

Week 5: IV Estimation and Inference

JJ Chen

February 13, 2015

Anoucement

► Today's Plan:

1. Some questions I received this week
2. Finite Sample Bias of the 2SLS estimator
3. IV Example Paper: Card (1997)

Questions, Questions, Questions

1. How to derive the Wald estimator from the IV formula?
2. Why do we do IV again?
 - ▶ How to think about possible threat to internal validity?
3. How to think about source of identification in a regression setting?
4. What happens if we have multiple endogenous variable?
 - ▶ One answer
 - ▶ Or if you just want to know the answer to the homework question, run a simple simulation
5. More questions?

2SLS set up

- ▶ Note: We are back to matrices and sample reg
- ▶ Consider a simple setting: one endog $\mathbf{x}_{N \times 1}$ and a few instruments $\mathbf{Z}_{N \times Q}$:
 - ▶ $\mathbf{y} = \rho \mathbf{x} + \boldsymbol{\eta}$
 - ▶ $\mathbf{x} = \mathbf{Z}\boldsymbol{\pi} + \boldsymbol{\xi}$
- ▶ We have the followings:
 - ▶ η_i is correlated with ξ_i
 - ▶ By construction, ξ_i is orthogonal to the instruments \mathbf{z}'_i
 - ▶ By assumption (ER), η_i is orthogonal to the instruments \mathbf{z}'_i
- ▶ The 2SLS estimator is defined as
 - ▶ $\hat{\rho}^{2SLS} = (\mathbf{x}' \mathbf{P}_Z \mathbf{x})^{-1} \mathbf{x}' \mathbf{P}_Z \mathbf{y} = \rho + (\mathbf{x}' \mathbf{P}_Z \mathbf{x})^{-1} \mathbf{x}' \mathbf{P}_Z \boldsymbol{\eta}$
 - ▶ $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}'$ is the projection matrix; how does it relate to the IV estimator?

The Bias of 2SLS

- ▶ Substitute the 1st for x in $x'P_Z\eta$:
 - ▶ $\hat{\rho}^{2SLS} - \rho = (x'P_Zx)^{-1}\pi'Z\eta + (x'P_Zx)^{-1}\xi'P_Z\eta$
- ▶ Take expectation and Bekker approximation:
 - ▶ $E(\hat{\rho}^{2SLS} - \rho) \approx (E(x'P_Zx))^{-1}E(\pi'Z\eta) + (E(x'P_Zx))^{-1}E(\xi'P_Z\eta) = (E(x'P_Zx))^{-1}E(\xi'P_Z\eta)$
 - ▶ The first term of RHS can be canceled since z_i is orthogonal to η_i
- ▶ Substitute the 1st for x again in $x'P_Zx$ gives
 - ▶ $E(\hat{\rho}^{2SLS} - \rho) \approx [E(\pi'Z'Z\pi) + E(\xi'P_Z\xi)]^{-1}E(\xi'P_Z\eta)$
 - ▶ Thus we see 2SLS is biased because η_i is correlated with ξ_i (derivation)

Simplify the bias term

- ▶ We can do some algebra to simplify the term and get a more intuitive formula
 - ▶ $E(\hat{\rho}^{2SLS} - \rho) \approx \frac{\sigma_{\eta\xi}}{\sigma_{\xi}^2} [F + 1]^{-1}$
 - ▶ Where F is the population F-stat for the joint significance of all regressors in the 1st
- ▶ Notice that when the 1st F-stat is small, the bias term gets larger; and if the instruments are too weak, the 2SLS bias term gets closer to OLS bias
 - ▶ More on week IV next week

Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Immigration

- ▶ Question: What is the effect of immigration on the wages and employment rates of natives?
- ▶ The “area studies” approach
 - ▶ Think about the correlation between natives’ wages in a city and the share of the work force that is foreign born
 - ▶ A long line of research finds that this correlation is small.
 - ▶ What are the problems in interpreting this correlation as a causal effect of immigration on natives’ wages?

Biases in “area studies”

- ▶ Natives may move out in response to immigrant inflows
 - ▶ Upward bias or downward bias?
- ▶ Local demand shocks may draw in population and also raise wages
 - ▶ Upward bias or downward bias?

A Model

- ▶ Each city produce a single output good, Y , by combining capital, K , and different types of labor, L :
 - ▶ $Y_c = F(K_c, L_c)$ Aggregate labor L_c is a combination of people from j different skill/occupation groups:
 - ▶ $L_c = \left[\sum_j (e_{jc} N_{jc}^{\frac{\sigma-1}{\sigma}}) \right]^{\frac{\sigma}{\sigma-1}}$
 - ▶ e_{jc} : Productivity of group j in city c
 - ▶ N_{jc} : Number of workers in group j in city c
 - ▶ σ is the substitutability between groups

The Relationship between Labor Supply/Demand and Wages

- ▶ Solving the model and setting the marginal product of each group equal to that group's wage. This delivers:
 - ▶ $\log N_{jc} = \theta_c + (\sigma - 1) \log e_{jc} - \sigma \log w_{jc}$
 - ▶ where $\theta_c = \sigma \log \left[q_c F_L(K_c, L_c) L_c^{\frac{1}{\sigma}} \right]$
- ▶ Let's also posit a labor supply function for each group:
 - ▶ $\log(N_{jc}/P_{jc}) = \varepsilon \log w_{jc}$
 - ▶ P_{jc} : Number of workers in group j in city c

A Little Bit of Algebra

- ▶ Solve the model for wage and employment equations

- ▶ $\log w_{jc} = \frac{1}{\varepsilon + \sigma} \left[(\theta_c - \log P_c) + (\sigma - 1) \log e_{jc} - \log\left(\frac{P_{jc}}{P_c}\right) \right]$

- ▶ $\log\left(\frac{N_{jc}}{P_{jc}}\right) = \frac{\varepsilon}{\varepsilon + \sigma} \left[(\theta_c - \log P_c) + (\sigma - 1) \log e_{jc} - \log\left(\frac{P_{jc}}{P_c}\right) \right]$

- ▶ Jump to regressions

- ▶ $\log w_{jc} = u_j + u_c + d_1 \log\left(\frac{P_{jc}}{P_c}\right) + u_{jc}$

- ▶ $\log\left(\frac{N_{jc}}{P_{jc}}\right) = v_j + v_c + d_2 \log\left(\frac{P_{jc}}{P_c}\right) + v_{jc}$

- ▶ u and v terms are fixed group and city effects
 - ▶ The coefficients d 's are functions of the underlying elasticities
 - ▶ What is the key source of bias in these equations?

Occupation groups

- ▶ The theoretical model assumes that workers within an occupation are perfect substitutes for one another.
- ▶ Card works with probabilistic occupation groups
 - ▶ Run multinomial logits, by gender and immigrant status, where the dependent variable indicates one of six broad occupation groups and the independent variables are age, education, race, marital status, and disability status (and country of origin and years in the US)
 - ▶ Table 2 and Table 3
- ▶ OLS results: Figure 1 and Table 4

IV

- ▶ Why IV?
 - ▶ Recall the biases we've discussed
- ▶ We need an instrument for the inflow of new immigrants
 - ▶ Immigrants tend to establish enclaves in certain cities.
- ▶ So we can imagine that the flow of immigrants to, say, Chicago is composed of two groups:
 - ▶ Those who move to Chicago because of a local demand shock (demand pull)
 - ▶ Those who move to Chicago because of the existing immigrant network (the supply push)
 - ▶ The second group is arguably exogenous

The Instrument

- ▶ M_g is the number of immigrants from country g who enter the US between 1985 and 1990
- ▶ λ_{gc} is the fraction of immigrants who arrived prior to 1985 from country g who live in city c in 1985
- ▶ τ_{gj} is the fraction of new immigrants from country g who work in occupation j
- ▶ The predicted “supply push” component of immigration by people in occupation j , from country g , to city c , is $\tau_{gj}\lambda_{gc}M_g$
- ▶ Summing across countries g yields the total occupation-city supply push
- ▶ How good is this instrument?

IV Results

- ▶ IV construction: Table 5
- ▶ 1st: Figure 2
- ▶ RF: Figure 3
- ▶ IV results: Table 4 and Table 6

References I

Card, David. 1997. *Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration*. National Bureau of Economic Research.