Business Analytics

Session 7b. Instrumental Variables

Renyu (Philip) Zhang

New York University Shanghai

Spring 2019

Introduction

- We are interested in establishing the causal effect of independent variable(s) X on outcome Y.
- Both X and Y are correlated with an omitted variable X'.
  - Directly regressing Y on X will result in some biased estimation.
  - X' is not observable.
  - Also called endogeneity in the literature (the variation in X is not completely exogenous).
  - E.g., more education leads to higher income (individual ability is an unobservable covariate).
- Solution: Find another covariate Z, correlated with X, but (conditionally) independent of X' (and, thus, Y).
  - Intuition: Establish the causal effect of Z on Y, and use this to infer the causal effect of X on Y.
  - Strong first stage: Z and X are correlated. Can be verified with data (weak-instrument test).
  - Exclusion restriction: Given X, Z is uncorrelated with X' and Y. Difficult to verify with data.

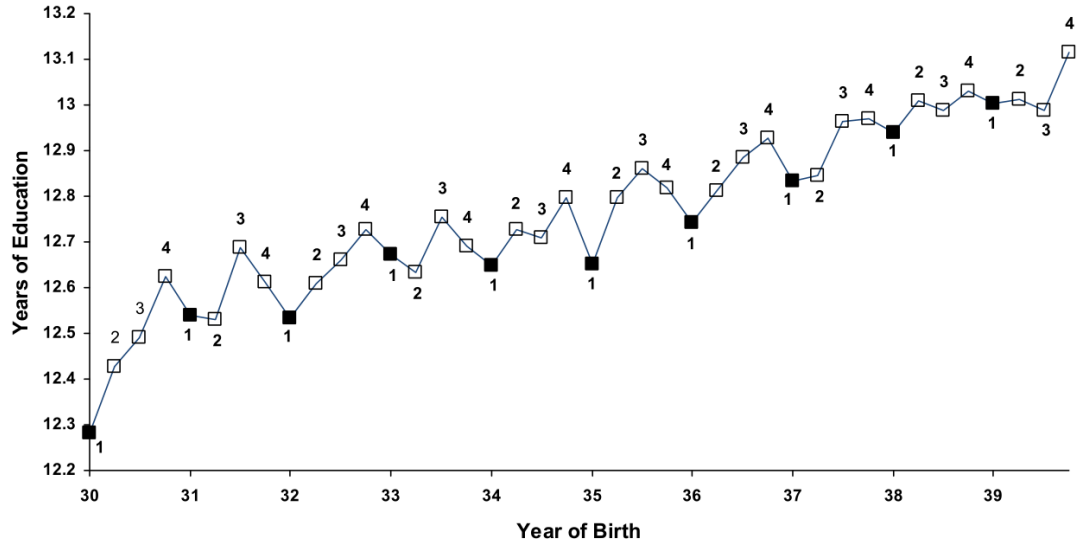
Examples of Instrumental Variables

Example: Education and Earnings

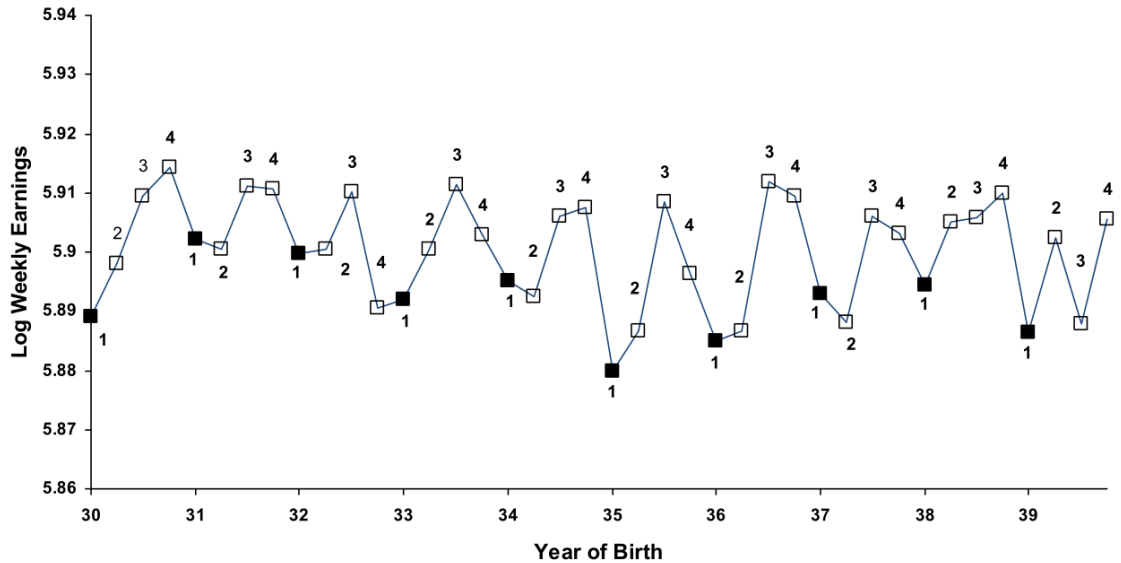
- Question: What is the effect of schooling on earnings?
- Omitted variable bias: We cannot directly observe individual ability, which is correlated with both schooling and earning.
- Instrumental Variable: Quarter of Birth.
  - IV is positively correlated with schooling, but independent with individual ability.
  - Studying causal effect in education: IVs related to costs and institutional constraints.

Example: Education and Earnings

A. Average Education by Quarter of Birth (first stage)



B. Average Weekly Wage by Quarter of Birth (reduced form)



Example: Supply Shocks

- Question: How to measure the sensitivity of price to the supply of the product?

$$Price \approx \hat{f}(Supply)$$

- Supply is endogenous: High prices will encourage greater supply.

- Instrumental Variable: Weather.
  - Weather will influence supply
  - Weather is not directly correlated with price

Example: Encouragement Design

- You launch a new product and want to measure satisfaction from use.
  - Randomized experiment is difficult because you cannot directly ask a customer to use your product, or not.
- You can use randomized encouragement to use the product: Randomly encourage half the population to use the product.
  - Discounts/promotions may be used.
- As long as:
  - encouragement is strongly correlated with usage; and
  - the act of encouragement itself is uncorrelated with satisfaction with the product,

  encouragement acts as an instrument for usage.

Causal Inference with Instrumental Variables

Estimating Causal Effect with IVs: Single Covariate Case

- X and Z are 1-dimensional.
- One unit of change in Z is associated with $\hat{\delta}$ unit change in X.
  - $X \approx \hat{\delta}_0 + \hat{\delta} Z$
- One unit of change in Z is associated with $\hat{\gamma}$ unit change in Y.
  - $Y \approx \hat{\gamma}_0 + \hat{\gamma} Z$
- Since Z is independent with Y, its effect on Y results from a causal effect of X on Y.
- The effect of X on Y is $\hat{\gamma}/\hat{\delta}$.
  - $Y \approx \hat{\gamma}/\hat{\delta} X + \cdots$

2-Stage Least Squares

- $\vec{X} = (X_1, X_2, \cdots, X_p)$ and $\vec{Z} = (Z_1, Z_2, \cdots, Z_q)$ are multi-dimensional.

Two-Stage Least Squares (2SLS)

1. Regress each covariate X_j ($j = 1, 2, \cdots, p$) on the IVs $\vec{Z}$.
2. Get the fitted covariate vector $\hat{\vec{X}}_i$ for each data point i ($i = 1, 2, \cdots, n$).
3. Regress the outcome Y on the fitted covariates $\hat{\vec{X}}$. The coefficients $(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p)$ are unbiased estimates of the causal effects of $\vec{X}$ on Y.

- Education and Earnings Example:
  - $Y =$ Weekly earning
  - $X =$ Year of schools
  - $\vec{Z} =$ Indicator for birth quarter

Concluding Thoughts about IVs

Challenge of Finding IVs

- IVs are typically found where there is exogenous variation that leads to changes in the covariate of interest.
- Exclusion restriction is hard to verify, especially when one only has observational data.
- Variation induced by randomized experiments can often be used as an IV. This is a common use case in IV analysis in high-tech industry.
  - To measure the satisfaction from using a product, use the randomized encouragement of adopting this product as an IV.

Homework

- Finish Homework 7 (NO need to submit it).

- We will briefly review predictive analytics and causal inference in Session 8b.
  - Prepare your questions.