Instrumental Variables

Part I

Fernando Hoces la Guardia

07/27/2022

Brief Reactions to Midterm 2 1/2

- Let me repeat my expectations. If you:
  - (1) actively engage with class material, **and**
  - (2) read the book, **and**
  - (3) did section + practice questions, **and**
  - (4) check with classmates and/or us in OH, **then**
  - you should have done well in the midterm.
- If you did this, and did not do well in the midterm, then I made some mistakes when designing the midterm.

Brief Reactions to Midterm 2 2/2

- Concerns over Midterm 2: time and clarity.
  - If the performance of Midterm 2 is substantially worse than Midterm 1, we will grade on a curve (upward).
  - I consider Midterm 1 a good test, and will make corrections on time and clarity so exam follows that template.
  - What will be different for the exam:
    - We will look for classroom where you can take the exam over longer period of time without adding much more questions. (possibly 8:10am - 11am, Thursday August 11th).
    - GSIs and I will work to further review and test the exam before administering it.

Suggestion on How to Approach the Exam

- If you are doing (1) - (4) and feeling comfortable with the material don't change anything.

- If you are struggling: think of the next two weeks as a very high intensity book club. Focus on reading Ch 3 - 6 in parallel to the lectures.

    - Write summaries.
    - Bring questions to lecture and OH.
    - Discuss with classmates.
    - Don't focus on the appendices.
    - Focus on understanding the key concepts. Apply them to different examples.
    - 60-70% of the exam will be about these chapters, and 30-40% will be small variations from previous practice questions.

Motivation for the Last Third of the Course

- So far we have explore:
  - 1 big problem when answering causal questions (selection bias)
  - 1 framework to understand causality (potential outcomes),
  - 2 research design tools to disentangle causal effects from selection bias (RCTs and Regression).
- Up to here the scenario is a bit bleak as the two options have significant drawbacks:
  - In many cases RCTs can be prohibitively expensive, logistically impossible, and/or ethically wrong.
  - Regression seems to require a very strong assumption: that we are controlling for all relevant variables, or put another way, that within each cell defined by the specific values of all controlling variables, we are conducting one RCT per-cell.

Three Research Designs That Give Us Hope!

- Today we start the last, and most hopeful, third of the course, where we explore three of the most commonly used research design tools to measure causality when the previous two fail. They are:
  - Instrumental Variables
  - Regression Discontinuity Design
  - Differences in Differences.
- All three tools rely on finding a plausible story to justify that, at some point, the assignment of treatment was very similar to a randomized experiment. Hence they are usually referred to as quasi-experimental methods, or natural experiments.

Instrumental Variables

- We will learn about IV by looking at three examples:
  - Charter Schools and Test Scores.
  - Policing and Domestic Violence.
  - Family Size and Earnings.
- We start with IV as a solution to imperfect RCTs, solidify key concepts and then move into IV scenarios where there is not an explicit random assignment strategy, and use IV to find as-good-as-random variation in the real world.
- Data sets that come from studies that do not have randomization as an explicit strategy are called observational data.

IV As a Solution to RCTs With Imperfect Compliance 1/2

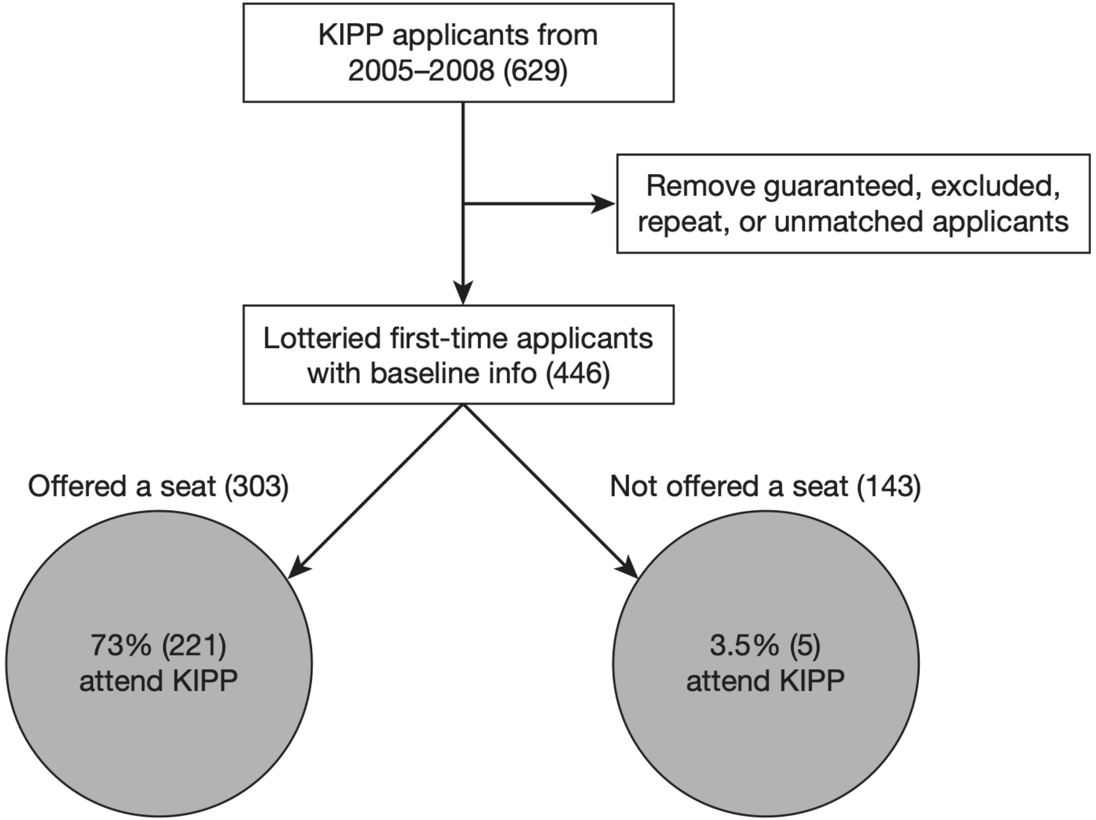
- General policy debate: increase or decrease the role of charter schools in the US.
- Supporters: its additional flexibility allows them to provide high quality education to undeserved communities.
  - Evidence cited: simple difference in groups between charter and traditional public schools shows positive "effects" of charter on test scores.
- Opponents: charters are basically selecting the more prepared students within this communities and distracting from the overall goal of high quality public education for all.
  - Interpretation of the evidence: selection bias!

IV As a Solution to RCTs With Imperfect Compliance 2/2

- Knowledge is Power Program (KIPP): Large charter school provider (140 schools).
  - No excuse philosophy, expand instruction time, non-unionized teachers
  - 95% black or hispanic.
  - 80% qualifies for federal government's subsidized lunch program.
- KIPP Lynn, MA
  - Initially undersubscribed, later oversubscribed: lottery to assign slots.
  - **Winners were *offered* a slot, not all winners took it. Additionally some lottery losers managed to get in anyway.**

Application and Enrollment to KIPP, Lynn

FIGURE 3.1
Application and enrollment data from KIPP Lynn lotteries



Note: Numbers of Knowledge Is Power Program (KIPP) applicants are shown in parentheses.

Table 3.1

		KIPP applicants			
	Lynn public fifth graders (1)	KIPP Lynn lottery winners (2)	Winners vs. losers (3)	Attended KIPP (4)	Attended KIPP vs. others (5)
		Panel A. Baseline characteristics			
Hispanic	.418	.510	−.058 (.058)	.539	.012 (.054)
Black	.173	.257	.026 (.047)	.240	−.001 (.043)
Female	.480	.494	−.008 (.059)	.495	−.009 (.055)
Free/Reduced price lunch	.770	.814	−.032 (.046)	.828	.011 (.042)
Baseline (4th grade) math score	−.307	−.290	.102 (.120)	−.289	.069 (.109)
Baseline (4th grade) verbal score	−.356	−.386	.063 (.125)	−.368	.088 (.114)
		Panel B. Outcomes			
Attended KIPP	.000	.787	.741 (.037)	1.000	1.000 —
Math score	−.363	−.003	.355 (.115)	.095	.467 (.103)
Verbal score	−.417	−.262	.113 (.122)	−.211	.211 (.109)
Sample size	3,964	253	371	204	371

Notes: This table describes baseline characteristics of Lynn fifth graders and reports estimated offer effects for Knowledge Is Power Program (KIPP) Lynn applicants. Means appear in columns (1), (2), and (4). Column (3) shows differences between lottery winners and losers. These are coefficients from regressions that control for risk sets, namely, dummies for year and grade of application and the presence of a sibling applicant. Column (5) shows

A Note on Standarizing Outcomes

- In many studies, it is increasingly common to standardize (non-binary) outcomes to facilitate interpretation and comparison across studies.

- The outcome Y is subtracted to have mean zero, and standard deviation of one. The mean and standard deviation can be those of the sample or of a target population.

- This is sounds similar to the process of standardizing a variable to compute the t-statistic, but is different in that the standard deviation corresponds to that of the underlying variable, **not** of the sample mean.

- As a result any change in the outcome now has the interpretation of "fraction of a standard deviation".

Table 3.1: Comparison to Lynn and Balance

	Lynn public fifth graders (1)	KIPP Lynn lottery winners (2)	Winners vs. losers (3)
	Panel A. Baseline characteristics		
Hispanic	.418	.510	−.058
			(.058)
Black	.173	.257	.026
			(.047)
Female	.480	.494	−.008
			(.059)
Free/Reduced price lunch	.770	.814	−.032
			(.046)
Baseline (4th grade) math score	−.307	−.290	.102
			(.120)
Baseline (4th grade) verbal score	−.356	−.386	.063
			(.125)

Table 3.1: Comparison to Lynn and Balance. Notes

- KIPP are different from rest of town:
    - more hispanic, more black, more female,
    - more low income,
    - slightly better scores in math,
    - substantially worse scores in verbal skills
- KIPP lottery winners and losers are indistinguishable from each other

	Lynn public fifth graders (1)	KIPP Lynn lottery winners (2)	Winners vs. losers (3)
	Panel A. Baseline characteristics		
Hispanic	.418	.510	−.058 (.058)
Black	.173	.257	.026 (.047)
Female	.480	.494	−.008 (.059)
Free/Reduced price lunch	.770	.814	−.032 (.046)
Baseline (4th grade) math score	−.307	−.290	.102 (.120)
Baseline (4th grade) verbal score	−.356	−.386	.063 (.125)

Table 3.1: Effects of Treatment "Offer"

		Panel B. Outcomes	
Attended KIPP	.000	.787	.741
			(.037)
Math score	−.363	−.003	.355
			(.115)
Verbal score	−.417	−.262	.113
			(.122)

Table 3.1: Effects of Treatment "Offer". Notes

- This is an average causal effect of receiving an offer to go to KIPP
- Balance in all pre-treatment characteristics
- Attended KIPP: 74 percentage points (pp) more to attend KIPP
- Math score: 0.36 standard deviations (σ)
- Verbal score: 0.11σ, but not statistically significant.
- Calculation of effects is done using regression that controls for year of application, sibling applicants and grade of application.
- But sometimes the key policy question is not the effect of an invitation, but the effect of receiving the actual treatment (attending KIPP in this case)

	Panel B. Outcomes		
Attended KIPP	.000	.787	.741
			(.037)
Math score	−.363	−.003	.355
			(.115)
Verbal score	−.417	−.262	.113
			(.122)

Enter Instrumental Variables (IV)

- What IV does, in this context, is that it takes these "effects of offers" and it transforms them into "effect of attendance".
- The instrument in this case is the lottery result of being offer a slot at KIPP (or not).
- The effect of offers is $\mathbf{0.36\sigma}$ over the **entire population** of applicants
- The fraction of applicants who ended up attending because of the offer is 74% of the population
- This $\mathbf{0.36\sigma}$ combines a positive effect on the 74% above, and an likely zero effect on those unaffected by the offer.
- How do you think we can obtain the estimator for the effect of attending?

Instrumental Variables Estimation

$$\text{Effect of offers on scores} = (\{\text{Effect of offers on attendance}\} \times \{\text{Effect of attendance on scores}\})$$

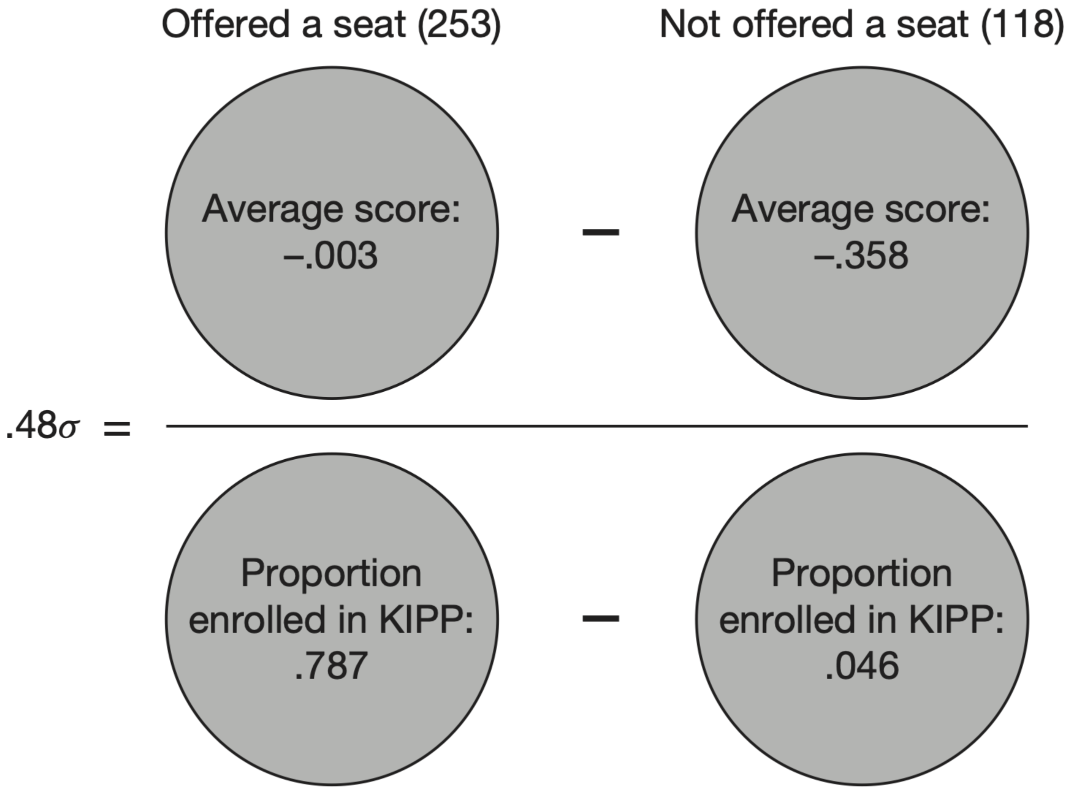
Rearranging:

$$\text{Effect of attendance on scores} = \frac{\{\text{Effect of offers on scores}\}}{\{\text{Effect of offers on attendance}\}}$$

Instrumental Variables Estimation

Effect of attendance on scores =

FIGURE 3.2
IV in school: the effect of KIPP attendance on math scores



Note: The effect of Knowledge Is Power Program (KIPP) enrollment described by this figure is $.48\sigma = .355\sigma/.741$.

IV: Three Assumptions

1. **Relevancy:** The instrument has a causal effect on the variable of interest. In our example: offers have a causal effect on attendance.

2. **Independence:** The instrument is randomly assigned or "as good as randomly assigned". Unrelated to omitted variables we might want to control for. In our example: the instrument is not related with other ommited variables because of randomization.

3. **Exclusion Restriction:** the instrumented treatment (attendance) is the only channel through which the instrument affects the outcome. In our example: this is equivalent to claiming that the entire effect of offers on scores $(\mathbf{0.36\sigma})$ is completely attributable to the difference in attendance (74%)

Estimating Effect of KIPP with OLS

- Forget about offers and focus only on attendance.
- If we run a regression on outcome, binary for attendance and same controls as before, we get the results from (4) and (5).
- This comparison effectively mixes the effects due to randomizing offers and the effects from parents that chose to ignore the lottery. Potential threat of selection bias.
- Given that (a) there is balance and (b) results are close to IV, we can argue that selection bias is not a threat **among applicants**.

plicants

Attended KIPP (4)	Attended KIPP vs. others (5)
.539	.012 (.054)
.240	−.001 (.043)
.495	−.009 (.055)
.828	.011 (.042)
−.289	.069 (.109)
−.368	.088 (.114)
1.000	1.000 —
.095	.467 (.103)
−.211	.211

IV: Definitions

- We now take all the concepts explore in the KIPP example connect them to formal definitions in the IV framework.

- **Instrumental variable, Z_i:** variable that induces random ("exogenous") variation in the treatment of interest. In our example: result of the lottery.

- **Treatment variable, D_i:** variable that indicates treatment of interest (historically refer to as "endogenous variable of interest" in non-RCT settings). In the example: assisting to KIPP, Lynn.

- **Outcome Variable, Y_i:** variable where we are interested in evaluating the causal effects of the treatment. In the example: 5th grade scores (reading and math).

IV: Key Steps (Chain Reaction)

- First check if there is a link from the instrument (Z_i) to the treatment (D_i). This step is called the **first stage**.

- Second check if there is a link from the instrument (Z_i) to the main outcome (Y_i). This is called the **reduced form**.

- Compute the causal effect of interest, the effect of Z on Y through D, as the ratio of the reduce form to the first stage. This effect estimates the causal effect for a specific group (comming up in a few slides), hence its named **Local Average Treatment Effect (LATE)**

- All this steps can be represented using conditional expectations.

IV: Key Steps Using Conditional Expectations

$$\textbf{THE FIRST STAGE: } E[D_i|Z_i = 1] - E[D_i|Z_i = 0]; \textbf{call it } \phi$$

- In the KIPP example this is the 74% difference in attendance.

IV: Key Steps Using Conditional Expectations

$$\textbf{THE FIRST STAGE: } E[D_i|Z_i = 1] - E[D_i|Z_i = 0]; \textbf{call it } \phi$$

- In the KIPP example this is the 74% difference in attendance.

$$\textbf{THE REDUCED FORM: } E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]; \textbf{call it } \rho$$

- In the KIPP example this is the $\mathbf{0.36\sigma}$ effect on math scores.

IV: Key Steps Using Conditional Expectations

$$\textbf{THE FIRST STAGE: } E[D_i|Z_i = 1] - E[D_i|Z_i = 0]; \textbf{call it } \phi$$

- In the KIPP example this is the 74% difference in attendance.

$$\textbf{THE REDUCED FORM: } E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]; \textbf{call it } \rho$$

- In the KIPP example this is the $\mathbf{0.36\sigma}$ effect on math scores.

$$\textbf{THE LOCAL AVERAGE TREATMENT EFFECT (LATE):}$$

$$\frac{\rho}{\phi} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}; \textbf{call it } \lambda$$

  - In the KIPP example this is the $\mathbf{0.48\sigma}$ effect on math scores (Figure 3.2).
  - Each of this expectations is estimated using sample averages or using regression and the correct standard errors to account for sampling variation (R and Stata do it!).

Key Policy Question: Who Does This LATE Applies to? 1/3

- Four types of individuals from the perspective of the instrument and treatment:
  - **Always Takers:** goes to KIPP $(D_i = 1)$ regardless of lottery outcome $(Z_i = 1, Z_i = 0)$.
  - **Never Takers:** does not go to KIPP $(D_i = 0)$ regardless of lottery outcomes $(Z_i = 1, Z_i = 0)$.
  - **Compliers:** Goes to KIPP $(D_i = 1)$ only if wins the lottery $(Z_i = 1)$. Defiers: Goes to KIPP only if they lose the lottery.
  - **Defiers:** Goes to KIPP $(D_i = 1)$ only if loses the lottery $(Z_i = 0)$ (!)

Key Policy Question: Who Does This LATE Applies to? 2/3

- *LATE property*: if an instrument has a non-zero first stage, has no defiers, and only affects outcome through treatment (exclusion restriction), then ratio of reduce to first stage captures the LATE, or the effect on compilers.

- For our purposes, this property says that when IV is well behaved, it measures the causal effect on compliers.

Table 3.2
The four types of children

		Lottery losers $Z_i = 0$	
		Doesn't attend KIPP $D_i = 0$	Attends KIPP $D_i = 1$
Lottery winners $Z_i = 1$	Doesn't attend KIPP $D_i = 0$	Never-takers (*Normando*)	Defiers
	Attends KIPP $D_i = 1$	Compliers (*Camila*)	Always-takers (*Alvaro*)

Note: KIPP = Knowledge Is Power Program.

Populations of Interest

- Population of compilers is usually a population of interest from a policy perspective
- Sometimes there is interest in other populations. For example we could care about the effect of the treatment in the entire treated population (attended KIPP or $E[Y_{1i} - Y_{0i}|D_i = 1]$). The corresponding estimate is called the Treatment on the Treated (TOT).
  - There are two ways to receive treatment: through Z (compliers) or regardless of Z (always takers). The TOT is a weighted average of LATE and effect on always takers. Depending on the story you have in mind $TOT \gtreqless LATE$ (example: of parents that are willing to do anything to get their kids into KIPP could make TOT>LATE)
- Beyond the specifics of TOT, the important takeaway is that IV measures different things and we always need to keep in mind whose causal effect are we measureing (i.e., who are the

External Validity

- The effect of a similar treatment in different settings might not be the same, this is the issue of **external validity**.
- To assess external validity we need to think about what is behind the causal effect.

- In the KIPP example: maybe a structured educational environment helps some but not all (fast learners for example might struggle with this).

- To explore external validity we can:

1. Explore how the effects vary across different characteristics (e.g., high/low math scores at baseline).
2. Look for different instruments and how they affect different types of compliers. Similar to RCTs, the best comparisons here are those that have similar LATEs across populations.

For Tomorrow:

- We will start with example #2, but will not pay too much attention to how LATE is TOT in this case.
- We will use example #2 primarily to reinforce concepts and to see how selection bias can contaminate the simple difference (of outcomes from treatment and control) such that it differs drastically from IV estimates.

Acknowledgments

- MM