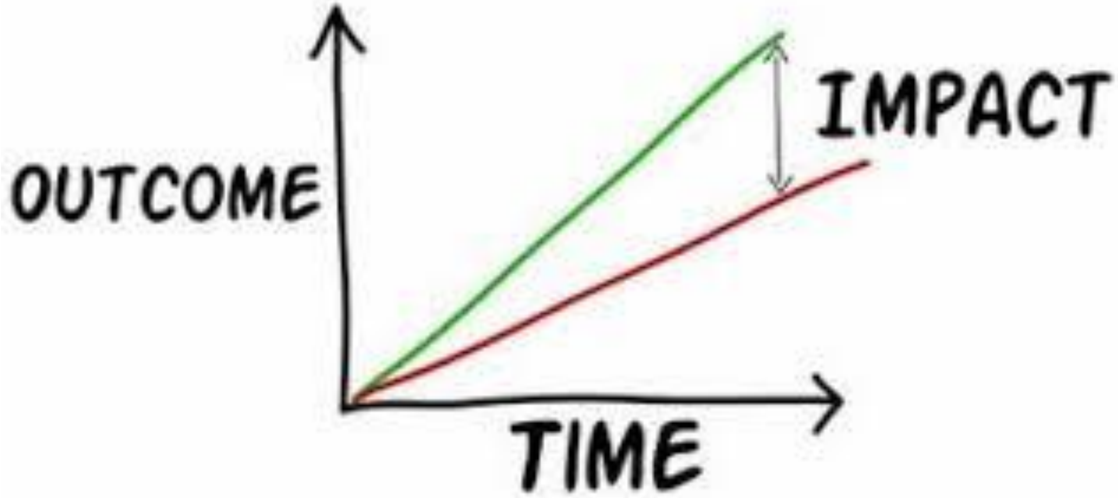
SUFINANCIRANO SREDSTVIMA EUROPSKE UNIJE
EUROPSKI POLJOPRIVREDNI FOND ZA RURALNI RAZVOJ –
EUROPA ULAŽE U RURALNA PODRUČJA

Impact evaluation tools



Agenda

1. Little recap from yesterday
2. Definitions and basic concepts:
    1. Correlation and causality
    2. Treated and control groups
    3. Intended and unintended effects
    4. Selection bias
3. Impact evaluations
4. Most common impact evaluation methods
    - Random selection
    - Instrumental variables
    - Regression discontinuity design
    - Difference in differences
    - Propensity score matching
5. Examples and case studies

Aims of the lecture

1. Getting familiar with the impact evaluation and its main tools
2. Being able to interpret basic evaluation methods

SUFINANCIRANO SREDSTVIMA EUROPSKE UNIJE
EUROPSKI POLJOPRIVREDNI FOND ZA RURALNI RAZVOJ –
EUROPA ULAŽE U RURALNA PODRUČJA

Recap from yesterday

Any questions or pending issues

Basic statistical concepts

1. Population
2. Sample
3. Variable
4. Parameter
5. Distribution

Basic concepts

1. Correlation and causality
2. Treated and control groups
3. Intended and unintended effects
4. Selection bias

Motivation

- Estimate the CAUSAL effect (impact) of
  - intervention P (program or treatment)
  - on
  - outcome Y (indicator, measure of success)
- Example: what is the effect of
  - RDP interventions (P)
  - on
  - Economic performance of supported farms (Y)?

Causality and correlation

Goal of impact evaluation: finding association between variables – causal effects => X: explanatory variable(s), Y: dependent variable

- X: explanatory variable, the independent variable, the explanatory variable, the control variable, the predictor variable, or the regressor
- Y: dependent variable, explained variable, response variable, the predicted variable, or the regressand

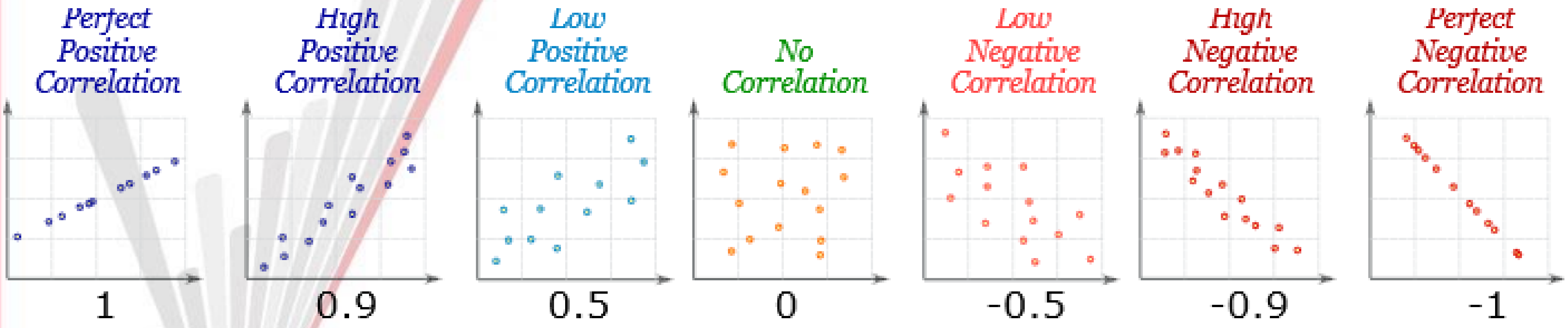
How to measure causal effects with statistical tools? Correlation

- Correlation = „a quantity measuring the extent of the interdependence of variable quantities." other words: how they fluctuate together

Correlation

Correlation coefficient: measures the <u>strength</u> of the relationship between two variables: $-1 \leq r_{XY} \leq 1$

positive (+), negative (–) or no correlation



Causality vs.correlation

Correlation ≠ causality

- Causality = changing one factor affects another one – it is unilateral: X affects Y
- Correlation: symmetric: X affect Y and vica versa ($r_{XY} = r_{YX}$)

Ex post evaluations aims to differentiate correlation from causality.

- In complex settings: not easy to identify the link, because of:
  1. Inverse causality/alternate reasoning:.
  2. Hidden causes/confounding factor = associated both with the outcome of interest and with the intervention of interest
  3. No real relationship: accidental fluctuation
- Primary knowledge is required to discover proper links between factors
- Ceteris paribus approach: how does the change of Xi affect Y, Holding other factors fixed?

Clear up the confusion with an example:

Ice cream sales and Temperature

- Causal relation does exist: warmer weather caused more sales
- But: ice cream sales did not caused hot weather

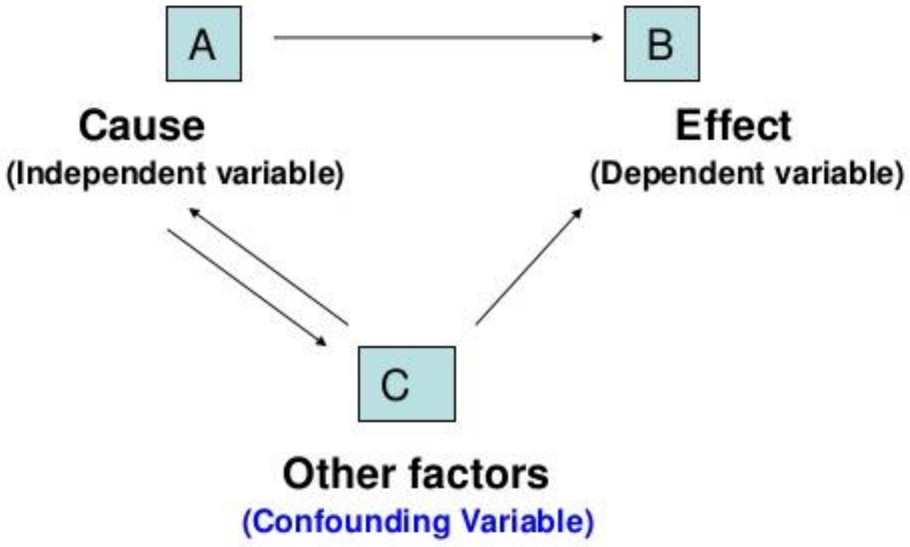
Ice cream sales and Sunglasses sold

- Positive correlation can be found. Is it a causal relationship?
- No, causal relation does NOT exist

Does ice cream sales affect the number of sunglasses?

- The cause for both is the outdoor temperature

Confounding factor



(The apparent association between A and B may be due to a third variable, C which associates with both A and B)

Soruce: http://dunesguesthouse.com/confound-psychology/

Confounding factor = associated both with the outcome of interest and with the intervention of interest

Do Police Reduce Crime? (Levitt, 1997)

X = Number of policeman in a city
Y = Level of crime

Hypothesis – *Does the presence of more policeman reduce the level of crime? Does the presence of more policeman cause a reduced level of crime? Is there unilateral relationship?*

Answer: Causal relation does not exist.

Inverse causality: Why is there more police?

- More crime might have led to the enhanced police presence
- More policeman might have reduced the level of crime

Simple Linear Regression Model

Two-variable linear regression model

Error term/disturbance = factors other than x affecting y

$$y = \beta_0 + \beta_1 x + u$$

Intercept parameter

Slope parameter

If the other factors in *u* are held fixed, so that the change in *u* is zero, $\Delta u = 0$, then *x* has a linear effect on *y*:

$$\Delta y = \beta_1 \Delta x \; if \; \Delta u = 0$$

= Holding all other factors fixed

Ceteris paribus (CP) approach: other relevant factors being equal

- Does the modell allow us to draw ceteris paribus conclusions about how x affects y?

Is this the end of the causality issue?

Simple Linear Regression Model

CP approach is applicable only if there is a restricting assumption on how the unobservable u is related to the explanatory variable x

Reason: u and x are random variables => any concept should be based on probability.

Crucial assumption: **zero conditional mean assumption**

$$E(u|x) = E(u) = 0$$

u does not depend on the value of x

Simple Linear Regression Model

What if zero conditional mean assumption is violated?

If $E(u|x) \neq 0$

$\Rightarrow$ Estimation is biased

$\Rightarrow$ β_1 does not show real relationships

$\Rightarrow$ CP approach is not applicable

- Mosly caused by hidden variables – „confounding factor"

What are impact evaluations?

Impact evaluation = an assessment of how the intervention affects outcomes, whether these effects are intended or unintended. (OECD definition)

- Qualitative or quantitative (to quantify these effects)
- Ex ante or ex post: methodology depends on the approach

Counterfactual: what the outcomes would have been in the absence of the intervention:

- Theoretical scenario

How to measure if it is only theoretical?

Intended and unintended impact

Interventions usually have intended and unintended impacts

Most common types of unintended impacts:

- Crowding out effects
- Deadweight loss
- Multiplier effect
- Etc.

SUFINANCIRANO SREDSTVIMA EUROPSKE UNIJE
EUROPSKI POLJOPRIVREDNI FOND ZA RURALNI RAZVOJ –
EUROPA ULAŽE U RURALNA PODRUČJA

Treated and control groups

- **Treated group**: participants of a program, or direct beneficiaries of an intervention – observed units
- **Control group**: non-participants of the same program

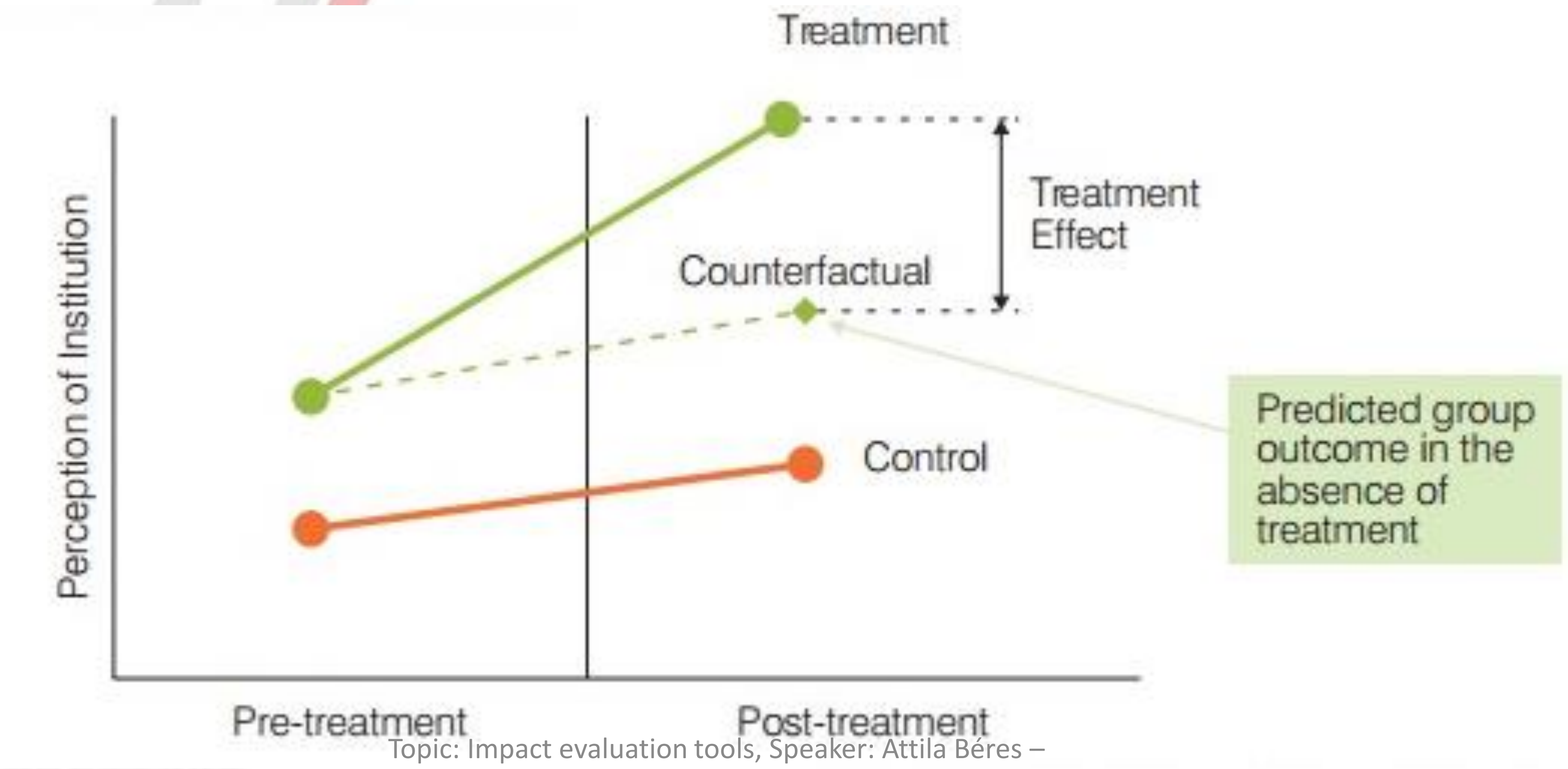
  Assumption: the same would have been happened to the treated group members in the absence of the program!

Average treatment effect (ATE) = ATE measures the difference in mean outcomes between units assigned to the treatment and units assigned to the control

Average Treatment Effect on Treated (ATT)

Counterfactual

= what would have been the outcomes in the absence of the intervention?



source: https://jeffbloem.wordpress.com/2015/12/24/the-economics-of-its-a-wonderful-life/counterfactual/

Selection bias

Problem: lack of random selection => selection bias

Selection bias = there is an error with the sampling and having a selection for analysis that is not properly randomized =>

subgroups differ from the population in some systematic and important way affecting the outcome.

- Observable and non-observable differences.
- Most common types: self-selection and skimming

Estimated effect is biased: bias leads to over- or underestimate the true value, since some effect are attributed to the program, although those impacts are the results of systematic differences between the treated and control groups

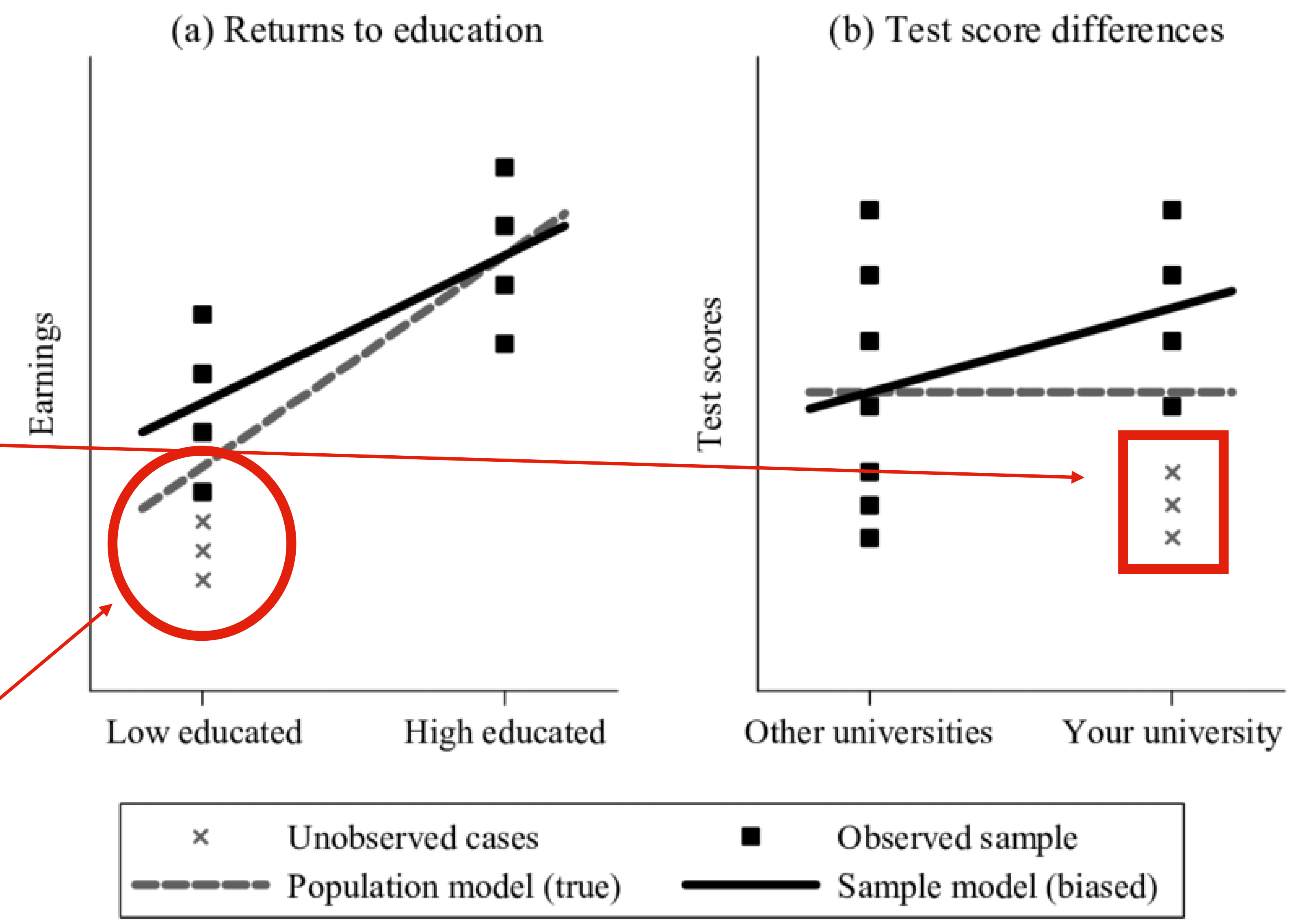
E.g.: training programs – participants are supposed to be more motivated, or people with higher skills are enrolled in the early stages of the program. Accordingly: the longer a program takes, the less efficient the overall outcome.

SUFINA
EUROPSKI P
EU

Selection bias
Examples

Did not take part on the exam

Non-working, therefore not observed



Relevance of a well-chosen control group

Control group: comparison group – Finding a proper control group is a precondition of drawing a valid conclusion!

= providing the validity of the methodology!

Control group is required to have (almost) the same features as treated group has.

In order to eliminate confounding factors: *we need to make sure that the only difference between the treated and the control groups is participation in the programme*

- **Why?** In this case, if the two groups are identical, the only expected difference between the control and treated group is the outcome variable being studied (see next slide).
- **How?** Randomized controlled trials (RCTs)
- With randomization treatment effect can be estimated

SUFINANCIRANO SREDSTVIMA EUROPSKE UNIJE
EUROPSKI POLJOPRIVREDNI FOND ZA RURALNI RAZVOJ –
EUROPA ULAŽE U RURALNA PODRUČJA

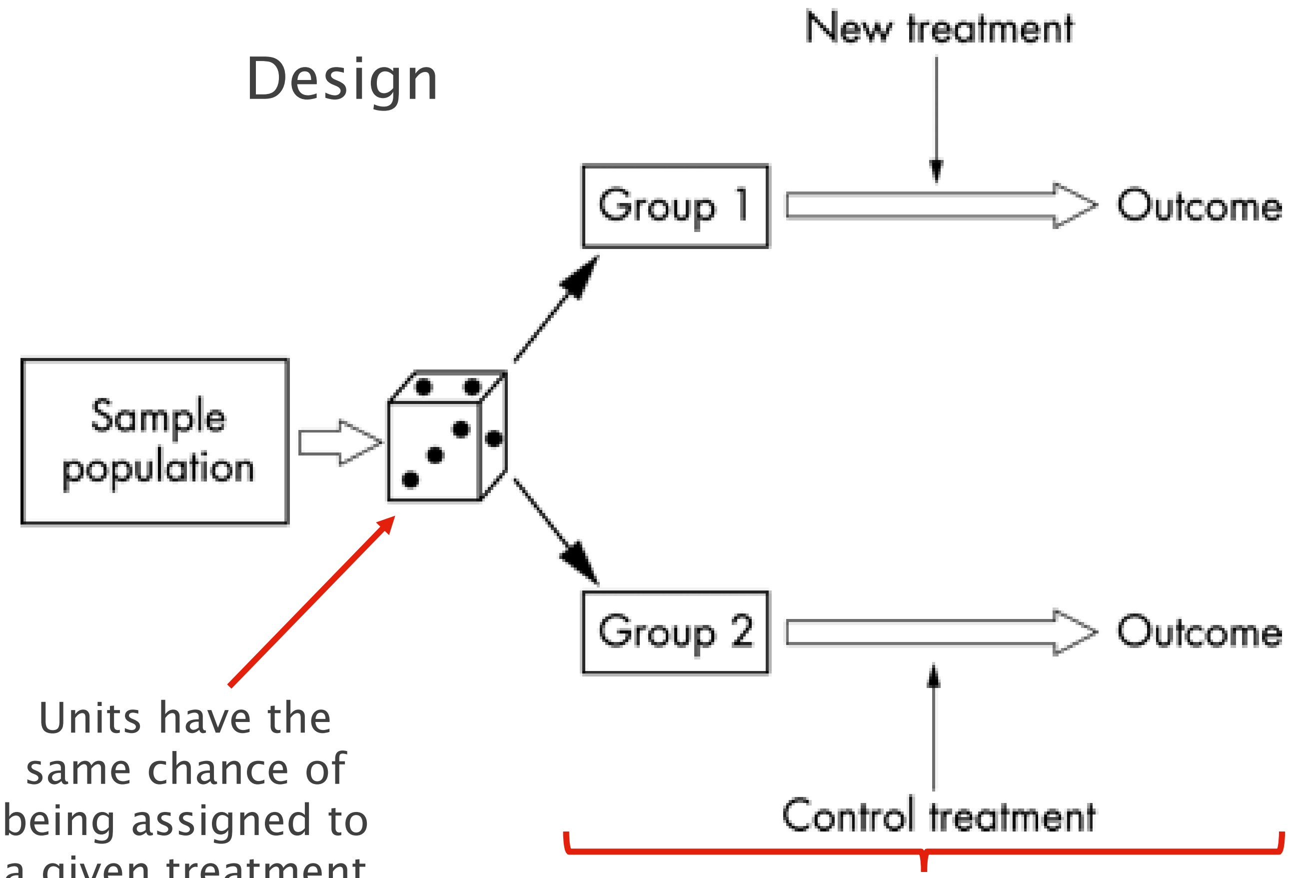
Randomized controlled trials

RCT = „a trial in which subjects are randomly assigned to one of two groups: the treated group receiving the intervention that is being tested, and the control group. The two groups are then followed up to see if there are any differences between them in outcome."

⇒ RCT can determine whether a cause-effect relation exists without knowing possible confounding factors!

E.g. Placebo-controlled study

Design



Units have the same chance of being assigned to a given treatment condition

As such, random assignment ensures that both the experimental and control groups are equivalent.

Randomized controlled trials

Disadvantages	Advantages
• Expensive in terms of time and money • Volunteer biases: the population that participates may not be representative of the whole	• Good randomization will "wash out" any population bias • Results can be analyzed with well known statistical tools • Populations of participating individuals are clearly identified

Source: https://himmelfarb.gwu.edu/tutorials/studydesign101/rcts.cfm

Quasi-experimental design

In real settings, randomization is not easy to implement, but need to manage them. In this case:

Quasi-experiments = „empirical interventional studies used to estimate the causal impact of an intervention on target population without random assignment"

- QE is similar to RCT, without randomity
- QE uses some criterion other than random assignment (e.g. age)

Commonly used evaluation methods (e.g. Diff-in-Diff, PSM,…) are based on quasi-experiments

Quasi-experimental design

Disadvantages	Advantages
• Potential for Non-Equivalent Groups • Potential for Low Internal Validity (Because of the lack of random assignment) • The impact are subject to contamination by confounding variables.	• Logistically Easy to Conduct • Control Group Comparisons Possible • Since quasi-experiments are natural experiments, findings in one may be applied to other subjects and settings, allowing for some generalizations to be made about population.

Commonly used impact evaluation methods

1. Instrumental variables
2. Regression discontinuity design
3. Difference in differences
4. Propensity Score Matching

Instrumental variables

$$y = \beta_0 + \beta_1 x + u,$$

where y:outcome, x: treatment

- Problem: units that are assigned to the treatment and comparison groups comply with their assignment (=imperfect compliance)

In other words: there is a hidden factor affecting:

1. whether a subject receives treatment or not (x)
2. the outcome (y)

- Consequence: OLS estimation is biased (and inconsistent)
- Solution: instrument – it helps to evaluate programs with imperfect compliance

Instrumental variables

Concept behind the IV:

- IV provides with some external source of variation to determine treatment status
- An instrumental variable influences the likelihood of participating in a program, but is outside of the participant's control and is unrelated to the participant's characteristics
- Intuitively, we can think of an IV as something outside the control of the individual that influences her likelihood of participating in a program, but is otherwise not associated with her characteristics

SUFINANCIRANO SREDSTVIMA EUROPSKE UNIJE
EUROPSKI POLJOPRIVREDNI FOND ZA RURALNI RAZVOJ –
EUROPA ULAŽE U RURALNA PODRUČJA

IV definitions

1. **Instrumental Variable (IV)** = „In an equation with an endogenous explanatory variable, an IV is a variable that
   - does not appear in the equation
   - is uncorrelated with the error in the equation
   - and is (partially) correlated with the endogenous explanatory variable."
2. **Instrumental Variables (IV) Estimator** = „An estimator in a linear model used when instrumental variables are available for one or more endogenous explanatory variables."

Instrumental variables

Cons	Pros
• Finding strong and valid instrumental variables that affect participation in the treatment but do not have a direct effect on the outcome of interest is difficult • Estimated treatment effects do not generally apply to the whole population, nor even to all the treated observations. • Estimated treatment effects may vary across different instruments. • For small sample sizes, and in case of “weak” instruments, instrumental variable estimates are biased. • Exclusion restriction is not testable	• Valid instrumental variables help to establish causality, even when using observational data • Using instrumental variables helps to address omitted variable bias • Instrumental variables can be used to address simultaneity bias • To address measurement error in the treatment variable, instrumental variables can be used

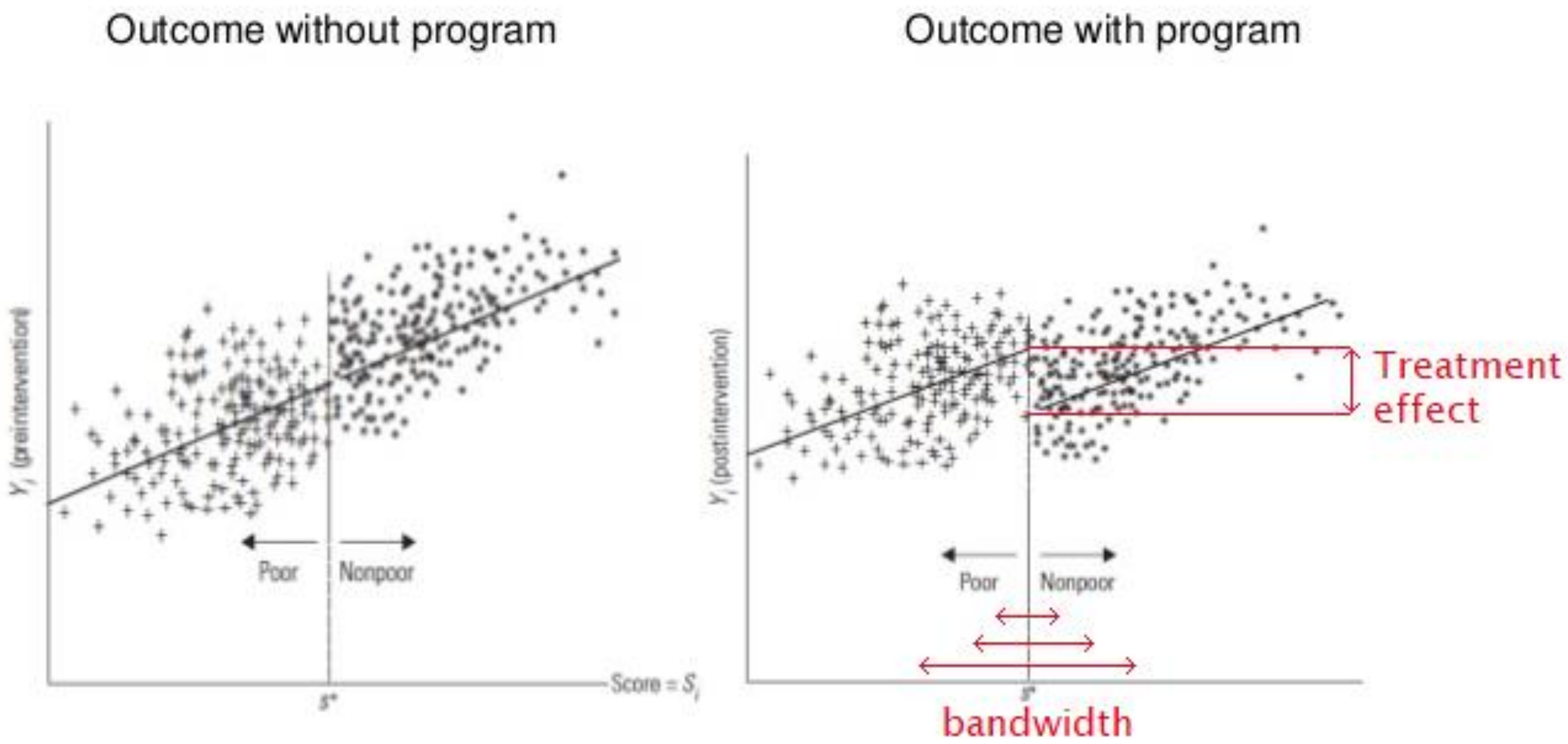
Regression Discontinuity Design (RDD)

- Treatment and control groups are distinguished based on a cutoff point => Assignment is **not** random
- Concept behind: for the population around the cutoff: RDD ~ randomized experiment
  - Units just below the threshold are similar to units just above the threshold! => variation in treatment near the threshold is randomized => LOCAL effect only – results depend on the width of the window around the threshold
    - X is an observed variable assigning the treatment
    - Treated if $X \geq C$ and subjects are not treated if $X < C$, where C is the cutoff value
- Idea behind: probability of participation is not continuous at the cutoff point
- RDD allows to identify causal effects using observational data

Regression Discontinuity Design

Possible cutoff points:

- Age
- Test scores
- Geography
- Poverty index
- Scoring
- Etc.



Regression Discontinuity Design

Sharp versus Fuzzy design:

Sharp design

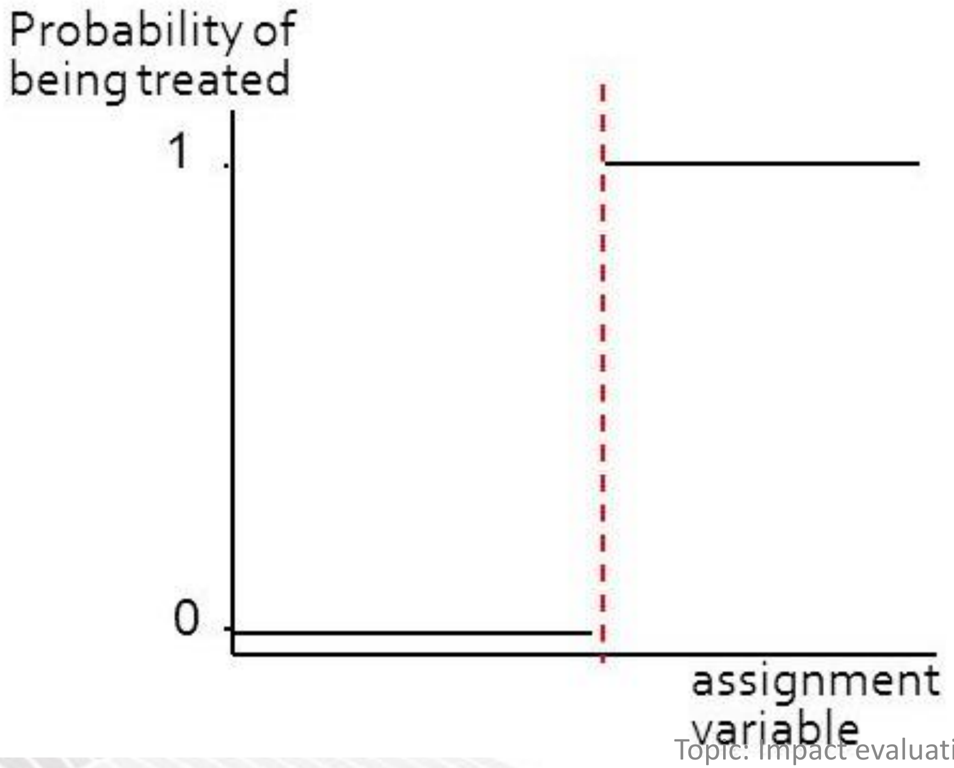
= Eligibility rules are strictly enforced and the probability of participation is either zero or one (of the two sides of the threshold)

Fuzzy design

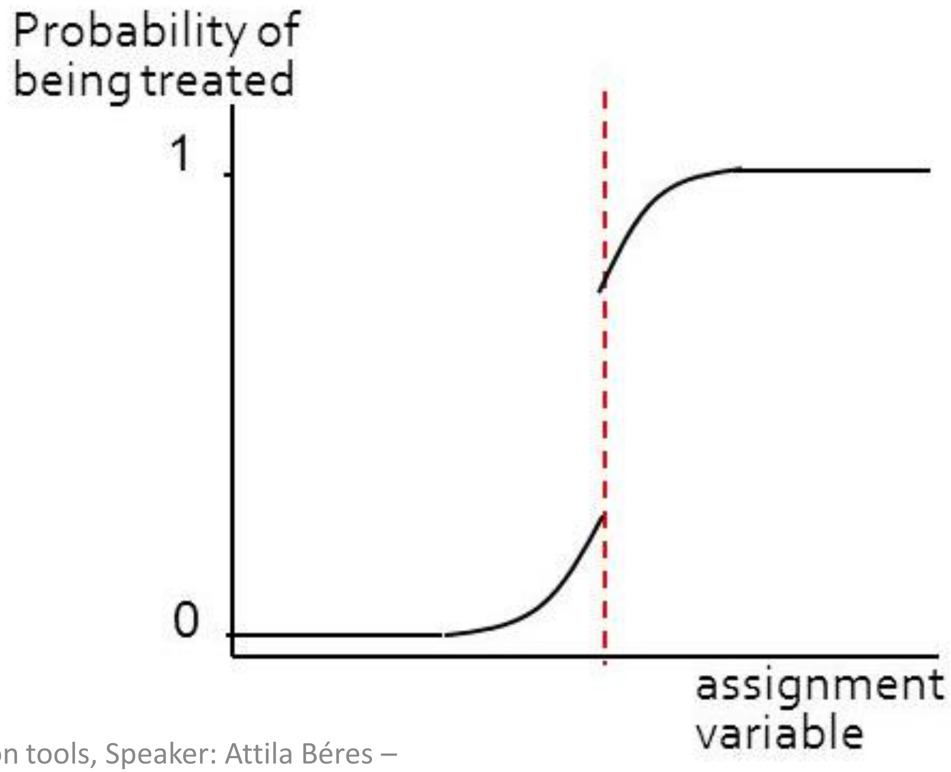
= Eligibility rules are NOT strictly enforces, and further factors affect participation => tehre is some probability of participation below the threshold, but still there is a discontinuous jump around the threshold

Regression Discontinuity Design

SHARP DISCONTINUITY



FUZZY DISCONTINUITY



SUFINANCIRANO SREDSTVIMA EUROPSKE UNIJE
EUROPSKI POLJOPRIVREDNI FOND ZA RURALNI RAZVOJ –
EUROPA ULAŽE U RURALNA PODRUČJA

RDD and/or IV?

- RDD = special IV
  - Exogenous variable: dummy of eligibility
  - It affects participation, but does not directly affect the independent variable, conditional on controlling the continuous effects of the eligibility variable

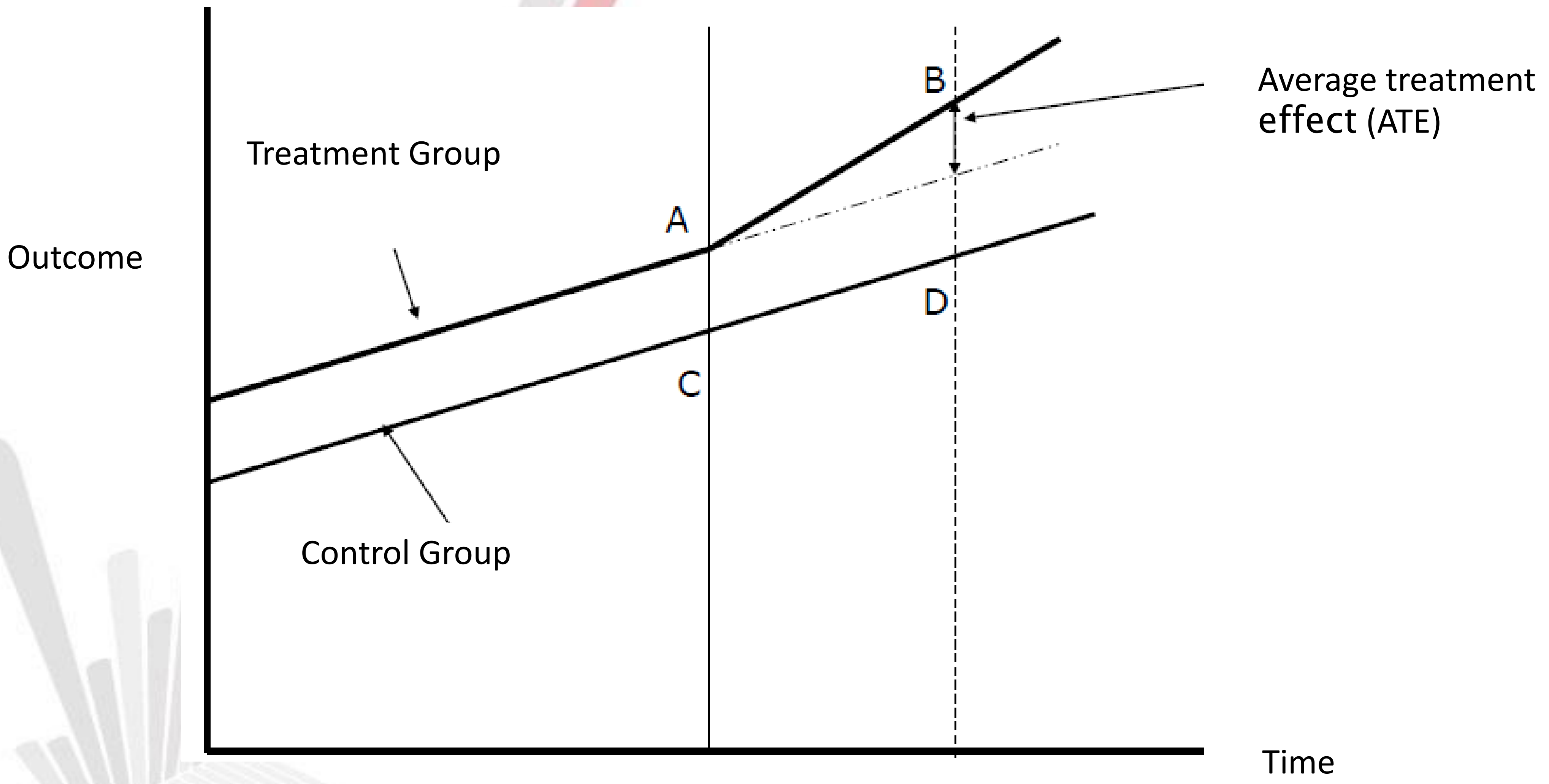
Why?

„The cutoff induces a change in the probability of treatment. If treatment matters, this induces a change in the outcome. Since the treatment doesn't affect all units, the jump at the cutoff in the outcome needs to be rescaled by the jump at the cutoff in the probability of treatment => standard IV"

Regression Discontinuity Design

Cons	Pros
• Treatment effects are local (LATE) • Limits external validity	• Internal validity: some key identifying assumptions can be empirically verified; specifically the absence of other discontinuities • Easy to estimate (like Randomized Treatment-Control), and can be analyzed (and tested) like randomized experiments • Credible causal estimates of treatment effects • Transparency: treatment and outcomes can be illustrated using graphical methods

Difference in differences



Difference in differences

Annual net income of farmers

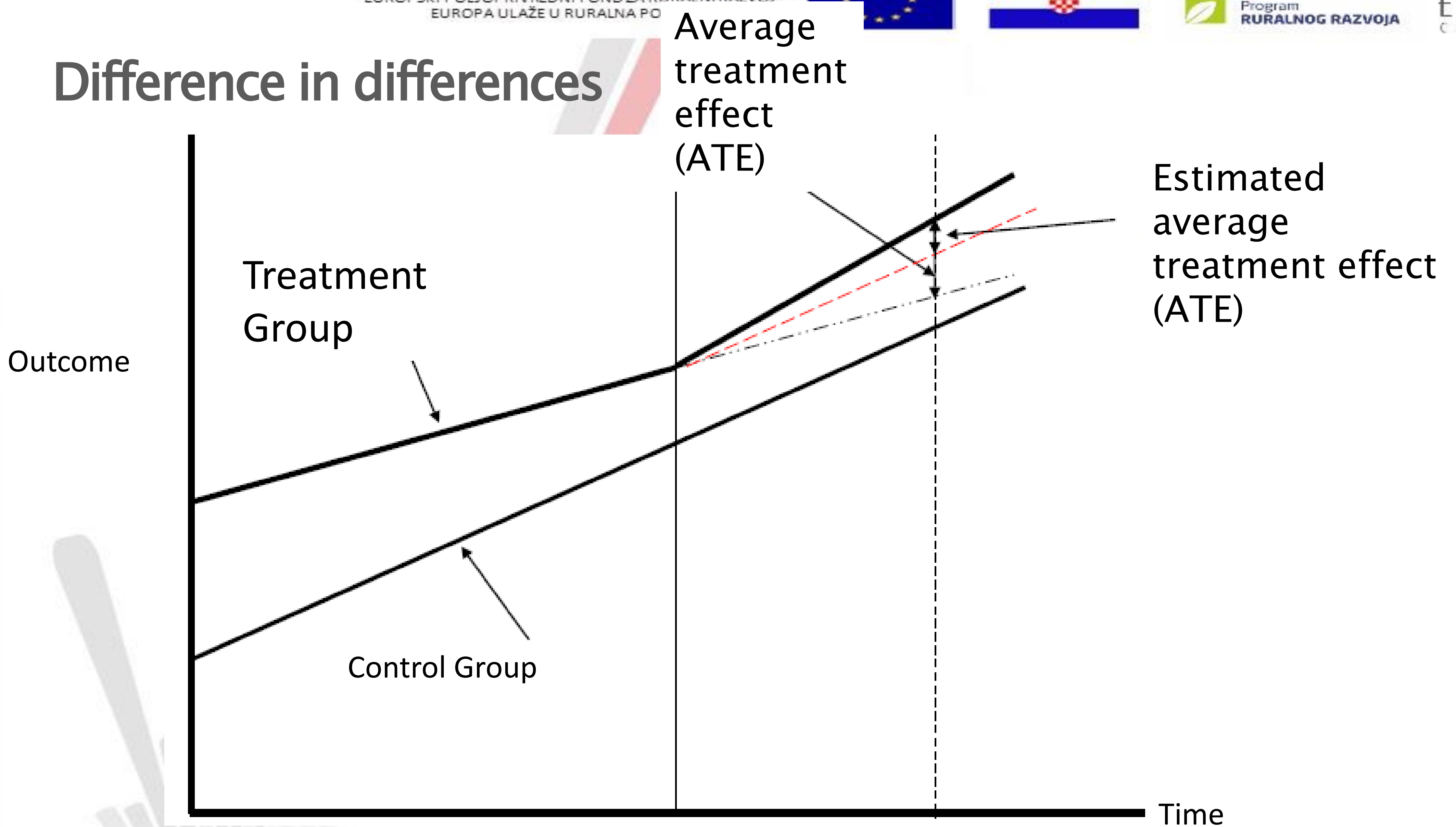
	Treated	Control	
After	65,000	60,000	
Before	50,000	55,000	
Difference	15,000	5,000	10,000

Difference in differences

Annual net income of farmers

	Treated	Control	Difference
After	65,000	60,000	5,000
Before	50,000	55,000	-5,000
			10,000

Difference in differences



Difference in differences

- Fundamental assumption:
- Trends (slopes) are the same in treatments and controls

- Need at least three observations in time
- Two observations “before”
- One observation “after”

SUFINANCIRANO SREDSTVIMA EUROPSKE UNIJE
EUROPSKI POLJOPRIVREDNI FOND ZA RURALNI RAZVOJ –
EUROPA ULAŽE U RURALNA PODRUČJA

Difference in differences

Strongly recommended:

1. White, H. (2006) Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank, World Bank, Washington, D.C. – p.71–82
2. Wooldridge, J. M. (2012). Introductory Econometrics: A Modern Approach. – Chapter 13

Propensity Score Matching

- Idea:
  - For each treated unit
  - Pick up the "best" comparison unit ("match")
  - from a larger survey
- How?
  - Matches are selected on the basis of similarities in observed characteristics
- Issue?
  - If there are differences unobservable characteristics
  - And those unobservables influence participation
  - => selection bias!

Propensity Score Matching

- Controls: non-participants with same observable characteristics as participants
  - In practice, it is very hard.
  - There may be many important characteristics!

- Solution proposed by Rosenbaum and Rubin:

  Match on the basis of the "propensity score":
  - Compute everyone's probability of participating, based on their observable characteristics
  - Choose matches that have the same probability of participation as the treatments

Propensity Score Matching

- Experiment ensures that participation is uncorrelated with:
  - Observable characteristics
  - AND unobservable characteristics
  - => no selection bias
- Matching
  - allows to control for the correlation between participation and observable characteristics
  - but if participation is also correlated with observable characteristics, then we may have SELECTION BIAS

Propensity Score Matching

- Matching ex-post
  - When randomization, RD, other options are not possible
  - Because there is no baseline
  - Be careful: Matching on endogenous variables gives BAD results
- Matching at baseline can be very useful:
  - combine with other techniques (i.e. diff in diff)
  - Know the assignment rule and match based on it
- Matching requires large samples and good data
  - Common support can be a problem

Propensity Score Matching

Strongly recommended:

1. White, H. (2006) Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank, World Bank, Washington, D.C. – p. 53-66

Regression Discontinuity Design: a case study

Title: Measuring the impact of Structural and Cohesion Funds using regression discontinuity design in EU27 in the period 1994–2011

Authors: Pellegrini & Tarola

Aim of the study: to assess the effects of Cohesion Policy (CP) on economic growth in the EU-27 regions receiving financial assistance in the programming periods of 1994-99, 2000-06 and 2007-13.

Methodology: counterfactual method – the regression discontinuity design (RDD)

Regression Discontinuity Design: a case study

Design:

- Cut-off point: allocation rule of regional EU transfer = regions with GDP/cap level below 75% of the EU average receive a huge amount of structural funds transfers
  - Sharp RDD: based on the jump in the probability of EU transfer receipt
- Contrafactual: what would have happened if the policies were not implemented
  - Close to the cut-off point: one can easily put apart any confounding factor by comparing the units belonging to the treated and non-treated groups
- Overall result: their findings show a positive impact of Regional Policy on economic growth

Complex case study in the Czech Republic

Title: EU Cohesion Policy attribution to employment: a case of the Czech Republic

Authors: Potluka & Brůha

Aim of the study: to test whether the European Social Fund (ESF) assistance attributed to employment (sustain jobs or attribute to create them)

- ESF support for training employees in companies
- The aim of the intervention in question is to "increase the adaptability of workers and employers." (HRE OP, p. 107)

Methodology: Counterfactual impact evaluation

Design:

- Companies are divided by size (micro, small, medium and large), regions (NUTS II), sectors of industry, forms of support (grant/system projects) and types of realized training.
- The support from the ESF was not allocated by random selection => selection biases can occur
  - It was necessary to use the quasi experimental methods to measure the effects: use of control group of companies, which did not receive the support

Complex case study in the Czech Republic

Design 1: Instrumental variables (IV)

Theory: Support is determined through intervention not only due to the decision of applicants, but also the processes which are not under their control. Two conditions have to be satisfied for using this method:

1. IV is a significant predictor of the probability of obtaining support
2. IV does not affect the examined indicator any other way than through the support.

IV in this case: strictness of appraisal experts which:

- influences the project approval – different level of „strictness"
- does not influence the final results of the support

Complex case study in the Czech Republic

Design 2: Regression discontinuity design (RDD)

Theory: The estimated impact is based on the comparison of the supported applicants with the unsupported applicants who are close to the selected cut-off point. Applicants on one side of the border are exposed to the intervention and on the other side they are not, even though they are very similar to those supported applicants.

- Cut-off criterium: scoring = points obtained during the appraising process
- Threshold: 65 points to get support – Sharp design

Estimation of Local Effect: The effectiveness of the subsidy was determined by comparing the results of applied companies located just below the threshold and with applied companies located just above the threshold.

Results of the complex case study

- Positive impact on employment in supported large and medium companies
- Negative impact of intervention was found in small companies

- Different methods can be used to test the same question. In this case: „We statistically confirm the impact of the ESF on employment in supported companies only for grant projects by the method of instrumental variables, for each call also by the RDD method."

Scarp Transition Pilot Project – Pakistan

Salinity Control and Reclamation Projects (SCARPs)

Project: In order to solve irrigation problem in Pakistan – SCARPs were launched = close public tubewells and open private ones

- Technically: successful
- Financially: unsustainable burden on the government's budget

Method: Diff-in-Diff based on survey – 391 results in treatment and 100 results in control areas

Findings:

- Success: close of public tubewells without public protest
- Private tubewells grew more rapidly in the control area => no real impact on agricultural productivity or incomes
- Behind this phenomenon: demonstration effect + other factors
- Still: positive rate of return by virtue of the savings in government revenue.

Second Rural Credit Program – Philippines

Second Rural Credit Projects (SRCP)

Project: Crediting small and (medium rice and sugar) farmers purchasing special farm equipment

- Credits were channeled through local banks
- 10% contribution was required from both rural banks and farmers

Method: survey of 738 borrowers, household questionnaires + 47 banks => before-after comparison

Findings:

- The mechanization of farming did not produce an expansion of holding sizes
- No change in cropping intensity was observed, but production and productivity were found to be higher at the end of the project.
- The project increased the demand for both family and hired labor.
- Farmers reported an increase in incomes and savings, and in several other welfare indicators, as a result of the project.

Kurunegala Rural Development Project, and Second Rural Development Project – Sri Lanka

Integrated Rural Development Projects

Project: Multisectoral planning project

- 1. KRDP: focus on paddy and coconut production, productive services (inputs supplies, extension services and credit were also provided)
- 2. SRDP: expanded the approach to two other districts

Method: Using secondary data in project and non-project areas within the same district in order to:

1. construct farm models to calculate the return to the project. With this method: Non-project areas in the same districts benefited from interventions from other donors => returns are under-estimated
2. use before-intervention figures as the counterfactual. With this method: attributing all increases to the project => returns are over-estimated

Kurunegala Rural Development Project, and Second Rural Development Project – Sri Lanka

Integrated Rural Development Projects

Findings:

- Project targets for increased area and productivity were not met. Still, the project did contribute to production increases of the targeted crops, and so higher incomes for beneficiaries.
- Other aspects of the quality of life were not captured by counterfactual analysis. The beneficiary assessment pointed in particular to the benefits from improved roads.

References

Becker, S. O. (2016). Using instrumental variables to establish causality. *IZA World of Labor*, (250).

Deschacht, N., & Goeman, K. (2015). Selection bias in educational issues and the use of Heckman's sample selection model. *Contemporary Economic Perspectives in Education*.

Draganus. Correlation vs Causality – Differences and Examples. https://www.georanker.com/correlation-vs-causality-differences-and-examples

Fondazione Giacomo Brodolini (2015). Evaluating the employment impact of hiring incentives in Italy. https://ec.europa.eu/social/BlobServlet?docId=15040&langId=en

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice*. The World Bank.

Kendall, J. M. (2003). Designing a research project: randomised controlled trials and their principles. *Emergency Medicine Journal*, *20*(2), 164-168.

Khandker, S., B. Koolwal, G., & Samad, H. (2009). *Handbook on impact evaluation: quantitative methods and practices*. The World Bank.

Levitt, Steven D. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." American Economic Review, June 1997, 87(3), pp. 270-90.

White, H. (2006) Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank, World Bank, Washington, D.C.

Wooldridge, J. M. (2012). Introductory Econometrics: A Modern Approach.