Instrumental Variables for Dummies

January 2011

Instrumental Variables (2SLS) Methodology

- Hypothesized structural model:

$$\begin{aligned} Y_i &= \alpha + \beta X_i + \varepsilon_i \\ X_i &= \gamma + \delta Y_i + \theta Z_i + \eta_i, \end{aligned}$$

where

$$\begin{aligned} Y_i &= \text{dependent variable (e.g. price)} \\ X_i &= \text{key explanatory variable (e.g. quantity)} \\ Z_i &= \text{vector of exogenous instrumental variables (e.g. costs)} \end{aligned}$$

- Reduced form for X_i:

$$X_i = \frac{\gamma + \delta\alpha + \theta Z_i + \delta\varepsilon_i + \eta_i}{1 - \delta\beta}$$

- **If Z_i is uncorrelated with ε_i and η_i** then we can estimate the "first stage regression"

$$X_i = a + bZ_i + u_i$$

using OLS where

$$a = \frac{\gamma + \delta\alpha}{1 - \delta\beta} \text{ and } b = \frac{\theta}{1 - \delta\beta}$$

- Then run "second-stage regression"

$$Y_i = \alpha + \beta\hat{X}_i + \varepsilon_i$$

using the fitted value

$$\hat{X}_i = \hat{a} + \hat{b}Z_i$$

- Estimate of β should reflect impact of variations in X_i that are due to exogenous variation in $Z_i's$ only

- Three key requirements of "good instruments":

↪ R^2 in first stage regression must be reasonably high

↪ must clearly be an exogenous determinant of X_i

↪ no other theoretical channels through which Z_i effects Y_i (i.e. Z_i is not correlated with ε_i in theory)

- Testing identification restrictions

↪ the last requirement can be tested for if the system is "over-identified": if there are more $Z's$ than $X's$

↪ Sargan test