

# Instrumental Variables

Felipe Balcazar

NYU

May , 2023



NEW YORK UNIVERSITY

# Common sources of bias

- Reverse causality: Outcome cause the treatment - not the other way around.
- Omitted variables: existence of unobservable confounder.
- Non-random measurement error: measurement correlated with unobservables.

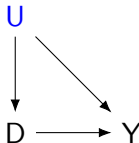
$$D \longrightarrow Y$$

- Reverse causality: Outcome cause the treatment - not the other way around.
- Omitted variables: existence of unobservable confounder.
- Non-random measurement error: measurement correlated with unobservables.

$$D \longleftrightarrow Y$$

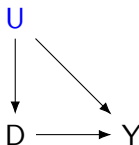
# Common sources of bias

- Reverse causality: Outcome cause the treatment - not the other way around.
- Omitted variables: existence of unobservable confounder.
- Non-random measurement error: measurement correlated with unobservables.



# Common sources of bias

- Reverse causality: Outcome cause the treatment - not the other way around.
- Omitted variables: existence of unobservable confounder.
- Non-random measurement error: measurement correlated with unobservables.



# What is an instrument?

An instrument is a variable  $Z$  that satisfies:

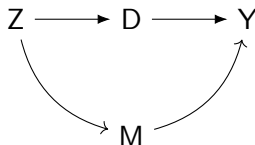
- **Relevance:** The instrument and the treatment need to be HIGHLY related. (No weak instrument!)

$$Z \longrightarrow D \longrightarrow Y$$

# What is an instrument?

An instrument is a variable  $Z$  that satisfies:

- **Relevance:** The instrument and the treatment need to be HIGHLY related. (No weak instrument!)
- **Exogeneity/exclusion restriction:** The instrument cannot affect the outcome through any other channel that is not the treatment.



# Two-stage least squares (2SLS)

- 1st stage we run the treatment against the instrument ( $Z$ ):

$$D_i = \alpha_1 + \beta_1 Z_i + \delta_1 X_i + \epsilon_1$$



# Two-stage least squares (2SLS)

- 1st stage we run the treatment against the instrument ( $Z$ ):

$$D_i = \alpha_1 + \beta_1 Z_i + \delta_1 X_i + \epsilon_1$$

- Then obtain the predicted values of the treatment, that is

$$\hat{D}_i = \alpha_1 + \beta_1 Z_i + \delta_1 X_i$$

F-test of excluded instrument needs to be greater than 10 (about 100 nowadays!).

# Two-stage least squares (2SLS)

- 1st stage we run the treatment against the instrument ( $Z$ ):

$$D_i = \alpha_1 + \beta_1 Z_i + \delta_1 X_i + \epsilon_1$$

- Then obtain the predicted values of the treatment, that is

$$\hat{D}_i = \alpha_1 + \beta_1 Z_i + \delta_1 X_i$$

F-test of excluded instrument needs to be greater than 10 (about 100 nowadays!).

- Using these values we run the regression we are interested in - the second-stage regression:

$$Y_i = \alpha_2 + \beta_2 \hat{D}_i + \delta_2 X_i + \epsilon_{i,2}$$

# What do instrumental variables do?

# Local Average Treatment Effect

$$Y_i = \alpha_2 + \beta_2 \hat{D}_i + \delta_2 X_i + \epsilon_{i,2}$$

$\beta$  is the Local Average Treatment Effect (LATE)!

$$Y_i = \alpha_2 + \beta_2 \hat{D}_i + \delta_2 X_i + \epsilon_{i,2}$$

$\beta$  is the Local Average Treatment Effect (LATE)!

- The effect on the compliers in regards to the instrument.

$$Y_i = \alpha + \beta_1 D_i + \beta_2 D_i \times X_i + \delta X_i + \epsilon_i$$

$$Y_i = \alpha + \beta_1 D_i + \beta_2 D_i \times X_i + \delta X_i + \epsilon_i$$

Then in your first stage you must estimate:

$$D_i = \alpha + \varphi_1 Z_i + \varphi_2 Z_i \times X_i + \tau X_i + \nu_i$$

$$Y_i = \alpha + \beta_1 D_i + \beta_2 D_i \times X_i + \delta X_i + \epsilon_i$$

Then in your first stage you must estimate:

$$D_i = \alpha + \varphi_1 Z_i + \varphi_2 Z_i \times X_i + \tau X_i + \nu_i$$

and

$$D_i \times X_i = \theta + \phi_1 Z_i + \phi_2 Z_i \times X_i + \pi X_i + \eta_i.$$



$$Y_i = \alpha + \beta_1 D_i + \beta_2 D_i \times X_i + \delta X_i + \epsilon_i$$

Then in your first stage you must estimate:

$$D_i = \alpha + \varphi_1 Z_i + \varphi_2 Z_i \times X_i + \tau X_i + \nu_i$$

and

$$D_i \times X_i = \theta + \phi_1 Z_i + \phi_2 Z_i \times X_i + \pi X_i + \eta_i.$$

- Do not forget to instrument your interaction!

# Theory is extremely important!

- Theory establishes if the instrument is relevant.

# Theory is extremely important!

- Theory establishes if the instrument is relevant.
- Theory establishes if we can believe the exclusion restriction!

# Theory is extremely important!

- Theory establishes if the instrument is relevant.
- Theory establishes if we can believe the exclusion restriction!
- Gives us a sense of who are the compliers and how important our results are.

# Theory is extremely important!

- Theory establishes if the instrument is relevant.
- Theory establishes if we can believe the exclusion restriction!
- Gives us a sense of who are the compliers and how important our results are.
- Gives us a sense of the bias that the instrument is addressing.

# Frequently asked questions

- Using an instrumenting someone has used before, but for a different outcome
  - Is this a violation of the exclusion restriction?
  - No, if an only if is path of the causal path.

# Frequently asked questions

- Using an instrumenting someone has used before, but for a different outcome
  - Is this a violation of the exclusion restriction?
  - No, if an only if is path of the causal path.
- Should I estimate the 2SLS by hand?
  - No! Your inference will be wrong! Use ivreg2.
  - Moulton problem - usually SE are underestimated.

- Using an instrumenting someone has used before, but for a different outcome
  - Is this a violation of the exclusion restriction?
  - No, if an only if is path of the causal path.
- Should I estimate the 2SLS by hand?
  - No! Your inference will be wrong! Use `ivreg2`.
  - Moulton problem - usually SE are underestimated.
- Can I use more than one instrumental variable?
  - Yes! However finding one instrument is already difficult enough.



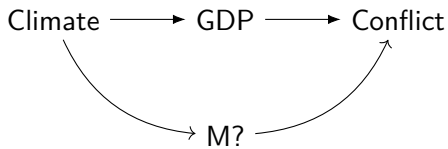
# Effect of GDP on conflict

GDP → Conflict

# Effect of GDP on conflict

Climate → GDP → Conflict

# Effect of GDP on conflict

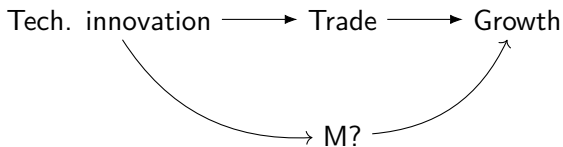


# Effect of trade on Growth

Trade  $\longrightarrow$  Growth

Tech. innovation  $\longrightarrow$  Trade  $\longrightarrow$  Growth

# Effect of trade on Growth



## Temporary page!

$\text{\LaTeX}$  was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it. If you rerun the document (without altering it) this surplus page will go away, because  $\text{\LaTeX}$  now knows how many pages to expect for this document.