Instrumental variable

Motivation: i.e. wage v.s. education.

OLS (Biased.) $\log(\text{wage}) = \beta_0 + \beta_1 ed + \beta_2 exp + \beta_3 exp^2 + \varepsilon$ ← pure noise

Endogenous. is $E(\varepsilon | ed, exp) = 0$? maybe not.

⇒ unobser vable corr with exp & ed.

- intelligence
- persistence

] examples

⇒ $E(\hat{\beta}_1] \neq \beta_1$

Basic idea

IV is independent from ε.

↓ effect

(ed) Indepent varibles ↓ (wage) bependent variables. ✗ $z \perp \varepsilon$.

example: ① Distance to nearest college. (lost d)

② quater of Birth

Requirements

1. Must be corr. with endogenous variable
2. Must only affect the dependent variable through its ~~effect~~ on the endogenous variable. (No direct effect)

ed — part1 → 1. corr. with ε (Bad)

ed — part2 → 2. corr. IV. (Not ε) ← Allow to only use this.

ed — part3 → 3. corr. everything else

Two stage least squares.

stage 1: $ed = \gamma_0 + \gamma_1 z + \gamma_2 exp + \gamma_3 exp^2 + u$

predict $\hat{ed} = \hat{\gamma}_0 + \hat{\gamma}_1 z + \hat{\gamma}_2 exp + \hat{\gamma}_3 exp^2$

- $\hat{ed}$ is unbaised estimate of ed.
- $\hat{ed}$ uncorrelated with ε.

stage 2: $\log(wage) = \beta_0 + \beta_1 \hat{ed} + \beta_2 exp + \beta_3 exp^2 + v$

($\hat{ed}$: unbais.)

Problem: weak IV.

- corr of IV and endogenous variable must be sufficiently strong.
- F-Test:
  - null hypothesis that the coff. on the IV in the 1st stage are 0.
  - $F \sim 12$ (at least 10).

More

① Mulipe IV → sargan test / wu-Hausman test

② IV does not give you the ATE.
↳ local ATE.

③ 2SLS ⇒ estimation error in 1st stage ⇒ SE wrong in 2nd