ECON220B Discussion Section 6

Instrumental Variable

Lapo Bini

UC San Diego

March 18, 2026

ROADMAP

1. **Endogeneity**: Problem & Solution
2. **Wald** and **TSLS**
3. **Weak** IV
4. **Multiple** Instruments
5. **Control Function**

INTRODUCTION

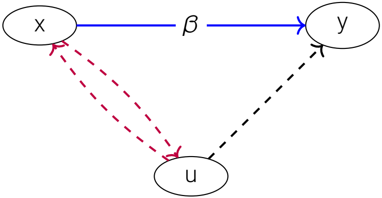
- **Problem**

  We cannot observe all the **confounders** (or we don't know them) and our regressor of interest is **correlated** with those unobservable characteristics.

  Model: $y_i = \alpha + x_i\beta + u_i$ Issue: $\text{Cov}(x_i, u_i) \neq 0$ Consequence: $\hat{\beta} = \beta + \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}$

- **Regression Endogeneity**

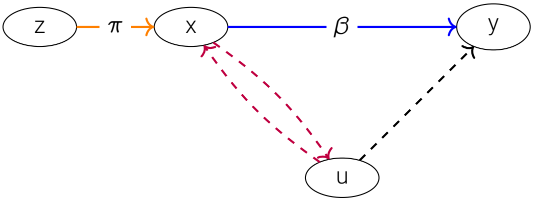
  OLS does not recover the structural parameter β. Instead, it converges to β plus a bias term proportional to the **covariance** between x_i and the error term.



INSTRUMENTAL VARIABLE

- **Intuition**

  Variation in x_i can be decomposed into two components: (i) one correlated with u_i; (ii) one uncorrelated with u_i. Let's find a third variable, the **instrument** z_i, to isolate the second one.



- **IV Conditions**

  What conditions must hold for a valid IV design?

  IV.1: **Relevance**: $\text{Cov}(x_i, z_i) \neq 0$

  IV.2: **Exogeneity**: $\text{Cov}(z_i, u_i) = 0$

  IV.3: **Exclusion**: z_i affects y_i only through u_i

DERIVING WALD ESTIMATOR

- **Identification**

  Under IV.1-IV.3, the structural parameter β is identified as the ratio of the **slope parameters** of two **reduced-form projections**:

$$y_i = \delta_0 + \delta z_i + \varepsilon_i$$
$$x_i = \pi_0 + \pi z_i + \eta_i$$

  Remember that **intercept** and **slope** parameters of a linear projection are defined as:

$$\begin{bmatrix} \delta_0 \\ \delta \end{bmatrix} := \underset{(d_0, d)' \in \mathbb{R}^2}{\arg\min} \; E[(y_i - d_0 - dz_i)^2] = \begin{bmatrix} E[y_i] - E[z_i]\delta \\ \text{Cov}(y_i, z_i)/\text{Var}(z_i) \end{bmatrix}$$

  $\text{Cov}(y_i, z_i) =$

IMPORTANCE OF EXCLUSION RESTRICTION

- **Violation**

  Exclusion restriction is violated if z_i has direct effect on y_i, i.e. we are **misspecifying** the **population causal model**:

$$y_i = \alpha + x_i\beta + \underbrace{z_i\phi + \nu_i}_{\text{previous error term}}$$

- **What Happen to Wald Estimator?**

  Suppose conditions IV.1 and IV-2 hold, let's study the probability limit of the Wald estimator:

$$\hat{\beta}_{\text{WALD}} =$$

TWO STAGE LEAST SQUARE

- **Identification**

  Under IV.1-IV.3, we use the instrument to **predict** the variation of x_i not associated with the unobservables, and use the predicted value to identify the **structural parameter** β.

$$y_i = \rho_0 + \rho\,\hat{x}_i + \xi_i \quad \text{(Second Stage)}$$

$$x_i = \underbrace{\pi_0 + \pi z_i}_{:=\hat{x}_i} + \eta_i \quad \text{(First Stage)}$$

$$\rho = \frac{\text{Cov}(y_i, \hat{x}_i)}{\text{Var}(\hat{x}_i)} =$$

- **Finite Sample Equivalence**

  Are $\hat{\beta}_{\text{TSLS}}$ and $\hat{\beta}_{\text{WALD}}$ equivalent in finite sample?

WEAK INSTRUMENT (I/II)

- Suppose the instrument has **no predictive power** ($\pi = 0$). What are the consequences for $\hat{\beta}_{TSLS}$? We want to derive an expression for $(\hat{\beta}_{TSLS} - \beta)$

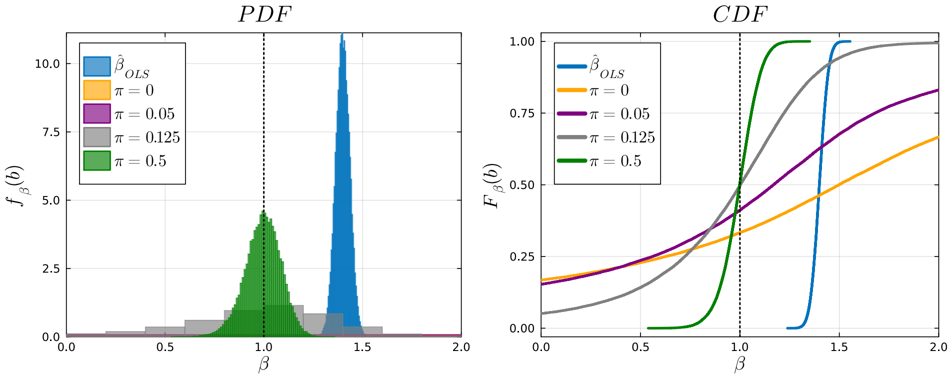
  $\hat{\beta}_{TSLS} =$

- Finite moments + iid assumption = CLT: $\left(\frac{1}{\sqrt{n}}\sum z_i u_i, \frac{1}{\sqrt{n}}\sum z_i \eta_i\right)' \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$,

  thus we can apply **Continuous Mapping Theorem**

WEAK INSTRUMENT (II/II)

Simulation: suppose $\beta = 1$ but $\text{Cov}(x_i, u_i) \neq 0$. We study the asymptotic distribution of $\hat{\beta}_{\text{OLS}}$ and $\hat{\beta}_{\text{IV}}$ under **different instrument strengths** (π). $N = 500$, for 5000 Monte Carlo replications.



- If $\pi \approx 0$ (local-to-zero), then $\hat{\beta}_{\text{IV}}$ is **inconsistent** (since β is not indentified)
- If $\pi \approx 0$ (local-to-zero), then $\hat{\beta}_{\text{IV}} - \beta \Longrightarrow \xi_u/\xi_\eta \sim \mathcal{N}/\mathcal{N}$
- If instrument is weak, $\hat{\beta}_{\text{IV}}$ is consistent but convergence is slow: **finite sample problem**.

MULTIPLE TREATMENT (I/IV)

- **Multidimensional model**

  Suppose we are now interested in $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i$ with $\dim(\boldsymbol{x}_i) = d$ and $\dim(\boldsymbol{z}_i) = k$.
  If $d = k$, the model is **just identified**. By method of moments:

$$E[\boldsymbol{z}_i u_i] = \mathbf{0} \qquad E[\boldsymbol{z}_i y_i] = E[\boldsymbol{z}_i \boldsymbol{x}_i']\boldsymbol{\beta} \qquad \boldsymbol{\beta} = E[\boldsymbol{z}_i \boldsymbol{x}_i']^{-1} E[\boldsymbol{z}_i y_i] =: \boldsymbol{\beta}_{IV}$$

  If $d < k$ model is **over-identified**, we can't take the inverse of $E[\boldsymbol{z}_i \boldsymbol{x}_i']$:

$$E[\boldsymbol{z}_i u_i] = \mathbf{0} \qquad \underset{(k\text{x}1)}{E[\boldsymbol{z}_i y_i]} = \underset{(k\text{x}d)}{E[\boldsymbol{z}_i \boldsymbol{x}_i']} \underset{(d\text{x}1)}{\boldsymbol{\beta}}$$

- **New First Stage**

  We have d endogenous variables (equations), and k predictors:

$$\boldsymbol{x}_i = \boldsymbol{\pi}\, \boldsymbol{z}_i + \boldsymbol{\eta}_i$$

MULTIPLE TREATMENT (II/IV)

- **TSLS Solution**

  The idea is to make the right-hand side invertible:

$$E[z_i y_i] = E[z_i x_i']\boldsymbol{\beta} \qquad M\, E[z_i y_i] = M\, E[z_i x_i']\boldsymbol{\beta} \qquad \boldsymbol{\beta} = \left(M\, E[z_i x_i']\right)^{-1} M\, E[z_i y_i]$$

  How do we chose M? Combine IVs according to their **predictiveness** of $\boldsymbol{x}_i$

MULTIPLE TREATMENT (III/IV)

- **Weighted Regression Representation**

  Claim 1: $\boldsymbol{\beta}_{TSLS}$ is equivalent to the weighted regression of reduced-form coefficients $\boldsymbol{\rho}$ on first stage coefficiens $\boldsymbol{\pi}$. Remember:

$$
\begin{aligned}
y_i &= z_i' \, \boldsymbol{\rho} + \varepsilon_i \\
\boldsymbol{x_i} &= \boldsymbol{\pi} \, z_i + \boldsymbol{\eta}_i
\end{aligned}
$$

  Claim 2: if $d = 1$, the TSLS coefficient is just a **linear combination** of k one-at-a-time Wald IVs, i.e. $\beta_{TSLS} = \sum_{j=1}^{k} \omega_j \beta_{WALD}^{(j)}$.

MULTIPLE TREATMENT (IV/IV)

- **Projection Interpretation of TSLS**

  Claim 3: Let $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, The TSLS estimator is equal to the OLS estimator obtained by regressing $\boldsymbol{Y}$ on the **predicted values** of $\boldsymbol{X}$ from the **first-stage** regression, that is:

$$\hat{\boldsymbol{\beta}}_{TSLS} = \left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1}\left(\hat{\boldsymbol{X}}'\boldsymbol{Y}\right)$$

CONTROL FUNCTION

- **Intuition**

  We can also decompose the **structural error term** into a component that is **correlated** with x_i and a component that is **orthogonal** to x_i.

$$y_i = \alpha + \beta x_i + u_i$$
$$x_i = \pi_0 + \pi z_i + \eta_i$$
$$u_i = \lambda \eta_i + \varepsilon_i$$

- **Identification**

  Under IV.1-IV.3, β is identified as the **slope coefficient** in the population regression of y_i on x_i **controlling for** the reduced-form **residual** η_i from the first-stage regression.

CONTROL FUNCTION

- **Finite Sample Equivalence**

  Can we establish that $\hat{\beta}_{WALD} = \hat{\beta}_{TSLS} = \hat{\beta}_{CF}$?