

Class 16 Instrumental Variables

Dr Wei Miao

UCL School of Management

Thu, Nov 24 2022

Section 1

Instrumental Variable

Causal Inference from OLS

- The necessary condition for OLS to reveal causal effect is: all confounding variables (i.e., variables that are correlated with X and affect outcome variable) are controlled in the regression.
- Otherwise, the OLS estimator will be biased and we can only obtain the **total effect (correlation)** rather than the **direct effect (causal effect)**.
- From secondary data, we will **never** be able to control all confounding factors, which means we can never obtain causation from OLS regressions.
- So, 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 variable z that satisfies two requirements:

- ① z is uncorrelated with ϵ ; that is, $cov(z, \epsilon) = 0$
- ② z is correlated with x ; that is, $cov(z, x) \neq 0$

- Point 1 is called **exogeneity requirement**: the instrumental variable should be beyond an individual's control, such that the instrumental variable will not be correlated with any individual's choices/omitted variables.
 - The spirit is similar to RCT's randomization
- 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 variations in X .

Graphical Illustration of IV



A Classic Example of Instrumental Variable

Return of Military Service to Lifetime Income¹

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

- OLS suffers from endogeneity problems, for example
 - individual ability correlates with military service and affects income
 - individual health status correlates with military service and affects income
- A draft lottery was used to determine if a soldier with a certain birthday goes to the war.
- The date of birth (z) is an **instrumental variable**
 - is correlated with military service: $cov(z, x) \neq 0$
 - but does not directly affect income: $cov(z, \epsilon) = 0$

¹Angrist, 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 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 causation questions?²

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

²See html version for answers.

Section 2

Two-Stage Least Square

Solving Endogeneity Using IV

- Given an endogenous OLS regression,

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

- 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 $X \sim Z$. The predicted X is predicted by Z , which should be uncorrelated with the error term ϵ .
 - \hat{X} (predicted X from Z) is exogenous, because Z is exogenous
 - All endogenous parts are now left 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.

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

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.
- A common data science interview question: how would you evaluate causal impact of COVID-19 on the company's business and profits?
 - Can we collect data on the *COVID cases* and *KPI measures*, and run an OLS regression to get the causal effect? $KPI \sim NumCovid$
 - What would hinder us from causal inference from the above OLS regressions?

Causal Impact of COVID-19 on UberEat Delivery Drivers' Labor Supply

- In this case workshop, we will see an application of instrumental variable in evaluating the causal impact of COVID-19 on UberEat delivery 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.
 - The causal impact of COVID-19 on Uber/Bolt drivers' labor supply
 - The causal impact of COVID-19 on customer demand for offline shopping
 - etc.
- 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](#)