Testing Exogeneity of Instrumental Variables Using Pretest-Posttest Designs

Naram Gwak
Yongnam Kim

Department of Education
Seoul National University

Overview

Exogeneity of Instrumental Variables (IV)

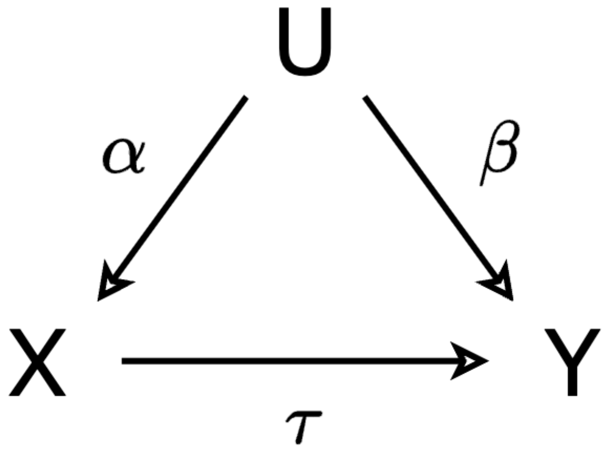
- Exogeneity assumption determines the validity of IV
- Empirically verifying this assumption is infeasible

Developing Approach for Testing the Validity of Exogeneity

- By utilizing pretest-posttest designs, possible to assess exogeneity
- A new approach for empirically testing violations of exogeneity

IV in Causal Graph

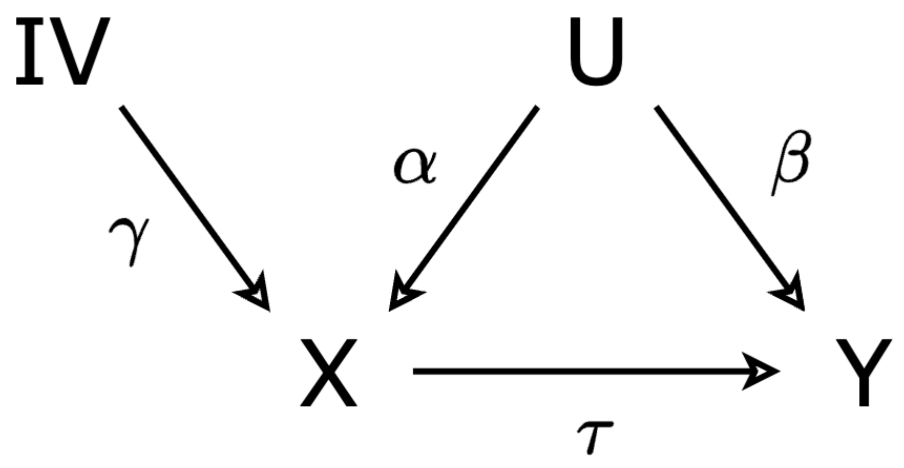
We cannot identify the causal effect of X on Y, due to unobserved confounders



IV in Causal Graph

Using IV,
we can identify the causal effect of X on Y

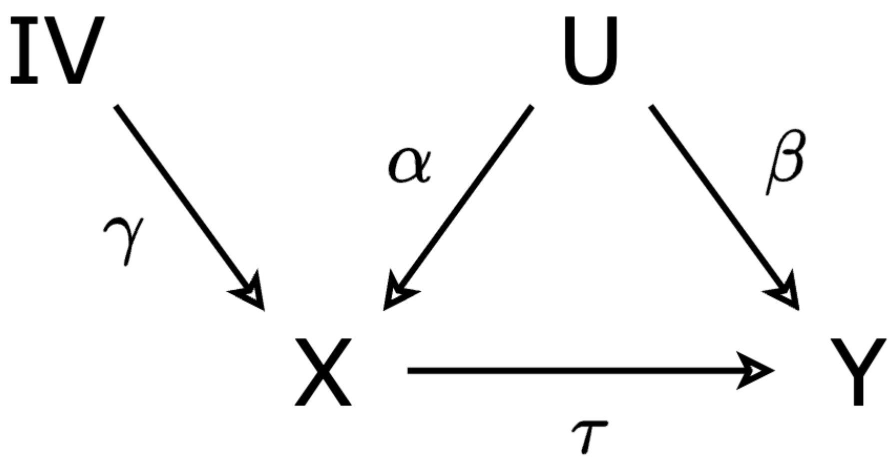
Steiner et al. (2017)



IV in Causal Graph

1. Effect of IV on Y

$$\frac{Cov(IV,Y)}{Var(IV)} = \frac{Var(IV)\gamma\tau}{Var(IV)} = \gamma\tau$$

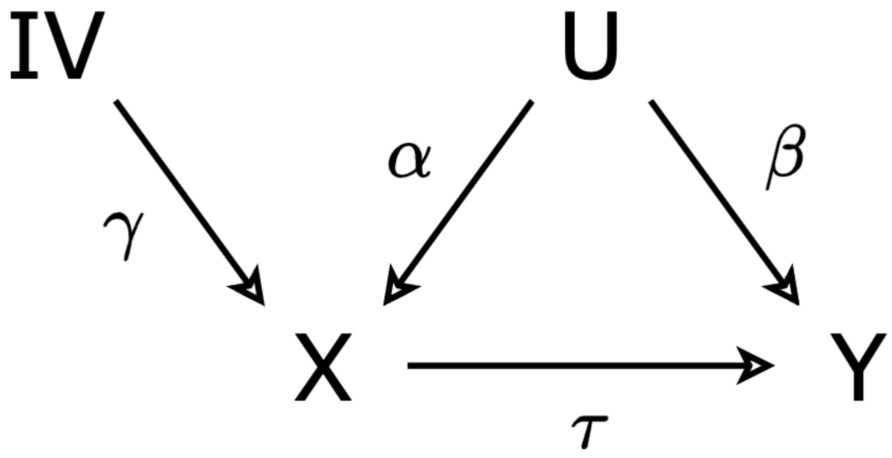


IV in Causal Graph

1. Effect of IV on Y: $\gamma\tau$

2. Effect of IV on X

$$\frac{Cov(IV,X)}{Var(IV)} = \frac{Var(IV)\gamma}{Var(IV)} = \gamma$$



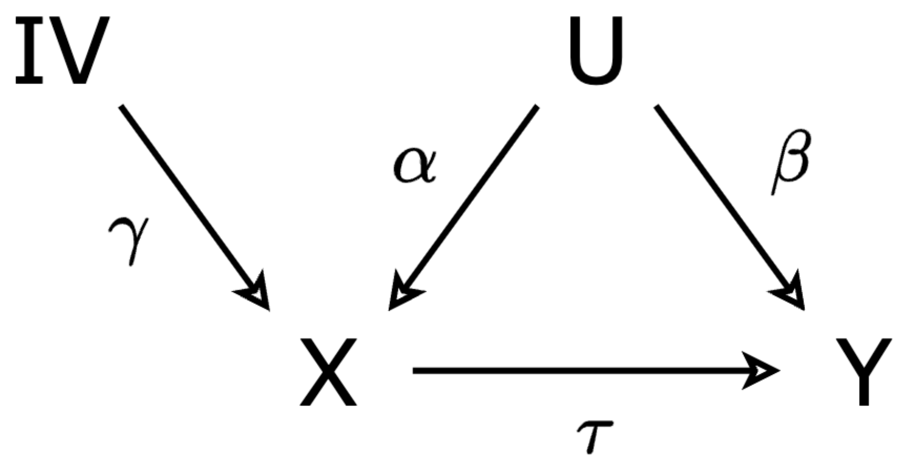
IV in Causal Graph

1. Effect of IV on Y: $\gamma\tau$

2. Effect of IV on X: γ

3. Ratio between the two effects

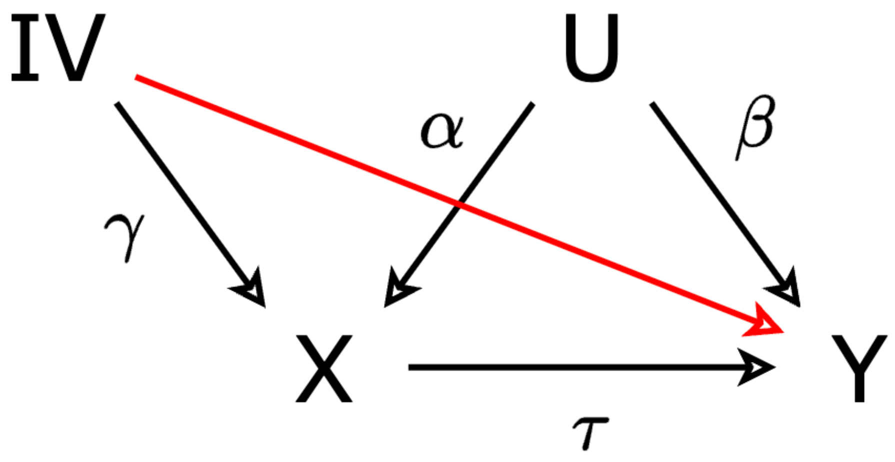
$$\frac{\gamma\tau}{\gamma} = \tau$$



Exogeneity Assumption

Instrument Exogeneity – Exclusion Restriction

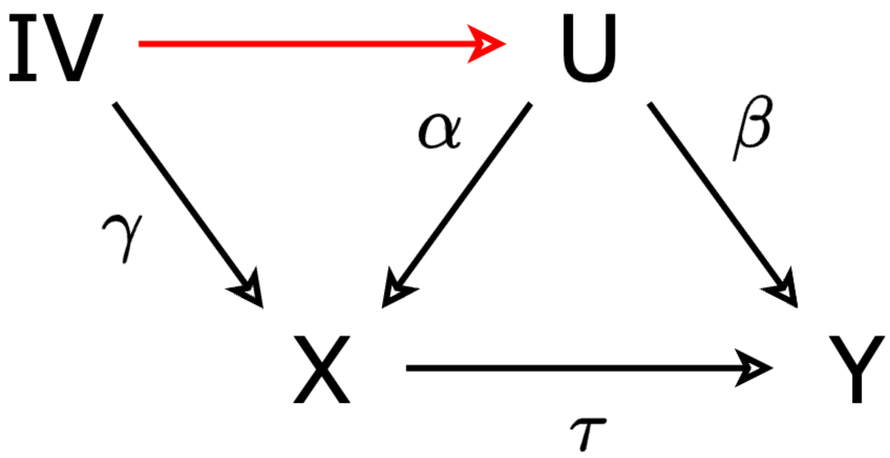
Wooldridge (2019)



IV should have no effect on Y,
after X and confounders have been controlled for

Exogeneity Assumption

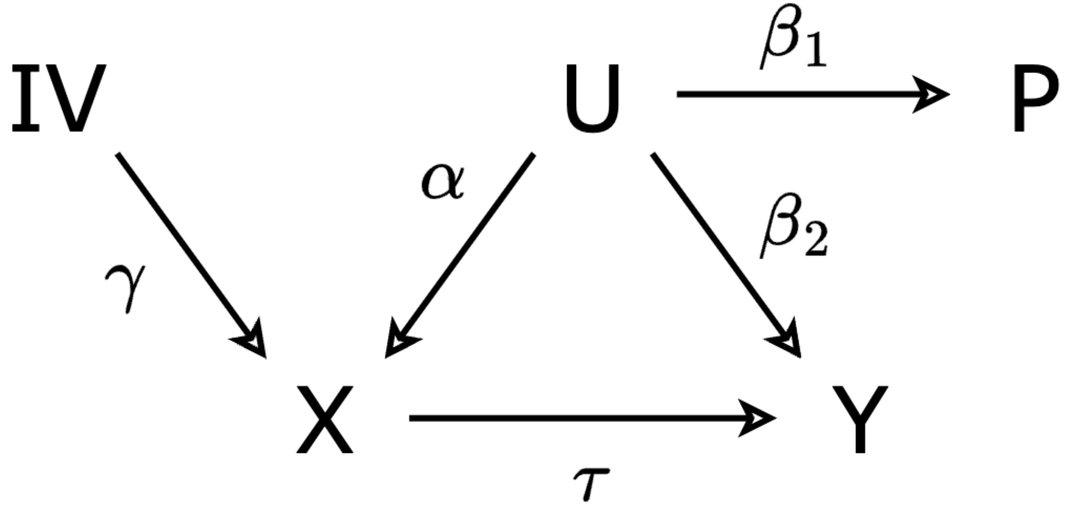
Instrument Exogeneity – Independence



IV should be uncorrelated with the confounders

IV in Pretest-Posttest Designs

Still, we can identify the causal effect of X on Y



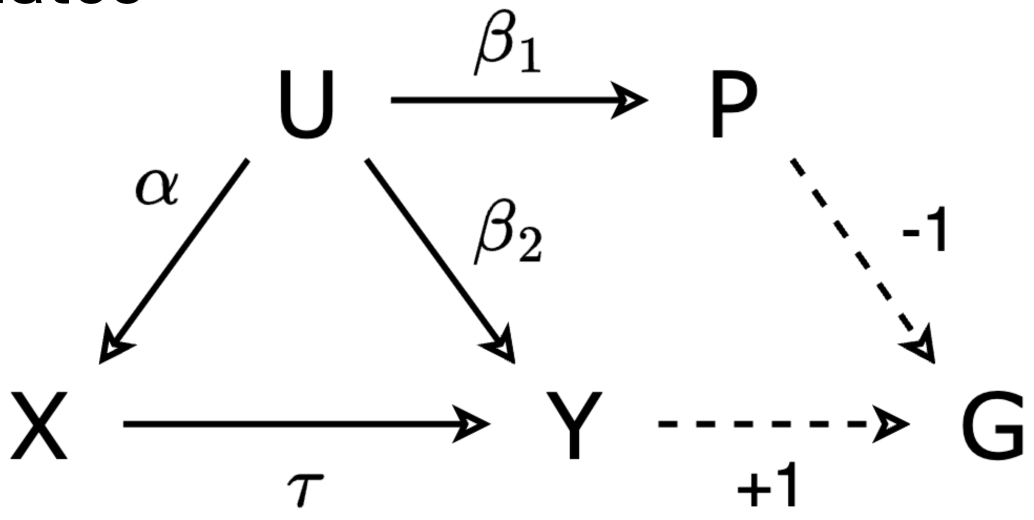
$$\frac{\text{Effect of } IV \text{ on } Y}{\text{Effect of } IV \text{ on } X} = \tau$$

DiD and Compass Variables

Kim & Steiner (2021)

We can identify the causal effect of X on Y, based on the common trend assumption ($\beta_1 = \beta_2$)

If the common trend assumption is violated ($\beta_1 \neq \beta_2$), we obtain biased estimates

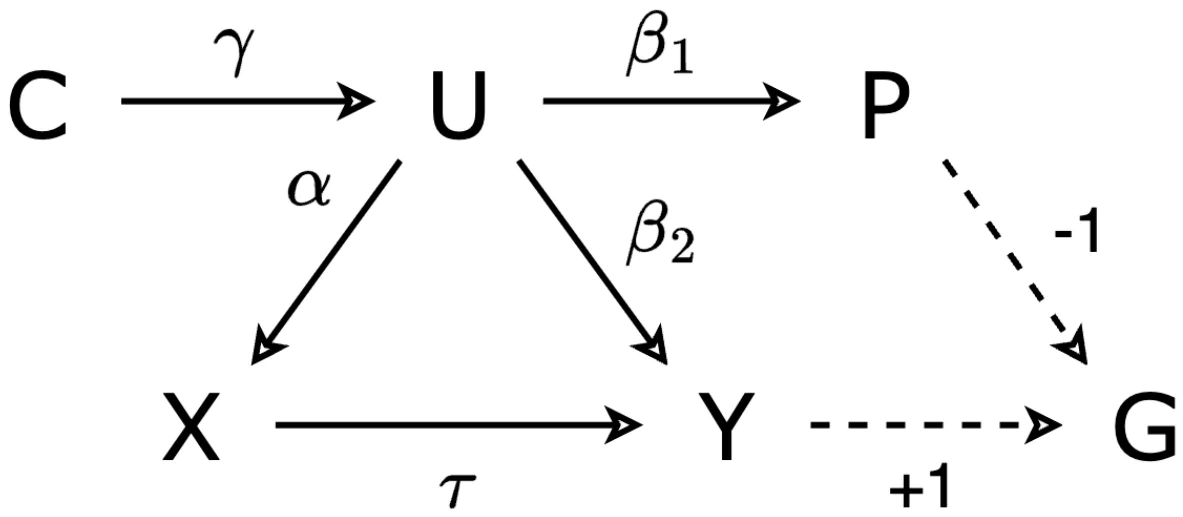


$$\hat{G} = a + bX,\ \ b = \tau + \alpha\beta_2 - \alpha\beta_1$$

DiD and Compass Variables

Compass Variable

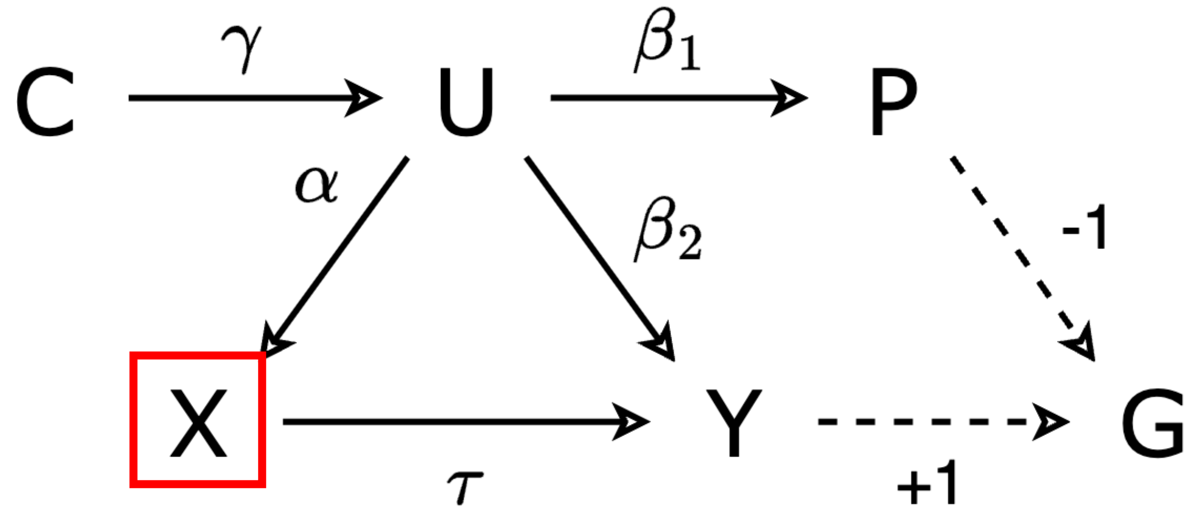
Kim, Gwak, & Lee (2022)



A variable that is associated with P and Y only via U

DiD and Compass Variables

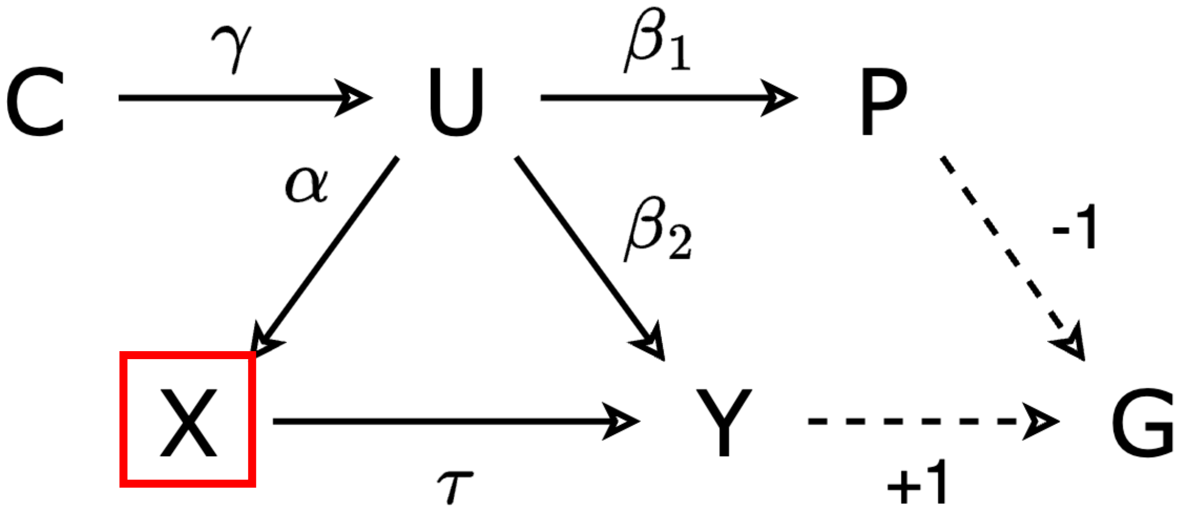
Quantify the difference between β_1 and β_2



$$\hat{P} = a_1 + b_1 C + c_1 X,\ b_1 = \gamma\beta_1 K$$

DiD and Compass Variables

Quantify the difference between β_1 and β_2

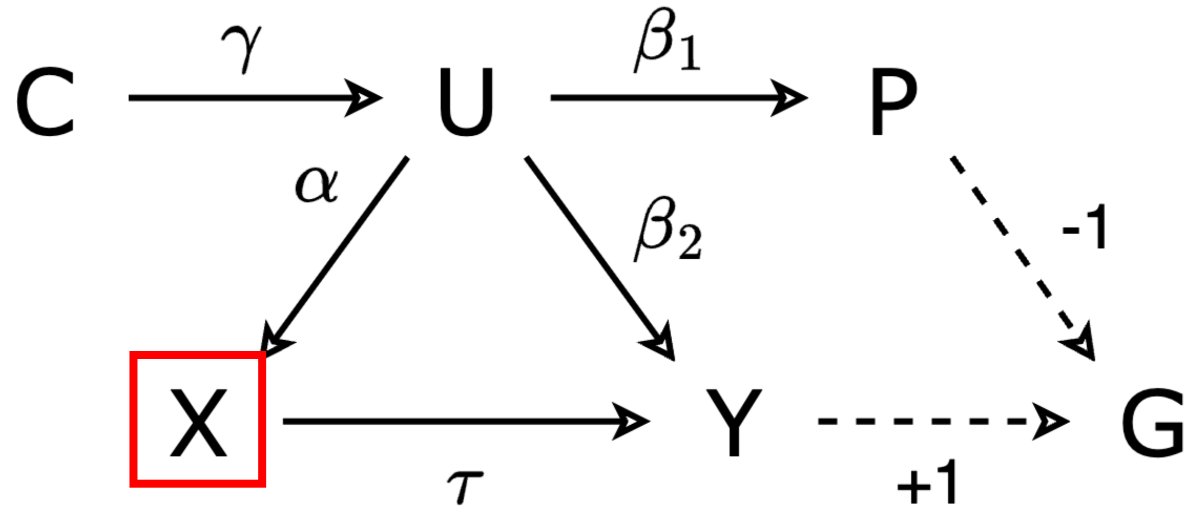


$$\hat{P} = a_1 + b_1 C + c_1 X, \; b_1 = \gamma \beta_1 K$$

$$\hat{Y} = a_2 + b_2 C + c_2 X, \; b_2 = \gamma \beta_2 K$$

DiD and Compass Variables

The difference between β_1 and β_2: $\delta = \frac{\beta_2}{\beta_1}$



$$\hat{P} = a_1 + b_1 C + c_1 X, \; b_1 = \gamma \beta_1 K$$

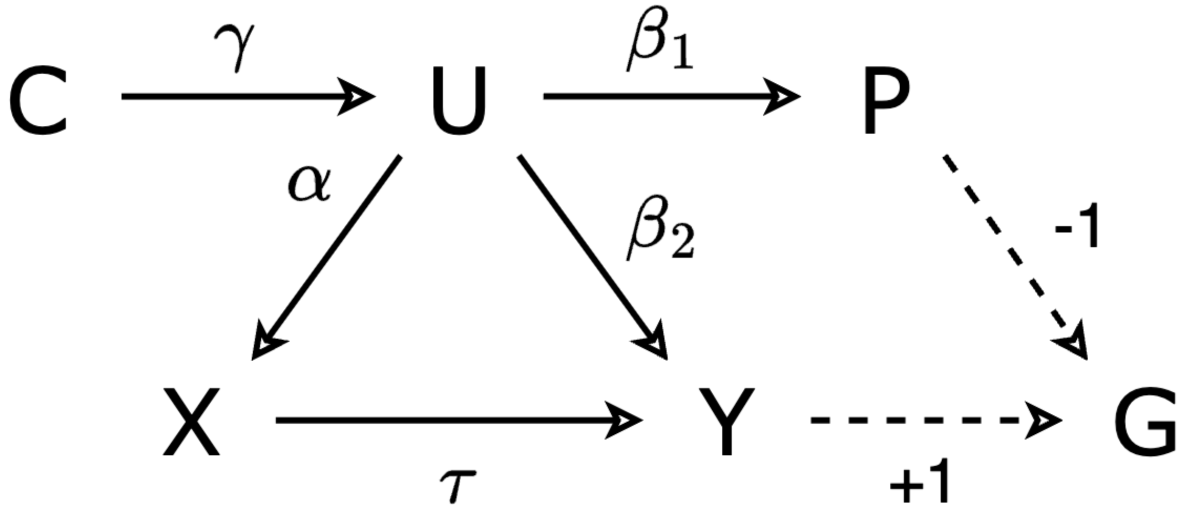
$$\hat{Y} = a_2 + b_2 C + c_2 X, \; b_2 = \gamma \beta_2 K$$

DiD and Compass Variables

Compass Variable

$$C \not\perp\!\!\!\perp U | X$$

$$C \perp\!\!\!\perp P | U, X \quad C \perp\!\!\!\perp Y | U, X$$

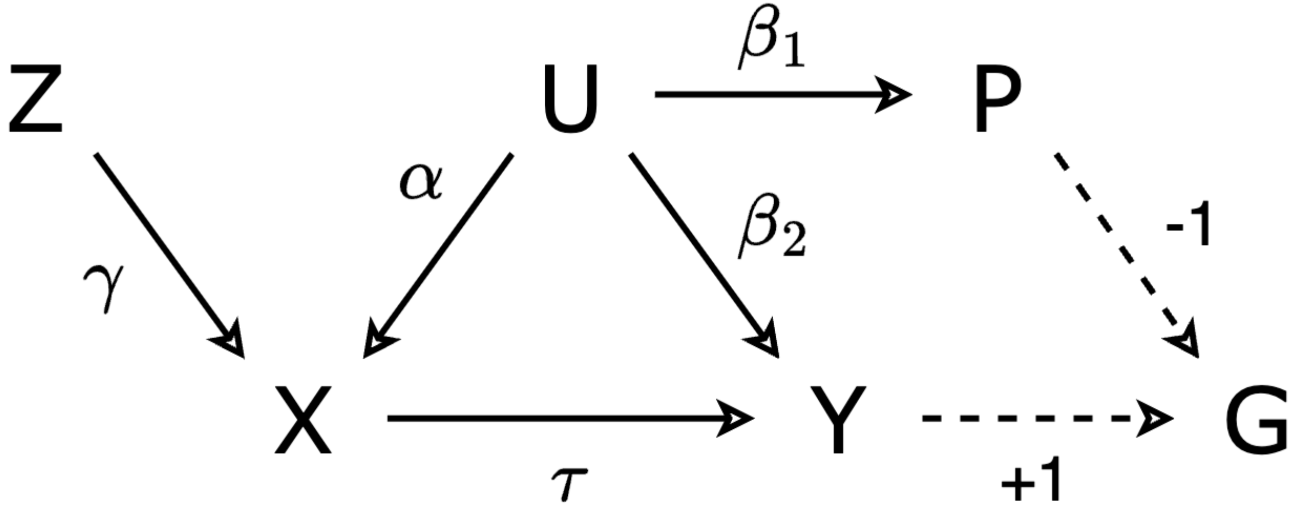


A variable that is associated with P and Y only via U

DiD and Compass Variables

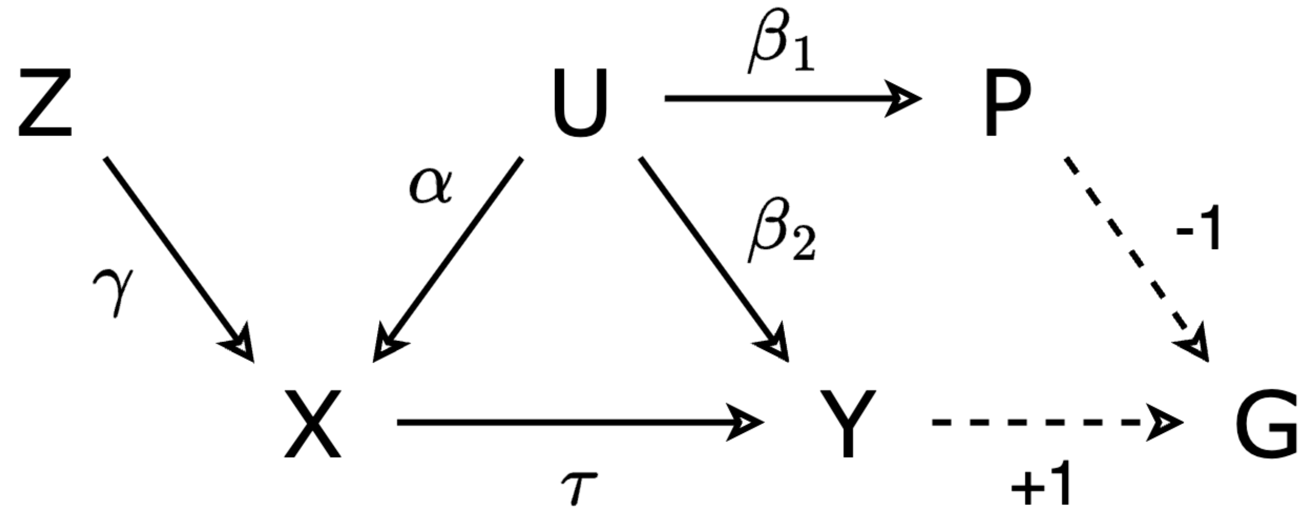
$$C \not\perp\!\!\!\perp U | X$$

$$C \perp\!\!\!\perp P | U, X \quad C \perp\!\!\!\perp Y | U, X$$



DiD and Compass Variables

An IV can also be a compass variable



Comparing IV Estimates to Adjusted DiD Estimates

The IV estimate when Z is a valid IV

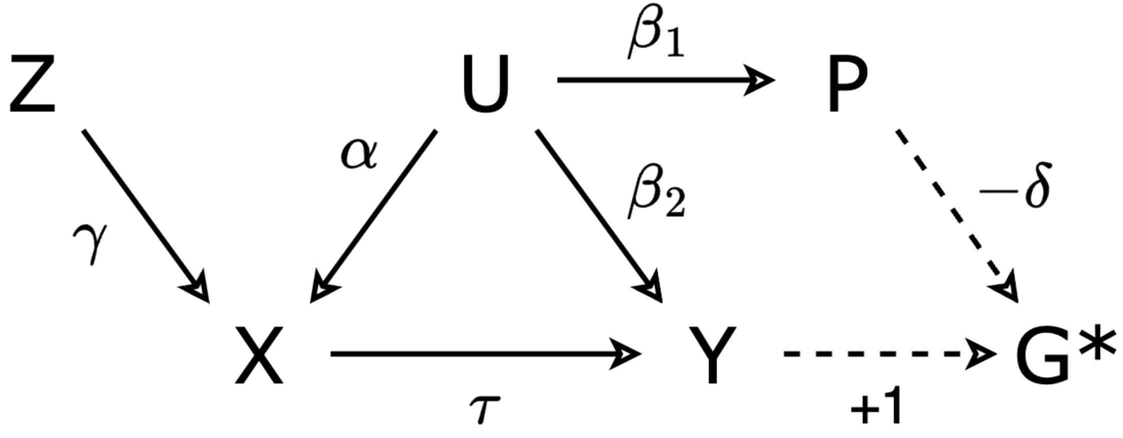
1) $\hat{X} = a_1 + b_1 Z$

   $(b_1 = \gamma)$

2) $\hat{Y} = a_2 + b_2 Z$

   $(b_2 = \gamma\tau)$

3) $\frac{b_2}{b_1} = \tau$



Comparing IV Estimates to Adjusted DiD Estimates

The adjusted DiD estimate when Z is a valid IV

1) $\hat{P} = a_1 + b_1 Z + c_1 X$

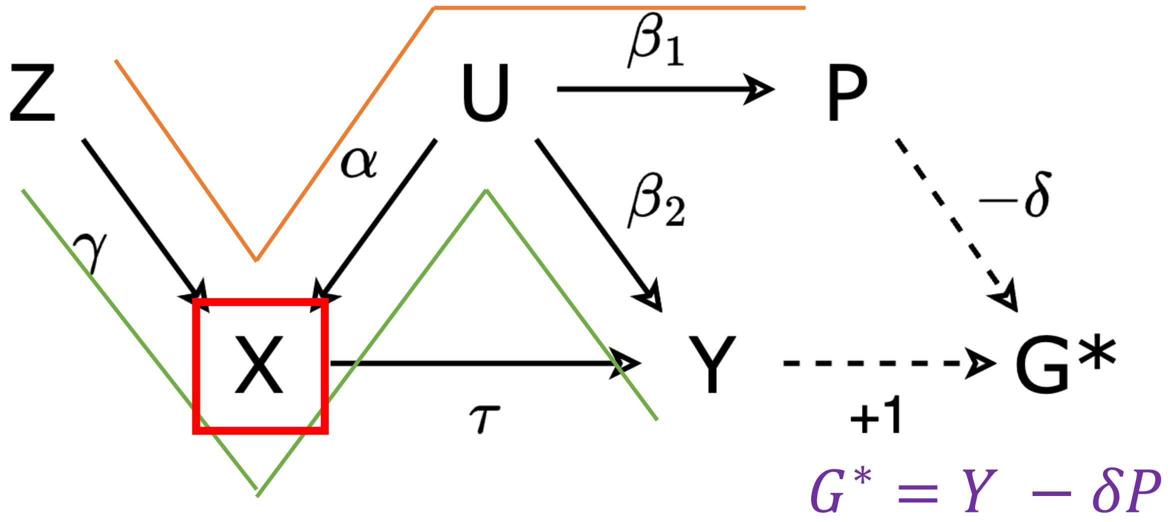
$b_1 = \gamma\alpha\beta_1$

2) $\hat{Y} = a_2 + b_2 Z + c_2 X$

$b_2 = \gamma\alpha\beta_2$

3) $\widehat{G^*} = a_3 + b_3 X$

$$b_3 = \tau + \alpha\beta_2 - \alpha\beta_1\delta$$
$$= \tau + \alpha\beta_2 - \alpha\beta_1 \frac{\beta_2}{\beta_1}$$
$$= \tau$$



Comparing IV Estimates to Adjusted DiD Estimates

If Z is a valid IV,

(Z is also a valid compass variable)

Then the IV estimate and the adjusted DiD estimate should equal

Comparing IV Estimates to Adjusted DiD Estimates

The IV estimate when Z is not a valid IV

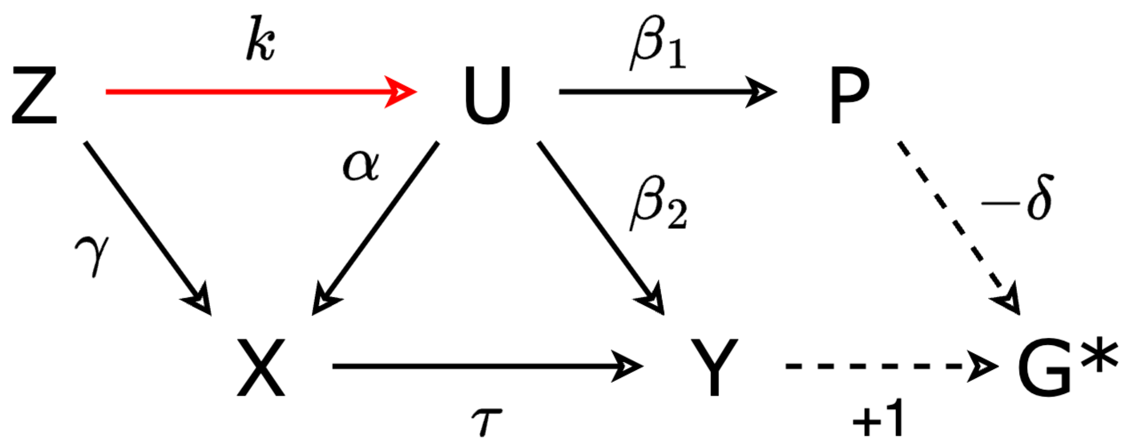
1) $\hat{X} = a_1 + b_1 Z$

   $(b_1 = \gamma + k\alpha)$

2) $\hat{Y} = a_2 + b_2 Z$

   $(b_2 = \gamma\tau + k\alpha\tau + k\beta_2)$

3) $\frac{b_2}{b_1} = \tau + \boxed{\frac{k\beta_2}{\gamma + k\alpha}}$ bias



Comparing IV Estimates to Adjusted DiD Estimates

The adjusted DiD estimate when Z is not a valid IV

1) $\widehat{P} = a_1 + b_1 Z + \ c_1 X$

$(b_1 = \ \gamma\alpha\beta_1 + k\beta_1)$

2) $\widehat{Y} = a_2 + b_2 Z + \ c_2 X$

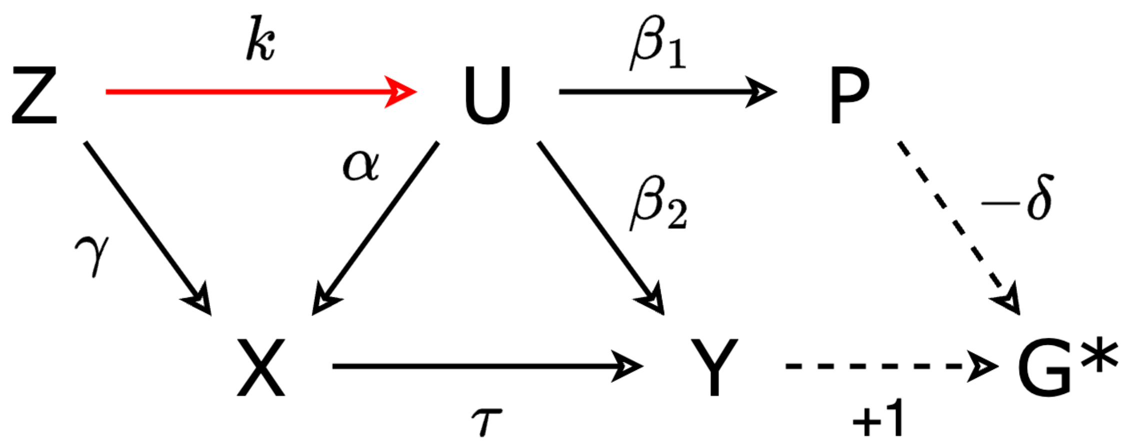
$(b_2 = \ \gamma\alpha\beta_2 + k\beta_2)$

3) $\widehat{G^*} = a_3 + b_3 X$

$$b_3 = \ \tau + \ \alpha\beta_2 \ - \alpha\beta_1\delta$$

$$= \ \tau + \ \alpha\beta_2 - \alpha\beta_1 \frac{(\gamma\alpha+k)\beta_2}{(\gamma\alpha+k)\beta_1}$$

$$= \ \tau$$



$G^* = Y \ - \delta P$

Comparing IV Estimates to Adjusted DiD Estimates

If Z is a valid instrumental variable,
(Z is also a valid compass variable)
Then the IV estimate and the adjusted DiD estimate should equal

If IV estimate and adjusted DiD estimate are different,
Then Z is not a valid instrumental variable

Comparing IV Estimates to Adjusted DiD Estimates

R simulation

- Data-generating model

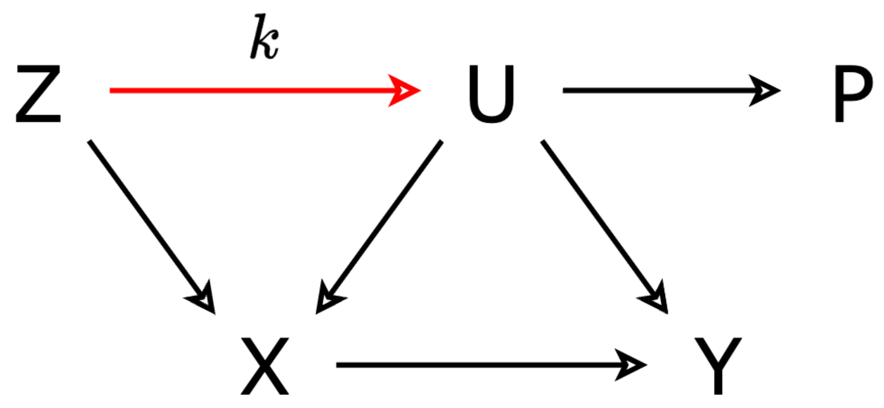
$$Z = \epsilon_Z$$

$$U = k \times Z + \epsilon_U$$

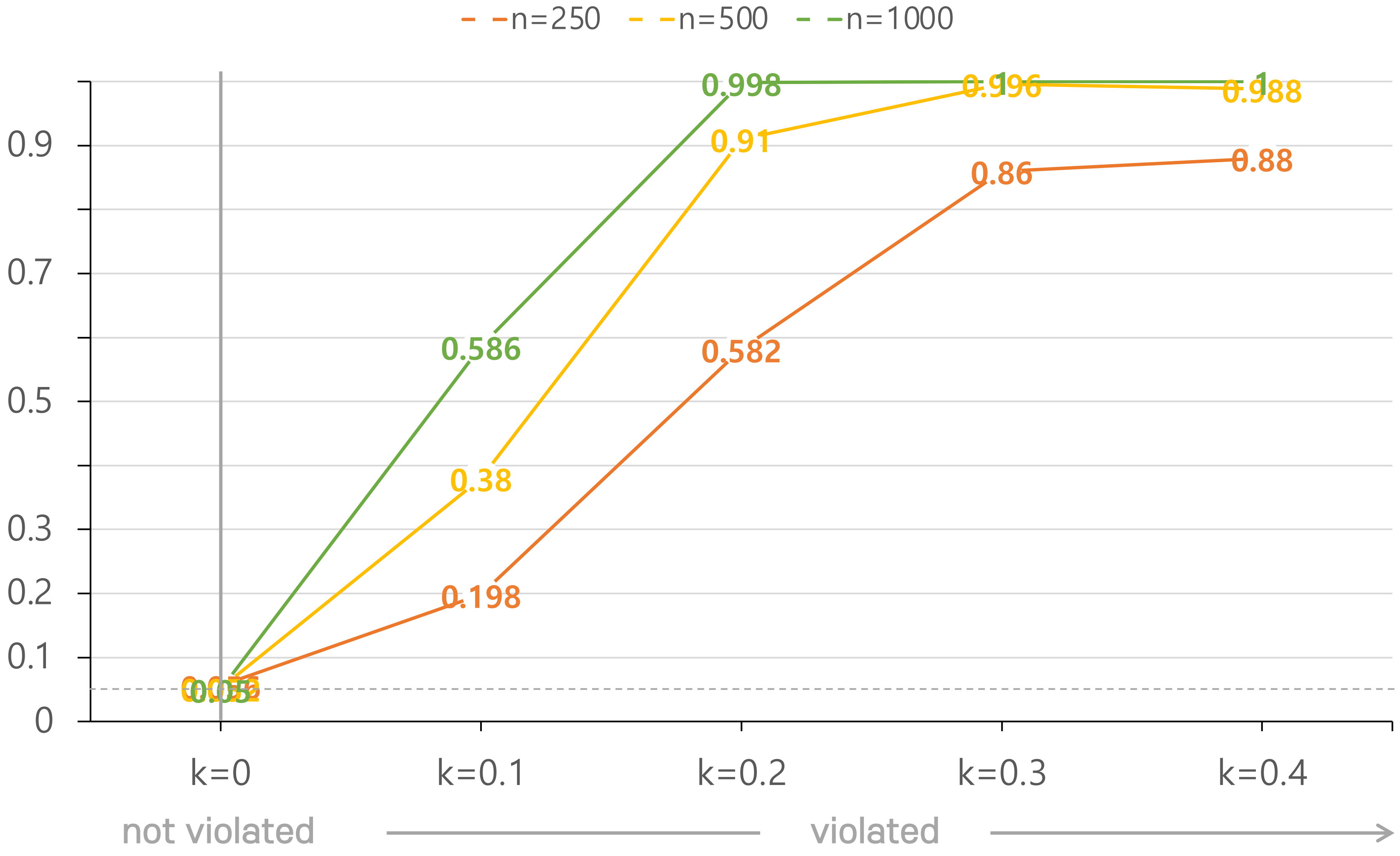
$$A = Z + U + \epsilon_A$$

$$P = U + \epsilon_P$$

$$Y = .777 \times X + 2 \times U + \epsilon_Y$$



- The possibility that IV estimates are different with adjusted DiD estimates



Discussion

- Tests empirically whether exogeneity is violated
- Encourages the conventional use of IV,
  especially by educational researchers

Discussion

- Tests empirically whether exogeneity is violated
- Encourages the conventional use of IV,
  especially by educational researchers

- Limitations
  - IV estimates and adjusted DiD estimates can both be biased
  - We cannot ascertain whether an IV is valid,
    but we can identify when it is invalid

References

- Kim, Y., Gwak, N., & Lee, S. (2022). Detection of and Correction for Violation of the Common Trend Assumption in Gain Score Analysis. *Journal of Education Evaluation, 35*(4), 743-761.
- Kim, Y. & Steiner, P. M. (2021). Gain scores revisited: A graphical models perspective. *Sociological Methods & Research, 50*(3), 1353-1375.
- Steiner, P. M., Kim, Y., Hall, C. E., & Su, D. (2017). Graphical Models for Quasi-experimental Designs. *Sociological Methods & Research, 46*(2), 155–188.
- Wooldridge, J. M. (2019). *Introductory econometrics: A modern approach*. (7th Ed.). Cengage Learning.