150C Causal Inference

Instrumental Variables: Modern Perspective with Heterogeneous Treatment Effects

Jonathan Mummolo

May 22, 2017

Two Views on Instrumental Variables

1. Traditional Econometric Framework
   - Constant treatment effects
   - Linearity in the case of a multivalued treatment
2. Potential Outcome Model of IV
   - Heterogeneous treatment effects
   - Focus in Local Average Treatment Effect (LATE)

Identification with Traditional Instrumental Variables

Definition

Two equations:

- $Y = \gamma + \alpha D + \varepsilon$ (Second Stage)
- $D = \tau + \rho Z + \eta$ (First Stage)

Identification Assumption

1. *Exogeneity and Exclusion:* $Cov(Z, \eta) = 0$ *and* $Cov(Z, \varepsilon) = 0$
2. *First Stage:* $\rho \neq 0$
3. $\alpha = Y_{1,i} - Y_{0,i}$ *constant for all units i.*
   Or in the case of a multivalued treatment with s levels we assume $\alpha = Y_{s,i} - Y_{s-1,i}$.

Instrumental Variable Estimator

- True model: $Y = D\alpha + X\beta + \varepsilon$
- Given the IV assumptions, we could regress: $Y = Z\rho + \omega$ and obtain an unbiased effect $\hat{\rho}$, the effect of Z on Y
- But we can also obtain an unbiased estimate of β, the effect of D on Y by using only the exogenous variation in D that is induced by Z

Assume $Cov[\nu = \varepsilon + X\beta, Z] = 0$.

Outline

1. Instrumental Variables with Potential Outcomes (No Covariates)
   - Identification
   - Estimation
   - Examples
   - Size of Complier Group

Potential Outcome Model for Instrumental Variables

Definition (Instrument)

Z_i: Binary instrument for *unit i*.

$$Z_i = \begin{cases} 1 & \text{if unit } i \text{ ``encouraged'' to receive treatment} \\ 0 & \text{if unit } i \text{ ``encouraged'' to receive control} \end{cases}$$

Definition (Potential Treatments)

D_z indicates *potential* treatment status given $Z = z$

- $D_1 = 1$ encouraged to take treatment and takes treatment

Assumption

Observed treatments are realized as

$$D = Z \cdot D_1 + (1 - Z) \cdot D_0 \text{ so } D_i = \begin{cases} D_{1i} & \text{if } Z_i = 1 \\ D_{0i} & \text{if } Z_i = 0 \end{cases}$$

Potential Outcome Model for Instrumental Variables

Following Angrist, Imbens, and Rubin (1996), we can define:

Definition

- *Compliers*: $D_1 > D_0$ ($D_0 = 0$ and $D_1 = 1$).
- *Always-takers*: $D_1 = D_0 = 1$.
- *Never-takers*: $D_1 = D_0 = 0$.
- *Defiers*: $D_1 < D_0$ ($D_0 = 1$ and $D_1 = 0$).

Problem

Only one of the potential treatment indicators (D_0, D_1) *is observed, so we cannot identify which group any particular individual belongs to*

Who are the Compliers?

Study	Outcome	Treatment	Instrument
Angrist and Evans (1998)	Earnings	More than 2 Children	Multiple Second Birth (Twins)
Angrist and Evans (1998)	Earnings	More than 2 Children	First Two Children are Same Sex
Levitt (1997)	Crime Rates	Number of Policemen	Mayoral Elections
Angrist and Krueger (1991)	Earnings	Years of Schooling	Quarter of Birth
Angrist (1990)	Earnings	Veteran Status	Vietnam Draft Lottery
Miguel, Satyanath and Sergenti (2004)	Civil War Onset	GDP per capita	Lagged Rainfall
Acemoglu, Johnson and Robinson (2001)	Economic performance	Current Institutions	Settler Mortality in Colonial Times
Cleary and Barro (2006)	Religiosity	GDP per capita	Distance from Equator

Potential Outcome Model for Instrumental Variables

Definition (Potential Outcomes)

Given the binary instrument $Z_i \in (1, 0)$ and the binary treatment $D_i \in (1, 0)$ every unit now has four potential outcomes $Y_i(D, Z)$:

- $Y(D = 1, Z = 1)$; $Y(D = 1, Z = 0)$; $Y(D = 0, Z = 1)$; $Y(D = 0, Z = 0)$

e.g. the causal effect of the treatment given the unit's realized encouragement status is given by $Y(D = 1, Z_i) - Y(D = 0, Z_i)$.

Assumption (Ignorability)

Ignorability of the Instrument: $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z$

- *Independence:* $(Y(D, Z), D_1, D_0) \perp\!\!\!\perp Z$ *which implies that causal effects of* Z *on* Y *and* Z *on* D *are identified.*
- *Exclusion:* $Y(D, 0) = Y(D, 1)$ *for* $D = 0, 1$ *so we can simply define potential outcomes indexed solely by treatment status:* (Y_1, Y_0)

Potential Outcome Model for Instrumental Variables

Estimand (LATE)

$\alpha_{LATE} = E[Y_1 - Y_0 | D_1 > D_0]$ *is defined as the Local Average Treatment Effect for Compliers*

- *This estimand varies with the particular instrument Z*

Proposition (Special Cases)

- *When the treatment intake, D, is itself randomized,*

Potential Outcome Model for Instrumental Variables

Estimand (LATE)

$\alpha_{LATE} = E[Y_1 - Y_0 | D_1 > D_0]$ *is defined as the Local Average Treatment Effect for Compliers*

- *This estimand varies with the particular instrument Z*

Proposition (Special Cases)

- *When the treatment intake, D, is itself randomized, then* $Z = D$ *and every individual is a complier*
- *Given one-sided noncompliance,* $D_0 = 0$*:*

Potential Outcome Model for Instrumental Variables

Estimand (LATE)

$\alpha_{LATE} = E[Y_1 - Y_0 | D_1 > D_0]$ *is defined as the Local Average Treatment Effect for Compliers*

- *This estimand varies with the particular instrument Z*

Proposition (Special Cases)

- *When the treatment intake, D, is itself randomized, then $Z = D$ and every individual is a complier*
- *Given one-sided noncompliance, $D_0 = 0$:*

$$E[Y_1|D_1 > D_0] = E[Y_1|D_1 = 1] = E[Y_1|Z = 1, D_1 = 1] = E[Y_1|D = 1]$$

, *and*

$$E[Y_0|D_1 > D_0] = E[Y_0|D = 1]$$

so $\alpha_{LATE} = E[Y_1 - Y_0|D_1 > D_0] = E[Y_1 - Y_0|D = 1] = \alpha_{ATET}$

Outline

1. Instrumental Variables with Potential Outcomes (No Covariates)
   - Identification
   - Estimation
   - Examples
   - Size of Complier Group

Identification with Instrumental Variables

Identification Assumption

1. *Ignorability of the Instrument:* $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z$
2. *First Stage:* $0 < P(Z = 1) < 1$ *and* $P(D_1 = 1) \neq P(D_0 = 1)$
3. *Monotonicity:* $D_1 \geq D_0$

Identification Result

$$\begin{aligned} E[Y_1 - Y_0 | D_1 > D_0] &= \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} \\ &= \frac{\textit{Intent to Treat Effect of Z on Y}}{\textit{First Stage Effect of Z on D}} \\ &= \frac{\textit{Intent to Treat Effect}}{\textit{Proportion of Compliers}} \end{aligned}$$

Identification with Instrumental Variables

Identification Assumption

1. *Ignorability of the Instrument:* $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z$
2. *First Stage:* $0 < P(Z = 1) < 1$ *and* $P(D_1 = 1) \neq P(D_0 = 1)$
3. *Monotonicity:* $D_1 \geq D_0$

Proof.

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} = \frac{E[Y_0 + (Y_1 - Y_0)D_1|Z=1] - E[Y_0 + (Y_1 - Y_0)D_0|Z=0]}{E[D_1|Z=1] - E[D_0|Z=0]}$$

$$= \frac{E[Y_0 + (Y_1 - Y_0)D_1] - E[Y_0 + (Y_1 - Y_0)D_0]}{E[D_1] - E[D_0]} = \frac{E[(Y_1 - Y_0)(D_1 - D_0)]}{E[D_1 - D_0]}$$

$$= \frac{E[Y_1 - Y_0|D_1 > D_0]P(D_1 > D_0) - E[Y_1 - Y_0|D_1 < D_0]P(D_1 < D_0)}{E[D_1 - D_0]} \text{ as } (D_1 - D_0) = (1, 0, -1)$$

$$= \frac{E[Y_1 - Y_0|D_1 > D_0]P[D_1 > D_0]}{P(D_1 > D_0)} = E[Y_1 - Y_0|D_1 > D_0]$$

□

Identification Assumptions

- Ignorability of the Instrument: $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z$
  - Implies that Z is randomly assigned so that the intent to treat effect and first stage effect are causally identified
  - $Y(d, z)$ implies exclusion restriction so that $Y(d, 0) = Y(d, 1)$ for $d = (1, 0)$. Rules out independent effect of Z on Y
  - Allows to attribute correlation between Z and Y to the effect of D alone; assumption is not testable
    - Random assignment is not a sufficient condition for exclusion.
- First Stage: $0 < P(Z = 1) < 1$ and $P(D_1 = 1) \neq P(D_0 = 1)$
  - Implies that the instrument Z induces variation in D
  - Testable by regressing D on Z
- Monotonicity: $D_1 \geq D_0$
  - Rules out defiers
  - Often easy to assess from institutional knowledge

Outline

Instrumental Variable: Estimators

Estimand (LATE)

$$E[Y_1 - Y_0 | D_1 > D_0] = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} \left(= \frac{cov(Y,Z)}{cov(D,Z)} \right)$$

Estimator (Wald Estimator)

The sample analog estimator is:

$$\left(\frac{\sum_{i=1}^{N} Y_i Z_i}{\sum_{i=1}^{N} Z_i} - \frac{\sum_{i=1}^{N} Y_i (1 - Z_i)}{\sum_{i=1}^{N} (1 - Z_i)} \right) \Big/ \left(\frac{\sum_{i=1}^{N} D_i Z_i}{\sum_{i=1}^{N} Z_i} - \frac{\sum_{i=1}^{N} D_i (1 - Z_i)}{\sum_{i=1}^{N} (1 - Z_i)} \right)$$

Instrumental Variable: Estimators

Estimand (LATE)

$$E[Y_1 - Y_0 | D_1 > D_0] = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} \left(= \frac{cov(Y,Z)}{cov(D,Z)} \right)$$

Estimator (Wald Estimator as IV Regression)

Can also implement Wald Estimator using an IV regression:

$$Y = \mu + \alpha D + \varepsilon$$

where $E[\varepsilon|Z] = 0$*, so* $\alpha = cov(Y,Z)/cov(D,Z)$

To estimate α *we run the simple IV regression of* Y *on a constant and* D *and instrument* D *with* Z*.*

Instrumental Variable: Estimators

Estimand (LATE)

$$E[Y_1 - Y_0 | D_1 > D_0] = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} \left(= \frac{cov(Y,Z)}{cov(D,Z)} \right)$$

Estimator (Two Stage Least Squares)

If identification assumptions only hold after conditioning on X, covariates are often introduced using 2SLS regression:

$$Y = \mu + \alpha D + X'\beta + \varepsilon,$$

where $E[\varepsilon|X,Z] = 0$. Now α and β are computed regressing Y on D and X, and using Z and X as instruments.

In general, α estimated in this way does not necessarily have a clear causal interpretation (see Abadie (2003))

Outline

Example: The Vietnam Draft Lottery (Angrist (1990))

- Effect of military service on civilian earnings
- Simple comparison between Vietnam veterans and non-veterans are likely to be a biased measure
- Angrist (1990) used draft-eligibility, determined by the Vietnam era draft lottery, as an instrument for military service in Vietnam
- Draft eligibility is random and affected the probability of enrollment
- Estimate suggest a 15% effect of veteran status on earnings in the period 1981-1984 for white veterans born in 1950-51; although the estimators are quite imprecise

Wald Estimates for Vietnam Draft Lottery (Angrist (1990))

		Draft-Eligibility Effects in Current $				
Cohort	Year	FICA Earnings (1)	Adjusted FICA Earnings (2)	Total W-2 Earnings (3)	$\hat{p}^e - \hat{p}^n$ (4)	Service Effect in 1978 $ (5)
1950	1981	−435.8	−487.8	−589.6	0.159	−2,195.8
		(210.5)	(237.6)	(299.4)	(0.040)	(1,069.5)
	1982	−320.2	−396.1	−305.5		−1,678.3
		(235.8)	(281.7)	(345.4)		(1,193.6)
	1983	−349.5	−450.1	−512.9		−1,795.6
		(261.6)	(302.0)	(441.2)		(1,204.8)
	1984	−484.3	−638.7	−1,143.3		−2,517.7
		(286.8)	(336.5)	(492.2)		(1,326.5)
1951	1981	−358.3	−428.7	−71.6	0.136	−2,261.3
		(203.6)	(224.5)	(423.4)	(0.043)	(1,184.2)
	1982	−117.3	−278.5	−72.7		−1,386.6
		(229.1)	(264.1)	(372.1)		(1,312.1)
	1983	−314.0	−452.2	−896.5		-2,181.8
		(253.2)	(289.2)	(426.3)		(1,395.3)
	1984	−398.4	−573.3	−809.1		−2,647.9
		(279.2)	(331.1)	(380.9)		(1,529.2)
1952	1981	−342.8	−392.6	−440.5	0.105	−2,502.3
		(206.8)	(228.6)	(265.0)	(0.050)	(1,556.7)
	1982	−235.1	−255.2	−514.7		−1,626.5
		(232.3)	(264.5)	(296.5)		(1,685.8)
	1983	−437.7	−500.0	−915.7		−3,103.5
		(257.5)	(294.7)	(395.2)		(1,829.2)
	1984	−436.0	−560.0	−767.2		−3,323.8

Example: Minneapolis Domestic Violence Experiment

- Minneapolis Domestic Violence Experiment was first field experiment to examine effectiveness of methods used by police to reduce domestic violence (Sherman and Berk 1984)
- **Sample**: 314 cases of male-on-female spousal assault in two high-density precincts, in which both parties present at scene. 51 patrol officers participated in the study.
- **Treatments**: Random assignment of cases to one of three approaches:
  - Send the abuser away for eight hours
  - Advice and mediation of disputes
  - Make an arrest
- **Outcome**: 6-month follow-up period, with both victims and offenders, as well as official records consulted to determine whether or not re-offending had occurred

Non-Compliance In Minneapolis Experiment

Table 1: Assigned and Delivered Treatments in Spousal Assault Cases

Assigned Treatment	Delivered Treatment			
		Coddled		
	Arrest	Advise	Separate	Total
Arrest	98.9 (91)	0.0 (0)	1.1 (1)	29.3 (92)
Advise	17.6 (19)	77.8 (84)	4.6 (5)	34.4 (108)
Separate	22.8 (26)	4.4 (5)	72.8 (83)	36.3 (114)
Total	43.4 (136)	28.3 (89)	28.3 (89)	100.0(314)

Notes: The table shows statistics from Sherman and Berk (1984), Table 1.

ITT Effect in Minneapolis Experiment

Table 2. First stage and reduced forms for Model 1.

	Endogenous variable is coddled			
	First stage		*Reduced form (ITT)*	
	(1)	*(2)**	*(3)*	*(4)**
Coddled-assigned	0.786 (0.043)	0.773 (0.043)	0.114 (0.047)	0.108 (0.041)
Weapon		−0.064 (0.045)		−0.004 (0.042)
Chem. influence		−0.088 (0.040)		0.052 (0.038)
Dep. var. mean	0.567 (Coddled–delivered)		0.178 (V Failed)	

The table reports OLS estimates of the first-stage and reduced form for Model 1 in the text.
*Other covariates include year and quarter dummies, and dummies for non-white and mixed race.

Treatment Effect in Minneapolis Experiment

Table 3. OLS and 2SLS estimates for Model 1.

	Endogenous variable is coddled			
	OLS		*IV/2SLS*	
	(1)	*(2)**	*(3)*	*(4)**
Coddled–delivered	0.087 (0.044)	0.070 (0.038)	0.145 (0.060)	0.140 (0.053)
Weapon		0.010 (0.043)		0.005 (0.043)
Chem. influence		0.057 (0.039)		0.064 (0.039)

The Table reports OLS and 2SLS estimates of the structural equation in Model 1.
*Other covariates include year and quarter dummies, and dummies for non-white and mixed race.

Outline

Estimating the Size of the Complier Group

- Since we never observe both potential treatment assignments for the same unit, we cannot identify individual units as compliers
- However, we can easily identify the proportion of compliers in the population using the first stage effect:

$$\begin{aligned} P(D_1 > D_0) &= E[D_1 - D_0] = E[D_1] - E[D_0] \\ &= E[D|Z=1] - E[D|Z=0] \end{aligned}$$

- Using a similar logic we can identify the proportion of compliers among the treated or controls only. For example:

$$P(D_1 > D_0|D=1) = \frac{P(Z=1)(E[D|Z=1] - E[D|Z=0])}{P(D=1)}$$

Size of Complier Group

TABLE 4.4.2
Probabilities of compliance in instrumental variables studies

Source (1)	Endogenous Variable (D) (2)	Instrument (Z) (3)	Sample (4)	$P[D=1]$ (5)	First Stage, $P[D_1 > D_0]$ (6)	$P[Z=1]$ (7)	Compliance Probabilities: $P[D_1 > D_0 \mid D=1]$ (8)	Compliance Probabilities: $P[D_1 > D_0 \mid D=0]$ (9)
Angrist (1990)	Veteran status	Draft eligibility	White men born in 1950	.267	.159	.534	.318	.101
			Non-white men born in 1950	.163	.060	.534	.197	.033
Angrist and Evans (1998)	More than two children	Twins at second birth	Married women aged 21–35 with two or more children in 1980	.381	.603	.008	.013	.966
		First two children are same sex		.381	.060	.506	.080	.048
Angrist and Krueger (1991)	High school graduate	Third- or fourth-quarter birth	Men born between 1930 and 1939	.770	.016	.509	.011	.034
Acemoglu and Angrist (2000)	High school graduate	State requires 11 or more years of school attendance	White men aged 40–49	.617	.037	.300	.018	.068

Notes: The table computes the absolute and relative size of the complier population for a number of instrumental variables. The first stage, reported in column 6, gives the absolute size of the complier group. Columns 8 and 9 show the size of the complier population relative to the treated and untreated populations.

Precision for LATE Estimation

- When N is large the standard error on the instrumental variable estimator of the LATE is approximately

$$SE_{\widehat{LATE}} \approx \frac{SE_{\widehat{ITT}}}{Compliance\ Ratio}$$

  - In JTPA data we get $330/.62 = 532$ which is close to the standard error estimate from the instrumental variable regression of 526.
- Two estimates converge if there is perfect compliance
- Otherwise, all else equal, the standard error on the LATE decreases linearly with the compliance!
  - If compliance ratio drops from 100% to 10%, the LATE standard error increases by a factor of 10
- Always wise to conduct a pilot to test the encouragement
- Design it to boost compliance, but do not violate exclusion restriction