Class 16 Instrumental Variables and Two-Stage Least Squares

Dr. Wei Miao

UCL School of Management

November 18, 2026

Section 1

Instrumental Variable

Class Objectives

- The requirements of a valid instrumental variable and how to find good instruments
- Intuition of why instrumental variables solve endogeneity problems
- Apply the two-stage least squares method to estimate the causal effects using instrumental variables

Causal Inference from OLS

- From non-experimental secondary data, it is impossible to control all confounding factors, which means we can never obtain causal effects from OLS regressions.
- Can we still obtain causal inference from secondary data?

What is an Instrumental Variable

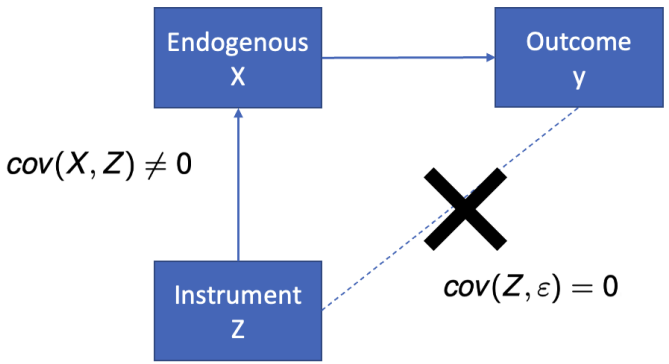
Instrumental Variable

An instrumental variable is a set of variables Z that satisfies the following requirements:

1. z is exogenous and uncorrelated with ϵ; that is, $cov(Z, \epsilon) = 0$
2. z only affects Y through X, but not directly affect Y
3. z affects x to some extent, that is, $cov(Z, x) \neq 0$

- Point 1 is the **exogeneity** requirement: the instrumental variable should be beyond an individual's control, such that the instrumental variables are uncorrelated with any of the individual's unobserved confounding factors.
  - Potential IVs: government policy; natural disasters; randomised experiment; birthdays; etc.
- Point 2 is the **exclusion restriction**: the instrumental variable should only affect Y through X, and not affect Y directly.
- Point 3 is the **relevance requirement**: although beyond an individual's control, the instrumental variable should still affect the individual's X, causing some exogenous changes in X that are beyond individual control.
  - If the correlation between z and x is too small, we have a **weak IV** problem.

Graphical Illustration of IV



A Classic Example of Instrumental Variable

Return of Military Service to Lifetime Income[1]

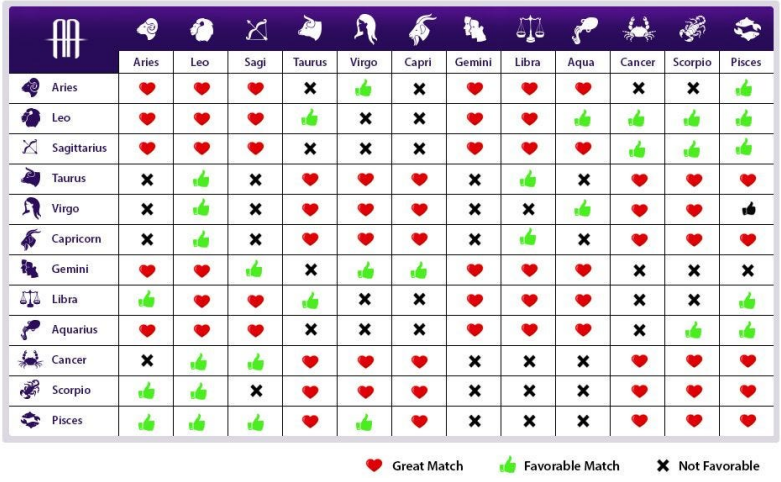
$$Income = \beta_0 + \beta_1 MilitaryService + \epsilon$$

- OLS suffers from endogeneity problems. What are the potential endogeneity issues?
- A lottery was used to determine whether soldiers with certain birthdays were drafted to the frontline.

[1] Angrist, Joshua D., Stacey H. Chen, and Jae Song. "Long-term consequences of Vietnam-era conscription: New estimates using social security data." *American Economic Review* 101, no. 3 (2011): 334-38.

A Classic Example of Instrumental Variable

- The date of birth (z) or zodiacs can be an **instrumental variable** for military service (x) in this case.
  - Relevance requirement: Affects years of military service: $cov(z, x) \neq 0$
  - Exogeneity requirement: Randomly drawn and thus uncorrelated with any confounders: $cov(z, \epsilon) = 0$
  - Exclusion restriction: z only affects Y through X, but not directly affect Y.



More Examples of IVs

Can you come up with IV candidates for the following causal questions?[2]

- Number of restaurants on UberEat => Number of orders on UberEat
- Retail price => Sales

[2] See html version for answers.

Section 2

Two-Stage Least Squares

Solving Endogeneity Using IV

- Given an endogenous OLS regression,

$$y_i = X_i\beta + \varepsilon_i, \quad \operatorname{cov}(X_i, \varepsilon_i) \neq 0$$

- Find instrumental variables Z_i that do not (directly) influence y_i , but are correlated with X_i

Two-Stage Least Squares: Stage 1

1. Run a regression of `X` on `Z`. The predicted value $\hat{X}$ is predicted by Z, and so should be uncorrelated with the error term ϵ.
   - $\hat{X}$ (the part of the changes in X that is due to Z) is exogenous, because Z is exogenous
   - All endogenous parts are now left over in the error term in the first-stage regression ϵ_i

$$X_i = Z_i\eta + \epsilon_i$$

Two-Stage Least Squares: Stage 2

2. Run a regression of Y on $\hat{X}$: now $\hat{X}$ is uncorrelated with the error term, and thus we can get a causal inference from the second-stage regression.

$$y_i = \hat{X}\beta + \varepsilon_i, \quad \text{cov}\left(\hat{X}_i, \varepsilon_i\right) = 0$$

After-Class Readings

- (optional) Econometrics with R: Instrumental Variables Regression