Class 16 Instrumental Variables

Dr Wei Miao

UCL School of Management

November 22, 2023

Section 1

Instrumental Variable

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.
- Is there still a way for us to obtain causal inference from secondary data?

What is an Instrumental Variable

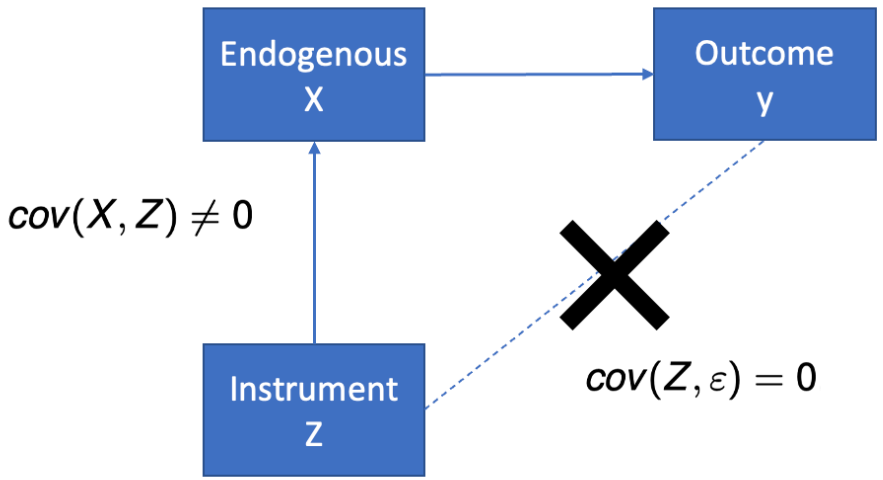
Instrumental Variable

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

1. z is exogeneous and uncorrelated with ϵ; that is, $cov(Z, \epsilon) = 0$
2. z affects x to some extent, that is, $cov(Z, x) \neq 0$

- Point 1 is called **exogeneity** requirement (**exclusion restriction** requirement): the instrumental variable should be beyond individual's control, such that the instrumental variables are uncorrelated with any individual's unobserved confounding factors.
  - Potential IVs: government policy; natural disasters; randomized experiment; etc.
- Point 2 is called **relevance requirement**: though beyond an individual's control, the instrumental variable should still affect the individual's X, causing some exogenous changes in X that is beyond individual control.

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, for example
  - Individual ability/health correlates with military service and affects income
- A lottery was used to determine if soldiers with certain birthdays are drafted.
- The date of birth (z) is an **instrumental variable**
  - Affects years of military service: $cov(z, x) \neq 0$
  - Randomly drawn and thus uncorrelated with any confounders: $cov(z, \epsilon) = 0$

[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.

IV Requirement I: Exogeneity

- Exogeneity requires that z should only affect Y through X, but not directly affect Y.
- The instrumental variable should be beyond an individual's control. Because omitted variable bias is often caused by individual's own selection, instrumental variables are thus not correlated with any omitted variables.

IV Requirement II: Relevance

- The instrumental variable must be sufficiently correlated with x.
- If the correlation between z and x is too small, we have a **weak IV** problem.
- For more mathematical details of the weak IV issue, refer to this resource.

More Examples of IVs

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

- COVID-19 cases => Uber Driver Supply
- Number of restaurants on UberEat => Number of orders on UberEat
- Retail price => Sales

[2] See html version for answers.

Section 2

Two-Stage Least Square

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 with `X ~ Z`. The predicted $\hat{X}$ is predicted by Z, which should be uncorrelated with the error term ϵ.
   - $\hat{X}$ (the part of changes in X 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 with $Y \sim \hat{X}$: now $\hat{X}$ is uncorrelated with the error term and thus we can get causal inference from the second stage regression.

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

Section 3

Application: Causal Effects of COVID-19

Causal Impact of COVID-19

- The COVID-19 pandemic has brought unprecedented disruptions to many industries, and platform businesses, especially sharing economy platforms, are among the most disrupted ones.
- How would you evaluate the causal impact of COVID-19 cases on the company's business and profits?
  - Can we collect data on the *COVID cases* and *sales* and run an OLS regression to get the causal effect? What would hinder us from causal inference from the above OLS regressions?

Causal Impact of COVID-19 on Uber Drivers' Labor Supply

- In this case study, we will see an application of instrumental variable in evaluating the causal impact of COVID-19 on Uber drivers' labor supply decision.
- Let's take out the Quarto document.

Beyond the Impact of COVID-19 on Labor Supply

- You can follow this case study and propose similar topics for your term 3 dissertation project, depending on the company you work with.
- For similar causal inference interview questions/data science tasks, when RCTs are difficult to implement, instrumental variable method can be a very powerful solution.

After-Class Readings

- (optional) Econometrics with R: Instrumental Variables Regression
- (highly recommended) Encouragement Designs and Instrumental Variables for A/B Testing at Spotify