

# Instrumental Variables

## Part II

---

Fernando Hoces la Guardia

07/28/2022

## Example #2: Solidifying Concepts and Exploring Selection Bias

- Policy issue: what was the most effective police response to reduce domestic violence in the US in the 1980s?
- Treatment: Advise or separate aggressor in case of domestic violence ("coddle").
- Control: Arrest the aggressor
- Outcome: 6-month recurrence of domestic violence in the same address (recidivism).
- Population: Minneapolis, 1980s, volunteer police, non-felony cases (probable cause for misdemeanor assault, not felonies).

# (Not so) Random Assignment

- Random assignment was done with colored coded pad reports. Let's think of hypothetical arrival of the police to the scene of crime.
- Two reasons why treatment assigned (“offered” in the KIPP example) was not the same as treatment delivered (“attended” in KIPP example):
  1. Some situations required judgment calls from police (e.g. arrest when assigned to coddle) and there was an understanding, between research team and the policy, that they could make such calls.
  2. In some cases officers forgot their color coded report pads (logistics matter!)

# Table 3.3 (Simplified)

	<b>Delivered</b>		
<b>Assigned</b>	Arrest (D=0)	Coddle (D=1)	total
Arrest (Z=0)	91	1	92
percent of Z=0	0.99	0.01	1
Coddle (Z=1)	45	177	222
percent of Z=1	0.203	0.797	1

# Non-Compliance

- The problem is lack of perfect compliance, this looks like a problem that IV can fix.
- Subset of assigned coddlers that ended up arrested are a non random subset of coddlers (think specially aggressive individuals, and how this might relate to recidivism).
- Simple comparison between coddlers and arrested are contaminated by selection bias

	Delivered		
Assigned	Arrest (D=0)	Coddle (D=1)	total
Arrest (Z=0)	91	1	92
percent of Z=0	0.99	0.01	1
Coddle (Z=1)	<b>45</b>	177	222
percent of Z=1	<b>0.203</b>	0.797	1

# Results: First Stage

- In English: effect of instrument on treatment
- Where is the instrument here? Where is the treatment?
- Where would you find the "average treatment, for those that received the instrument"?
- Now write it the term above as expectation.
- Repeat for "average treatment, for those that did not received the

	Delivered		
Assigned	Arrest (D=0)	Coddle (D=1)	total
Arrest (Z=0)	91	1	92
percent of Z=0	0.99	0.01	1
Coddle (Z=1)	45	177	222
percent of Z=1	0.203	0.797	1

# Results: First Stage

- In English: effect of instrument on treatment
- Where is the instrument here? Where is the treatment?
- Where would you find the "average treatment, for those that received the instrument"?
- Now write it the term above as expectation.
- Repeat for "average treatment, for those that did not received the

	Delivered		
Assigned	Arrest (D=0)	Coddle (D=1)	total
Arrest (Z=0)	91	1	92
percent of Z=0	0.99	<b>0.01</b>	1
Coddle (Z=1)	45	177	222
percent of Z=1	0.203	<b>0.797</b>	1

$$\begin{array}{rcl} E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = & & \\ 0.797 \quad \quad \quad - 0.011 & & = 0.786 \end{array}$$

# Results: Reduced Form

- Average recidivism was 18% (18% in the sample addresses reported another incident of domestic violence in the following 6 months).
- Recidivism was larger for those *assigned* to coddled ( $Z = 1$ ) than those *assigned* to arrested ( $Z = 0$ ):
  - $E[Y|Z = 1] - E[Y|Z = 0] = 0.211 - 0.097 = 0.114$
  - Given the overall mean of 18%, a 11.4% reduction is substantial
- The effect of this intention to treat (assignment) is called Intention-to-treat effect (ITT) and is the difference in outcomes between group assigned and not assigned (regardless of actual delivery). In the case of IV for RCTs: ITT = Reduced Form



# Results: LATE

The LATE (effect on compliers) is :

$$\begin{aligned}\lambda = \frac{\rho}{\phi} &= \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \\ &= \frac{0.211 - 0.097}{0.797 - 0.011} = \frac{0.114}{0.786} = 0.145\end{aligned}$$

- ITT is in general, smaller than LATE because it does not take non-compliance into account
- (When we are on a situation with no “always-takers”  $TOT = LATE$ )

# How About a Regression

- Write down reg
- Match to CEs
- Present result
- Compare with reduced form
- Point out that most of the difference comes from a higher fraction of recidivism among control group (arrested)
- This is the type of selection bias that we where looking for, and didn't find,

	Delivered		
Assigned	Arrest (D=0)	Coddle (D=1)	total
Arrest (Z=0)	91	1	92
percent of Z=0	0.99	0.01	1
Coddle (Z=1)	<b>45</b>	177	222
percent of Z=1	<b>0.203</b>	0.797	1

# Example #3: Family Size and Years of Education

- In economics there is an old (and potentially problematic) debate around whether families are choosing the "correct" number of children.
- "Quantity-quality trade-off" in family size: reduction in family size might lead to higher parental investment in children.
- One dimension where this can be measured is whether having larger families affect the highest level of education obtained by children.
- This example is useful for us in that it clearly shows how IV can be used outside of RCTs, but we should also think critically about its policy relevancy.

# Causal Question

- What is the effect of having a larger family on the educational outcomes of the older child in that family?
- What's the problem with a simple difference in groups? (average education of older child in large families - average education of older child in small families)

# Causal Question

- What is the effect of having a larger family on the educational outcomes of the older child in that family?
- What's the problem with a simple difference in groups? (average education of older child in large families - average education of older child in small families)

$$\mathbb{E}(Y_i|D = 1) - \mathbb{E}(Y_i|D = 0) = \kappa + \underbrace{\mathbb{E}(Y_{i0}|D_i = 1) - \mathbb{E}(Y_{i0}|D_i = 0)}_{\text{Selection bias}}$$

(assuming constant effects)

- You can describe the same problem using OVB. Think of a variable like income.

# Using RCTs as a Thought Experiment

- How would an RCT look like in this case?
  1. Draw a sample of families with one child.
  2. In some of this families, randomly assign them a second child ( $D_i = 1$ ).
  3. Wait 20-30 years and collect data on educational attainment of the firstborn (who did and did not got a second child)
- Given randomization,  $\mathbb{E}(Y_{i0}|D_i = 1) = \mathbb{E}(Y_{i0}|D_i = 0)$  and a simple difference in groups measures the causal effect:

$$\mathbb{E}(Y_i|D = 1) - \mathbb{E}(Y_i|D = 0) = \kappa$$

# What is an Instrument in This Case? 1/3

- A good instrument must satisfy:

1. **Relevancy:** The instrument has a causal effect on the variable of interest. In this example: something that affects number of children in the family.

# What is an Instrument in This Case? 1/3

- A good instrument must satisfy:

1. **Relevancy:** The instrument has a causal effect on the variable of interest. In this example: something that affects number of children in the family.
2. **Independence:** The instrument is randomly assigned or “as good as randomly assigned”. Unrelated to omitted variables we might want to control for. In this example: the instrument must not be related to other factors that explain (a) number of children and (b) education of first born. **This is the main challenge.**



# What is an Instrument in This Case? 1/3

- A good instrument must satisfy:

1. **Relevancy:** The instrument has a causal effect on the variable of interest. In this example: something that affects number of children in the family.
2. **Independence:** The instrument is randomly assigned or “as good as randomly assigned”. Unrelated to omitted variables we might want to control for. In this example: the instrument must not be related to other factors that explain (a) number of children and (b) education of first born. **This is the main challenge.**
3. **Exclusion Restriction:** the instrumented treatment (number of children) is the only channel through which the instrument affects the outcome. In this example: something that affects education only through its effect on family size.

# What is an Instrument in This Case? 2/3

- Instrument #1: having twins in the second birth.
- $Z_{1i} = 1$  if second birth consist of twins
- $Z_{1i} = 0$  if second birth consist of one child (singleton)
- **Relevancy:** twins affects number of children in the family. Very plausible and verifiable in the data:  $E(D|Z_1 = 1) = 3.92$  while  $E(D|Z_1 = 0) = 3.6$ . Why is this? Why not 1 full child of a difference?

# What is an Instrument in This Case? 2/3

- Instrument #1: having twins in the second birth.
- $Z_{1i} = 1$  if second birth consist of twins
- $Z_{1i} = 0$  if second birth consist of one child (singleton)
- **Relevancy:** twins affects number of children in the family. Very plausible and verifiable in the data:  $E(D|Z_1 = 1) = 3.92$  while  $E(D|Z_1 = 0) = 3.6$ . Why is this? Why not 1 full child of a difference?
- Some families that where planing for 2 end up with three, but other families that were plannig for 3 (or more) are not affected by the instrument. What are the names of these two groups?

# What is an Instrument in This Case? 3/3

- **Independence of Instrument #1:** plausible as twins occur more or less at random, but maybe not because of age (also less plausible today due to IVF).
- **Exclusion restriction of Instrument #1:** twins affect education only through family size. Maybe? This would not work if, for example, there was a cultural belief that twins are, on average, better students than the rest of the population (in this case, positive stigma would create a link from twins to education in addition to the channel of family size).

# Results Using Twins IV

- **First Stage:**  $\phi = E(D|Z_1 = 1) - E(D|Z_1 = 0) = 3.92 - 3.6 = 0.32$ .
- **Reduce Form:**  $\rho = E(Y|Z_1 = 1) - E(Y|Z_1 = 0) = \text{zero}$  (no  $\hat{\rho}$  provided).
- **LATE:**  $\lambda = \rho/\phi = 0/0.32 = 0$ .
- Notice that treatment need not be binary (nor the instrument)
- Given that having a zero reduce form implies zero LATE, it is come to present reduce form results first (if its 0 in the RF, it will also be in LATE).
- It seems there is no quantity quality trade-off. **When increasing family size from 2 to 3 children, among the compliers.**
- An OLS estimate with Education ( $Y_i$ ) on treatment ( $D_i$ ) and controls, yields  $\hat{\beta} = -0.25$ . This is pure selection bias!

# Second Possible Instrument

- Instrument #2: Cultural preference for mixed gender in children (girl-boy preferred to boy-boy or girl-girl)
- $Z_{2i} = 1$  if second birth is same gender as first, and  $Z_{1i} = 0$  if second birth is different gender than first.
- **Relevancy:**  $E(D|Z_1 = 1) = 3.68$  while  $E(D|Z_1 = 0) = 3.60$ . Maybe?
- **Independence:** similar rationale as twins (assuming no gender selective abortions).
- **Exclusion:** Maybe? One possible additional channel is that same sex siblings share more resources (room, clothes, etc.) than mixed gender siblings. If this savings for

# Instrument #2: How to Check for Exclusion Restriction 1/2

- To check for relevancy: look in the data if there instrument explains the treatment variable.
- To check for independence: similar to RCTs check for balance in covariates (covariates = regressors or characteristics).
- The check for exclusion restriction: cannot be done directly. Look for an effect where there shouldn't be one:
  - Focus on groups where there is no link between instrument and treatment (e.g. always-takers)
  - If the instrument is still having an effect on the outcome, this would suggest that there is an additional channel connecting instrument and outcome and this assumption is violated.

# Instrument #2: How to Check for Exclusion Restriction 2/2

- Example: religious families are more likely to have three or more children (always takers). Or highly educated families are less likely to have more than one child (never takers).
- Effects of  $Z$  on  $Y$  (reduced form) in samples with few compliers are suggestive evidence that that exclusion does not hold.
- Looking at the formula for LATE:  $\lambda = \rho/\phi$ , rearranging  $\lambda\phi = \rho$ . Hence, when there is not first stage ( $\phi = 0$ ), there should not be a reduce form effect either ( $\rho = 0$ ). Observing no first stage with a strong reduce form relationship is suggestive evidence that exclusion is not holding (other factors behind the reduce form link).
- The study that used Instrument #2 did this check and found no reduce form effects. Hinting at exclusion holding.



# Combining IV and Regression: 2SLS

- Two reasons to combine IV with regression:
  1. Sometimes we might have more than one instrument and combining them in one regression improves statistical precision (because of a smaller variance in the residual).
  2. Our instruments might not be "as-good-as-random" but might achieve independence after controlling for a few observable characteristics (e.g. age of the mother in case of the twins instrument).
- The procedure that combines regression and IV is called **Two Stage Least Squares (2SLS)**

# First Stage and Reduce Form in Regression

- For the case of a binary instrument, we can write the first stage and reduce form as the following regression (end of lecture on CEF):

$$\text{THE FIRST STAGE: } D_i = \alpha_1 + \phi Z_i + e_{1i}$$

$$\text{THE REDUCED FORM: } Y_i = \alpha_0 + \rho Z_i + e_{0i}$$

- Where we can evaluate each conditional expectation from the previous formulation (of FS and RF) and obtain:

$$\text{THE FIRST STAGE: } E[D_i | Z_i = 1] - E[D_i | Z_i = 0] = \phi$$

$$\text{THE REDUCED FORM: } E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0] = \rho$$

- Where  $LATE = \lambda$  is the ratio the slopes of both regressions.
- 2SLS offers an alternative way of computing this ratio (and getting the SEs right!)

# 2SLS Procedure

- First step: estimate the regression equation for the first stage and generate fitted values  $\widehat{D}_i$ :

$$\widehat{D}_i = \alpha_1 + \phi Z_i$$

- Second step: regress  $Y_i$  on  $\widehat{D}_i$ :

$$\widehat{Y}_i = \alpha_2 + \lambda_{2SLS} \widehat{D}_i + e_{2i}$$

- The regression estimate for  $\lambda_{2SLS}$  is **identical** to the ratio  $\rho/\phi$ ! (proved in the appendix of Ch3)

# 2SLS With Multiple Regressors

- Now that we have the regression setup ready, it is straight forward to add control.
- The most important thing to remember is that you need to include the additional controls in all the equations (otherwise we would be inducing a type of OVB).
- Using the example of the additional control of maternal age,  $A_i$ :

$$\text{THE FIRST STAGE: } D_i = \alpha_1 + \phi Z_i + \gamma_1 A_i + e_{1i}$$

$$\text{THE REDUCED FORM: } Y_i = \alpha_0 + \rho Z_i + \gamma_0 A_i + e_{0i}$$

And in the 2SLS estimate:

$$\text{FIRST STAGE FITS: } \widehat{D}_i = \alpha_1 + \phi Z_i + \gamma_1 A_i$$

$$\text{SECOND STAGE: } Y_i = \alpha_2 + \lambda_{2SLS} \widehat{D}_i + \gamma_2 A_i + e_{2i}$$

- 2SLS gets the SEs right for  $\lambda_{2SLS}$  (more on appendix of Ch3).

# 2SLS With Multiple Instruments

- In addition the twins instrument ( $Z_i$ ), we can add now the siblings gender instrument. Let's label this last one  $W_i$  to avoid confusions. We can also bring the additional controls (Age,  $A_i$ , First born boy  $B_i$ ) and get new first stage:

$$\begin{aligned}\text{FIRST STAGE:} \quad D_i &= \alpha_1 + \phi_t Z_i + \phi_s W_i + \gamma_1 A_i + \delta_1 B_i + e_{1i} \\ \text{REDUCED FORM:} \quad Y_i &= \alpha_0 + \rho_t Z_i + \rho_s W_i + \gamma_0 A_i + \delta_0 B_i + e_{0i}\end{aligned}$$

- And the corresponding 2SLS estimation:

$$\begin{aligned}\text{FIRST STAGE FITS:} \quad \widehat{D}_i &= \alpha_1 + \phi_t Z_i + \phi_s W_i + \gamma_1 A_i + \delta_1 B_i \\ \text{SECOND STAGE:} \quad Y_i &= \alpha_2 + \lambda_{2SLS} \widehat{D}_i + \gamma_2 A_i + \delta_2 B_i + e_{2i}\end{aligned}$$

- Ready to read results from most IV papers!

# IV Results for Family Size and Education: First Stage

TABLE 3.4  
Quantity-quality first stages

	Twins instruments		Same-sex instruments		Twins and same- sex instruments
	(1)	(2)	(3)	(4)	(5)
Second-born twins	.320 (.052)	.437 (.050)			.449 (.050)
Same-sex sibships			.079 (.012)	.073 (.010)	.076 (.010)
Male		-.018 (.010)		-.020 (.010)	-.020 (.010)
Controls	No	Yes	No	Yes	Yes

*Notes:* This table reports coefficients from a regression of the number of children on instruments and covariates. The sample size is 89,445. Standard errors are reported in parentheses.

# IV Results for Family Size and Education: Second Stage + OLS

TABLE 3.5  
OLS and 2SLS estimates of the quantity-quality trade-off

Dependent variable	OLS estimates (1)	2SLS estimates		
		Twins instruments (2)	Same-sex instruments (3)	Twins and same- sex instruments (4)
Years of schooling	−.145 (.005)	.174 (.166)	.318 (.210)	.237 (.128)
High school graduate	−.029 (.001)	.030 (.028)	.001 (.033)	.017 (.021)
Some college (for age $\geq 24$ )	−.023 (.001)	.017 (.052)	.078 (.054)	.048 (.037)
College graduate (for age $\geq 24$ )	−.015 (.001)	−.021 (.045)	.125 (.053)	.052 (.032)

*Notes:* This table reports OLS and 2SLS estimates of the effect of family size on schooling. OLS estimates appear in column (1). Columns (2), (3), and (4) show 2SLS estimates constructed using the instruments indicated in column headings. Sample sizes are 89,445 for rows (1) and (2); 50,561 for row (3); and 50,535 for row (4). Standard errors are reported in parentheses.

# IV Results for Family Size and Education: Second Stage + OLS

TABLE 3.3  
OLS and 2SLS estimates of the quantity-quality trade-off

Dependent variable	OLS estimates (1)	2SLS estimates		
		Twins instruments (2)	Same-sex instruments (3)	Twins and same- sex instruments (4)
Years of schooling	-.145 (.005)	.174 (.166)	.318 (.210)	.237 (.128)
High school graduate	-.029 (.001)	.030 (.028)	.001 (.033)	.017 (.021)
Some college (for age $\geq 24$ )	-.023 (.001)	.017 (.052)	.078 (.054)	.048 (.037)
College graduate (for age $\geq 24$ )	-.015 (.001)	-.021 (.045)	.125 (.053)	.052 (.032)

Notes: This table reports OLS and 2SLS estimates of the effect of family size on



## IV - Final Considerations 1/2

- Quick intuitions why SE of  $\lambda_{2SLS}$  are wrong if estimated with OLS:  $\widehat{D}_i$  is an estimated variable that has more uncertainty than  $D_i$ , we know that, but the software doesn't. Hence it generates fictitiously small SEs (SE from 2SLS > SE from OLS).
- When assessing the relevance of one instrument use t-test as usual. When assessing the relevance of multiple ( $K$ ) instruments use a joint hypothesis test  $\phi_1 = \phi_2 = \phi_K = 0$ . The rule of thumb here is that the F-statistic reported for these type of tests has to be greater than 10 (p-hacking alert!).
- Beware of studies that are *instrument driven* ("I just found a new cool and clever instrument! Now, which policy could I use this instrument for?") as oppose to *policy driven* ("Policy X is of high relevance, let's look for IVs to identify its causal effect").

# IV - Final Considerations 2/2

- When it comes to external validity never forget that LATE is the effect on compliers (MM constantly does!).
- There is a twitter account that emphasizes this extrapolation problem in bio-medical sciences by adding the proper caveat at the end of each new flashy result:



**justsaysinmice** @justsaysinmice · Mar 17  
IN MICE



neurosciencenews.com

Live Fast, Die Young? Or Live Cold, Die Old? - Neuroscience News

Body temperature exerts a greater effect on longevity and lifespan than metabolic rate, researchers report.



# Acknowledgments

- MM