Estimating treatment effect. Difference-in-difference.

Jakub Mućk
SGH Warsaw School of Economics

Treatment Effects

- For researchers, it is important to avoid the faulty line of reasoning known as **post hoc, ergo propter hoc**
  - One event's preceding another does not necessarily make the first the cause of the second.
  - **Correlation is not the same as causation**.
- **Selection bias** is an issue when a proper randomization is not achieved, i.e., the sample is not random.
- **Selection bias** arises also when the method of collecting samples is not appropriate.
- **Selection bias** may be critical for measuring a **causal effect (treatment effect)**. If selection bias is not taken into account the results of statistical inference may be not valid.

- **Randomized controlled experiment**. To avoid selection bias researchers would like to randomly assign items to a **treatment group**, with others being treated as a **control group**. As a result, two groups can be compared.
- In economics, performing randomized controlled experiment is limited.

- **Assumption #1 Unconfoundedness**

$$(y(0), y(1)) \perp w | X \tag{1}$$

where
  - $y(0), y(1)$ are the potential/counterfactual outcomes.
  - w is the assignment.

- **Assumption #2: Overlap**

$$0 < P(w = 1 | X = x) < 1 \tag{2}$$

Estimating the treatment effect

Three general strategies:

1. Regression-based methods,
2. Propensity score methods,
3. Matching methods.

- The indicator variable d:

$$d = \begin{cases} 1 & \text{if} \quad \text{individual in treatment group,} \\ 0 & \text{if} \quad \text{individual in control group.} \end{cases}$$

- The regression function is conditional to treatment:

$$\mathbb{E}(y_i) = \begin{cases} \beta_1 + \beta_2 & \text{if} \quad \text{individual in treatment group}, d_i = 1, \\ \beta_1 & \text{if} \quad \text{individual in control group}, d_i = 0. \end{cases}$$

- The econometric model:

$$y_i = \beta_1 + \beta_2 d_i + \varepsilon_i, \qquad i = 1, \ldots, N. \tag{3}$$

- In this simplified case the least square estimator for the **treatment effect** β_2:

$$\hat{\beta}_2^{LS} = \frac{\sum_{i=1}^{N} \left(d_i - \bar{d}_i\right)(y_i - \bar{y})}{\sum_{i=1}^{N} \left(d_i - \bar{d}_i\right)^2} = \bar{y}_1 - \bar{y}_0, \tag{4}$$

while $\bar{y}_1$ and $\bar{y}_0$ are the samples averages in treatment and control group, respectively.

- $\hat{\beta}_2^{LS}$ is also called the **difference estimator**.

- **Unbiasedness**. The expected value of the difference estimator β_2^{LS}:

$$\hat{\beta}_2^{LS} = \beta_2 + \frac{\sum_{i=1}^{N}\left(d_i - \bar{d}_i\right)\left(\varepsilon_i - \bar{\varepsilon}\right)}{\sum_{i=1}^{N}\left(d_i - \bar{d}_i\right)^2} = \beta_2 + \bar{\varepsilon}_1 - \bar{\varepsilon}_2, \tag{5}$$

  the potential bias:

$$\mathbb{E}\left(\bar{\varepsilon}_1 - \bar{\varepsilon}_2\right) = \mathbb{E}\left(\bar{\varepsilon}_1\right) - \mathbb{E}\left(\bar{\varepsilon}_2\right) = 0. \tag{6}$$

- However, If we allow individuals to *self-select* into treatment and control groups then $\mathbb{E}\left(\bar{\varepsilon}_1\right) - \mathbb{E}\left(\bar{\varepsilon}_2\right)$ is the selection bias of the treatment effect.

- If conditioning factors are omitted in regression then difference estimator is biased.
- Popular strategy is to introduce fixed effects which captures unobservable factors determining outcome:.
- [Example] Project STAR

$$TOTALSCORE = \beta_0 + \beta_1 SMALL + \varepsilon, \tag{7}$$

where

- $TOTALSCORE$ – the combined reading and math achievement scores,
- $SMALL$ – indicator variable which takes 1 if the student was assigned to a small class.

- **School fixed effects:**

$$school_j = \begin{cases} 1 & \text{if} \quad \text{student is in school } j, \\ 0 & \text{if} \quad \text{otherwise.} \end{cases}$$

The extended regression:

$$TOTALSCORE = \beta_0 + \beta_1 SMALL + \sum_{j=1}^{S} \delta_j school_j + \varepsilon, \tag{8}$$

where δ_j is the fixed effect for the j-th school..

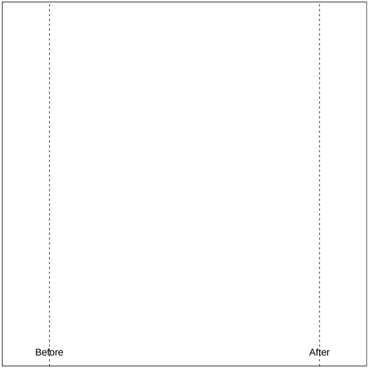
- Key idea in the estimation of treatment effect is comparison of counterfactual outcomes.
- One might use the individuals with the same characteristics ($\implies$ exact matching).
- Another possibility is to use propensity score, i.e., the estimated probability of treatment conditional to an observed X.
- Practically, it can be done with logit/probit estimation of the treatment assignment on some individual features/X.
- Next, the counterfactual outcome is estimated by comparison individuals from the treated and control group that are similar in terms of the propensity score.

- The common strategy is to control for effect of conditioning factors. This could be done with adding explanatory variables.
- Another way to check for random assignment is to regress treatment variable on these characteristics and check for any significant coefficients.
  - This is equivalent with the linear probability model for the treatment variable.
- If there is random assignment, we should not find any significant relationships

Differences-in-Differences estimator

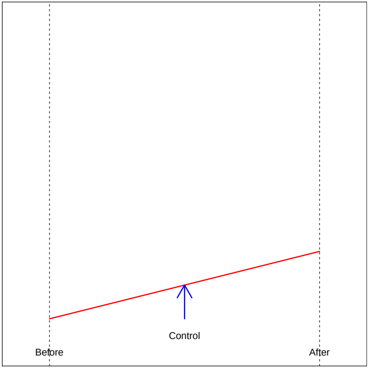
In **Differences-in-Differences** approach estimation of the treatment effect is based on data averages for the two groups (treatment (T) and control (C)) in the two periods (before (B) and after (A)):

$$\hat{\delta} = \left(\bar{y}_{T,A} - \bar{y}_{C,A}\right) - \left(\bar{y}_{T,B} - \bar{y}_{C,B}\right).$$



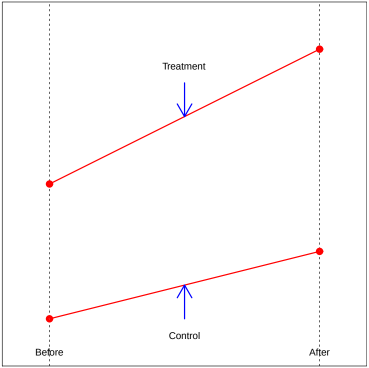
In **Differences-in-Differences** approach estimation of the treatment effect is based on data averages for the two groups (treatment (T) and control (C)) in the two periods (before (B) and after (A)):

$$\hat{\delta} = \left(\bar{y}_{T,A} - \bar{y}_{C,A}\right) - \left(\bar{y}_{T,B} - \bar{y}_{C,B}\right).$$



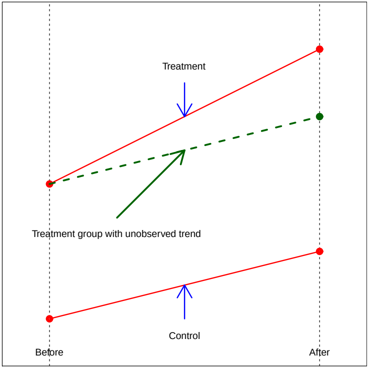
In **Differences-in-Differences** approach estimation of the treatment effect is based on data averages for the two groups (treatment (T) and control (C)) in the two periods (before (B) and after (A)):

$$\hat{\delta} = \left(\bar{y}_{T,A} - \bar{y}_{C,A}\right) - \left(\bar{y}_{T,B} - \bar{y}_{C,B}\right).$$



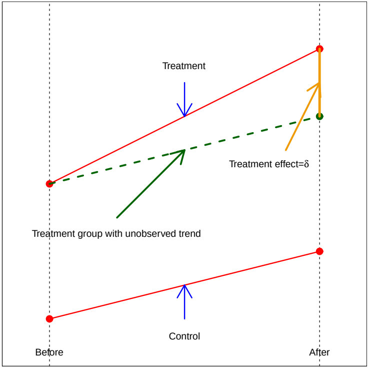
In **Differences-in-Differences** approach estimation of the treatment effect is based on data averages for the two groups (treatment (T) and control (C)) in the two periods (before (B) and after (A)):

$$\hat{\delta} = \left(\bar{y}_{T,A} - \bar{y}_{C,A}\right) - \left(\bar{y}_{T,B} - \bar{y}_{C,B}\right).$$



In **Differences-in-Differences** approach estimation of the treatment effect is based on data averages for the two groups (treatment (T) and control (C)) in the two periods (before (B) and after (A)):

$$\hat{\delta} = \left(\bar{y}_{T,A} - \bar{y}_{C,A}\right) - \left(\bar{y}_{T,B} - \bar{y}_{C,B}\right).$$



- Consider the regression:

$$y_{it} = \beta_1 + \beta_2 TREAT_i + \beta_3 AFTER_t + \delta \left(TREAT_i \times AFTER_t \right) + \varepsilon_{it} \quad (9)$$

- the expected outcome

$$\mathbb{E}(y_{it}) = \begin{cases} \beta_1 & \text{if} \quad TREAT = 0 \quad AFTER = 0 \\ \beta_1 + \beta_2 & \text{if} \quad TREAT = 1 \quad AFTER = 0 \\ \beta_1 + \beta_3 & \text{if} \quad TREAT = 0 \quad AFTER = 1 \\ \beta_1 + \beta_2 + \beta_3 + \delta & \text{if} \quad TREAT = 1 \quad AFTER = 1 \end{cases} \quad (10)$$

- The least squares estimates of treatment effect:

$$\hat{\delta}^{LS} = \left(\bar{y}_{T,A} - \bar{y}_{C,A} \right) - \left(\bar{y}_{T,B} - \bar{y}_{C,B} \right). \quad (11)$$

- D-i-D can be applied in more general regression.
- Using the panel data techniques we can control for unobserved heterogeneity.
- **Parallel trends assumption**. In the D-i-D approach key assumption is that before treatment/intervention there was common trend in both treated and control group.