

Exercises and Questions around IV

Econ 140, Section 10

Jonathan Old

Roadmap

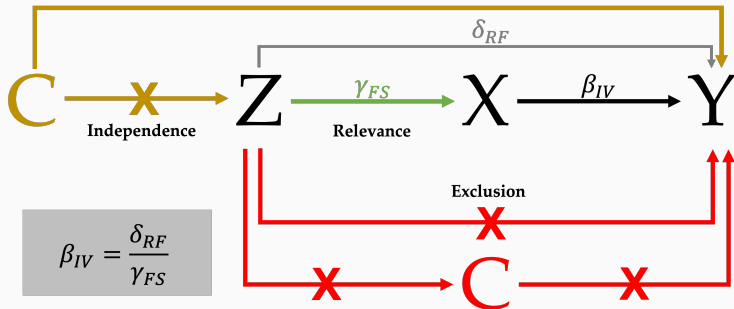
1. Recap: IV
2. Group work
3. IV Exam Exercise
4. Questions

Recap: IV

Recap: IV summary

We need the following three assumptions for IV to work:

- 1 **Relevance:** Z must truly affect X
- 2 **Independence:** Z is as good as randomly assigned
- 3 **Exclusion Restriction:** The **only** way that Z affects Y is via X .



How IV estimates can be different from OLS estimates

- **OVV:** If OLS had omitted variable bias, and our instrument is valid, then the IV estimate should be different. We can check whether this is plausible with the OVB formula
- **Measurement error:** If we have random measurement error in the independent variable (X), then we can use IV to overcome this. In that case, the IV coefficient will be larger (in absolute value) than the OLS coefficient
- **LATE:** OLS gives us ATT + Selection Bias, while IV gives us the treatment effect on the compliers (LATE). The ATT may be different from LATE, even without selection bias.
- **Invalid IV:** Hard to determine what exactly is going on
- **Sampling variation:** This can just happen by chance

- There is a catch in IV: independence and exclusion restriction only need to hold **conditionally**
- This means: If we **know** C, we can control for it in the IV regression, and the concerns go away
- But then, we have the old OLS problems again
- So we should only do this in cases where we really know all the C's.

Group work

Group work

Group 1: We are interested in the effect of being in the army on crime. We instrument being in the army with a lottery (paper)

Group 2: We are interested in the effect of income on conflict. We instrument income with rainfall (paper)

Group 3: We are interested in the effect of air pollution on mortality. We instrument local air pollution with wind direction (paper)

- 1 Relevance: Z must truly affect X
- 2 Independence: Z is as good as randomly assigned
- 3 Exclusion restriction: The **only** way that Z affects Y is via X

Your job: Discuss whether these assumptions hold!

IV Exam Exercise

IV Exam exercise i

The land of Econometrea consists of 34 islands. Each island has their own university and students attend uni on their home island. Students in Econometrea either have to attend lectures in person or watch them online. Priyanka has obtained data for the students from all the universities and would like to study the effect of watching lectures online on the students' exam scores. For each student she has the exam score (0 - 100, 70+ is a first, 40 and below a fail), which she uses as her left hand side variable, the fraction of lectures watched online, and how on how many days the student visited the library each week. Priyanka obtains the following regression results:

IV Exam exercise ii

Regressor	Estimation Method				
	OLS (1)	IV (2)	IV (3)	IV (4)	IV (5)
Fraction online	-4.91 (0.11)	-2.08 (0.22)	-0.56 (0.40)	-0.75 (0.90)	6.09 (3.22)
Library visits			0.91 (0.08)		2.09 (0.43)
Sample	All	All	All	Unis with lottery	Unis with lottery

All regressions also contain a constant term and a dummy variable for each but one of the universities. Robust standard errors are reported in parentheses.

- (a) What is the interpretation of the coefficient in column (1). If this were a causal effect, would it be big or small? Explain whether this estimate is likely to have a causal interpretation.
- (b) Wei Min observes that some halls of residence are close to the lecture rooms while others are further away and suggests to use the distance of halls from lecture halls as an instrumental variable for the fraction of lectures watched online. Results are in column (2). Explain which assumptions need to be satisfied for the IV to deliver better causal estimates than OLS here. Discuss the validity of the

IV Exam exercise iv

assumptions in this case.

Regressor	Estimation Method				
	OLS (1)	IV (2)	IV (3)	IV (4)	IV (5)
Fraction online	-4.91 (0.11)	-2.08 (0.22)	-0.56 (0.40)	-0.75 (0.90)	6.09 (3.22)
Library visits			0.91 (0.08)		2.09 (0.43)
Sample	All	All	All	Unis with lottery	Unis with lottery

All regressions also contain a constant term and a dummy variable for each but one of the universities. Robust standard errors are reported in parentheses.

Solution

- Students who watch all lectures online, on average, have around 4.9 fewer points on the exam than students who went to all lectures in person. This is a moderately strong effect, but likely not causal (think of omitted variables, such as student motivation)

Solution

- **Relevance:** Proximity to lecture halls must correlate strongly with lecture attendance. This is plausible from our experience, and we could test this using an F-test.
- **Independence:** Proximity must be as good as randomly allocated. This might not hold here, for example, if more motivated students live closer to campus.
- **Exclusion restriction:** Proximity to lecture halls only influences exam scores because it has an effect on lecture attendance. This can be violated if proximity also affects the probability to go and study in the library, or if residence halls closer to campus are also closer to bars

IV Exam exercise vii

(c) Alma points out to Wei Min that students can pick which hall they want to live in and that some halls are located in Study Village, close to lecture halls and the library while others are located further away in Party Town, surrounded by pubs. How does this information affect your assessment of the IV strategy?

(d) Manuel realises that the data also include a variable for the number of times a student has checked into the library per week. He suggests to rerun the instrumental variables regression adding this variable as a control. Results for this regression are displayed in column (3). Assess Manuel's strategy.

Solution

- This would be a violation of independence: The students who live closer to the campus are just inherently different than students who live far away
- Library visits are a bad control: They are another outcome of the instrument, and so controlling for them introduces additional bias.

IV Exam exercise ix

Regressor	Estimation Method				
	OLS (1)	IV (2)	IV (3)	IV (4)	IV (5)
Fraction online	-4.91 (0.11)	-2.08 (0.22)	-0.56 (0.40)	-0.75 (0.90)	6.09 (3.22)
Library visits			0.91 (0.08)		2.09 (0.43)
Sample	All	All	All	Unis with lottery	Unis with lottery

All regressions also contain a constant term and a dummy variable for each but one of the universities. Robust standard errors are reported in parentheses.

- (e) Jenny notices that there are five universities which assign students to their halls of residence by a lottery. She suggests to run the IV model from columns (2) and (3) for the subsample of students from these universities only. Results are displayed in columns (4) and (5). Assess Jenny's regressions.
- (f) Drawing on the results in the table above, what have you learned from this exercise about the causal effect of watching lectures online on students' exam results?

Solution

- For these universities, the distance is as good as randomly assigned, and so we have no violations of the independence restriction. However, the exclusion restriction may still not hold.
- In a simply OLS regression, it may look as if watching lectures online causes worse grades. However, we can correct for this using a valid IV, and find that the effect is only very small and statistically insignificant.

Questions
