Econometrics

Instrumental Variables

Paul P. Momtaz

The Anderson School
UCLA

Outline

Introduction
The basic idea

Idea: Partition x in a part that is uncorrelated with y and a part that is correlated with u, and use only the former part to estimate β.

Conditions:

- Relevance $Corr(z_i, x_i) \neq 0$.
- Exogeneity $Corr(z_i, u_i) = 0$

$$\begin{aligned} Cov(y_i, z_i) &= Cov(\beta_0 + \beta_1 x_i + u_i, z_i) \\ &= \beta_1 Cov(x_i, z_i) \end{aligned}$$

$$\beta_1 = \frac{Cov(y_i, z_i)/Var(z_i)}{Cov(x_i, z_i)/Var(z_i)}$$

Introduction

Taxonomy

$$\hat{\beta}_1^{IV} = \frac{S_{yz}}{S_{xz}} \text{ where } S_{yz} \xrightarrow{P} Cov(y_i, z_i) \text{ and } S_{xz} \xrightarrow{P} Cov(x_i, z_i)$$

$$\begin{aligned} \text{First-Stage: } & x_i = \alpha_0 + \alpha_1 z_i + \xi_i \\ \text{Second-Stage: } & y_i = \gamma_0 + \gamma_1 \hat{x}_i + v_i \\ \text{Reduced-Form: } & y_i = \delta_0 + \delta_1 z_i + \varepsilon_i \end{aligned}$$

- In large samples, $\hat{\beta}_1^{IV} \sim N(\beta_1, \hat{\sigma}^2_{\beta_1^{IV}})$
- Estimates generally not unbiased
- Second-stage errors incorrect in OLS $\rightarrow$ hekroskedasticity-adjustment

Wald estimator for binary instrument $\hat{\beta}_{wald} = \frac{\overline{y}_1 - \overline{y}_0}{\overline{x}_1 - \overline{x}_0}$

General Form $\hat{\beta}_{IV} = \frac{dy/dz}{dx/dz}$

(Wu-)Hausman Test for Endogeneity

Idea: Test consistency of estimator (OLS) to less efficient estimator (IV).

$$H_0 : \mathbb{E}[u|x] = 0$$

Under the null, $Var[\hat{\beta}_{IV} - \hat{\beta}_{OLS}] = Var[\hat{\beta}_{IV}] - Var[\hat{\beta}_{OLS}]$

$$\xi_H = (\hat{\beta}_{IV} - \hat{\beta}_{OLS})'[\hat{Vas}[\hat{\beta}_{IV}] - \hat{Var}[\hat{\beta}_{OLS}]]^{-1}(\hat{\beta}_{IV} - \beta_{OLS})$$

$$\xi_H \sim_a \chi^2(k)$$

under the null, $plim(\hat{\beta}_{IV} - \hat{\beta}_{OLS}) = 0$ since both estimators are consistent.

Multivariate IV

$$\hat{\beta}_{2SLS} = (x' P_z x)^{-1} x' P_z y$$
$$P_z = z'(z'z)^{-1} z'$$
$$\text{if } u = Q, \qquad \hat{\beta}_{IV} = (z'x)^{-1} z'y$$

$$u \equiv \text{ Number of explanatory variables}$$
$$Q \equiv \text{ Number of instruments}$$

$u = Q$: exactly identified

$u > Q$: underidentified (not identified)

$u < Q$: overidentified

Weak Instruments

OLS bias

IV consistent but biased. Bias increases in weak instruments and overidentification.

Bias towards OLS

Consider $y = \beta x + \eta$ and $x = 2\pi + \xi$.

OLS biased since $Corr(\eta_i, \xi_i) \neq 0$, Note that

- $Corr(z_i, \xi_i) = 0$ by construction.
- $Corr(z_i, \eta_i) = 0$ by assumption.

OLS bias: $\mathbb{E}[\hat{\beta}_{OLS} - \beta] = \frac{Cov(\eta, x)}{Var(x)} \overset{\text{if } Corr(\xi_i,\eta_i)\neq 0}{=} \frac{\sigma^2_{\eta\varepsilon}}{\sigma^2_x}$

Weak Instruments
2SLS bias

2SLS bias:

$$\mathbb{E}[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\varepsilon}}{\sigma_{\xi}^{2}} \frac{1}{F+1}$$

Where F is population analogue of F-stat of the first stage.

- Weak first -stage ($F \to 0$), 2SLS bias $\to$ OLS bias. Since, if $\pi = 0, \sigma_x^2 = \sigma_\xi^2$
- $F \to \infty$, 2SLS bias $\to 0$.
- Adding more weak instruments reduces F further.
- Bias in just-identified models leads to large SE in second - stage.

Sargan Test

Testing Overidentifying Restric

Idea: Cannot test exogeneity restriction, $Cov(z, \varepsilon) = 0$ since it involves the unobserved error. But with $Q > k$, can estimate $\hat{\varepsilon}_i$ with z_1 and then check $Corr(z_2, \hat{\varepsilon}_i)$.

1. Estimate structural equation by 2SLS using all instruments and obtain $\hat{\varepsilon}_i$.
2. Regress $\hat{\varepsilon}_i$ on all independent variables and get R_1^2.
3. Under the null, $NR_1^2 \sim \chi^2(f)$, where f is the number of overidentifying restrictions. Check test statistic

Mitigating Weak Instruments

- Just-identified model with strongest IV.
- Limited information maximum likelihood estimator (LIML), which provides some asymptotic distributions as 2SLS but finite sample bias reduction.

Practical Tips for IV Papers

- Report first-stage results.
- Report F-stats on excluded instruments.
  - F-stats $>$10 means no weak instrument problem.
  - If more than one dependent variable, use Cragg-Donald minimum eigenvalue statistic to test for weak instruments.
- Focus on strongest IV, prefer just-identified models.
- Check overidentified 2SLS models with LIML.
- Check reduced form (unbiased since OLS). Can you see a casual relationship?

Special IV Models

- Two - Sample IV (Angrist, 1990)
- Split - Sample IV (Angrist and Krueger, 1995)
- Grouped Data (Angrist, 1991; Imbens and Van der klaauw, 1995)