THE UNIVERSITY of EDINBURGH

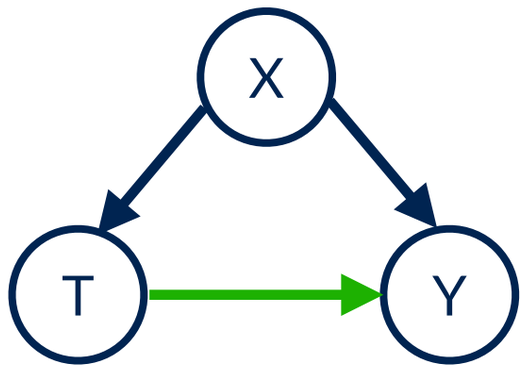
Methods for Causal Inference
Lecture 8: Difference in difference

Ava Khamseh

School of Informatics
2025-2026

Average Treatment Effect (ATE)

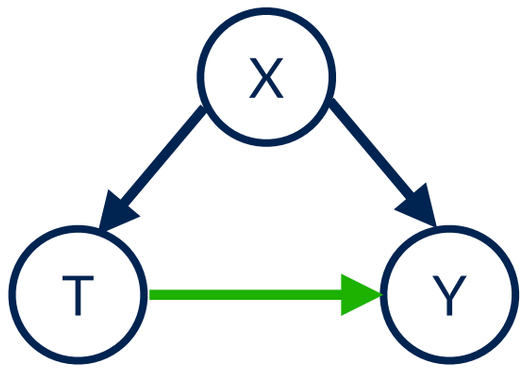
$$\mathbb{E}[Y_1 - Y_0]$$



Average Treatment Effect (ATE)

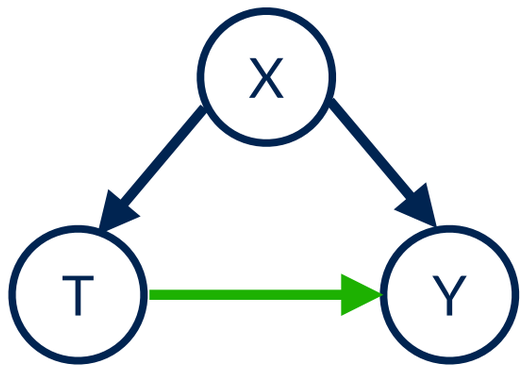
$$\mathbb{E}[Y_1 - Y_0] = \mathbb{E}_X\left[\mathbb{E}[(Y_1 - Y_0|X)]\right]$$

Iterated expectation



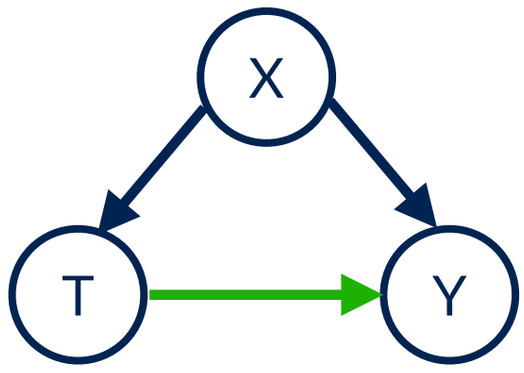
Average Treatment Effect (ATE)

$$
\begin{aligned}
\mathbb{E}[Y_1 - Y_0] &= \mathbb{E}_X\left[\mathbb{E}[(Y_1 - Y_0|X)]\right] && \text{Iterated expectation} \\
&= \mathbb{E}_X\left[\mathbb{E}[(Y_1|X)] - \mathbb{E}[(Y_0|X)]\right] && \text{Linearity} \\
&= \mathbb{E}_X\left[\mathbb{E}[(Y_1|T=1,X)] - \mathbb{E}[(Y_0|T=0,X)]\right] && \text{Unconfoundedness} \\
& && Y_1, Y_0 \perp\!\!\!\perp T \mid X
\end{aligned}
$$



Average Treatment Effect (ATE)

$$
\begin{aligned}
\mathbb{E}[Y_1 - Y_0] &= \mathbb{E}_X\left[\mathbb{E}[(Y_1 - Y_0|X)]\right] && \text{Iterated expectation} \\
&= \mathbb{E}_X\left[\mathbb{E}[(Y_1|X)] - \mathbb{E}[(Y_0|X)]\right] && \text{Linearity} \\
&= \mathbb{E}_X\left[\mathbb{E}[(Y_1|T=1,X)] - \mathbb{E}[(Y_0|T=0,X)]\right] && \text{Unconfoundedness} \\
&= \mathbb{E}_X\left[\mathbb{E}[Y|T=1,X] - \mathbb{E}[Y|T=0,X]\right] && Y_1, Y_0 \perp\!\!\!\perp T \mid X
\end{aligned}
$$

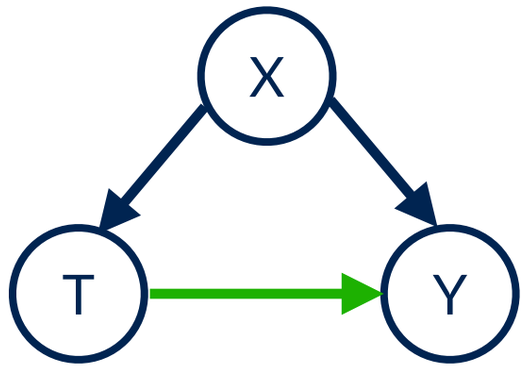


Average Treatment Effect (ATE)

$$\mathbb{E}[Y_1 - Y_0]$$

Observed Confounders only, ATE estimated via:

1) Regression (covariate) adjustment



$$ATE = \mathbb{E}_X\Big[\mathbb{E}[Y|T=1,X] - \mathbb{E}[Y|T=0,X]\Big]$$

Estimated as β_t (under e.g. **linearity** assumption) via:

$$\mathbb{E}[Y|T,X] = \alpha_0 + \beta_x X + \beta_t T$$

(Note: The estimate will be different under a different model, be aware of model-misspecification!!)

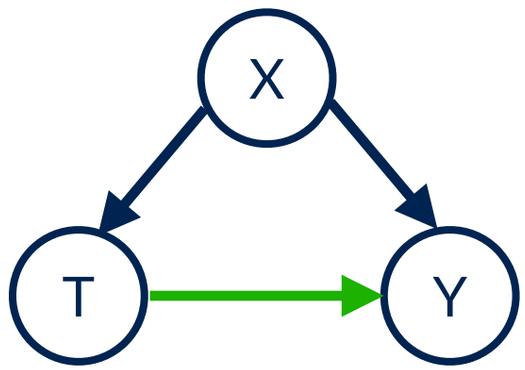
Average Treatment Effect (ATE)

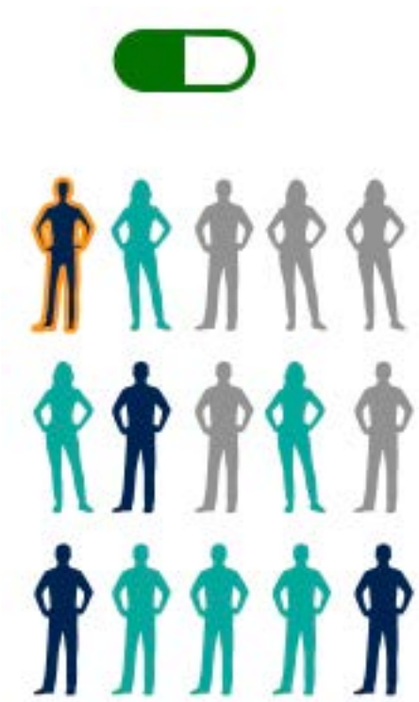
$$\mathbb{E}[Y_1 - Y_0]$$

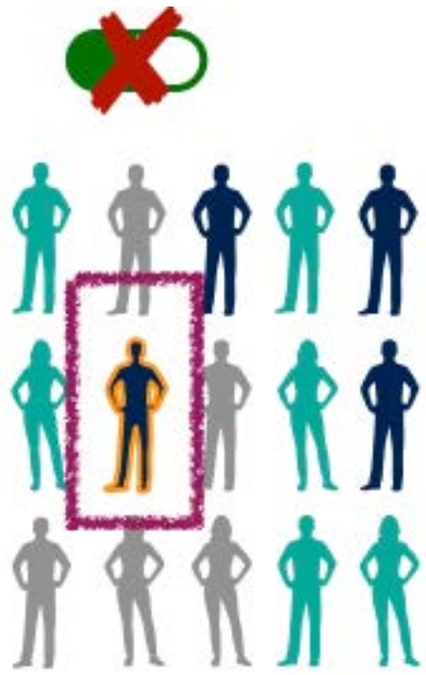
Observed Confounders only, ATE estimated via:

1) Regression (covariate) adjustment
2) Propensity score matching and IPTW

$$e(x) = p(t = 1|x)$$





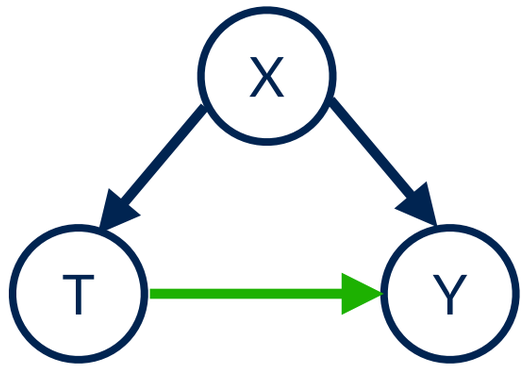


Average Treatment Effect (ATE)

$$\mathbb{E}[Y_1 - Y_0]$$

Observed Confounders only, ATE estimated via:

1) Regression (covariate) adjustment
2) Propensity score matching and IPTW



$$\boxed{e(x) = p(t = 1|x)}$$

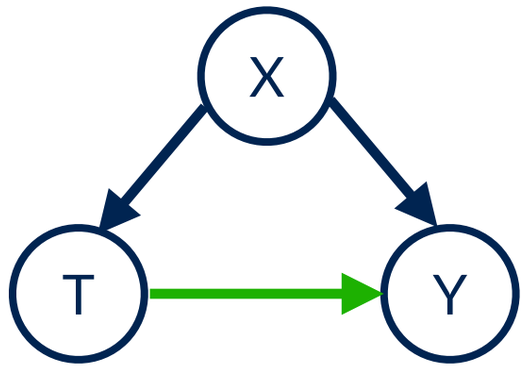
$$\frac{1}{N_1} \sum_{\text{treated}} y_{\underline{1}}^{(i)} \frac{1}{e(x_i)} - \frac{1}{N_0} \sum_{\text{not treated}} y_{\underline{0}}^{(i)} \frac{1}{1 - e(x_i)}$$

Average Treatment Effect (ATE)

$$\mathbb{E}[Y_1 - Y_0]$$

Observed Confounders only, ATE estimated via:

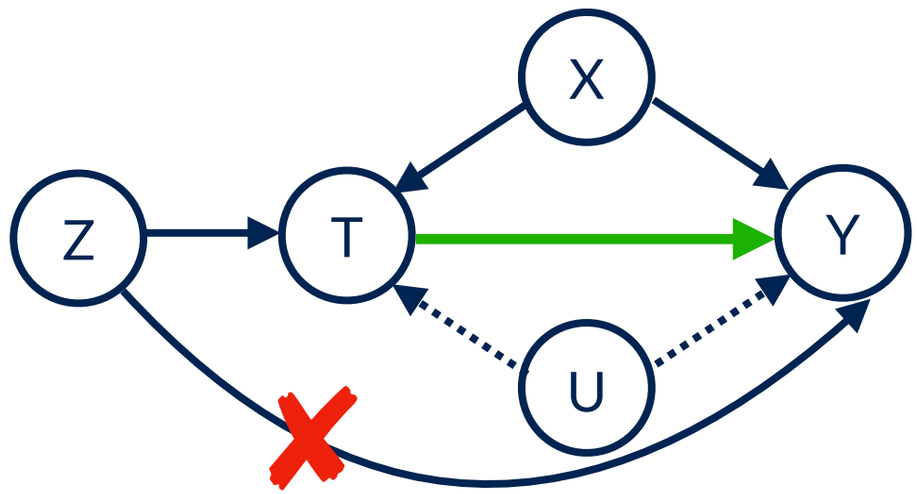
1) Regression (covariate) adjustment
2) Propensity score matching and IPTW



Unobserved Confounders, ATE estimated via:

1) Instrumental variable

$$\tau = \frac{\mathbb{E}\left[(Y|z=1) - (Y|z=0)\right]}{\mathbb{E}\left[(T|z=1) - (T|z=0)\right]}$$

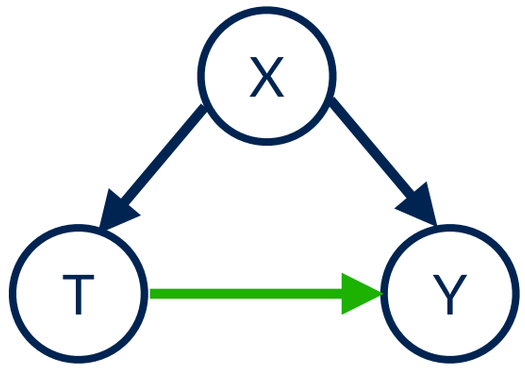


Average Treatment Effect (ATE)

$$\mathbb{E}[Y_1 - Y_0]$$

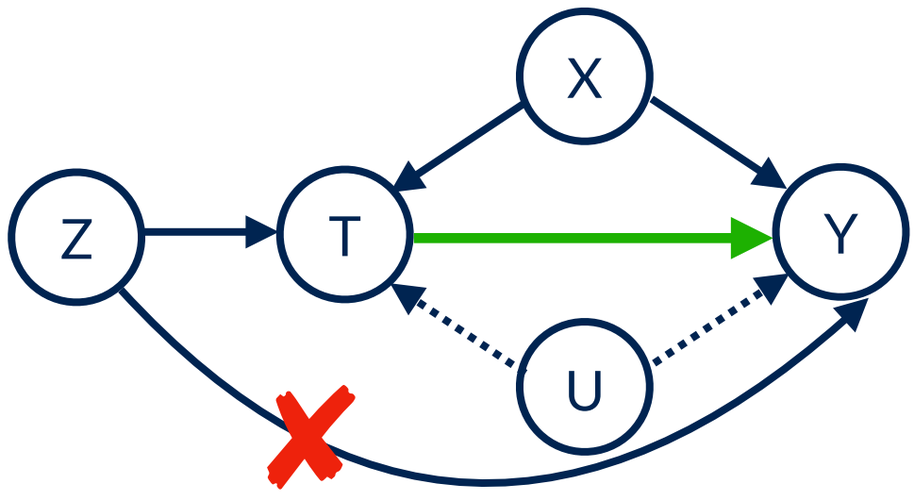
Observed Confounders only, ATE estimated via:

1) Regression (covariate) adjustment
2) Propensity score matching and IPTW



Unobserved Confounders, ATE estimated via:

1) Instrumental variable
2) Front door criterion (Pearl, see later)



Difference in Difference: Examples

We wish to estimate the impact of a treatment/policy T applied at time s
on some outcome Y by using information **before** and **after** the treatment

Examples:

- What is the effect of government policy for renewable energy (e.g. T=subsidies, advertisement, …) on people switching to renewable energy sources (e.g Y= installing solar panels, switching to hydro, …)

- What is the effect of NHS funding for research in AI
  on improving hospital wait-time

- What is the effect of a internship on students' in getting a job

Any **issues** with the above set up?

Difference in Difference: Examples

We wish to estimate the impact of a treatment/policy T applied at time s
on some outcome Y by using information **before** and **after** the treatment

Examples:

- What is the effect of government policy for renewable energy (e.g. T=subsidies, advertisement, …) on people switching to renewable energy sources (e.g Y= installing solar panels, switching to hydro, …)

- What is the effect of NHS funding for research in AI
on improving hospital wait-time

- What is the effect of a internship on students' in getting a job

Issue: Control group? Was the increase/decrease due to the treatment or would we have seen the same increase/decrease even without the intervention?

Difference in Difference

We wish to estimate the impact of a treatment/policy T applied at time s on some outcome Y by using information **before** and **after** the treatment

We have two components:

1) Treatment group and control groups.
   Treatment group: Receives treatment at some point in time s
   Control group: Does **not** receive treatment (within considered time frame)

2) **Before** vs **after** (the treatment is applied),
   i.e. we have data from before and after the intervention is applied

Difference in Difference

We wish to estimate the impact of a treatment/policy T applied at time s on some outcome Y by using information **before** and **after** the treatment

We have two components:

1) Treatment group and control groups.
   Treatment group: Receives treatment at some point in time s
   Control group: Does **not** receive treatment (within considered time frame)

2) **Before** vs **after** (the treatment is applied),
   i.e. we have data from before and after the intervention is applied

If the treatment assignment is completely random, then could consider treatment and control to only differ based on the intervention (all else being the same) …

Difference in Difference

We wish to estimate the impact of a treatment/policy T applied at time s on some outcome Y by using information **before** and **after** the treatment

We have two components:

1) Treatment group and control groups.
   Treatment group: Receives treatment at some point in time s
   Control group: Does **not** receive treatment (within considered time frame)

2) **Before** vs **after** (the treatment is applied),
   i.e. we have data from before and after the intervention is applied

If the treatment assignment is completely random, then could consider treatment and control to only differ based on the intervention (all else being the same) …
but most of the time this is not random …

Difference in Difference

Treatment assignment is not random and the two groups may well have different **starting points**

Examples: Perhaps …

- People in California are already more likely to switch to solar panels than people in Scotland (e.g. due to Tesla's offer, affordability, sun, …)

- NHS Scotland might have a shorter wait time than NHS England (e.g. due to number of patients, funding available, …)

- Students at university A who get an internship already have a better performance at the university (e.g. due prior experience with coding, more advance courses, …) than students in university B

Difference in Difference

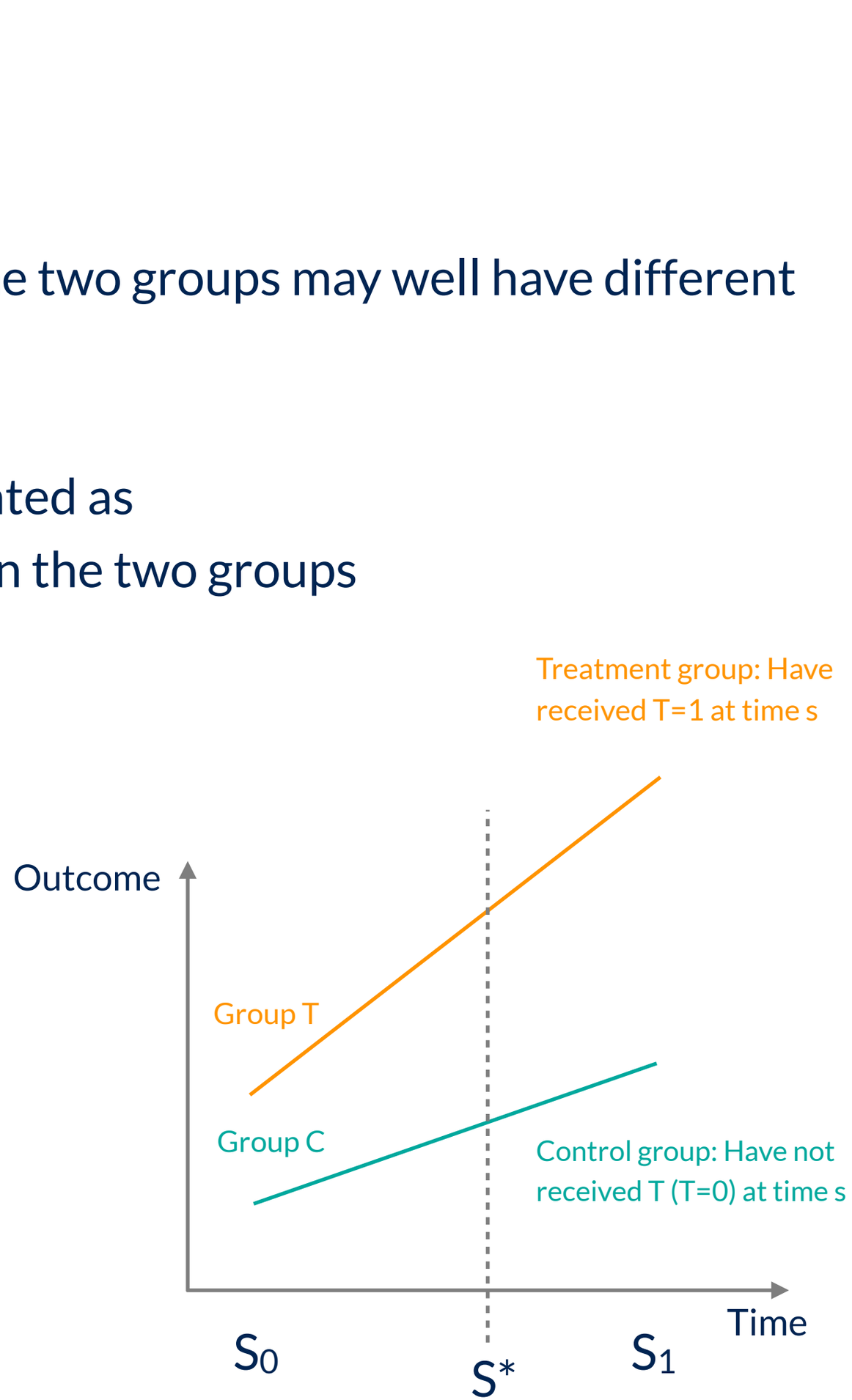
Treatment assignment is not random and the two groups may well have different **starting points**

DiD: Treatment effect on outcome is estimated as
the difference in changes over time between the two groups

i.e. difference in trends

Before time S, neither group has received a treatment

After time S, Group T has received the treatment but group C has not



Difference in Difference

Treatment assignment is not random and the two groups may well have different **starting points**

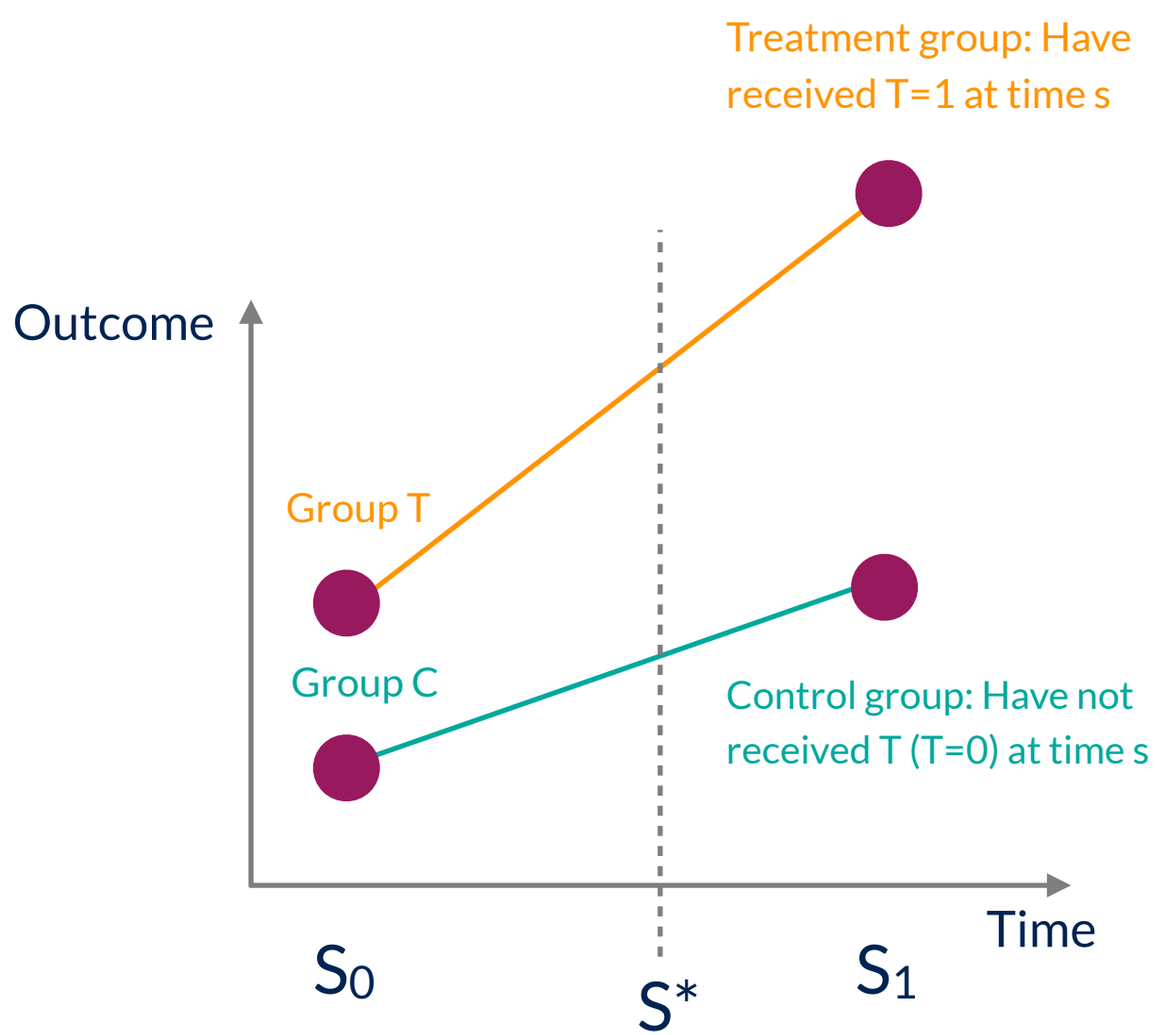
DiD: Treatment effect on outcome is estimated as
the difference in changes over time between the two groups

i.e. difference in trends

Before time S, neither group has received a treatment

After time S, Group T has received the treatment but group C has not

● = measured (line is for visualisation only!)



Average Treatment effect of the Treated (ATT)

ATE: $\mathbb{E}[Y_1 - Y_0] = \mathbb{E}_X\left[\mathbb{E}[Y|T=1,X] - \mathbb{E}[Y|T=0,X]\right]$ Slide 1

ATT:

$$\mathbb{E}[Y_1 - Y_0|T=1] = \mathbb{E}_X\left[\mathbb{E}[(Y_1 - Y_0|T=1,X)]\Big|T=1\right]$$

Iterated expectation

$$= \sum_x \mathbb{E}[Y_1 - Y_0|T=1, X=x]p(X=x|T=1)$$

Average Treatment effect of the Treated (ATT)

ATE: $\mathbb{E}[Y_1 - Y_0] = \mathbb{E}_X\left[\mathbb{E}[Y|T=1,X] - \mathbb{E}[Y|T=0,X]\right]$ Slide 1

ATT:

$$\mathbb{E}[Y_1 - Y_0|T=1] = \mathbb{E}_X\left[\mathbb{E}[(Y_1 - Y_0|T=1,X)]\Big|T=1\right]$$ Iterated expectation

$$= \sum_x \mathbb{E}[Y_1 - Y_0|T=1,X=x]p(X=x|T=1)$$

Weaker Unconfoundedness to either remove T or set it to zero

$$= \sum_x \mathbb{E}[Y|T=1,X=x]p(X=x|T=1) - \sum_x \mathbb{E}[Y|T=0,X=x]p(X=x|T=1)$$

$Y_0 \perp\!\!\!\perp T \mid X$

Average Treatment effect of the Treated (ATT)

ATE: $\mathbb{E}[Y_1 - Y_0] = \mathbb{E}_X\left[\mathbb{E}[Y|T=1,X] - \mathbb{E}[Y|T=0,X]\right]$ Slide 1

ATT:

$$\mathbb{E}[Y_1 - Y_0|T=1] = \mathbb{E}_X\left[\mathbb{E}[(Y_1 - Y_0|T=1,X)]\Big|T=1\right]$$

Iterated expectation

$$= \sum_x \mathbb{E}[Y_1 - Y_0|T=1,X=x]p(X=x|T=1)$$

$$= \sum_x \mathbb{E}[Y|T=1,X=x]p(X=x|T=1) - \sum_x \mathbb{E}[Y|T=0,X=x]p(X=x|T=1)$$

$$= \sum_{x,y} yp(Y|T=1,X=x)p(X=x|T=1) - \sum_x \mathbb{E}[Y|T=0,X=x]p(X=x|T=1)$$

Average Treatment effect of the Treated (ATT)

ATE: $\mathbb{E}[Y_1 - Y_0] = \mathbb{E}_X\left[\mathbb{E}[Y|T=1,X] - \mathbb{E}[Y|T=0,X]\right]$ Slide 1

ATT:

Iterated expectation

$$\begin{aligned}
\mathbb{E}[Y_1 - Y_0|T=1] &= \mathbb{E}_X\left[\mathbb{E}[(Y_1 - Y_0|T=1,X)]\Big|T=1\right] \\
&= \sum_x \mathbb{E}[Y_1 - Y_0|T=1,X=x]p(X=x|T=1) \\
&= \sum_x \mathbb{E}[Y|T=1,X=x]p(X=x|T=1) - \sum_x \mathbb{E}[Y|T=0,X=x]p(X=x|T=1) \\
&= \sum_{x,y} yp(Y|T=1,X=x)p(X=x|T=1) - \sum_x \mathbb{E}[Y|T=0,X=x]p(X=x|T=1) \\
&= \sum_{x,y} yp(Y,X=x|T=1) - \sum_x \mathbb{E}[Y|T=0,X=x]p(X=x|T=1) \\
&= \sum_y yp(Y|T=1) - \sum_x \mathbb{E}[Y|T=0,X=x]p(X=x|T=1) \\
&= \mathbb{E}[Y|T=1] - \sum_x \mathbb{E}[Y|T=0,X=x]p(X=x|T=1)
\end{aligned}$$

Average Treatment effect of the Treated (ATT)

ATE: $\mathbb{E}[Y_1 - Y_0] = \mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0]$ No confounders

ATT = ATE only when there are no confounders!

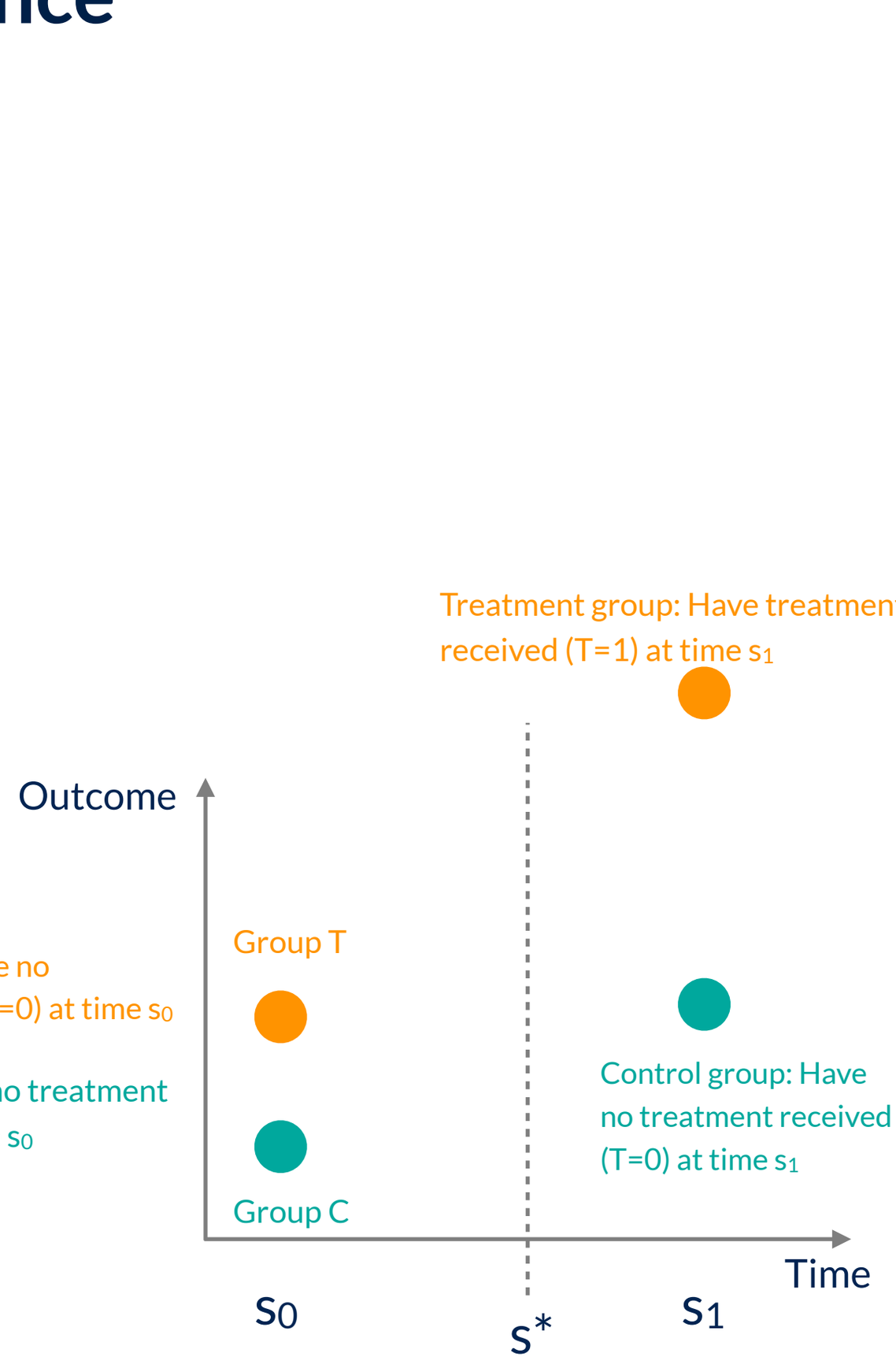
Back to Difference in Difference

Treatment group: Have no treatment received (T=0) at time s_0

Control group: Have no treatment received (T=0) at time s_0

Treatment group: Have treatment received (T=1) at time s_1

Control group: Have no treatment received (T=0) at time s_1



● = measured

Back to Difference in Difference

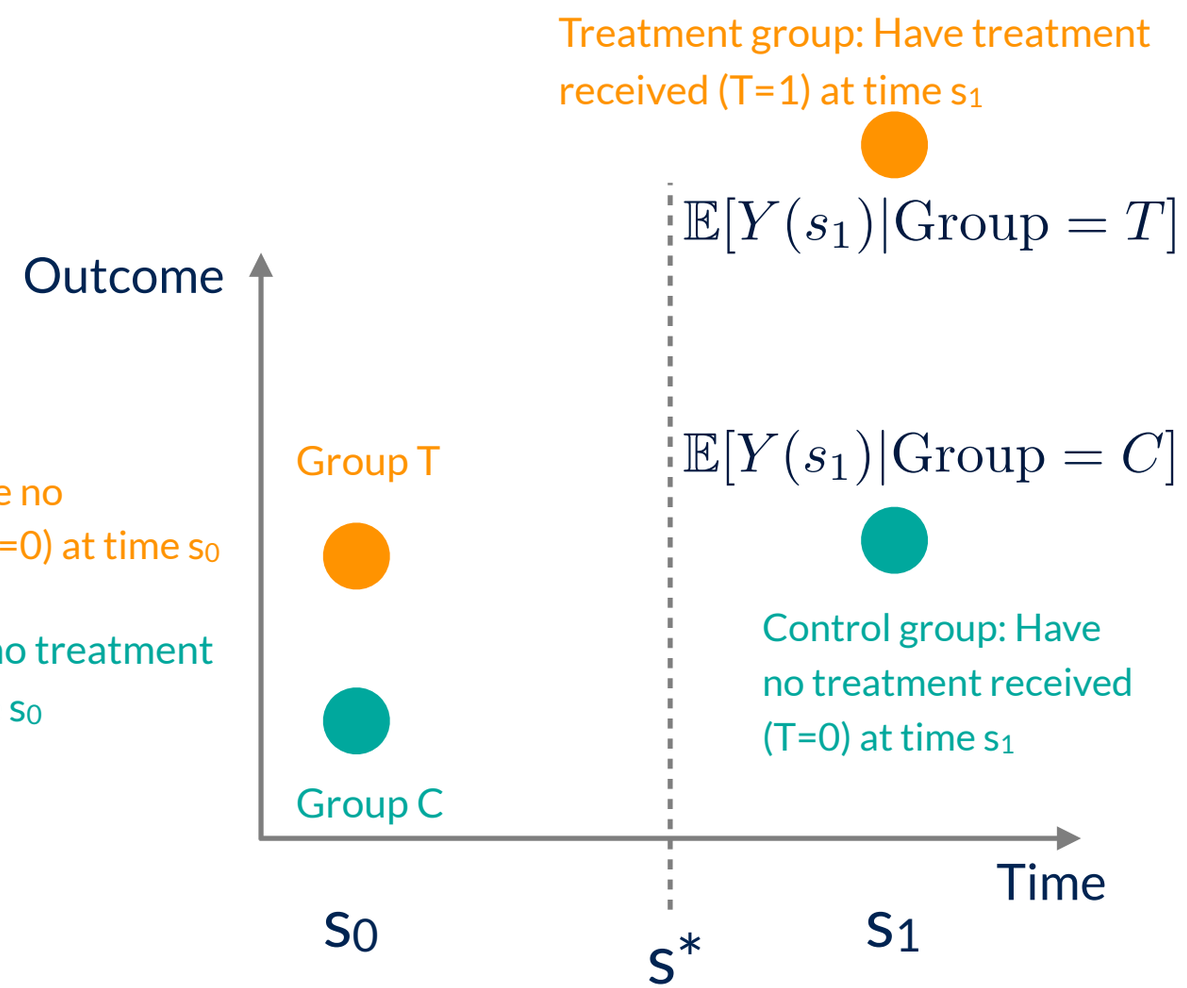
$$\mathbb{E}[Y(s_0)|\text{Group} = T]$$

Treatment group: Have no treatment received (T=0) at time s_0

$$\mathbb{E}[Y(s_0)|\text{Group} = C]$$

Control group: Have no treatment received (T=0) at time s_0

● = measured



Back to Difference in Difference

We want the ATT: $\mathbb{E}[Y_1 - Y_0 | \text{Group} = T]$

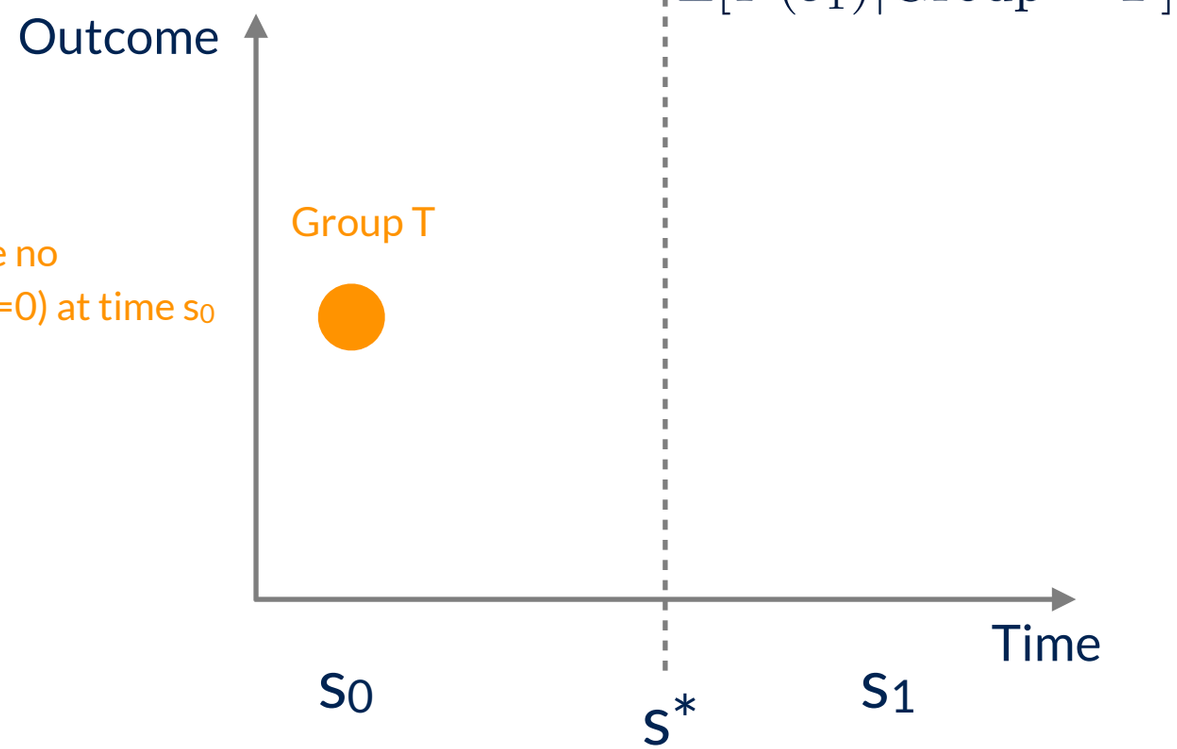
i.e. amongst the treated, how does giving the treatment compare with not giving the treatment (causally)

$\mathbb{E}[Y(s_0)|\text{Group} = T]$

Treatment group: Have no treatment received (T=0) at time s_0

Treatment group: Have treatment received (T=1) at time s_1

$\mathbb{E}[Y(s_1)|\text{Group} = T]$



Back to Difference in Difference

We want the ATT: $\mathbb{E}[Y_1 - Y_0 | \text{Group} = T]$

i.e. amongst the treated, how does giving the treatment compare with not giving the treatment (causally)

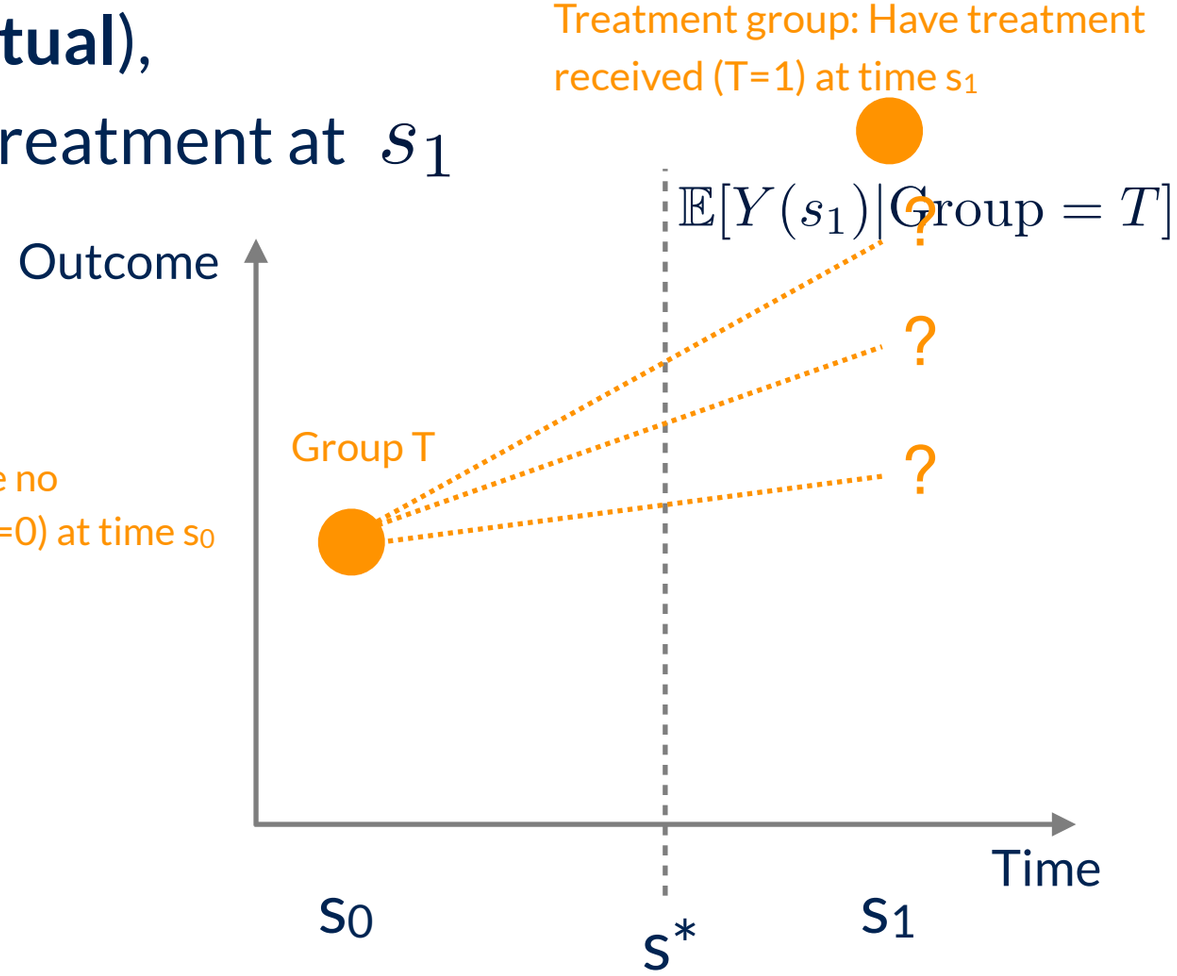
BUT: $\mathbb{E}[Y_1 | \text{Group} = T]$ is actually measured: $\mathbb{E}[Y(s_1) | \text{Group} = T]$

while $\mathbb{E}[Y_0 | \text{Group} = T]$ is not (**counterfactual**),

Because in Group T everyone receives the treatment at s_1

$\mathbb{E}[Y(s_0) | \text{Group} = T]$

Treatment group: Have no treatment received (T=0) at time s_0



Back to Difference in Difference

We want the ATT: $\mathbb{E}[Y_1 - Y_0 | \text{Group} = T]$

i.e. amongst the treated, how does giving the treatment compare with not giving the treatment (causally)

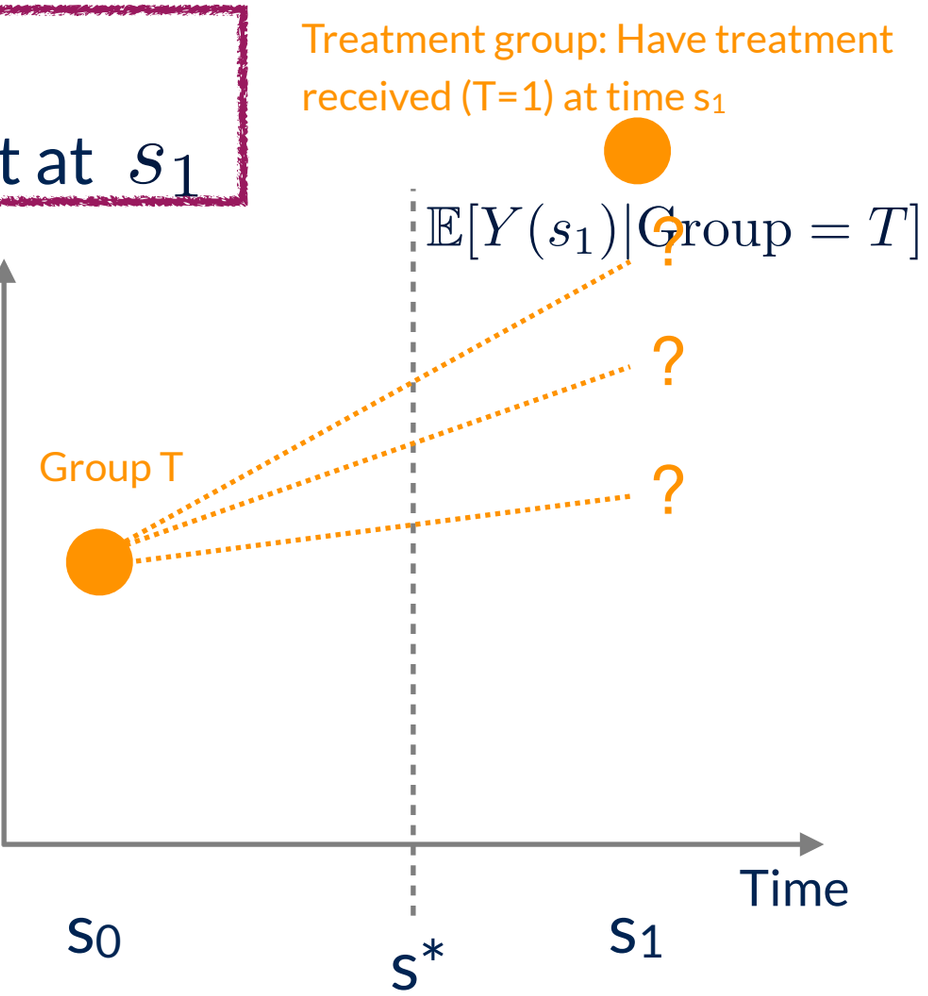
BUT: $\mathbb{E}[Y_1 | \text{Group} = T]$ is actually measured: $\mathbb{E}[Y(s_1) | \text{Group} = T]$

while $\mathbb{E}[Y_0 | \text{Group} = T]$ is not (**counterfactual**),
Because in Group T everyone receives the treatment at s_1

$\mathbb{E}[Y(s_0) | \text{Group} = T]$

Treatment group: Have no treatment received (T=0) at time s_0

Treatment group: Have treatment received (T=1) at time s_1



Utilise the control for this!

Back to Difference in Difference

We want the ATT: $\mathbb{E}[Y_1 - Y_0 | \text{Group} = T]$

i.e. amongst the treated, how does giving the treatment compare with not giving the treatment (causally)

BUT: $\mathbb{E}[Y_1 | \text{Group} = T]$ is actually measured: $\mathbb{E}[Y(s_1) | \text{Group} = T]$

while $\mathbb{E}[Y_0 | \text{Group} = T]$ is not (**counterfactual**),

Because in Group T everyone receives the treatment at s_1

$\mathbb{E}[Y(s_0) | \text{Group} = T]$

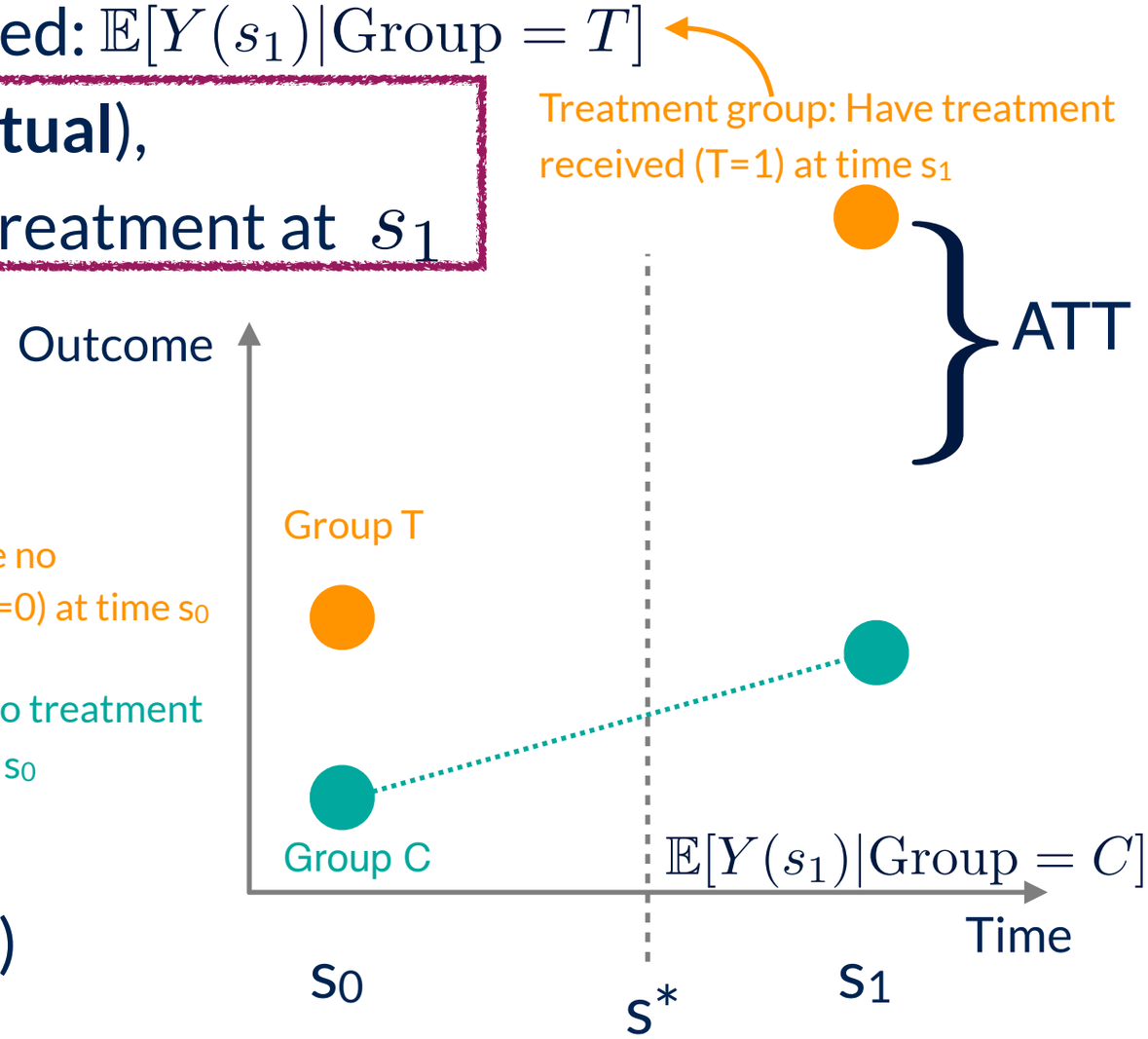
Treatment group: Have no treatment received (T=0) at time s_0

$\mathbb{E}[Y(s_0) | \text{Group} = C]$

Control group: Have no treatment received (T=0) at time s_0

Utilise the control for this!

Assumption: **Parallel trends** (same gradient)



Back to Difference in Difference

We want the ATT: $\mathbb{E}[Y_1 - Y_0 | \text{Group} = T]$

i.e. amongst the treated, how does giving the treatment compare with not giving the treatment (causally)

BUT: $\mathbb{E}[Y_1 | \text{Group} = T]$ is actually measured: $\mathbb{E}[Y(s_1) | \text{Group} = T]$
while $\mathbb{E}[Y_0 | \text{Group} = T]$ is not (**counterfactual**),

Treatment group: Have treatment received (T=1) at time s_1

Because in Group T everyone receives the treatment at s_1

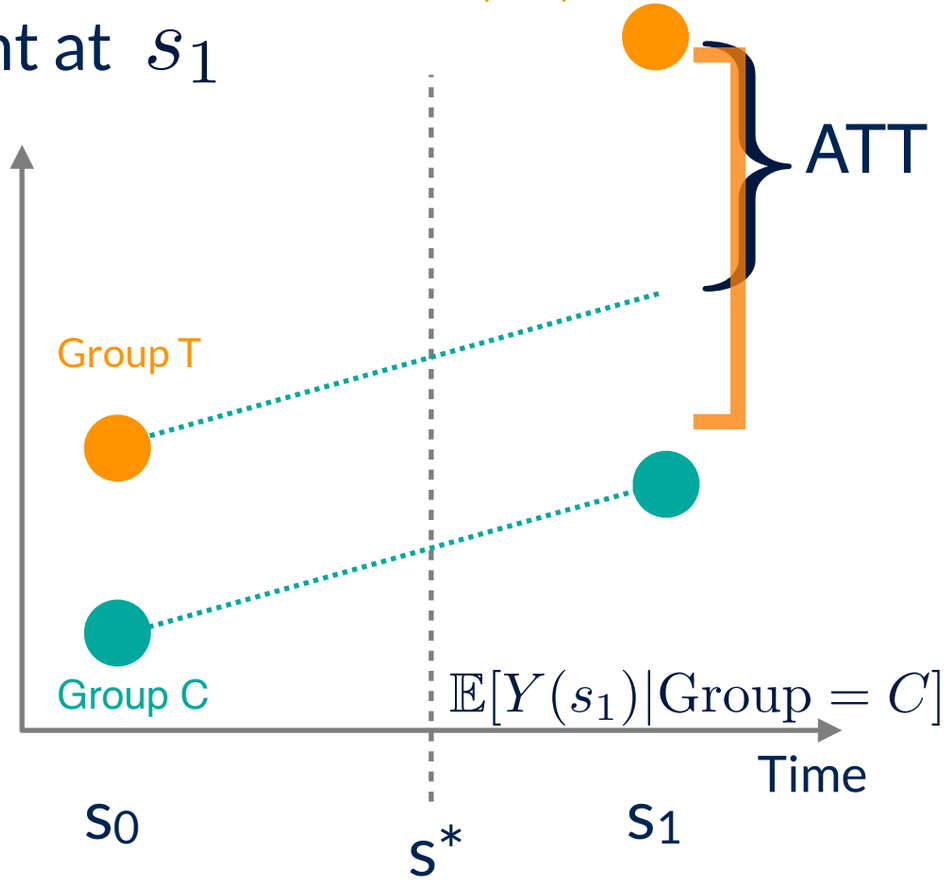
$$\mathbb{E}[Y(s_0) | \text{Group} = T]$$

Treatment group: Have no treatment received (T=0) at time s_0

$$\mathbb{E}[Y(s_0) | \text{Group} = C]$$

Control group: Have no treatment received (T=0) at time s_0

$$\mathbb{E}[Y_1 - Y_0 | \text{Group} = T] = (\mathbb{E}[Y(s_1) | \text{Group} = T] - \mathbb{E}[Y(s_0) | \text{Group} = T]) - (\mathbb{E}[Y(s_1) | \text{Group} = C] - \mathbb{E}[Y(s_0) | \text{Group} = C])$$



Back to Difference in Difference

We want the ATT: $\mathbb{E}[Y_1 - Y_0|\text{Group} = T]$

i.e. amongst the treated, how does giving the treatment compare with not giving the treatment (causally)

BUT: $\mathbb{E}[Y_1|\text{Group} = T]$ is actually measured: $\mathbb{E}[Y(s_1)|\text{Group} = T]$
while $\mathbb{E}[Y_0|\text{Group} = T]$ is not (**counterfactual**),

Treatment group: Have treatment received (T=1) at time s_1

Because in Group T everyone receives the treatment at s_1

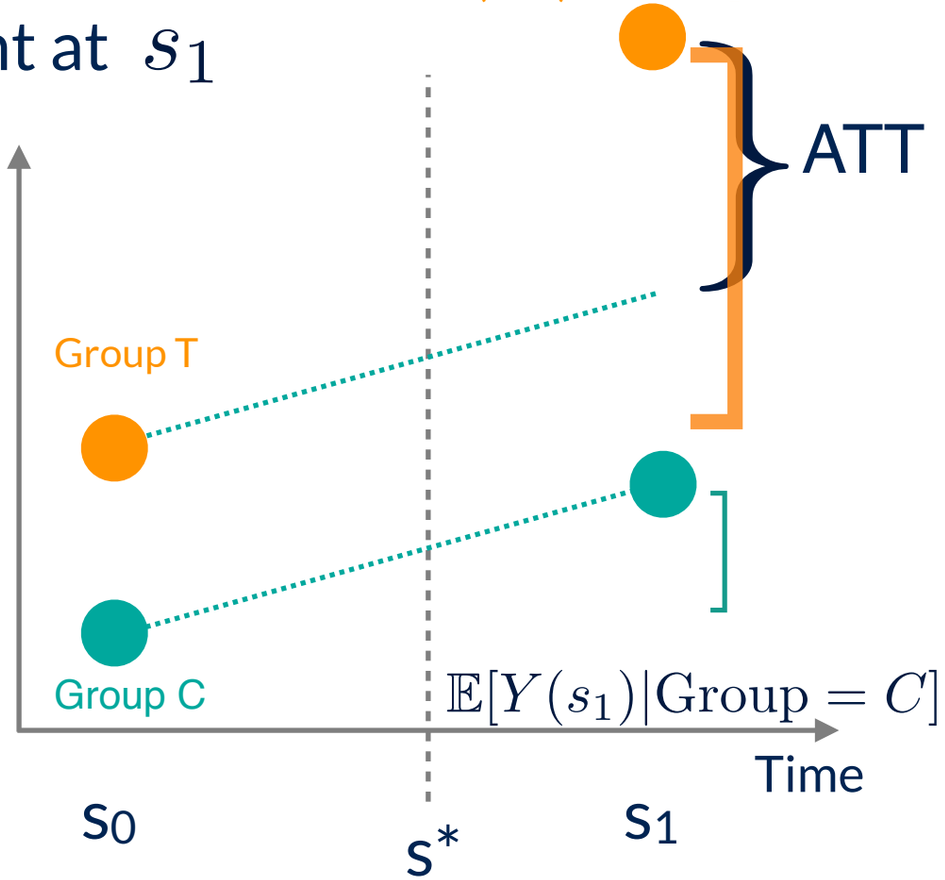
$$\mathbb{E}[Y(s_0)|\text{Group} = T]$$

Treatment group: Have no treatment received (T=0) at time s_0

$$\mathbb{E}[Y(s_0)|\text{Group} = C]$$

Control group: Have no treatment received (T=0) at time s_0

$$\mathbb{E}[Y_1 - Y_0|\text{Group} = T] = (\mathbb{E}[Y(s_1)|\text{Group} = T] - \mathbb{E}[Y(s_0)|\text{Group} = T]) - (\mathbb{E}[Y(s_1)|\text{Group} = C] - \mathbb{E}[Y(s_0)|\text{Group} = C])$$



Back to Difference in Difference

We want the ATT: $\mathbb{E}[Y_1 - Y_0|\text{Group} = T]$

i.e. amongst the treated, how does giving the treatment compare with not giving the treatment (causally)

BUT: $\mathbb{E}[Y_1|\text{Group} = T]$ is actually measured: $\mathbb{E}[Y(s_1)|\text{Group} = T]$
while $\mathbb{E}[Y_0|\text{Group} = T]$ is not (**counterfactual**),

Treatment group: Have treatment received (T=1) at time s_1

Because in Group T everyone receives the treatment at s_1

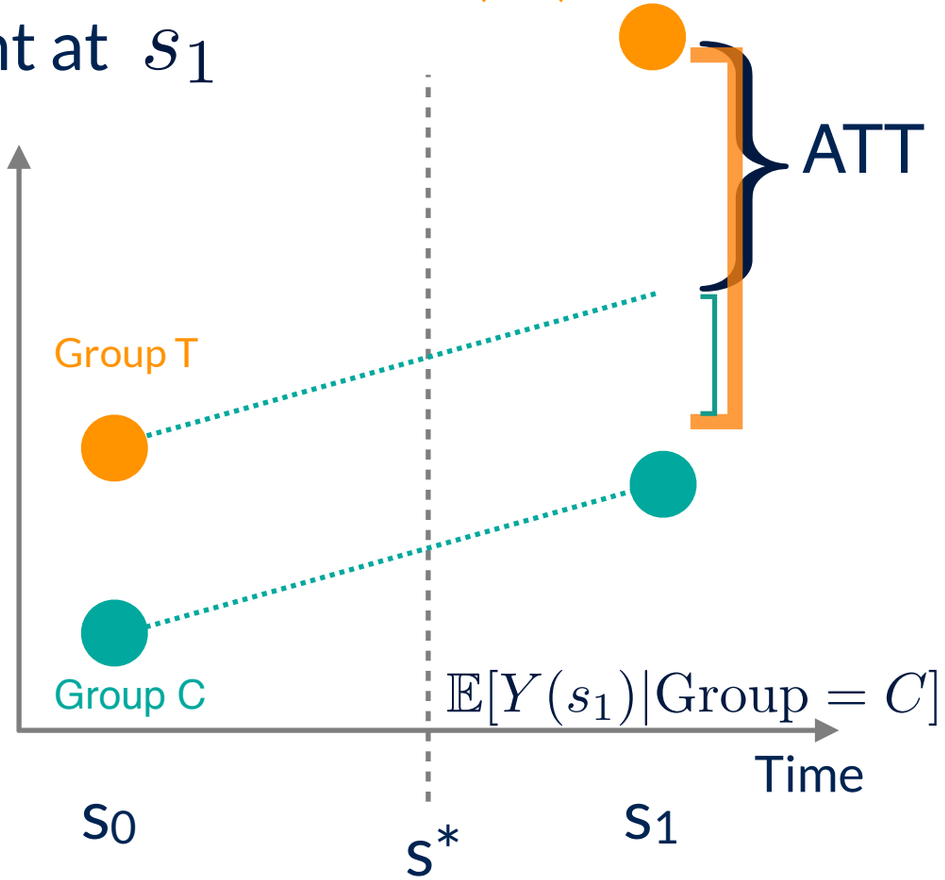
$$\mathbb{E}[Y(s_0)|\text{Group} = T]$$

Treatment group: Have no treatment received (T=0) at time s_0

$$\mathbb{E}[Y(s_0)|\text{Group} = C]$$

Control group: Have no treatment received (T=0) at time s_0

$$\mathbb{E}[Y_1 - Y_0|\text{Group} = T] = (\mathbb{E}[Y(s_1)|\text{Group} = T] - \mathbb{E}[Y(s_0)|\text{Group} = T]) - (\mathbb{E}[Y(s_1)|\text{Group} = C] - \mathbb{E}[Y(s_0)|\text{Group} = C])$$



Proof of estimator DiD

Parallel trends assumption: $\mathbb{E}[Y_0(s_1) - Y_0(s_0)|\text{Group} = T] = \mathbb{E}[Y_0(s_1) - Y_0(s_0)|\text{Group} = C]$

i.e., $\Big(Y_0(s_1) - Y_0(s_0)\Big) \perp\!\!\!\perp \text{Group } T \text{ or } C$

$$\mathbb{E}[Y_1 - Y_0|\text{Group} = T] = (\mathbb{E}[Y(s_1)|\text{Group} = T] - \mathbb{E}[Y(s_0)|\text{Group} = T]) - (\mathbb{E}[Y(s_1)|\text{Group} = C] - \mathbb{E}[Y(s_0)|\text{Group} = C])$$

Proof of estimator DiD

Parallel trends assumption: $\mathbb{E}[Y_0(s_1) - Y_0(s_0)|\text{Group} = T] = \mathbb{E}[Y_0(s_1) - Y_0(s_0)|\text{Group} = C]$

i.e., $\Big(Y_0(s_1) - Y_0(s_0)\Big) \perp\!\!\!\perp \text{Group } T \text{ or } C$

$$\Rightarrow \mathbb{E}[Y_0(s_1)|\text{Group} = T] = \mathbb{E}[Y_0(s_0)|\text{Group} = T] + \mathbb{E}[Y_0(s_1) - Y_0(s_0)|\text{Group} = C]$$

And no pre-treatment effect (participants knowing they have been assigned to the treatment group do not change the behaviour based on this knowledge).

$$\Rightarrow \mathbb{E}[Y_0(s_1)|\text{Group} = T] = \mathbb{E}[Y(s_0)|\text{Group} = T] + \mathbb{E}[Y(s_1)|\text{Group} = C] - \mathbb{E}[Y(s_0)|\text{Group} = C]$$

Therefore:

$$\begin{aligned}\mathbb{E}[Y_1 - Y_0|\text{Group} = T] &= \mathbb{E}[Y_1(s_1) - Y_0(s_1)|\text{Group} = T] \\ &= \mathbb{E}[Y_1(s_1)|\text{Group} = T] - \mathbb{E}[Y_0(s_1)|\text{Group} = T] \\ &= (\mathbb{E}[Y(s_1)|\text{Group} = T] - \mathbb{E}[Y(s_0)|\text{Group} = T]) \\ &\quad - (\mathbb{E}[Y(s_1)|\text{Group} = C] - \mathbb{E}[Y(s_0)|\text{Group} = C])\end{aligned}$$

$$\begin{aligned}\mathbb{E}[Y_1 - Y_0|\text{Group} = T] = &(\mathbb{E}[Y(s_1)|\text{Group} = T] - \mathbb{E}[Y(s_0)|\text{Group} = T]) \\ - &(\mathbb{E}[Y(s_1)|\text{Group} = C] - \mathbb{E}[Y(s_0)|\text{Group} = C])\end{aligned}$$

Regression Discontinuity Design

Recall the **positivity** assumption:

$$0 < p(T = 1|X = x) < 1 \qquad \text{when } p(X = x) > 0$$

It is essential to apply propensity score matching or regression adjustment in order to identify the causal effect of treatment T on outcome Y.

Under certain assumptions on the data (or the **design** generating it), we can still identify this causal effect when positivity is *completely violated*.

Regression Discontinuity Design

Recall the **positivity** assumption:

$$0 < p(T = 1|X = x) < 1 \qquad \text{when } p(X = x) > 0$$

It is essential to apply propensity score matching or regression adjustment in order to identify the causal effect of treatment T on outcome Y.

Under certain assumptions on the data (or the **design** generating it), we can still identify this causal effect when positivity is *completely violated*.

Data structure:

- Outcome Y
- Binary treatment T
- Special confounder W
- (Extra confounders X possible but here omitted, same argument)

Regression Discontinuity Design

Example

Students with average mark ≥ 80 in undergraduate degree, receive a scholarship for postgrad degree. Causal effect of scholarship on performance?

- Higher mark typically higher performance anyway, so upward bias
- **Positivity violation:** All such students are "treated", so

$$p(T = 1|\text{mark} \geq 80) = 1$$

Regression Discontinuity Design

Example

Students with average mark ≥ 80 in undergraduate degree, receive a scholarship for postgrad degree. Causal effect of scholarship on performance?

- Higher mark typically higher performance anyway, so upward bias
- **Positivity violation:** All such students are "treated", so

$$p(T = 1 | \text{mark} \geq 80) = 1$$

W = mark

Regression Discontinuity Design

Example

Students with average mark ≥ 80 in undergraduate degree, receive a scholarship for postgrad degree. Causal effect of scholarship on performance?

- Higher mark typically higher performance anyway, so upward bias
- **Positivity violation:** All such students are "treated", so

$$p(T = 1|\text{mark} \geq 80) = 1$$

W = mark

Example

Admission to a museum is often cheaper for children (e.g., under 12 years) or for pensioners (e.g., above 65 years). What is the causal effect of this lower price on admission numbers?

- Children and pensioners tend to go to a museum more often: upward bias
- Positivity violation: All children and pensioners are "treated", so

$$p(T = 1|\text{Age} < 12 \text{ or } \text{Age} > 65) = 1$$

Regression Discontinuity Design

Example

Students with average mark ≥ 80 in undergraduate degree, receive a scholarship for postgrad degree. Causal effect of scholarship on performance?

- Higher mark typically higher performance anyway, so upward bias
- **Positivity violation:** All such students are "treated", so

$$p(T = 1 | \text{mark} \geq 80) = 1$$

W = mark

Example

Admission to a museum is often cheaper for children (e.g., under 12 years) or for pensioners (e.g., above 65 years). What is the causal effect of this lower price on admission numbers?

- Children and pensioners tend to go to a museum more often: upward bias
- **P**ositivity violation: All children and pensioners are "treated", so

$$p(T = 1 | \text{Age} < 12 \text{ or } \text{Age} > 65) = 1$$

W = Age

Sharp Regression Discontinuity (SRD)

How can we identify the causal effect here despite positivity violations?

Sharp Regression Discontinuity (SRD)

How can we identify the causal effect here despite positivity violations?

Sharp Regression Discontinuity (SRD) design, where Wdetermines T:

$$T(W) = \mathbb{I}\{W \geq c\} = \begin{cases} 1 & \text{if } W \geq c \\ 0 & \text{if } W < c \end{cases}$$

Here c is the cut-off value (e.g., Mark ≥ 80).

Sharp Regression Discontinuity (SRD)

How can we identify the causal effect here despite positivity violations?

Sharp Regression Discontinuity (SRD) design, where Wdetermines T:

$$T(W) = \mathbb{I}\{W \geq c\} = \begin{cases} 1 & \text{if } W \geq c \\ 0 & \text{if } W < c \end{cases}$$

Here c is the cut-off value (e.g., Mark ≥ 80).

Aim: Estimate the causal effect $\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$

Example

Mark in undergraduate is (at best) a fuzzy representation of scholarship merit (e.g., bad day): Students around cut-off c = 80 (e.g., 75-85) are comparable!

Sharp Regression Discontinuity (SRD)

How can we identify the causal effect here despite positivity violations?

Sharp Regression Discontinuity (SRD) design, where Wdetermines T:

$$T(W) = \mathbb{I}\{W \geq c\} = \begin{cases} 1 & \text{if } W \geq c \\ 0 & \text{if } W < c \end{cases}$$

Here c is the cut-off value (e.g., Mark ≥ 80).

Aim: Estimate the causal effect $\tau_{\text{SRD}} = \mathbb{E}\left[Y_1 - Y_0 \mid W = c\right]$

Example

Mark in undergraduate is (at best) a fuzzy representation of scholarship merit (e.g., bad day): Students around cut-off c = 80 (e.g., 75-85) are comparable!

Real goal is to give scholarship to strong/talented/hard-working students

Sharp Regression Discontinuity (SRD)

How can we identify the causal effect here despite positivity violations?

Sharp Regression Discontinuity (SRD) design, where Wdetermines T:

$$T(W) = \mathbb{I}\{W \geq c\} = \begin{cases} 1 & \text{if } W \geq c \\ 0 & \text{if } W < c \end{cases}$$

Here c is the cut-off value (e.g., Mark ≥ 80).

Aim: Estimate the causal effect $\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$

Example

Mark in undergraduate is (at best) a fuzzy representation of scholarship merit (e.g., bad day): Students around cut-off c = 80 (e.g., 75-85) are comparable!

Idea: SRD looks at the **discontinuity** in outcome at the cut-off, i.e., the outcome *just above* the cut-off minus the outcome *just below* the cut-off

Sharp Regression Discontinuity (SRD)

<u>Idea:</u> SRD looks at the **discontinuity** in outcome at the cut-off, i.e., the outcome *just above* the cut-off minus the outcome *just below* the cut-off

$$\hat{\tau}_{\text{SRD}} = \lim_{w \downarrow c} \mathbb{E}[Y \mid W = w] - \lim_{w \uparrow c} \mathbb{E}[Y \mid W = w]$$

Imbens-Lemieux (2007)

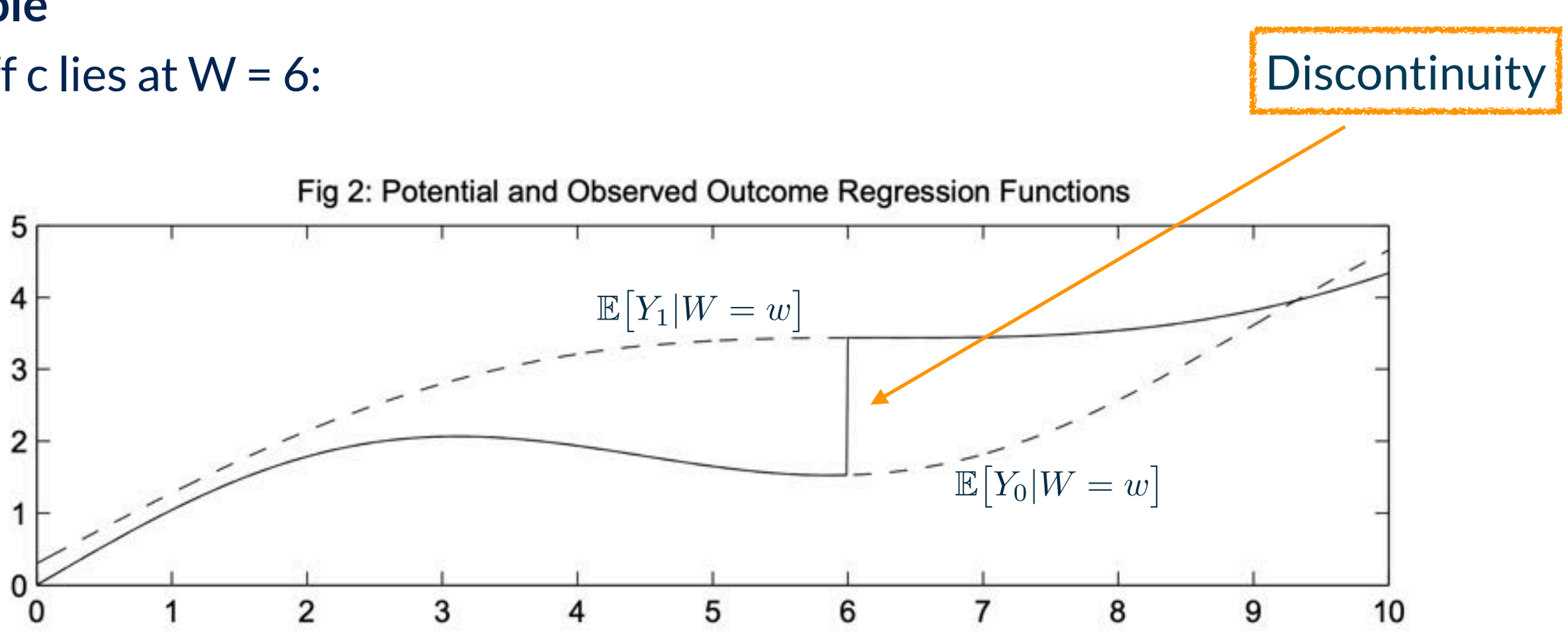
Sharp Regression Discontinuity (SRD)

<u>Idea:</u> SRD looks at the **discontinuity** in outcome at the cut-off, i.e., the outcome *just above* the cut-off minus the outcome *just below* the cut-off

$$\hat{\tau}_{\text{SRD}} = \lim_{w \downarrow c} \mathbb{E}[Y \mid W = w] - \lim_{w \uparrow c} \mathbb{E}[Y \mid W = w]$$

Example

Cut-off c lies at W = 6:



Imbens-Lemieux (2007)

Sharp Regression Discontinuity (SRD)

<u>Idea:</u> SRD looks at the **discontinuity** in outcome at the cut-off, i.e., the outcome *just above* the cut-off minus the outcome *just below* the cut-off

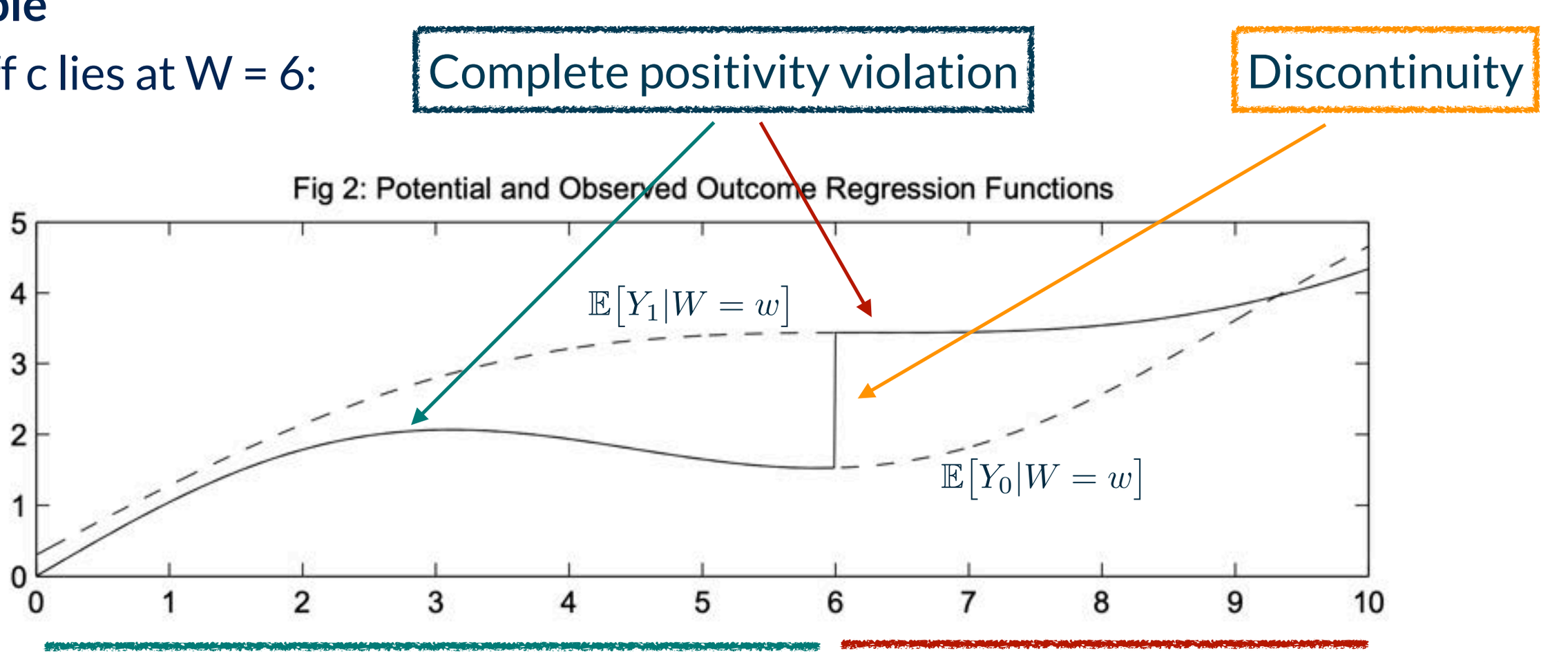
$$\hat{\tau}_{\text{SRD}} = \lim_{w \downarrow c} \mathbb{E}[Y \mid W = w] - \lim_{w \uparrow c} \mathbb{E}[Y \mid W = w]$$

Example

Cut-off c lies at W = 6:

Complete positivity violation

Discontinuity



Imbens-Lemieux (2007)

SRD: Identifiability

Aim: Assumptions to identify the causal effect $\tau_{\mathrm{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$?

SRD: Identifiability

Aim: Assumptions to identify the causal effect $\tau_{\mathrm{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$?

Exchangeability / unconfoundedness holds because W determines T

$$Y_0, Y_1 \perp\!\!\!\perp T \mid W$$

However, not directly useful due to complete positivity violation!

SRD: Identifiability

Aim: Assumptions to identify the causal effect $\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$?

Exchangeability / unconfoundedness holds because W determines T

$$Y_0, Y_1 \perp\!\!\!\perp T \mid W$$

However, not directly useful due to complete positivity violation!

Similarly, the second term here has no support in the data:

$$\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big] = \mathbb{E}\big[Y_1 \mid W = c\big] - \mathbb{E}\big[Y_0 \mid W = c\big]$$

SRD: Identifiability

Aim: Assumptions to identify the causal effect $\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$?

Exchangeability / unconfoundedness holds because W determines T

$$Y_0, Y_1 \perp\!\!\!\perp T \mid W$$

However, not directly useful due to complete positivity violation!

Similarly, the second term here has no support in the data:

$$\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big] = \mathbb{E}\big[Y_1 \mid W = c\big] - \mathbb{E}\big[Y_0 \mid W = c\big]$$

Key assumption:

The functions $\mu_t(w) = \mathbb{E}\big[Y_t \mid W = w\big]$ are **continuous** (at least at w = c)

Consequence:

Any jump / discontinuity at $W = c$ is *only* due to the treatment

SRD: Identifiability

Aim: Assumptions to identify the causal effect $\tau_{\mathrm{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$?

Key assumption:
The functions $\mu_t(w) = \mathbb{E}\big[Y_t \mid W = w\big]$ are **continuous** (at least at w = c)

Now, we can identify the causal effect of T on Y at $W = c$:

SRD: Identifiability

Aim: Assumptions to identify the causal effect $\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$?

Key assumption:
The functions $\mu_t(w) = \mathbb{E}\big[Y_t \mid W = w\big]$ are **continuous** (at least at w = c)

Now, we can identify the causal effect of T on Y at $W = c$:

$$\begin{aligned}\tau_{\text{SRD}} &= \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big] \\ &= \mathbb{E}\big[Y_1 \mid W = c\big] - \mathbb{E}\big[Y_0 \mid W = c\big] \qquad \text{(Expectation linear)}\end{aligned}$$

SRD: Identifiability

Aim: Assumptions to identify the causal effect $\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$?

Key assumption:
The functions $\mu_t(w) = \mathbb{E}\big[Y_t \mid W = w\big]$ are **continuous** (at least at w = c)

Now, we can identify the causal effect of T on Y at $W = c$:

$$
\begin{aligned}
\tau_{\text{SRD}} &= \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big] \\
&= \mathbb{E}\big[Y_1 \mid W = c\big] - \mathbb{E}\big[Y_0 \mid W = c\big] && \text{(Expectation linear)} \\
&= \lim_{w \downarrow c} \mathbb{E}\big[Y_1 \mid W = w\big] - \lim_{w \uparrow c} \mathbb{E}\big[Y_0 \mid W = w\big] && \text{(Continuity)}
\end{aligned}
$$

SRD: Identifiability

Aim: Assumptions to identify the causal effect $\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$?

Key assumption:
The functions $\mu_t(w) = \mathbb{E}\big[Y_t \mid W = w\big]$ are **continuous** (at least at w = c)

Now, we can identify the causal effect of T on Y at $W = c$:

$$
\begin{aligned}
\tau_{\text{SRD}} &= \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big] \\
&= \mathbb{E}\big[Y_1 \mid W = c\big] - \mathbb{E}\big[Y_0 \mid W = c\big] && \text{(Expectation linear)} \\
&= \lim_{w \downarrow c} \mathbb{E}\big[Y_1 \mid W = w\big] - \lim_{w \uparrow c} \mathbb{E}\big[Y_0 \mid W = w\big] && \text{(Continuity)} \\
&= \lim_{w \downarrow c} \mathbb{E}\big[Y \mid W = w\big] - \lim_{w \uparrow c} \mathbb{E}\big[Y \mid W = w\big]
\end{aligned}
$$

$Y_1 = Y$ if $w > c$ $\qquad$ $Y_0 = Y$ if $w < c$

SRD: Identifiability

Aim: Assumptions to identify the causal effect $\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$?

Key assumption:
The functions $\mu_t(w) = \mathbb{E}\big[Y_t \mid W = w\big]$ are **continuous** (at least at w = c)

Now, we can identify the causal effect of T on Y at $W = c$:

$$
\begin{aligned}
\tau_{\text{SRD}} &= \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big] \\
&= \mathbb{E}\big[Y_1 \mid W = c\big] - \mathbb{E}\big[Y_0 \mid W = c\big] && \text{(Expectation linear)} \\
&= \lim_{w \downarrow c} \mathbb{E}\big[Y_1 \mid W = w\big] - \lim_{w \uparrow c} \mathbb{E}\big[Y_0 \mid W = w\big] && \text{(Continuity)} \\
&= \lim_{w \downarrow c} \mathbb{E}\big[Y \mid W = w\big] - \lim_{w \uparrow c} \mathbb{E}\big[Y \mid W = w\big] = \hat{\tau}_{\text{SRD}}
\end{aligned}
$$

Causal identifiability!

Fuzzy Regression Discontinuity (FRD) [non-exam.]

A similar approach exist to identify the causal effect around a cut-off c

$$\tau_{\text{SRD}} = \mathbb{E}\big[Y_1 - Y_0 \mid W = c\big]$$

when treatment below and above c is not deterministic

This is called **Fuzzy** Regression Discontinuity (FRD)

Example

FRD can incorporate (non-)compliers and never-takers. For example, if grades on an assignment are below cut-off c, extra help is offered to students —> may or may not take the help. What is the causal effect of this help?

Causal identifiability and estimation of FRD requires an approach similar to instrumental variables (see reference)

Imbens-Lemieux (2007)

Overview of the course

- **Lecture 1**: Introduction & Motivation, why do we care about causality? Why deriving causality from observational data is non-trivial.
- **Lecture 2:** Recap of probability theory, variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3**: Recap of regression, multiple regression, graphs, SCM
- **Lecture 4-20:**