Statistics for impact analysis

Differences in differences (DID) & Regression discontinuity designs (RDD)

Antoine Leblois

INRAE (Institut Agro / *ex. Supagro*, campus la Gaillarde)
antoine.leblois@inrae.fr
http://antoine.leblois.free.fr

UM, Richter
12/2022

How far have you got ?
Some stats / empirists jargon

Correlation vs. causality / Counterfactual / credibility revolution

Statistical significance (standard-error and confidence interval), power of a test

Observables vs. unobservables / mean/avg vs. expected value

linear regressions (OLS, 2-way FE) / modèle de regression linéaire multiple: $Y_{it} = a + b.X_{it} + \varepsilon_{it}$ / clusters / STATA → ?? do file ?

Randomization / RCT / natural experiment

Internal vs. external validity

Why?

* Gold standard in impact evaluation = randomized experiments

* When it is not possible to run a randomized experiment (natural experiment), four alternative possibilities:

- instrumental variable (IV)

- matching

- Difference in differences (= double differences)

- Regression discontinuity (RD)

* Contrary to randomized experiments, these methods are not

assumption free. Rely on assumptions which cannot be tested

→ should be carefully discussed.

What ? Before after / control interv° analysis

Based on the existence of a **countrefactual**

How would have evolved the treated in absence of a treatment?
Impact of treatment=(total change – change that would have happened anyway)

Originally called BACI (Before / After … Control / Intervention), but:

* C/I does not control for **selection bias** (treatment not randomly allocated) &

* B/A does not control for **time trends**

Diff in diff

Effect of a treatment with counterfactual

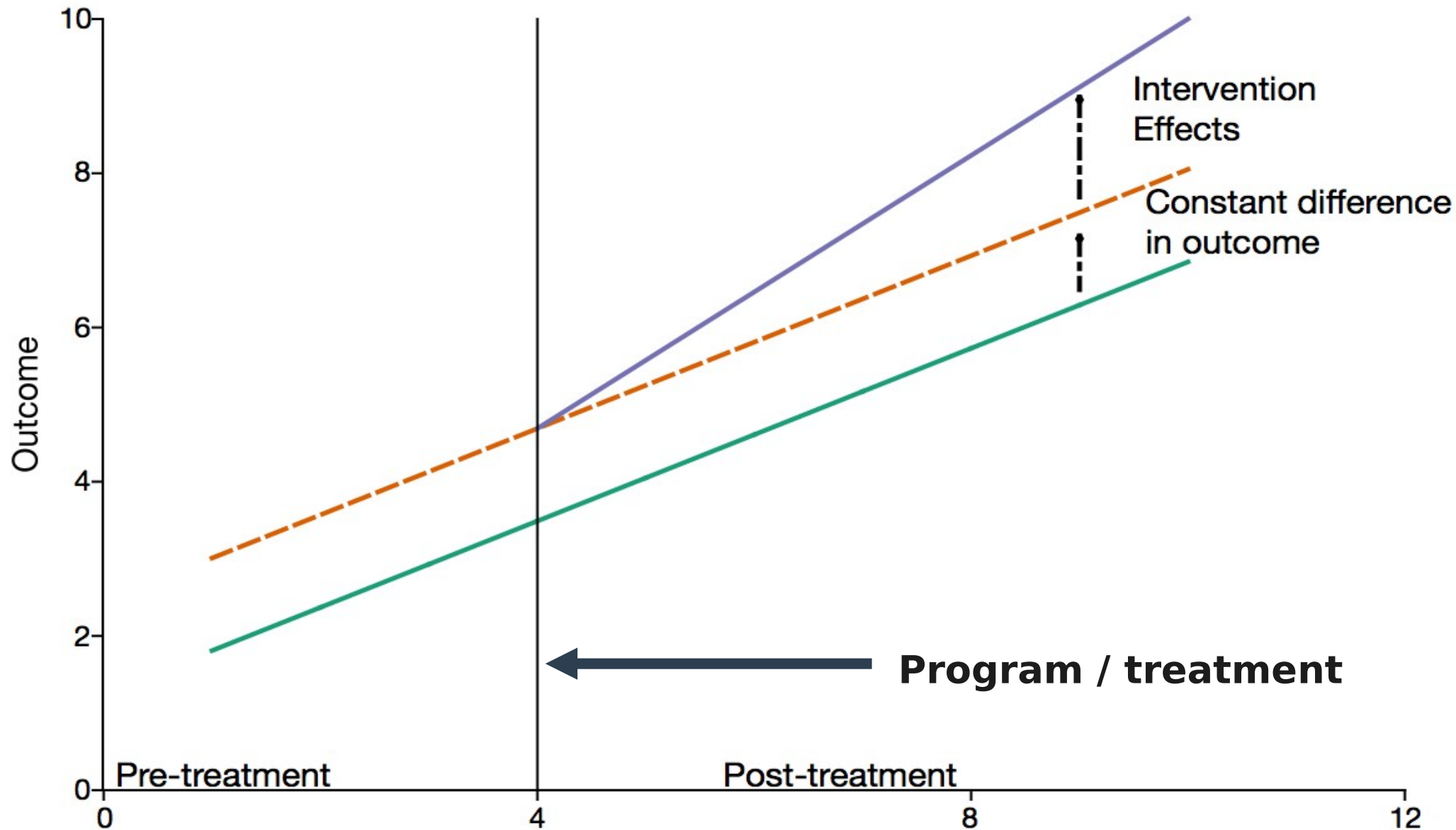Since the work by Ashenfelter and Card (1985), widespread use of difference-in-differences methods.

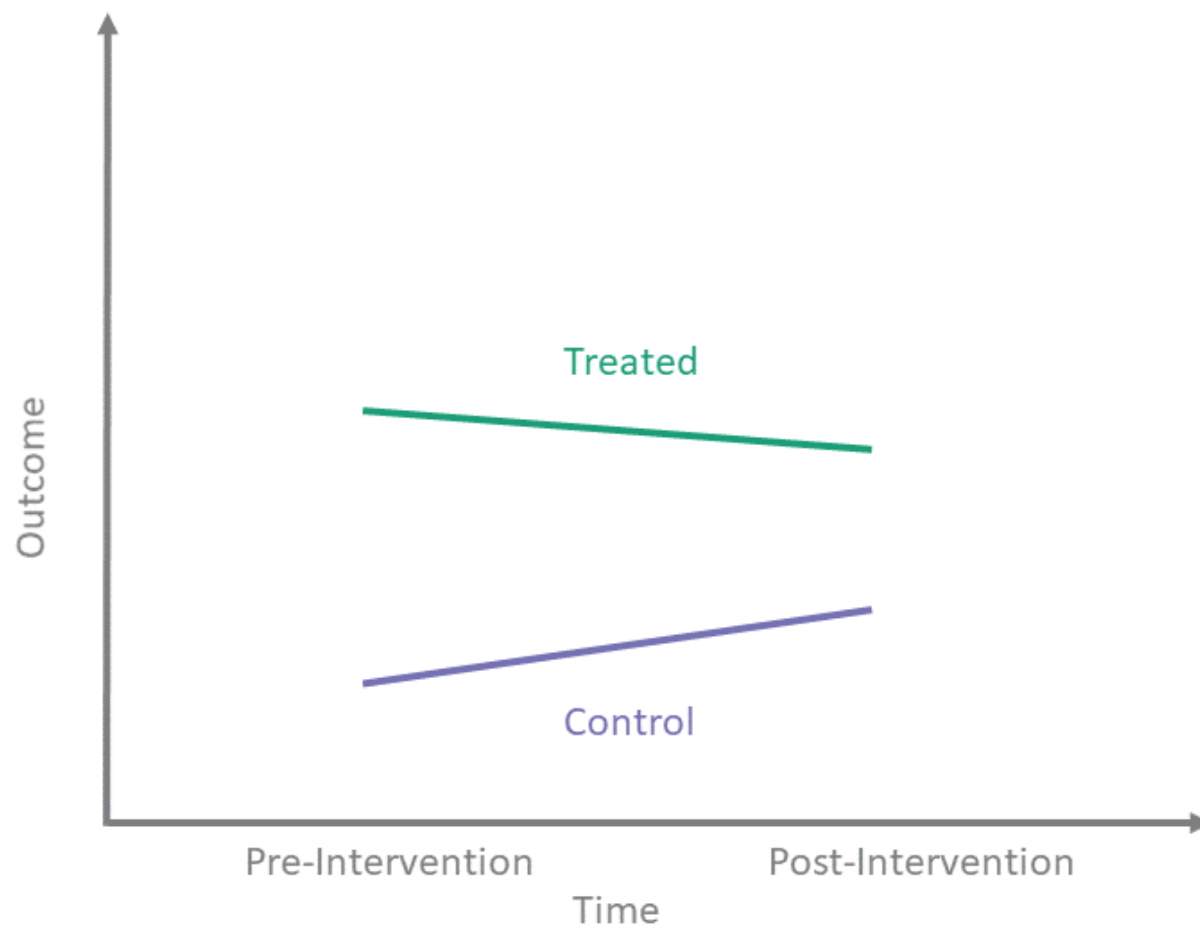Basically a BACI, controlling for time trends

Compatibility with the 2way fixed efffects (FE) panel framework

Difference-In-Differences Estimation

	Treatment	Comparison
Pre-Program	$\bar{Y}_{pre}^{treatment}$	$\bar{Y}_{pre}^{comparison}$
Post-Program	$\bar{Y}_{post}^{treatment}$	$\bar{Y}_{post}^{comparison}$

	$T = 0$	$T = 1$
$G = 0$	0% treated	0% treated
$G = 1$	0% treated	100% treated

Y = observed outcome
Y (1): what happens to someone if he receives the treatment.
Y (0): what happens to someone if he does not receive receives the treatment.

$$E(Y(1) - Y(0) \mid T = 1, G = 1) = E(Y(1) \mid T = 1, G = 1) - \mathbf{E(Y(0) \mid T = 1, G = 1)}$$
$$= \delta$$

Counterfactual
(unobserved)

Objective of **impact analysis**: construct the (unobserved) **counterfactual**

| | | Potential outcomes | | | |
| | | Pre-intervention | | Post-intervention | |
Individual	Group	Untreated	Treated	Untreated	Treated
1	Treated		✓	?	✓
2	Treated		✓	?	✓
3	Treated		✓	?	✓
...					
N-2	Control		✓	✓	?
N-1	Control		✓	✓	?
N	Control		✓	✓	?

4 DiD estimators

- Direct

$$DD = \bar{Y}^{treatment}_{post} - \bar{Y}^{treatment}_{pre} - \left(\bar{Y}^{comparison}_{post} - \bar{Y}^{comparison}_{pre} \right)$$

- Pooled OLS

$$Y_{i,t} = \alpha + \beta . D_i + \zeta . Post_t + \delta . (D_i * Post_t) + \varepsilon_{i,t}$$

- Fixed effects

$$Y_{it} = \alpha_i + \gamma_t + \beta^{DD} D_{it} + \varepsilon_{ti}$$

- First difference

$$Y_i\,{}^{post} - Y_i\,{}^{pre} = \alpha + \beta^{FD} . D_i + \varepsilon_i$$

With panel data and 2 periods, they are all the same

Intuitively

Simplest diff in diff estimator is:

$$DD = \bar{Y}_{post}^{treatment} - \bar{Y}_{pre}^{treatment} - \left(\bar{Y}_{post}^{comparison} - \bar{Y}_{pre}^{comparison} \right)$$

or $\quad \tau_2 = \mathbb{E}\left[Y_{i2}(1) - Y_{i2}(0) \,|\, D_i = 1\right].$

but $Y_{i2}(0)$ never observed

$$\tau_2 = \underbrace{\mathbb{E}\left[Y_{i2} - Y_{i1} \,|\, D_i = 1\right]}_{\text{Change for } D_i = 1} - \underbrace{\mathbb{E}\left[Y_{i2} - Y_{i1} \,|\, D_i = 0\right]}_{\text{Change for } D_i = 0},$$

Hypothesis

DID estimation requires that:

- **Intervention unrelated to outcome** at baseline (allocation of intervention was not determined by outcome)

Example of a famous issue: *Ashenfelter Dip*

Lower outcome just before treatment (earning before training), notably training and labor market (income before getting a training or a job)

- Treatment/intervention and control groups have **Parallel Trends** in outcome (see below for details)

Testing samples & hyp

- **Comparable samples: balance table**

Compares **means** of outcome and characteristics (not necessarily covariates) of **treated & control** groups before treatment

(STATA command: balancetable, version >=STATA14)

+ parallel trends *(next slides)*

- exchangeability, positivity, and **Stable Unit Treatment Value Assumption (SUTVA)**

Sutva

Composition of intervention and comparison groups is stable for repeated cross-sectional design (part of SUTVA)

The potential outcomes for each individual are unaffected by the treatment status of other individuals *i.e.*: no spillover effects of treatment onto untreated units (part of SUTVA)

Balance table

* = significance

*** : p-value <0.01

Table 2: Balance Table for Pre-Treatment variables for children in 2013 DHS

	Control group	Treatment group	Difference
Child variables :			
Height-to-age standard deviation	-1.973	-2.337	-0.365***
	(2.063)	(1.974)	(0.065)
Child's food diversity scale (from 0 to 10)	1.701	1.630	-0.071
	(1.801)	(1.653)	(0.051)
Birth order number	4.513	4.562	0.049
	(2.832)	(2.856)	(0.075)
1 if child is a girl, 0 if not	0.498	0.487	-0.011
	(0.500)	(0.500)	(0.013)
Age of child (months)	27.720	27.871	0.152
	(17.285)	(17.353)	(0.501)
Preceding birth interval (months)	33.949	33.446	-0.503
	(16.743)	(15.828)	(0.480)
1 if child is twin, 0 if not	0.044	0.036	-0.008
	(0.269)	(0.242)	(0.007)
Mother variables:			
Mother's body mass index	2,188.108	2,102.707	-85.401***
	(383.704)	(328.123)	(10.036)
Household variables:			
Number of household members	7.862	7.540	-0.322***
	(3.668)	(3.482)	(0.097)
1 if female headed-household, 0 if not	0.041	0.031	-0.010*
	(0.197)	(0.173)	(0.005)
Age of household head	41.202	40.260	-0.941***
	(11.735)	(11.267)	(0.309)
Education of household head (years)	1.297	0.553	-0.744***
	(3.005)	(1.937)	(0.075)
1 if respondent is Christian, 0 if not	0.056	0.003	-0.053***
	(0.230)	(0.053)	(0.006)
1 if respondent is Muslim, 0 if not	0.929	0.991	0.063***
	(0.258)	(0.092)	(0.006)
1 if respondent is currently married, 0 if not	0.974	0.991	0.017***
	(0.159)	(0.092)	(0.004)
Time to get to water source (minutes)	19.338	22.041	2.702***
	(28.120)	(24.261)	(0.730)
Cluster variables:			
Drought episodes	6.419	4.621	-1.798***
	(2.276)	(1.638)	(0.058)
Distance to GGW project	69.938	8.567	-61.371***
	(37.772)	(3.952)	(0.904)
Observations	7,420	1,751	9,171

Treatment group includes all the rural children who are less than 15 km from any Great Green Wall project site, including orchards, shelterbelts and boreholes. Children residing 15-30 km from a project are excluded from the sample. *** p<0.01, ** p<0.05, * p<0.1.

Normalized differences
in average covariates

Imbens and Rubin (2015) → look at the normalized differences

$$\frac{\bar{X}_{k,t} - \bar{X}_{k,c}}{\sqrt{(s^2_{k,t} + s^2_{k,c})/2}}.$$

$X_{k,t}$ & $s_{k,t}$ are the sample mean and sample standard deviation of the kth covariate of the treatment group, and are the analogous statistics for the control group
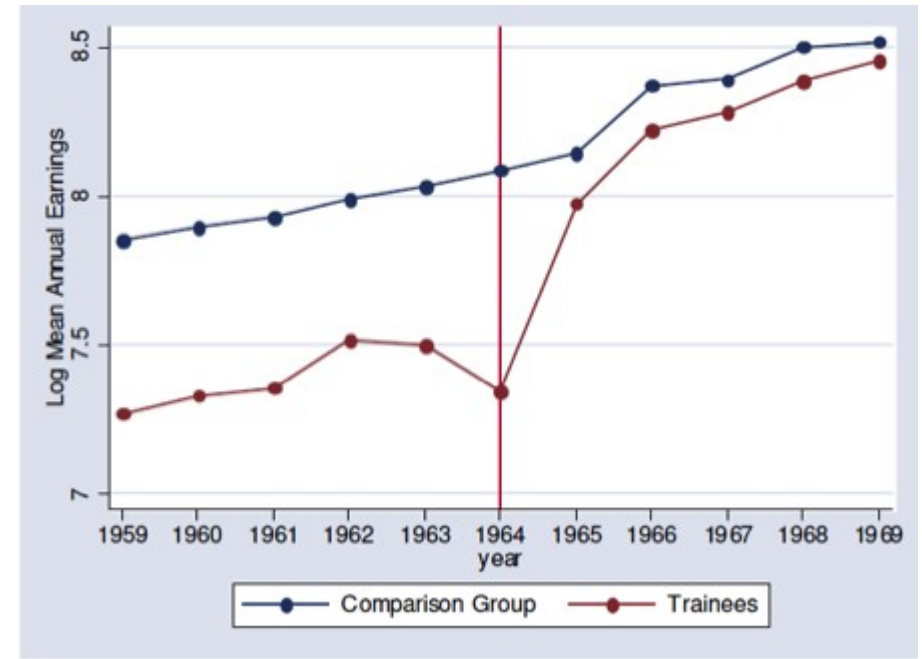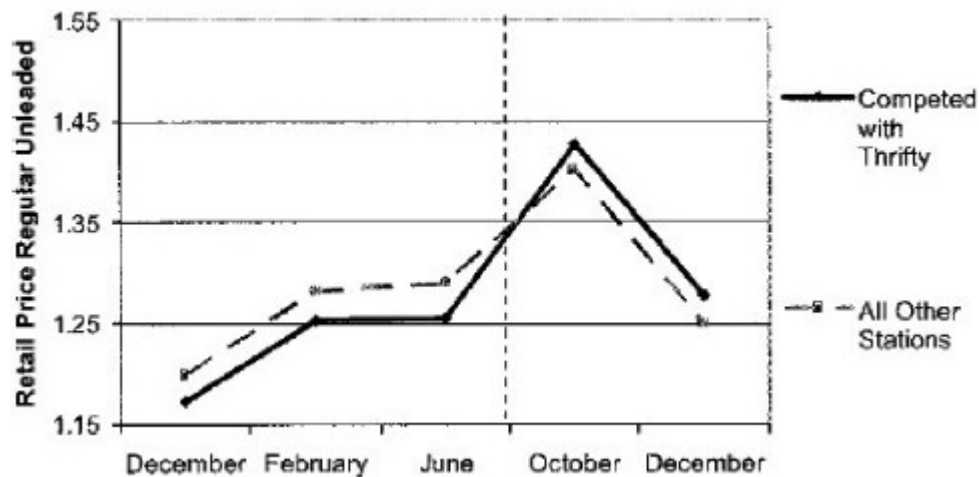
Rule of thumb: covariate imbalance if several covariates demonstrate a normalized difference of above 0.5

<.25 → balanced covariates

Imbens, G. & Rubin, D. (2015). Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press.
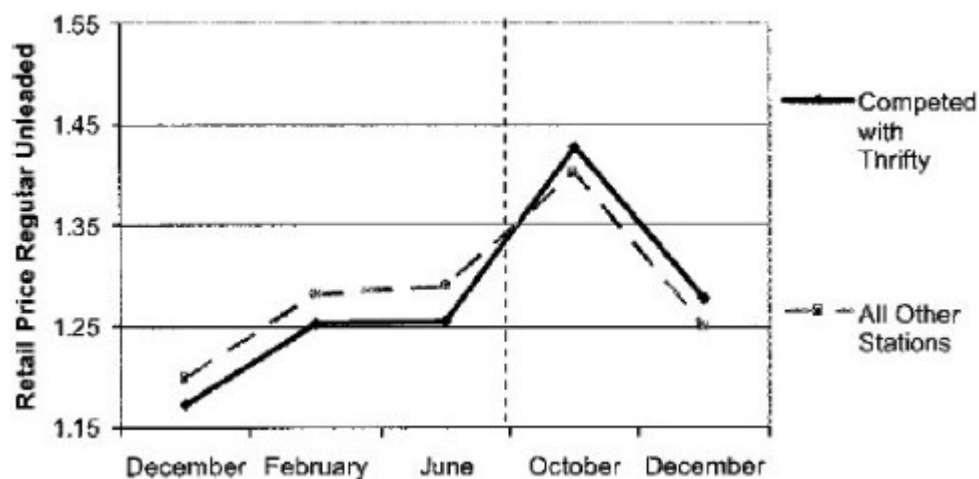
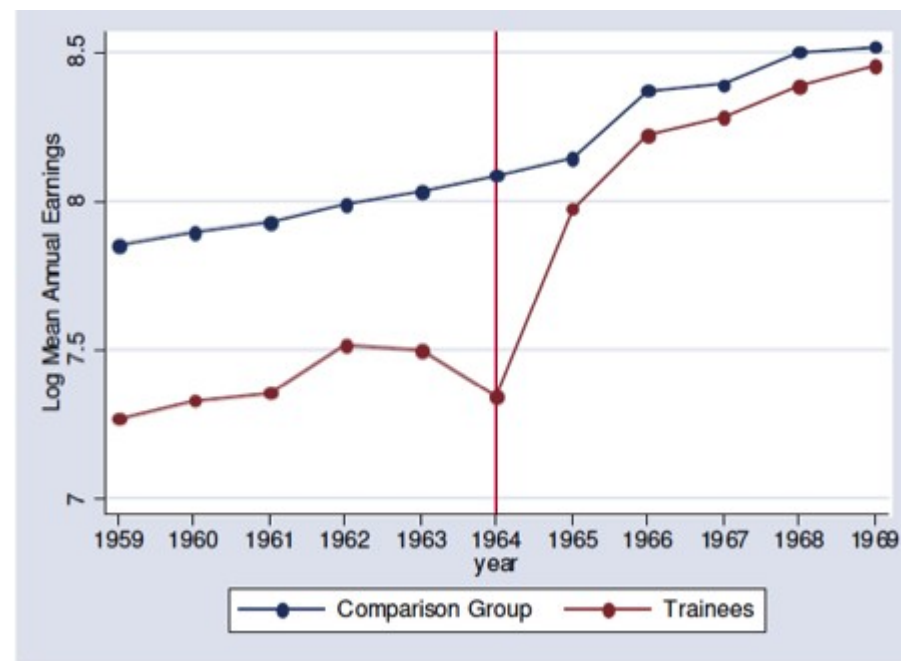# Parrallel trend (visual examples)

Assumption

$$\mathbb{E}\left[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0\right].$$

Parrallel trend (visual examples)

Assumption

validated   **not validated**

$$\mathbb{E}\left[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0\right].$$
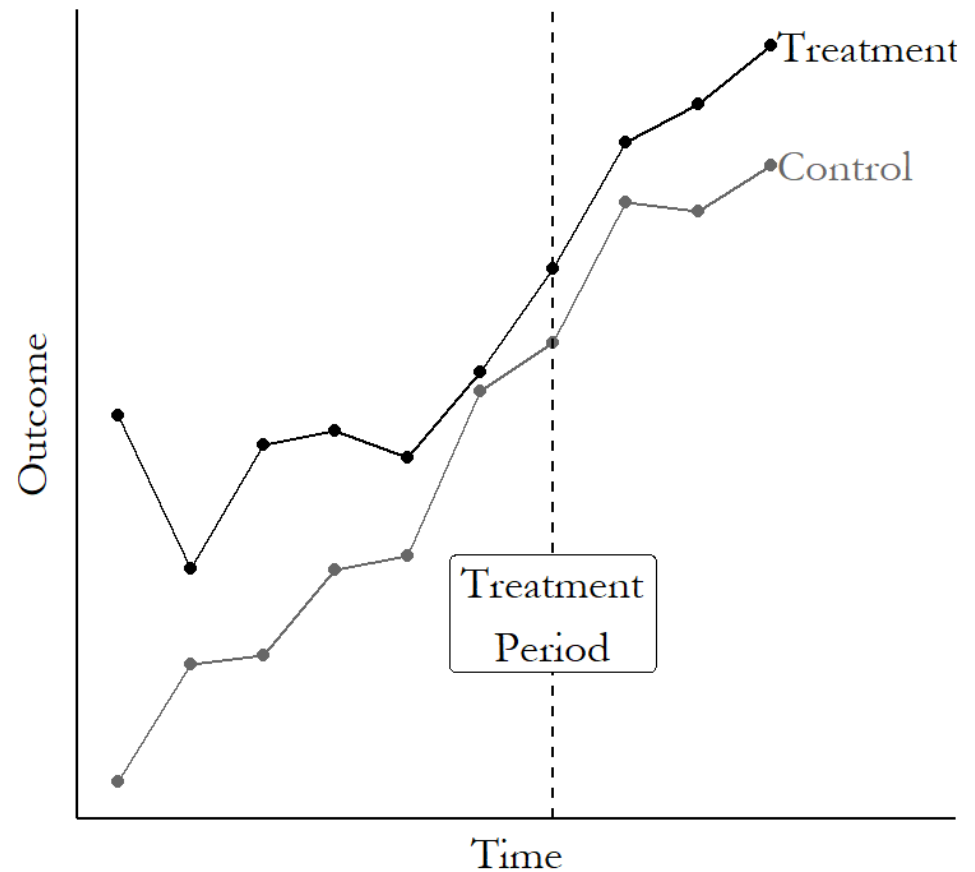
Parrallel trend (visual examples)

Assumtion

(a) Parallel Prior Trends
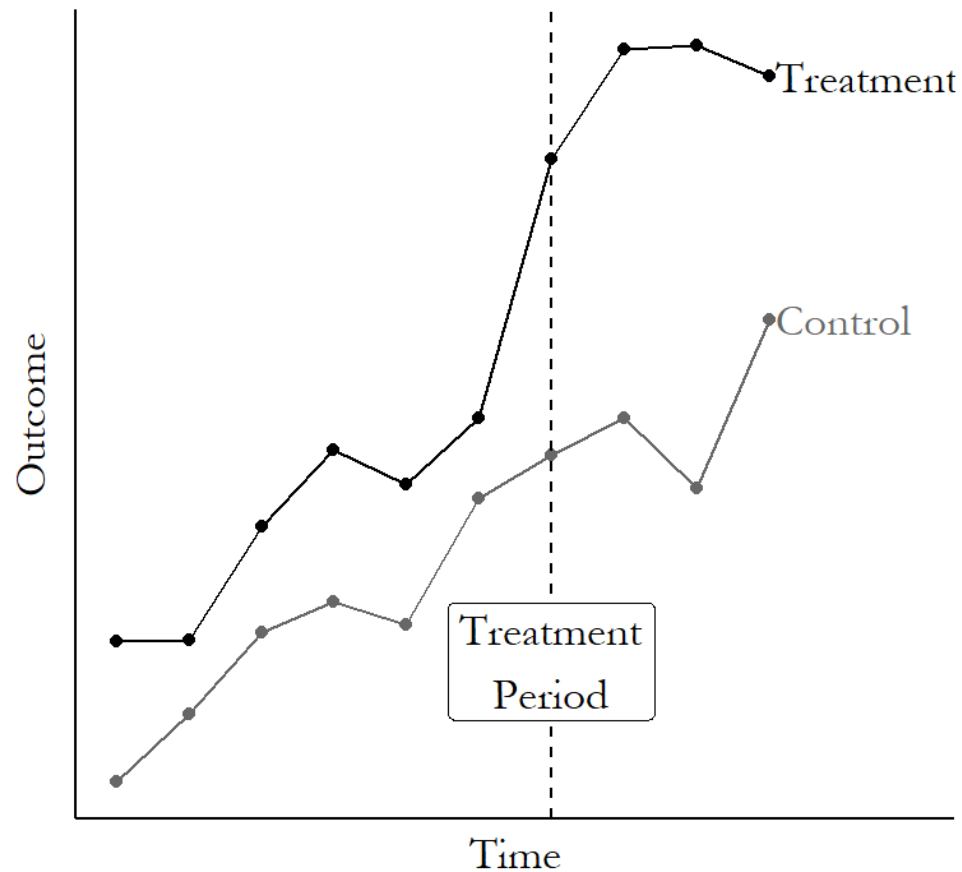
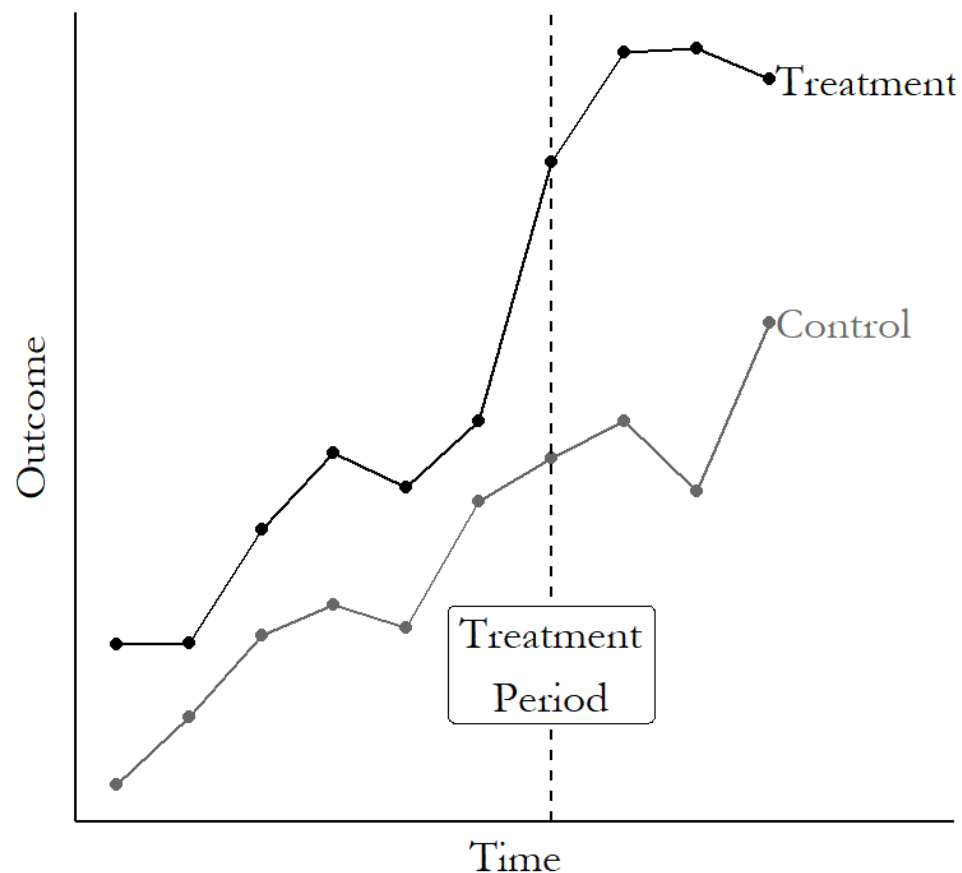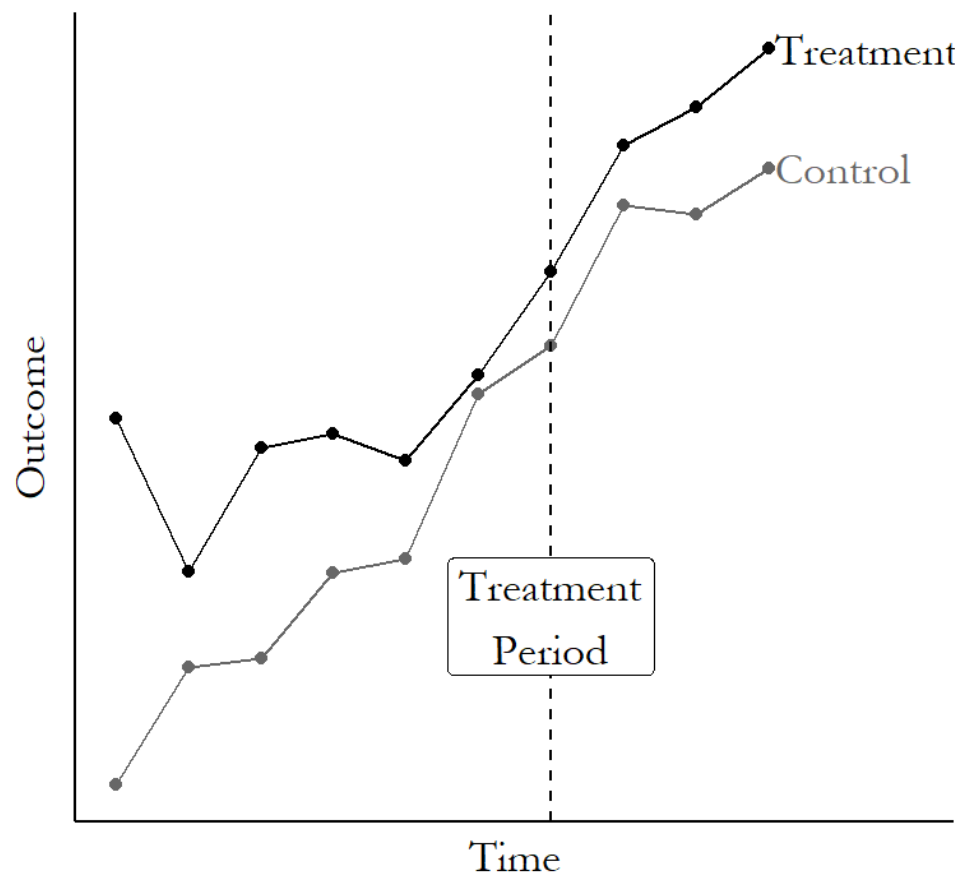(b) Converging Prior Trends

Parrallel trend (visual examples)

Assumtion

(a) Parallel Prior Trends

(b) Converging Prior Trends

validated

not validated

Careful: visual assessment + depends on specification (log transformation etc.)

Common trend assumption

Assumption: $E[Y_i \mid D_i = 0, t = \tau] = \gamma_i + \lambda_\tau$

- Selection bias relates to fixed characteristics of individuals (γ_i, amplitude does note change in time)

- Time trend (λ_t), same for treatment and control groups

$$DD = \bar{Y}_{post}^{treatment} - \bar{Y}_{pre}^{treatment} - \left(\bar{Y}_{post}^{comparison} - \bar{Y}_{pre}^{comparison} \right)$$

$= E[Y_{1i} \mid D_i = 1, t = 2] - E[Y_{0i} \mid D_i = 1, t = 1] - \big[E[Y_{0i} \mid D_i = 0, t = 2] - E[Y_{0i} \mid D_i = 0, t = 1] \big]$

$= E[\gamma_i \mid D_i = 1] + \delta + \lambda_2 - (E[\gamma i \mid Di = 1] + \lambda_1) - E[\gamma_i \mid D_i = 0] + \lambda 2 - E[\gamma_i \mid D_i = 0] + \lambda_1$

$= \boldsymbol{\delta} \qquad E[\gamma_i \mid D_i=1] - E[\gamma_i \mid D_i=0] - \big(E[\gamma_i \mid D_i=1] - E[\gamma_i \mid D_i=0] \big) \quad + \lambda_2 - \lambda_1 - (\lambda_2 - \lambda_1)$

      ---> **Selection biases**   = 0   --→ (same) **time trends** = 0

Alternative demo & notations

Parallel trends assumption:

$E(Y(0) \mid T = 1, \underline{G = 1}) - E(Y(0) \mid T = 0, \underline{G = 1}) =$

$E(Y(0) \mid T = 1, \underline{G = 0}) - E(Y(0) \mid T = 0, \underline{G = 0})$

Under this assumption:

$E(Y(0) \mid T = 1, G = 1) = E(Y(0) \mid T = 0, G = 1) + E(Y(0) \mid T = 1, G = 0) - E(Y(0) \mid T = 0, G = 0)$

and therefore :

$E(Y(1) - Y(0) \mid T = 1, G = 1) =$ ← **Impact of treatment**

$E(Y(1) \mid T = 1, G = 1) - E(Y(0) \mid T = 0, G = 1) -$

$[E(Y(0) \mid T = 1, G = 0) - E(Y(0) \mid T = 0, G = 0)]$ ← **Diff In Diff**

Identification (via no anticipation assumption)

$$\tau_2 = \mathbb{E}\left[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1\right].$$

Common trend assumption:

$$\mathbb{E}\left[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0\right].$$

Re-arranging terms

$$\mathbb{E}\left[Y_{i2}(0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i1}(0) \mid D_i = 1\right] + \mathbb{E}\left[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0\right].$$

by the no anticipation assumption, $\mathbb{E}\left[Y_{i1}(0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i1}(1) \mid D_i = 1\right]$

$$\mathbb{E}\left[Y_{i2}(0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i1}(1) \mid D_i = 1\right] + \mathbb{E}\left[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0\right]$$
$$= \mathbb{E}\left[Y_{i1} \mid D_i = 1\right] + \mathbb{E}\left[Y_{i2} - Y_{i1} \mid D_i = 0\right],$$

Since *Y(1)* observed for treated units

& *Y(0)* observed for control (untreated) units

$$\tau_2 = \mathbb{E}\left[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1\right].$$

$$\tau_2 = \underbrace{\mathbb{E}\left[Y_{i2} - Y_{i1} \mid D_i = 1\right]}_{\text{Change for } D_i = 1} - \underbrace{\mathbb{E}\left[Y_{i2} - Y_{i1} \mid D_i = 0\right]}_{\text{Change for } D_i = 0},$$

Robustness: placebo test & alternative outcomes

In order to test if the **significant impact** found is **robust**, we may test:

- false (placebo) treatments on untreated period: if found significant → clue that something may be awry about the parallel trends assumption

- alternative outcomes: if significant impacts found on other outcomes that may not be impacted by the policy, then something (other transmission channels or impacts of other, sometimes unobservable, variables)

Type of effect estimated

If you have true randomization in a representative sample and don't need to do any adjustment, you have an average treatment effect (**ATE**). If people selected to receive the treatment had a choice whether to use it or not (selection bias), you can only properly estimate the intention to treat (**ITT**), i.e. the average impact of being selected among the treated (for a sample of users and non-users).

If you have true randomization only within a certain group, and you isolate that group so you can take advantage of that randomization, you have a **conditional average treatment effect**.

If you are identifying your effect by assuming that some untreated group is what the treated group would look like if they hadn't been treated, then we have the average treatment on the treated (**ATT**).

If part of the variation in treatment is driven by an exogenous variable, and you isolate just the part driven by that exogenous variable, then you have a local average treatment effect (**LATE**).

Type of effect estimated

Average Treatment Effect. The average treatment effect across the population.

Average Treatment on the Treated. The average treatment effect among those who actually received the treatment in your study.

Average Treatment on the Untreated. The average treatment effect among those who did not actually receive the treatment in your study.

Conditional Average Treatment Effect. The average treatment effect among those with certain values of certain variables (for example, the average treatment effect among women).

Heterogeneous Treatment Effect. A treatment effect that differs from individual to individual.

Intent-to-Treat. The average treatment effect of assigning treatment, in a context where not everyone who is assigned to receive treatment receives it (and maybe some people not assigned to treatment get it anyway).

Local Average Treatment Effect. A weighted average treatment effect where the weights are based on how much more treatment an individual would get if assigned to treatment than if they weren't assigned to treatment.

Inference in a regression framework

To implement diff-in-diff in a regression framework, we estimate:

$$Y_{i,t} = \alpha + \beta \cdot D_i + \zeta \cdot Post_t + \delta \cdot (D_i * Post_t) + \varepsilon_{i,t}$$

where:

- $D_i = 1$ indicator of treatment group (dummy)

- $Post_i$ is an indicator (dummy) $= 1$ if t $= 2$

- δ is the coefficient of interest (the treatment effect)

$= $ ATT estimand

- $\alpha = E[\gamma_i | D_i = 0] + \lambda_1 \rightarrow$ pre-program mean in comparison group

- $\beta = E[\gamma_i | D_i = 1] - E[\gamma_i | D_i = 0] \rightarrow$ selection bias

- $\zeta = \lambda_2 - \lambda_1 \rightarrow$ time trend (common trend assumption)

Visual interpretation of coefficients

Alternatively:

$Y = B0 + B1*[Time] +$

$B2*[Intervention] +$

$B3*[Time*Intervention] +$

$B4*[Covariates] + \varepsilon$
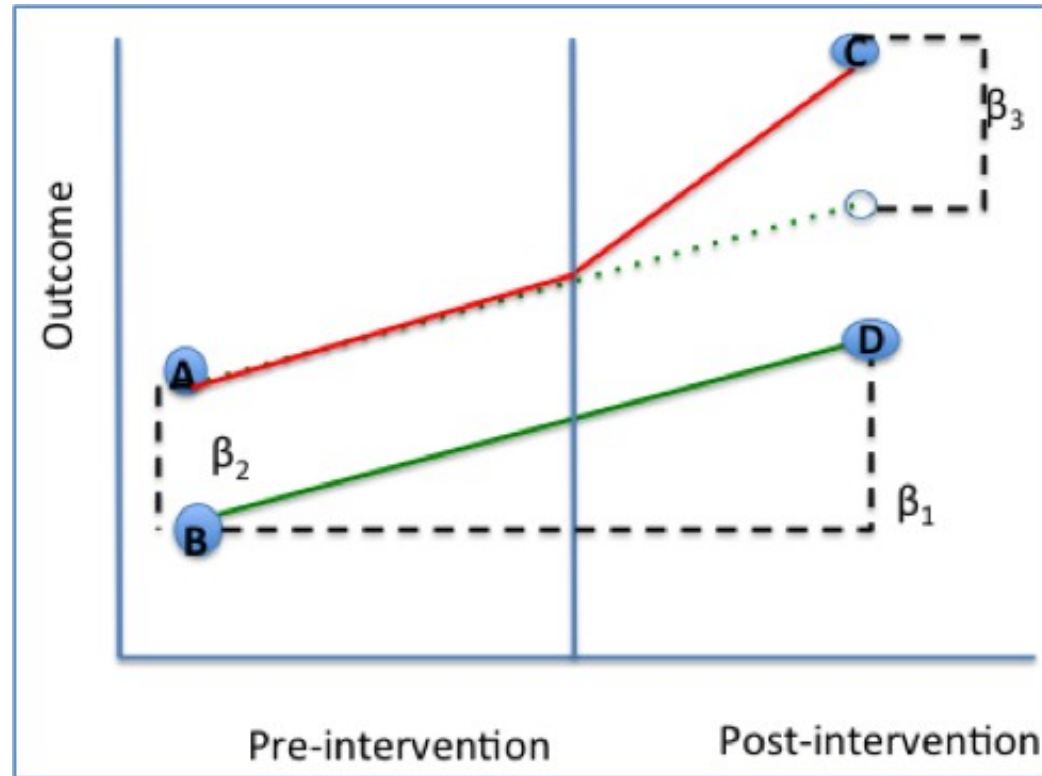
Coefficient	Calculation	Interpretation
β_0	B	Baseline average
β_1	D-B	Time trend in control group
β_2	A-B	Difference between two groups pre-intervention
β_3	(C-A)-(D-B)	Difference in changes over time

More than 2 periods: panel with 2way FE

$$Y_{it} = \alpha_i + \gamma_t + \beta^{DD} D_{it} + \varepsilon_{ti}$$

With a balanced panel, the OLS coefficient on β is also numerically identical to the coefficient from a regression that replaces α_i and γ_t with treatment group and period fixed effects (The POST and TREAT variables are thus subsumed by the fixed effects structure),

$$Y_{it} = \alpha + D_i \cdot \theta + 1[t = 2]\cdot\zeta + (1[t = 2] \cdot D_i)\cdot\beta + \varepsilon_{it}$$

The latter regression generalizes to repeated cross-sectional data.

Panel with two way FE

Diff in diff is similar to 2 ways FE:

$Y_{it} = \alpha_i + \gamma_t + \beta.X_{it} + e_{it}$

but when > 1 pre and 1 post period; Bertrand, Duflo and Mullainathan (2004) → standard errors are biased downward (i.e., too small, over reject)

But → Compare in different timing of treatment & control after treatment (before is ok) C. de Chaisemartin & X. D'Haultfœuille / Borusyak / Callaway & Sant'anna, → methods which does not rely on constant treatment effects, only conditional on common trend assumption

https://econ2017.sites.olt.ubc.ca/files/2019/05/pdf_594lecture3_thomas-lemieux.pdf

http://economics.ozier.com/econ626/lec/econ626-L03-slides-2019.pdf

https://personal.utdallas.edu/~d.sul/Econo2/lect_10_diffindiffs.pdf

The most popular way to account for clustered data in diff-in-diff is clustered standard errors (Cameron and Miller 2015; Abadie et al. 2022), nb clusters >30-40 (or fewer with wild bootstrap Cameron, Gelbach, and Miller, 2008,cf. Canay, Santos and Shaikh (2021) for validity).

Diff-in-dif in panel & assumption of homogeneous treatment effects

When treatment effects are homogeneous (across units & time), DD estimation yields **average treatment effect on the treated** (ATT)

If not, it averages across treated units and over time

When treatment effects are heterogeneous across units (not time): variance-weighted treatment effect ($\neq$ATE)

When treatment effects change over time, biased estimation (DD estimate of treatment effect may depend on choice of evaluation window)

$\Rightarrow$ Changes in treatment effect bias DD coefficient

$\Rightarrow$ Event study, stacked DD more appropriate

Long term effects

As mentioned above: late and early treatment analysed similarly in 2way FE.

Look at annual impact rather the average

Interact year x post x treatment

Event studies (either matching or DiD) in the case of **staggered** entry design

TAB. – Impact by public-private partnership model

	F/T support	Co-Management	Delegated
ATT overall			
PPP = TRUE	-0.0573	-0.1486	-0.1382**
ATT by group			
Group = 2003	-0.0767*		-0.1952***
Group = 2005	0.1287	0.0819	-0.3550***
Group = 2007		-0.0189	
Group = 2008		-0.0791**	-0.0278
Group = 2010			-0.1566***
Group = 2012		0.0208***	
Group = 2013		-0.2264***	
Group = 2014		-0.2131***	
Group = 2015		-0.3556***	-0.1368**
Group = 2016		-0.1477	
Group = 2017	-0.0695	-0.1119	-0.0063
Group = 2019		0.1699**	-0.1244

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1
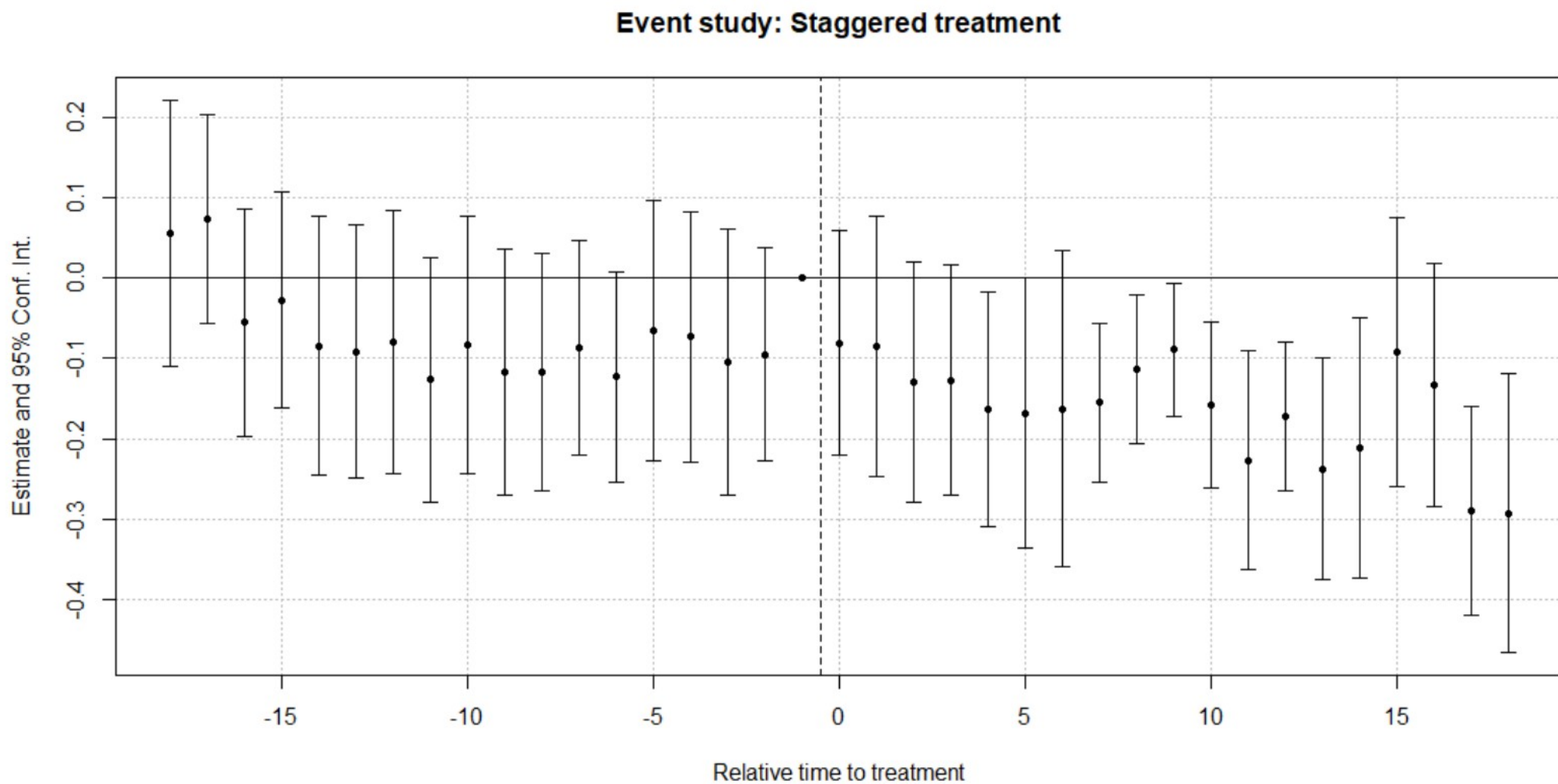
Fɪɢ. – Estimation of the event study on matched data
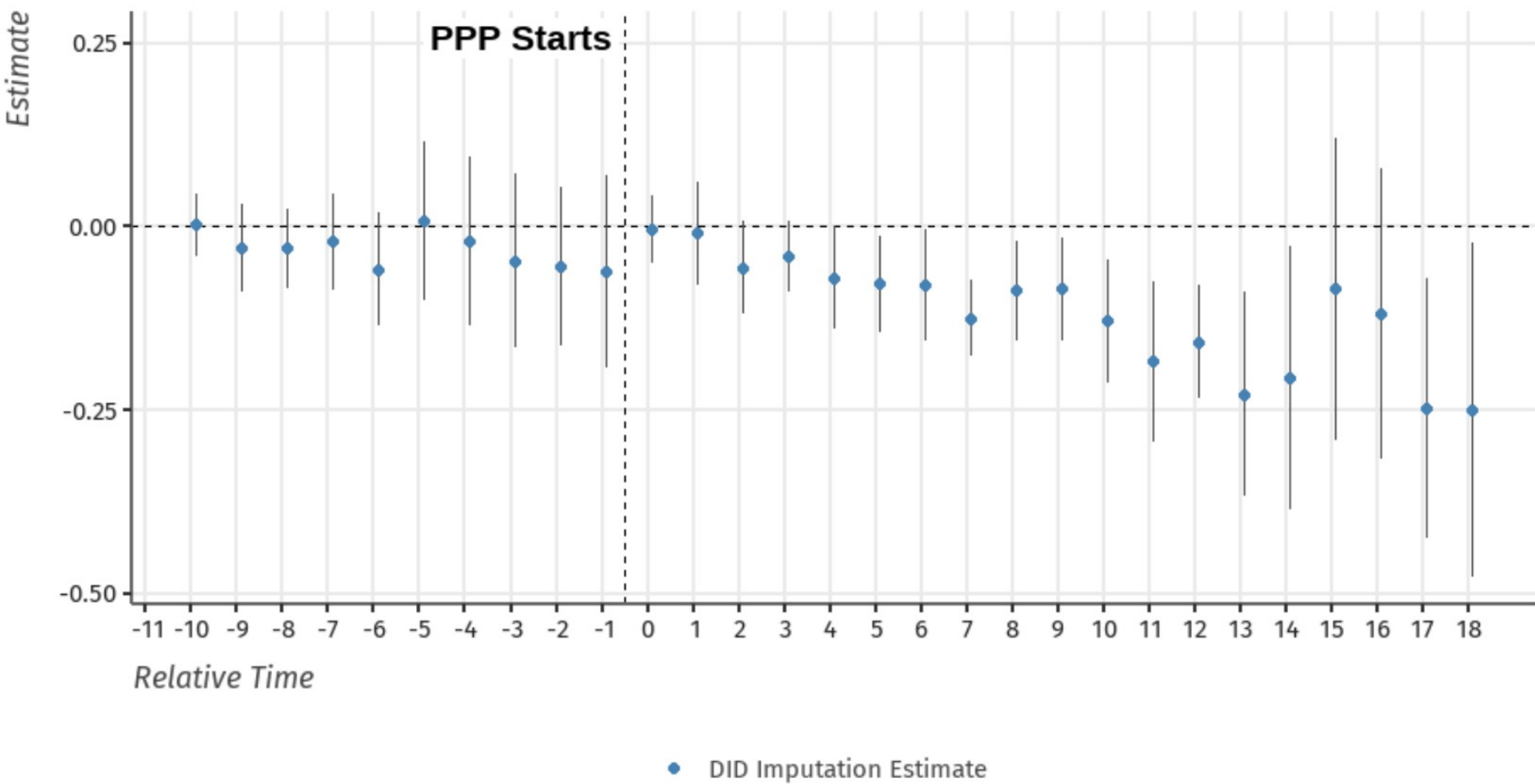
FIG. – Event study inspired by Borusyak et al. (2021)

What to remember ?

- Causal assessment requires the building of a (unobservable) counterfactual (how would treated individuals react in absence of a treatment?)

- In absence of a random control group (RCT) → need to convince about the similarity of control and treatment group

- DiD may be computed either directly (no SD, no significance associated to coef.) or through a regression.

- Timing of the treatment (2 periods or more)

- Treatment effect varies across time & individuals

Checklist fo DiD practitioners

Everyone treated at the same time?

Yes → panel balanced?   Yes → TWFE

No → "heterogeneity-robust" estimator for staggered treatment timing, TWFE only if confident in treatment effect homogeneity

Validity of parallel trends (PT)?

Yes → why ? Justify the choice of functional form

No →

Conditional on covariates

Event study plot

Diagnostics of the power of pre-test + sensitivity analysis to violations of PT

Recent advances in DiD
(J. Roth; P. H. C. Sant'Anna; A. Bilinski; J. Poe, 2022*)

→ (i) allow for multiple periods and variation in treatment timing

→ (ii) consider potential violations of parallel trends

→ (iii) depart from the assumption of observing a sample of many independent clusters sampled from a super-population

***https://psantanna.com/files/RSBP_DiD_Review.pdf**

(i) Multiple periods and variation in treatment timing

TWFE regressions make both "clean" comparisons between treated and not-yet treated units as well as "forbidden" comparisons between units who are both already-treated.

When treatment effects are heterogeneous, these "forbidden" comparisons potentially lead to severe drawbacks (false coef sign…)

(ii) Non-parallel trends

- conditional parallel trends assumption: holds only conditional on observed covariates

* low power of these 'pre-trend' tests

* pre-test bias (selection effect)

* treatment effect still interesting

Sample Sizes Effect on Power
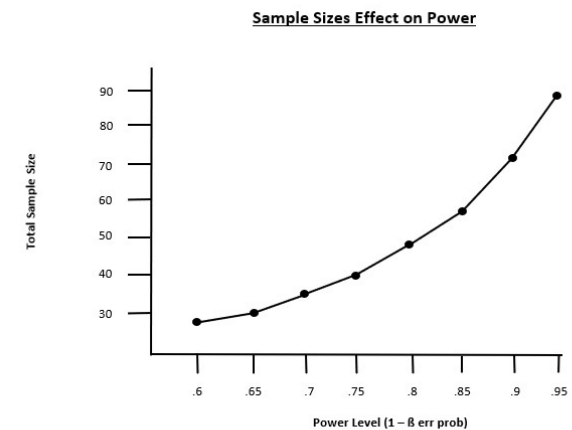
Note: As the sample size increases in the model, so does power.

→ pre-test only "passes" if there is evidence that the pre-trend is small

→ the post-treatment violation of parallel trends is assumed to

be no larger than the maximal pre-treatment violation of parallel trends

(iii) Alternative sampling

- When large number of treated and untreated clusters sampled from a super-population:

permutation (recompute the statistic under many permutations of the treatment assignment (@ the cluster level) → reject the null hypothesis of no effect)

& bootstrap procedures

- design-based inference randomness in the treatment assignment vector (considered to be random)

Regression Discontinuity Design (RDD)

Exploiting a discontinuity in the outcome, depending on a variable

- either in space (distance / border)

- or threshold in policy implementation: trigger

A **threshold/cutpoint (c)**, of a continuous eligibility index is clearly defined to determine treatment (beneficiaries & non-beneficiaries of the policy).

D = 1 $\{x \geq c\}$

The person goes from treated with probability=0, to treated with probability=1, when x crosses the threshold.

outcome

A

B

A
B = IMPACT
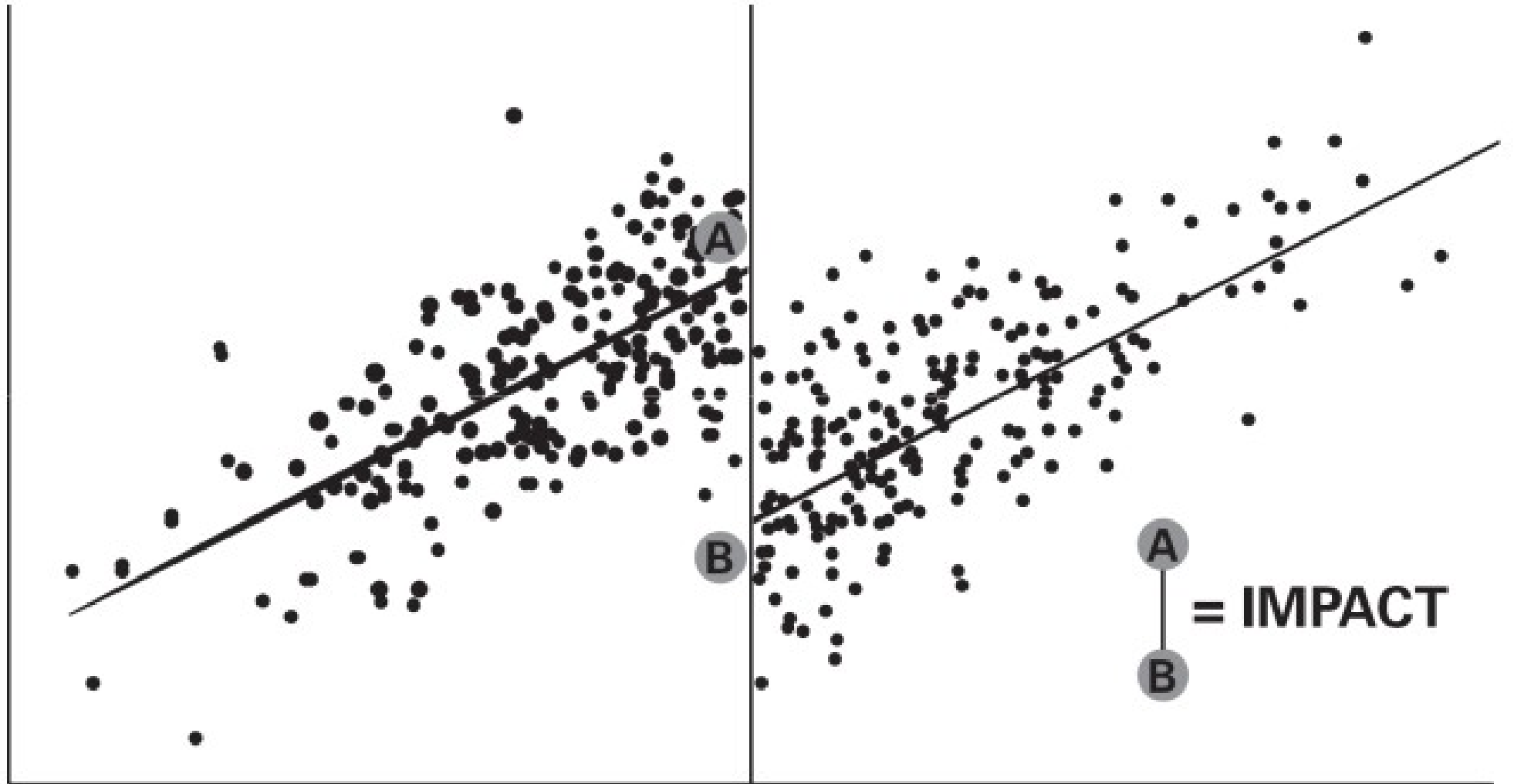
$[c - h \qquad \mathbf{c} \qquad c + h]$

index

threshold

$[\qquad \text{interval} \qquad]$

- Anti-poverty programmes (index: poverty index / threshold: poverty line or half median income)

- retirement policies (index: age / threshold: 60 years old)

- Environmental pollution control programmes, e.g. European fines for air or water pollutants (index: concentration of pollutants / threshold: defined by health or environmental risks)

- Agro-environmental policies (index: ha of land use or environmental practices / threshold defined for subsidies)

How does it work

Near the threshold, beneficiaries are similar & may not manipulate the index or the threshold.

Define an interval around the threshold $[c - h \leq x_i \leq c + h]$

The discontinuity simply acts as a random experiment at the cutpoint $x = c$

The outcome level nearby the threshold is used as the counterfactual level (Y_0) of beneficiaries

For estimation, since $\Delta \mathit{ATE}$ is defined as x approaches c from above and below and holds in the limit, one restricts attention to observations around the threshold.

- **Two approaches:**

– Kernels

- Nonparametric technique that can be used to calculate the weighted average outcome on both sides of c (weight more obs. close to c)

- As the number of observations used in the weighted average increases (you get further from c), the bias increases

– Local linear regression

Sharp vs. fuzzy discontinuity

Sharp: probability of being treated goes from 0 to 1 when crossing the threshold

Fuzzy: jump in treatment probability (but some treated below threshold and some not treated above it)

Example (C. de Chaisemartin)

You get 2 years unemployment benefits after 50 in Austria, and only 6 months before that age

$$\lim_{A \to 50, A \geq 50} E(Y \mid A) - \lim_{A \to 50, A < 50} E(Y \mid A) =$$

$$\lim_{A \to 50, A \geq 50} E(Y_1 \mid A) - \lim_{A \to 50, A < 50} E(Y_0 \mid A) =$$

$$E(Y_1 \mid A = 50) - E(Y_0 \mid A = 50) =$$

$$E(Y_1 - Y_0 \mid A = 50),$$

thanks to the continuity assumption.

Sharp design

to estimate $\lim_{A \to 50, A \geq 50} E(Y|A) - \lim_{A \to 50, A < 50} E(Y|A)$, run the

following regression:

$$Y = \alpha + \beta_1(A - 50) + \beta_2(A - 50)^2 + \ldots + \beta k (A - 50)^k$$
$$+\beta'_1 (A - 50) 1_{\{A \geq 50\}} + \beta'_2 (A - 50)^2 1_{\{A \geq 50\}} + \ldots + \beta'_k (A - 50)^k 1_{\{A \geq 50\}} + \gamma 1_{\{A \geq 50\}} + \varepsilon$$

You can check that under this model,

$$\gamma = \lim_{A \to 50, A \geq 50} E(Y|A) - \lim_{A \to 50, A < 50} E(Y|A).$$

RDD: hypothesis & design checking

Hyp. 1: Continuity of the Conditional Regression Function (in x, at c, for all Y)

Hyp. 2: Monotonicity of treatment (probability of treatment is not increasing in c), weak monotonicity in the case of fuzzy design.

Check graphically:

- Treatment by forcing variable

(SRD: avg from 0 to 1, FRD: just a discrete jump

- Outcomes by forcing variable (effect?)

- Other covariates by forcing variable (not due to programme?)

- Density of forcing variable (enough N_{obs} around c, threshold manipulation → violation hyp 1), McCrary test (DCdensity in Stata)

Outcome before the programme

Éligibles

Non éligibles
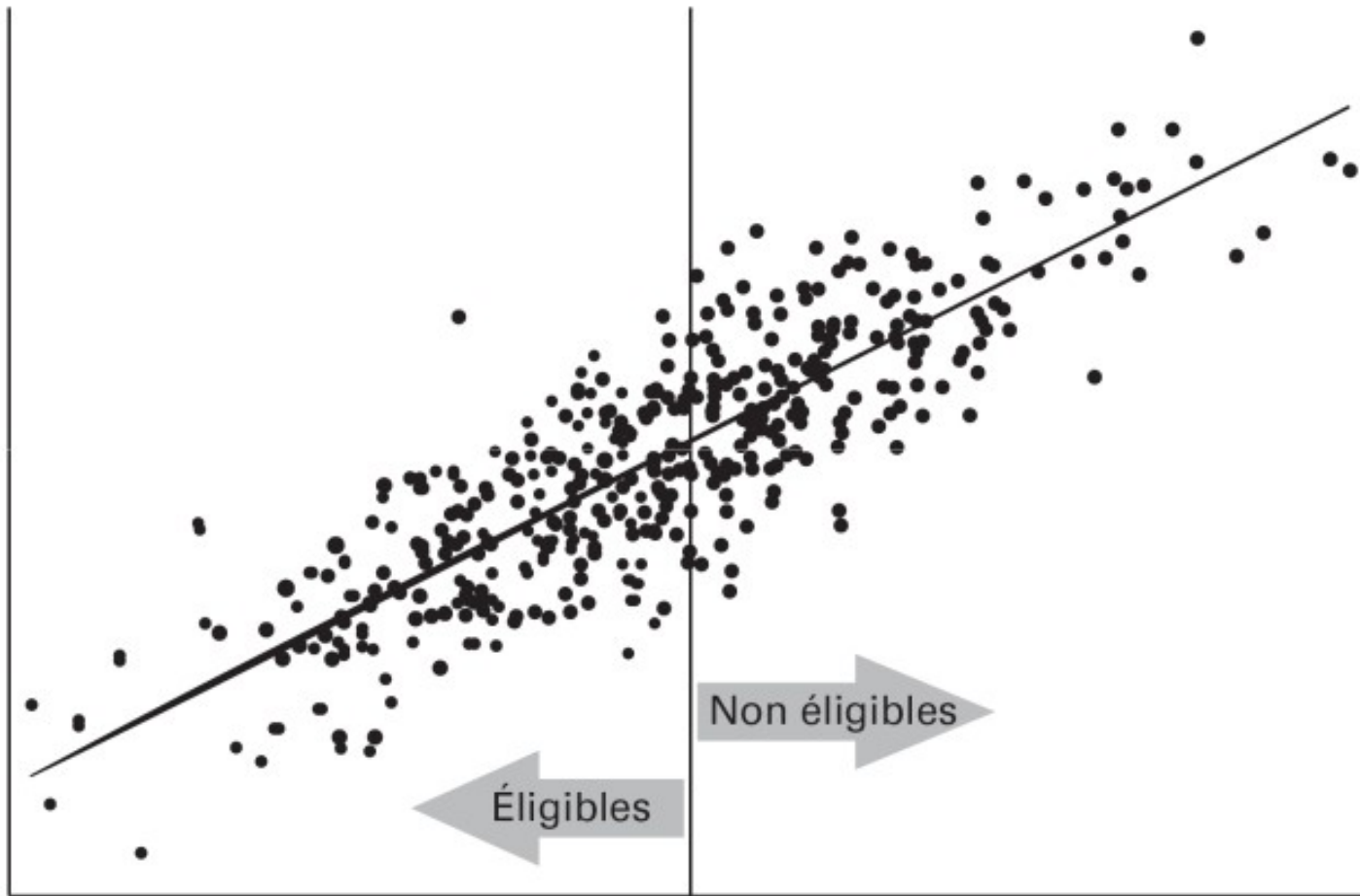
Outcome after the programme
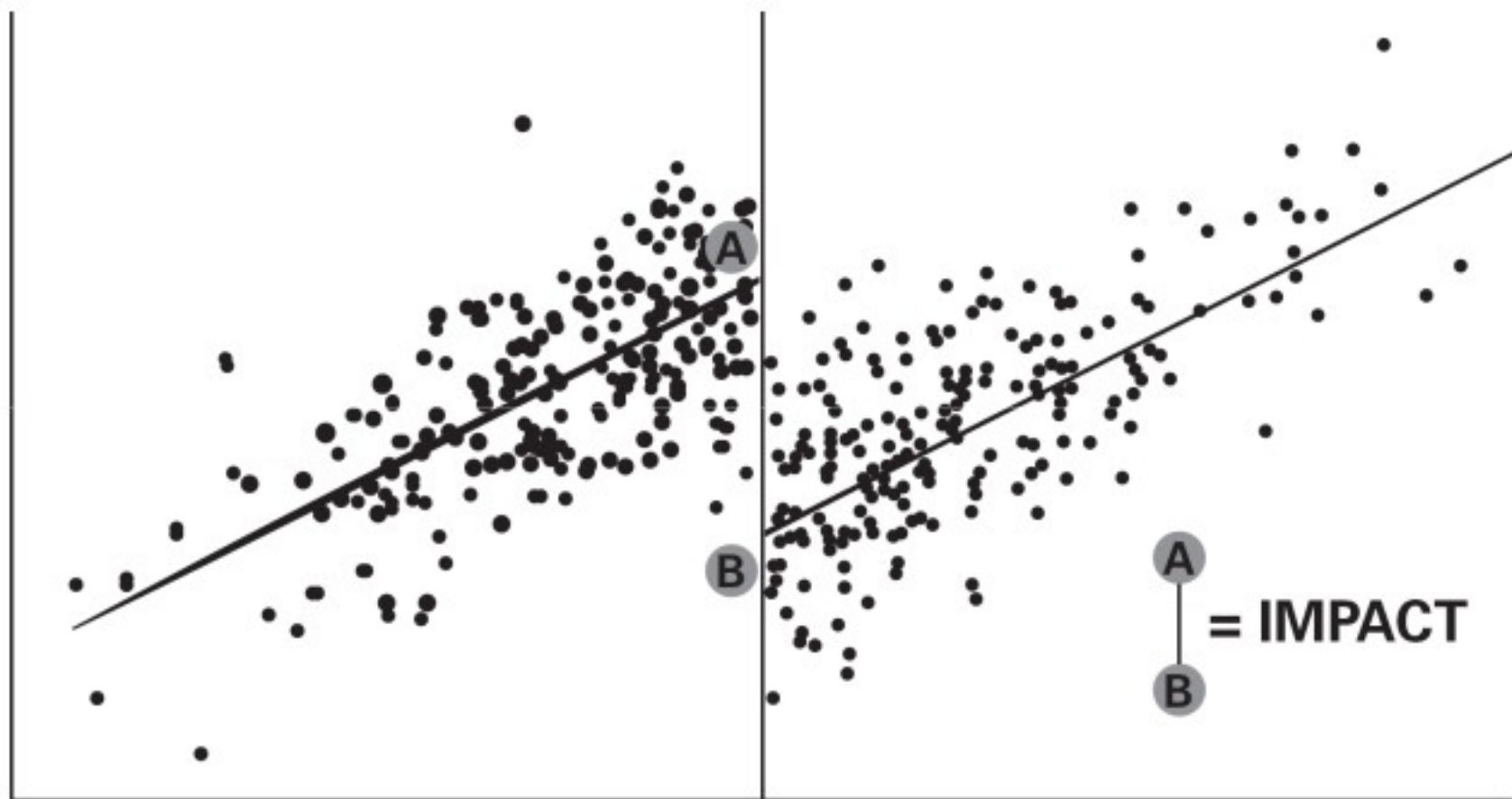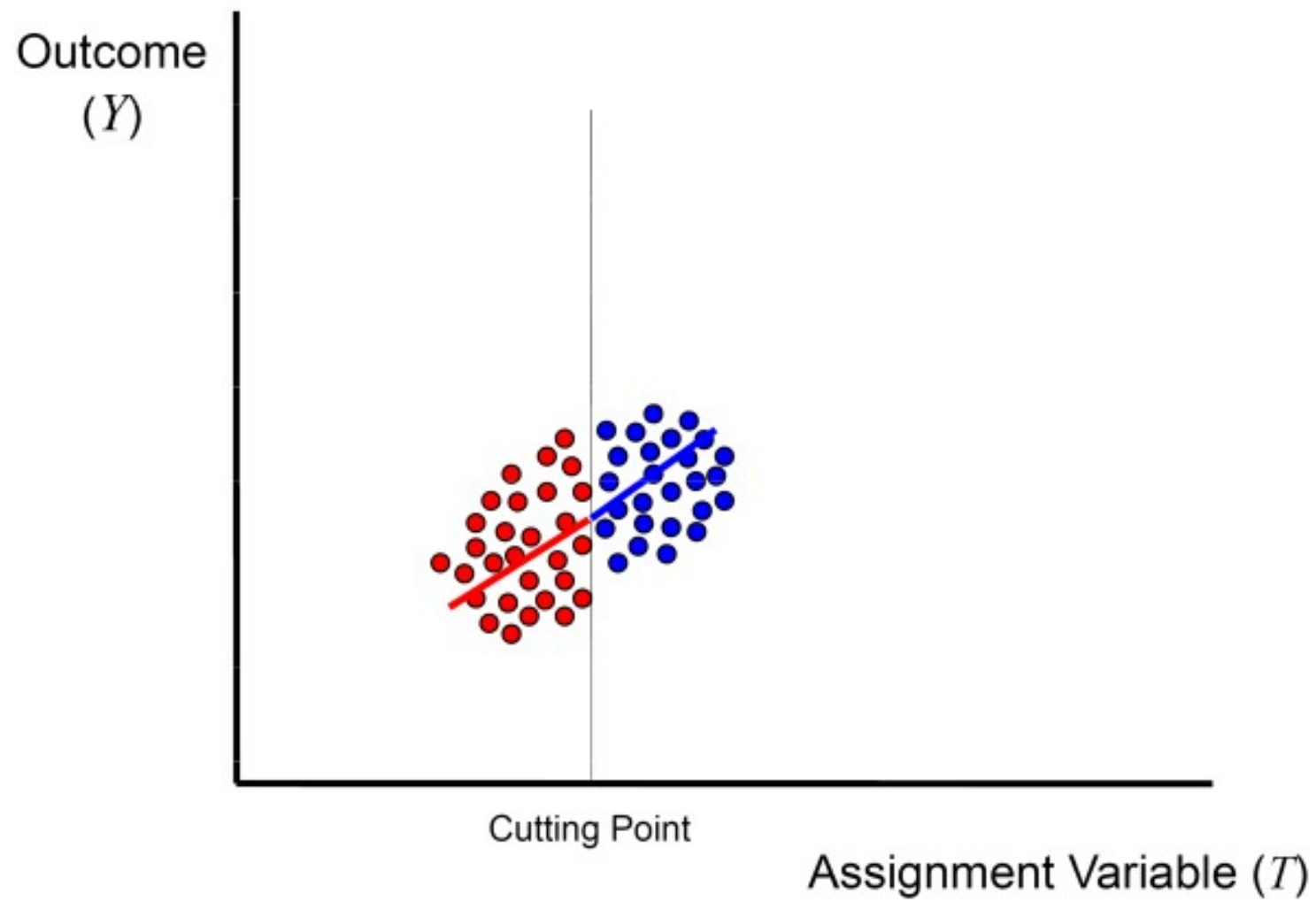

Impact = E(y ^{1} − y ^{0} | X = c)

Non-linearity ≠ discontinuity
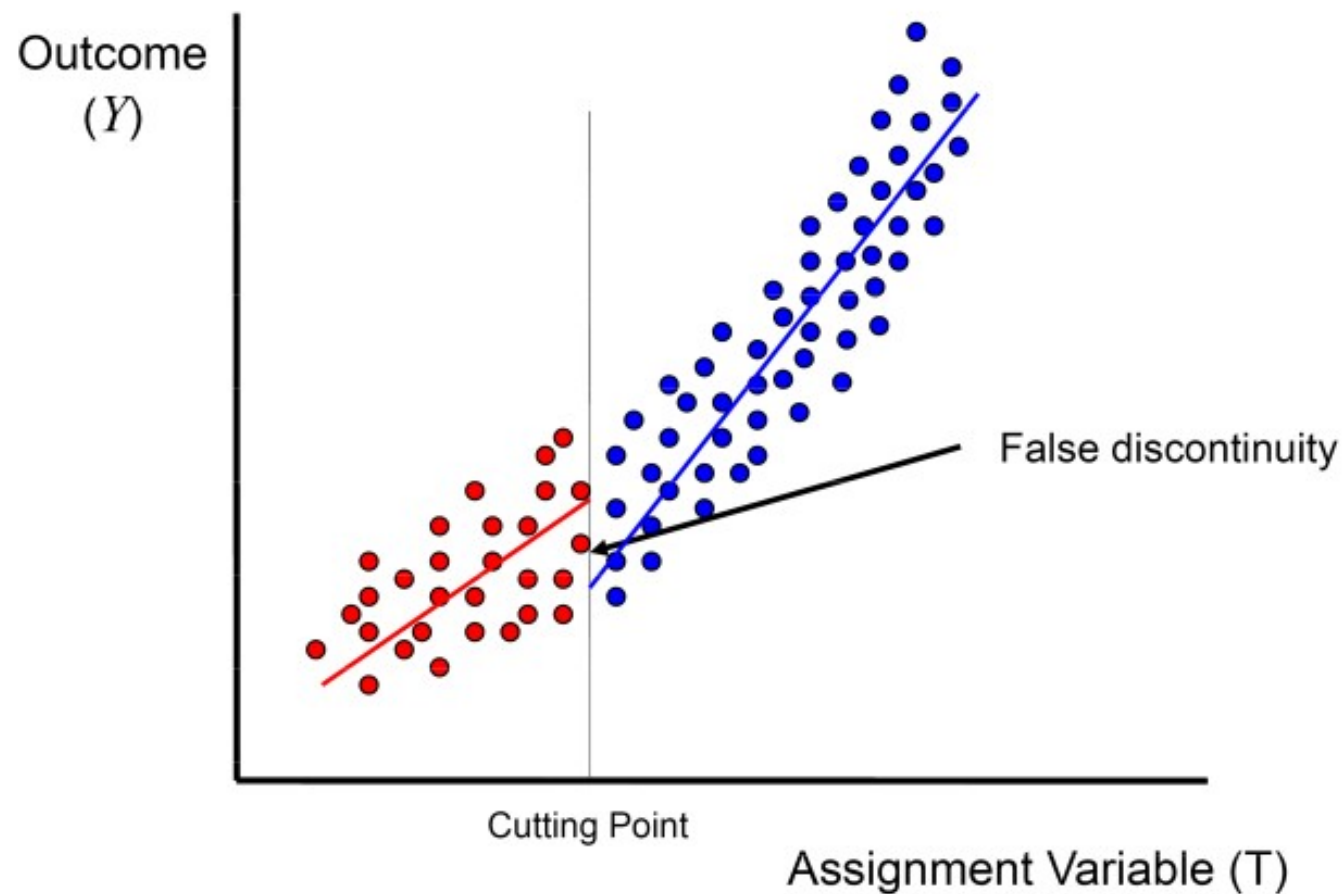
Non-linearity interpreted as a polynomial function (regression)

Outcome (Y)

Cutting Point

Assignment Variable (T)
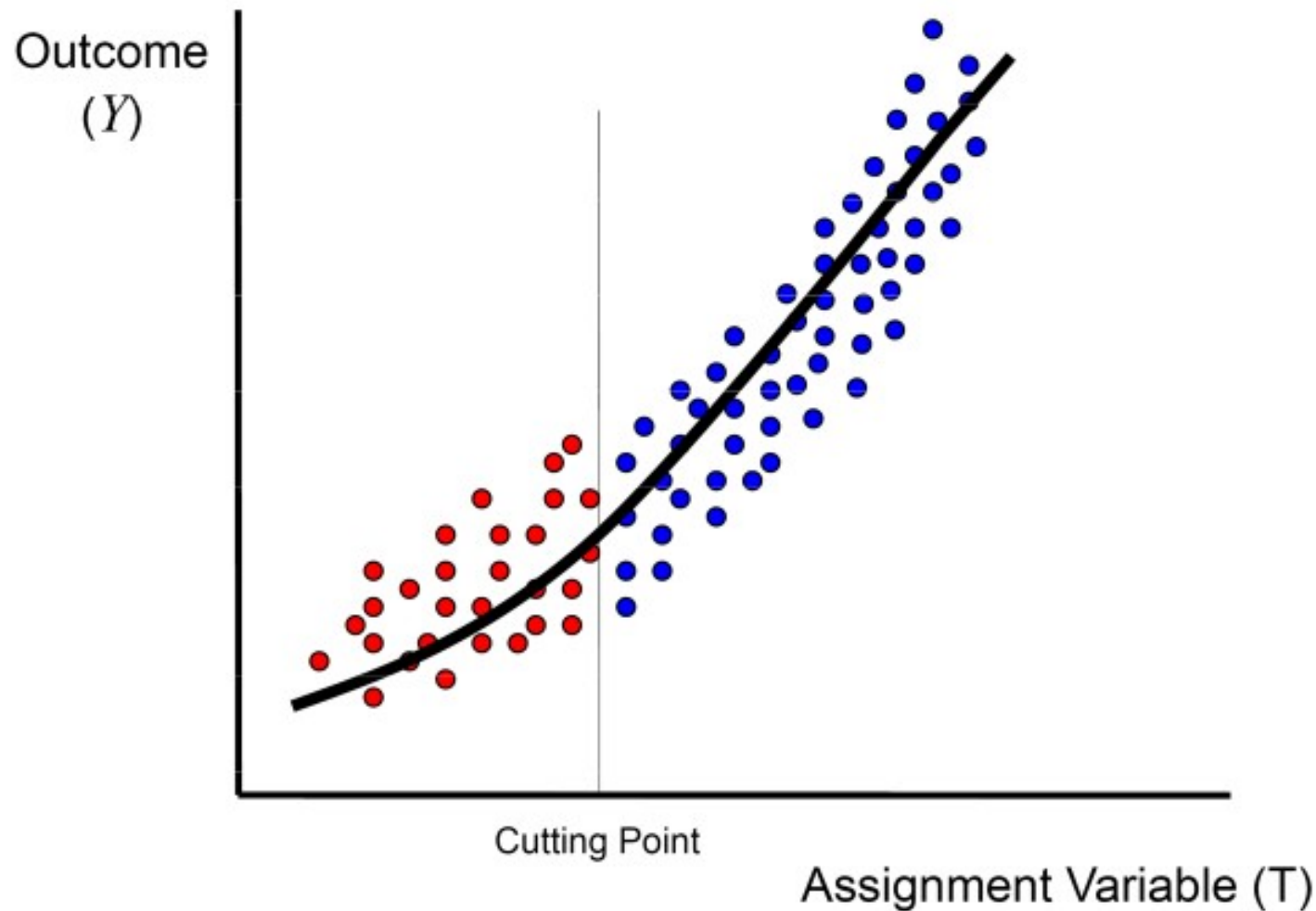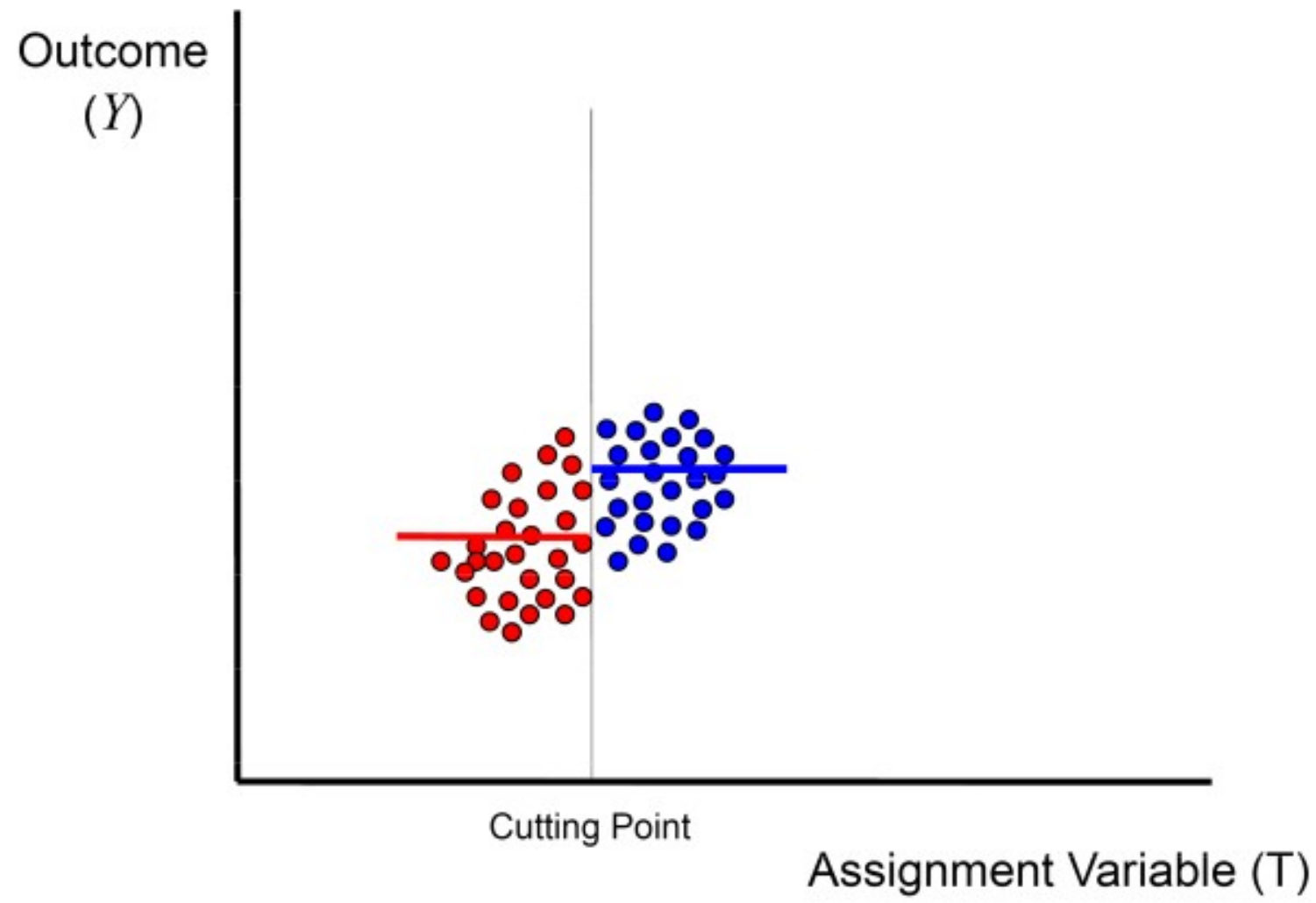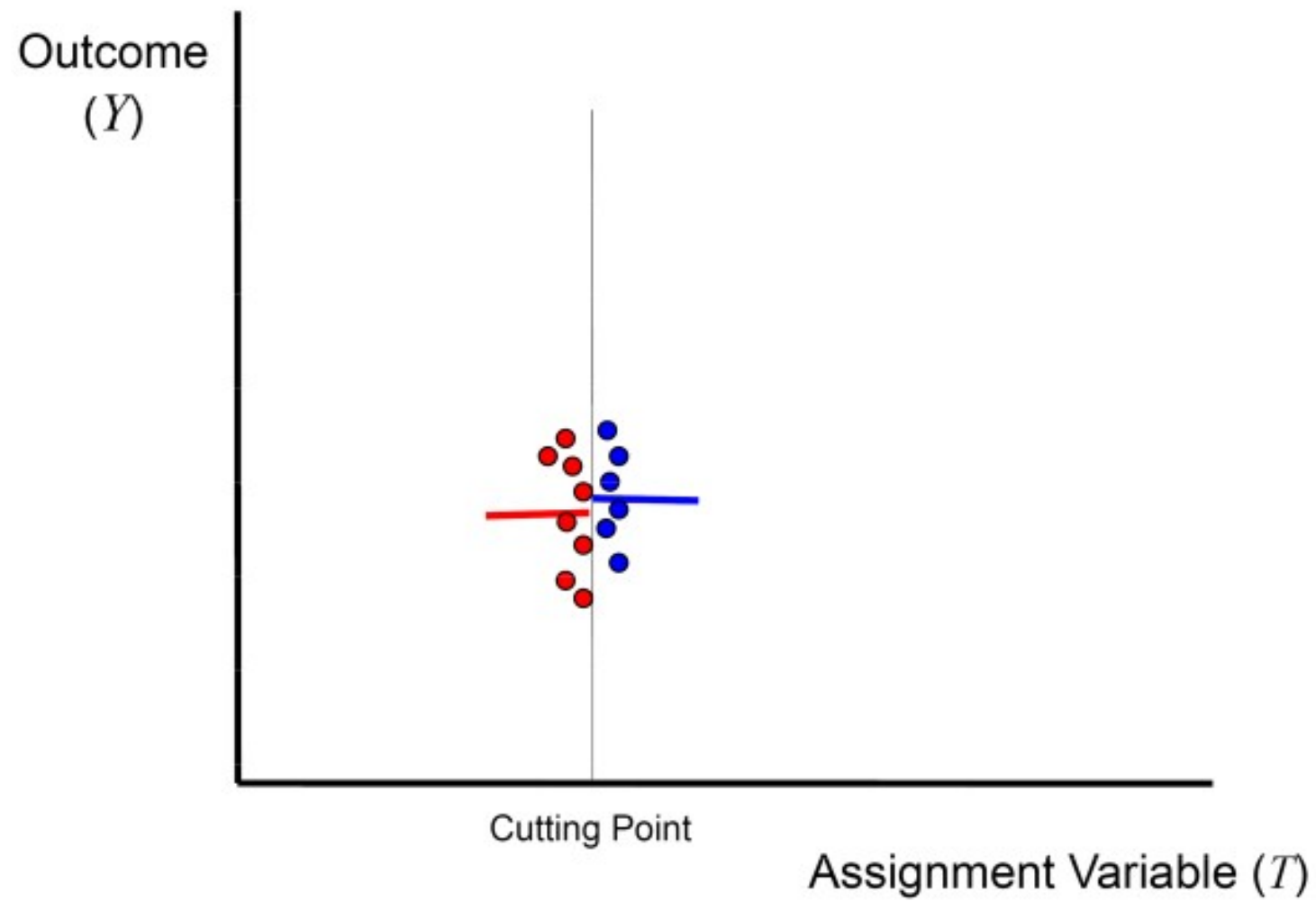
What to show (graphical analysis)

Treatment probability along the threshold

Outcome along the threshold

Check no other covariate also jumps (at this threshold)

Density of the forcing variable at the threshold, check for manipulation (McCrary test, DCdensity in Stata), i.e. violation of hyp. 1

Bandwith settings (h)

Trade off between number of observations and distance from the threshold.

- Trimmed (outliers) cross validation with respect to h

- Imbens & Lemieux 2008 (drop half of the sample)

- simply use sensitivity check to h choice

2 examples

- Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design

Jens Ludwig & Douglas L. Miller, *The Quarterly Journal of Economics*, Volume 122, Issue 1, February 2007, Pages 159–208.

Data & replication files (codes): https://web.archive.org/web/20190619165949/http://faculty.econ.ucdavis.edu:80/faculty/dlmiller/statafiles/

- Do Fiscal Rules Matter?

Veronica Grembi, Tommaso Nannicini and Ugo Troiano, *American Economic Journal: Applied Economics*, Vol. 8, No. 3 (2016), pp. 1-30.

Data & replication files (codes):

https://www.openicpsr.org/openicpsr/project/113637/version/V1/view?path=/openicpsr/113637/fcr:versions/V1/APP2015-0076_data&type=folder

Ludwig & Miller, 2007

Head Start: grant for poor children in preschool years

* Parent involvement

* Nutrition

* Social services

* Health & mental health service

Spring 1965: technical assistance for the 300 poorest counties ($P_c < P_{300}$) of year 1960
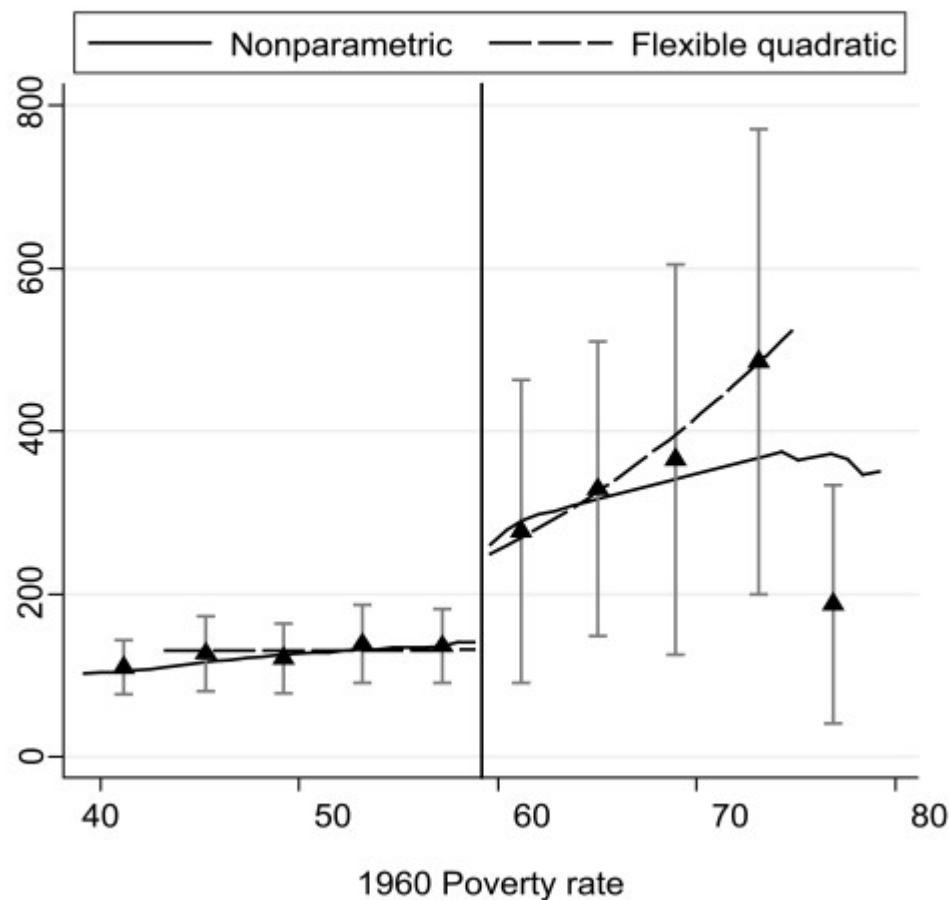
→ 50-100% higher funding rates

→ **ITT** (RDD) estimation, local linear regressions [Fan 1992] to estimate the left and right limits of the discontinuity, where the difference between the two is the estimated treatment impact.
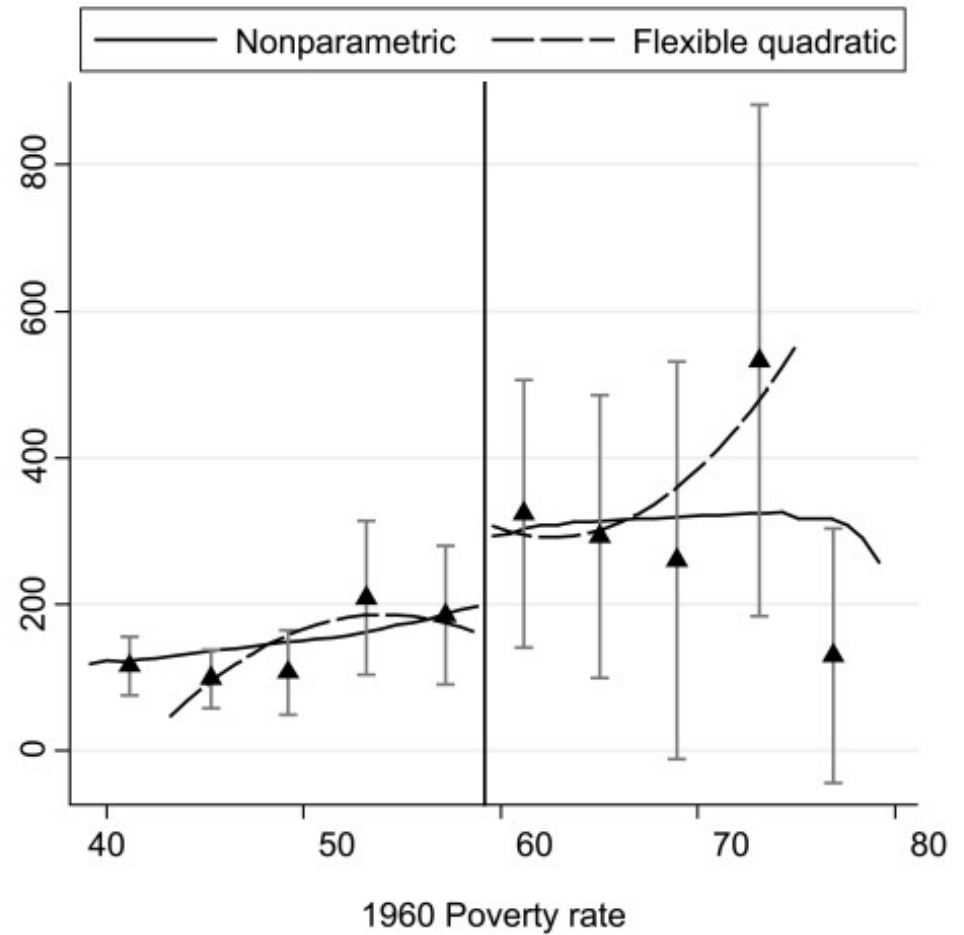
Kernel weights $w_c = K((P_c - P_{300})/h)$ for chosen bandwidth h

$$Y_c = b_0 + b_1.(P_c - P_{300}) + \alpha.G_C + b_2.G_c.(P_c - P_{300}) + \varepsilon_c .$$

Panel A: 1968 Head Start funding per 4 year old   Panel B: 1972 Head Start funding per 4 year old
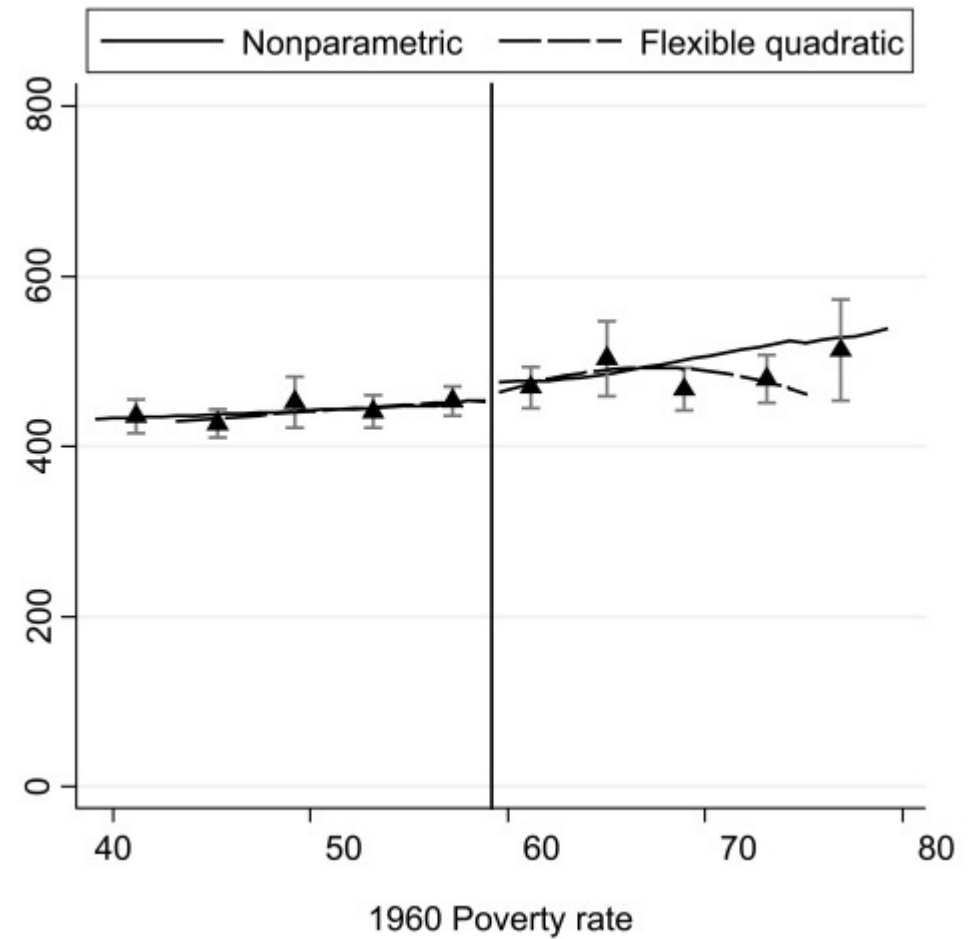
REGRESSION DISCONTINUITY ESTIMATES OF THE EFFECT OF HEAD START ASSISTANCE ON HEAD START SPENDING AND PARTICIPATION

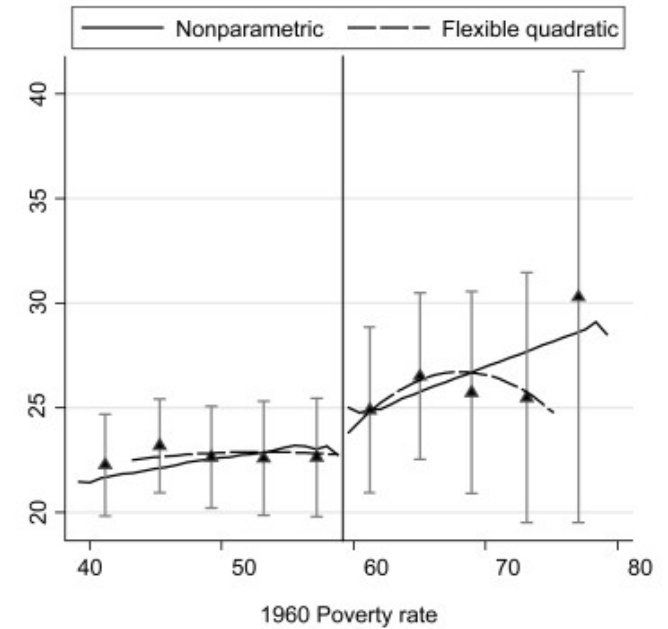| | | | | | Parametric | |
| | | | | | Flexible linear | Flexible quadratic |
Variable	Control mean	Nonparametric				
Results from the National Education Longitudinal Study						
Bandwidth or poverty range		9	18	36	8	16
Number of observations (counties) with nonzero weight		43	96	288	43	82
Head Start participation (base year sample)	0.259	0.168	0.135	0.126	0.168	0.217
		(0.142)	(0.094)	(0.056)	(0.148)	(0.142)
		[0.317]	[0.129]	[0.113]	[0.277]	[0.216]
Head Start participation (first year follow-up sample)	0.286	0.238	0.172*	0.148*	0.238	0.316*
		(0.150)	(0.100)	(0.061)	(0.156)	(0.151)
		[0.189]	[0.090]	[0.089]	[0.154]	[0.087]
Results from county-level federal spending data						
Bandwidth or poverty range		9	18	36	8	16
Number of observations (counties) with nonzero weight		527	961	2,177	484	863
1968 Head Start spending per child	137.505	137.251	114.711	134.491**	127.389	112.706
		(128.968)	(91.267)	(62.593)	(120.098)	(112.603)
		[0.157]	[0.138]	[0.045]	[0.225]	[0.297]
1972 Head Start spending per child	198.396	182.119*	88.959	130.153*	175.773	155.899
		(148.321)	(101.697)	(67.613)	(142.363)	(132.292)
		[0.085]	[0.352]	[0.090]	[0.109]	[0.152]
1972 other social spending, per capita	452.384	14.474	19.590	14.506	−2.361	5.659
		(28.356)	(19.612)	(14.929)	(25.624)	(27.471)
		[0.459]	[0.222]	[0.478]	[0.979]	[0.755]

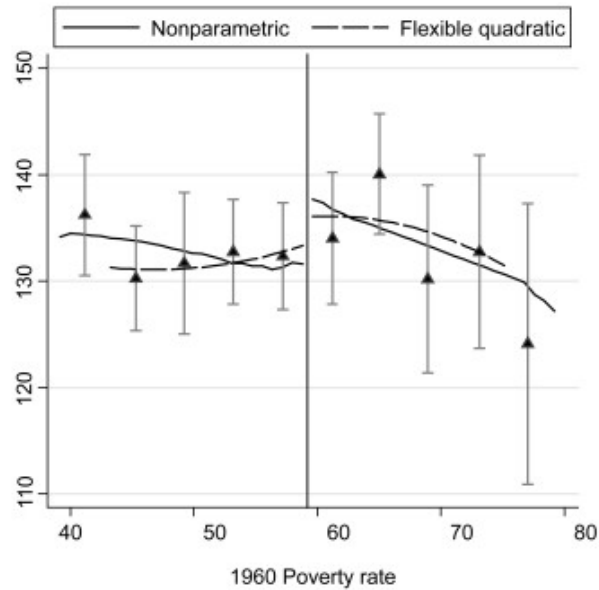# Other federal spending

Why checking this ?

& this ?

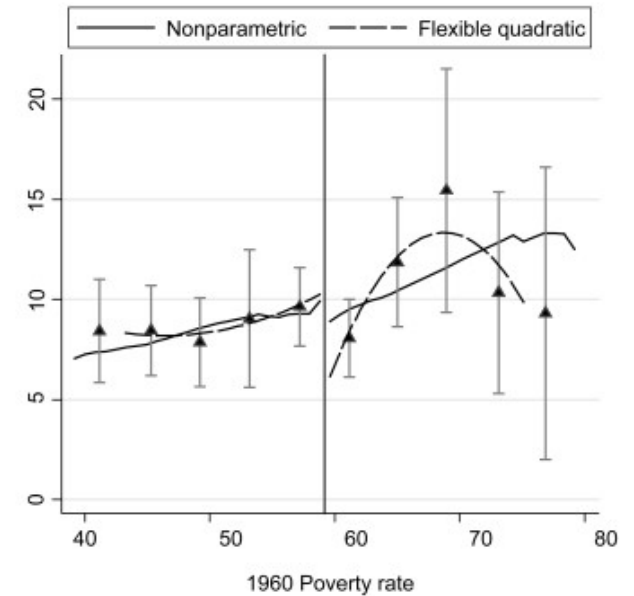Panel A: Children 5-9, Head Start susceptible causes, 1973-83
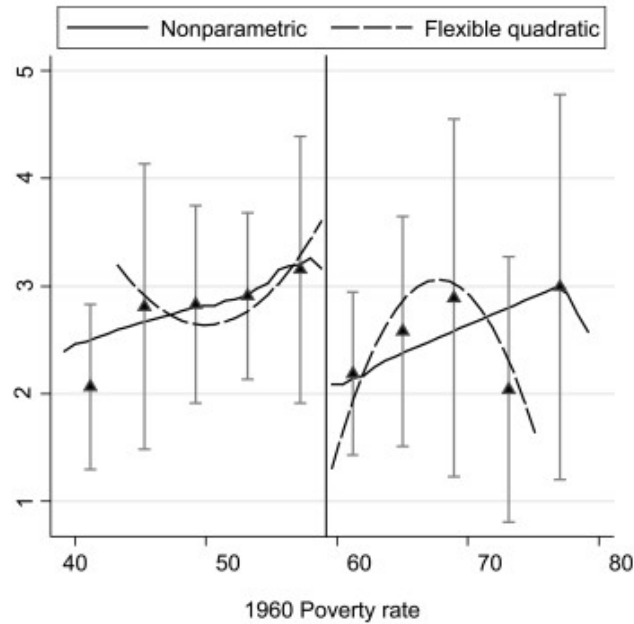
Panel B: Children 5-9, Injuries, 1973-83

Panel C: Adults 25+, Head Start susceptible causes, 1973-83

Panel D: Children 5-9, Head Start susceptible causes, 1959-64

REGRESSION DISCONTINUITY ESTIMATES OF THE EFFECT OF HEAD START ASSISTANCE ON MORTALITY

| | | | | | Parametric | |
| | | | | | Flexible linear | Flexible quadratic |
Variable	Control mean	Nonparametric estimator				
Bandwidth or poverty range		9	18	36	8	16
Number of observations (counties) with nonzero weight		527	961	2,177	484	863
Main results						
Ages 5–9, Head Start-related causes, 1973–1983	3.238	−1.895**	−1.198*	−1.114**	−2.201**	−2.558**
		(0.980)	(0.796)	(0.544)	(1.004)	(1.261)
		[0.036]	[0.081]	[0.027]	[0.022]	[0.021]
Specification checks						
Ages 5–9, injuries, 1973–1983	22.303	0.195	2.426	0.679	−0.164	0.775
		(3.472)	(2.476)	(1.785)	(3.380)	(3.401)
		[0.924]	[0.345]	[0.755]	[0.998]	[0.835]
Ages 5–9, all causes, 1973–1983	40.232	−3.416	0.053	−1.537	−3.896	−2.927
		(4.311)	(3.098)	(2.253)	(4.268)	(4.295)
		[0.415]	[0.982]	[0.558]	[0.317]	[0.505]
Ages 25+, Head Start-related causes, 1973–1983	131.825	2.204	6.016	5.872	2.091	2.574
		(5.719)	(4.349)	(3.338)	(5.581)	(6.415)
		[0.700]	[0.147]	[0.114]	[0.749]	[0.689]

Cours RDD dispos en ligne

De Chaisemartin:

https://www.parisschoolofeconomics.eu/docs/de-chaisemartin-clement/natural_experiments_rdd.pdf

Chabé-ferret:

https://drive.google.com/file/d/1eN9HNx2p2WQSEe3kDmn_HT3NXUNVDMEU/view

Conclusion

Regression discontinuity is a very popular and useful way to estimate causal impacts

• Often discontinuity could be seen as an instrumental variable, but works well as a reduced form analysis

• Quite a lot of parameters to choose (e.g. polynomial order, bandwidth), but at least relatively easy to check robustness

How to design a RCT / survey

RCT = Random selection of treatment (solve selection bias)

Compute minimal detectable effect (MDE): power calculation using the baseline study (survey data before) but specific cases:

What is the treatment level (individual, village), ==> clusters

- within groups ?

- compliance in encouragement designs

Common criticism (to RCTs)

It is costly (once dropping the policy implementation, which should be considered as such, randomization is not)

It is unfair since it should prevent a (significant) share of the population targeted from receiving treatment

It often suffers from similar issues than non random policy allocation (spatial/network spillovers…)

Limitations & critics

We have seen how **design-based strategies** that rely on **internal validity** help to deductively identify **causal effects**

Limited **external validity** (not specific to causal identification) ➞

* Local effects, lack of generalization

* Conditions required for the cause to occur

* Data quality (simple but often lacking)

Answer: what external validity without internal validity ?

Stata commands: introduction

preserve

collapse

set seed

by: sum

ttest

reg

nnmatch

psmatch2

foreach x of {

}

Coding (concepts)

Loop

Global and local variables

Local path (to file, through folders: /.../...)

Set memory, log

Import, ssc install