Lecture 5: Diff-in-diff

Lecturer: Ksenia Kasianova
xeniakasianova@gmail.com

December 4, 2023

Plan

Plan

1) Causal inference, treatment effect estimation

2) Diff-in-diff

3) Matching + Diff-in-diff

Estimating treatment effect

The binary variable D_i:

$D_i = 1$ for**Treatment group** – the group of objects that were exposed to the treatment

$D_i = 1$ for **Control group** – the group that was not exposed to the treatment

To indicate that an object belongs to a particular group, we will use

For example,

$D_i = 1$ if the ith individual was hospitalized. The health level of the i-th individual will be denoted by $Y_i(1)$;

$D_i = 0$ – if the individual was not hospitalized. The health level of the i-th individual will be denoted by $Y_i(0)$.

Then the change in an individual's health as a result of hospitalization can be defined as follows:

$$Y_i(1) - Y_i(0)$$

This quantity is called **the treatment effect or causal effect for the ith individual.**

In our example, the treatment effect for some individual is the amount by which his level of health will change if he is subjected to treatment, compared to the case if he is not treated.

Estimating treatment effect

If we average the treatment effect over all individuals from the population, we obtain the so-called **average treatment effect (ATE)**.

Here are examples of ATE :

- How much does the health of individuals increase on average as a result of their hospitalization?

- How much will students' performance increase on average if they are taught in a small class instead of a normal-sized class?

- How much will employment in fast food restaurants change on average as a result of the adoption of the minimum wage law?

To calculate the treatment effect, you need to calculate the difference $Y_i(1) - Y_i(0)$.

In practice, **this is not possible**, since for no object we observe both $Y_i(1)$ and $Y_i(0)$.

Estimating treatment effect

The problem of counterfactual

The main challenge of an impact evaluation is to determine what would have happened to the beneficiaries if the program had not existed.

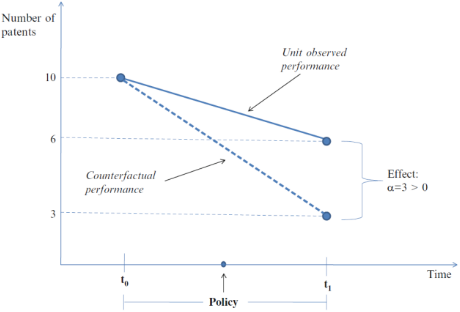


Figure:

Without information on the counterfactual, the next best alternative is to compare outcomes of treated individuals with those of a comparison group that has not been treated.

Estimating treatment effect

The level of health that we actually observe is the value of the dependent variable available to us in the data for a particular individual (observed outcome):

$$Y_i = \begin{cases} Y_i(1), & \text{if } D_i = 1; \\ Y_i(0), & \text{if } D_i = 0. \end{cases}$$

Sometimes it is convenient to write Y_i as follows:

$$Y_i = Y_i(0) + D_i \cdot (Y_i(1) - Y_i(0)).$$

Since we cannot directly calculate the impact effect for an individual object $Y_i(1) - Y_i(0)$, we cannot calculate its mathematical expectation $E(Y_i(1) - Y_i(0))$, i.e. ATE, as is.

Instead, we can try to estimate this effect using observed data.

Let's consider comparing the expected health levels of those who were hospitalized with the expected health levels of everyone else:

$$E(Y_i \mid D_i = 1) - E(Y_i \mid D_i = 0)$$

where $E(Y_i \mid D_i = 1)$ is the expected value of the dependent variable for objects that were exposed; $E(Y_i \mid D_i = 0)$ is the expected value of the dependent variable for objects that were not exposed.

Estimating treatment effect

To find out how this difference in mathematical expectations relates to the value ATE of interest to us, let us carry out the following transformations:

$$
\begin{gathered}
E\left(Y_i \mid D_i=1\right)-E\left(Y_i \mid D_i=0\right)=E\left(Y_i(1) \mid D_i=1\right)-E\left(Y_i(0) \mid D_i=0\right)= \\
=E\left(Y_i(1) \mid D_i=1\right)-E\left(Y_i(0) \mid D_i=1\right)+E\left(Y_i(0) \mid D_i=1\right)-E\left(Y_i(0) \mid D_i=0\right)= \\
=\underbrace{E\left(Y_i(1)-Y_i(0) \mid D_i=1\right)}_{\text{ATET}}+\underbrace{E\left(Y_i(0) \mid D_i=1\right)-E\left(Y_i(0) \mid D_i=0\right)}_{\text{selection bias}} .
\end{gathered}
$$

The last expression consists of two terms:

- $E\left(Y_i(1)-Y_i(0) \mid D_i=1\right)$ is the average treatment effect on the treated, ATET

- $E\left(Y_i(0) \mid D_i=1\right)-E\left(Y_i(0) \mid D_i=0\right)$ - this expression is called selection bias

– the expected level of health of hospitalized people $(D=1)$ if they had not gone for treatment $(Y_i(0))$.

– **minus** the expected level of health of people who did not go for treatment.

Thus, the initial difference in mathematical expectations can be written as follows:

$$
E\left(Y_i \mid D_i=1\right)-E\left(Y_i \mid D_i=0\right)=ATET+\text{ selection bias.}
$$

Example: people with low levels of health are more likely to enter the experiment $=>$ $Bias<0$

Estimating treatment effect

Under conditions of random distribution into groups (**random assignment**), whether an object falls into one group or another will not depend on its characteristics.

In terms of mathematical expectations, this would mean that:

$$E\left(Y_i(0) \mid D_i = 1\right) = E\left(Y_i(0) \mid D_i = 0\right) = E\left(Y_i(0)\right).$$

In this situation, there is no self-selection bias:

$$\text{selection bias } = E\left(Y_i(0) \mid D_i = 1\right) - E\left(Y_i(0) \mid D_i = 0\right) = 0.$$

Consequently, the difference in conditional mathematical expectations is equal to the average impact effect of interest to us:

$$E\left(Y_i \mid D_i = 1\right) - E\left(Y_i \mid D_i = 0\right) = ATET.$$

Estimating treatment effect

In case of no bias, mathematical expectations can be consistently estimated by averages, a consistent estimate of the average impact effect can be calculated as follows:

$$\bar{Y}_1 - \bar{Y}_0 = \widehat{ATET}$$

where $\bar{Y}_1$ is the sample average value of Y_i for objects included in the test group;

$\bar{Y}_0-$ sample average value of the dependent variable for objects included in the control group.

Estimating treatment effect

An estimate of the treatment effect can also be obtained using ordinary pairwise regression. To do this, you need to estimate the model parameters:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot D_i,$$

where $\hat{\beta}_2 = \bar{Y}_1 - \bar{Y}_0$.

If the experiment is constructed correctly, then in ordinary pairwise regression the explanatory variable is exogenous.

Therefore, ordinary pairwise regression provides an unbiased and consistent estimate of the average treatment effect. Therefore, it is not necessary to use **control variables** in the regression.

However, there are two reasons why their use can still be useful:

1. **Increased estimation precision:** Including control variables allows us to better describe the dependent variable, reduce the standard error of the regression, and obtain more accurate coefficient estimates.

2. **Checking the quality of randomization:** if the experiment is constructed correctly and the binary variable D is truly exogenous, then the estimates of the coefficient for this variable in paired and multiple regression should not differ much (since both estimates are consistent).

Estimating treatment effect

The average treatment effect on the treated (ATET) and the average treatment effect on the untreated (ATENT), defined respectively as:

$$\mathrm{ATET} = \mathrm{E}\,(Y_1 - Y_0 \mid D = 1)$$
$$\mathrm{ATENT} = \mathrm{E}\,(Y_1 - Y_0 \mid D = 0)$$

The ATET is the average treatment effect calculated within the subsample of treated units.

The ATENT is the average treatment effect calculated within the subsample of untreated units.

$$\mathrm{ATE} = \mathrm{ATET} \cdot p(D = 1) + \text{ATENT} \cdot p(D = 0)$$

We can also define the previous parameters as conditional on x as "individual-specific average treatment effects".

Distribution of ATE(x), ATET(x), and ATENT(x)

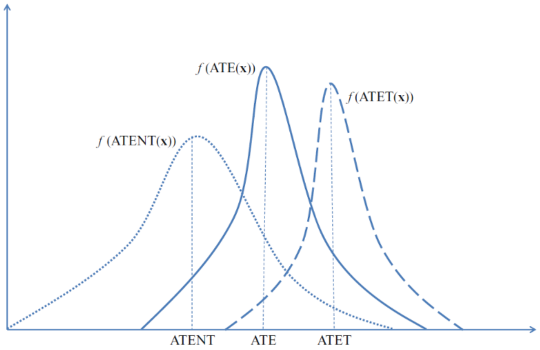


Figure:

SUTVA

We exclude the possibility that the treatment of one unit affects the outcome of another unit.

In the literature (Rubin 1978) this occurrence is called SUTVA—or stable unit treatment value assumption.

However, assuming that units are independent might be rather restrictive in many evaluation contexts.

Example

COVID-19 restrictions directly affected firms in tourism industry. This, however, caused a decline in firms of other business activity types located close. They were affected indirectly (a spillover effect).

Here SUTVA is not valid.

Estimating treatment effect

Regression-Adjustment

$$\text{ATE}(\mathbf{x}) = \mathrm{E}(Y \mid \mathbf{x}, D = 1) - \mathrm{E}(Y \mid \mathbf{x}, D = 0)$$

that can be interpreted as a conditional DIM estimator. By simply denoting:

$$m_1(\mathbf{x}) = \mathrm{E}(Y \mid \mathbf{x}, D = 1)$$

and

$$m_0(\mathbf{x}) = \mathrm{E}(Y \mid \mathbf{x}, D = 0)$$

we have that:

$$\text{ATE}(\mathbf{x}) = m_1(\mathbf{x}) - m_0(\mathbf{x})$$

If consistent estimators of $\mathrm{m}_1(\mathrm{x})$ and $\mathrm{m}_0(\mathrm{x})$ are available, causal parameters ATEs can be estimated through the sample equivalents:

Sample equivalents of ATEs in RA

$$\begin{aligned}
\widehat{\mathrm{ATE}} &= \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \left[\widehat{m}_1(\mathbf{x}_i) - \widehat{m}_0(\mathbf{x}_i)\right] \\
\widehat{\mathrm{ATET}} &= \tfrac{1}{N_1} \textstyle\sum_{i=1}^{N} D_i \cdot \left[\widehat{m}_1(\mathbf{x}_i) - \widehat{m}_0(\mathbf{x}_i)\right] \\
\widehat{\mathrm{ATENT}} &= \tfrac{1}{N_0} \textstyle\sum_{i=1}^{N} (1 - D_i) \cdot \left[\widehat{m}_1(\mathbf{x}_i) - \widehat{m}_0(\mathbf{x}_i)\right]
\end{aligned}$$

Both $m_1(x)$ and $m_0(x)$ can be estimated either parametrically, semi-parametrically, or non-parametrically.

Estimating treatment effect

The linear parametric RA assumes that $m_0(\mathbf{x}) = \mu_0 + \mathbf{x}\boldsymbol{\beta}_0$ and $m_1(\mathbf{x}) = \mu_1 + \mathbf{x}\boldsymbol{\beta}_1$, where μ_0 and μ_1 are scalars and $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are two vectors of parameters.

In such a case, applying RA implies estimating two distinct OLS regressions:

$Y_i = \mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0$ only on untreated

and

$Y_i = \mu_1 + \mathbf{x}_i\boldsymbol{\beta}_1$ only on treated units,

thus getting the predicted values $\widehat{m}_1(\mathbf{x}_i)$ and $\widehat{m}_0(\mathbf{x}_i)$.

These quantities can be used to recover all the causal parameters of interest by inserting them into the RA formulas for $\widehat{\mathrm{ATE}}, \widehat{\mathrm{ATET}}, \widehat{\mathrm{ATENT}}$.

Estimating treatment effect

Case 1 Homogenous reaction function of Y_0 and Y_1 to x and equivalent fraction of treated and control: $\beta_1 = \beta_2$

$$\mathrm{ATE} = \mathsf{ATE}(\mathbf{x}) = \mathsf{ATET} = \mathsf{ATET}(\mathbf{x}) = \mathsf{ATENT} = \mathsf{ATENT}(\mathbf{x})$$
$$= \mu_1 - \mu_0)\mathrm{E}(Y \mid D, \mathbf{x}) = \mu_0 + D \cdot \mathrm{ATE} + \mathbf{x}\beta$$

The ATE can be estimated by OLS:

$$\text{OLS}: \quad Y_i = \mu_0 + D_i\alpha + \mathbf{x}_i\boldsymbol{\beta}_0 + \text{ error }_i, \quad i = 1, \ldots, N$$

where $\alpha = \mathsf{ATE}$.

Partial randomization implies that $\mathrm{cov}(D, \varepsilon) = 0$, which is one of the key assumptions of ordinary least squares in obtaining unbiased estimates, independence of regressors from the disturbance term ε.

Its violation causes bias.

Estimating treatment effect

Imputation is based on conditioning over the values of one single variable x, which is supposed to take on four values: A, B, C, D.

Unit	D	x	$m_1 = E(Y\|D=1;x)$	$m_0 = E(Y\|D=0;x)$	$m_1 - m_0$	ATET	ATENT	ATE
1	1	A	25	**68**	**−43**			
2	1	B	65	**25**	**40**	**−1.5**		
3	1	C	36	**74**	**−38**			
4	1	D	47	**12**	**35**			
5	0	B	**65**	25	**40**			**6.3**
6	0	D	**47**	12	**35**			
7	0	D	**47**	12	**35**		**11.5**	
8	0	A	**25**	68	**−43**			
9	0	C	**36**	74	**−38**			
10	0	B	**65**	25	**40**			

Figure:

Some minimal units' overlap over x is necessary for imputation to be achieved (and, thus, for identifying treatment effects).

Difference-in-differences method

Motivating example [Card, Kreuger, 1994]:

In 1992, the state of New Jersey, USA, increased the minimum wage from \$4.25 to \$5.05.

Hypothesis: Economic theory suggests that such a decision should affect the employment of low-skilled workers (after all, it is their work that is often paid at the minimum rate).

A rise in minimum wages is believed to cause an increase in unemployment since employers would not be willing to pay wages at higher rates to the same number of employees.

Data: employment of workers in fast food restaurants, i.e. the dependent variable is the number of full-time workers employed at that restaurant.

Estimated ATE: the average change in fast food restaurant employment in New Jersey resulting from the passage of the new minimum wage law.

How could this change be calculated?

1) Compare data for New Jersey before and after the minimum wage change => Time effects are not considered.

2) Compare employment in the average restaurant in New Jersey (i.e., the treatment

Difference-in-differences method

The main factors that can influence occupancy in a typical restaurant:

- specific features of the state in which the restaurant is located (state effect);

- features of different periods of time, say, changes in economic conditions (temporary effect);

- the effect of changing the minimum wage (the same effect that we are trying to estimate).

Formally we can write it like this:

$$Y_{ist} = \alpha_s + \mu_t + \delta \cdot D_{ist} + \varepsilon_{ist},$$

where index i – restaurant number; Y_{ist} – the number of workers employed in this restaurant; the variable $D = 1$, for New Jersey after wages changed, and $D = 0$ otherwise;

α_s – state effect: α_{control} , for Pennsylvania; $\alpha_{\text{treatment}}$ for New Jersey;

μ_t – time effect: μ_{before} before the wage change and μ_{after} after the change;

δ **is the effect of a wage increase on employment**;

ε_{ist} – random errors of the model.

Difference-in-differences method

New Jersey before the wage change:

$$E(Y_{ist} \mid s = \text{ treatment }, t = \text{ before }) = \mu_{\text{before}} + \alpha_{\text{treatment}} \,.$$

New Jersey after the change:

$$E(Y_{ist} \mid s = \text{ treatment }, t = \text{ after }) = \mu_{\text{after}} + \alpha_{\text{treatment}} + \delta.$$

The expected change in employment in New Jersey:

$$\Delta_{\text{treatment}} = \mu_{\text{after}} - \mu_{\text{before}} + \delta$$

Pennsylvania before the wage change:

$$E(Y_{ist} \mid s = \text{ control}, t = \text{ before }) = \mu_{\text{before}} + \alpha_{\text{control}} \,.$$

Pennsylvania after the change:

$$E(Y_{ist} \mid s = \text{ control }, t = \text{ after }) = \mu_{\text{after}} + \alpha_{\text{control}} \,.$$

The expected change in employment in Pennsylvania:

$$\Delta_{\text{control}} = \mu_{\text{after}} - \mu_{\text{before}} \,.$$

The expected change in employment in New Jersey:

$$\Delta_{\text{treatment}} = \mu_{\text{after}} - \mu_{\text{before}} + \delta$$

The expected change in employment in Pennsylvania:

$$\Delta_{\text{control}} = \mu_{\text{after}} - \mu_{\text{before}} .$$

Finally,

$$\Delta_{\text{treatment}} - \Delta_{\text{control}} = \delta$$

Thus, the treatment effect can be represented as the difference between the differences of conditional mathematical expectations:

$$\begin{gathered}\delta = \Delta_{\text{treatment}} - \Delta_{\text{control}} = \\ = [E\,(Y_{\text{ist}} \mid s = \text{ treatment }, t = \text{ after }) - E\,(Y_{\text{ist}} \ \ mids = \text{ treatment }, t = \text{ before })] - \\ - [E\,(Y_{\text{ist}} \mid s = \text{ control }, t = \text{ after }) - E\,(Y_{\text{ist}} \ \ mids = \text{ control }, t = \text{ before })] .\end{gathered}$$

And by LLN, a consistent estimate of each of these mathematical expectations is the corresponding average value. Therefore:

$$\hat{\delta} = \left[\bar{Y}_{\text{treatment, after}} - \bar{Y}_{\text{treatment, before}}\right] - \left[\bar{Y}_{\text{control, after}} - \bar{Y}_{\text{control, before}}\right],$$

Difference-in-differences method

Estimates of an increase in the minimum wage on employment using the difference-in-differences method

Variable	Pennsylvania (1)	New Jersey (2)	(3) = (2) − (1)
1. Average number of employees before the change	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)
2. Average number of employees after the change	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)
3. Change in average number of employees	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

The treatment effect:

$$\hat{\delta} = \left[\bar{Y}_{\text{treatment,after}} - \bar{Y}_{\text{treatment, before}}\right] - \left[\bar{Y}_{\text{control, after}} - \bar{Y}_{\text{control, before}}\right] =$$
$$= [21.03 - 20.44] - [21.17 - 23.33] = 2.76.$$

We can conclude that raising the minimum wage increased the equilibrium level of employment in New Jersey fast food restaurants by an average of 2.76 employee.

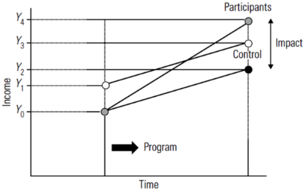
This result contradicts the conclusions of standard theoretical models from microeconomics (or labor economics), and therefore has caused widespread discussion in the literature.

Described example is **a quasi-experiment**

Graphical illustration:

$$DD = E(Y_1^T - Y_0^T \mid T_1 = 1) - E(Y_1^C - Y_0^C \mid T_1 = 0)\cdot \qquad (5.1)$$

Figure 5.1 An Example of DD



$$DD = (Y_4 - Y_0) - (Y_3 - Y_1).$$

$$DD = (Y_4 - Y_2)$$

Figure:

Difference-in-differences method

The difference-in-differences method is directly related to the estimation of models using regressions.

$$Y_{it} = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot z_t + \delta \cdot x_i \cdot z_t + \varepsilon_{it},$$

where x_i is a binary variable that equals one if the ith restaurant is located in New Jersey (i.e., belongs to the test group);

$z_t -$ is a binary variable that is equal to one for all observations related to the second period (the period after the minimum wage increase).

Difference-in-differences method

In this case, applying the least squares method, we obtain the following equation:

$$\hat{Y}_{it} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i + \hat{\beta}_2 \cdot z_t + \hat{\delta} \cdot x_i \cdot z_t.$$

where FE-estimator:

$$\begin{gathered}
\hat{\beta}_0 = \bar{Y}_{\text{control, before}}\ ; \\
\hat{\beta}_1 = \bar{Y}_{\text{treatment,before}} - \bar{Y}_{\text{control, before}}\ ; \\
\hat{\beta}_2 = \bar{Y}_{\text{control, after}} - \bar{Y}_{\text{control, before}}\ ; \\
\hat{\delta} = \left[\bar{Y}_{\text{treatment, after}} - \bar{Y}_{\text{treatment,before}}\right] - \left[\bar{Y}_{\text{control, after}} - \bar{Y}_{\text{control, before}}\right].
\end{gathered}$$

Thus, the coefficient of the product $x_i \cdot z_t$ is equal to the same estimate of the impact effect that we derived above.

An equivalent estimation method is to apply OLS to the following pairwise regression:

$$\Delta Y_i = \beta_2 + \delta \cdot x_i + u_i,$$

where x_i is still a binary variable that is equal to one if the i-th object belongs to the test group;

hence, DD-estimator: $\Delta Y_i = Y_{i1} - Y_{i0}$.

Difference-in-differences method

DD estimator by OLS is unbiased if

1. The model in equation (outcome) is correctly specified. For example, the additive structure imposed is correct.

2. The error term consists of time-invariant and time-varying parts $\eta_i + \varepsilon_{it}$.

– ε_{it} is supposed to be uncorrelated with the other variables in the equation, including treatment.

– η_i though could be correlated with treatment and other unobserved characteristics in ε_{it} can be ruled out of the equation if estimated with fixed effects or in first differences.

3. An assumption that besides the treatment nothing would affect the difference ρ between the treated and the control groups is known as the parallel-trend assumption (PTA). It is the most critical in DD.

It means that unobserved characteristics affecting program participation do not vary over time with treatment status.

A Comparison Between DD and FE Estimator

DD is the estimator in first differences.

FE is the estimator in differences from the average.

Therefore, they have different form of conditional independence assumption.

For DD we require

$$\begin{aligned} &\text{cov}\left(T_i, \Delta\varepsilon_{it}\right) = 0 \text{ or} \\ &\text{cov}\left(T_i, \varepsilon_{it-1}\right) = \text{cov}\left(T_i, \varepsilon_{it}\right) \end{aligned}$$

Whereas for FE we require more restrictive assumption (that also implies CIA for DD)

$$\begin{aligned} \text{cov}\left(\Delta T_{it}, \Delta\varepsilon_{it}\right) = {} & \text{cov}\left(T_{it}, \varepsilon_{it}\right) - \text{cov}\left(T_{it-1}, \varepsilon_{it}\right) \\ & - \text{cov}\left(T_{it}, \varepsilon_{it-1}\right) + \text{cov}\left(T_{it-1}, \varepsilon_{it-1}\right) = 0 \end{aligned}$$

PSM with DD (Doubly-robust estimand)

DD can be combined with PSM to better match control and project units on pre-program characteristics.

Combining different methods may sometimes lead to an estimation of the treatment effects having **better properties in terms of robustness**.

The robustness of this approach lies in the fact that either the **conditional mean** or the **propensity-score** needs to be correctly specified but not both.

The propensity score can be used to match participant and control units, and the treatment effect is calculated across participant and matched control units **within the common support**.

PSM with DD (Doubly-robust estimand)

The idea is to find, from a large group of nonparticipants, individuals who are observationally similar to participants in terms of characteristics not affected by the program.

- When a treatment cannot be randomized, the next best thing to do is to try to mimic randomization—that is, try to have an observational analogue of a randomized experiment.

- Matching is not the only way to eliminate bias (e.g. regressions with control variables and/or instruments).

- Matching is a non-parametric method. You do not need to assume any functional form for the causal relationship being investigated.

PSM with DD (Doubly-robust estimand)

Propensity score matching

If the vector of explanatory variables has a large dimension, or if there are continuous variables among the variables, then an exact comparison is not entirely convenient.

In this case, a propensity score is used - the conditional probability that an object will be affected given the given values of the regressors.

The propensity score is usually estimated using a logit or probit model.

$$P\left(D_i = 1 \mid x_i^{(1)}, x_i^{(2)} \ldots, x_i^{(k)}\right) = \\ = p\left(x_i^{(1)}, x_i^{(2)} \ldots, x_i^{(k)}\right)$$

Thus, matching is carried out in two stages:

1. For each observation, the value of the propensity measure is estimated (for example, based on a logit model)

2. Then a comparison of objects with similar values of the propensity measure is carried out using

– Nearest neighbor matching

– Comparison with stratification

– Radial matching

PSM with DD (Doubly-robust estimand)

The validity of PSM depends on two conditions:

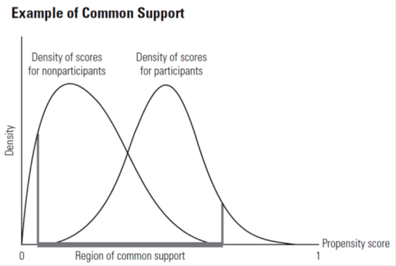
A1: conditional independence (namely, that unobserved factors do not affect participation)

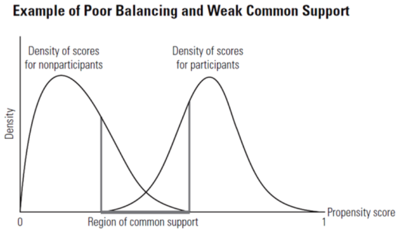
A2: sizable common support or overlap in propensity scores across the participant and nonparticipant samples.

If one assumes that **differences in participation** are based solely on **differences in observed characteristics**, and if enough nonparticipants are available to match with participants, the corresponding treatment effect can be measured even if treatment is not random.

A3: Independent observations ensure that the outcome and treatment for one individual has no effect on the outcome or treatment for any other individual.

PSM with DD (Doubly-robust estimand)





Counterfactual

With matching methods, one tries to develop a **counterfactual or control group** that is as similar to the treatment group as possible in terms of observed characteristics.

Each participant is matched with an observationally similar nonparticipant, and then the average difference in outcomes across the two groups is compared to get the program treatment effect

The problem is to **credibly identify groups** that look alike.

PSM with DD (Doubly-robust estimand)

Example:

"Determining risk premium for green bonds"

Summary

1) Causal effect inference: problem of counterfactual

2) DiD estimator for reducing sampling biases

3) Matching + DiD for a doubly robust estimator