Applied Microeconometrics

Lecture 3: Empirical strategies (‘refresher lecture’)

Manuel Bagues

August 2022

Lecture Slides

Overview

1. Introduction
   - Structural and treatment effect approaches
   - Guidance of theory
   - ‘Identification strategy’ vs. ‘estimation method’
2. Identification strategies
   - RCTs
   - Selection based on observables
   - Instrumental variables
   - Differences in differences

Previous lecture

- Statistical power
  - false positives and negatives
  - magnitude bias and sign bias
  - example 1: Gelman and Weakliem (2009), ‘On Beauty, Sex and Power’
  - example 2: Oster (2005), ‘Hepatitis B and the Case of the Missing Women’
  - example 3: Danzer and Lavy (2018), “Paid Parental Leave and Children’s Schooling’s Outcomes,’
- Post-study probability
  - simplified bayesian approach to update our priors
  - example 1: Gelman and Weakliem (2009), ‘On Beauty, Sex and Power’
  - example 2: serological tests

- When we have low statistical power:
  - Most likely we will be unable to detect the effect, even if it exists
  - But is can be worse: if we find significant results, their magnitude will be implausibly large (or even with the wrong sign)
- A way to quantify our ignorance is to calculate the post-study probability (PSP)
  - consider several plausible sizes of the effect based on the literature and common sense
  - calculate the PSP for different priors - let the reader choose his/hers

Structural and treatment effect approaches

- The classic approach to quantitative policy evaluation in economics is the *structural approach*.
- Its goals are to specify a class of theory-based models of individual choice, choose the one within the class that best fits the data, and use it for ex-post or ex-ante policy simulation.
- During the last two decades the treatment effect approach has established itself as a formidable competitor that has introduced a different language, different priorities, techniques and practices in applied work.
- Not only that, it has also changed the perception of evidence-based economics among economists, public opinion, and policy makers.
- The ambition in a structural exercise is to use data from a particular context to identify, with the help of theory, deep rules

- A treatment effect (TE) exercise is context-specific and addresses less ambitious policy questions.
- The goal is to evaluate the impact of an existing policy by comparing the distribution of a chosen outcome variable for individuals affected by the policy (treatment group) with the distribution of unaffected individuals (control group).
- The aim is to choose the control and treatment groups in such a way that membership of one or the other, either results from randomization or can be regarded as if they were the result of randomization.
- In this way one hopes to achieve the standards of empirical credibility on causal evidence that are typical of experimental biomedical studies.

- The TE literature has expressed dissatisfaction with the existing structural approach along several dimensions:
  - Between theory, data, and estimable structural models there is a host of untestable functional form assumptions that undermine the force of structural evidence by:
    1. Having unknown implications for results.
    2. Giving researchers too much discretion.
    3. Complexity affects transparency and replicability.
  - By being too ambitious on the policy questions we get very little credible evidence from data. Too much emphasis on "external validity" at the expense of the more basic "internal validity".

- The TE literature sees the role of empirical findings as one of providing bits and pieces of hard evidence that can help the assessment of future policies in an informal way.
- Main gains in empirical research are not expected to come from the use of formal theory or sophisticated econometrics, but from understanding the sources of variation in data with the objective of identifying policy parameters.

Research process

- When we do research we typically proceed in the following way:
  1. Causal question of interest (e.g. the impact of treatment X on Y)
  2. Theory (how should we think about it)
  3. Empirical strategy
  4. Estimation method

Guidance of theory

- We need theory to guide our analysis: there is always, explicitly or not, a theory that we are testing.
- a good knowledge of the context and the correct empirical tools might provide you adequate empirical strategy to estimate the effect of the treatment
- but you need some theoretical framework in order to understand results
    - know what to make out of the evidence
    - external validity (you need the mechanism!)
- design appropriately the empirical analysis: testable implications of each theory

Some examples from my own experience

- Example 1: Azmat, Bagues, Cabrales and Iriberri (2019), "What You Don't Know... Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education"
- Example 2: Bagues, Sylos-Labini and Zinovyeva (2017), 'Does the Gender Composition of Scientific Committees Matter?'
- Example 3: Bagues and Esteve-Volart (2016): Politicians' Luck of the Draw: Evidence from the Spanish Christmas Lottery
- Example 4: Zinovyeva and Bagues (2015): The Role of Connections in Academic Promotions

Identification strategies

The identification strategy refers to the manner in which a researcher uses observational data (i.e. data not generated by a randomized trial) to approximate a real experiment and identify causal effects. In other words, it is the way in which we try to find a control group that is comparable to the treatment group.

- Main empirical strategies:
  1. *Random assignment*
  2. *Selection based on observables*
  3. *Instrumental variables*
  4. *Difference-in-differences*
  5. *Regression discontinuity design*

- Do not confuse your 'identification strategy' with your 'estimation method'
- Identification (or empirical) strategy:
  - How you manage to find a control group that provides information about what would have happened to the treatment group in the absence of the treatment
- Examples:
  1. Random assignment (e.g. RCTs)
  2. Selection based on observables
  3. Instrumental variables (IV)
  4. Difference-in-differences (DID)
  5. Regression discontinuity design (RDD)

- Estimation method: the statistical method that we use to estimate a given β
- Examples:
    - *t-test* to compare two sample means
    - OLS to estimate some conditional correlation
    - Probit, logit... (limited dependent variables)
    - Propensity score matching
    - 2SLS
    - Fixed effects estimation
    - Local linear estimation
    - ....

- In general, each identification strategy tends to be associated to a particular estimation method, but it is not always necessarily the case.
- For instance, OLS is often used both when the identification relies on random assignment (either to account for controls or because the treatment is a continuous variable) and when it is based on observables
- In your research projects, you should always have a clear understanding of what is the empirical strategy and what is the method
  - Example 1: Identification based on observables + OLS
  - Example 2: Identification based on observables + Probit (e.g. if outcome variable is a dummy)
  - Example 3: Identification based on observables + Propensity score matching

Potential drawbacks of RCTs

Feasibility

- Problems of implementation
  - Not possible to randomize many relevant treatments (e.g. monetary policy)
  - Cost (e.g. impact of mediterranean diet)
  - Political issues (policy makers need to acknowledge ignorance),...
  - Compliance and attrition
- Ethical issues
  - Examples: STAR, babies...
  - The ethical argument is not obvious when (i) the treatment cannot be applied to everybody (maybe due to some budget constraints) and (ii) the optimal assignment rule is unknown.
  - Note: Universities typically have a Research Ethics Committee (also known as Institutional Review Board or IRB) that ensures that all research involving participants and/or their data addresses relevant ethical considerations.

Potential drawbacks of RCTs

Internal validity: Hawthorne effect

- Hawthorne effect
  - The Illumination Experiment (Landsberger 1950, Levitt and List 2011)
  - Audit study in France (Behaghel et al. 2015)

Potential drawbacks of RCTs

Internal validity: SUTVA

- Potential violation of the *Stable Unit Treatment Value Assumption (SUTVA)*
  - This framework is generally not well suited to the evaluation of system-wide reforms which are intended to have substantial general equilibrium effects.
- Example: Crepon, Duflo, Gurgand, Rathelot and Zamora 2013
  Program of job placement assistance to young unemployed workers in France to improve their labor market outcomes
  - Can you think of potential ways in which the treatment may affect the control group? (hint: displacement effects)

Potential drawbacks of RCTs

- Solution $\rightarrow$ Clustered Randomized Experiment (or two-step randomized design)
  1. 1st step: Partition of the covariate space, and randomize the number of units that are assign to treatment in each cluster
  2. 2nd step: Randomized which units receive the treatment within each cluster
- Example: Crepon, Duflo, Gurgand, Rathelot and Zamora 2013
  1. each of 235 local employment areas are randomly assigned a proportion P of job seekers to be assigned to treatment: either 0%, 25%, 50%, 75%, or 100%.
  2. a fraction P of all the eligible job seekers is randomly selected to participate in the job placement program
- Interpretation of results?
  1. Within each cluster, the treatment group has better labor market outcomes
  2. However, on average job seekers are as likely to find a job in

Potential drawbacks of RCTs

External validity

- Many authors are concerned with the potential lack of external validity of RCTs (e.g. Deaton, Manski...)
    - Treatment heterogeneity
- Problem also applies to non-experimental empirical strategies
- Possible solution:
    - Estimate heterogeneity of treatment effects and combine RCTs with structural models
    - Example: Duflo, Hanna and Ryan (AER 2012), ‘Incentives Work: Getting Teachers to Come to School’

Selection based on observables

- We may not have a controlled experiment, but maybe the treated group and the non-treated group differ only by a set of **observable** characteristics.
- The crucial assumption of the identification based on observables strategy is that, conditional on observables, the assignment of the treatment is as good as random.
- This assumption, which would justify the causal interpretation of our estimates, is known as the **Conditional Independence Assumption** (CIA), also called selection-on-observables

Causality and the CIA

- In general, would you expect individuals in the control and treatment group with similar observable characteristics to be similar in all other relevant unobserved characteristics? In other words, how relevant is the selection problem in practice?
- Selection would not be an issue if agents were fully irrational or fully uninformed. Unfortunately, individuals are unlikely to randomize their choices (even conditional on observables)
- What may drive selection?
    - Information, differences in preferences...
    - These (unobserved) selection factors may affect the outcome variable

Causality and the CIA

- For relevant ‘treatments’, selection is usually a relevant problem. Note that there is crucial paradox in empirical studies that rely on an identification strategy based on observables. In order to estimate the effect of a certain treatment, we need to assume that, conditional on certain observables, this treatment was assigned in a random way. However, those who benefit more from the treatment probably may have tried to get more "treatment".
- Corolary: we should be always be very cautious about the interpretation of estimates when the identification relies on observables
- Suggestion: use Altonji, Elder, Taber (2005) and Oster (2016) to bound your estimates

Oster 2016

'A common approach to evaluating robustness to omitted variable bias is to observe coefficient movements after inclusion of controls. This is informative only if selection on observables is informative about selection on unobservables. Although this link is known in theory (i.e. Altonji, Elder and Taber (2005)), very few empirical papers approach this formally. I develop an extension of the theory which connects bias explicitly to coefficient stability. I show that it is necessary to take into account coefficient and R-squared movements. I develop a formal bounding argument. I show two validation exercises and discuss application to the economics literature.'

Identification based on observables

Ideal experiment

- To assess the reliability of an empirical strategy that relies on observables, it is useful to reflect about the ‘ideal experiment’ that the author is hoping to capture. Is it plausible?
- Let us consider a particular example: The effect of having a distinct black name (Fryer and Levitt 2004)
- How would you address this question using a identification based on observables approach? What would you control for?
- What is the ideal RCT that the authors are hoping to capture with their empirical strategy?
- Note: if you have a paper which is based on observables, make sure always to think about your implicit RCT

Main threats to the validity of OLS

- Main threats to validity
  1. Omitted variables
  2. Bad controls
  3. Measurement error in the independent variable
  4. Measurement error in the dependent variable

Instrumental variables

- Let us consider the following model:
$$Y = X\beta + U$$
- where we are concerned about the possibility that $E[X \cdot U] \neq 0$
- Would it be a good idea to use OLS to estimate β?
  - $\beta_{OLS} = [X'X]^{-1}X'Y$
- What about using an instrument Z? Which features should it have?
  - $\beta_{IV} = [Z'X]^{-1}Z'Y$

What is a valid instrumental variable?

- Two main assumptions:
  1. **Relevance**: The instrument is correlated with the causal variable of interest, S_i,
     $\text{Cov}(Z_i, S_i) \neq 0$
  2. **Independence:** The instrument is uncorrelated with any other determinants of Y_i
     $\text{Cov}(Z_i, \eta_i) = 0$
     This requirement can be decomposed in two:
     - 2.1 **Exogeneity**: The instrument is as good as random, none of the unobserved factors affects it [$\eta_i \not\rightarrow Z_i$]
     - 2.2 **Exclusion restriction**: Z_i only affects Y_i through its effect on S_i [$Z_i \not\rightarrow \eta_i$]

Note: Some authors may refer to the independence assumption as the exogeneity condition or the exclusion restriction. However, it is useful to consider exogeneity and exclusion restriction as two distinct requirements.

Why IV works?

- Intuitive idea behind IV is as follows:
  1. **Relevance**: You found a variable (the instrument) that affects who is assigned to the treatment
  2. **Exogeneity**: The instrument is as good as randomly assigned
  3. **Exclusion restriction**: And you know that your instrument only affect the outcome through its impact on the treatment (cannot be correlated with any other factor that affects the outcome, or affect directly the outcome).
- Note that an IV strategy is equivalent to an RCT where there is no full compliance

Can we test validity of IV?

- Can we test the assumptions required for the validity of a IV:
  1. Is the instrument correlated with the treatment? [Cov$(Z_i, S_i) \neq 0$]
     - YES: Significance of first stage, F-statistics
  2. Exogeneity [$\eta_i \not\rightarrow Z_i$]
     - SOMETIMES YES: Is the instrument as good as random?
  3. Exclusion restriction? [$\eta_i \not\leftarrow Z_i$]
     - Is it plausible that the instrument only affects the outcome variable through its impact on the treatment?
     - We can try to show that it is not satisfied looking at the impact of the instrument on the outcome variable among always-takers or never-takers
     - But we will never be able to proof that there is no unobserved channel through which the instrument affects the outcome

Three ways to calculate the IV estimator

1. Standard formula: $\hat{\beta}_{IV} = [Z'X]^{-1}Z'Y$
2. $\hat{\beta}_{IV} = \frac{\hat{\beta}_{\text{reduced form}}}{\hat{\beta}_{\text{1st stage}}}$
3. Two-stage least squares (2SLS)
    - 1st stage: regress the treatment on the instrument (+ covariates)
    - 2nd stage: regress the outcome variable on the predicted treatment (+ covariates)

Local average treatment effects

- The impact of the treatment might heterogenous
- Therefore, it is important to understand the type of treatment effect that we are identifying
- Let $\mathbf{D_{i1}} \in \{0, 1\}$ indicate whether an individual **affected** by the instrument would receive or not the treatment. Similarly, let $\mathbf{D_{i0}} \in \{0, 1\}$ indicate whether an individual **not affected** by the instrument would receive or not the treatment
- We can consider four instrument-dependent subgroups, defined by the manner in which members of the population react to the instrument:
  - Compliers: $D_{i1} = 1, D_{i0} = 0$
  - Always-takers: $D_{i1} = 1, D_{i0} = 1$
  - Never-takers: $D_{i1} = 0, D_{i0} = 0$
  - Defiers: $D_{i1} = 0, D_{i0} = 1$

Example: Angrist and Krueger 1991

- Verbalize who are the members of each group in the following setup:
  - Does compulsory school attendance affect schooling and earnings? (Angrist and Krueger 1991)
    - Outcome: earnings
    - Treatment: years of schooling
    - Instrument: quarter of birth
- Who are in this case the (i) compliers, (ii) always-takers, (iii) never-takers, and (iv) defiers?

- Groups
  - Compliers: kids who drop out or not depending on their month of birth (they only drop out if they were born at the beginning of the year). Eg: the instrument (*month of birth*) affects (in the expected way) whether they receive the treatment or not (*years of education*)
  - Always-takers: kids who would always finish the academic year, independently of their month of birth
  - Never-takers: kids who would never finish the academic year, independently of their month of birth
  - Defiers: kids who would drop out if born at the end of the year, but not if they were born at the beginning of the year

- Typically we assume that there are no defiers (*monotonicity assumption*)
- Under this assumption, with an instrumental variable strategy we learn about the impact of the treatment for the group of compliers.
- IV strategy identifies the Local average treatment effect (LATE), in the sense that we learn about the impact of the treatment for a very particular group of individuals.
- Discussion: are compliers a relevant group?

Example: the link between absentism and students' performance

- Andrietti, D'Addazio and Velasco (2008)
  - OLS estimates: students that attend class tend to obtain better grades
  - IV strategy:
    - distance to reach campus from the student's house
    - dummy variable that indicates if the student works
  - Both instruments correlate with absentism
  - But what about:
    - exogeneity?
    - exclusion restriction?
- Other possible instruments?
  - Arulampalam, Naylor and Smith (2012): "Am I missing something? The effects of absence from class on student performance"

- Arulampalam, Naylor and Smith (2012) study the causal effects of class absence on student performance in a paper titled ‘Am I missing something? The effects of absence from class on student performance’. They use data from 2nd year Economics undergraduate students and they focus on the absenteeism in tutorial classes. First, they show that being absent from 10% of classes is associated with around a 1.3 percentage point lower mark in the subject (st. error=0.4). Second, to deal with the potential endogeneity of absenteeism, they use an instrumental variables strategy which exploits (i) that students are randomly assigned to different tutorial groups and (ii) that attendance tends to be lower on certain days and on certain periods of the day (e.g. Monday morning). Their IV estimate suggests that being absent from 10% of classes has a causal negative impact on performance, which is reduced by 1.6 percentage points (st. error=0.7).

1. Discuss possible explanations for why the OLS point estimate is smaller than the IV point estimate (1.3 vs. 1.6).
2. Would you expect the exogeneity assumption to be satisfied? How would you test this assumption?
3. Propose some possible violation of the exclusion restriction
4. Discuss verbally who are the always-takers, the never-takers, the compliers and defiers.

Instrumental variables as an RCT without full compliance

1. So far we have introduced IV as a way to solve a problem of endogeneity in the variable of interest
2. Another insightful interpretation is as an RCT without full compliance

RCT without full compliance

- Example: providing information on coworkers salaries (Card et al. 2012, Inequality at Work: The Effect of Peer Salaries on Job Satisfaction)
- A random sample of workers of the University of California received an email with information on their coworkers salary...
- ... but not everybody checks this information, only 50%
- Moreover, some people in the control group also got access to the information (around 20%)
- There is an increase in the likelihood that workers who received the email look for a new job (e.g. X p.p. increase)
- How do we figure out what is the impact of

Example: The Effect of Peer Salaries on Job Satisfaction

- Let us consider a simple set up where there is a binary instrument and a binary treatment:
  - Card et al. 2012 "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction"
- Email to random sample of employees of University of California with link to a webpage with information on salaries.
- treatment: information; outcome: job satisfaction; instrument: email
- Impact of the instrument on the treatment:
  - People that receive the information email: 50% check the webpage
  - People that do not receive the information email: 20% check the webpage

Example: The Effect of Peer Salaries on Job Satisfaction

- Two questions
  1. Explain verbally who are the compliers, always-takers, never-takers and defiers.
  2. Let us assume that there are no defiers. What is the share of always-takers, never-takers and compliers?

Good instruments are hard to find

Good instruments come from a combination of three ingredients:

- Good institutional knowledge
- Economic theory
- Last but not least: Originality

Some usual sources of instruments:

- Nature
  - Sometimes nature randomizes, providing exogenous variation for some variable of interest (e.g. gender of children, twins...)
- Assignment rules that rely on randomization
  - returns to medical school in Netherlands, time in prison, foster care, military service...
- ‘Natural’ experiments
  - Some exogenous variables may influence assignment to the treatment (the quarter of birth, electoral timing...)

Note that, in general, choice variables of the agent tend to be bad instruments

Sources of instruments: nature

- The effect of family size on children's education and female labor force participation
  - Twins, gender of the first born, gender of the two first born (Black, Devereux and Salvanes, QJE 2005; Angrist, Lavy and Schlosser, JOLE 2009)
  - IVF treatments (Lundborg, Plug and Wurtz Rasmussen 2014)

Sources of instruments: assignment rules that rely on randomization

- Foster care
- Time in prison
- Children's custody
- Impact of geographical mobility
- Other ideas?

Sources of instruments: ‘natural’ experiments

- Immigration
  - Networks of immigrants (Card 1991)
- The effect of preschool television exposure on standardized test scores during adolescence:
  - Gentzkow and Shapiro 2008
- Influence of mass media on U.S. government response to natural disasters
  - Eisensee and Strömberg 2007

Roadmap so far:

- With every research question it is not possible to run a randomized controlled trial.
- Maybe we can look for an instrumental variable, but good instruments are difficult to find...
- We may also try to learn about the impact of a treatment using an empirical strategy based on observables:
  - We can compare individuals exposed to the treatment with other individuals that look alike in terms of observables.
  - Unfortunately, this evidence is likely to be subject to selection biases and often it is difficult to interpret.
- What else can we do $\rightarrow$ Difference-in-differences
  - We look for a control group such that its evolution provides a good counterfactual for how the treatment group would have evolved in the absence of the treatment
  - Note: the control is assumed to evolve similar, but does not need to be similar in levels to the treatment group

Differences in differences

- Canonical DID: 2 time periods and 2 groups
- Staggered DID
  - Athey and Imbens (2018), Borusyak and Jaravel (2017), Calway and Sant'Anna (2020), de Chaisemartin and DÕHaultfouille (2020), Goodman-Bacon (2019), Sun and Abraham (2020)
- Synthetic control methods
  - Xu (2017), Arkhangelsky et. al. (2019), Ben-Michael et al. (2019)
- Pre-trends
  - Kahn-Lang and Lang (2019), Roth (WP, 2020), Freyaldenhoven et al. (2019), Manski and Pepper (2018), Rambachan and Roth (2019))

Differences-in-differences (dif-in-dif)

- The above example captures the main intuition behind the differences-in-differences analysis.
- We use the evolution of the outcome variable in the control group to construct a counterfactual of what would have happened in the treatment group in the absence of the treatment.
- Parallel trends assumption: The fundamental identifying assumption is that, in the absence of the treatment, both groups would have followed parallel trends
- Note that this empirical strategy allows for the existence of time-invariant differences between the two groups, but it assumes that there are no time-variant relevant differences.

Main threats to validity of dif-in-dif estimates

1. If the groups are different in levels, maybe they evolve differently?
2. Why did the treatment group adopt the policy, and not the control group?
3. Policies are usually implemented in bundles (the timing of the treatment may not be by chance) $\rightarrow$ the outcome variable may be affected by these other policies
4. The treatment should not affect the control group (SUTVA)
5. The composition of the treatment and control groups should not change as a result of treatment

Usual checks

1. The two groups evolved similarly in the past (although note that this is neither a sufficient nor a necessary condition for the validity of the empirical strategy!)
   - Show it graphically
   - Report in a table the estimates from an event study estimation: interact ‘treatment dummy’ with lags and forwards of the ‘ time dummies’ (e.g. year dummies)
2. Make sure that the timing of the adoption of the policy was as good as random (e.g. not driven by expectations of economic shocks)
3. No other policies were adopted at the same time
4. Verify that there is no reason to belief that the control group might be affected by the treatment

Classic example: Card and Krueger (1994)

Effect of Minimum wages on employment

- On April 1, 1992, New Jersey raised the state minimum from \$4.25 to \$5.05, whereas in the bordering state of Pennsylvania the minimum wage stayed at \$4.25 throughout this period.
- Card and Krueger (1994) evaluated the effect of this change on the employment of low wage workers.
- They conducted a survey to some 400 fast food restaurants from the two states just before the NJ reform, and a second survey to the same outlets 7-8 months after.

Other Examples

- Impact of legalized abortion on crime (Donohue and Levitt 2001).
    - 5 states that allowed abortion in 1970 compared to the rest, which legalized in 1973.
- Economic diversity at school and social preferences and behavior (Rao 2013).
    - In 2007, a policy change in India that introduced quotas for poor children in new admissions.
- Redistribution policies and investment in education (Abramitzky and Lavy 2014).
    - Different Israeli kibbutzim shifted from equal sharing to productivity-based wages in different years

Final comments on diff-and-diff:

- Identification relies on a claim that is very often more of an act of faith than a assumption clearly grounded on theory
- Show that in the past trends were parallel (it does not guarantee that this will be satisfied in the future, but at least it is supportive)
- Discuss explicitly why it is a good assumption to believe that the timing of the treatment/policy was as good as random
- Discuss explicitly the existence of alternative policies that might contemporaneously affect the treatment or the control group
- Discuss the possibility that the control group is affected by the treatment.
- Note: next day we will discuss in detail how to calculate properly standard errors in a diff-and-diff setting.