Applied Microeconometrics (L3): Instrumental Variables Regression

Nicholas Giannakopoulos

University of Patras
Department of Economics

ngias@upatras.gr

October 27, 2015

Overview

Single Regressor & Single Instrument

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Aim in IV regression: breakdown X into
  - a part that might be correlated with u
  - a part that is not correlated with u
- Why?
  - in order get unbiased estimate of β_1
- How?
  - using an instrumental variable Z_i uncorrelated with u_i
  - Z_i is considered to be an exogenous variable
  - Note: Endogenous variable: "determined within the system"

Conditions for a valid instrument

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Instrument relevance: $corr(Z_i, X_i) \neq 0$
- Instrument exogeneity: $corr(Z_i, X_i) = 0$

How can you use Z_i to estimate β_1?

The IV Estimator, one X and one Z

Two Stage Least Squares (2SLS)

- First stage: isolates the part of X that is uncorrelated with u
  - regress X on Z using OLS: $X_i = \pi_0 + \pi_1 Z_i + v_i$
  - compute the predicted values: $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$
- Second stage: replace X_i by $\hat{X}_i$ in the regression of interest
  - regress Y on $\hat{X}_i$ using OLS: $Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i$
  - resulting estimator is called the "Two Stage Least Squares" (2SLS) estimator: $\hat{\beta}_1^{2SLS}$
  - $\hat{\beta}_1^{2SLS}$ is a consistent estimator of β_1

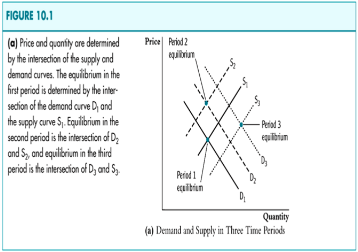
Example: Supply and demand for butter

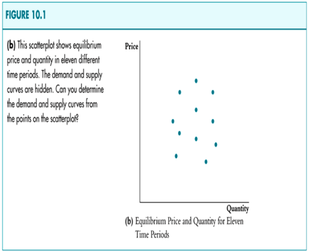
Estimate demand elasticities for agricultural goods

$$ln(Q_i) = \beta_0 + \beta_1 ln(P_i) + u_i$$

- β_1=price elasticity of butter=percent change in quantity for a 1% change in price
- Data: observations on price and quantity of butter for different years
- The OLS regression of $ln(Q_i)$ on $ln(P_i)$ suffers from simultaneous causality bias
- (!) price and quantity are determined by the interaction of demand and supply

Interaction of demand and supply



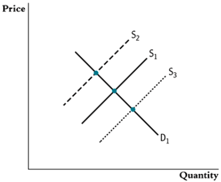


Source: J.H. Stock and M.W. Watson, Introduction to Econometrics (first edition), Addison-Wesley, 2003

What would you get if only supply shifted?

FIGURE 10.1

(c) When the supply curve shifts from S_1 to S_2 to S_3 but the demand curve remains at D_1, the equilibrium prices and quantities trace out the demand curve.



(c) Equilibrium Price and Quantity When Only the Supply Curve Shifts

Source: J.H. Stock and M.W. Watson, Introduction to Econometrics (first edition), Addison-Wesley, 2003

(1): 2SLS estimates the demand curve by isolating shifts in price and quantity that arise from shifts in supply

(2): Z is a variable that shifts supply but not demand.

Example: Supply and demand for butter

$$ln(Q_i) = \beta_0 + \beta_1 ln(P_i) + u_i$$

- Let Z=rainfall (rain) in dairy-producing regions
- Is Z a valid instrument?
  - Relevant? $corr(\text{rain}_i, ln(P_i)) \neq 0$
    Plausibly: insufficient rainfall means less grazing means less butter
  - Exogenous? $corr(\text{rain}_i, u_i) = 0$
    Plausibly: whether it rains in dairy-producing regions should not affect demand

2SLS in the supply-demand example

$$ln(Q_i) = \beta_0 + \beta_1 ln(P_i) + u_i,\ Z_i = \text{rain}_i,\ i = 1, ..., R$$

- First stage: isolates changes in log price that arise from supply
  - regress $ln(P_i)$ on rain$_i$ using OLS: $ln(P_i) = \pi_0 + \pi_1 \text{rain}_i + v_i$
  - compute the predicted values: $ln(\hat{P_i})$
- Second stage
  - regress $ln(Q_i)$ on $ln(\hat{P_i})$ using OLS: $ln(Q_i) = \beta_0 + \beta_1 ln(\hat{P_i}) + u_i$
  - resulting estimator $\hat{\beta}_1^{2SLS}$
  - The regression counterpart of using shifts in the supply curve to trace out the demand curve

Inference

1. Statistical inference proceeds in the usual way
2. The justification is (as usual) based on large samples
3. This all assumes that the instruments are valid (vs. weak instruments)
4. Standard errors
   - The OLS standard errors from the second stage regression are not right they do not take into account the estimation in the first stage ($\hat{X}_i$ is estimated)
   - Instead, use a single specialized command that computes the 2SLS estimator and the correct SEs
   - as usual, use heteroskedasticity-robust SEs