Instrumental Variables: Theory

Huan Deng

May 14, 2025

Outline

- Introduction
  - Overview
  - Situations Where Endogeneity Arises
- Estimation
  - IV Estimator
  - Two Stage Least Squares
  - Control Function
  - GMM
- IV Under the Potential Outcome Framework
  - Going Beyond
  - Better LATE Than Nothing
- Discussion on Identification Assumptions
  - Exclusion Restriction
  - Monotonicity

Introduction

Introduction

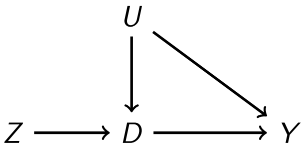
Overview

Overview

- Instrumental variable (IV) methods are fundamental to causal inference because it allows for unobserved confounders.
- We will first discuss "traditional" results on identification and estimation of IV under linear functional from and constant treatment effect
- Then we discuss the nonparametric identification of IV using the potential outcome framework
- Welcome to the world of unobservable heterogeneity!

Overview

- We can also use a directed acyclic graph (DAG) to illustrate IV.

  where D is the treatment indicator, Y is the outcome variable, U denotes unobserved confounder, and Z is the instrumental variable.
- IV invokes assumptions:
  - Relevance: there exists an arrow from Z to D
  - Exogeneity: no arrows between Z and U
  - Exclusion: no direct arrow from Z and Y

Introduction

Situations Where Endogeneity Arises

Simultaneous Equations of Supply and Demand

- Consider a simple model of supply and demand:

$$Q_d = \alpha_1 + \beta_1 P + \tau_1 Z_d + u_1 \tag{1}$$

$$Q_s = \alpha_2 + \beta_2 P + \tau_2 Z_s + u_2 \tag{2}$$

  where Q and P denote log of quantity and price
- We are interested in estimating the elasticity of demand or supply: β_1 or β_2
- The *Equilibrium* values of price and quantity can be of positive, negative, or no correlation at all, thus offering no interpretation as either demand or supply elasticities
- We can solve the system and easily find that $E[Pu_1] \neq 0$ and $E[Pu_2] \neq 0$
- We need a demand shifter to identify the supply equation, and vice versa.

Self Selection

- Suppose we want to estimate the return of college attendance
- D denotes college attendance and Y denotes the labor market outcomes.
- Y and D are positively correlated, can we interpret this as causal?
- College attendance is a deliberate choice, not an assignment.
- Those who have higher ability will be more likely to choose college and have higher wages.
- Even if college attendance has zero effect on wages, we can still observe a positive correlation between college attendance and wage

Measurement Error

- Consider a simple bivariate regression model: $Y = X'\beta + \epsilon$ where $E[X\epsilon] = 0$
- If X is measured with error, for which we specify a classical measurement model: $X^* = X + u$ where u is a measurement error. Then we estimate the below model:

$$Y = X^{*\prime}\beta + v, \quad v = \epsilon - u'\beta. \tag{3}$$

- Here, $E[X^* v] \neq 0$, leading to attenuation bias (the estimate shrinks towards 0):

$$\beta^* = \beta \left(1 - \frac{E[u^2]}{E[X^{*2}]}\right). \tag{4}$$

Estimation

Estimation

IV Estimator

Basics

- Consider a linear model:

$$Y_i = X_i'\beta + \epsilon_i \tag{5}$$

  where $E(X_i\epsilon_i) \neq 0$

- We can partition X_i into two components $[X_{i1}, X_{i2}]$, where $E[X_{i2}'\epsilon_i] = 0$ but $E[X_{i1}'\epsilon_i] \neq 0$
- We call X_{i1} endogenous regressors.
- We also have instruments $\tilde{Z}_i$ such that $E[\tilde{Z}_i\epsilon_i] = 0$

Basics

- Denote $Z_i = [\tilde{Z}_i, X_{i2}]$, so $E[Z_i \epsilon_i] = 0$
- We denote the length of $X_i, X_{i1}, X_{i2}, \tilde{Z}_i$ as K, K_1, K_2, J
- In some literature, identification is loosely defined as comparing the number of variables
  - $J < K_1$, under-identified
  - $J = K_1$, just-identified
  - $J > K_1$, over-identified

Basics

- Using matrix notations, we express three models:

$$Y = X\beta + \epsilon \quad \text{(the structural equation)} \tag{6}$$

$$X = Z\alpha + \eta \quad \text{(the first stage)} \tag{7}$$

$$Y = Z\gamma + \nu \quad \text{(the reduced form)} \tag{8}$$

- "Reduced form" here just refers to the model where we regress Y on all the exogenous variables Z.

IV Estimator

- Consider the case where $K_1 = J$
- We know that $E[Z_i(Y_i - X_i'\beta)] = 0$, which can be rewritten as:
$$E[Z_iX_i']\beta = E[Z_iY_i] \tag{9}$$
- Assume that $rank(E[Z_iX_i']) = K$, then we have $\beta = E[Z_iX_i']^{-1}E[Z_iY_i]$
- Then we can use the sample analogy:
$$\hat{\beta}_{IV} = (Z'X)^{-1}ZY \tag{10}$$
- We can also rewrite $\hat{\beta}$ as:
$$\begin{aligned}\hat{\beta}_{IV} &= (Z'X)^{-1}ZY \\ &= ((Z'Z)^{-1}Z'X)^{-1}(Z'Z)^{-1}Z'Y \\ &= \hat{\alpha}^{-1}\hat{\gamma} \end{aligned} \tag{11}$$

Estimation

Two Stage Least Squares

TSLS

- Consider the case where $K_1 < J$
- Remember that the first stage is just a projection of X on Z. This means that the correlation between X and ϵ is actually all captured by η and $P_Z X$ is uncorrelated with ϵ
- Regress Y on $P_Z X$, we get the below TSLS estimate:

$$\begin{aligned}\hat{\beta}_{TSLS} &= (X' P_Z P_Z X)^{-1} X' P_Z Y \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1} X'Z(Z'Z)^{-1}Z'Y \qquad (12)\end{aligned}$$

- We can easily show that $\hat{\beta}_{TSLS}$ is consistent:

$$\begin{aligned}\hat{\beta}_{TSLS} &= (X'Z(Z'Z)^{-1}Z'X)^{-1} X'Z(Z'Z)^{-1}Z'Y \\ &= \beta + (X'Z(Z'Z)^{-1}Z'X)^{-1} X'Z(Z'Z)^{-1}Z'\epsilon \\ &\xrightarrow{p} \beta \qquad (13)\end{aligned}$$

Estimation

Control Function

Control Function

- If we can decompose ϵ into two parts and control the part that is causing the endogeneity problem, then we are done!
- Project ϵ on η, we have the below model:

$$\epsilon = \eta\alpha + \delta \tag{14}$$

- Substitute the above equation back into the structural equation, we have:

$$Y = X\beta + \eta\alpha + \delta \tag{15}$$

- It is easy to check that now we have $E[X_i\delta_i] = E[\eta_i\delta_i] = 0$
- The above model is infeasible since we don't observe η. But we can estimate η

Control Function

- Thus we can use a two-step estimation method:
  - First, regress X on Z and get the residual $\hat{\eta}$
  - Second, regress Y on X and $\hat{\eta}$, get $\hat{\beta}_{cf}$
- We can prove that $\hat{\beta}_{cf} = \hat{\beta}_{TSLS}$. Below we will repeatedly use the project and annihilation matrix.
- By FWL, we know that

$$\hat{\beta}_{cf} = (X'M_{\eta}X)^{-1}X'M_{\eta}Y \tag{16}$$

- It suffices to prove that $M_{\eta}X = P_Z X$, which is true because

$$\begin{aligned} M_{\eta}X &= (I - P_{\eta})X \\ &= X - \eta(\eta'\eta)^{-1}\eta'X \\ &= X - M_z X(X'M_z X)^{-1}X'M_z X \\ &= P_z X \end{aligned} \tag{17}$$

Estimation

GMM

Generalized Methods of Moments

- Remember we have the moment condition:

$$E[Z_i\epsilon_i] = E[Z_i(Y_i - X_i'\beta)] = 0 \tag{18}$$

- Given a positive definite matrix W, we specify the below criterion function:

$$J(\beta) = N(Z'Y - Z'X\beta)'W(Z'Y - Z'X\beta) \tag{19}$$

- Take first-order derivative, we get the GMM estimate:

$$\hat{\beta}_{GMM} = (X'ZWZ'X)^{-1}X'ZWZ'Y \tag{20}$$

- When $W = (Z'Z)^{-1}$, $\hat{\beta}_{GMM} = \hat{\beta}_{TSLS}$

IV Under the Potential Outcome Framework

IV Under the Potential Outcome Framework

Going Beyond

Overview

- So far, we have assumed constant treatment effect (β is the same for all individuals) and a linear model
- Is the constant treatment effect assumption reasonable?
- Mogstad and Torgovitsky (2024): "Interesting treatment variables are often choices. Interesting outcome variables often reflect substantive consequences for the human beings under consideration. Human beings don't make choices randomly; they likely consider, at least in part, the effect that their choices may have on the outcome. These choices then become treatment variables that are associated with their effects on the outcome. Unless there's a compelling domain-specific reason to believe that the effect of the treatment cannot vary for some physical or institutional reason, then there will be UHTE that is systematically associated with the observed treatment choices."

Overview

- In addition, we want to go beyond the linear specification and ask: can we non-parametrically identify some causal estimands?
- Imbens and Angrist (1994) extend the potential outcome framework to IV
- You might have heard LATE (Local Average Treatment Effect), which can be shown to be identical with the IV estimate under the binary instrument and binary treatment scenario.

IV Under the Potential Outcome Framework

Better LATE Than Nothing

Overview

- Consider a binary instrument Z, binary treatment D, Y is the outcome variable
- We need to slightly modify the notation for potential outcomes
- Define $Y_i(D_i(Z_i), Z_i)$ and $D_i(Z_i)$ as two forms of potential outcomes, one for the outcome variable and one for the treatment.
- We need below four assumptions:
  - Exclusion Restriction: $Y_i(D_i(Z_i), Z_i) = Y_i(D_i(Z_i))$
  - Relevance: $E[D_i(1) - D_i(0)] \neq 0$
  - Exogeneity: $\{Y_i(0), Y_i(1), D_i(0), D_i(1)\} \perp Z_i$
  - Monotonicity: $D_i(1) \geq D_i(0)$ for all i or $D_i(0) \geq D_i(1)$ for all i

Choice Groups

- We partition each individual i into latent groups based on how his/her treatment status varies with the instrument $G_i = (D_i(0), D_i(1))$, for our simple case, we have below four groups:

	$D_i(1) = 0$	$D_i(1) = 1$
$D_i(0) = 0$	never taker	complier
$D_i(0) = 1$	defier	always taker

- The monotonicity assumption rules out either the compliers or the defiers.
- If we see the instrument as some "nudge", then it would be reasonable to assume away defiers.

Identifying LATE

- Consider the difference in means estimator for the outcome, $E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)$

$$\begin{aligned}
E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0) &= E(D_i(1)Y_i(1) + (1 - D_i(1))Y_i(0)|Z_i = 1) - \\
&\quad E(D_i(0)Y_i(1) + (1 - D_i(0))Y_i(0)|Z_i = 0) \\
&= E((D_i(1) - D_i(0))(Y_i(1) - Y_i(0))) \\
&= P(D_i(1) - D_i(0) = 1) \times E(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1) - \\
&\quad P(D_i(1) - D_i(0) = -1) \times E(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = -1)
\end{aligned} \tag{21}$$

- Monotonicity rules out defiers:

$$E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0) = P(D_i(1) - D_i(0) = 1) \times E(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1) \tag{22}$$

Identifying LATE

- Consider the difference in means estimator for the treatment, $E(D_i|Z_i = 1) - E(D_i|Z_i = 0)$

$$\begin{aligned} E(D_i|Z_i = 1) - E(D_i|Z_i = 0) &= E(D_i(1)|Z_i = 1) - E(D_i(0)|Z_i = 0) \\ &= E((D_i(1) - D_i(0))) \\ &= P(D_i(1) - D_i(0) = 1) \end{aligned} \tag{23}$$

- Thus we show that we can identify the average treatment effect for compliers:

$$\begin{aligned} \tau_{LATE} &= E(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1) \\ &= \frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)} \end{aligned} \tag{24}$$

Causal Interpretation of IV

- consider a linear model of IV

$$Y_i = \alpha + \beta D_i + \epsilon_i \tag{25}$$

$$D_i = \gamma + \theta Z_i + \eta \tag{26}$$

- We know that the IV estimator gives:

$$\begin{aligned}
\beta_{IV} &= \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)} \\
&= \frac{E(Y_i Z_i) - E(Y_i)E(Z_i)}{E(D_i Z_i) - E(D_i)E(Z_i)} \\
&= \frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)} \\
&= \tau_{LATE}
\end{aligned} \tag{27}$$

Causal Interpretation of IV

- Can we generalize the “IV is LATE” result to more general case where we have non-binary treatment or non-binary instrument?
- This is tricky. When we have multi-valued treatment or instrument, the simple IV estimand might not have a causal interpretation, i.e., we can’t express IV estimand as a positively weighted average of basic causal estimands.
- Mogstad and Torgovitsky (2024) offer an up-to-date discussion on this topic.

Discussion on Identification Assumptions

Discussion on Identification Assumptions

Exclusion Restriction

Randomness Doesn't Guarantee Exclusion

- The exclusion restriction is controversial, even if the instrumental is randomly assigned.
- Random assignment of Z is only enough to identify the treatment effect of Z on D and Y, but not for LATE.
- Angrist, Imbens and Rubin (1996, JASA) gives an example.
- D is military service, Y is lifetime earnings, and Z is draft lottery.
- The lottery is randomly assigned based on birthday, but people with a unfavorable lottery can choose to go to school to avoid military service, and schooling can affect earnings.
- Although the lottery is purely random, it may fail to satisfy the exclusion restriction.

Rain, Rain, Go Away

- Another example comes from Sarsons (2015)
- Rainfall is a popular IV for studies on the effects of income on conflicts
- Sarsons (2015) find that rainfall is uncorrelated with production in dam-fed districts but still impacts riot incidence.
- This suggests that rainfall affects conflict through some other channel.
- Mellon (2024) conducts a meta-analysis of 289 studies that use weather as instruments and find 194 potential exclusion-restriction violations

Discussion on Identification Assumptions

Monotonicity

Monotonicity Implies Strong Behavioral Restrictions

- Remember that monotonicity requires that the instrument push *everyone* in the same direction.
- When the instrument is money as in some incentivized experiments, it is plausible to assume that more money will weakly push everyone from nonparticipation to participation.
- But in other contexts, especially when instrument has no natural ordering, monotonicity can be controversial.

Gender Composition on Fertility

- In Angrist and Evans (1998), to study the effect of fertility on labor supply, the authors used the sex composition of a family's existing children as an instrument for further childbearing, assuming that families prefer to have both a boy and a girl than two boys or two girls
- This is a strong restriction on people's preference.
- It is reasonable to doubt that some parents prefer to have two boys or two girls, and they will have a third baby only if the first two are not of the same gender.

Judge IV

- Judge designs are based on institutionally-prescribed random assignment of a judge to cases in which the judge chooses treatment.
- Some judges are systematically more/less likely to assign treatment (incarnation), then the judge identities serve as an instrument for treatment.
- One judge is stricter than another judge on *every case*
- This effectively prevents judges from systematically disagreeing based on types of crime or features of the suspects.

Multiple Instruments

- The credibility of the monotonicity condition is not about the cardinality of the instrument, but about whether it has a natural ordering.
- This ordering issue arises when Z is a vector containing multiple distinct components.
- Monotonicity becomes unattractive with multiple instruments even if each instrument has a natural ordering.

Multiple Instruments

- Consider two binary instruments for school attendance where one indicates monetary incentive and another denotes distance
- We know people prefer a short distance holding monetary incentive constant; and people prefer a higher monetary incentive holding distance constant.
- But would it be plausible to assume that everyone prefers a combination of a short distance and low monetary incentive to a combination of a long distance and high monetary incentive?
- Monotonicity here assumes away heterogeneity in people's opportunity cost of time.