

# MOOC Econometrics

## Lecture 4.4 on Endogeneity: Testing for endogeneity

Dennis Fok

## Outline

Given

- Model  $y = X\beta + \varepsilon$
- Instruments  $Z$

Two important things to test

- 1  $Z$  satisfies assumptions for instruments?
- 2  $X$  exogenous or endogenous?

## Testing the validity of instruments

Valid instruments satisfy three conditions

- 1 There are enough instruments  
→ Easy! Just count.
- 2 Instruments are correlated (enough) with  $X$   
→ Check significance of instruments in first stage regression
- 3 Instruments are not correlated with  $\varepsilon$   
→ Perform *Sargan test*

## Test correlation $Z$ vs. $X$

- $X_1$  potentially endogenous variables
- $X_2$  exogenous variables
- $Z = (Z^*, X_2)$  instruments

First-stage regression: apply OLS to  $X_1 = Z^*\gamma_1 + X_2\gamma_2 + \eta$

### Test

Why does 2SLS require  $\gamma_1 \neq 0$ ?

If  $\gamma_1 \approx 0$ :

- $\hat{X}_1 \approx X_2\hat{\gamma}_2$   
→  $\hat{X}_1$  almost perfectly correlated with  $X_2$
- (Extremely) large estimation uncertainty

Test for sufficient correlation:

- Test  $H_0 : \gamma_1 = 0$  in first-stage regression.

## Sargan test

Ingredients:

- Model:  $y = X\beta + \varepsilon$
- Explanatory variables:  $X = (X_1, X_2)$   
 $X_1$  (endogenous),  $X_2$  (exogenous)
- Instruments:  $Z = (Z^*, X_2)$

Null hypothesis ( $H_0$ ): Correlation  $Z$  and  $\varepsilon$  equals 0

Test procedure:

- Rewrite to  $H_0 : \delta = 0$  in

$$\varepsilon = Z\delta + \xi$$

- $\varepsilon$  cannot be observed  
→ Estimate  $\varepsilon$  using 2SLS



Lecture 4.4, Slide 5 of 1, Erasmus School of Economics

## Sargan test

Procedure:

- 1 Use  $Z$  to obtain 2SLS estimator  $b_{2SLS}$  for  $\beta$
- 2 Calculate  $e_{2SLS} = y - Xb_{2SLS}$
- 3 Regress  $e_{2SLS}$  on  $Z$
- 4  $nR^2 \approx \chi^2(m - k)$  under  $H_0$  (valid instruments)
  - ▶  $m$  instruments in  $Z$
  - ▶  $k$  explanatory variables in  $X$

### Test

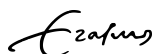
The Sargan test requires  $m > k$ . What happens when  $m = k$ ?



Lecture 4.4, Slide 6 of 1, Erasmus School of Economics

## Notes on the Sargan test

- Test only works when there are “too many” instruments ( $m > k$ )
- At least  $k$  of the instruments should be valid
- Test cannot indicate which instruments are invalid!



Lecture 4.4, Slide 7 of 1, Erasmus School of Economics

## Testing for exogeneity of variables – Hausman test

Intuition:

- Use the instruments to split potentially endogenous variables into
  - 1 a guaranteed exogenous part
  - 2 a potentially endogenous part
- Check whether the endogenous and exogenous part affect  $y$  differently.



Lecture 4.4, Slide 8 of 1, Erasmus School of Economics

Ingredients:

- Explanatory variables:  $X = (X_1, X_2)$
- Potentially endogenous:  $X_1$  ( $k_1$  variables)
- Exogenous variables:  $X_2$  ( $k_2$  variables)
- Instruments:  $Z$

Null hypothesis ( $H_0$ ):  $X_1$  is exogenous

Formal procedure:

- 1 Regress  $y$  on  $X \rightarrow$  calculate  $e = y - Xb$
- 2 Regress  $X_1$  on  $Z \rightarrow$  calculate residuals  $V$
- 3 Regress  $e$  on  $X$  and  $V$
- 4  $nR^2 \approx \chi^2(k_1)$  under  $H_0$  of exogeneity



- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).

