Lecture 7: Quasi-experimental design

Difference-in-Difference

Xi Chen

Rotterdam School of Management
Erasmus University Rotterdam

June 9, 2023

Outline

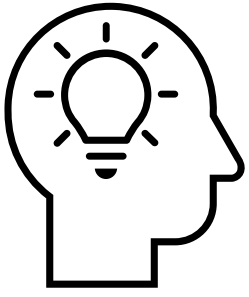
RSM Erasmus

Outline

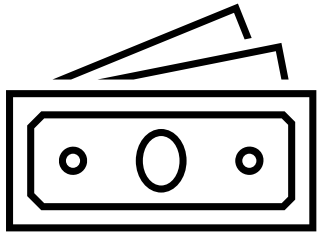
RSM Erasmus

Randomized experiments are powerful, but...

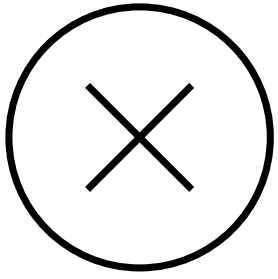
In many scenarios, we cannot do randomized experiments.



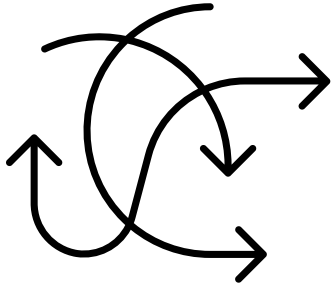
Unethical



Costly



Physically Impossible



Infeasible

Quasi-experiments

Definition (Quasi-experiments)

Research designs that resemble experimental designs but lack full random assignment of participants to groups.

Different designs of quasi-experiments

- Interrupted time-series design
- Event studies
- Case-control design
- **Difference-in-difference (DID)**
- **Regression discontinuity design (RDD)**
- ...

Quasi-experiments

Definition (Quasi-experiments)

Research designs that resemble experimental designs but lack full random assignment of participants to groups.

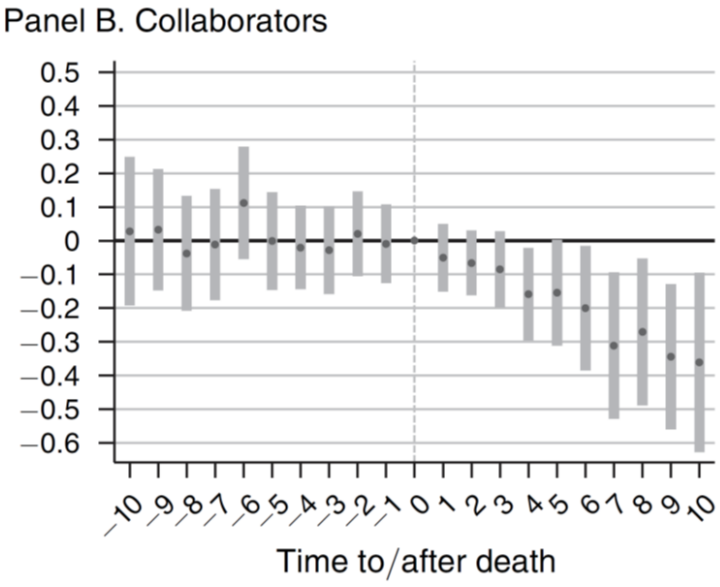
Different designs of quasi-experiments

- Interrupted time-series design
- Event studies
- Case-control design
- **Difference-in-difference (DID)**
- **Regression discontinuity design (RDD)**
- ...

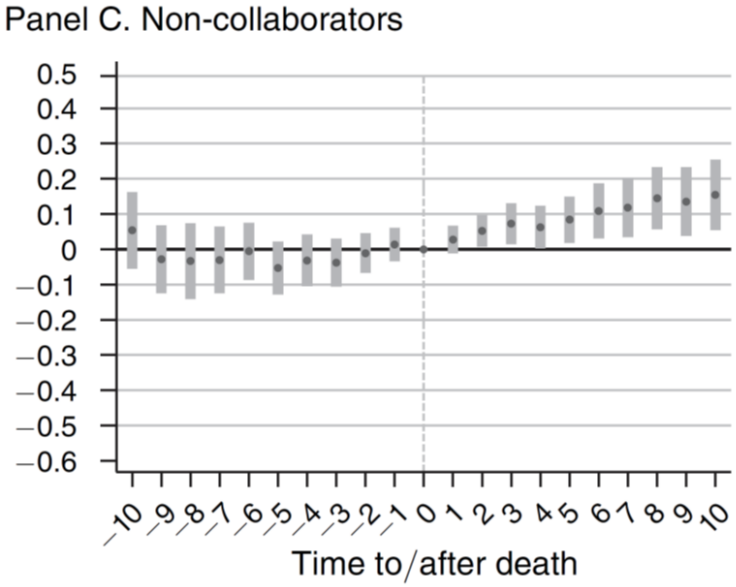
Numerous examples of quasi-experiments...

Azoulay et al. (2019) investigated how premature death of eminent life scientists alter the vitality of their fields.

- The flow of articles by collaborators into affected fields decreases after the death of a star scientist.
- The flow of articles by non-collaborators increases markedly.



(a) Collaborators

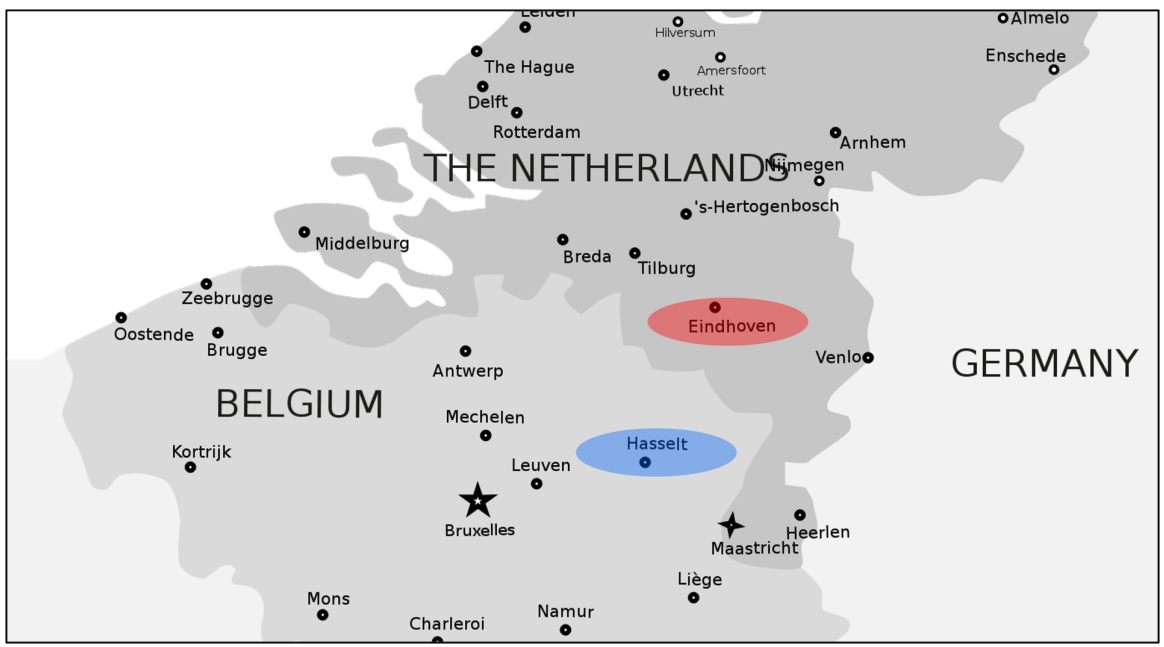


(b) Non-collaborators

Numerous examples of quasi-experiments...

Bol.com - a Dutch e-retailer of consumer electronics - examined the effect of display ads on sales.

- They compared two cities: one in Holland and one in Belgium.
- They stopped the display ads in the Belgian city.



RSM Erasmus

Outline

RSM Erasmus

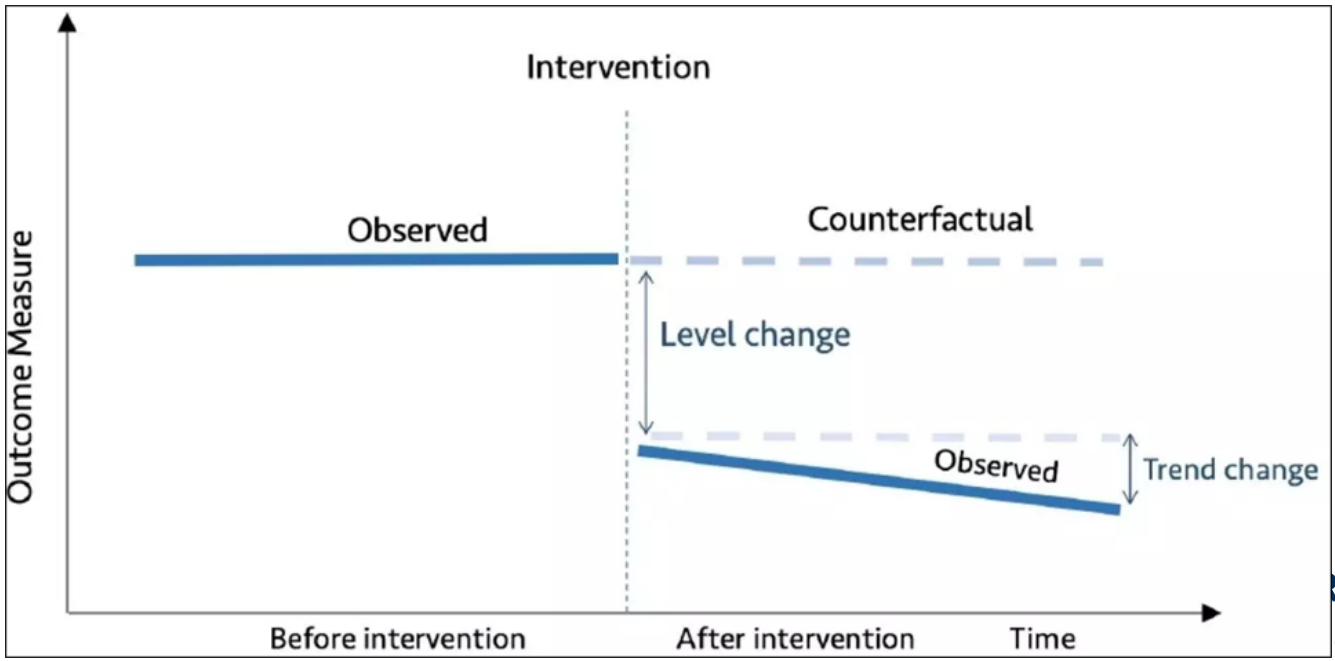
Outline

RSM Erasmus

Before-after design

In practice, we observe an event happens at certain time point.

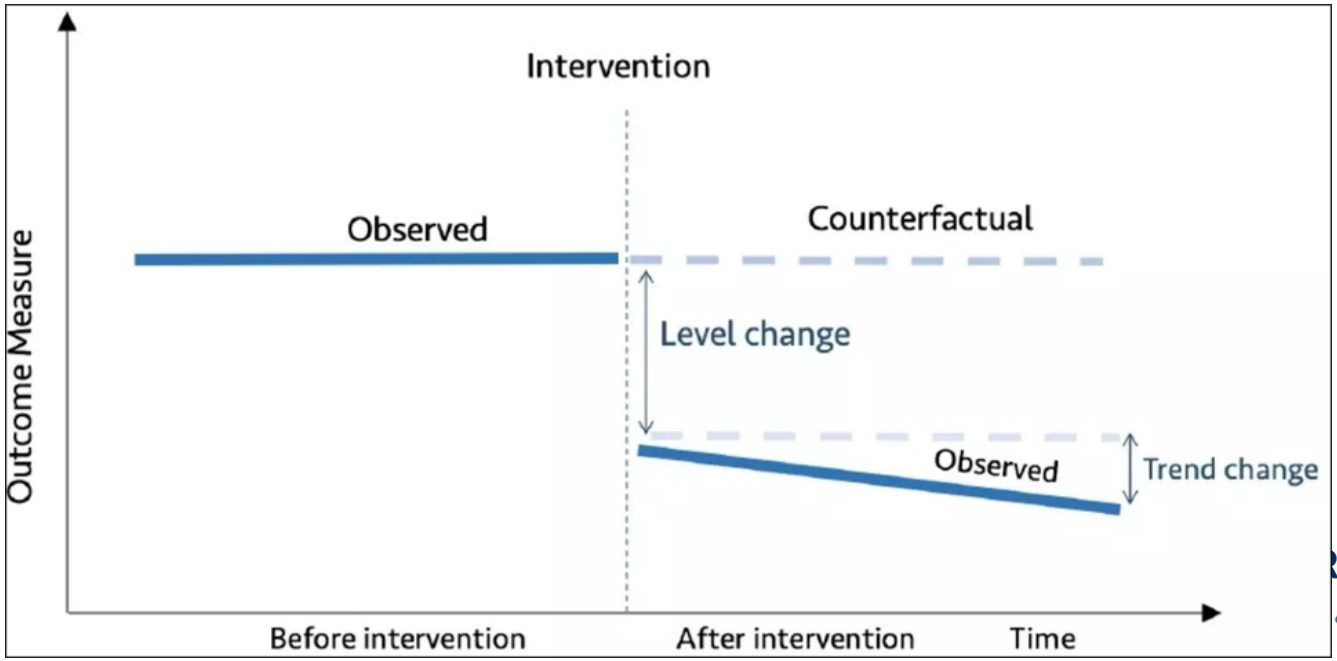
- Predict the "after" outcome based on the before data.
- Compare the "after" outcome with the predicted one.
  - **Level changes**: up and down shift.
  - **Trend changes**: changes in the trajectory.



Before-after design

In practice, we observe an event happens at certain time point.

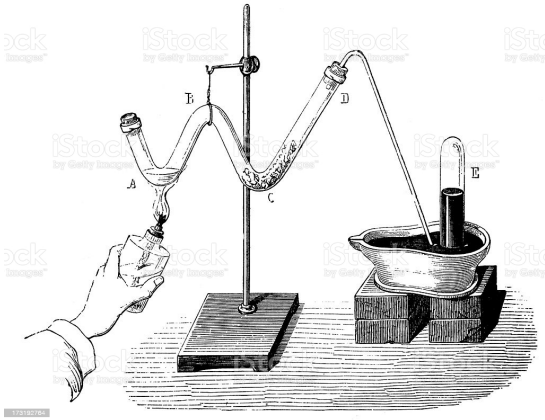
- German reunification
- The implementation of GDPR
- The Great Recession in 2008
- ...



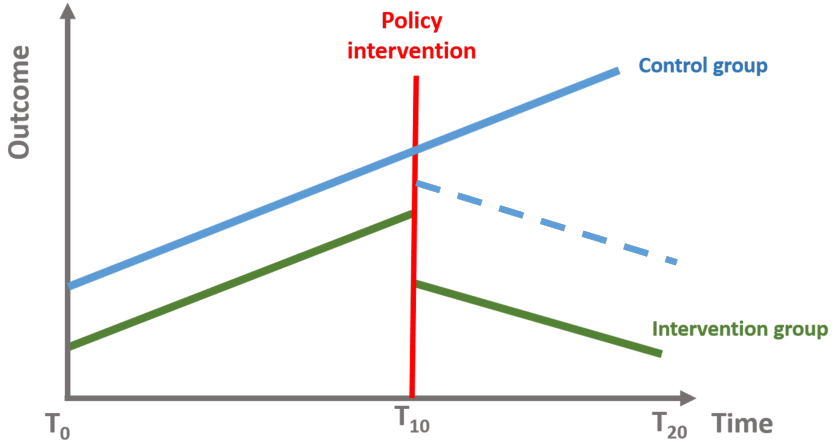
Before-after design

The use of "before" to predict the counterfactual of "after?"

For some scenarios, it's credible:

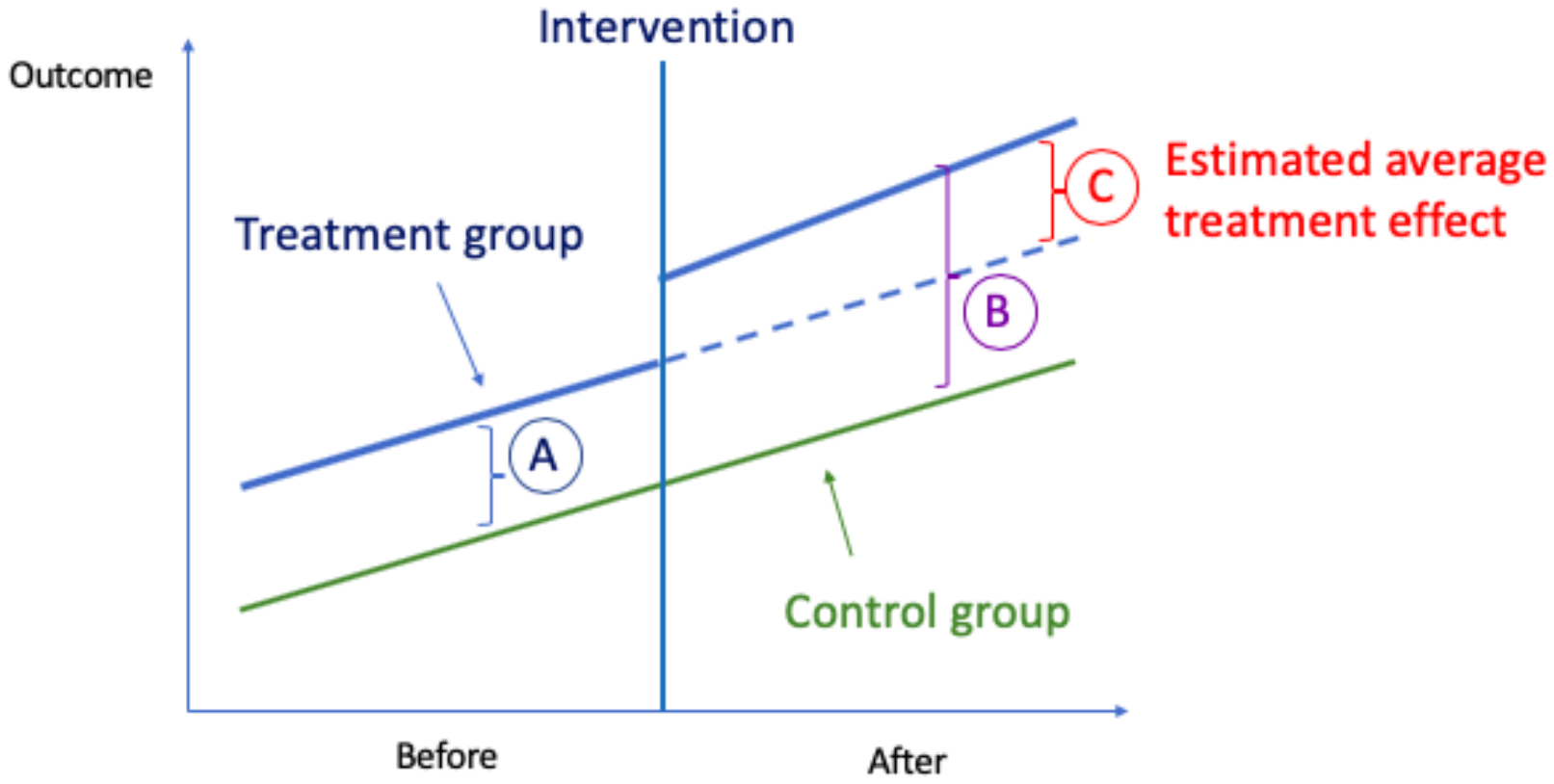


What if the counterfactual also goes through a "regime shift?"



RSM Erasmus

Two sources of variations



- **Time variation**: Before vs. after the intervention.
- **Treatment variation**: Treatment vs. control group.

RSM Erasmus

Outline

RSM Erasmus

The canonical DID: basic setup

- N units (i) across $T = 2$ time periods (t).
- A binary treatment $D_{it} = \{0, 1\}$.
- Outcome Y_{it}.

- Treatment happens in period 2 $(t = 2)$.
- A group of always untreated units $(D_{it} = 0, \forall t)$

Potential outcomes

2 time periods $\times$ 2 treatment status $=$ 4 potential outcomes.

$Y_{it}(D_i)$	Time 1	Time 2
Untreated	$Y_{i1}(0)$	$Y_{i2}(0)$
Treated	$Y_{i1}(1)$	$Y_{i2}(1)$

The canonical DID: potential outcomes

Observed vs. unobserved potential outcomes.

$Y_{it}(D_i)$	Time 1	Time 2
Untreated	$Y_{i1}(0)$	~~$Y_{i2}(0)$~~
Treated	~~$Y_{i1}(1)$~~	$Y_{i2}(1)$

(a) Treatment Group

$Y_{it}(D_i)$	Time 1	Time 2
Untreated	$Y_{i1}(0)$	$Y_{i2}(0)$
Treated	~~$Y_{i1}(1)$~~	~~$Y_{i2}(1)$~~

(b) Control Group

RSM Erasmus

The canonical DID: the target of identification

- **Primary interests**: the treatment effects of the intervention ($t = 2$).

$$\tau_{\text{ATE}} = E\left(Y_{i2}\left(1\right) - Y_{i2}\left(0\right)\right)$$

- But, we only observe:

$$\begin{cases} E\left(Y_{i2}\left(1\right) \mid D_{i2} = 1\right) & \text{For treatment group} \\ E\left(Y_{i2}\left(0\right) \mid D_{i2} = 0\right) & \text{For control group} \end{cases}$$

- Assuming **unconfoundedness**, we can identify τ_{ATE}:

$$\begin{cases} E\left(Y_{i2}\left(1\right) \mid D_{i2} = 1\right) & = E\left(Y_{i2}\left(1\right)\right) \\ E\left(Y_{i2}\left(0\right) \mid D_{i2} = 0\right) & = E\left(Y_{i2}\left(0\right)\right) \end{cases}$$

RSM Erasmus

The canonical DID: the target of identification

- The unconfoundedness assumption is unrealistic for quasi-experiments.
- By definition, the treatment is not randomly assigned!

The average treatment effect on the treated (ATT)

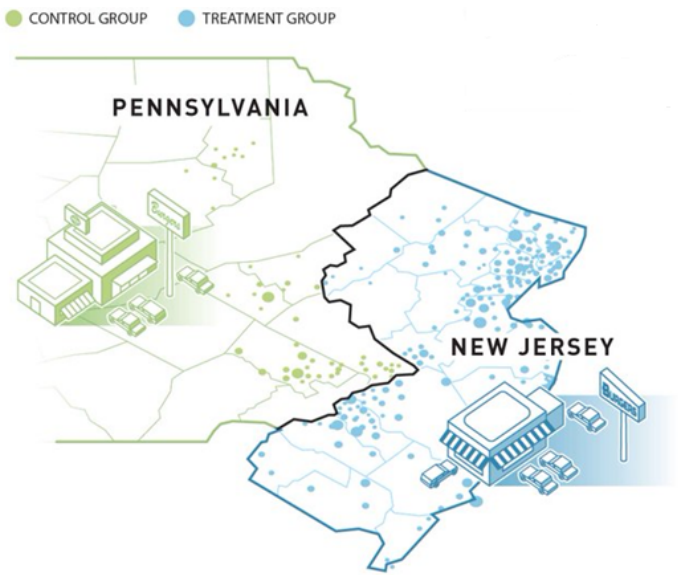
Researchers or policymakers may be more interested in ATT, because:

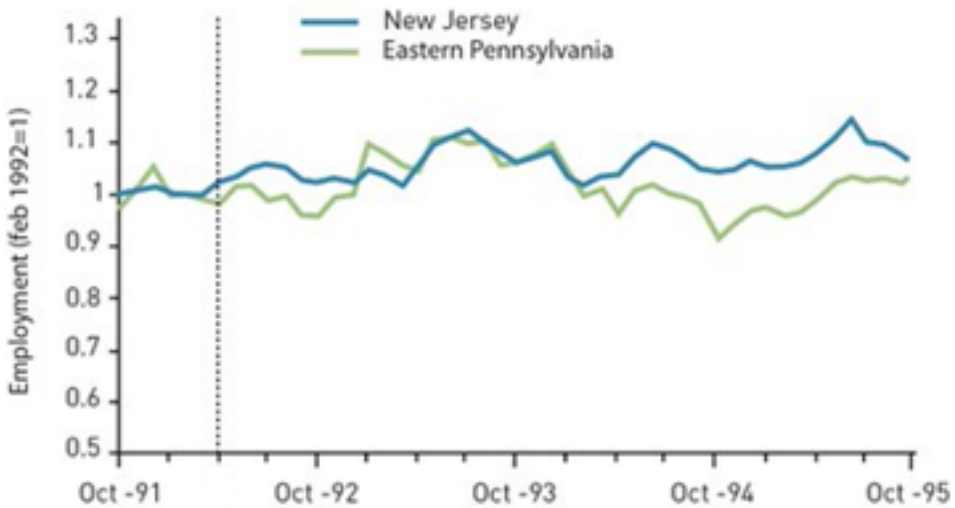
1. It is uncontrollable who will be in the treatment;
2. Participants of the treatment are more relevant for the policy.

RSM Erasmus

The canonical DID: examples

The influence of minimum wage on employment from Card and Krueger (1994)[1].





[1] Figures are accredited to Jahan Jarnestad at the Royal Swedish Academy of Sciences.

The canonical DID: examples

Snapchat did a lot of field experiments of the app features and recommendation algorithms.

- They mainly focused on the US market.
- Always used New Zealand as the control group.

The canonical DID: identification assumptions

The target is ATT: $\tau_{\text{ATT}} = E\left[Y_{i2}\left(1\right) - Y_{i2}\left(0\right) \mid D_{i2} = 1\right]$

- We observe $E\left[Y_{i2}\left(1\right) \mid D_{i2} = 1\right]$ for the treatment group.
- But, we do not observe $E\left[Y_{i2}\left(0\right) \mid D_{i2} = 1\right]$.

One assumption to make is **ignorability**:

$$E\left[Y_{i2}\left(0\right) \mid D_{i2} = 1\right] = E\left[Y_{i2}\left(0\right) \mid D_{i2} = 0\right]$$

- The treatment status is independent of $Y_{i2}\left(0\right)$.
- Yet, this is still a strong assumption.
- People do not participate because they see little benefits.

RSM Erasmus

The canonical DID: identification assumptions

It turns out we need an assumption than ignorability.

The Parallel Trend Assumption

The before-after difference of the potential outcomes under no treatment is ignorable:

$$\underbrace{E\left[Y_{i2}\left(0\right)-Y_{i1}\left(0\right)\mid D_{i2}=1\right]}_{\text{Before-after differences of treatment group}} = \underbrace{E\left[Y_{i2}\left(0\right)-Y_{i1}\left(0\right)\mid D_{i2}=0\right]}_{\text{Before-after differences of control group}}$$

Under the the parallel trend assumption, **we can identify the ATT of the treatment at** $t=2$!

RSM Erasmus

The canonical DID: proof of identification results

Technical proof:
If we rearrange the parallel trend assumption, we have:

$$\begin{aligned} &E\left[Y_{i2}\left(0\right) \mid D_{i2}=1\right]-E\left[Y_{i1}\left(0\right) \mid D_{i2}=1\right] \\ &-E\left[Y_{i2}\left(0\right) \mid D_{i2}=0\right]+E\left[Y_{i1}\left(0\right) \mid D_{i2}=0\right]=0 \end{aligned} \tag{1}$$

The ATT is:

$$\tau_{\text{ATT}}=E\left[Y_{i2}\left(1\right) \mid D_{i2}=1\right]-E\left[Y_{i2}\left(0\right) \mid D_{i2}=1\right]$$

If we add equation (1) to the ATT formula, we have:

$$\begin{aligned} \tau_{\text{ATT}}=&\left(E\left[Y_{i2}\left(1\right) \mid D_{i2}=1\right]-E\left[Y_{i1}\left(0\right) \mid D_{i2}=1\right]\right) \\ &-\left(E\left[Y_{i2}\left(0\right) \mid D_{i2}=0\right]-E\left[Y_{i1}\left(0\right) \mid D_{i2}=0\right]\right) \end{aligned} \tag{2}$$

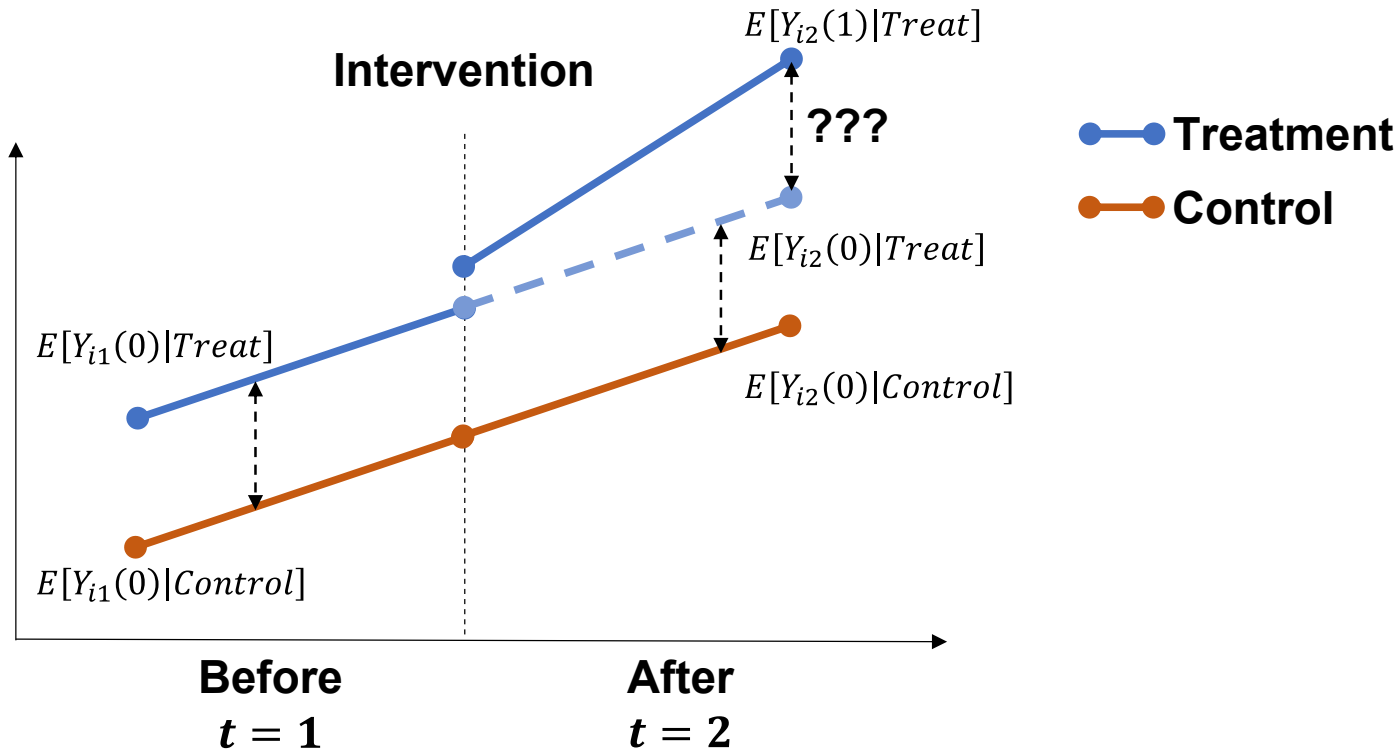
All the terms in equation (2) can be inferred from the observation.

RSM Erasmus

Outline

RSM Erasmus

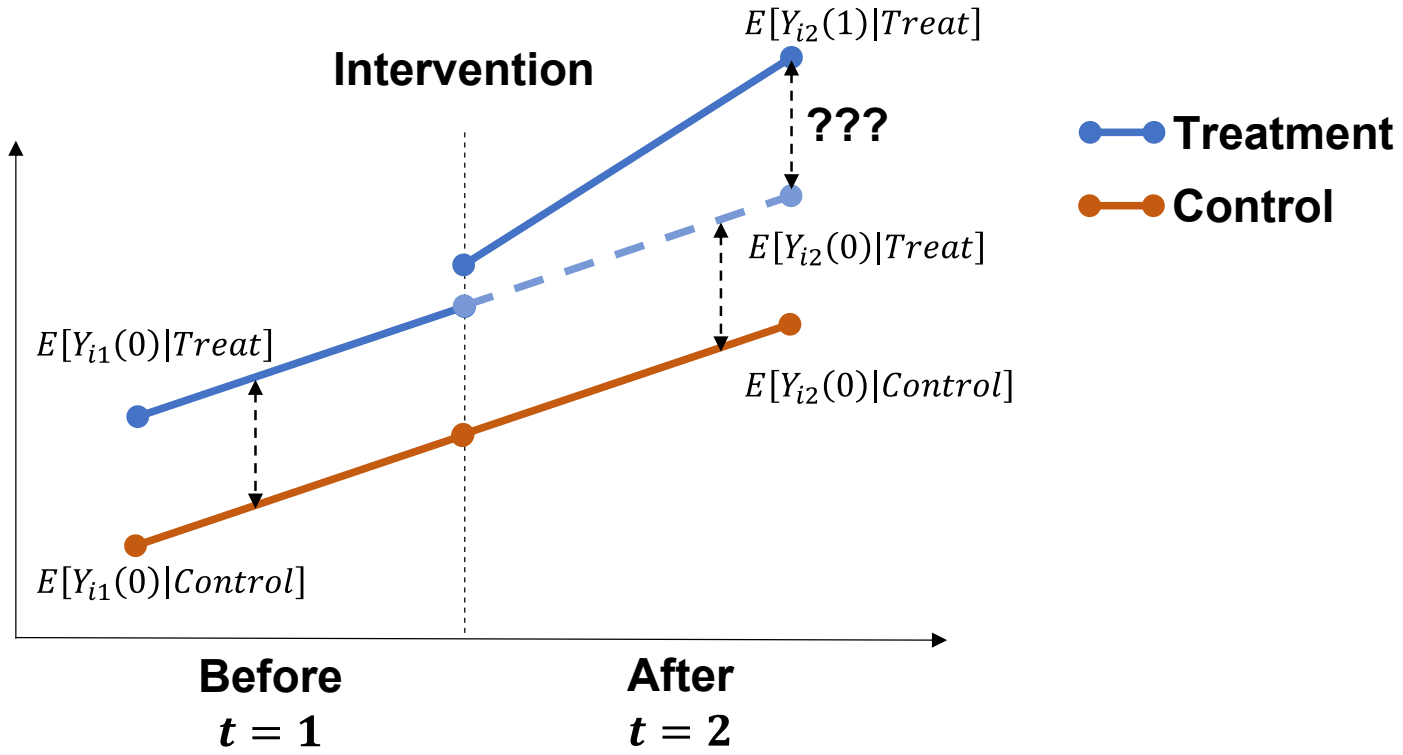
Understanding the parallel trend



We need to calculate: $\tau_{\text{ATT}} = E\left[Y_{i2}\left(1\right) \mid \text{Treat}\right] - \cancel{E\left[Y_{i2}\left(0\right) \mid \text{Treat}\right]}$.

RSM Erasmus

Understanding the parallel trend

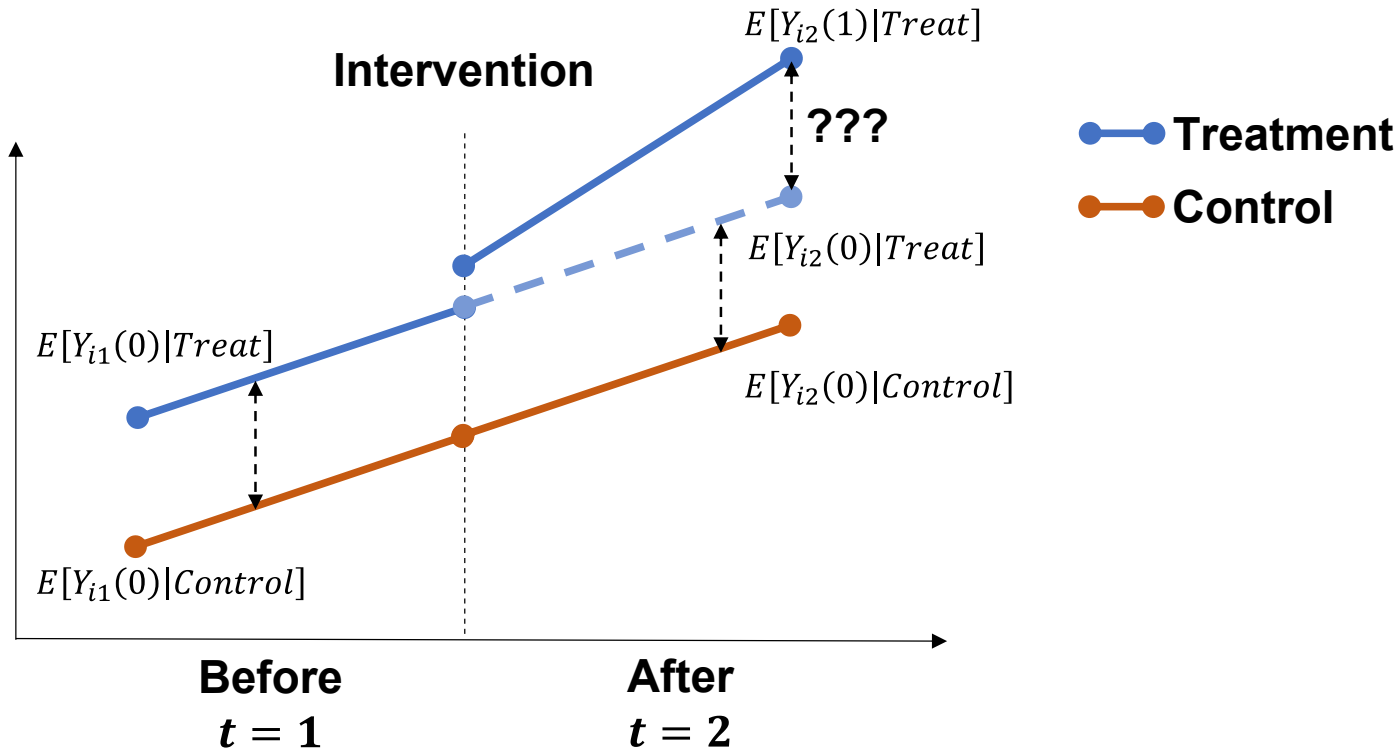


Counterfactual language: in the absence of the treatment, the difference between the treated and control group is constant over time.

$$E\left[Y_{i1}\left(0\right) \mid \text{Treat}\right] - E\left[Y_{i1}\left(0\right) \mid \text{Control}\right] = \cancel{E\left[Y_{i2}\left(0\right) \mid \text{Treat}\right]} - E\left[Y_{i2}\left(0\right) \mid \text{Control}\right]$$

RSM

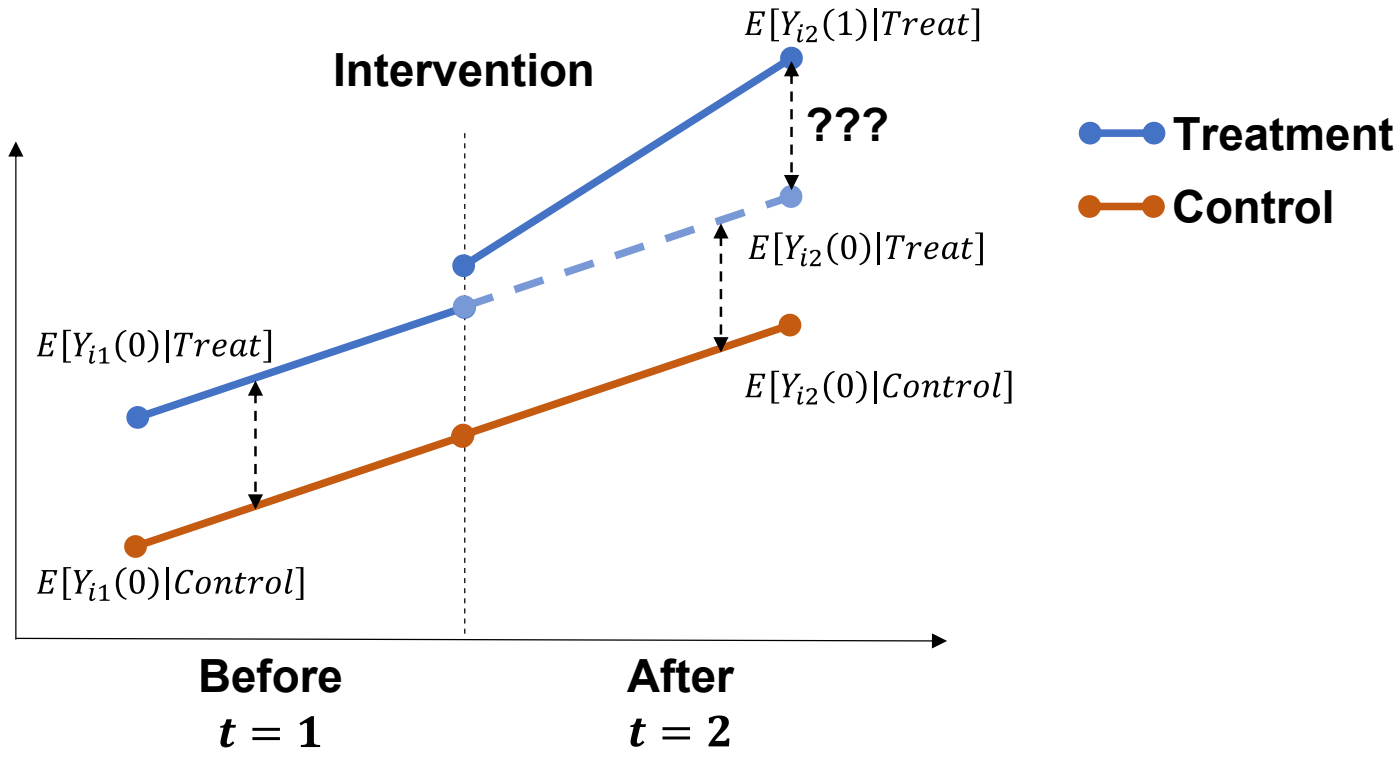
Understanding the parallel trend



Assume the parallel trend, we can solve for $E[Y_{i2}(0) \mid \text{Treat}]$ with:

$$E[Y_{i1}(0) \mid \text{Treat}] - E[Y_{i1}(0) \mid \text{Control}] = \underbrace{E[Y_{i2}(0) \mid \text{Treat}]}_{\text{Solve for this term}} - E[Y_{i2}(0) \mid \text{Control}]$$

Parallel trend assumption is untestable



- It's an assumption on an unobserved potential outcome $E[Y_{i2}(0) \mid \text{Treat}]$.
- The fundamental problem of causal inference $\rightarrow$ it's untestable!

Parallel trend assumption is scale dependent

The parallel trend holds for Y, but may not for a monotonic transformation of Y.

$Y_{it}(D_i)$	Time 1	Time 2
Untreated	$Y_{i1}(0) = 1$	$Y_{i2}(0) = 3$
Treated	$Y_{i1}(0) = 2$	$Y_{i2}(0) = 4$

(c) Original Values: $2 - 1 = 4 - 3$

$Y_{it}(D_i)$	Time 1	Time 2
Untreated	$log(Y_{i1}(0)) = log(1)$	$log(Y_{i2}(0)) = log(3)$
Treated	$log(Y_{i1}(0)) = log(2)$	$log(Y_{i2}(0)) = log(4)$

(d) Log-transformed Values: $log(2) \neq log\left(\frac{4}{3}\right)$

RSM Erasmus

Outline

RSM Erasmus

Outline

RSM Erasmus

Non-parametric estimation

From the identification results on page 24

$$\begin{aligned}\tau_{\text{ATT}} = &\left(E\left[Y_{i2}\left(1\right) \mid \text{Treat}\right] - E\left[Y_{i1}\left(0\right) \mid \text{Treat}\right]\right) \\ &- \left(E\left[Y_{i2}\left(0\right) \mid \text{Control}\right] - E\left[Y_{i1}\left(0\right) \mid \text{Control}\right]\right),\end{aligned}$$

A natural non-parametric estimator is

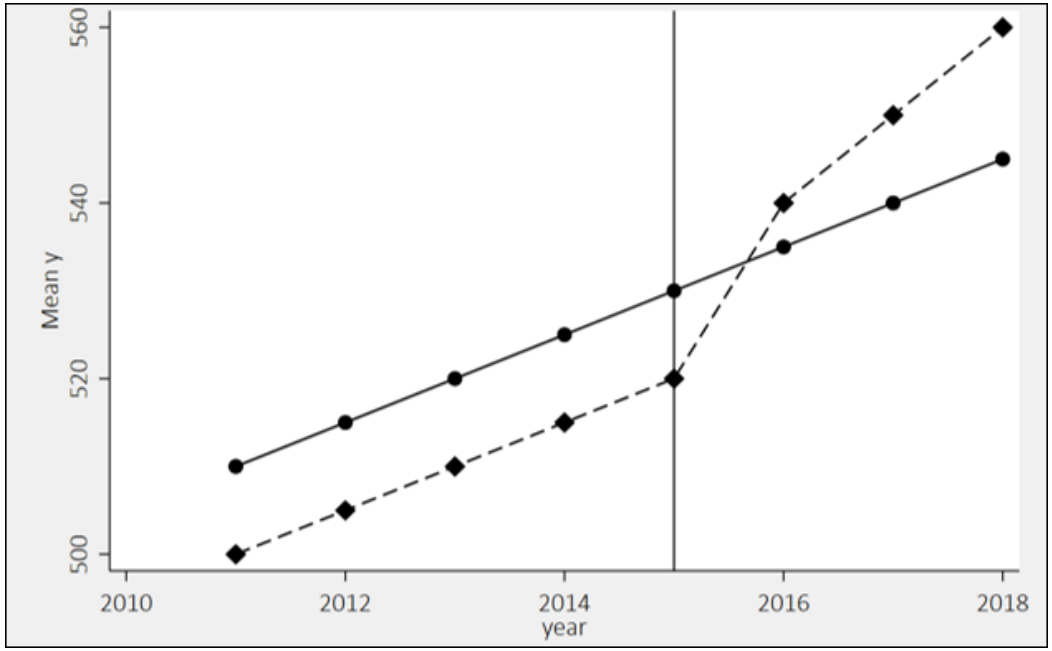
$$\begin{aligned}\hat{\tau}_{\text{ATT}} = &\left(\frac{1}{N_1}\sum_{i\in\text{Treat}} Y_{i2} - \frac{1}{N_1}\sum_{i\in\text{Treat}} Y_{i1}\right), \\ &- \left(\frac{1}{N_0}\sum_{i\in\text{Control}} Y_{i2} - \frac{1}{N_0}\sum_{i\in\text{Control}} Y_{i1}\right)\end{aligned}$$

Two differences $\Rightarrow$ "Difference-in-difference"

RSM Erasmus

Multiple time periods

- Average the time periods and get 4 outcomes (treated vs. control and before vs. after).
  - Not efficient as aggregation loses information.
- **Model-based estimation**: two-way fixed effects regression.



Model-based estimation

A two-way fixed effects regressioin:

$$y_{it} = \alpha_i + \lambda_t + \underbrace{\tau}_{\text{ATT}}(\text{Treated}_i \times \text{After}_t) + \varepsilon_{it}$$

- α_i: individual fixed effects.
- λ_t: time fixed effects.
- After_t: a time dummy of before vs. after treatment.
- Treated_i: an individual dummy of treated vs. control.

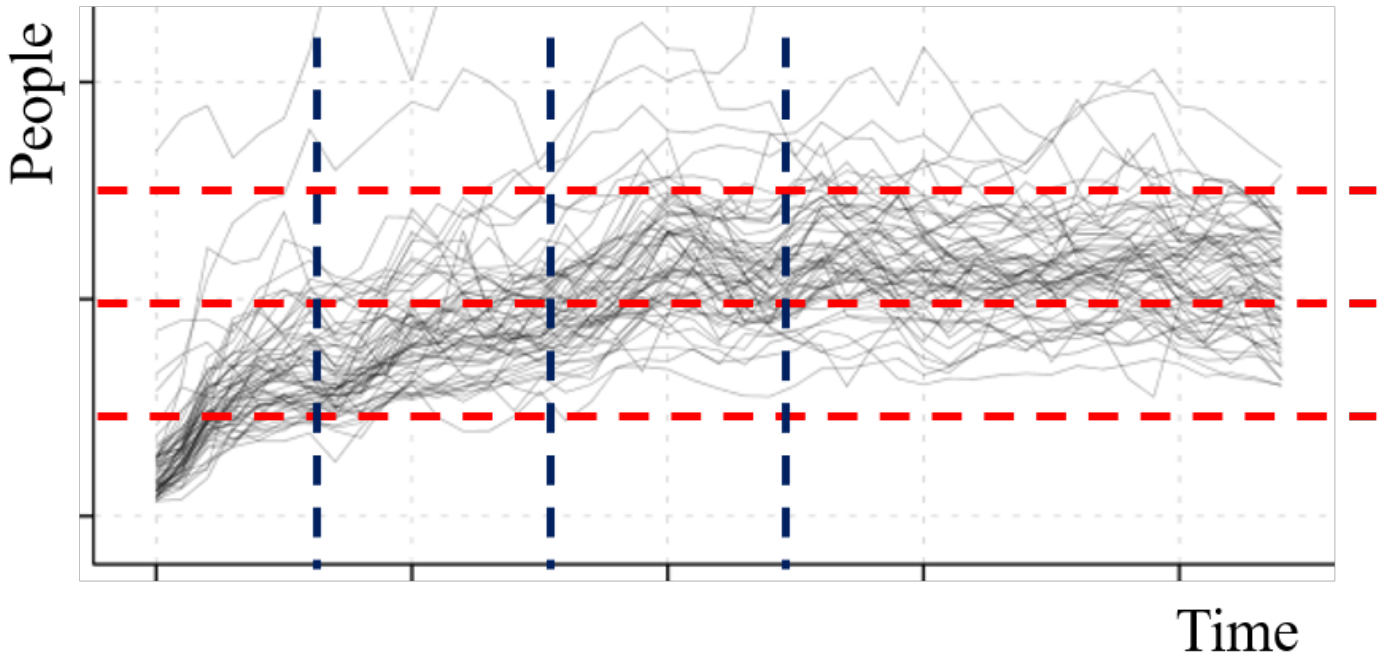
Model-based estimation

To better understand the model:

$y_{it} = \alpha_i + \lambda_t + \tau\,(\text{Treated}_i \times \text{After}_t) + \varepsilon_{it}$

	Before $(t = 1)$	After $(t = 2)$	Difference
Control $(i = 1)$	$\alpha_1 + \lambda_1 + \varepsilon_{11}$	$\alpha_1 + \lambda_2 + \varepsilon_{12}$	$E(\Delta Y_C) = \lambda_2 - \lambda_1$
Treated $(i = 2)$	$\alpha_2 + \lambda_1 + \varepsilon_{21}$	$\alpha_2 + \lambda_2 + \tau + \varepsilon_{22}$	$E(\Delta Y_T) = \tau + \lambda_2 - \lambda_1$
DID			$E(\Delta Y_T) - E(\Delta_C) = \tau$

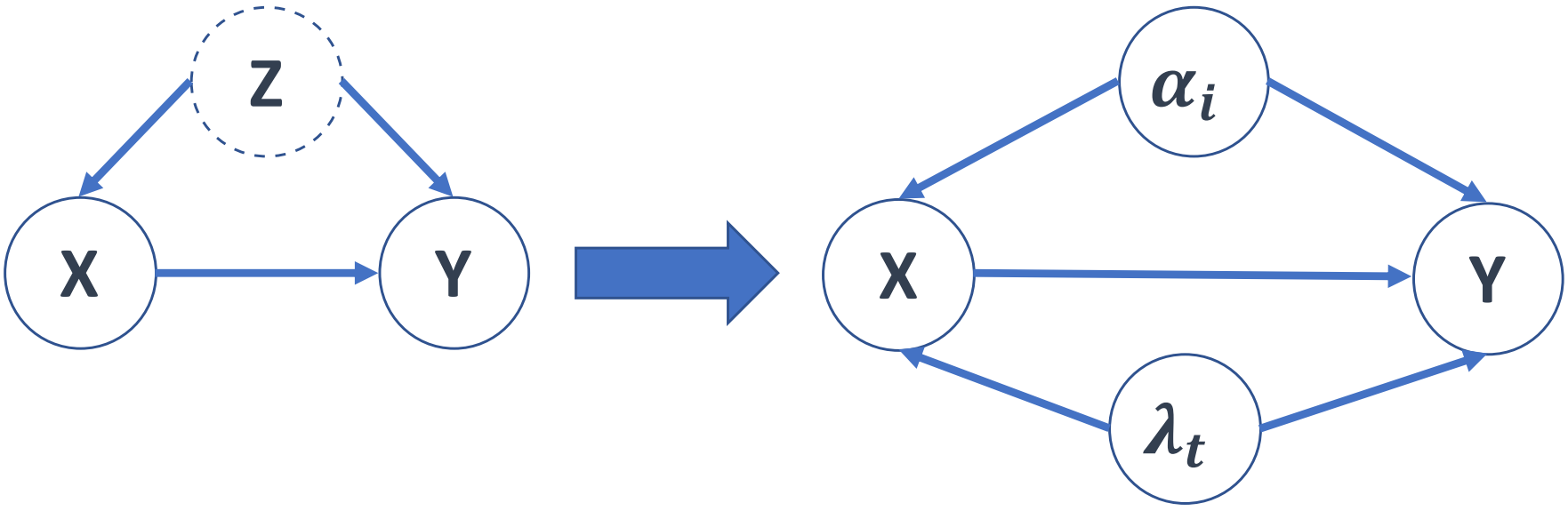
Model-based estimation: a link to data



With the two dimensional data (time-person), we can "control for"

1. Individual fixed effects (α_i) from repeated time periods;
2. Time fixed effects (λ_t) from multiple people.

Another look at the identification of DID



- DID approach is by no means safe-proof!
- Threat-to-identification: **confounders that vary across individuals and time**.

RSM Erasmus

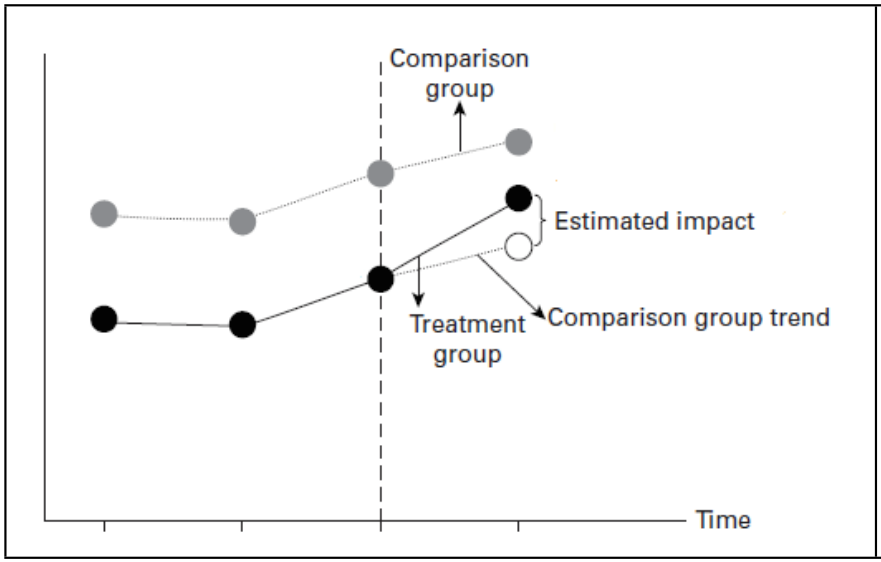
Outline

RSM Erasmus

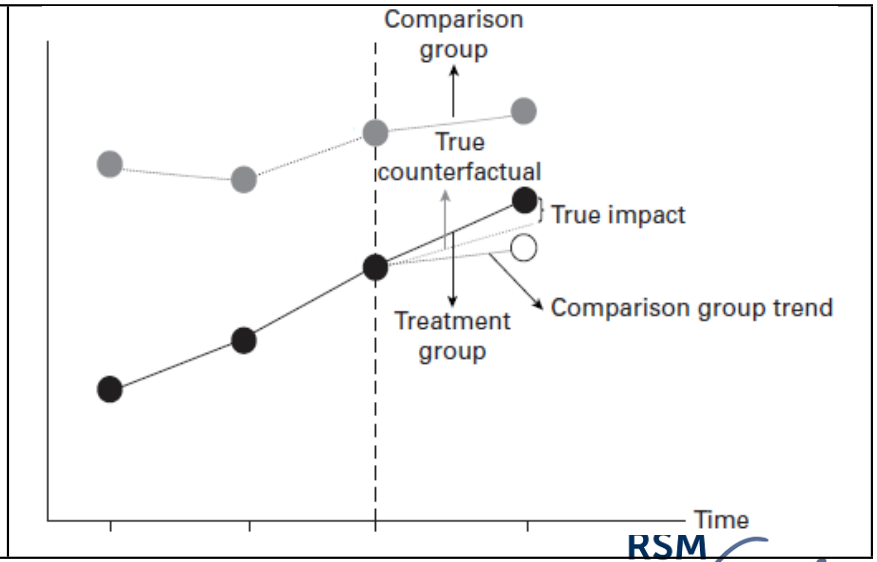
How to examine the parallel trend assumption?

The parallel trend assumption is untestable! What to do then?

- It has testable implications: **the pre-treatment trends should be parallel**.
- Therefore, we can test the pre-treatment trends.



Parallel Pre-trends

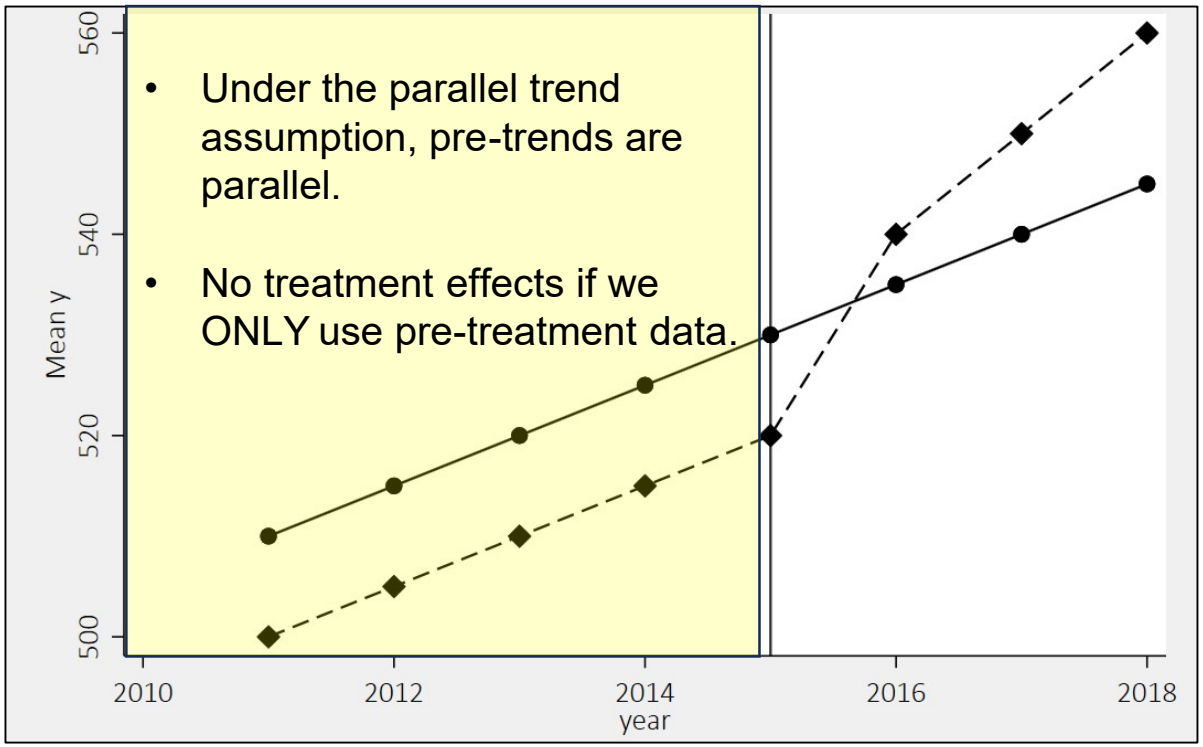


Non-parallel Pre-trends

Placebo tests in DID

Definition (Placebo Tests)

A pseudo-treatment (like a placebo in a drug trial) should not or cannot have an effect. Therefore, analyses with the pseudo-treatment should yield null results.



RSM Erasmus

Placebo tests in DID

The procedure of placebo tests in DID

1. Obtain the pre-treatment data.
2. Construct a pseudo treatment $\widetilde{\text{After}}_t$ by assuming the treatment happens at t in the pre-data.
3. Run a two-way fixed effects regression with $\widetilde{\text{After}}_t$ on pre-data.

$$y_{it} = \alpha_i + \lambda_t + \widetilde{\tau}\left(\text{Treated}_i \times \widetilde{\text{After}}_t\right) + \varepsilon_{it}$$

4. Repeat for different times t and collect the pseudo-ATT $\widetilde{\tau}$.

Under the parallel trend, $\widetilde{\tau}$ should be statistically 0.

RSM Erasmus

Event study graphs

Another way is to look at the dynamics of treatment effects using **event study graphs**.

Event Study Graphs

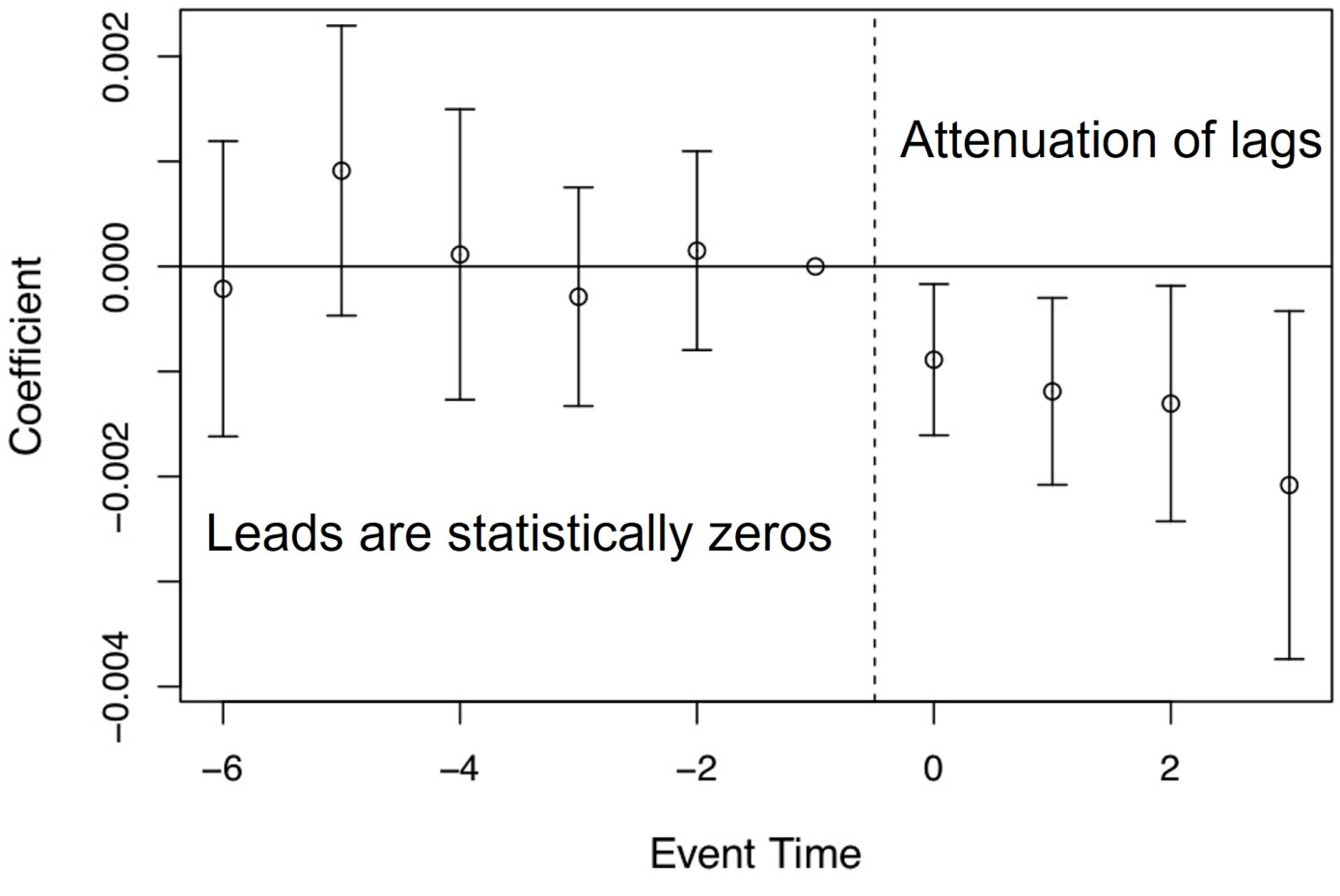
- Estimate a two-way fixed effects regression with q leads, m lags and the original treatment timing After_t.

$$y_{it} = \alpha_i + \lambda_t + \sum_{s=-q}^{m} \widetilde{\tau}_s \left(\text{Treated}_i \times \left(\underbrace{L^{-s}}_{\text{Lag}} \widetilde{\text{After}}_t \right) \right) + \varepsilon_{it}$$

- Plot the coefficients $\widetilde{\tau}_s$ and check the patterns.

RSM Erasmus

Event study graphs



The plot is accredited to Miller et al. (2021).

What if pre-trends are not parallel?

You need extra assumptions / information for a more credible identification.

- **Scenario 1**: With pre-treatment characteristics X_i.
- **Scenario 2**: With outcomes Y_{it} of many pre-treat periods.

Parallel pre-trends

Parallel pre-trends are **neither necessary or sufficient** for the parallel trend assumption!

RSM Erasmus

Scenario 1: With pre-treatment variables X_i

Conditional Parallel Trends Assumption

Conditioning on X_i, we have the following:

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

Identification results

Given the assumption, we can identify and conditional ATT:

$$\tau_{\text{ATT}}(x) = \underbrace{E\left[Y_{i2} - Y_{i1} \mid D_i = 1, X_i = x\right]}_{\text{Changes for the treated at } x} - \underbrace{E\left[Y_{i2} - Y_{i1} \mid D_i = 0, X_i = x\right]}_{\text{Changes for the control at } x}$$

Scenario 1: With pre-treatment variables X_i

Conditional Parallel Trends Assumption

Conditioning on X_i, we have the following:

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

Identification results

Given the assumption, we can identify and conditional ATT:

$$\tau_{\text{ATT}}\left(x\right) = \underbrace{E\left[Y_{i2} - Y_{i1} \mid D_i = 1, X_i = x\right]}_{\text{Changes for the treated at } x} - \underbrace{E\left[Y_{i2} - Y_{i1} \mid D_i = 0, X_i = x\right]}_{\text{Changes for the control at } x}$$

Scenario 1: With pre-treatment variables X_i

Matching (see Heckman et al. 1997):

$$\hat{\tau}_{\text{ATT}} = \frac{1}{N_1} \sum_{i \in \text{Treated}} \Big((Y_{i2} - Y_{i1}) - \hat{E}_X \left[Y_{i2} - Y_{i1} \mid D_i = 0, X_i \right] \Big)$$

- $\hat{E}_X \left[Y_{i2} - Y_{i1} \mid D_i = 0, X_i \right]$ is the estimated expectation function fitted on the control units.
- It takes in values of X_i and predicts the counterfactual changes at X_i for a treated unit.
- In theory, any function form can be use, e.g., a polynomial function.

RSM Erasmus

Scenario 1: With pre-treatment variables X_i

Weighting (see Abadie 2005):

$$\hat{\tau}_{\text{ATT}} = \frac{1}{N_1} \sum_{i \in \text{Treated}} (Y_{i2} - Y_{i1}) - \frac{1}{N_1} \sum_{i \in \text{Control}} \left(\frac{\hat{e}(X_i)}{1 - \hat{e}(X_i)} \right) (Y_{i2} - Y_{i1})$$

- $\hat{e}(X_i)$ is the estimated propensity score function.
- We first estimate $\hat{e}(X_i)$ and then plug it in.

RSM Erasmus

Scenario 1: With pre-treatment variables X_i

With multiple time periods, we can run a two-way fixed effects regression with X_i as controls:

$$y_{it} = \alpha_i + \lambda_t + \tau\left(\text{Treated}_i \cdot \text{After}_t\right) + \beta\left(X_i \cdot \text{After}_t\right) + \varepsilon_{it}$$

- Note that this specification may not be consistent – no treatment heterogeneity.
- The coefficient β is the same across different levels of X_i.
- With too may X_i, we can use sub-classification based on X_i.

RSM Erasmus

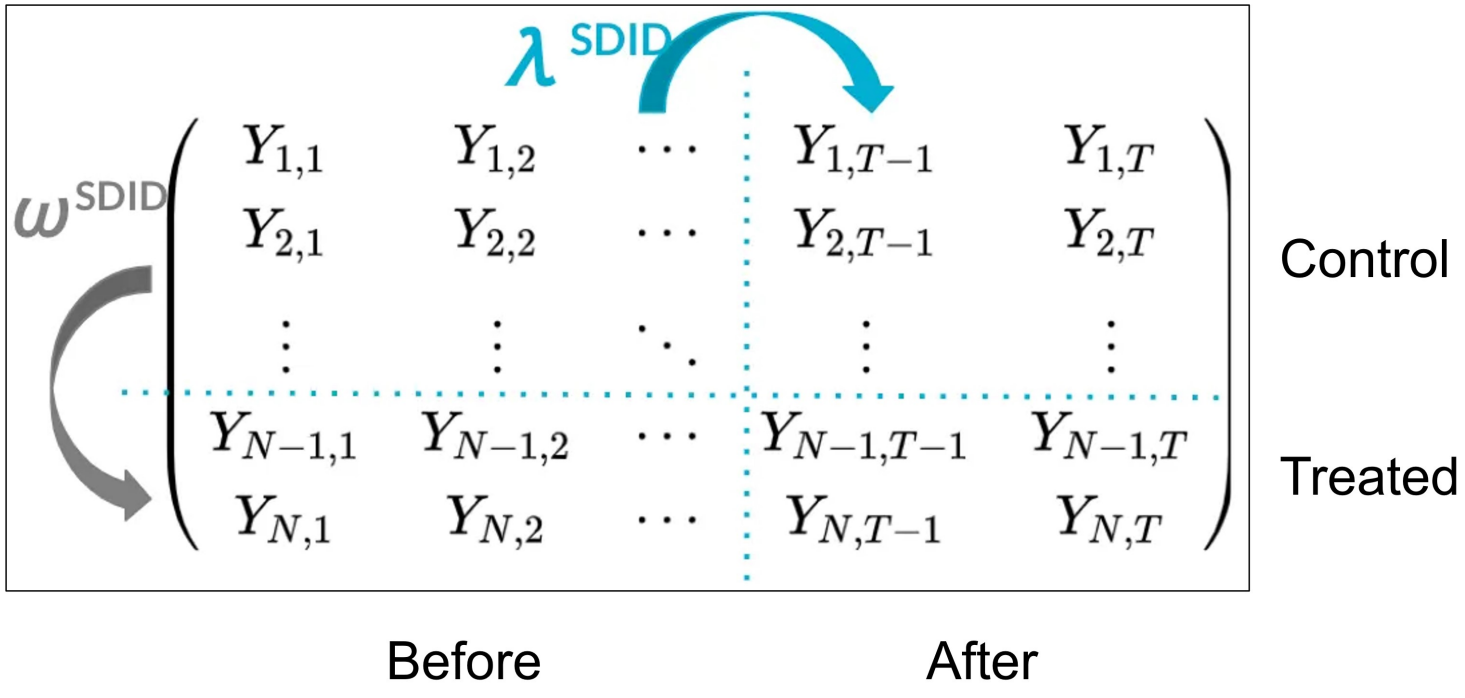
Scenario 2: With many pre-treatment periods

In weighting approach, we weight the control units for ATT (ω^{SDID}).

- Inadequate if the pre-trends are non-parallel.

Arkhangelsky 2021 proposes to also weight the time periods (λ^{SDID}).

- To calculate λ^{SDID} with long pre-treatment periods.
- The idea is similar to traditional event studies.



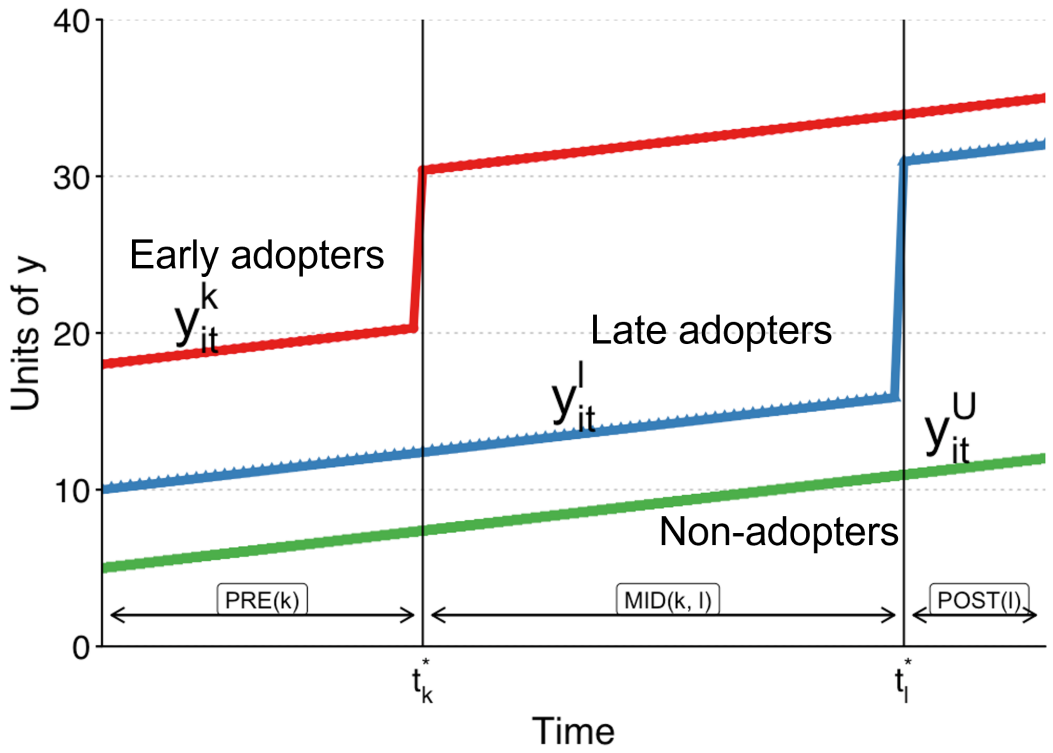
Outline

RSM Erasmus

Differential timing of the treatment

So far, we assume two things about the treatment:

1. Treatment happens at t for all people.
2. Once treated, remain treated[2].

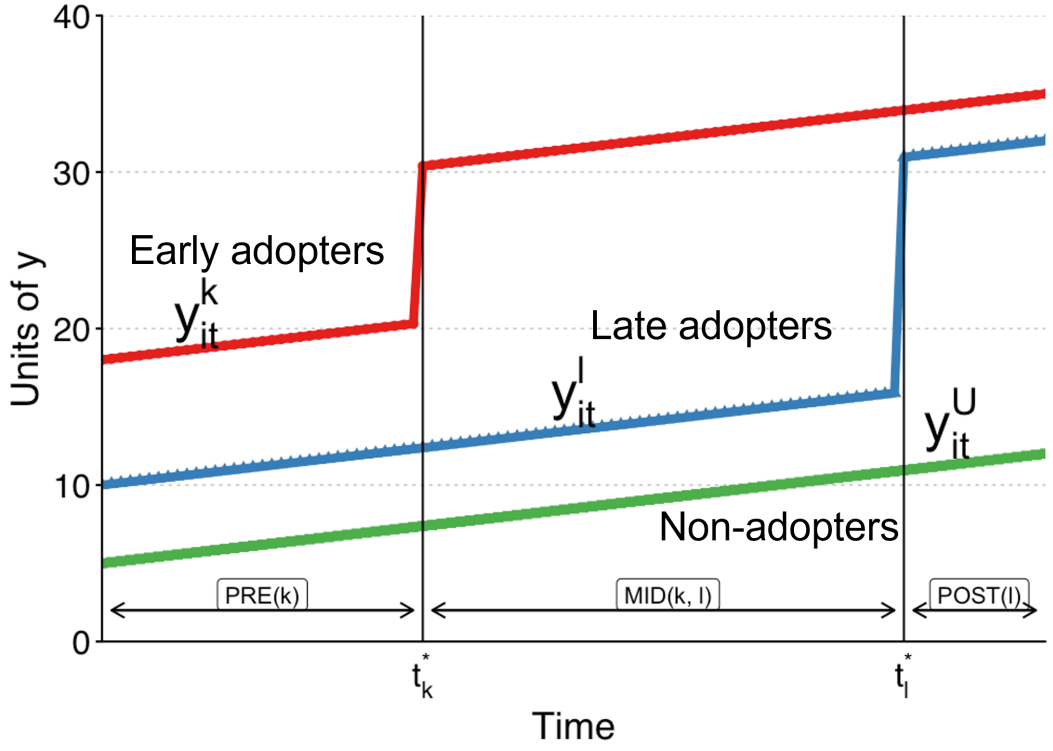


[2] In statistical term, treatment is an absorbing state.

Differential timing of the treatment

In practice, we have "staggered adoptions"

- Different groups receive the treatments at different times.
- E.g., roll-out strategies in marketing or policy testing.

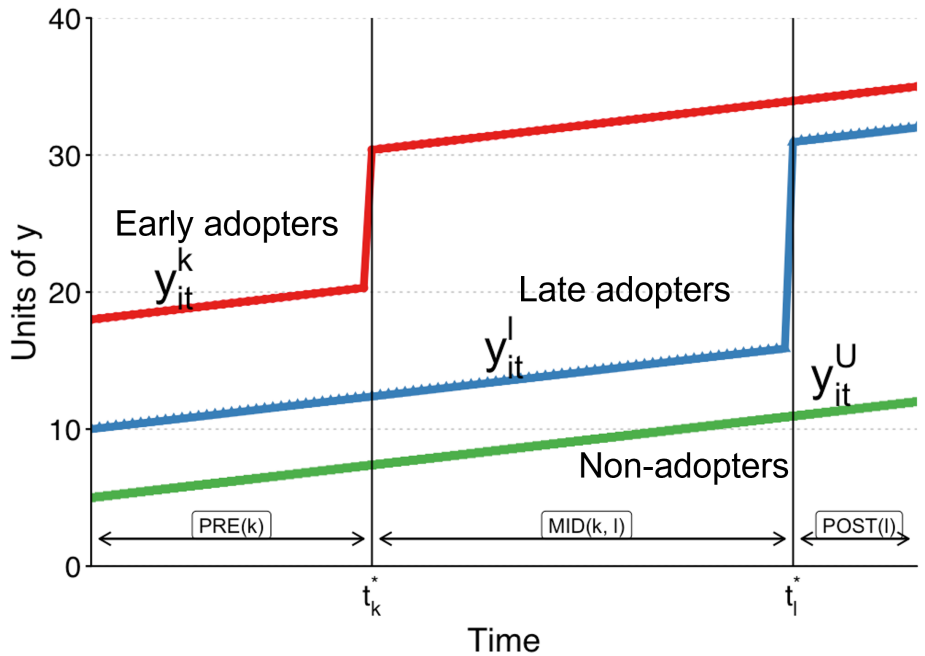


Differential timing of the treatment

Can we still use the go-to model – two-way fixed effects regression?

$$y_{it} = \alpha_i + \lambda_t + \tau \underbrace{D_{it}}_{\text{Treatment}} + \varepsilon_{it}$$

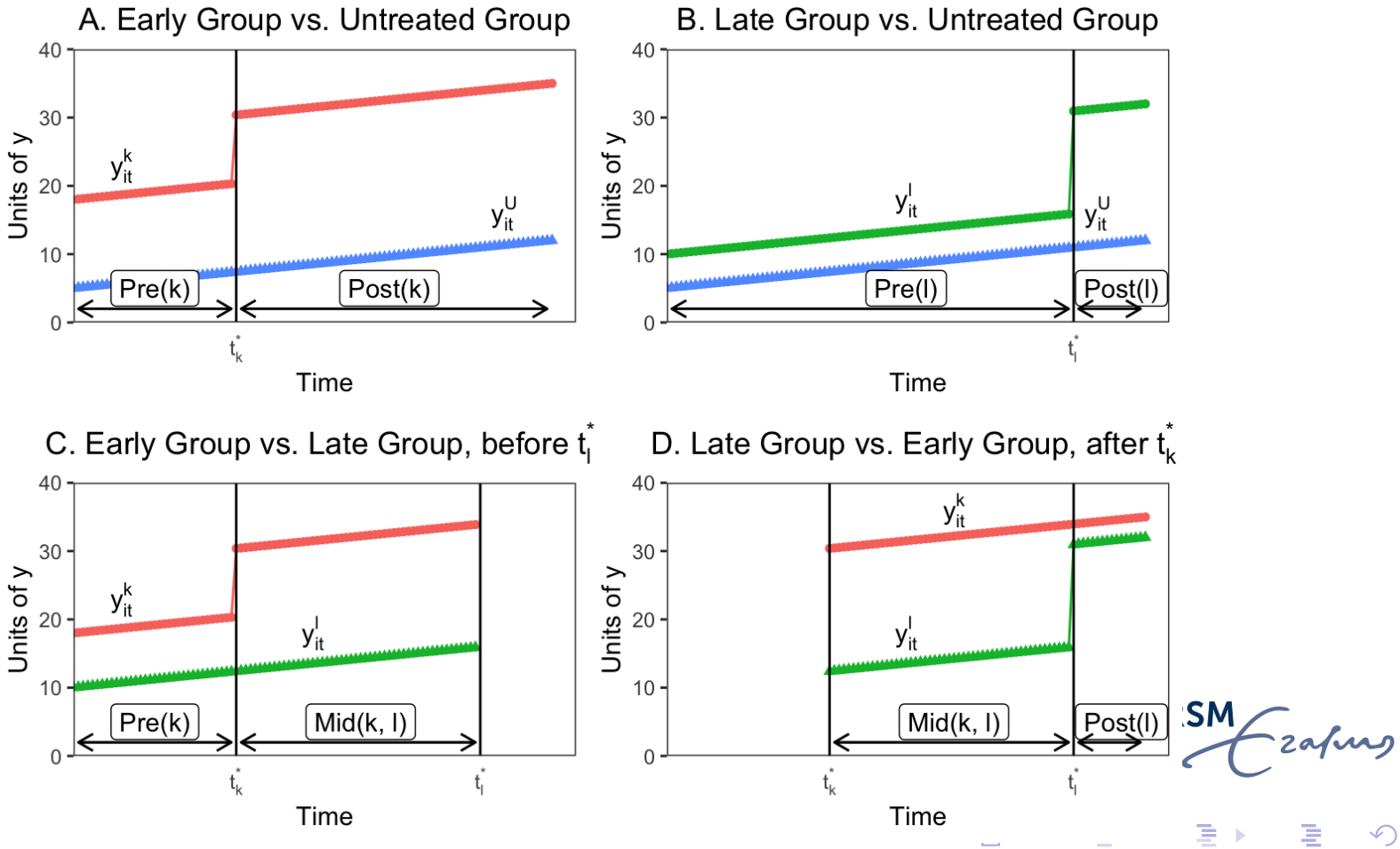
- Instead of $(\text{Treated}_i \cdot \text{After}_t)$, we now have a variable D_{it}.
- The treated people receive the treatment at different times.



RSM Erasmus

The two-ways fixed effects regression is biased!

Goodman-Bacon (2021) formalizes this and shows a nice intuition. ATT from the regression is a weighted average of group*time comparisons.



How to correct for the bias?

A few approaches proposed recently:

- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200-230.
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2), 175-199.
- Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. The Quarterly Journal of Economics, 134(3), 1405-1454.

RSM Erasmus

References I

- Azoulay, P., Fons-Rosen, C., & Zivin, J. S. G. (2019). Does science advance one funeral at a time?. American Economic Review, 109(8), 2889-2920.
- Card, D., & Krueger, A. B. (1994). Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. The American Economic Review, 84(4), 772-793.
- Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. Foundations and Trends® in Econometrics, 4(3), 165-224.
- Miller, S., Johnson, N., & Wherry, L. R. (2021). Medicaid and mortality: new evidence from linked survey and administrative data. The Quarterly Journal of Economics, 136(3), 1783-1829.

References II

- Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. Review of Economic Studies, 72(1), 1-19.
- Heckman, J., Ichimura, H., & Todd, P. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. Review of Economic Studies, 64(4), 605-654.
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. American Economic Review, 111(12), 4088-4118.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254-277.

RSM Erasmus