IV Examples

Yuta Toyama

May 24th, 2019

Example 1: Wage regression

- Consider a regression model

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \underbrace{\beta_3 ability_i + u_i}_{\equiv \epsilon_i}$$

- $ability_i$
  - Not observed in data → part of error term ϵ_i
  - Correlated with $educ_i$
  - Leads to endogeneity issue. Biased estimates of β

- How to get rid of bias of β_1 ?

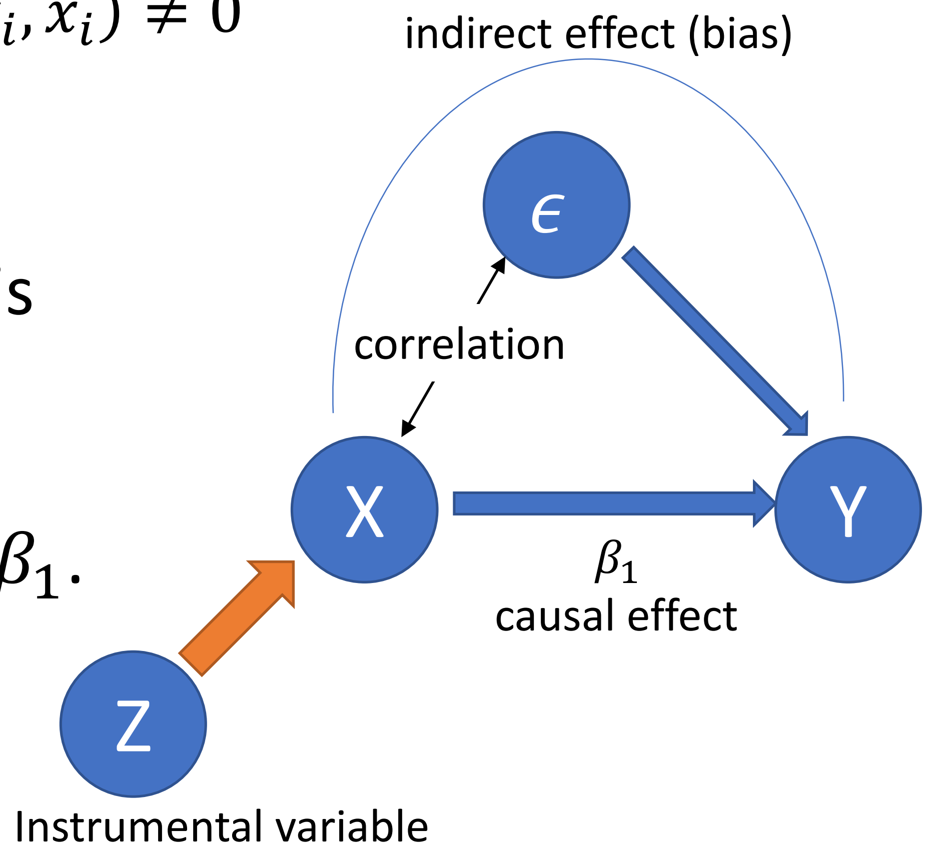
  → Use instrumental variable estimator!

Idea of Instrumental Variable (IV)

- Consider $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
  - x_i : endogenous variable. $Cor(x_i, \epsilon_i) \neq 0$
- Instrumental variable (IV) z_i
  - $Cor(z_i, \epsilon_i) = 0$ and $Cor(z_i, x_i) \neq 0$

Idea:

- Pick variation of x_i that is explained by z_i.
- Use this variation to explain y_i and estimate β_1.



Two Conditions for IV and Example

1. **Independence**: Uncorrelated with error term
2. **Relevance**: Correlated with endogenous variable

- Example:

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \epsilon_i$$

  - $educ_i$ is correlated with ϵ_i (through unobserved ability).

- IV: father's education $fathereduc_i$
  - Worker's ability (ϵ_i) is affected by her education, not her father's education. → uncorrelated with ϵ_i.
  - More educated father is likely to invest in education of his children. → correlated with $educ_i$

Example: Wage regressions (MROZ.dta)

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \epsilon_i$$

IV: father's education ($fathereduc_i$)

- OLS might have an upward bias:
  - Education level is positively correlated with ability.

- OLS: 1 additional year of schooling → 11% increase in wage.
- IV: 7.5% increase in wage. IV helps to eliminate the upward bias.

Dependent variable: log(wage)

	OLS	IV
educ	0.109	0.075
	(0.013)	(0.036)
exper	0.016	0.016
	(0.004)	(0.004)
constant	-0.400	0.036
	(0.183)	(0.466)
Observations	428	428
R-squared	0.15	0.14

Example 2: Measurement Error

- Consider the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ x_i^* = x_i + v_i$$

- We observe (y_i, x_i^*)
- The regression equation

$$y_i = \beta_0 + \beta_1 x_i^* + (\epsilon_i - \beta_1 v_i)$$

- Consider the second measurement with error

$$z_i = x_i^* + u_i$$

where u_i is a classical measurement error. z_i can be used as an IV.

- Example: Ashenfelter and Krueger (1994). Each twin was asked about his or her sibling's years of education: a second measure that can be used as an IV for self-reported education.

Example 3: Demand and Supply (**Simultaneous equation**)

- Let t be an index for "market" (geographic and/or time)
- Demand equation:

$$q_t = \alpha_0 + \alpha_1 p_t + \alpha_2 Y_t + \epsilon_t^d$$

where Y_t is **demand shifter** (GDP, income,etc)

- Supply equation:

$$q_t = \beta_0 + \beta_1 p_t + \beta_2 w_t + \epsilon_t^s$$

where w_t is **cost shifter** (oil price, wage, etc..)

- Y_t as an IV for p_t in supply equation
- w_t as an IV for p_t in demand equation

Demand estimation from Ryan (2012, Econometrica)

- Estimate the demand for cement in market j in year t

$$\log(Q_{jt}) = \alpha_0 + \alpha_1 \log(P_{jt}) + \alpha_3 X_{jt} + \epsilon_{jt}$$

  - Panel data!
- IV: wage, electricity price, coal price, gas price
- Elasticity is under-estimated in OLS.

	[1]		[2]		[3]		[4]	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
log(price)	-1.04	0.21	-2.69	0.31	-0.66	0.14	-1.77	0.24
Log(population)					0.44	0.03	0.38	0.03
Constant	12.15	0.85	19.07	1.29	3.63	0.72	9.20	1.28
Method	OLS		IV		OLS		IV	
Sample size	483		483		483		483	

1^{st} stage regression

- Regress endogenous variable on IV and exogenous variables.

```
. reg logp gas96 wage96 elec96 coal96 logpop, robust

Linear regression                               Number of obs     =        483
                                                F(5, 477)         =      35.28
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2844
                                                Root MSE          =     .16336

------------------------------------------------------------------------------
             |               Robust
        logp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gas96 |     .03786   .0036738    10.31   0.000     .0306411    .0450789
      wage96 |   .0013112   .0023254     0.56   0.573    -.003258     .0058804
      elec96 |  -.2668457   .0385111    -6.93   0.000    -.3425181   -.1911733
      coal96 |   .0329132    .004789     6.87   0.000     .0235029    .0423234
      logpop |  -.0157024   .0116415    -1.35   0.178    -.0385774    .0071726
       _cons |   4.807489   .1881122    25.56   0.000     4.437858     5.17712
------------------------------------------------------------------------------

. test  gas96 = wage96 = elec96 = coal96 = 0

 ( 1)  gas96 - wage96 = 0
 ( 2)  gas96 - elec96 = 0
 ( 3)  gas96 - coal96 = 0
 ( 4)  gas96 = 0

       F(  4,   477) =   35.96
            Prob > F =    0.0000
```