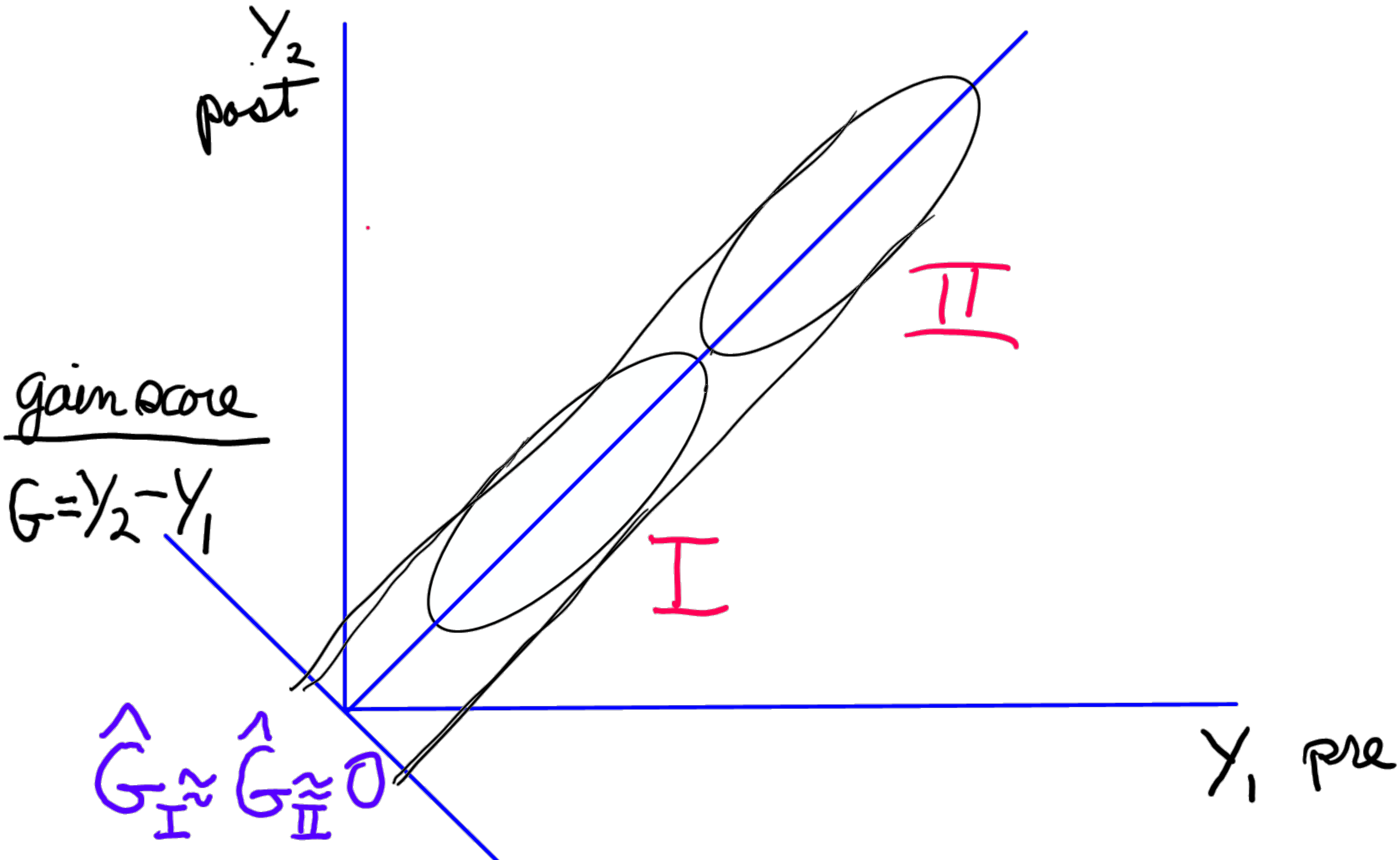
$Y_2 \sim T + Y_1$



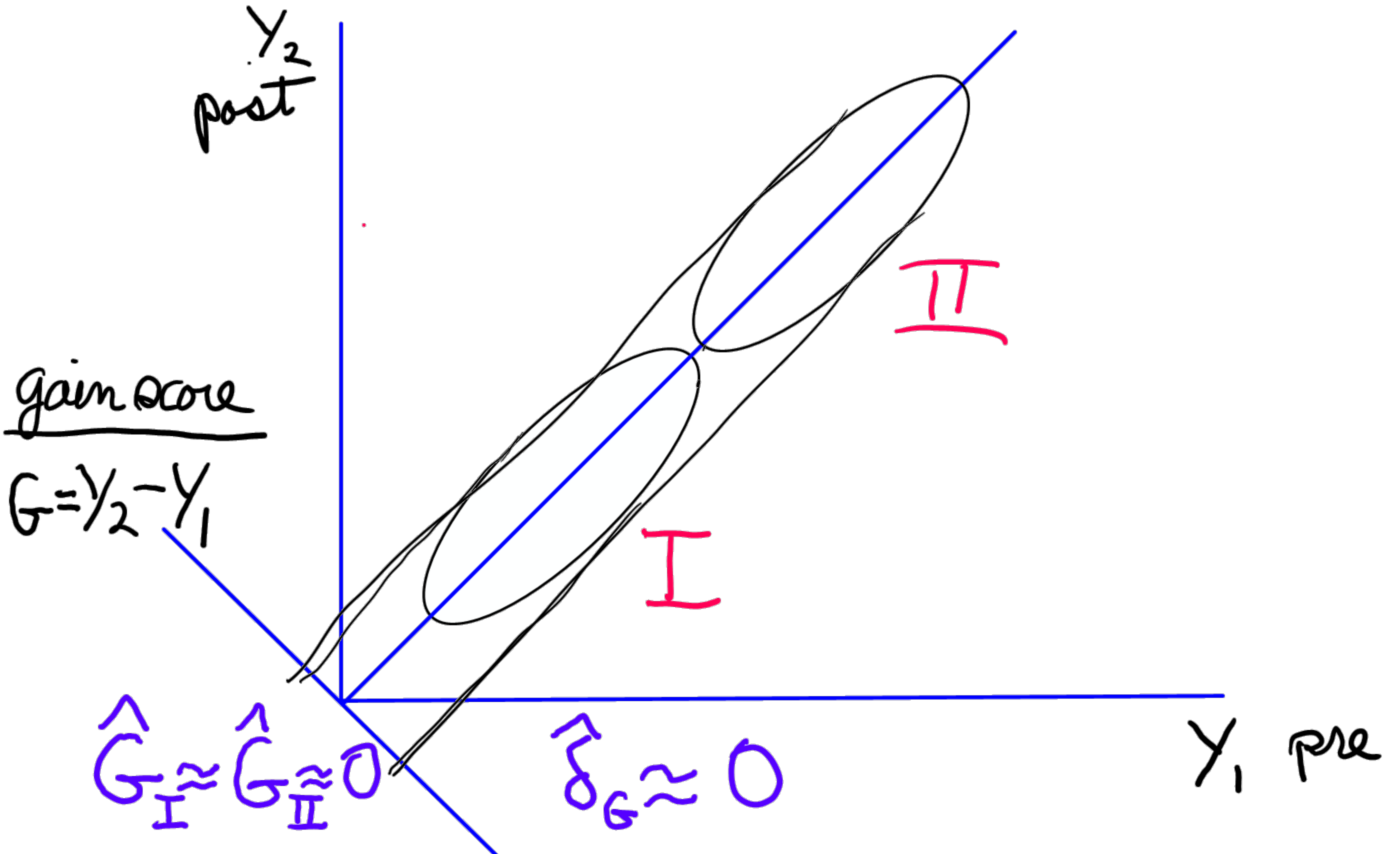
Use pretest as covariate
$Y_2 \sim T + Y_1$
Y_2 post
Y_1 pre
$\hat{\beta}_T$
spurious estimate of effect of T



Y_2 post
gain score
$G = Y_2 - Y_1$
Y_1 pre



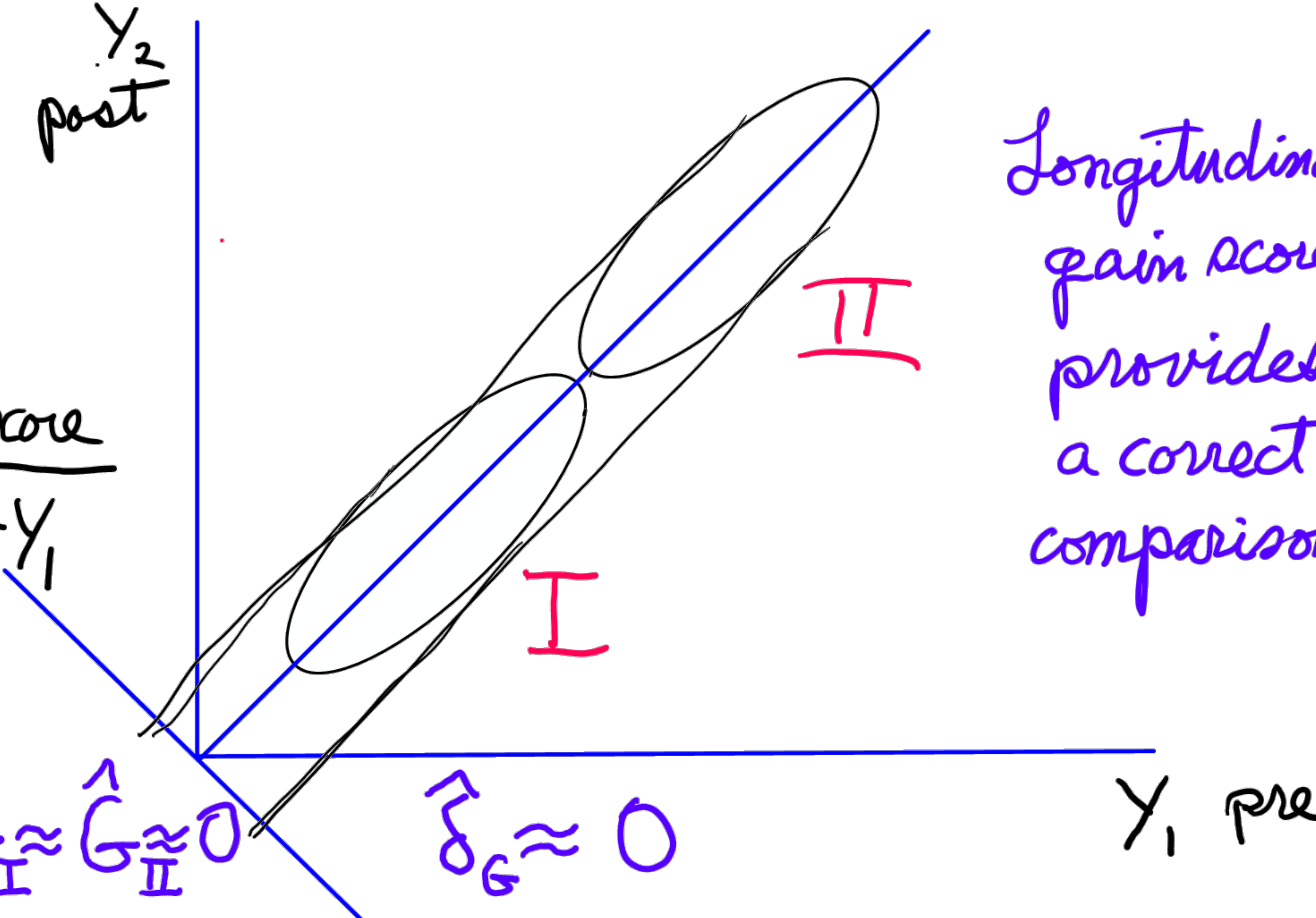
Y_2 post
gain score
$G = Y_2 - Y_1$
II
I
Y_1 pre
$\hat{G}_{I} \approx \hat{G}_{II} \approx 0$



Y2 post
gain score
G = Y2 − Y1
II
I
Ĝ_I ≈ Ĝ_II ≈ 0
δ̂_G ≈ 0
Y1 pre

gain score

$G = Y_2 - Y_1$



Longitudinal gain score provides a correct comparison

$\hat{G}_{I} \approx \hat{G}_{II} \approx 0$ $\quad$ $\hat{\delta}_G \approx 0$

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(Yres)/SD(Xres)
   Small SD(Yres), Large SD(Xres)
3) Honest SE

4) Robust   Propensity scores – focus on X

Use the AVP to compare models.

Using confounders close to Y

$\downarrow SD(Y_{res})$, $\uparrow SD(X_{res})$ } $\downarrow SE(\hat{\beta}_T)$

But may not have knowledge about structure of model for Y

Using confounders close to X

$\uparrow SD(Y_{res})$, $\downarrow SD(X_{res})$ } $\uparrow SE(\hat{\beta}_T)$

But may have better understanding of assignment model for X

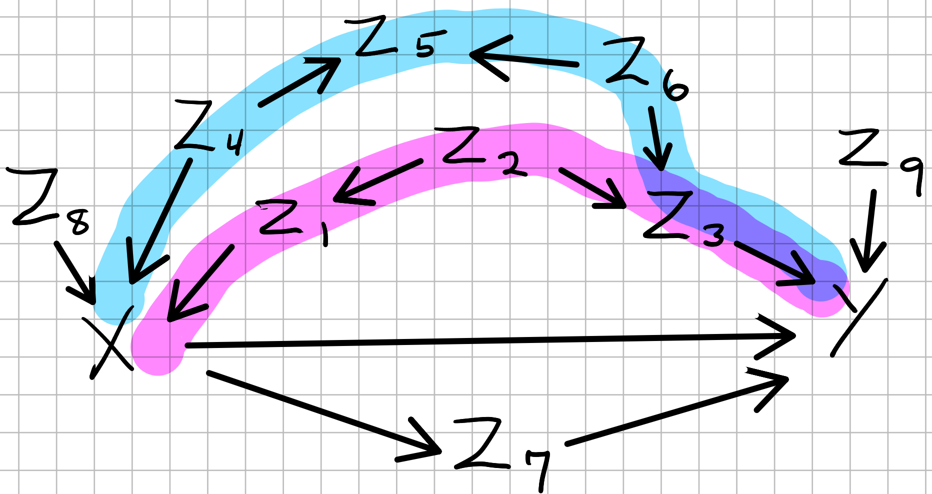
Propensity score methods focus on predicting X with $\hat{X}$

- 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:



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

(blue) Back door path #1

(pink) " " " #2

- Not including Z_5 blocks #1
- any of Z_1, Z_2, Z_3 blocks #2

$$SD(\hat{\beta}_X) = ? = \frac{1}{\sqrt{n}} \frac{S_e}{S_{X|others}}$$

Comparing models, consider impact on S_e & $S_{X|others}$