Week 4: IVs with Constant Effects

JJ Chen

February 6, 2015

Anoucement

- Week 4:
  - IV Motivations, Assumptions, Mean Comparison (Wald Estimator)
  - IV Implementations: 2SLS
  - IV Example Paper: Card (1997), Hoxby (1994) (we will not cover the hoxby paper)
- Week 5: IV Estimation and Inference
- Week 6: Week IV and Specification Tests
- Week 7: LATE
- Week 8: Midterm Review?

(One) Traditional IV setting

- In class, we started by discussing IVs with constant effect
  - Close to traditional IV practice; allow us to focus on the first order issues: identification
  - But it imposes strong assumptions (will be relaxed)
- Example:
  - Y-Eqn: $\mathrm{Y}_i = f(\mathrm{D}_i, \varepsilon_i) = \beta_0 + \beta_1 \mathrm{D}_i + \varepsilon_i$
  - D-Eqn: $\mathrm{D}_i = f(\mathrm{Z}_i, \upsilon_i) = \alpha_0 + \alpha_1 \mathrm{Z}_i + \upsilon_i$
- Key Assumptions to identify β_1 (more on them later)
  - $\mathrm{E}(\mathrm{Z}_i \varepsilon_i) = 0$, $\mathrm{E}(\mathrm{Z}_i \upsilon_i) = 0$, $\mathrm{C}(\mathrm{D}, \mathrm{Z}_i) \neq 0 \implies \beta_1^{IV} = \frac{\mathrm{C}(\mathrm{Y}_i, \mathrm{Z}_i)}{\mathrm{C}(\mathrm{D}_i, \mathrm{Z}_i)}$
- Implicit Assumptions:
  - Constant effect; additive separability; point identification...

Motivations

- We are interested in a single structural eqn:
  - $\text{Y}_{si} = \alpha + \rho \text{S}_i + \text{A}_i'\boldsymbol{\gamma} + \varepsilon_i$
  - $\rho = \text{Y}_{12,i} - \text{Y}_{11,i} = \text{Y}_{16,i} - \text{Y}_{15,i} \dots$
- But we can't run a dream pop regression because of "unavailable" A_i'
  - $\text{Y}_i = \alpha + \rho \text{S}_i + \text{A}_i'\boldsymbol{\gamma} + e_i$
- Assume we have an IV z_i, problem solved!
  - Should we end the review section and just go find IVs?
- More broadly, IV is a general method of studying causality
  - *"Clean"* variations in "treatment" variable
  - Other ID strategies link closely to the IV idea: DID, RD
- Other Motivations: measurement error, reversal causality, random assignment and actual exposure in experiments

Assumptions and Why Does IV Work

- In a nutshell, we want to capture a single channel (IV chain):
  - $(\text{Z} \rightarrow \text{Y}) = (\text{Z} \rightarrow \text{D}) \times (\text{D} \rightarrow \text{Y})$
  - RF $=$ 1st $\times$ Causal Effect of Interest *(maybe)*
- Conceptual ID Assumptions:
  1. *Exclusion Restriction*: Only one channel through which Z affects Y
     - One implication: In a no-1st sample, we should expect RF $= 0$, example?
     - Another implication; Look at a sample that 1st is active, but the primary channel is not, example?
     - Can we run a reg of Y_i on both D_i and Z_i to test whether ER hold?
  2. *Relevance Condition*: Z has a causal impact on D (with decent size)
- Ex: Suppose you want to know the impact of school per-pupil

IV Mean Comparision: Binary IV I

- RF = 1st $\times$ ρ^{IV} $\Longrightarrow$
- $\rho^{IV} = \frac{RF}{1st} = \frac{\mathrm{C}(\mathrm{Y}_i,\mathrm{Z}_i)}{\mathrm{C}(\mathrm{D}_i,\mathrm{Z}_i)} = \frac{\mathrm{C}(\mathrm{Y}_i,\mathrm{Z}_i)/\mathrm{V}(\mathrm{Z}_i)}{\mathrm{C}(\mathrm{D}_i,\mathrm{Z}_i)/\mathrm{V}(\mathrm{Z}_i)}$
- When IV is a dummy, both RF and 1st can be simplified
  - RF: $\mathrm{C}\left(\mathrm{Y}_i,\mathrm{Z}_i\right) = \left[\mathrm{E}\left(\mathrm{Y}_i \mid \mathrm{Z}_i = 1\right) - \mathrm{E}\left(\mathrm{Y}_i \mid \mathrm{Z}_i = 0\right)\right] p(1-p)$
  - 1st: $\mathrm{C}\left(\mathrm{D}_i,\mathrm{Z}_i\right) = \left[\mathrm{E}\left(\mathrm{D}_i \mid \mathrm{Z}_i = 1\right) - \mathrm{E}\left(\mathrm{D}_i \mid \mathrm{Z}_i = 0\right)\right] p(1-p)$
- The ratio ρ^{IV} is the Wald estimator ρ^{Wald}, which provides the essential idea of IV:
  - *RF mean comparisons across groups defined by the IV, scaled by the 1st*

IV Mean Comparision: Binary IV II

- Another way to derive the Wald estimator is through the single structural eqn with constant effect:
  - $\mathrm{Y}_i = \alpha + \rho \mathrm{D}_i + \eta_i$
  - $\mathrm{E}(\mathrm{Y}\,|\mathrm{Z}_i = 1) = \alpha + \rho\,\mathrm{E}(\mathrm{D}_i\,|\mathrm{Z} = 1) + \mathrm{E}(\eta_i\,|\mathrm{Z}_i = 1)$
  - $\mathrm{E}(\mathrm{Y}\,|\mathrm{Z}_i = 0) = \alpha + \rho\,\mathrm{E}(\mathrm{D}_i\,|\mathrm{Z} = 0) + \mathrm{E}(\eta_i\,|\mathrm{Z}_i = 0)$
  - $\mathrm{E}(\mathrm{Y}\,|\mathrm{Z}_i = 1) - \mathrm{E}(\mathrm{Y}\,|\mathrm{Z}_i = 0) = \rho\left[\mathrm{E}(\mathrm{D}_i\,|\mathrm{Z}_i = 1) - \mathrm{E}(\mathrm{D}_i\,|\mathrm{Z}_i = 0)\right]$
    - ER $\implies \mathrm{E}(\eta_i\,|\mathrm{Z}_i = 1) = \mathrm{E}(\eta_i\,|\mathrm{Z}_i = 0)$
- So ρ^{Wald} capture the ratio of mean diff
- $\rho^{Wald} = \frac{RF}{1st} = \frac{\mathrm{E}\left(\mathrm{Y}\,|\mathrm{Z}_i = 1\right) - \mathrm{E}\left(\mathrm{Y}\,|\mathrm{Z}_i = 0\right)}{\mathrm{E}\left(\mathrm{D}_i\,|\mathrm{Z}_i = 1\right) - \mathrm{E}\left(\mathrm{D}_i\,|\mathrm{Z}_i = 0\right)}$
- Further simplification: if D_i is binary, 1st becomes mean difference in conditional probability
  - $\Pr(\mathrm{D}\,|\mathrm{Z}_i = 1) - \Pr(\mathrm{D}\,|\mathrm{Z}_i = 0)$

IV Implementation: 2SLS

- What if we have
  1. a multi-valued IV and controls
  2. multiple IVs
  3. RF and 1st come from different sample
- 2SLS is the way to go, but just beware the usual mistakes

2SLS with One IV and Covariates

- Suppose a causal model with controls is
  - $\mathrm{Y}_i = \mathrm{X}_i'\boldsymbol{\alpha} + \rho \mathrm{S}_i + \eta_i$, $\eta_i = \mathrm{A}_i'\boldsymbol{\gamma} + \upsilon_i$
  - Even if the IV is insanely good (An amazing valid ER), we might still want to control for some "useless" covariates to reduce std.err.
- 2SLS: Get predicted value: $\mathrm{S}_i = \hat{\mathrm{S}}_i + \xi_{1i}$ and derive 2st:
  - 2st: $\mathrm{Y}_i = \mathrm{X}_i'\boldsymbol{\alpha} + \rho \hat{\mathrm{S}}_i + [\eta_i + \rho \xi_{1i}]$
- Notice when the IV is $\hat{\mathrm{S}}_i^*$, 2SLS = IV
  - $\hat{\mathrm{S}}_i^*$ is the res of running a reg of $\hat{\mathrm{S}}_i$ on X_i
  - $\rho^{2SLS} = \frac{\mathrm{C}\left(\mathrm{Y}_i, \hat{\mathrm{S}}_i^*\right)}{\mathrm{V}\left(\hat{\mathrm{S}}_i^*\right)} = \frac{\mathrm{C}\left(\mathrm{Y}_i, \hat{\mathrm{S}}_i^*\right)}{\mathrm{C}\left(\mathrm{S}_i, \hat{\mathrm{S}}_i^*\right)} = \rho^{IV}$
- Note: manual 2SLS won't get correct std.err.

ILS with One IV and Covariates

- ILS: Find a 1st $\mathrm{S}_i = \mathrm{X}_i'\boldsymbol{\pi}_{10} + \pi_{11}\mathrm{Z}_i + \xi_{1i}$ and derive RF
  - $\mathrm{Y}_i = \mathrm{X}_i'\boldsymbol{\alpha} + \rho[\mathrm{X}_i'\boldsymbol{\pi}_{10} + \pi_{11}\mathrm{Z}_i + \xi_{1i}] + \eta_i$
  - $\mathrm{Y}_i = \mathrm{X}_i'[\boldsymbol{\alpha} + \rho\boldsymbol{\pi}_{10}] + \rho\pi_{11}\mathrm{Z}_i + [\rho\xi_{1i} + \eta_i]$
  - RF: $\mathrm{Y}_i = \mathrm{X}_i'\boldsymbol{\pi}_{20} + \pi_{21}\mathrm{Z}_i + \xi_{2i} \implies \rho^{ILS} = \dfrac{\pi_{21}}{\pi_{11}}$
- Notice that one-IV 2SLS equals IV, when the IV is Z_i^*, and it also equals ILS
  - Z_i^* is the res of running a reg of Z_i on X_i
  - $\frac{\mathrm{C}\left(\mathrm{Y}_i, \hat{\mathrm{S}}_i^*\right)}{\mathrm{V}\left(\hat{\mathrm{S}}_i^*\right)} = \frac{\mathrm{C}\left(\mathrm{Y}_i, \hat{\mathrm{S}}_i^*\right)}{\mathrm{C}\left(\mathrm{S}_i, \hat{\mathrm{S}}_i^*\right)} = \frac{\mathrm{C}\left(\mathrm{Y}_i, \mathrm{Z}_i^*\right)}{\mathrm{C}\left(\mathrm{S}_i, \mathrm{Z}_i^*\right)} = \frac{\pi_{21}}{\pi_{11}}$

2SLS with Multiple IVs

- 2SLS yields a weighted average of estimates one would get using each of the IVs separately
- Let $\rho_j^{IV} = \frac{C(Y_i, Z_{ji})}{C(D_i, Z_{ji})}, j = 1, 2$
  - Two IV estimands using Z_{1i} and Z_{2i} to instrument D_i
- 2SLS estimand will be $\rho^{2SLS} = \phi\rho_1 + (1-\phi)\rho_2$
  - ϕ captures the relative strength of the IVs in the 1st
- 2SLS with multiple IVs will also give smaller std.err.
  - Use more "clean" variations

Visual IV

- When we have multiple IVs, each IV defines a causal effect (hopefully) for one group of people
  - $\mathrm{E}\left(\mathrm{Y}_i \mid \mathrm{R}_i\right) = \alpha + \rho\, \mathrm{E}\left(\mathrm{D}_i \mid \mathrm{R}_i\right) = \alpha + \rho \Pr\left(\mathrm{D}_i = 1 \mid \mathrm{R}_i\right)$, $\mathrm{R}_i \in j = 1, \cdots, J$
  - $\bar{y}_j = \alpha + \rho \hat{p}_j + \bar{\eta}_j$
  - Look like a 2SLS?
- In a constant effect linear model, GLS (in this case, WLS) would be efficient for grouped data
  - Recall from last semester, one motivation for learning GLS is heteroskedasticity
  - If $\mathrm{V}\left(\eta_i\right) = \sigma_\eta^2$, then the group variance is ${}^{\sigma_\eta^2}/{}_{n_j}$
- GLS is 2SLS, so 2SLS is the efficient linear combination of the underlying Wald estimates

Split Sample IV

- Recall the IV estimator is $\rho^{IV} = \frac{RF}{1st}$
- Nothing in the formula prevents us from estimating RF and 1st separately from different sample
- Example:
  - RF: One dataset on income and draft numbers, but no veteran status
  - 1st: Another dataset on draft numbers and veteran status, but no income information

2SLS Mistakes

1. Doing 2SLS manually
2. Doing 2SLS manually and/or forget to add the same covariates in both the 1st and 2st
3. Doing 2SLS manually and/or use nonlinear model for 1st

Questions and Discussions

- Q: Does Competition Among Public Schools Benefit Students and Taxpayers?
  - Competition is good, but also economics of scale
  - We can run a OLS, are we done?
  - Direction of Bias?
- How to measure the degree of competition among public school districts within metropolitan areas?
- IV: number of streams (cross section variations)
  - What are the key ID assumptions?
  - Table 1: What can we conclude from the 1st reg table?
- Table 2 and 2a, 3, 4 and 4a: Compare OLS with IV estimates
- Aside: many other competing policies: charter schools, voucher plans, catholic schools, private schools, competitions among school districts. . .

References I

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

Hoxby, Caroline Minter. 1994. *Does Competition Among Public Schools Benefit Students and Taxpayers?* National Bureau of Economic Research.