Econometrics 2

Sergey Lychagin

CEU

Winter 2020

Using IV when treatment effects are heterogeneous

Instrumental variables

- Issue: outcome is correlated with treatment, D_i, even after conditioning on observables, X_i.
- IV (Z_i): exogenous and relevant.
- We use IV to isolate exogenous variation in X_i. Use this variation only to identify the causal effect of X_i on Y_i.
- This works if the effect does not depend on i. But what if it does?

Imbens-Angrist-Rubin potential outcome model

Stripped down version: no controls, binary instrument and treatment. There is a population of agents

- We randomly select agent i. This agent comes with (Y_{0i}, Y_{1i}) and (D_{0i}, D_{1i}) — main outcome and treatment decision.
- We assign $Z_i = 0, 1$ – an instrument – independently of $(Y_{0i}, Y_{1i}, D_{0i}, D_{1i})$ (think of something that entices agents to take treatment)
- The treament is selected by the instrument: $D_i = D_{Z_i i}$
- The outcome is selected by the treatment: $Y_i = Y_{D_i i}$ (**implicit assumption: exclusion restriction**)

Compliance types

Types, T_i:

D_{0i}	D_{1i}	
	0	1
0	never-taker	complier
1	defier	always-taker

Variation in the instrument affects compliers and defilers.
Never-takers and always-takers do not respond to Z_i.

- We don't observe one's type
- We observe which row one belongs to if $Z_i = 0$
- We observe i's column if $Z_i = 1$

No-defiance assumption (or, monotonicity): $D_{1i} \geq D_{0i}$.

ATE for compliers?

Observables reveal the distribution of types:

	Z_i	
D_i	0	1
0	never-taker or complier	never-taker
1	always-taker	always-taker or complier

- Share of always-takers: $P_a = \Pr\{D_i = 1 | Z_i = 0\}$
- Share of never-takers: $P_n = \Pr\{D_i = 0 | Z_i = 1\}$
- The rest are compliers: $P_c = 1 - P_a - P_n$

This relies on Z_i being independent of type.

Outcomes for compliers

	Z_i	
D_i	0	1
0	Y_{0i} for P_n never-takers and P_c compliers	Y_{0i} for never-takers
1	Y_{1i} for always-takers	Y_{1i} for P_a always-takers and P_c compliers

Relate estimable statistics to potential outcomes

$$E[Y_i|Z_i = 0, D_i = 0] = \frac{P_n E[Y_{0i}|T_i = n] + P_c E[Y_{0i}|T_i = c]}{P_n + P_c}$$

$$E[Y_i|Z_i = 1, D_i = 0] = E[Y_{0i}|T_i = n]$$

This implies

$$E[Y_{0i}|T_i = c] = \frac{P_c + P_n}{P_c} E[Y_i|Z_i = 0, D_i = 0] - \frac{P_n}{P_c} E[Y_i|Z_i = 1, D_i = 0]$$

Outcomes for compliers

	Z_i	
D_i	0	1
0	Y_{0i} for P_n never-takers and P_c compliers	Y_{0i} for never-takers
1	Y_{1i} for always-takers	Y_{1i} for P_a always-takers and P_c compliers

Similarly

$$E[Y_{1i}|T_i = c] = \frac{P_c + P_a}{P_c} E[Y_i|Z_i = 1, D_i = 1] - \frac{P_a}{P_c} E[Y_i|Z_i = 0, D_i = 1]$$

Local average treatment effect

Definition: treatment effect for compliers

$$LATE = E[Y_{1i} - Y_{0i} | T_i = c]$$

This can be expressed via observables using previous 3 slides. After some re-arrangements (check by substitution, use the rule of iterated expectations),

$$LATE = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

We got the Wald estimator (IV with binary treatment & instrument)! IV estimates LATE.

But what about always- and never-takers? Can't say anything about Y_{0i} for the former, Y_{i1} for the latter. Thus, ATE, ATT, ATU are not point-identified in general.

What if defiers exist?

Expand expectations in the numerator, make everything conditional on type:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_{1i} - Y_{0i}|T_i = c] \Pr\{T_i = c\} - E[Y_{1i} - Y_{0i}|T_i = d] \Pr\{T_i = d\}$$

Denominator: $\Pr\{T_i = c\} - \Pr\{T_i = d\}$.
Even if the effect is roughly similar for compliers and defiers, anything is possible: wrong sign, wrong magnitude.

Extensions: multiple instruments?

LATE depends on the choice of instrument. For example

- Suppose we have two instruments for college education: Z_1, information treatment, and Z_2, a stipend
- Agents who respond to Z_1 may be very different from those who respond to Z_2
- If the effects are highly heterogeneous, don't be surprized if you get different IV estimates.
- Threat to external validity: LATE of education on earnings using an informational treatment IV may not be informative of returns to a stipend.

Extensions: multiple instruments?

What if estimates based on Z_1 and on Z_2 are very similar?

- If the respective compliers are *very* different, this is a sign of little heterogeneity in TE
- If TEs don't vary across the agents, $LATE \approx TE$ for always- and never-takers
- $ATE \approx ATT \approx ATU \approx LATE$.

How can one check if compliers for Z_1 and those for Z_2 are different?

Multiple instruments: who are the compliers?

Do compliers for Z_1 differ from those for Z_2? Use observables X_i to compare.

Suppose X_i is a dummy:

$$\begin{aligned}\Pr[X_i = 1|T_i = c] &= \frac{\Pr\{T_i = c|X_i = 1\}}{P_c}\Pr\{X_i = 1\} \\ &= (\Pr\{D_i = 1|Z_i = 1, X_i = 1\} \\ &\quad - \Pr\{D_i = 1|Z_i = 0, X_i = 1\})\frac{\Pr\{X_i = 1\}}{P_c}\end{aligned}$$

Everything on the RHS can be approximated using the data. Compute this for Z_1 and Z_2 as IV

In the wage-college example, do stipend and information treatment compliers have similar incomes?

IV — concluding remarks

- Things get really messy once TE heterogeneity is introduced
- Be careful extending your results to a different part of the population.
- E.g., you identify the effect of college education using an informational nudge as an instrument.
- Your results may be misleading for statements about a tuition waiver policy. The value added of college may be different for people who respond to tuition waivers!

Starting your first research project

Things people value in research projects

What is usually met with interest?

1. Practical relevance:
   When Y rises or falls, people are hurt or helped.
2. Challenging common beliefs:
   Prevailing wisdom says that X reduces Y, you find that X increases Y.
3. Controversy:
   Some argue one thing while other say another.
4. Size:
   Y is big (like the service sector) or common (like traffic jams).

Things people value in research projects

Two dimensions of quality:

1. Internal validity:
   Do we believe that we measure the true causal effect?
2. External validity:
   Can we apply the results to other markets, time periods, countries?

Different fields of economics put different emphasis on internal and external validity.

How to start an empirical project

Several strategies:

Question-driven. Pick a research theme, identify interesting questions. Work hard to get data.

- This is the 'proper' way, but can be risky (uninteresting results)
- Data may be unavailable/expensive
- Identification is very hard; no good natural experiments
- Works well for: entrepreneurial types, people with access to funding

How to start an empirical project

"Freakonomics." Search for natural experiments. Then look for a dataset and a question.

- Solid identification highly valued in the profession.
- Avoid narrow policy evaluation exercises ("Would my paper be interesting to a game theorist from Kazakhstan?")
- Bad for personal marketing ("What exactly is your field?")

How to start an empirical project

Data-driven. Find rich data that few people used before. Then, find a question.

- High sunk cost: collection/cleaning takes time. What if there is no good question?
- Before investing into data collection, ask yourself: What if I already have the ideal, cleaned dataset?
- Works well if: you already have some data.
- Uniqueness is a big plus. Quoting one editor:

> *[W]e all think that you have quite a unique data, and we are all confident that there exists a nice paper that can come out of these data.*

Workflow

Suppose you settled on the question, what's next? Write a research proposal (for yourself)

- What's the research question?
- Why is it interesting?
- How you are planning to answer it, intuitively?
- What data do you plan on using?
- What's new compared to the existing literature?

Putting ideas into writing highlights weak points. Abandon/put aside hopeless projects early on!

Workflow

At some point, the research proposal will morph into your paper's introduction.

As you go along, you may shift focus, switch to a different question. The process is iterative:

- Rewrite the introduction,
- Work on the analysis,
- Rewrite the introduction,
- Work on the analysis,
- ...
- Cut unnecessary details, clean up the draft

Some examples from the past

Rozovskaya (2012) “Can Observers Prevent Vote Fraud? Evidence from Russian Elections of 2011-2012 in Moscow”

- Policy relevance?
- Controversial topic?
- Size?

Threat to identification: Observers are mostly sent to urban polling stations, places with higher perception of fraud.
Research design — natural experiment: observers are sent to polling stations in Moscow whose numbers end with a 0.

Lower votes for the ruling party, lower recorded turnout. Suggestive of ballot stuffing? The estimates are too low to question election outcome.

Some examples from the past

Korganbekova (2014) “The effect of hosting an Olympic Game on the performance of the National sports team at the following Olympics”

- Policy relevance?
- Controversial topic?
- Size?

Threat to identification: Olympic hosts are far from average.
Selection on infrastructure, interest in sports?
Research design — diff-in-diff with a hand-picked control group:

- Countries who made bids to host in the same year,
- Countries who survived till the last round of selection.

The effect is $\approx$ 6 medals, but the estimates are very noisy.

Some examples from the past

Bardits (2016) “The effect of the 2013 school centralization on school performance in Hungary”

- Policy relevance?
- Controversial topic?
- Size?

Mechanisms?

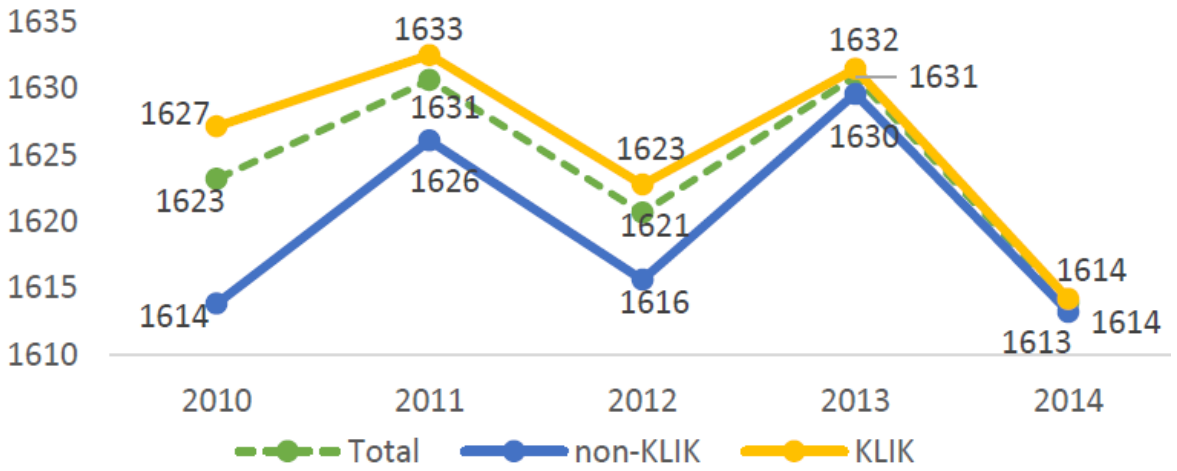
- Bureaucracy
- Change in funding

Threats to identification: Selection into treatment, by school, students, teachers.

Some examples from the past

Research design — diff-in-diff using school data

- All local public schools switch to central control in 2013
- Control group: church, private schools.



External validity? The more we can say about the mechanism, the broader the implications. Can't say much here.