# Econometrics

**TA Session 7**

Lucia Sauer

2025-11-14

## Overview

- Instrumental Variables (IV)
- Two-Stage Least Squares (2SLS)
- Probit (IV-Probit)
- Marginal Effects
- Local Average Treatment Effect (LATE)

---

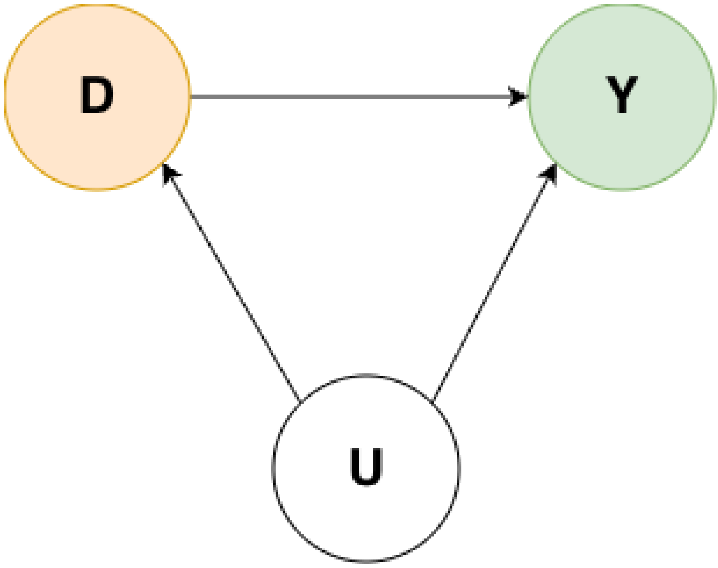
## Motivation

- Previously we have seen that RCT is the gold standard to identify causal effects.


---

## Ways to address endogeneity

However, in many situations, RCTs are not feasible or ethical and the presence of **confounders** can bias our estimates.

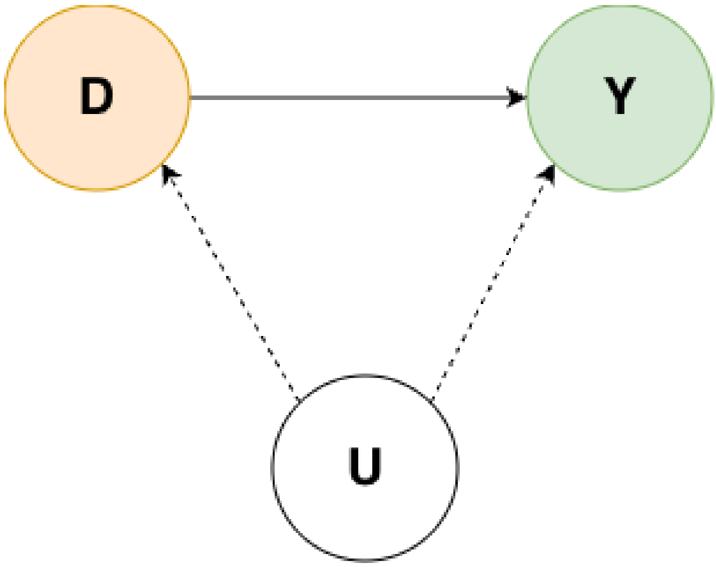



1. **Matching**: we try to control for all confounders, but this only works when the confounders are **observable**.

---

## Ways to address endogeneity

However, in many situations, RCTs are not feasible or ethical and the presence of **confounders** can bias our estimates.

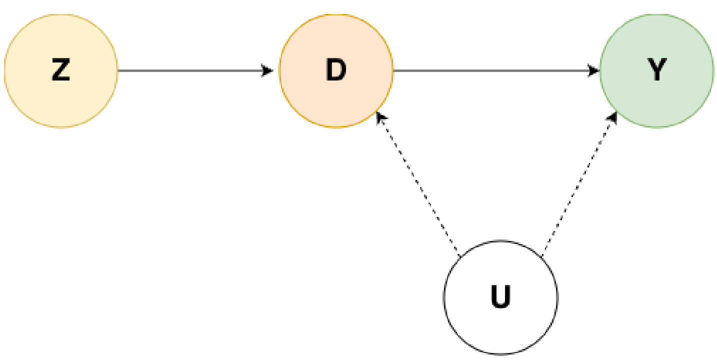



2. **Instrumental Variables (IV)**: use an external variable (instrument) that affects the treatment but is not directly related to the outcome, removing the bias from unobserved confounders.

---

### Instrumental Variables (IV)

- Assumptions for a valid instrument:
  1. Relevance: $Cov(Z, X) \neq 0$
  2. Exogeneity: $Cov(Z, u) = 0$




So when we critize IV, we usually focus on these two assumptions.

---

### Two-Stage Least Squares (2SLS)

1. First Stage: regress the endogenous variable X on the instrument Z and other exogenous variables to obtain the predicted values of X

$$X = \delta_0 + \delta_1 Z + v$$

2. Second Stage: regress the dependent variable Y on the predicted values $\hat{X}$ and other exogenous variables.

$$Y = \beta_0 + \beta_1 \hat{X} + \varepsilon$$

- Statistical software that supports 2SLS estimation directly (e.g., `ivreg` in R, `ivregress` in Stata, in python `linearmodels.iv`).

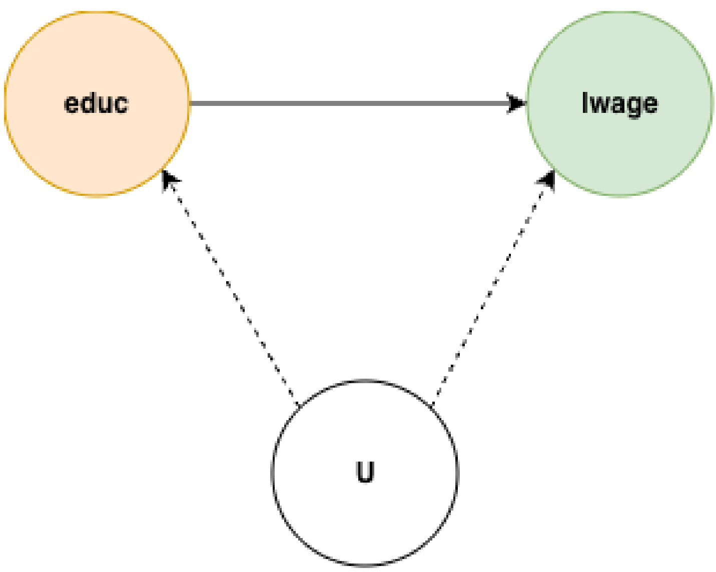
---

#### Returns of Education

We want to estimate the effect of education on wages.

$$\text{lwage}_i = \beta_1 + \beta_2 \text{exper}_i + \beta_3 \text{expersq}_i + \beta_4 \text{educ}_i + \varepsilon_i$$

However, `educ` may be endogenous due to omitted ability bias.



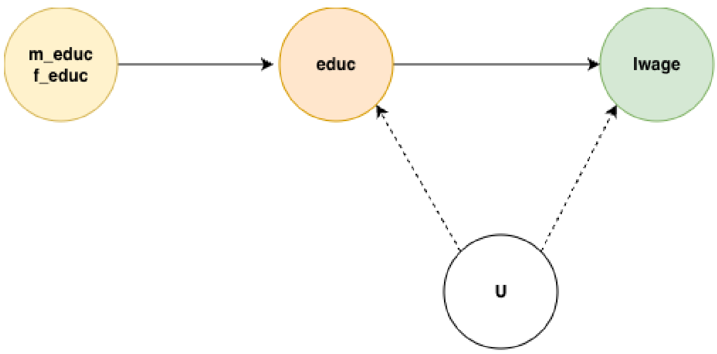

---

### Returns of Education - IV

We want to estimate the effect of education on wages.

$$\text{lwage}_i = \beta_1 + \beta_2 \text{exper}_i + \beta_3 \text{expersq}_i + \beta_4 \text{educ}_i + \varepsilon_i$$

- We can use `fatheduc` and `motheduc` as instruments for `educ`:
  - Relevance: Parents' education likely affects child's education.
  - Exogeneity: Parents' education likely does not directly affect child's wage, except through child's education.




---

### Implementation of 2SLS in Stata

1. First Stage

```
use http://fmwww.bc.edu/ec-p/data/wooldridge/mroz, clear
regress educ fatheduc motheduc exper expersq if !missing(lwage)
```

---

### Implementation of 2SLS in Stata

2. Second Stage

```
predict educhat
regress lwage educhat exper expersq
```

---

### Implementation of 2SLS in Stata

Or we can use the built-in command for 2SLS:

```
ivregress 2sls lwage exper expersq (educ= fatheduc motheduc)
```

---

## Probit (IV-Probit)

We are interested in estimating the effect of household income on female labor force participation.

$$\text{fem\_work}_i = \beta_1 + \beta_2 \text{other\_inc}_i + \beta_3 \text{fem\_educ}_i + \beta_4 \text{kids}_i + \varepsilon_i$$

where:

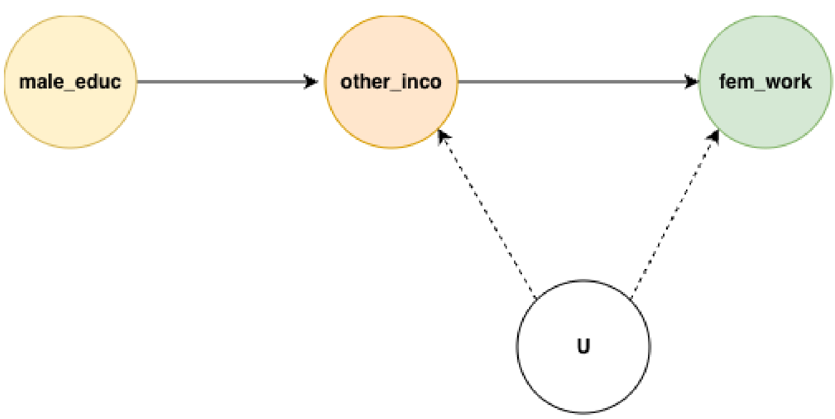
- fem_work: Female labor force participation
- other_inc: Other family income (endogenous)
- fem_educ: Female education level
- kids: Number of kids

But other_inc is likely endogenous due to omitted variable bias (e.g., family preferences)

---

### Endogeneity & Instrument

- `other_inc` correlated with unobserved family attitudes that also affect female labor participation.



- Use `male_educ` as instrument:
  - $\uparrow$ Education $\rightarrow \uparrow$ Income (**relevance**)
  - No direct effect on `fem_work` (**exogeneity**)
- Controls: `fem_educ`, `kids`

---

#### Non-Linear Model: Probit

- fem_work is a binary outcome (1 if working, 0 otherwise).
- OLS is not suitable for binary outcomes $\rightarrow$ may predict probabilities outside [0,1].
- Probit models the probability that $Y = 1$ using a normal CDF:

$$P(Y = 1|X) = \Phi(X\beta)$$

- Captures the **nonlinear** link between covariates and probability of working.

---

### Stata code for IV-Probit

```
clear all
webuse laborsup, clear
ivprobit fem_work fem_educ kids (other_inc = male_educ)
```

---

#### Interpretation of `kids = -0.18`

- A one-child increase reduces the **latent propensity** to work by **0.18 SD**, holding other factors constant.
- This is **not yet** a change in probability because the Probit model is **nonlinear**.

$$P(\text{fem\_work} = 1|X) = \Phi(X\beta)$$

The **marginal effect** of one more child on that probability is:

$$\frac{\partial P(\text{fem\_work} = 1|X)}{\partial \text{kids}} = \phi(X\beta) \cdot \beta_{\text{kids}}$$

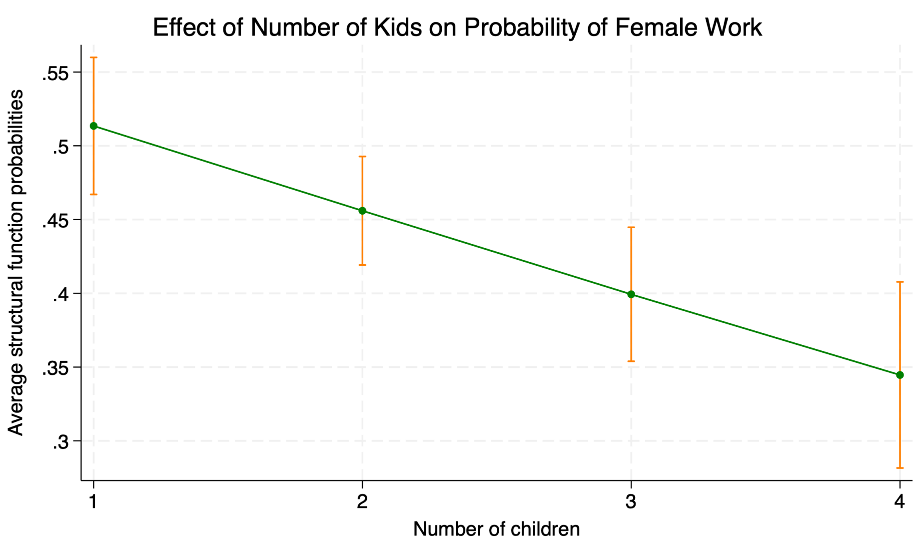
where $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and PDF of the standard normal.

---

## Marginal Effects

- **Treatment variable continuous**, we can compute the marginal effects to understand the impact of a one-unit increase in the treatment on the probability of the outcome.

```
margins, at(kids=(1(1)4)) predict(pr)
marginsplot
```




---

## Local Average Treatment Effect (LATE)

:auto-animate:

LATE

The IV estimate identifies the causal effect for compliers, i.e., those whose treatment decision responds to the instrument.

| | $\mathbf{Z = 0}$ | $\mathbf{Z = 1}$ |
|---|---|---|
| $\mathbf{D(0)}$ | complier o never-taker | defier o never-taker |
| $\mathbf{D(1)}$ | defier o always-taker | complier o always-taker |

---

## Local Average Treatment Effect (LATE)

:auto-animate:

LATE

The IV estimate identifies the causal effect for compliers, i.e., those whose treatment decision responds to the instrument.

| | $\mathbf{Z = 0}$ | $\mathbf{Z = 1}$ |
|---|---|---|
| $\mathbf{D(0)}$ | complier o never-taker | ~~defier~~ o never-taker |
| $\mathbf{D(1)}$ | ~~defier~~ o always-taker | complier o always-taker |

---

## Local Average Treatment Effect (LATE)

:auto-animate:

LATE

The IV estimate identifies the causal effect for compliers, i.e., those whose treatment decision responds to the instrument.

| | **Z = 0** | **Z = 1** |
|---|---|---|
| **D(0)** | **complier** o never-taker | ~~defier~~ o never-taker |
| **D(1)** | ~~defier~~ o always-taker | **complier** o always-taker |

---

### The Ghana Experiment

*"The Impact of Free Secondary Education: Experimental Evidence from Ghana"*, Duflo, Dupas & Kremer (2021), *QJE*

- Many students **qualify** for secondary school but **cannot afford to enroll**.
- Researchers implemented a **scholarship lottery**:
  - Surveyed students *who qualified for secondary school* but had not yet enrolled.
  - Randomly offered scholarships that covered school fees.

So instead of assigning education randomly, they conducted a randomized controlled trial (RCT) where they assigned scholarships to some students to attend secondary school.

---

### Variables

$$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$$

- **Outcome**:

$$Y_i = \text{post-intervention cognitive test score}$$

- **Treatment**:

$$D_i = 1 \text{ if the student actually attended secondary school}$$

- **Instrument**:

$$Z_i = 1 \text{ if student was randomly offered a scholarship}$$

---

### Imperfect Compliance

Getting a scholarship does not translate into **1 to 1** relationship with going to school:

> Defiers
>
> 1. Some scholarship winners ($Z = 1$) **still do not enroll** ($D = 0$)
> 2. Some students without scholarships ($Z = 0$) **manage to enroll anyway** ($D = 1$)

Therefore, the instrument $Z_i$ affects schooling $D_i$ **but imperfectly**, creating the classic compliance groups.

---

### Compliance Types

- **Always Takers** enroll **with or without** scholarship.

$$D(1) = 1, \quad D(0) = 1$$

- **Never Takers** do **not** enroll even if they receive a scholarship.

$$D(1) = 0, \quad D(0) = 0$$

- **Compliers (the key group)** enroll **only if** they receive the scholarship, these are the students **whose decisions change because of the instrument**.

$$D(1) = 1, \quad D(0) = 0$$

- **Defiers (usually ruled out)** do the **opposite** of the instrument assignment.

$$D(1) = 0, \quad D(0) = 1$$

  Assumption: **Monotonicity** $\rightarrow$ no defiers.

---

### 1. Effect of scholarship offer on enrollment

$$E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]$$

- Scholarship increases enrollment

- Many students enroll when offered
- Few enroll without scholarship

**2. Effect of scholarship offer on cognitive scores.**

$$E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]$$

Using expression for the outcome variable:

$$E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0] = \beta_1(E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0])$$

**Wald Estimator**

$$\hat{\beta}_{IV} = \frac{E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]}{E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]}$$

Causal effect of schooling.

---

## What Does IV Identify?

IV does **not** estimate the average effect for all students.

- Always Takers: scholarship does not change their attendance.
- Never Takers: scholarship does not change their attendance.
- Defiers: assumed not to exist.
- **Compliers**: the only group whose behavior is affected by the instrument → the only source of identifying variation.

Thus IV identifies a **Local** Average Treatment Effect.

$$\text{LATE} = E[Y(1) - Y(0) \mid \text{Compliers}]$$

**Interpretation:**

> The causal effect of secondary schooling **for students whose attendance decisions are actually changed by the scholarship offer**.

---

## Connection to Duflo-Dupas-Kremer (2021)

They write:

> "Our estimates capture the causal impact of secondary schooling for students whose enrollment decisions are affected by the scholarship offer."

This is **exactly** the definition of LATE.