Applied Econometrics
Lecture 2

Giovanni Marin[1]

[1]Università di Urbino

Università di Urbino
PhD Programme in Global Studies
Spring 2018

Issues to be considered when doing empirical research

- What is the causal relationship of interest?
  - The question should be something like: "does my research consider the comparison between an outcome and a (potential) counterfactual that would have emerged in absence of 'something' that actually happened?"
  - Counterfactual $\Rightarrow$ what would have happened to the outcome of individual i if he had done something different from what he actually did? $\Rightarrow$ as in Back to the Future...

Issues to be considered when doing empirical research

- Is an experimental approach (potentially) suitable?
  - With an experiment, you have a treatment group (randomly selected from a population) and a control group (randomly selected from the same population)
  - Random selection implies that the control group, that does not receive the treatment, mimics what would have happened to the treatment group if it was not treated
  - Question $\Rightarrow$ could your research question answered by an hypothetical experiment $\Rightarrow$ you should not consider the actual feasibility of the experiment (e.g. ethical concerns, cost of the experiment, time needed to run the experiment, etc)
  - If the problem can be evaluated by means of an experiment, then the relationship you have in mind is 'causal' $\Rightarrow$ notion of counterfactual

Issues to be considered when doing empirical research

- Development of the identification strategy
  - If we actually have true experimental data, the strategy to measure the causal effect is extremely trivial (usually a mean comparison)
  - If we rely on observational data (as you probably will), the identification strategy is the way these data are used in a research design that is able to identify a causal link
  - Using the hypothetical experiment as a benchmark is always very useful to develop a successful identification strategy
  - In the real world, there are many situations that (more or less) mimic a controlled experiment $\Rightarrow$ quasi-experimental

Naive policy evaluation

- Example: assessment of the impact of hospitalization on health outcome
- Prior belief about the impact of hospitalization $\Rightarrow$ positive effect
- Available data $\Rightarrow$ information about the health status of two groups of people: i. people that have been hospitalized in the last 12 months; ii. people that have not been hospitalized in the last 12 months
- Policy evaluation: comparison of the average health outcome of the two groups
- Result: people that was hospitalized in the last 12 months has a worse health status than people that was not hospitalized
- W H Y ? ? ?

Naive policy evaluation

- The result that hospitalization worsens health (net of actual, but not so frequent, cases in which this happens due to the contagion from other ill people) is not credible...
- Question: were people that were not hospitalized a good control group (counterfactual)?
- NO!
  - People goes to the hospital when is injured or ill
  - People that did not go the the hospital had, on average, a better health status than hospitalized people even before hospitalization
  - The assignment of the treatment is not random, but is actually correlated with the outcome variable
- The problem is that people self-select into the treatment...
- Ideal framework: we observe the same person both in the case in which it decided to go to the hospital and in the case in which it decided not to go to the hospital $\Rightarrow$ need for a time machine...

Policy evaluation: optimal framework

More formally:

$$\text{Potential outcome} = \begin{cases} Y_{1i} & \text{if} \quad D_i = 1 \\ Y_{0i} & \text{if} \quad D_i = 0 \end{cases}$$

where:

- Y_{0i} is the outcome if the individual i did not go to hospital
- Y_{1i} is the outcome if the SAME individual i did go to hospital
- D_i is the treatment status (1 is treated, 0 is not)
- The potential outcome could be written as $Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$
- The treatment effect would be $Y_{1i} - Y_{0i}$

Policy evaluation: selection bias

- However, with observational data (but also with experiments...) we just observe one of the potential outcomes $\Rightarrow$ the individual i either went or not to hospital (no time machine)
- This means that we would observe Y_{0i} for those is that did not go to hospital ($D_i = 0$) and Y_{1i} for those is that did go to hospital ($D_i = 1$)
- Naive mean comparison between treated and control individuals will be:

$$E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0)$$

- If we add and subtract $E(Y_{0i}|D_i = 1$, that is the potential outcome that we cannot observe:

$$\begin{aligned} E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0) &= \underbrace{E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)}_{\text{ATT}} \\ &+ \underbrace{E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)}_{\text{Selection bias}} \end{aligned}$$

ATT and selection bias

- $E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)$ is the average treatment effect on the treated (ATT)
  - Comparison of the two outcomes for the ones that were ultimately treated ($D_i = 1$)
- $E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)$ is the selection bias
  - Difference in the outcome between treated and control if they were not treated
- The main objective of 'causal' econometrics is to elaborate a design in which the selection bias is eliminated

Experiments, ATT and selection bias

$$
\begin{aligned}
E(Y_{1i}|D_i=1) - E(Y_{0i}|D_i=0) &= \underbrace{E(Y_{1i}|D_i=1) - E(Y_{0i}|D_i=1)}_{\text{ATT}} + \\
&+ \underbrace{E(Y_{0i}|D_i=1) - E(Y_{0i}|D_i=0)}_{\text{Selection bias}}
\end{aligned}
$$

- What experiments do is to assign the treatment D_i randomly
- Random assignment $\Rightarrow$ Y_i will be independent on D_i
- Independence means that the potential outcome Y_{0i} is expected to be the same (on average) between treated and control groups
- If Y_{0i} is independent on D_i, this means that on average $E(Y_{0i}|D_i=1)$, that is not observable, is equal to $E(Y_{0i}|D_i=0)$, that is observed
- In words, the control group that is randomly assigned with no treatment represents a good counterfactual (what-if)

Experiments, ATT and selection bias

$$E(Y_{1i}|D_i=1) - E(Y_{0i}|D_i=0) = \underbrace{E(Y_{1i}|D_i=1) - E(Y_{0i}|D_i=1)}_{\text{ATT}} + \underbrace{E(Y_{0i}|D_i=1) - E(Y_{0i}|D_i=0)}_{\text{Selection bias}}$$

- Substituting $E(Y_{0i}|D_i=1) = E(Y_{0i}|D_i=0)$, mean comparison in experiments is:

$$\begin{aligned} E(Y_{1i}|D_i=1) - E(Y_{0i}|D_i=0) &= \underbrace{E(Y_{1i}|D_i=1) - E(Y_{0i}|D_i=0)}_{\text{ATT}} + \\ &+ \underbrace{E(Y_{0i}|D_i=0) - E(Y_{0i}|D_i=0)}_{\text{Selection bias=0}} \\ &= \underbrace{E(Y_{1i}|D_i=1) - E(Y_{0i}|D_i=0)}_{\text{ATT}} \end{aligned}$$

Regressions to evaluate experiments

- Mean comparison would be enough to estimate the average treatment effect on the treated in randomized experiments
- Regression can be used to do mean comparison $\Rightarrow Y_i = \alpha + \beta D_i + \varepsilon_i$
- $\hat{\alpha}$ will be the average outcome for the control group
- $\hat{\alpha} + \hat{\beta}$ will be the average outcome for the treated group
- $\hat{\alpha} + \hat{\beta} - \hat{\alpha} = \hat{\beta}$ will be the ATT
- Adding covariates in randomized experiment has no impact on the ATT estimation, but only reduces the variance of the estimation

Figure: Source: Angrist and Pischke (2008)

Table 2.2.2: Experimental estimates of the effect of class-size assignment on test scores

Explanatory variable	(1)	(2)	(3)	(4)
Small class	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
Regular/aide class	.12	.29	.53	.31
	(2.23)	(1.13)	(1.09)	(1.07)
White/Asian (1 = yes)	–	–	8.35	8.44
			(1.35)	(1.36)
Girl (1 = yes)	–	–	4.48	4.39
			(.63)	(.63)
Free lunch (1 = yes)	–	–	-13.15	-13.07
			(.77)	(.77)
White teacher	–	–	–	-.57
				(2.10)
Teacher experience	–	–	–	.26
				(.10)
Master's degree	–	–	–	-0.51
				(1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.25	.31	.31

Note: Adapted from Krueger (1999), Table 5. The dependent variable is the Stanford Achievement Test percentile score. Robust standard errors that allow for correlated residuals within classes are shown in parentheses. The sample size is 5681.

Regression as conditional expectation function

- Multivariate regressions aims at partialling out the outcome variable from some observed features
- Linear regression can be seen as a way to evaluate the Conditional Expectation Function $\Rightarrow E(y_i|x_{1,i}, x_{2,i}) = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i$
- As long as the Conditional Expectation Function that we have in mind is causal (e.g. treatment effect in a policy evaluation framework), regression will have a causal interpretation
- The idea of multivariate regression analysis is to include control variables with the objective of reaching the conclusion that, conditional (partialling out) on the controls, the assignment of the variable of interest across individuals is random $\Rightarrow$ no selection bias!
- $E(y_i|Z_i, x_i = 1) - E(y_i|Z_i, x_i = 0) = E(y_{1i} - y_{0i}|Z_i) \Rightarrow$ no selection bias
- Idea of 'selection on observable'

Research design and policy evaluation

- Recall the FAQs ⇒ think about your research question as a potential experiment
- Almost all well developed research questions can be expressed in terms of 'treatment effect' (x causes y)
- Research designs developed to perform policy evaluation can be use in many different framework
- Objectives of the policy evaluation methods:
  - Eliminate the selection bias
  - In the end, the selection bias corresponds to the omitted variable bias ⇒ (non-random) selection into treatment driven by factors that enter the error terms

Additional online lectures ⇒
`http://www.nber.org/minicourse3.html`

Policy evaluation designs

- Matching
- Difference in differences
- Regression discontinuity design

- These are not estimators!
- All of them could be done without any regression (but regressions help)
- My suggestion: do not rely too much on pre-compiled module in statistical software
  - Obviously, these modules work perfectly
  - However, it is very important to understand what is happening into the module to (and even more what is happening with pencil and paper) to develop a robust research design

Assumptions

- SUTVA - Stable Unit Treatment Value Assumption
  - The expected impact on the treatment should not influence the outcome of the individuals belonging to the control group
  - Example 1: evaluate the effect of a subsidy to half of the class that was funded by taxing the other half of the class $\Rightarrow$ the 'taxed' half is not a good counterfactual as it is influenced by the treatment
  - Example 2: evaluate the effect of limits to polluting emissions to firms with more than 100 employees within a sector on their level of production $\Rightarrow$ firms with less than 100 employees are likely to occupy the market shares left free by the regulated firms within the same sector $\Rightarrow$ bad counterfactual
  - General equilibrium effects
- Common support
  - The distributions for treated and control units of observed variables should be overlapped

Selection bias: again

- Usually, individuals that enter into a treatment (policy, education, investments, etc) are not randomly drawn
- Sometimes the treatment is attribute to individuals that satisfy some given requirements (e.g. establishments into the EU Emission Trading Scheme)
- In other cases, individual self-select into the treatment status (e.g. application to calls to obtain government's subsidies)
- If what drives the assignment to treatment is expected to influence the outcome itself, untreated individuals do not represent a good counterfactual for the treatment group

Comparing treated and controls

- A way to get rid of the selection bias is to compare each treated individual with individuals that are similar ⇒ quasi-experimental approach
- Similar in what?
  - Employment? No (otherwise you would get, by construction, ATT=0)
  - Observable features ⇒ how many?
  - Unobservable features ⇒ not possible...
- Matching
  - Identify a group of similar individuals for each treated individual and compare the averages ⇒ selection on observables

Matching: unconfoundedness

- Selection into the treatment only depends on the observed characteristics
- Consequence $\Rightarrow$ for a given vector of observed characteristics, the fact of being treated or not is completely random
- The concept of unconfoundedness corresponds to the concept of Conditional Independence Assumption
- If unconfoundedness holds, the average treatment effect on the treated is computed as follows:
  1. For each cell, compute the average outcome variable separately for treated and untreated individuals
  2. Compute the difference between the two averages within each cell
  3. Compute the average of these differences across all the cells, weighted by the number of treated individuals belonging to each cell

Matching

- Example: subsidy to companies to hire young workers
- We have information on companies that receive the subsidy and companies that did not about:
  - Outcome variable (number of young employees hired)
  - Size of company (sales)
  - Industry of the company
  - Profitability (ROE) and productivity (value added per employee)
  - Region
- How do we combine all these different dimensions? $\Rightarrow$ matching

Matching

- Main objective ⇒ compare the average outcome variable of treated and (as similar as possible) untreated firms ⇒ if the firms in the two groups are very similar, any difference in the outcome variable can be attributed to the fact of receiving the subsidy
- We can start building cells that combine industry and sales' bands
- Within each cells, all companies will belong to the same sector and have a somewhat similar level of sales
- What if we also want to account for the region? ⇒ we need to further split each industry/sales cell into multiple cells, one for each region
- The more dimensions we add, the higher is the risk that within a cell we just find treated individuals (and no controls ⇒ dimensionality issue!
- Even finding just one control firm in a cell with many treated firms may be problematic (e.g. if the control firm is an outlier...)

Matching on the propensity score

- A solution to the (severe) dimensionality problem is the identification of a latent variable that combines all the different dimensions that influence the selection into treatment $\Rightarrow$ propensity score
- The propensity score is the conditional probability $P(T = 1|X)$ that an individual receives the treatment, given a set of covariates X
- How to compute the propensity score?
  1. Estimate a probit or logit (or LPM) regression with the treatment dummy as the dependent variable and observable features that drive selection into treatment as independent variables
  2. Estimate the predicted probability
- If the propensity score is the 'true' propensity score, assignment into treatment is random given the propensity score $\Rightarrow$ unconfoundedness

Matching on the propensity score

- Once the propensity score is estimated, each treated individual needs to be matched with one (or more) untreated individual that has a similar probability of being treated
- Given the propensity score (i.e. individuals with similar propensity score), the assignment into treatment is random
- Average treatment effect (on the treated)
  - Compute the difference in the outcome variable between each treated individual and its closest (untreated) individual(s) in terms of propensity score
  - Compute the average across all treated individuals

Matching on the propensity score

- There are different possible algorithms to be used to match treated and untreated individuals
  1. Nearest neighbour ⇒ only match one untreated individual, the closest in terms of PS
  2. Nearest neighbour ⇒ match the N closest individuals
  3. Caliper (combined - or not - with NN) ⇒ match all unmatched individuals that are within a certain range of estimated PS of the treated individual (e.g. 1 percent)
  4. Kernel ⇒ create a counterfactual for each treated unit that combines all potential untreated individuals, with weights that decrease with the distance in terms of estimated PS
- There is a trade off between
  - Bias ⇒ the higher the distance (in terms of propensity score) between treated and matched untreated, the larger the difference in terms of observable characteristics
  - Precision ⇒ the variance of the estimated ATT is larger the smaller is the size of the control group

Matching on the propensity score: tips

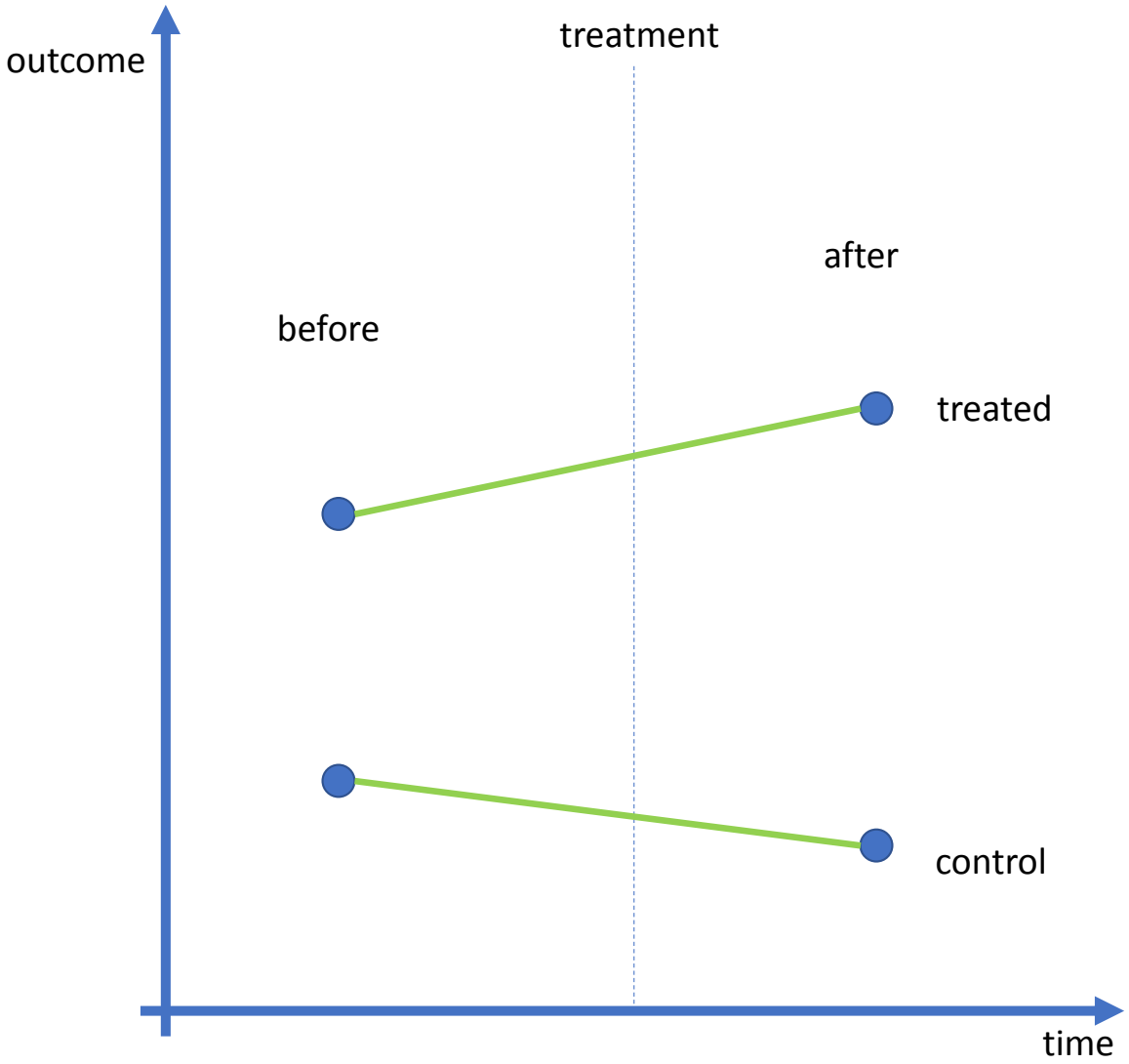
- The distributions of estimated propensity score between treated and control individuals need to have the same support ⇒ common support assumption ⇒ the support of the estimated PS of treated individuals should be contained in the support for untreated individuals
- Observable features should be measured, ideally, before the treatment
- A crucial diagnostic check is to evaluate whether the matching is good at eliminating differences (on average and by block of PS) in observable characteristics between treated and control individuals ⇒ useful to check the balancing properties for both variables that were included in the estimation of the PS and for other variables
- It is always possible to combine the propensity score matching with exact matching on certain features

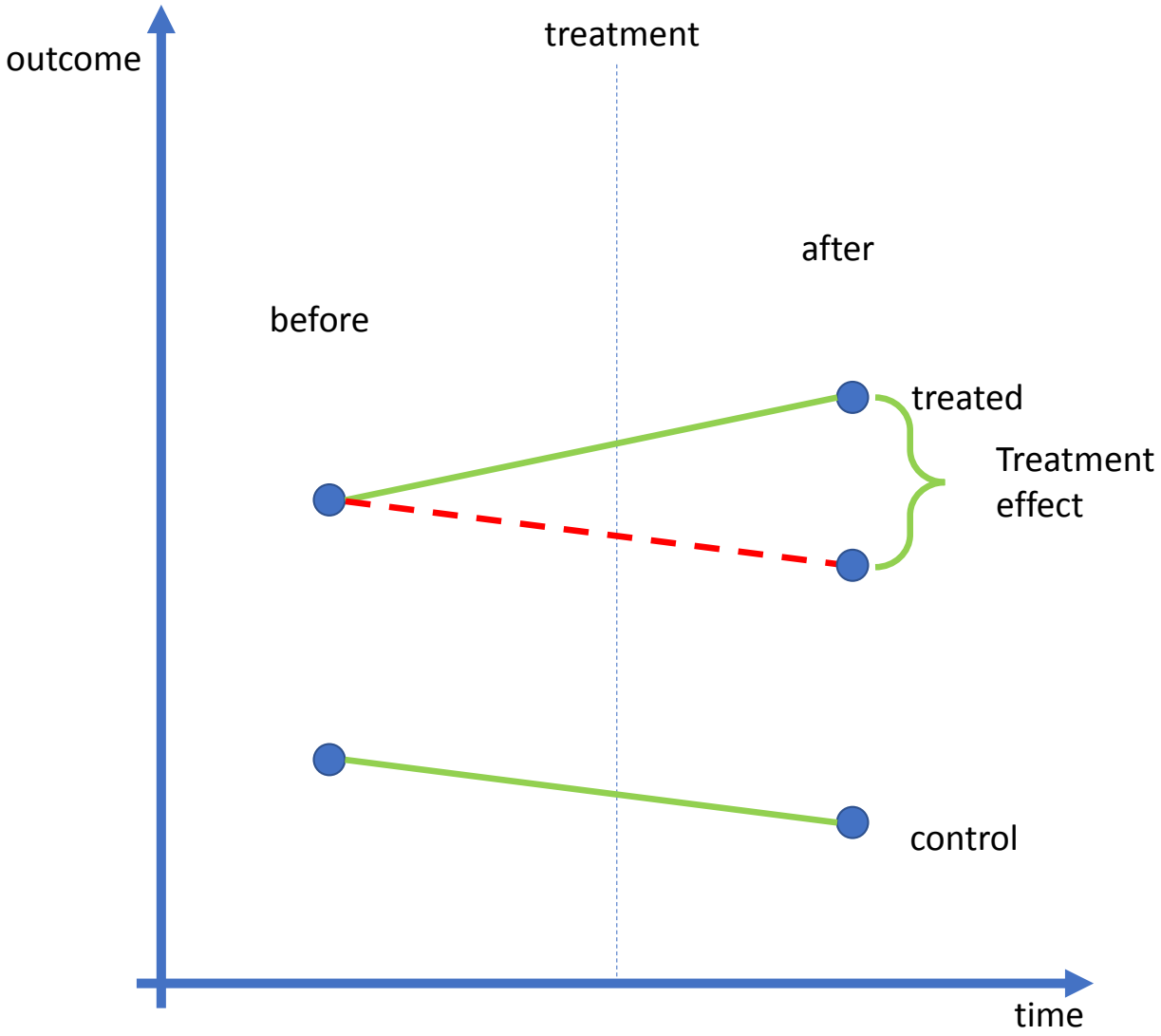
Before-after comparison

- Assume that you can observe your treatment group in two points in time: before and after the treatment
- The first temptation would be to estimate the treatment effect by comparing the average outcome before and after the treatment
- Why this is not a good idea? $\Rightarrow$ the change in average outcome in the treatment group would be driven by a large variety of factors (e.g. long run trends) different from the fact of receiving a treatment
- How to exploit the time dimension? $\Rightarrow$ difference in differences

Difference in differences

- Assume that the assignment to treatment is not random
  - Treated and controls differ in some specific features
  - Treatment and control group belong to different 'populations' of reference
- Comparing average outcomes of the two groups after the treatment is not enough, as the difference could be due to non-random features
- However, it could be reasonable to assume that, though different, the two groups would have evolved in the same way in absence of the treatment


outcome
treatment
after
before
treated
control
time


outcome
treatment
after
before
treated
Treatment
effect
control
time

Difference in differences

- How to compute the treatment effect, where the treatment is defined with the dummy variable D_i
  1. Compute the difference in outcome between treated and control group in the pre-treatment period $\Rightarrow$ $E(y_{it}|t = pre, D_i = 1) - E(y_{it}|t = pre, D_i = 0)$
  2. Compute the difference in outcome between treated and control group in the post-treatment period $\Rightarrow$ $E(y_{it}|t = post, D_i = 1) - E(y_{it}|t = post, D_i = 0)$
  3. Compute the difference between the two differences $\Rightarrow$ difference-in-differences!
- Alternatively (and equivalently) it is possible to compute the growth in the outcome variable of the treated, the growth in the outcome variable in the control and compute the difference between the two

Difference in differences

- The main identification assumption is that, in absence of the treatment, the two group would have evolved in the same way
- A first check would be to evaluate how different were the treatment and control groups in terms of characteristics that may have some influence on the dynamics (not the level) of the outcome variable $\Rightarrow$ possible to combine with matching
- A second check would be to assess, if possible, whether treated and control group had similar trends even before the treatment $\Rightarrow$ pre-treatment common trend assumption
  - If you have information on the outcome variable for two or more periods before the treatment, you can test this assumption

Difference in differences with regression

- The difference-in-difference treatment effect can be simply computed by comparing simple averages
- A regression framework is usually employed to estimate the treatment effect
  - Easy to do inference (i.e. estimating standard errors of the effect)
  - Add control variables that help the isolation of common trends (conditional on controls)
- Specification:

$$Y_{it} = \alpha + \beta D_i + \gamma Post_t + \delta D_i \times Post_t + \varepsilon_{it}$$

Difference in differences with regression

$$Y_{it} = \alpha + \beta D_i + \gamma Post_t + \delta D_i \times Post_t + \varepsilon_{it}$$

where:

- $E(Y_{it}|D_i = 0, Post_t = 0) = \hat{\alpha}$
- $E(Y_{it}|D_i = 1, Post_t = 0) = \hat{\alpha} + \hat{\beta}$
- $E(Y_{it}|D_i = 0, Post_t = 1) = \hat{\alpha} + \hat{\gamma}$
- $E(Y_{it}|D_i = 1, Post_t = 1) = \hat{\alpha} + \hat{\beta} + \hat{\gamma} + \hat{\delta}$

This means that the treatment effect is given by:

$$\begin{aligned}
&[E(Y_{it}|D_i = 1, Post_t = 1) - E(Y_{it}|D_i = 0, Post_t = 1)] - \\
&[E(Y_{it}|D_i = 1, Post_t = 0) - E(Y_{it}|D_i = 0, Post_t = 0)] = \\
&= \hat{\delta}
\end{aligned}$$

Quasi-experimental approach

- Assume that the treatment is assigned to all individual above a certain threshold of a continuous variable ⇒ e.g. ranking in a pass-list, where only individuals above 30pt pass
- Individuals right above the threshold (30pt) will be very similar to individuals right below the threshold (29pt)
- It is reasonable to assume that, around the threshold, assignment to treatment is random
- This discontinuity can be exploited to estimate the treatment effect ⇒ regression discontinuity design

Regression Discontinuity Design

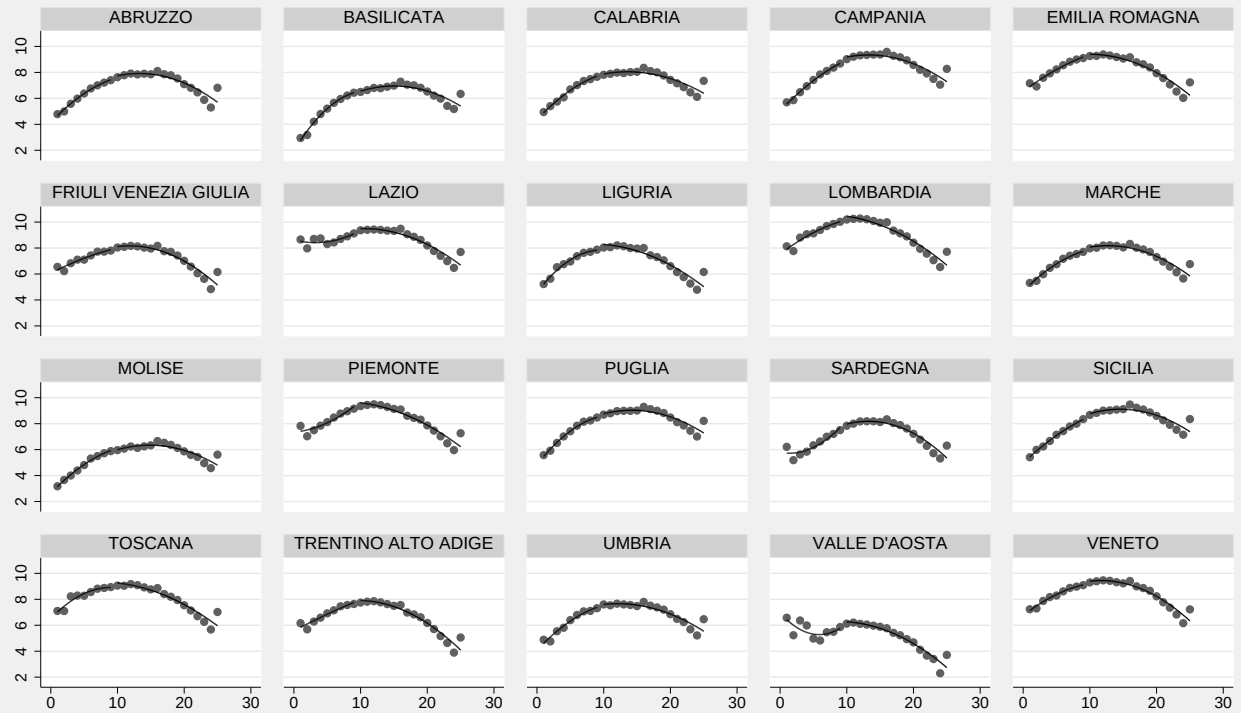
- The idea is that, in absence of treatment, the outcome variable y_i was some function of the assignment variable $x_i \Rightarrow y_i = f(x_i) + \varepsilon_i$
- If the policy had an effect on the outcome variable, however, we should expect a 'jump' in this function around the threshold $\Rightarrow$ $y_i = f(x_i) + \rho T_i + \varepsilon_i$
- Which function for $f(x_i)$?
  - Linear
  - Quadratic
  - Non-parametric (spline)
  - Common (or not) in the two sides of the threshold
- The estimated effect, however, cannot be extended to all treated individuals, but is likely to be 'valid' only for treated individuals that are close to the threshold
- The results of the RDD are often reported only graphically

Description of the scrapping scheme (2009)

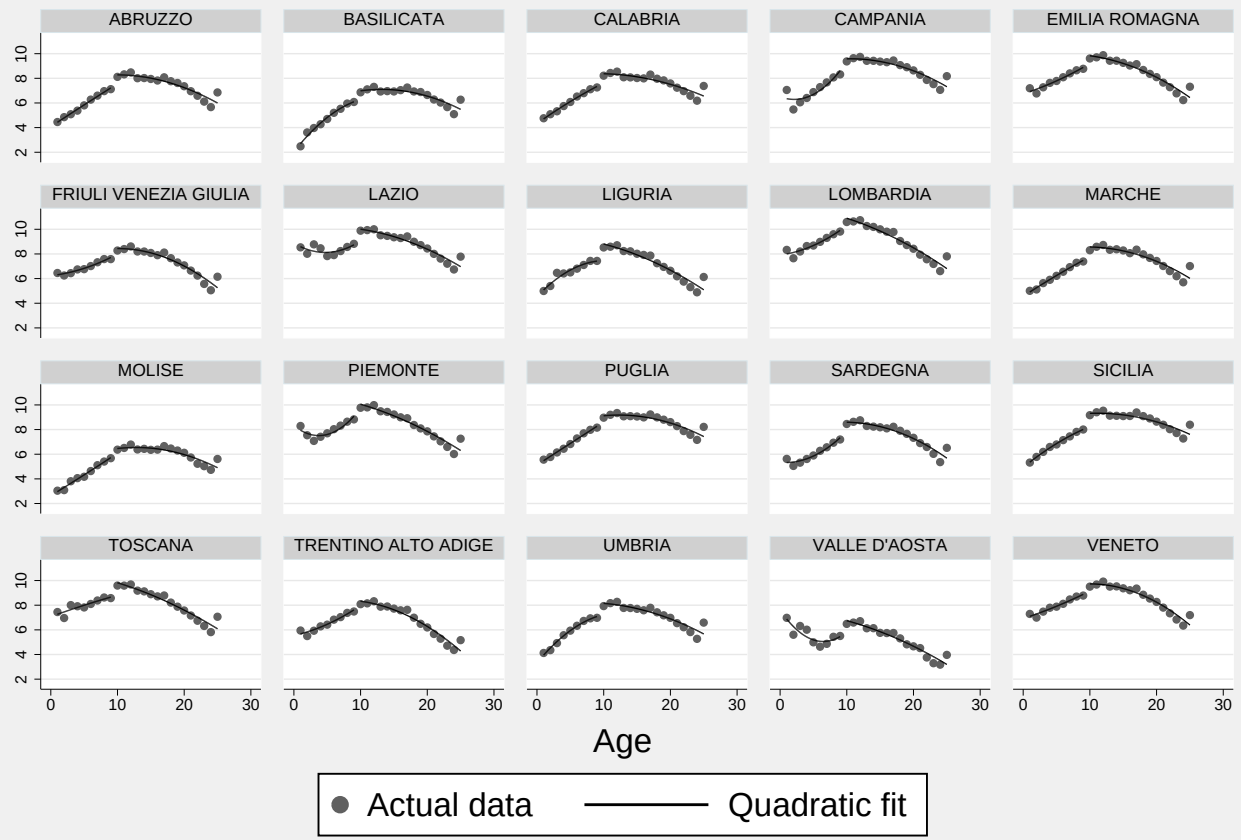
- **Scrapping scheme** introduced by the Italian government in **February 2009** (L. 33/09):
  - **Subsidy** of 1500 euro (with no budget limit) for **buying a new vehicle** after **scrapping** a **vehicle registered before January 2000** and compliant with **EURO2 or lower**
  - **Further increase** in the subsidy if the new car was fuelled with **LPG**
  - Programme active **until December 2009**
- The **scheme** is **national**, but targeted to **specific categories of cars** (i.e. older than 10 years)
- We exploit this **discontinuity** to **identify** the effect of the scheme
  - The **likelyhood** of scrapping a car that is **9 years old** is **similar** to the one of scrapping a car that is **10 years old** (in absence of the scheme)
  - **Before** (2008) and **after** (2010) the scheme there should have been **no particular discontinuity** around the **age of 10** for scrapping cars

$\Rightarrow$ **Regression Discontinuity Design** (**RDD**)

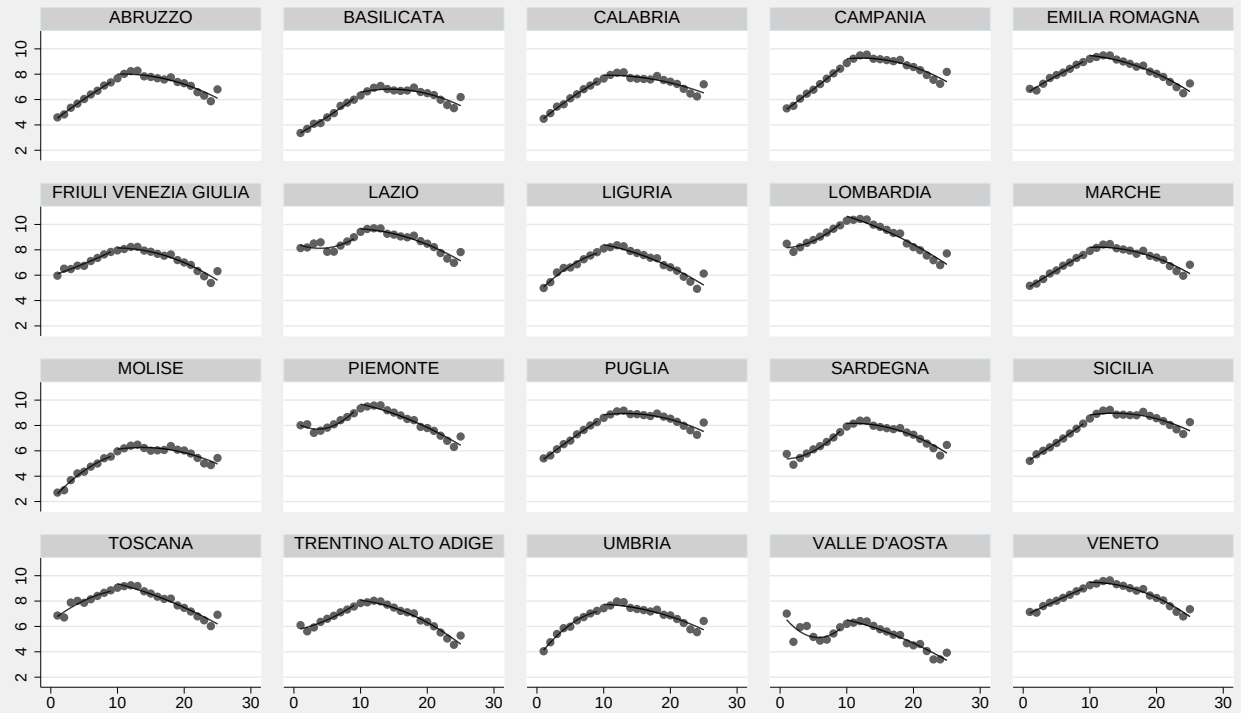
RDD - year 2008 (placebo)



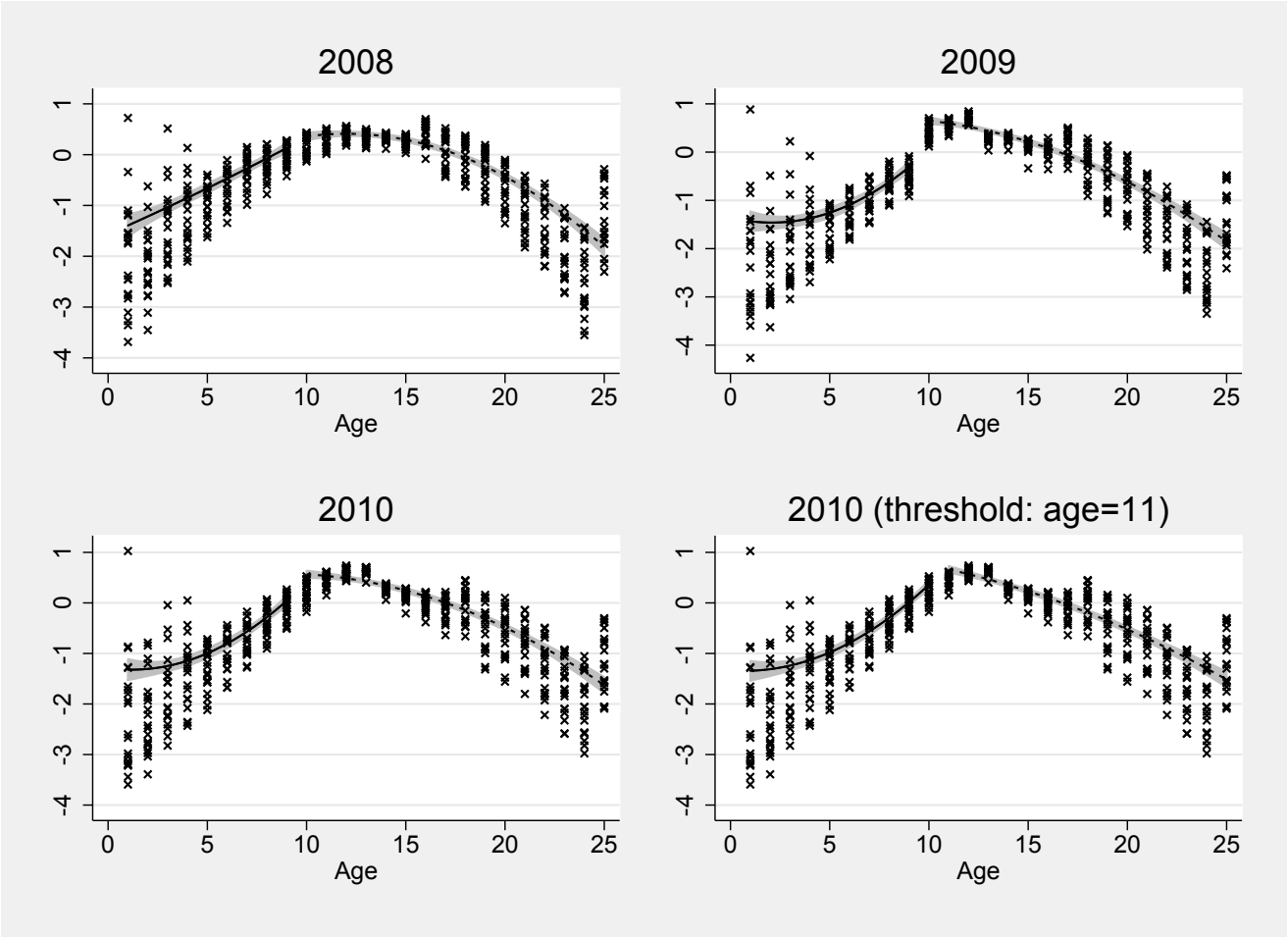
RDD - year 2009



RDD - year 2010 (placebo)



RDD - comparison



RDD estimates

Table: RDD for different years

RDD (quadratic)	2007	2008	2009
Dummy age≥10	-1.061	0.122	1.718***
	(0.710)	(0.534)	(0.458)
RDD (quadratic - region specific)	2007	2008	2009
Dummy age≥10	-0.691	0.336	1.887***
	(0.479)	(0.362)	(0.360)

N=500. Dependent variable: logarithm of deregistered cars by region and age. OLS model weighted by total deregistered cars by year and region. Standard errors clustered by region and age in parenthesis. * p<0.1, ** p<0.05, *** p<0.01. Quadratic fit (pooled or region-specific) is allowed to differ for cars with 9 or less years and cars with 10 or more years.

Table: RDD for different age thresholds (2009)

RDD (quadratic - region specific)	(1)	(2)	(3)
Dummy age≥8	-0.979*		
	(0.523)		
Dummy age≥9		0.171	
		(0.550)	
Dummy age≥10			1.887***
			(0.360)

N=500. Dependent variable: logarithm of deregistered cars by region and age for year 2009. OLS model weighted by total deregistered cars by year and region. Standard errors clustered by region and age in parenthesis. * p<0.1, ** p<0.05, *** p<0.01. Quadratic region-specific fit is allowed to differ for cars with 7, 8 and 9 or less years and cars with 8, 9 and 10 or more years in specification 1, 2 and 3 respectively.