ECON 4151
Lab Session 2: IV

Wonjin Lee

Washington University in St. Louis

October 9, 2020

Outline

1. Review of IV (Wooldridge's Introductory Ecconometrics)

2. Exercise (Gujarati's Econometrics by Example)

Why regressors and error term are correlated?

1. Measurement errors in the regressor(s) $\longleftarrow$ IV can fix!
2. Omitted variable bias $\longleftarrow$ IV can fix (modern application)!
3. Simultaneous equation bias $\longleftarrow$ IV can fix!
4. Dynamic regression model with serial correlation in the error term

Intro to IV

- Structural equation:

$$y_i = \alpha + \rho s_i + X_i'\beta + \underbrace{A_i'\gamma + \eta_i}_{\equiv \epsilon_i}$$

- ρ has a causal interpretation.
- A are unobserved variables $\implies$ a biased estimate of ρ.
- Need to find at least one instrument z with special characteristics.
  1. Relevance requirement: $\mathrm{Cov}(z, s) \neq 0$
     - We can test it! Just run a regression and do a t test!
  2. Exclusion restriction: $\mathrm{Cov}(z, \epsilon) = 0$
     - We can not test it... Then what should we do?

How IV works?

- We want to run the following regression:

$$y_i = \alpha + \rho s_i + \underbrace{A_i'\gamma + \eta_i}_{\equiv \epsilon_i}$$

- First-Stage:

$$s_i = \delta_0 + \delta_1 z_i + \nu_i$$

- Reduced Form:

$$\begin{aligned} y_i &= \alpha + \rho\left(\delta_0 + \delta_1 z_i + \nu_i\right) + A_i'\gamma + \eta_i \\ &= \alpha + \rho\delta_0 + \rho\delta_1 z_i + \rho\nu_i + A_i'\gamma + \eta_i \end{aligned}$$

How IV works?

- The IV estimator in the bivariate context:

$$\hat{\rho}_{IV} = \frac{\mathrm{Cov}(z, y)}{\mathrm{Cov}(z, s)} = \frac{\mathrm{Cov}(z, y) / \mathrm{Var}(z)}{\mathrm{Cov}(z, s) / \mathrm{Var}(z)} = \rho$$

- Why?
  - Numerator:

$$\begin{aligned} \frac{\mathrm{Cov}\,(z_i, y)}{\mathrm{Var}(z)} &= \frac{\mathrm{Cov}\,(z, \alpha + \rho\delta_0 + \rho\delta_1 z_i + \rho\nu_i + \epsilon_i)}{\mathrm{Var}(z)} \\ &= \frac{0 + 0 + \rho\delta_1 \mathrm{Var}(z) + 0 + 0}{\mathrm{Var}(z)} \\ &= \rho\delta_1 \end{aligned}$$

  - Denominator:

$$\frac{\mathrm{Cov}(z, s)}{\mathrm{Var}(z)} = \delta_1$$

One comment on IV

- You might want to do the followings:
    1. You might want to simply run the first-stage and reduced form separately, and do the division manually.
    2. You also might want to run the first-stage, predict $\hat{s}$, and run the structural regression with s replaced by $\hat{s}$.
- The coefficient estimates will be correct, but the standard errors will not.
- A better way is to let Stata do it for you.

One more comment on IV

- In small, or finite, samples, IV estimators are biased and their variances are less efficient than the OLS estimators.
- We want $\text{Corr}(z, s)$ to be high! (Relevance requirement)
  - $\text{Corr}(z, s)$ is high (low) $\implies$ IVs are strong (weak).
- Why? If the instruments are weak, IV estimators may not be normally distributed even in large samples.

One comment on Var$(\hat{\rho}_{IV})$

- Since $R^2_{s,z} < 1$, Var$(\hat{\rho}_{IV}) >$ Var$(\hat{\rho}_{OLS})$ (when OLS is valid)
  - $R^2_{s,z}$: the strength of the linear relationship b/w s and z (Relevance requirement)
  - $R^2_{s,z} \uparrow \implies$ Var$(\hat{\rho}_{IV}) \downarrow$
- Why?
  - The IV estimator in the bivariate context (p. 495):

$$\text{Var}(\hat{\rho}_{IV}) = \frac{\sigma^2_\epsilon}{n\sigma^2_s \rho^2_{s,z}} \implies \frac{\hat{\sigma}^2_\epsilon}{SST_s R^2_{s,z}}$$

  - The OLS estimator in the bivariate context (p. 51):

$$\text{Var}(\hat{\rho}_{OLS}) = \frac{\sigma^2_\epsilon}{SST_s}$$

  - $\sigma^2_\epsilon = \text{Var}(\epsilon)$, $\sigma^2_s = \text{Var}(s)$, and $\rho^2_{s,z} = (\text{Corr}(s,z))^2$

Heterogeneous Treatment Effects - LATE

- Treatment effects are not likely to equal across the entire population.
- In general, we cannot find the ATE or even the TOT using an IV strategy.
- IV only identifies the LATE.
- For convenience, assume that z_i is a dummy instrument and s_i is a dummy treatment.
- Compliance Sub-Populations: Four different types of individuals based on their treatment status in response to different values of the instrument.

	$s_{0i} = 0$	$s_{0i} = 1$
$s_{1i} = 0$	never-taker	defier
$s_{1i} = 1$	complier	always-taker

LATE $\neq$ TOT

- Local average treatment effect (LATE):

$$\rho = \frac{\rho\delta_1}{\delta_1} = \frac{\mathbb{E}\left[y_i|z_i = 1\right] - \mathbb{E}\left[y_i|z_i = 0\right]}{\mathbb{E}\left[s_i|z_i = 1\right] - \mathbb{E}\left[s_i|z_i = 0\right]} = \mathbb{E}\left[y_{1i} - y_{0i}|c_i = 1\right]$$

  where c_i is a dummy indicating whether i is a complier.
  - When we have heterogeneous treatment effects, we can only hope to identify a treatment effect for the sub-population whose treatment status is determined by the instrument.
  - The IV estimate is about compliers (the relevance requirement).
  - Monotonicity.

Hausman test

- How to test for endogeneity?
- Steps:
  1. From the First-Stage, get $\hat{\nu}_i$:

$$s_i = \delta_0 + \delta_1 z_i + \nu_i$$

     - Note that in the structural equation, $\delta_1 z_i$ is the exogenous part of s_i and ν_i is the endogenous part of s_i.
  2. Estimate the structural equation with $\hat{\nu}_i$:

$$y_i = \alpha + \rho s_i + \pi \hat{\nu}_i + \epsilon_i$$

  3. Test $H_0 : \pi = 0$ using a t statistic.
     - H_0 means s is not endogenous.

Identification

- # of instruments $=$ # endogenous regressors:
  - The regression coefficients are exactly identified.
  - We can obtain unique estimates of them.
- # of instruments $>$ # endogenous regressors:
  - The regression coefficients are overidentified.
  - We may obtain more than one estimate of one or more of the regressors.
- # of instruments $<$ # endogenous regressors:
  - The regression coefficients are underidentified.
  - We cannot obtain unique values of the regression coefficients.

Overidentification test

- Now we have two IVs for s: z and w.
- We want to test the exclusion restriction; are z and w uncorrelated with the structural error?
- Steps:
    1. Estimate the structural equation by 2SLS and obtain the 2SLS residuals, $\hat{\epsilon}$.
    2. Regress $\hat{\epsilon}$ on all exogenous variables (including IVs) and obtain R^2.
    3. Test H_0 that z and w uncorrelated with the structural error.
        - Under H_0, $nR^2 \sim \chi^2_q$ where q is the # of surplus instruments.

Example

- Run the following regressions:

$$\begin{aligned}\ln Earn_i = & B_1 + B_2 s_i + B_3 Wexp_i + B_4 Gender_i \\ & + B_5 Ethblack_i + B_6 Ethhisp_i + u_i\end{aligned}$$

- 3 instruments: mother's schooling, father's schooling, and number of siblings.