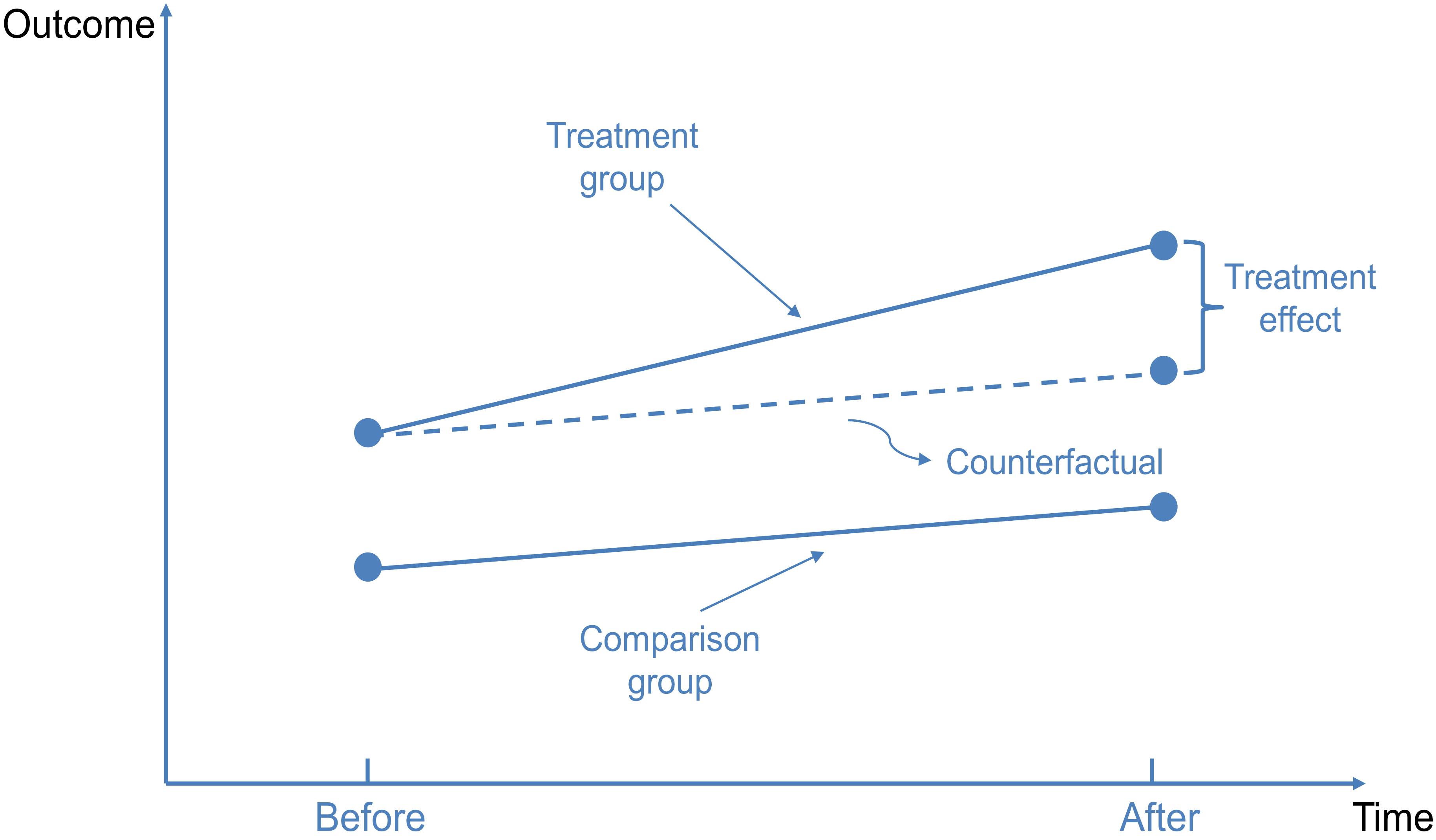
Difference-in-Difference Methods

Racha Ramadan

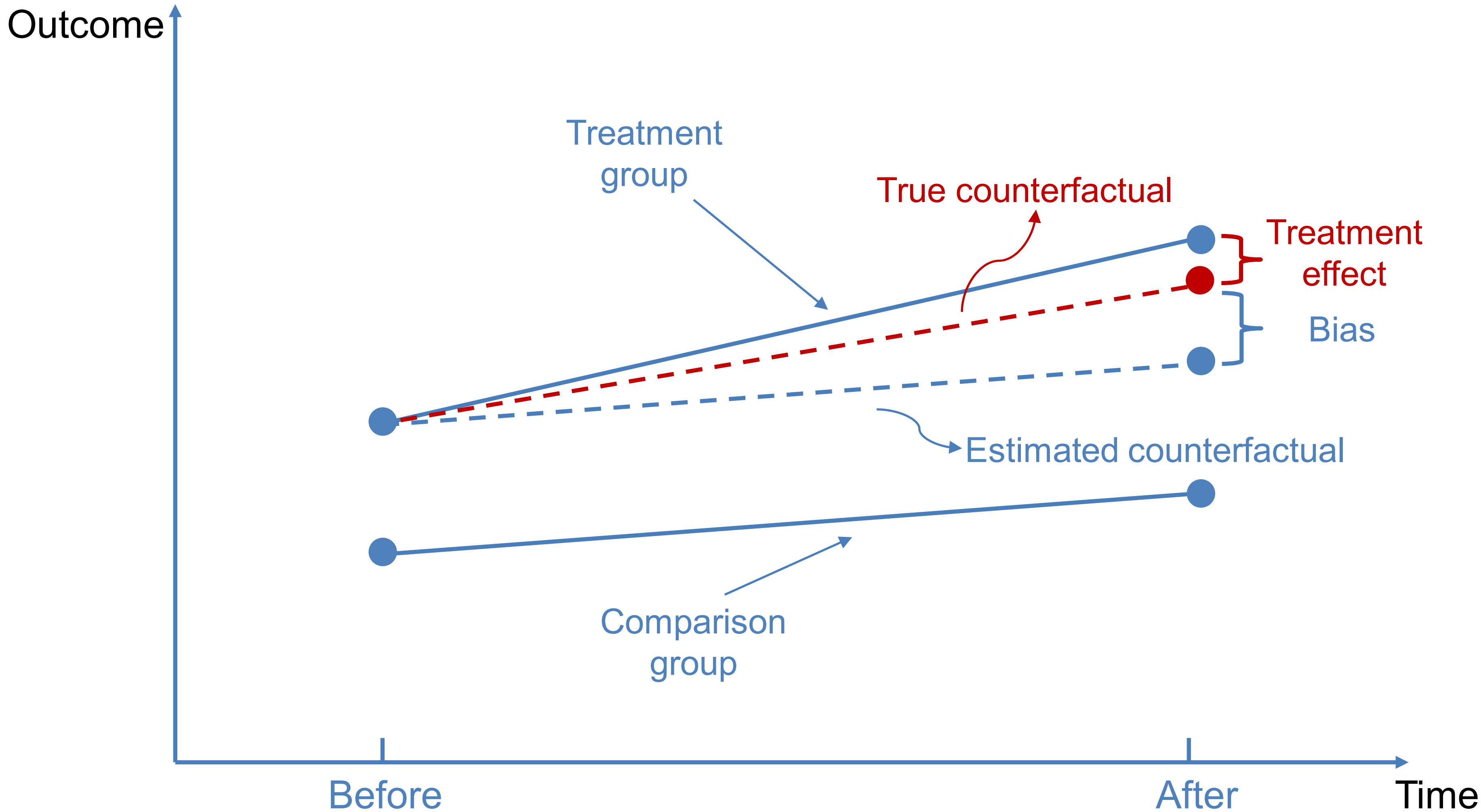
racha.ramadan@feps.edu.eg

Training on Applied Micro-Econometrics and Public Policy Evaluation
Economic Research Forum
2024

Parallel trends assumption: in the absence of treatment, outcomes in treatment and comparison groups display equal trends



Threat comes from any time-varying factor that differentially affects outcomes in treatment and control



Identification Assumptions

- Conditional Exogeneity of Treatment:

Treatment or placement into program does not depend on **unobservables** characteristics that may affect the outcome of interest.

- Conditional Exogeneity of Treatment to changes in outcomes:
  - It is a much weaker assumption.
  - Here we distinguish between two kinds of **unobservables**:
    1. time invariant unobservables.
    2. time varying unobservables.
  - Under this assumption, selection can depend on time invariant unobservables, but is assumed not to depend on time-varying unobservables.

	Outcome of interest before treatment	Outcome of interest after treatment
Treated Group	A	B
Control Group	C	D

First Difference

- Under the stronger assumption "*conditional exogeneity of Treatment*":
  - ✓ all we need to do is compare outcomes for a treatment and control group at one point in time controlling for observables X:

	Outcome of interest after treatment	**Difference**
Treated Group	B	B-D
Control Group	D	

- This is called a first difference approach

Difference-in Difference Approaches

- Under the weaker assumption "*exogeneity of treatment to changes in outcomes*":
  - ✓ we need to compare the difference from before and after the program for a treatment group to the same difference for a control group.
- This is called difference-in-difference (DD) or second difference approach.
- DD assumes that the bias selection is based on observed and unobserved characteristics that are time invariant.

	Outcome of interest before treatment	Outcome of interest after treatment	Difference
Treated Group	A	B	B-A
Control Group	C	D	D-C
Difference			**DD=(B-A)- (D-C)**

- DD method estimate the average program impact as follows:

$$\mathbf{DD = E\left(Y_1^T - Y_0^T \mid X, T_1 = 1\right) - \ E(Y_1^c - Y_0^c \mid X, T_1 = 0)}$$

- If the counterfactual expectations are not expected to change over time:

$$\mathbf{E(Y_1^c - Y_0^c \mid X, T_1 = 0) = 0}$$

- Under such an assumption, the DD estimator reduces to the Before and After (BA) estimator on the treated group, making measurement on a comparison group unnecessary:

$$\mathbf{BA\ (X) = DD = E\left(Y_1^T - Y_0^T \mid X, T_1 = 1\right)}$$

Such an assumption is unlikely to hold in practice!!

Estimating the DD Estimator: The Regression Approach

Step 1: Regress the outcome of interest (Y_{it}) on the treatment status (T_{i1}), a time dummy (t) and the interaction between the two variables

$$\boldsymbol{Y_{it} = \alpha + \gamma T_{i1} + \delta t + \beta T_{i1} t + \varepsilon_{it}}$$

Note:

- t=0 for time before the treatment;
- t=1, for time after treatment;
- $T_{i0} = 0$, by definition

Step 2: Calculate the following expectations:

- outcome of individual i before treatment if the person is not treated:

$$E(Y_{i0}|T_{i1} = 0) = \alpha$$

- outcome of individual i after treatment if the person is not treated :

$$E(Y_{i1}|T_{i1} = 0) = \alpha + \delta$$

- outcome of individual i before treatment if the person is treated

$$E(Y_{i0}|T_{i1} = 1) = \alpha + \gamma$$

- outcome of individual i after treatment if the person is treated

$$E(Y_{i1}|T_{i1} = 1) = \alpha + \gamma + \delta + \beta$$

- **Step 3:** Calculate DD estimator

$$DD = E\left(Y_1^T - Y_0^T \mid X, T_1 = 1\right) - E\left(Y_1^C - Y_0^C \mid X, T_1 = 0\right)$$

$$DD = (\beta + \delta) - \delta = \beta$$

Under the weaker identification assumption:
the effect of the treatment on the treated (ATT) is given by the regression coefficient β.
It is the coefficient of the interaction term between the treatment dummy and the time dummy.

Introducing Covariates

- It is easy to introduce covariates in DD estimation:

$$\boldsymbol{Y_{it} = \alpha + \gamma T_{i1} + \delta t + \beta T_{i1} t + \sum_{j} \theta_j X_{itj} + \varepsilon_{it}}$$

- All the expectations can now be written given X. ATT is still equal to β.

- To obtain heterogeneous effects, all you need to do is interact T and (T x t) with the covariates Xs.

Cross-Sectional Difference-in-Difference

- DD can be done across groups of individuals, with one group not expected to be affected by treatment.

- Example:
  - Program and non-program villages (instead of time)
  - Targeted and non-targeted households (instead of participants vs. non-participants)

- DD: difference between targeted and non-targeted individuals (or households) in program villages, minus the same difference in non-program villages.

Difference-in Difference and PSM

- One can use DD in the context of PSM as well.
- Matching is still done along pre-treatment characteristics using “wide” data.
- Use p score as weight in the DD regression

Data Structures for DD Estimation

- Multiple observations on the same units – *Panel Data*

- x_{ij} : the variable x for the individual i in time period j.

Examples: employment status, income, place of residence, sex…etc

- Some variables are 'time-varying' and can change over time

  Examples: employment status, income.

- Some variables are 'time invariant' and cannot change over time (i.e. sex)

- *Data entry errors or other issues may make time invariant variables time-varying in actual data!*

- Panel data can come in a number of different shapes:
  - "Long" data:
    - An individual shows up twice, once in time period 0 and once in time period 1.
    - An observation is an individual in a time period.
  - "Wide" data:
    - An observation is a person, and there are different variables for each time period.

- Example of Long data

i id	j year	sex	x_ij inc
1	80	0	5000
1	81	0	5500
1	82	0	6000
2	80	1	2000
2	81	1	2200
2	82	1	3300
3	80	0	3000
3	81	0	2000
3	82	0	1000

- An observation is a person (id=i) in a year(=j)
- A person has multiple observations (multiple years)
- inc (income) is time varying (x_ij)
- Sex is time invariant

- Example of Wide data

i		.... x_ij		
id	sex	inc80	inc81	inc82
1	0	5000	5500	6000
2	1	2000	2200	3300
3	0	3000	2000	1000

- An observation is a person (id=i)
- A person has only one observation
- There is still only one variable for sex (time invariant)
- Now there are three variables for income
- One for each year
- There is no more year variable—it is a suffix on income

❑ Reshaping Data from Long to Wide and vice versa

- You can use the “reshape” command in STATA to reshape data from long to wide and vice versa.

- However, before you do that, you need to know what form your data is in.

- The “duplicate report” command on the unique individual identifier can allow you to do that.

References

- Slides of Ragui Assaad and Caroline Krafft.
- Benmarhnia T, Bailey Z, Kaiser D, Auger N, King N, Kaufman J. 2016. "A difference-indifferences approach to assess the effect of a heat action plan on heat-related mortality, and differences in effectiveness according to sex, age, and socioeconomic status" (Montreal, Quebec). Environ Health Perspect 124:1694–1699; http://dx.doi.org/10.1289/EHP203
- Centre for Evaluation Seminar Series. 2018. "Difference-in-Differences in Action". Department of Global Health and Development London School of Hygiene and Tropical Medicine.
- Khandker, S.R., Koolwal, G.B. and Samad, H.A., 2009. "Handbook on Impact Evaluation Quantitative Methods and Practices". World Bank.