Public Policy Evaluation
Lecture 3: Instrumental variable

Yao Thibaut Kpegli

Licence 3 in Economics, ENS Paris-Saclay
2023–2024

March 2024

Outline

Context

- Unobserved heterogeneity: unobserved characteristics that explain both participation in the program and the outcome variable
- Double difference cannot be used:
  - heterogeneity is not fixed
  - no panel data
- Instrument(s): availability of exogenous sources of variation that explain the treatment

Context

Angrist et al,.(1996).

- effect of veteran status in the Vietnam era on mortality after conflicts
  - risk-attitudes explain both veteran status and mortality
- risk-attitudes explain veteran
  - Always-taker: "risk-lovers" who seek adventure and want to fight
  - Never-taker: "prudents and pacifists" who do everything to avoid having to fight
  - Compliers: "conformists" who will fight if they are compelled to do so but won't if they can avoid it
- risk-attitudes explain mortality
- "risk-takers" take more risks (drink, smoke, drive faster, etc.), and consequently have a higher mortality risk than "never-taker"

Context

- Some part of the variation of veteran status comes from endogenous source (self selection of always-taker and never-taker)
- "Compliers" are the basis of the identification
- Among the sub-population of compliers, any difference in the average mortality rate observed between veterans and others can be attributed to veteran status itself.

Context

- For each year of birth, random allocation of a number between 1 and 365 to each day of birth
- Army decided the proportion of recruits for each year of birth.
- This proportion determines the Veteran status
- This mechanism constraints some compliers to participate
- the assigned number explains veteran status but has no direct link with mortality – what we call instrument

Context

$$\boldsymbol{Y}_i = \alpha + \boldsymbol{X}_i\beta + \gamma\boldsymbol{T}_i + \epsilon_i$$

- value of the variable of interest for individual i.
- indicator taking the value 1 if the individual is "treated" and 0 otherwise
- $\boldsymbol{X}_i$ set of observable characteristics of the individual (education, age, gender, etc.).
- error term that measures everything that influences the value of $\boldsymbol{Y}_i$ and that we do not observe.
- γ : Effect of treatment
- Problem: γ is biased if $\boldsymbol{T}_i$ is endogenous, i.e. $\text{cov}(\boldsymbol{T}_i, \epsilon_i) \neq 0$

Context

- Omitted variable: an unobservable variable causes both treatment status and outcome variable
  - risk-attitudes in our example of the impact of veteran status on mortality – self-selection
- Reverse causality: treatment status is determined by the value of the outcome variable
  - excellence scholarship status and academic outcomes – policy placement
- Measurement errors: treatment status is measured with errors, some treated individuals are considered untreated, and vice versa

Outline

Instrumental variable

Instrumental variable is a remedy for endogeneity regardless of its source

Instrumental variable: exogenous source of variation that influences the probability of being treated without a direct relationship with the variable of interest

Instrumental variable

- Relevance: $\mathbf{Z}$ affects the treatment $\mathbf{T}$

$$Cov(T, Z) \neq 0 \tag{1}$$

- Exclusion: $\mathbf{Z}$ is independent of the unobserved component of the potential outcomes ϵ

$$Cov(Z, \epsilon) = 0 \tag{2}$$

- IV estimator:

$$\gamma_{IV} = \frac{Cov(Y, Z)}{Cov(T, Z)}$$

Principle

1. $\boldsymbol{Z}_i$ value of the instrument $\boldsymbol{Z}$ for individual i
   - the number assigned to the day of birth of individual
2. Extract variations of $\boldsymbol{T}_i$ that come from $\boldsymbol{Z}_i$
   - this exogenous part captures random assignments of "compliers" to treatment
3. Use this exogenous part of the variation of $\boldsymbol{T}_i$ to evaluate the impact

This procedure is known as Two-Stage Least Square (2SLS)

IV- 2SLS

1. Stage 1: OLS regression

$$\boldsymbol{T}_i = a + b\boldsymbol{Z}_i + u_i \tag{3}$$

2. Stage 2: OLS regression

$$\boldsymbol{Y}_i = \alpha + \beta\widehat{\boldsymbol{T}}_i + e_i \tag{4}$$

with $\widehat{\boldsymbol{T}}_i = \widehat{a} + \widehat{b}\boldsymbol{Z}_i$

Controls

1. Stage 1: OLS regression

$$\boldsymbol{T}_i = a + b\boldsymbol{Z}_i + \boldsymbol{X}_i c + u_i \tag{5}$$

2. Stage 2: OLS regression

$$\boldsymbol{Y}_i = \alpha + \beta\widehat{\boldsymbol{T}}_i + \boldsymbol{X}_i\gamma + e_i \tag{6}$$

with $\widehat{\boldsymbol{T}}_i = \widehat{a} + \widehat{b}\boldsymbol{Z}_i + \boldsymbol{X}_i\widehat{c}$

Limits

1. Vulnerability to Weak instrument: $Cov(T, Z)$ small

$$\gamma_{IV} = \frac{Cov(Y, Z)}{Cov(T, Z)}$$

2. Local Average Treatment Effect (LATE): the evaluation is only valid for those who are treated due to the exogenous origin represented by $\boldsymbol{Z}_i$ ("compliers")

Limits - more on LATE

- First stage function $T_i(Z_i)$ that assigns the value of the treatment $T_i \in \{0, 1\}$ to the value of the instrument $Z_i \in \{0, 1\}$
- Four instrumental populations :
  - Compliers: $T_i(Z_i) = Z_i$
  - Never-takers: $T_i(Z_i) = 0$
  - Always-takers: $T_i(Z_i) = 0$
  - Defiers: $T_i(Z_i) = 1 - Z_i$

	T=1 | Z=1	T=0 | Z=1
T=1 | Z=0	Always-takers	Defiers
T=0 | Z=0	Compliers	Never-takers

Limits - more on LATE

- Z independent to potential outcomes leads to:

$$\begin{aligned} E[Y_i|Z_i=1] &= E[Y_{0i}+T_i(Y_{1i}-Y_{0i})|Z_i=1] \\ &= E[Y_{0i}+T_i(1)(Y_{1i}-Y_{0i})] \end{aligned}$$

$$\begin{aligned} E[Y_i|Z_i=0] &= E[Y_{0i}+T_i(Y_{1i}-Y_{0i})|Z_i=0] \\ &= E[Y_{0i}+T_i(0)(Y_{1i}-Y_{0i})] \end{aligned}$$

- Difference in outcome

$$\begin{aligned} E[Y_i|Z_i=1]-E[Y_i|Z_i=0] &= E[\big(T_i(1)-T_i(0)\big)(Y_{1i}-Y_{0i})] \\ &= E[Y_{1i}-Y_{0i}|T_i(1)-T_i(0)=1] \\ &\quad + \\ &= E[-(Y_{1i}-Y_{0i})|T_i(1)-T_i(0)=-1] \end{aligned}$$

- Only compliers (assuming no defiers) contributes to the difference:

$$E[Y_i|Z_i=1]-E[Y_i|Z_i=0] = E[Y_{1i}-Y_{0i}|T_i(1)-T_i(0)=1]$$

Outline

Application

The impact of building secondary schools on the primary education level of children

Thank you for your attention !