Tutorial, Part 10 — Instrumental Variables

Instrumental variables in case of: $y = \beta_1 + \beta_2 x + u$

- simultaneous causality
- unobserved (omitted) variables
- measurement error bias
- ⋮

because of $corr(x, u) \neq 0$ in such cases, which causes inconsistent estimates of all parameters.

Instrumental variable Z conditions:

- exogeneity: $Cov(Z, u) = 0$
- relevance: $Cov(Z, X) \neq 0$

Example 1

a) structural model

$$\hat{Q} = a_1 + a_2 XPER + a_3 CAP + a_4 LAB$$ (Exclusion restriction)

b) (Wu-)Hausman test

Reduced model: instrument

$$\widehat{XPER} = d_1 + d_2 AGE + d_3 CAP + d_4 LAB$$

↳ e_{RM} (residuals from reduced model)

$$\hat{Q} = g_1 + g_2 XPER + g_3 CAP + g_4 LAB + g_5 e_{RM}$$

$H_0: \gamma_5 = 0$ (XPER is exogenous)

$H_1: \gamma_5 \neq 0$ (XPER is endogenous)

(Stata)

t-test : $t = -2.20$, $p = 0.031$ (< 0.05)

→ Reject H_0 and conclude from H_1 ⇒

Variable XPER is endogenous, OLS estimator from part a) is INCONSISTENT.

c) Instrument condition

- <u>relevance</u> : are XPER and AGE related?

e.g. reduced model ⇒ $H_0: \delta_2 = 0$ $H_1: \delta_2 \neq 0$

(Stata) P-value : 0.003 (< 0.05) → reject H_0

d) IV estimator

2SLS (2-stage least squares)

(built-in Wu-Hausman test: same P-value, $F = t^2$)

Very different result compared to using the instrumental variable instead of the instrumented one.

e) IV estimator manually

Reduced model: $\widehat{XPER} = d_1 + d_2\, AGE + d_3\, CAP + d_4\, LAB$

↳ fitted values into structural model instead of XPER

↳ $\hat{Q} = b_1 + b_2\, \widehat{XPER}_{RM} + b_3\, CAP + b_4\, LAB$

f) IV regression coefficient - manually for a bivariate case

Structural model: $\hat{Q} = a_1 + a_2\, XPER$

Reduced model: $\hat{XPER} = d_1 + d_2\, AGE$

$$\beta_{IV} = \frac{Cov(Y, Z)}{Cov(X, Z)}$$

$$b_{IV} = \frac{Cov(Q, AGE)}{Cov(XPER, AGE)} = \frac{9.884}{18.075} = 0.547 \qquad \neq a_2$$

$\neq c_2$ from $\hat{Q} = c_1 + c_2\, AGE$

$$\hat{Q} = 2.044 + 0.547\, \hat{XPER}_{RM}$$

g) Sargan test for extra instruments

if more than one instrument for an instrumented variable

IV regression: $Q = \beta_1 + \beta_2\, \hat{XPER} + \beta_3\, CAP + \beta_4\, LAB + u$

($\hat{XPER}$ from reduced model:

$\hat{XPER} = d_1 + d_2\, AGE + d_3\, AGE^2 + d_4\, CAP + d_5\, LAB$)

residuals from IV regression $\rightarrow \hat{u}$

$$\hat{u} = c_1 + c_2\, CAP + c_3\, LAB + c_4\, AGE + c_5\, AGE^2$$

H_0: exogenous variables are not correlated with the IV disturbance term

H_1: — " — are correlated — " —

↳ at least one instrument is endogenous (= invalid)

if fail to reject H_0: We have some confidence in this choice of instruments.

Sorgan test : $LM = n \cdot R^2$ (from $\hat{u}$ regression)

$= 75 \cdot 0.0255 =$

$= 1,91 \sim \chi^2_{(m = \#\text{ extra instruments})}$

$\chi^2_{c(m=1,\ \alpha=0.05)} = 3.84$

$LM < \chi^2_c \Rightarrow$ cannot reject $H_0 \Rightarrow$ We can have some confidence in this choice of instruments.

($P = 0.167 > 0.05$)

Example 2

a) • simultaneous causality ?

more kids → less time to work

more work → less time for kids

• or: more kids → need more money → work more

b) not consistent

c) Instrument condition :

– relevance (binary variables → χ^2-test for frequencies' goodness-of-fit) : $P = 0.000 < 0.05$

H_0: no association ⟶ reject H_0

SAMESEX can't be affected by working more or less. SAMESEX influences WEEKSM1 only through MOREKIDS.

d) Not the same results as in part b).

$$\widehat{WEEKSM1} = 21{,}421 - 6.314\ \widehat{MOREKIDS}$$
(binary)

Interpretation: Mothers with more than 2 kids worked on average by 6.3 weeks less compared to mothers with only 2 kids in 1979.

e) Manually : same coefficients, different $se(b_j)$

f) Yes.

Tutorial, Part 11

Stationarity, ARMA models

Example 1

Unit-root

$$y_t = \rho\, y_{t-1} + N_t \quad / - y_{t-1}$$

$$\Delta y_t = \underbrace{(\rho - 1)}_{\gamma}\, y_{t-1} + N_t$$

$\Rightarrow$ solution („root") for $\gamma = 0$

$\Rightarrow \rho = 1$ („unit")

$\rho = 1 \Rightarrow y_t$ is a non-stationary process

$\rho < 1 \Rightarrow y_t$ is a stationary process

($\rho > 1 \Rightarrow y_t$ is an explosive process $\rightarrow$ uncommon)

(Augmented) Dickey - Fuller test:

H_0: unit root ($\gamma = 0 \Leftrightarrow \rho = 1$)

H_1: stationarity ($\rho < 1$)

$\Rightarrow$

Perfectly random time series rw1 and rw2 appear to be correlated. This is a spurious regression, because they are both non-stationary.

Example 2

The difference operator manipulation

$$\Delta_j^i = (1 - L^j)^i$$, where L is the lag operator

$$\Delta Y_t = (1-L) Y_t = Y_t - Y_{t-1}$$

If y_t is given in logs:

$$\log y_t - \log y_{t-1} = \log \frac{y_t}{y_{t-1}} \Rightarrow \text{„growth rate"}$$

For example: prices (quarterly data) $\longrightarrow$ growth rate: inflation

a) $\Delta_4 Y_t = (1-L^4) Y_t = Y_t - L^4 Y_t = Y_t - Y_{t-4}$

E.g.: yearly change of price level for a certain quarter
if in logs → yearly inflation on a quarterly basis

b) $\Delta^2 Y_t = (1-L)^2 Y_t = (1-2L+L^2) Y_t = Y_t - 2L Y_t + L^2 Y_t =$

$= Y_t - 2 Y_{t-1} + Y_{t-2} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) =$

$= \Delta Y_t - \Delta Y_{t-1}$

E.g.: quarterly change of a quarterly change for two consecutive quarters

if in logs → change of quarterly inflation between two consecutive quarters → „acceleration"

c) $\Delta_1 \Delta_4 Y_t = (1-L)(1-L^4) Y_t = (1 - L - L^4 + L^5) Y_t =$

$= Y_t - Y_{t-1} - Y_{t-4} + Y_{t-5} =$

$= (Y_t - Y_{t-1}) - (Y_{t-4} - Y_{t-5}) =$ E.g.: yearly change of quarterly inflation

OR $= (Y_t - Y_{t-4}) - (Y_{t-1} - Y_{t-5})$ quarterly change of yearly inflation

d) $\Delta_4^2 y_t = (1-L^4)^2 y_t = (1-2L^4+L^8) y_t =$

$= y_t - 2y_{t-4} + y_{t-8} = (y_t - y_{t-4}) - (y_{t-4} - y_{t-8}) =$

$= \Delta_4 y_t - \Delta_4 y_{t-4}$

E.g.: yearly change of a yearly inflation (regarding the same quarter)