

# Principles of Empirical Analysis

## Lecture 9: Difference-in-differences part 2 Regression Discontinuity Design

*Spring 2024*  
*Miri Stryjan*

# Observational alternatives to experiments

- 1. Selection on observables:** treatment and control groups differ from each other only w.r.t. **observable characteristics**, there is no problematic **selection** into treatment
  - Lecture 7 discussed such a situation
- 2. Selection on unobservables:** treatment and control groups differ from each other in **unobservable characteristics**
  - **Treatment and controls are observed before and after treatment – difference-in-differences (DID).** Today we continue discussion from Lecture 8.
  - Selection mechanism is known – **regression discontinuity designs (RDD)**
  - Exogenous variable induces variation in treatment – **instrumental variables (IV)**

# Outline

- 1. DID continued: Last lecture, we discussed the basic idea of difference-in-difference designs: DID with two groups and two time periods**
  - **Today:** More general case with many time periods
  - **Application:** Currie and Walker paper on congestion and infant health
  - Some words on limitations of DID with several time periods and staggered treatment effects.
  - Other examples
- 2. Introduction to Regression discontinuity design (RDD)**

# DID recap

- **Idea:**
  - Even if treated and control groups differ in baseline characteristics, we can use observations on **treatment and control groups before and after the treatment** to estimate a causal effect.
- **Assumptions:**
  - The potential outcomes (not observed) would have **developed in a parallel manner** for both groups in the absence of treatment.
    - *This assumption includes the “Common shocks” assumption: There can be no differential changes over time for the treated and control groups.*
- **Testing for design validity:**
  - **Visualization and testing:** are trends in outcomes parallel before treatment?  
**Discuss:** Is there anything else that could have happened to one group but not the other? (know your institutional setting!)



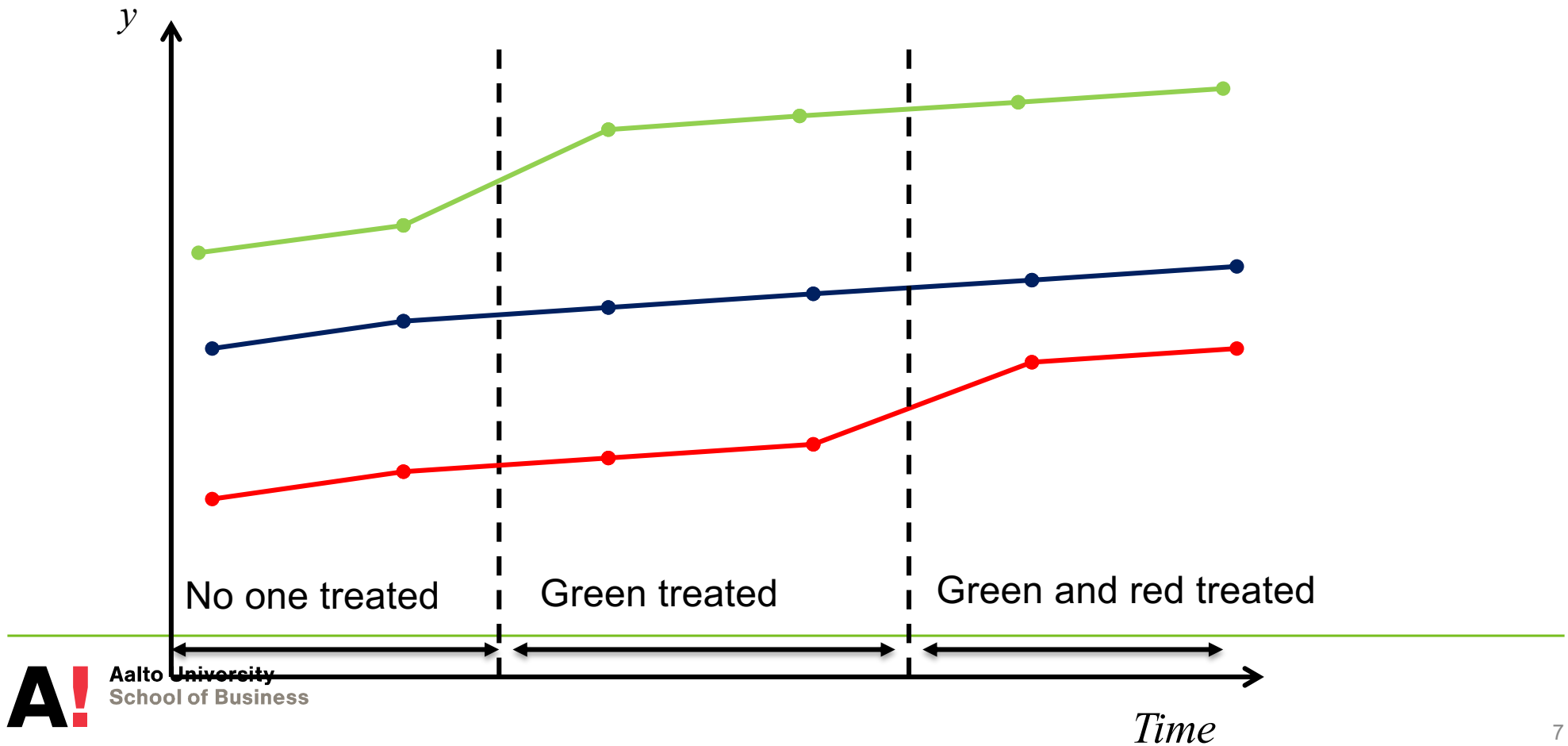
Aalto University  
School of Business

# Staggered or differential treatment timing

# Staggered or differential treatment timing

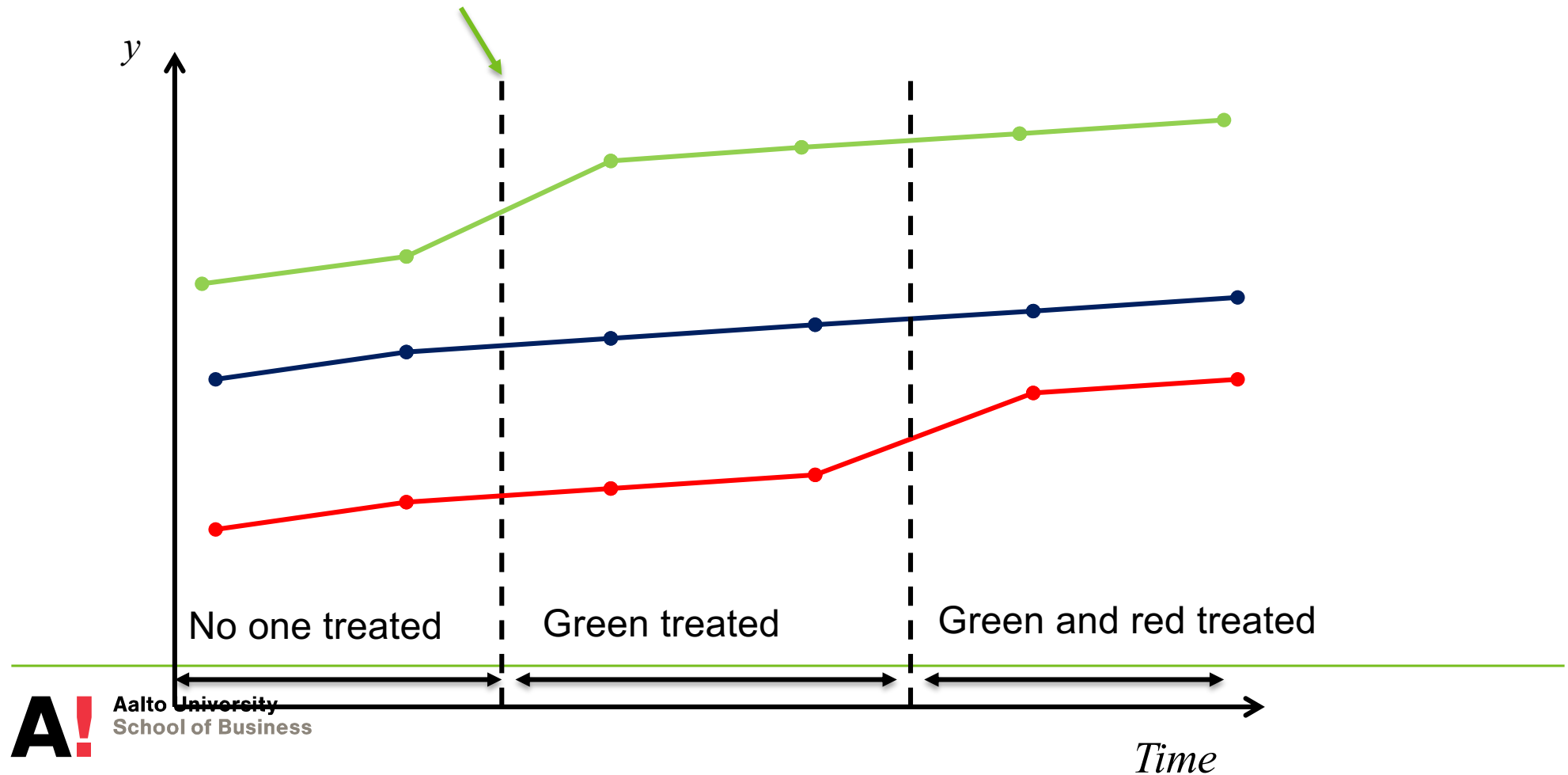
- In our examples thus far, we had **two time periods and two groups**, and the **treatment timing was the same** for all treated units
- Most DID applications, however, exploit variation across groups of units that **receive treatment at different times**
  - In some applications, all units are eventually treated while in others there is a control group that never gets treated

# Staggered timing - example



# Staggered timing - example

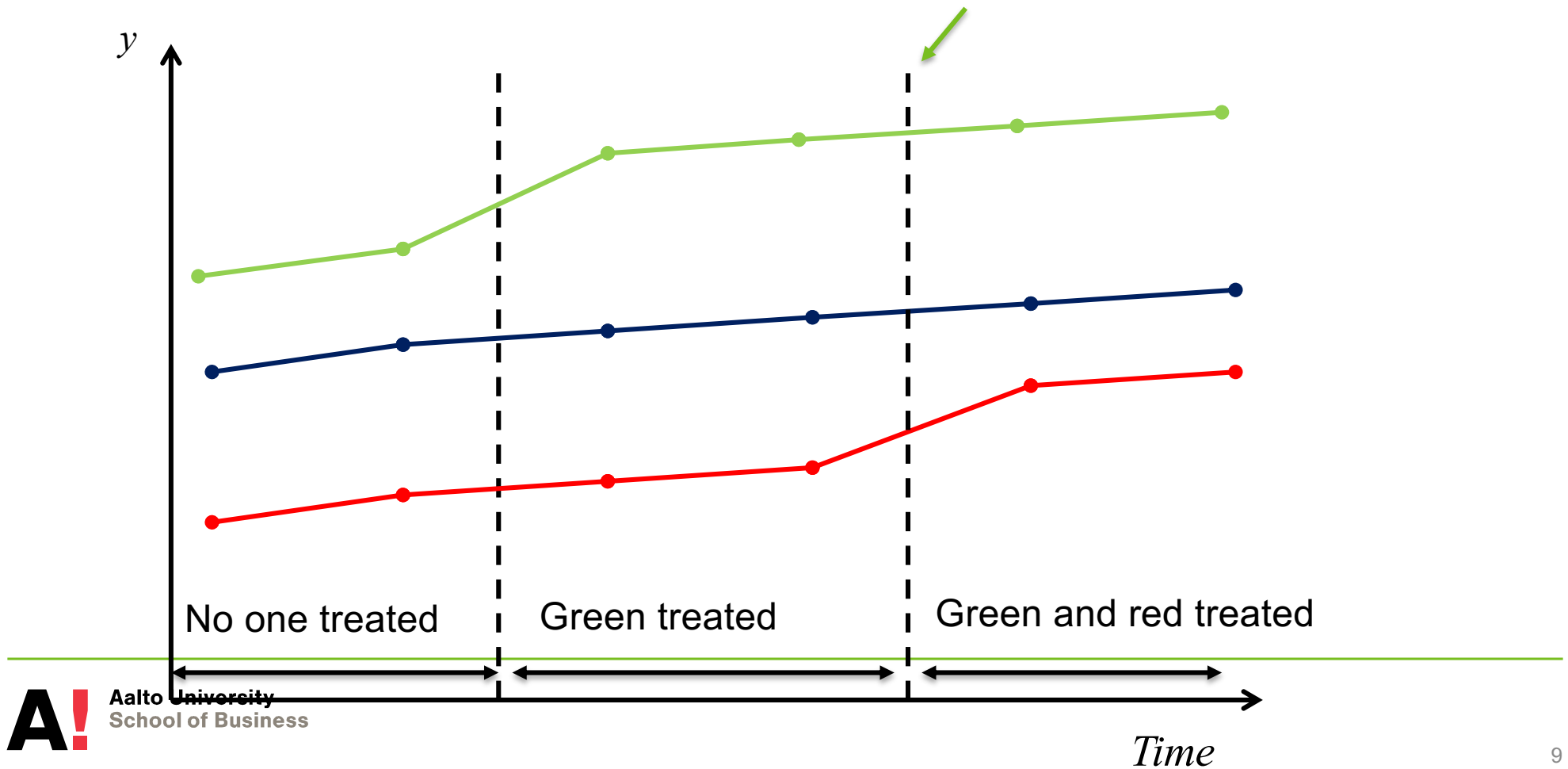
Comparison of early treated group to the control group and the late treated group





# Staggered timing - example

Comparison of late treated group to the control group and the early treated group



## Examples of “treatments” w. staggered implementation

- Effect of more education on later labor market outcomes: **country-wide reform of the education system that is gradually implemented** – late implementation areas can serve as controls for early implementation areas.
- Effect of access to abortion on womens’ and children’s socio-economic outcomes: **gradual legislation of abortion across US states** – late legalization state serve as controls to those where legalization happened earlier.
- Effect of the “green revolution” in agriculture in the 1960’s-70’s: (introduction of new improved seeds for food crops) on crop yields – **different crops had an improved version at different times**. Compare [yield difference] for early improved crops to [difference for] those improved later.

## Traffic Congestion and Infant Health: Evidence from E-ZPass<sup>†</sup>

By JANET CURRIE AND REED WALKER<sup>\*</sup>

*We exploit the introduction of electronic toll collection, (E-ZPass), which greatly reduced both traffic congestion and vehicle emissions near highway toll plazas. We show that the introduction of E-ZPass reduced prematurity and low birth weight among mothers within 2 kilometers (km) of a toll plaza by 10.8 percent and 11.8 percent, respectively, relative to mothers 2–10 km from a toll plaza. There were no immediate changes in the characteristics of mothers or in housing prices near toll plazas that could explain these changes. The results are robust to many changes in specification and suggest that traffic congestion contributes significantly to poor health among infants. (JEL I12, J13, Q51, Q53, R41)*

# Research question

How does a reduction in air pollution affect the health of infants? The “bigger” question is how does pollution affect health.

- Why is this interesting?
  1. There is increasing evidence of the long-term effects of **poor health at birth** on **future outcomes**
  2. There is also evidence that exposure to **polluted air** affects **health negatively**
  3. The study of **newborns** overcomes several difficulties in making the connection between pollution and health because the link between cause and effect is immediate

# Selection bias

T = being more vs less exposed to pollution

Y = premature birth and low birth weight

Since air pollution is **not randomly assigned**, studies comparing health outcomes for populations exposed to differing pollution levels may not adequately control for confounding determinants of health

- Families with higher incomes or preferences for cleaner air are likely to **sort into locations** with better air quality, and failure to account for this sorting **overestimates** of the effects of pollution
- Alternatively, pollution levels are higher in urban areas where there are often more educated individuals with better access to health care, which can cause **underestimates** of the true effects of pollution on health

# Natural/Policy experiment

**Studies the effect of E-Zpass in New Jersey and Pennsylvania, which led to sharp reductions in local traffic congestion, on the health of infants born to mothers living near toll plazas.**

**T = living near a Toll plaza**

**(used as a *proxy*\* for reduced air pollution)**

**Y = premature birth and low birth weight**

E-ZPass is an interesting policy experiment because, while pollution control was an important reason that policy makers implemented the pass, the main reasons that consumers signed up for E-ZPass was to reduce travel time.

*\*proxy = an (imperfect) measure of*

# Empirical strategy

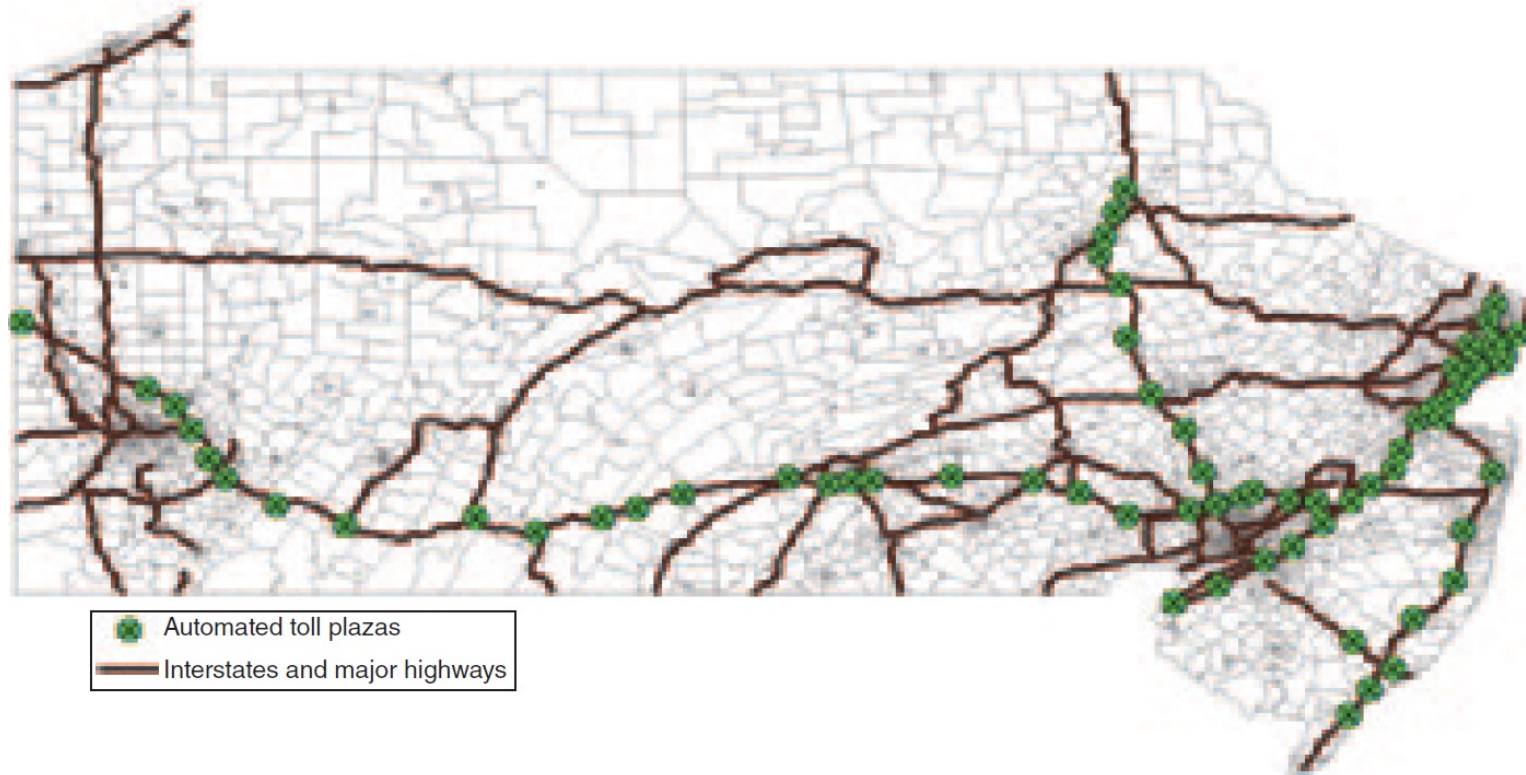


FIGURE 1. LOCATIONS OF TOLL PLAZAS AND MAJOR ROADWAYS IN NEW JERSEY AND PENNSYLVANIA

# Empirical strategy

- “we compare mothers within 2 km of a toll plaza to mothers who are between 2 km and 10 km from a toll plaza, but still within 3 km, of a major highway before and after the adoption of E-ZPass in New Jersey and Pennsylvania.”
- **Assumption:**
  - “Our difference in differences research design relies on the **assumption** that the characteristics of mothers near a toll plaza change over time in a way that is comparable to those of other mothers who live further away from a plaza, but still close to a major highway.”



# Main results

Premature birth by day before and after E-ZPass

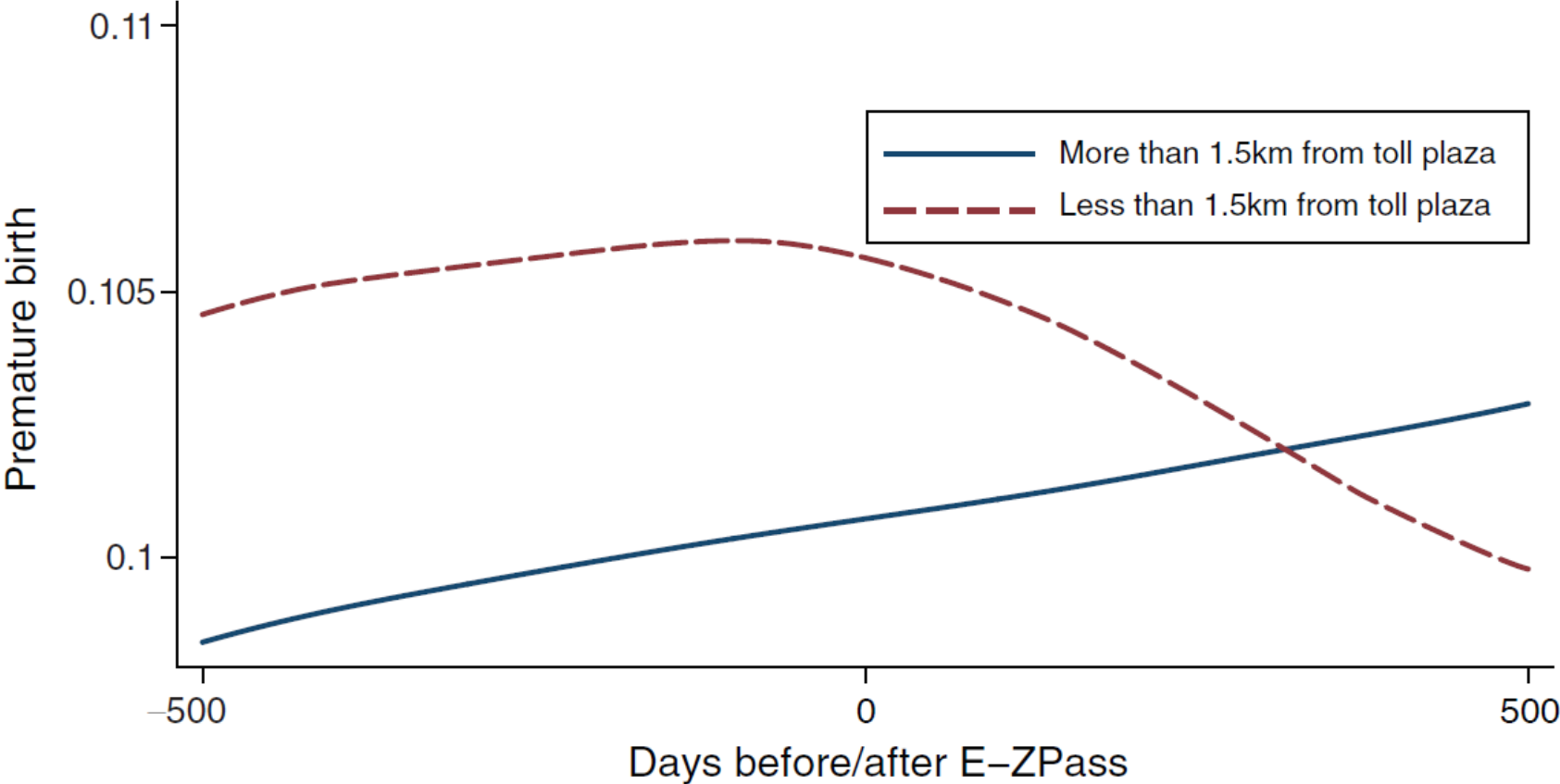


FIGURE 6

# Main results

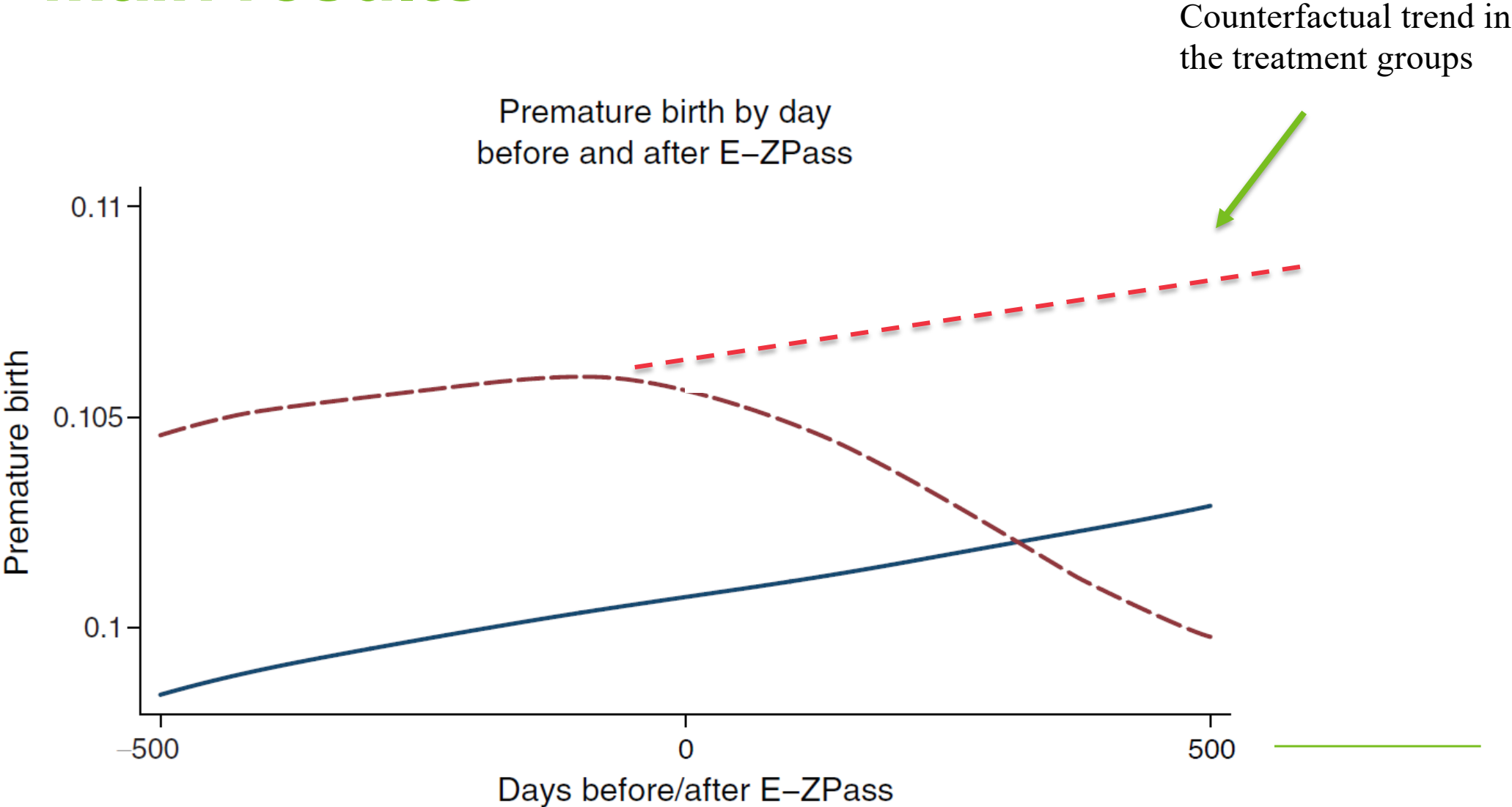


FIGURE 6

# Main results

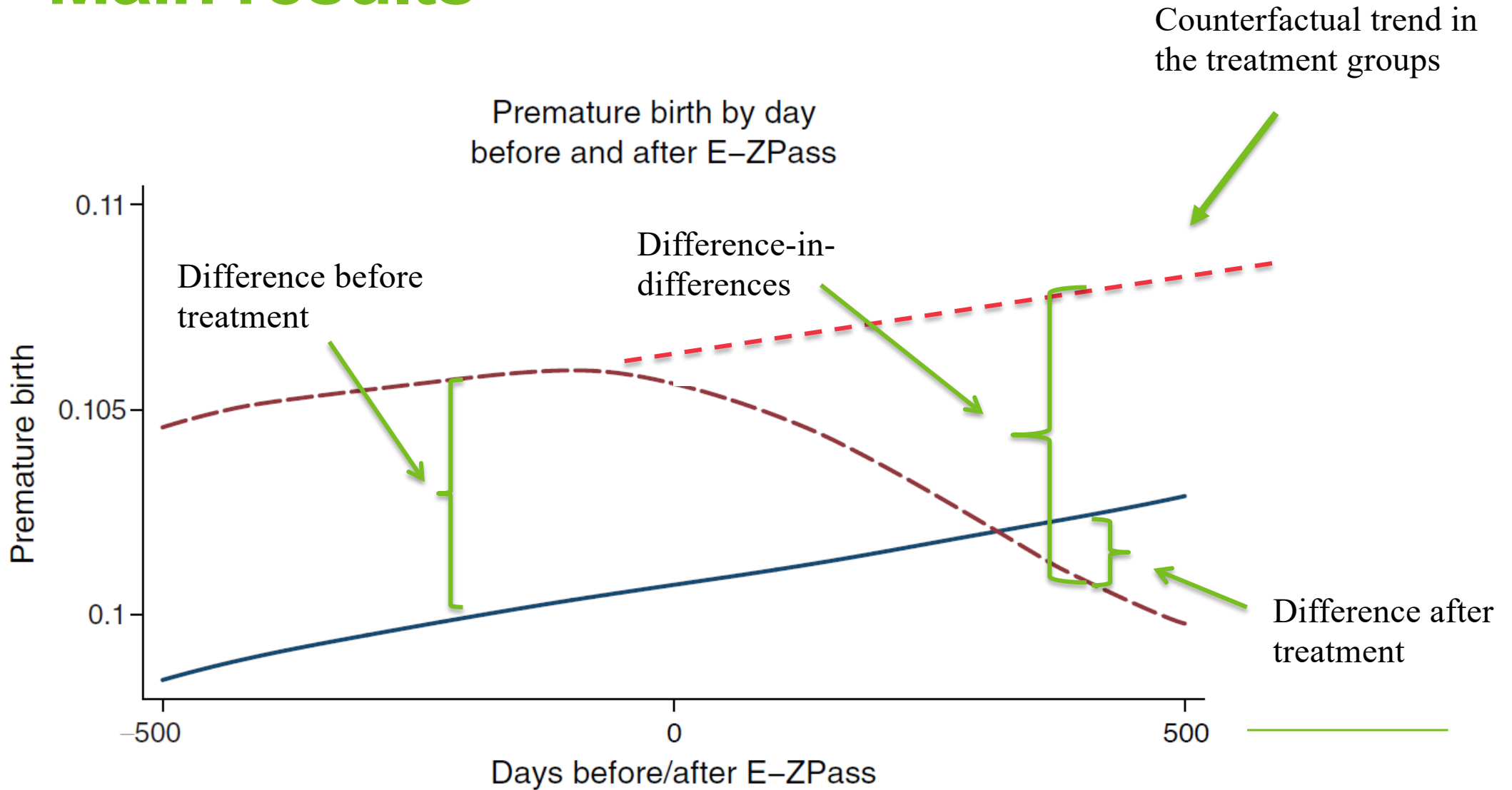


FIGURE 6

# Conclusions – Currie & Reed (2011)

- **The E-ZPass reduced the incidence of prematurity and low birth weight in the vicinity of toll plazas by 6.7–9.1 percent and 8.5–11.3 percent, respectively**
  - These are large but not implausible effects given the correlations between proximity to traffic and birth outcomes found in previous studies
- **Take away: policies intended to curb traffic congestion can have significant health benefits for local populations (in addition to the more often cited benefits in terms of reducing travel costs)**

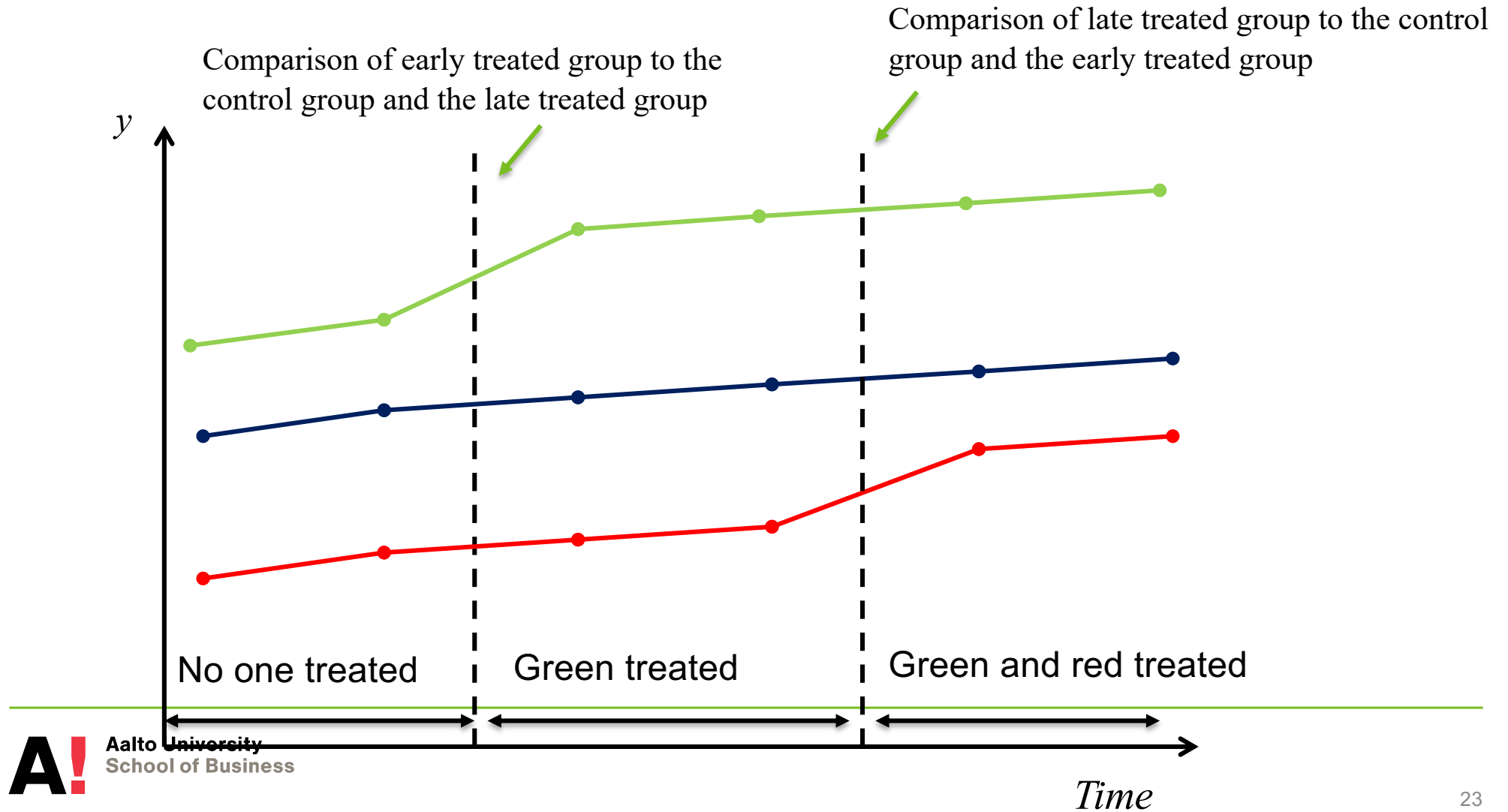
# Regression for multi-period DID\*

- Question last time: **Why is regression better than just comparing means?**
- For a simple 2X2 example as we saw in Lecture 8 (Card and Krueger, 1994) the difference is not that big.
- Regression provides both an estimate and standard error, and calculates the p-value (but it is of course possible to get similar information also with means and standard deviations for a known distribution).
- **Regression is easier to generalize to more demanding setups, such as with DID analysis with many periods and multiple comparisons.**

# DID with differential treatment times

- **Most DID applications exploit variation across groups of units that receive treatment at different times**
  - In some applications, all units are eventually treated, e.g. a rollout of a national policy
  - In others there is a control group that never gets treated
  - Rollout is a great way to make sure we learn something about the effects of the policy
- **In principle this is all fine**
  - But there are complicated issues concerning staggered designs and the literature is moving forward on this

# Problems with staggered T: Recall example



# Problems with staggered treatment

In the example on previous page, the Green group is both part of the control group and part of the treated group.

More generally, in DID with staggered treatment, the analysis consists of multiple 2X2 comparisons around the time windows when a unit is treated.

If treatment effects are heterogenous, i.e. differ over time, analyzing staggered treatments with “regular” DID methods (Two way fixed effects) can give biased results, and lead to both Type I and Type II errors (i.e. we do not know in what direction the bias goes!)

There is currently an active discussion and new method versions have been developed that can address these challenges.



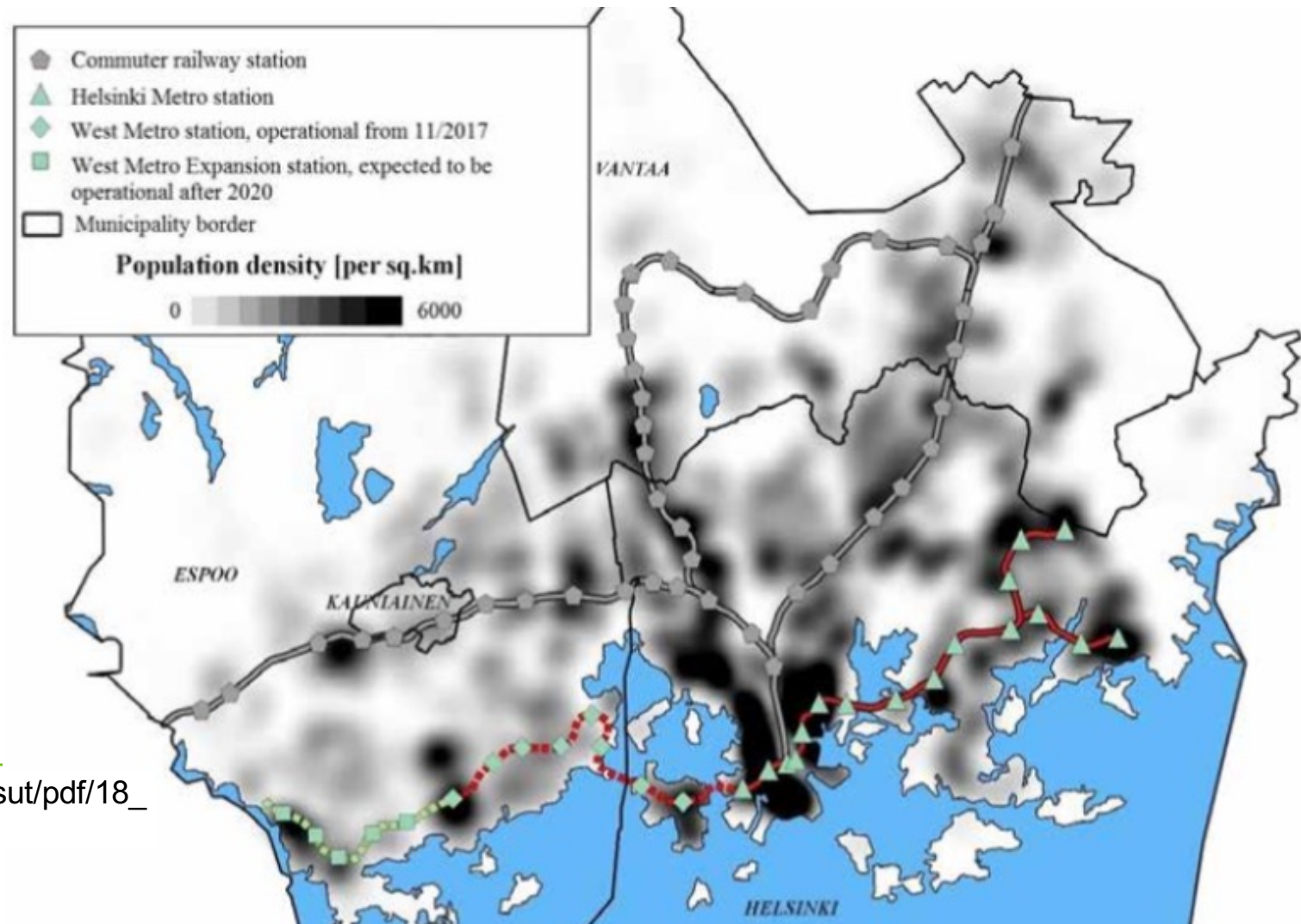


Aalto University  
School of Business

# Other Applications



# West Metro extension in the HMA



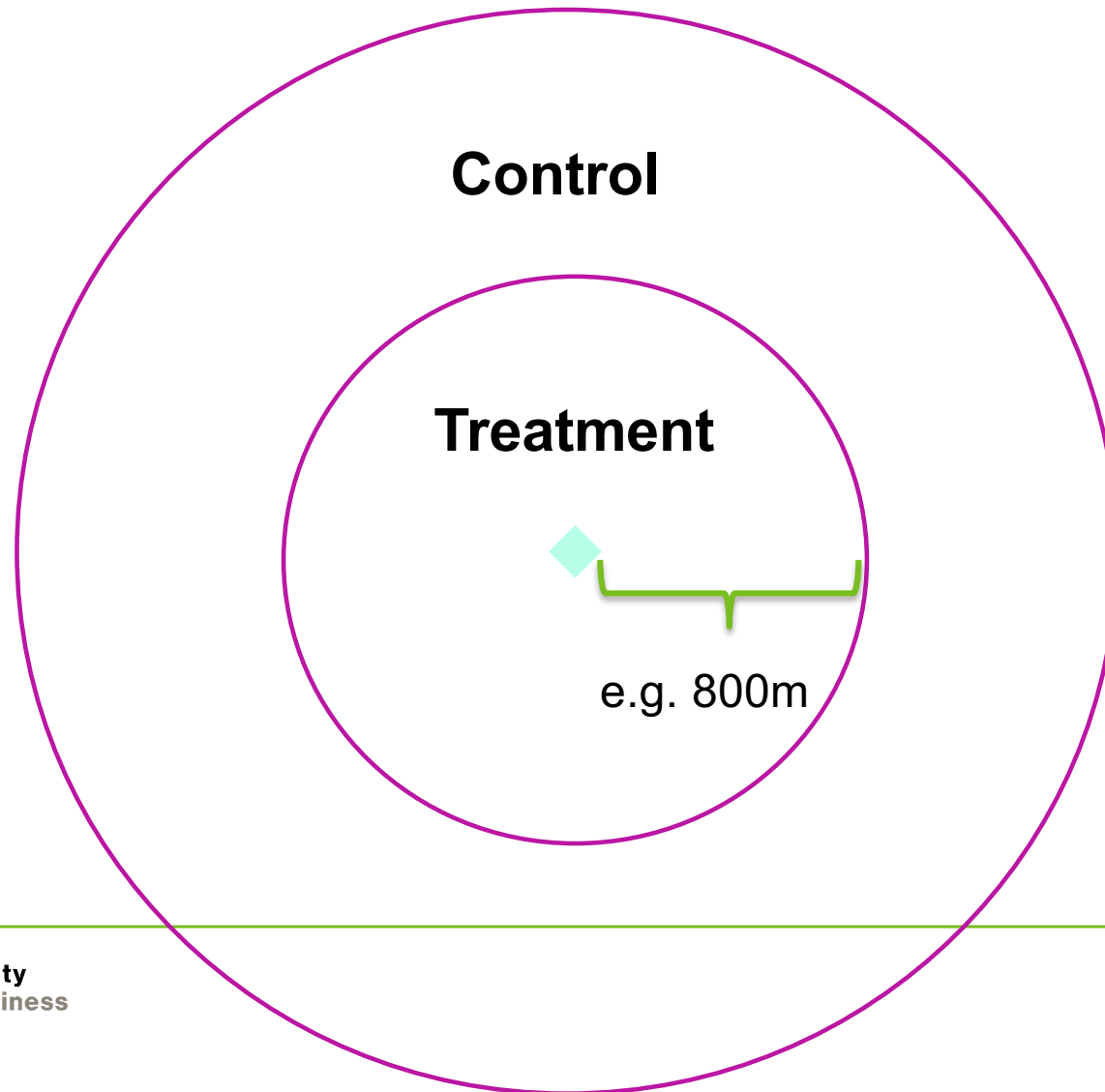
[https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/18\\_01\\_25\\_tyopapereita\\_02\\_Harjunen.pdf](https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/18_01_25_tyopapereita_02_Harjunen.pdf)

# Harjunen 2018 measures “anticipation effects”

“The estimated completion date and projected construction costs of the West Metro were adjusted during the construction period. Before **the tunnel works of the first stage of the West Metro began in 2009**, the aim was that the metro would start operating in fall 2014 and the budget was 714 million euros (index corrected budget for 2016 was 849 million euros). However, this estimate of the opening was year later postponed to 2015 and later on to 2016. Finally, the opening date was announced to be 15th of august 2016”

The date of treatment is in 2009 when the concrete plans of the metro were initiated, and expectations were that the metro stations would indeed open.

# DID design

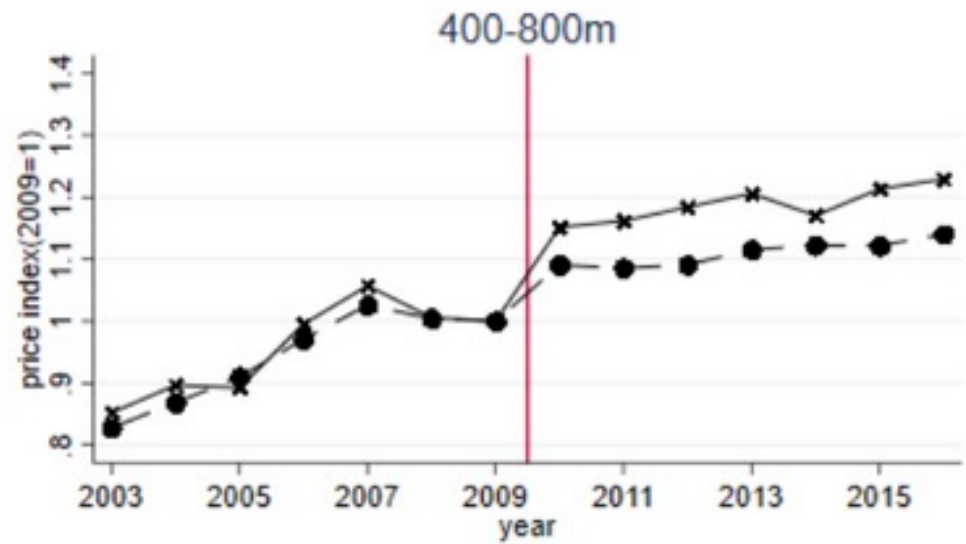
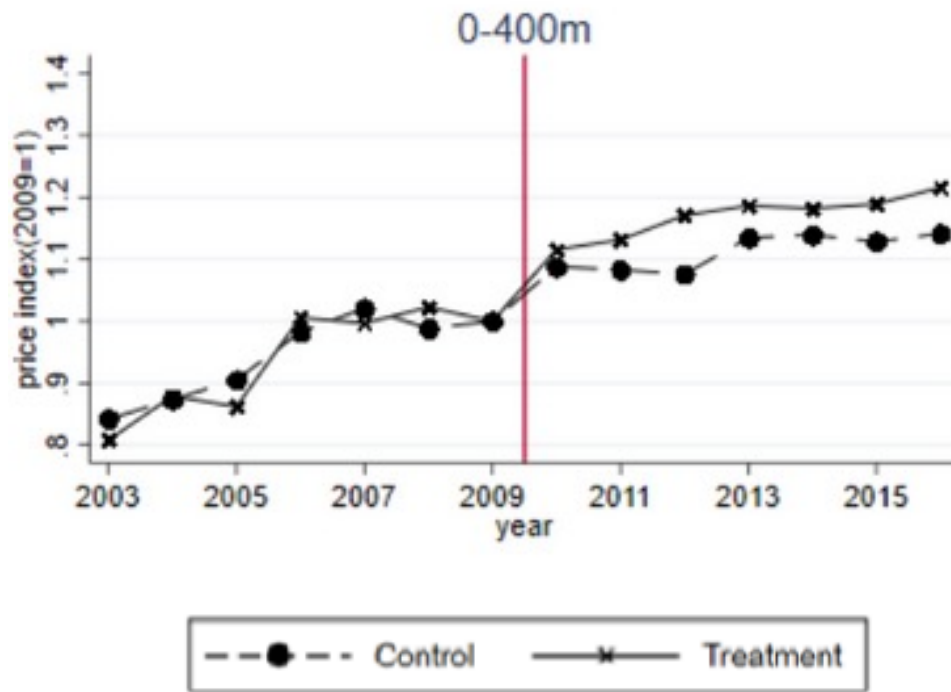


◆ New station

# Data

Sample	Whole data (Helsinki and Espoo)		0 to 800m		800 to 1 600m	
	Status		Treated	Control	Treated	Control
<b>N</b>		<b>43 025</b>	<b>6 868</b>	<b>15 640</b>	<b>4 429</b>	<b>11 267</b>
<b>Sale price</b>		223 668 [110 007]	252 024 [119 458]	196 154 [78 980]	311 661 [156 343]	199 122 [82 107]
<b>Square price</b>		3 506 [918]	4 181 [951]	3 325 [805]	3 877 [919]	3 242 [805]
<b>Area</b>		66 [29]	62 [27]	61 [25]	82 [38]	64 [27]
<b>Age</b>		37 [17]	43 [17]	32 [17]	32 [13]	39 [18]
<b>Maint. Charge (€/m<sup>2</sup>)</b>		3,5 [1.2]	3,8 [1.1]	3,5 [1.2]	3,5 [1.2]	3,5 [1.3]
<b>Floor number</b>		2,4 [1.6]	2,7 [1.7]	2,5 [1.5]	2,3 [1.5]	2,3 [1.4]
<b>Floors in building</b>		3,8 [3.0]	4,4 [2.2]	3,8 [2.1]	3,6 [2.3]	3,4 [1.9]
<b>Dist. to nearest station (m)</b>		869 [489]	482 [190]	484 [185]	1 168 [239]	1 134 [239]
<b>Dist to CBD (km)</b>		12 [4.6]	9 [3.6]	13 [4.8]	11,2 [3.2]	12,5 [4.6]

# Results



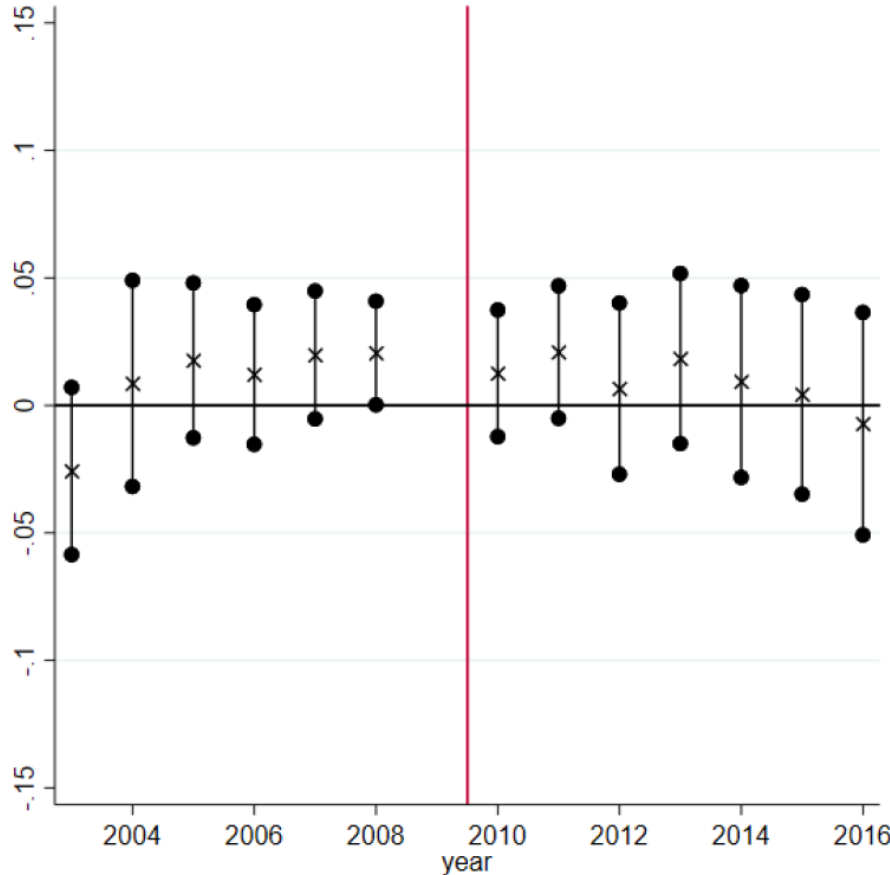
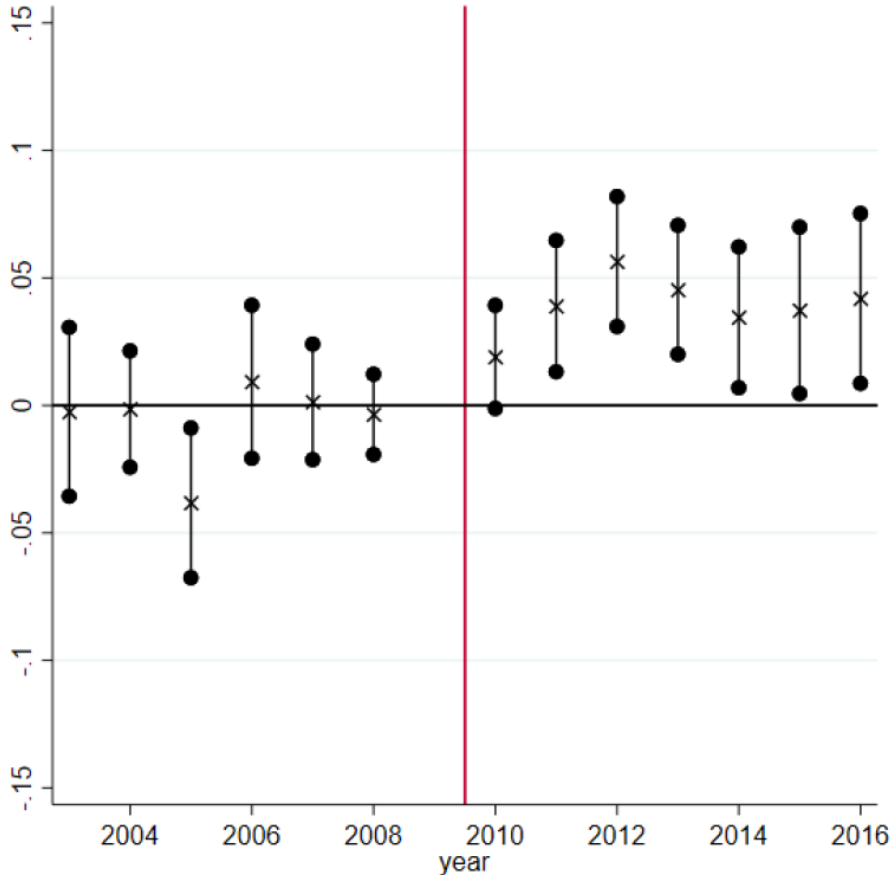


# Results



0-800m

800-1 600m





# Results

Table 1. Estimation results, average treatment effect during 2010 to 2016

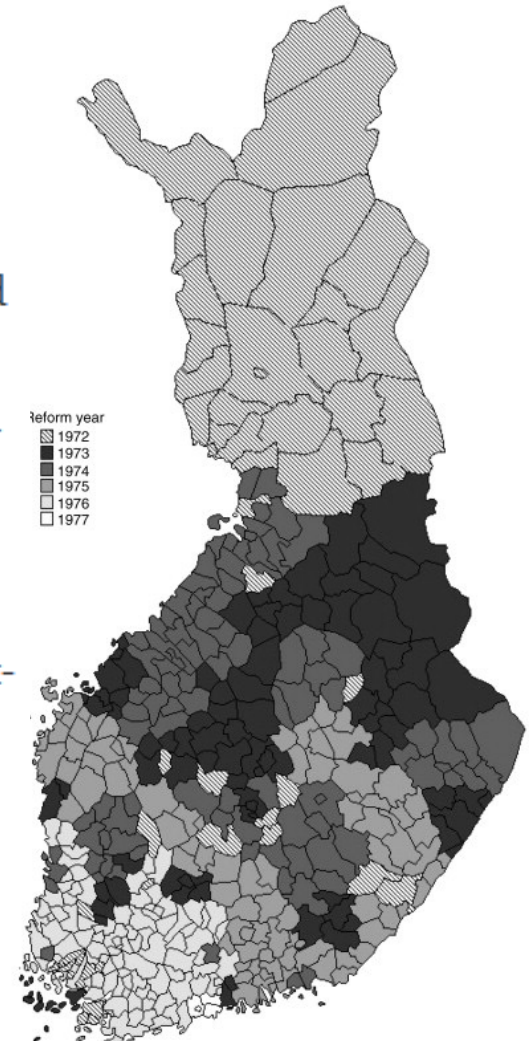
Dependent variable: ln(sale price)						
	(1)	(2)	(3)	(4)	(5)	(6)
Distance band	0-400m	400-800m	800-1 200m	1 200-1 600m	0-800m	800-1 600m
<b>treated</b>	<b>0.093***</b>	<b>0.136***</b>	<b>0.166***</b>	<b>0.159***</b>	<b>0.130***</b>	<b>0.161***</b>
	[0.026]	[0.023]	[0.026]	[0.032]	[0.022]	[0.024]
<b>treated*after</b>	<b>0.042***</b>	<b>0.036***</b>	<b>-0.010</b>	<b>0.014</b>	<b>0.040***</b>	<b>-0.001</b>
	[0.011]	[0.013]	[0.015]	[0.018]	[0.011]	[0.014]
<b>R-squared</b>	0,87	0,88	0,89	0,90	0,88	0,89
<b>N</b>	7 759	14 749	9 500	6 196	22 508	15 696
<b># clusters</b>	92	123	118	111	135	151

Notes: Estimated coefficient is statistically significant at \*\*\* 1% level, \*\* 5% level, \* 10% level. Standard errors are clustered by small city districts. House characteristics include all reported in table A4 in the Appendix III (area and age also in second power)

# Rollout implementation of a policy – Pekkarinen et al. (2009)

## Abstract

This paper estimates the effect of a major education reform on intergenerational income mobility. The Finnish comprehensive school reform of 1972–1977 replaced the old two-track school system with a uniform nine-year comprehensive school and shifted the selection of students to vocational and academic tracks from age 11 to age 16. We estimate the effect of this reform on the intergenerational income elasticity using a representative sample of males born between 1960 and 1966. The identification strategy relies on a differences-in-differences approach and exploits the fact that the reform was implemented gradually across the country during a six-year period. The results indicate that the reform reduced the intergenerational income elasticity by 23% from the pre-reform elasticity of 0.30 to post-reform elasticity of 0.23.



Aalto University  
School of Business

<https://www.sciencedirect.com/science/article/pii/S0047272709000619>

# DID recap

- **Idea:**
  - Even if treated and control groups differ in baseline characteristics, we can use observations on **treatment and control groups before and after the treatment** to estimate a causal effect.
- **Assumptions:**
  - The potential outcomes (not observed) would have **developed in a parallel manner** for both groups in the absence of treatment.
    - *This assumption includes the “Common shocks” assumption: There can be no differential changes over time for the treated and control groups.*
- **Testing for design validity:**
  - **Visualization and testing:** are trends in outcomes parallel before treatment?  
**Discuss:** Is there anything else that could have happened to one group but not the other? (know your institutional setting!)

# Regression discontinuity design (RDD)



# Observational alternatives to experiments

- 1. Selection on observables: treatment and control groups differ from each other only w.r.t. observable characteristics**
  - Lecture 8
- 2. Selection on unobservables: treatment and control groups differ from each other in unobservable characteristics**
  - Treatment and controls are observed before and after treatment – **difference-in-differences (DID)** Lecture 8-9
  - Selection mechanism is known – **regression discontinuity designs (RDD)**
  - Exogenous variable induces variation in treatment – **instrumental variables (IV)**

# Outline

- **Basic idea of regression discontinuity designs**
  - Setup and assumptions
  - Fuzzy and sharp RDD
- **Testing RDD assumptions**
  - Manipulation, covariate balance, fake cutoff placebos and other placebos

# RDD

- **Introduced by Thistlethwaite and Cambell (1960)**
  - Studied the impact of merit awards on future academic outcomes, where merit award was given if test score exceeds a cutoff
  - Idea: students can (of course) affect their test scores by studying, but they cannot manipulate their scores to be just above the cutoff because the cutoff is unknown to them *ex ante*
- **Reappeared and formalized in economics in late 90s and has proven to be a powerful causal tool in empirical economics and other disciplines**
  - Political science, education, epidemiology, criminology etc.
- **Strong internal validity, but very data intensive**
  - Need to have a lot of observations near the cutoff

# Estimating causality and RDD

**As usual, we start with some causal relationship in mind.**

We would like to estimate the **effect of some treatment, T on some outcome, Y.**

But we suspect that there is selection into treatment: treatment status T is correlated with unobservables that may also affect Y.

**RDD can be used to isolate the causal effect of treatment in certain situations where Individuals (or units) become treated after crossing some **arbitrary cutoff.****



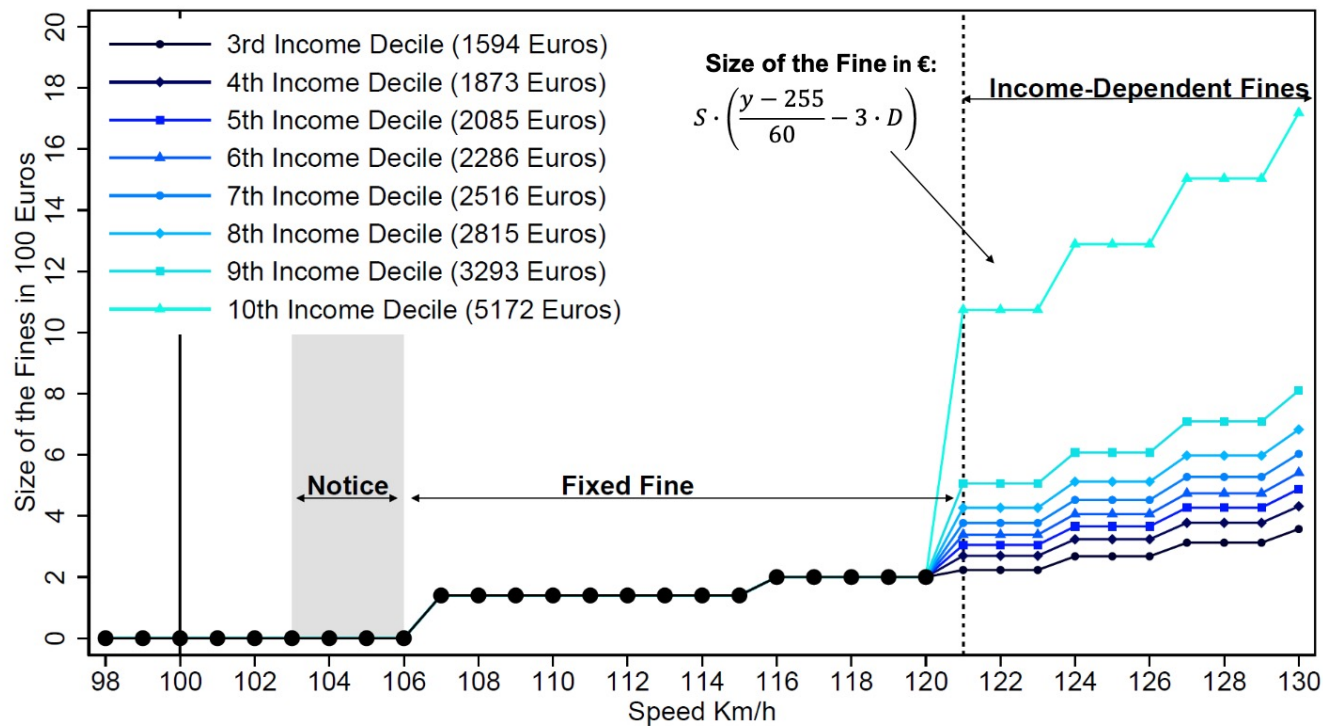
# Speeding tickets in Finland

(Martti Kaila, 2023)

“In Finland, speeding tickets become income-dependent if the driver’s speed exceeds the speeding limit by more than 20 km/h, leading to a substantial jump in the size of the fine.

**Figure 1:** Relationship Between Income and the Size of Speeding Ticket

(a) Theoretical example when the speeding limit is 100 Km/H



# RDD – the setup

- RDD has three fundamental components: **running variable**, **cut-off**, and **treatment**
- Individuals become treated after crossing some cutoff in the running (or forcing or score) variable
  - **Sharp RDD**: treatment received with probability zero below the cut-off (or threshold) and probability one above cut-off
  - **Fuzzy RDD**: The probability of receiving the treatment increases discontinuously at the threshold (**imperfect compliance**)
- **Assumption**: the potential outcomes evolve smoothly across the cutoff. In other words:
  - If there is **no precise manipulation** of the running variable, observations just below the threshold are very similar to those just above the threshold and therefore constitute **a valid control group**.

# Examples of running variables and cutoffs that can be used for RDD

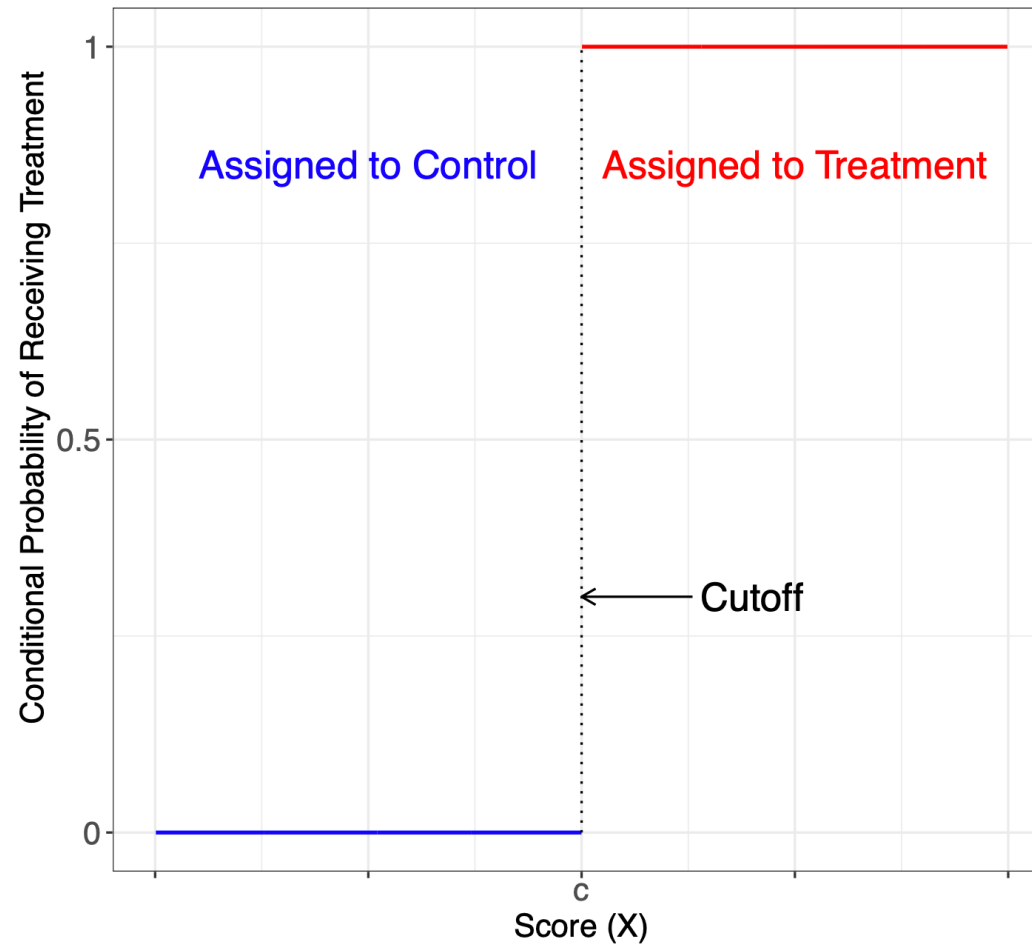
- **Test scores:** entry to high school/university depends on some test score/GPA
- **Age:** after some age (e.g. 18, 21), you become eligible to do something (vote, buy alcohol in US)
- **Geography:** access to services based on residential location and catchment areas; coordinates or distance to some boundary/border determines treatment
- **Elections:** candidate's vote share (running variable) determines election status (treatment)
- **Speed:** in Finland, after exceeding speed limit by more than 20 km, you get a different type of fine.

# Sharp RDD

For the following graphs, let's think of the example of **test scores**: admission to a course or school depends on some test score/GPA

- **Running variable**: test score
- **Cutoff**: a threshold score required for passing the exam, e.g. 50/100.
- **Treatment**: attending that course/school

# Sharp RDD



Source: Cattaneo et al. (2019): A Practical Introduction to Regression Discontinuity Designs: Foundations.

# The problem of causal inference in RDD

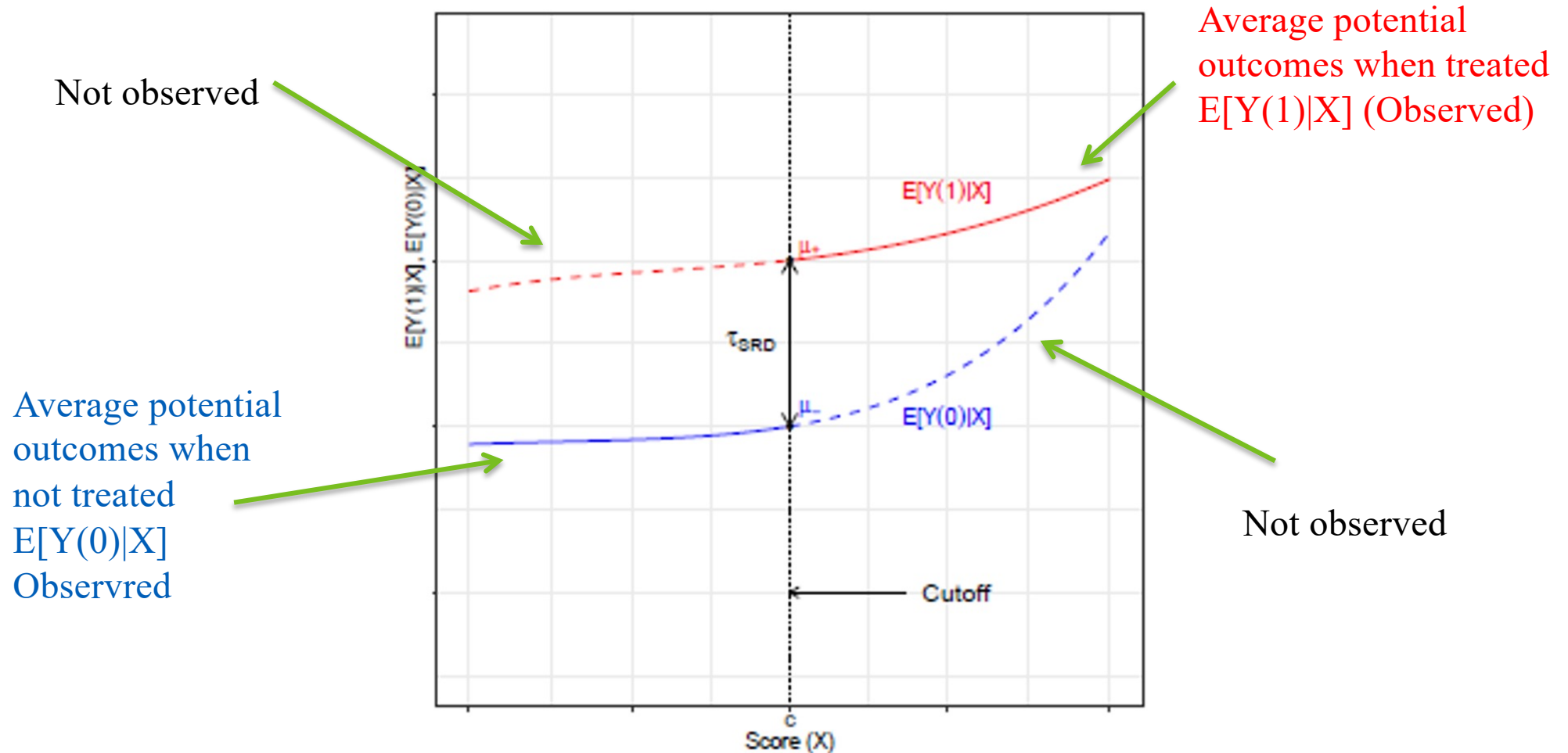
”The fundamental problem of causal inference [in RDD] occurs because we only observe the outcome under control,  $Y_i(0)$ , for those units whose score is below the cut-off, and we only observe the outcome under treatment,  $Y_i(1)$ , for those units whose score is above the cut-off”

Source: Cattaneo et al. (2019): A Practical Introduction to Regression Discontinuity Designs: Foundations.

In the example of test scores that determine entry to education: we never observe two people with the same score, but where one was accepted and one was rejected. --> There is **no common support in scores between the accepted and rejected group**.

The potential outcomes (based on pre treatment characteristics such as ability and motivation) are likely to be different between those who score high and those who score low.

# Potential outcomes for those with high scores

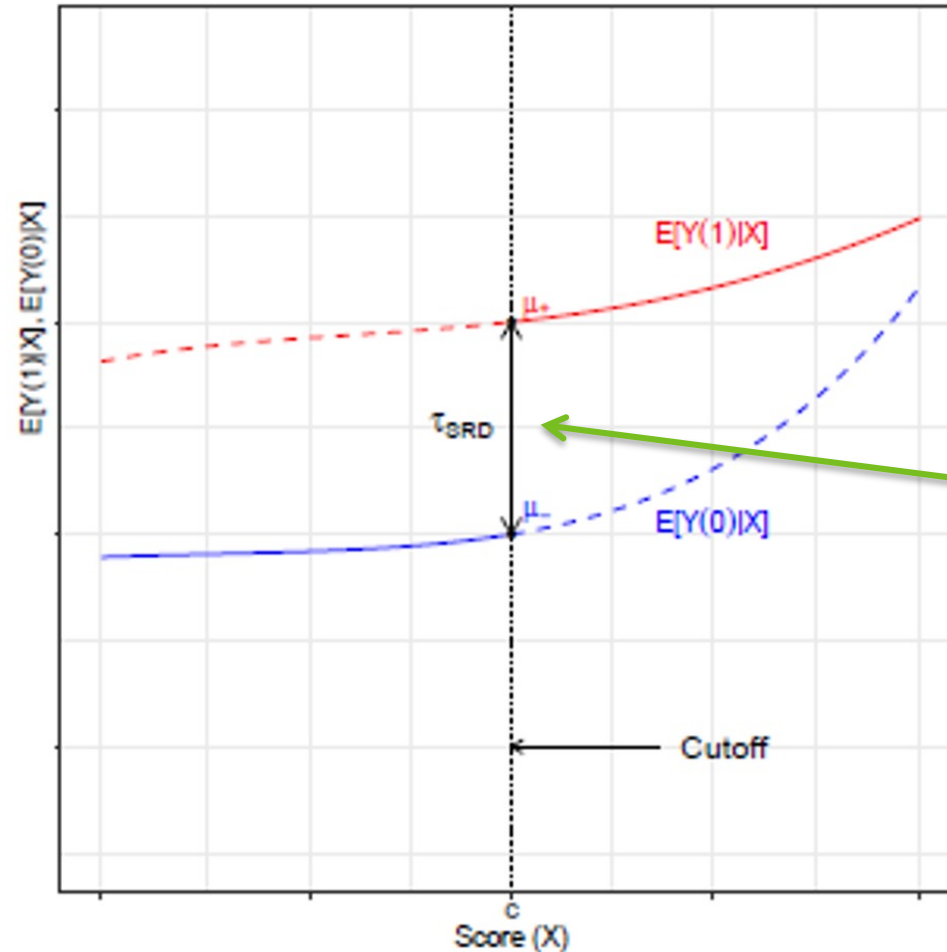


(b) RD Treatment Effect

# Local causal effect

If units CANNOT perfectly “sort” around the cutoff, the discontinuous change in the probability of treatment can still be used to learn about the **local causal effect** of the treatment

Units with scores barely below the cutoff can be used as a control group for units with scores barely above it



Local causal effect

$$\tau_{\text{SRD}} \equiv \mathbb{E}[Y_i(1) - Y_i(0) | X_i = c]$$

(b) RD Treatment Effect



# Notes to figures on p 12-13

Source: Cattaneo et al. (2019): A Practical Introduction to Regression Discontinuity Designs: Foundations.

The Sharp RD design exhibits an extreme case of lack of common support, as units in the control and treatment groups **cannot** have the same value of the running variable ( $X_i$ ). This feature makes RD designs different from other non- experimental settings.

→ RD analysis fundamentally relies on **extrapolation** towards the cutoff point in order to compare control and treatment units.

As shown in the Figures, the average treatment effect at a given value of the score,  $E[Y_i(1) | X_i = x] - E[Y_i(0) | X_i = x]$ , is the vertical distance between the two regression curves at that value. **This distance cannot be directly estimated because we never observe both curves for the same value of x.**

# Key point and assumption of RDD

Units that receive very similar score values on opposite sides of the cut-off are comparable to each other in all relevant aspects, except for their treatment status.

There are different ways of thinking of what this “comparability” means formally

1. in a small neighborhood around the cut-off we obtain conditions that mimic a randomized experiment, meaning that **units on each side of the cut-off are as good as randomly assigned to either receive treatment or not** (local randomization framework)
2. There is a continuity of average potential outcomes near the cut-off (continuity-based framework). This is a conceptually more difficult to think about since potential outcomes cannot be observed.