Instrumental Variables and Two-stage Least Squares

Chayanee Chawanote

Lecture 3/3 - Semester 2/2014

Multiple Endogenous Explanatory Variables

- We may have more than one endogenous variable:

  $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 y_3 + \beta_3 z_1 + \beta_4 z_2 + \beta_5 z_3 + u_1$ (1)

  where $E(u_1) = 0$ and u_1 is uncorrelated with z_1, z_2, z_3. We have y_2 and y_3 are endogenous explanatory variables

- To estimate (1) by 2SLS, we need at least two exogenous varibles that is not included in (1) but are correlated with y_2 and y_3.
  - We need either z_4 or z_5 to appear in each reduced form for y_2 and y_3
  - This is a necessary condition for identification, called the order condition

- The sufficient condition, called the rank condition, will be discussed in Simultaneous Equations Model.

Multiple Hypotheses testing with 2SLS

- R^2 from 2SLS can be negative.
- Using 2SLS residuals to compute SSR for both restricted and unrestricted models cannot guarantee that $SSR_R \geq SSR_{UR}$
  - If $SSR_R < SSR_{UR}$, then F stat would be negative.
- We can combine the sum of squared residuals from the 2nd stage regression with SSR_{UR} to obtain a statistic with an approximate F distribution in larger samples.
  - We let STATA do this.

2SLS with Heteroskedasticity

- Same issues as with OLS
- Test for heteroskedasticity: use similar method of the Breusch-Pagan test
- Let $\hat{u}$ denote 2SLS residuals
- Let $z_1, z_2, ..., z_m$ denote all the exogenous variables (including IVs)
- Under reasonable assumptions, an asymptotically valid statistic is the usual F stat for joint significance in a regression of $\hat{u}^2$ on $z_1, z_2, ..., z_m$
  - Regress $\hat{u}^2$ on $z_1, z_2, ..., z_m$, then (overall) F-test
  - H_0: all $z_j = 0 \rightarrow$ homoskedasticity
  - If reject H_0, then there is heteroskedasticity. We might use heteroskedasticity-robust standard errors
- We can also use a weighted 2SLS procedure if we know how error variance depends on the exogenous variables
  - Estimate a model for $Var(u|z_1, z_2, ..., z_m) = h(z)$
  - Do GLS/WLS: Calcuate the estimated variance, $\hat{h}$, from above. Divide the dependent variable, the explanatory variable, and all instrumental variables by $\sqrt{\hat{h}_i}$
  - Apply 2SLS on the transformed equation using the transformed instruments.