Statistical Learning
Difference-in-Differences

Outline

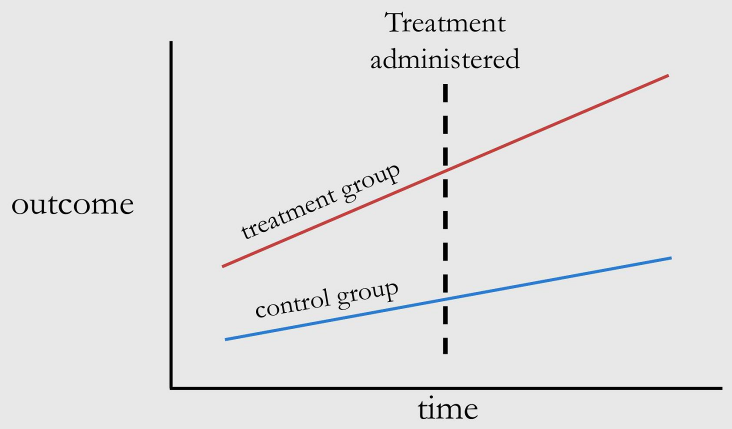
Motivation

Context

Treatment and control group both before and after treatment administered

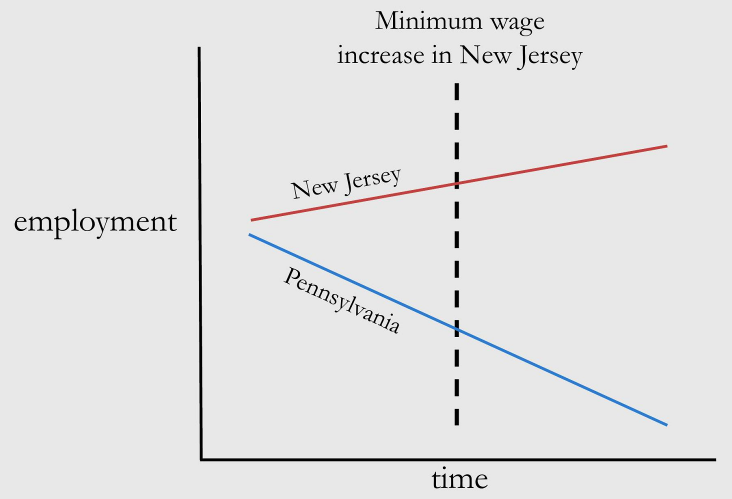
Advantage

Use time dimension to help with identification



Motivating Example

from Card & Krueger (1994)



Average Treatment Effect on the Treated (ATT)

ATE

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0]$$

Unconfoundedness

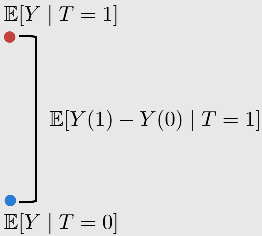
$$(Y(0), Y(1)) \perp\!\!\!\perp T$$

ATT

$$\begin{aligned}
\mathbb{E}[Y(1) - Y(0) \mid T = 1] &= \mathbb{E}[Y(1) \mid T = 1] - \mathbb{E}[Y(0) \mid T = 1] \\
&= \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y(0) \mid T = 1] \\
&= \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y(0) \mid T = 0] \qquad (Y(0) \perp\!\!\!\perp T) \\
&= \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0]
\end{aligned}$$

Outline

Introducing Time

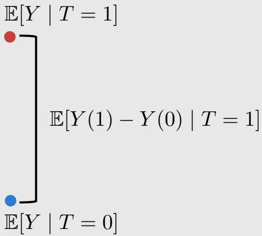


Unconfoundedness:

$$Y(0) \perp\!\!\!\perp T$$

ATT estimand without time: $\mathbb{E}[Y(1) - Y(0) \mid T = 1]$

Introducing Time

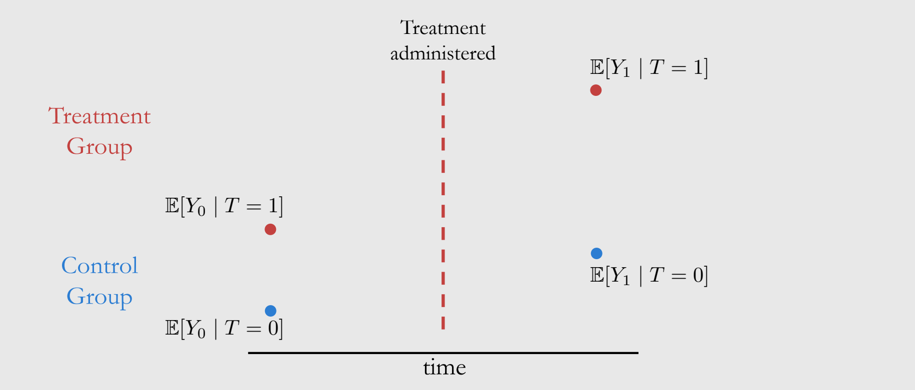


~~Unconfoundedness:~~

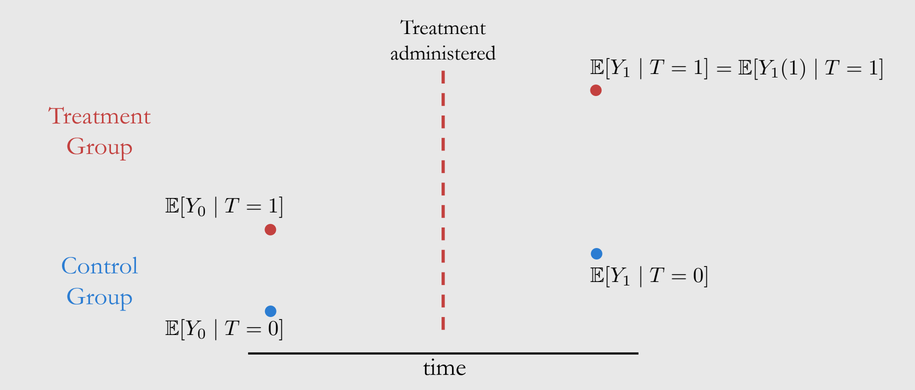
$\cancel{Y(0) \perp T}$

ATT estimand without time: $\mathbb{E}[Y(1) - Y(0) \mid T = 1]$

Introducing Time

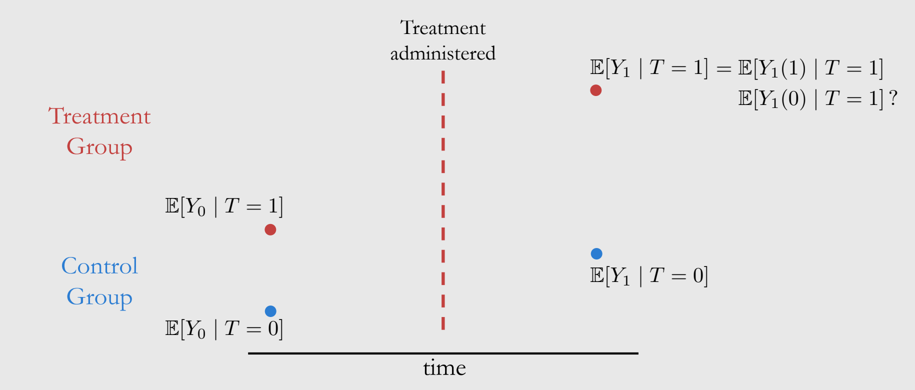


Introducing Time



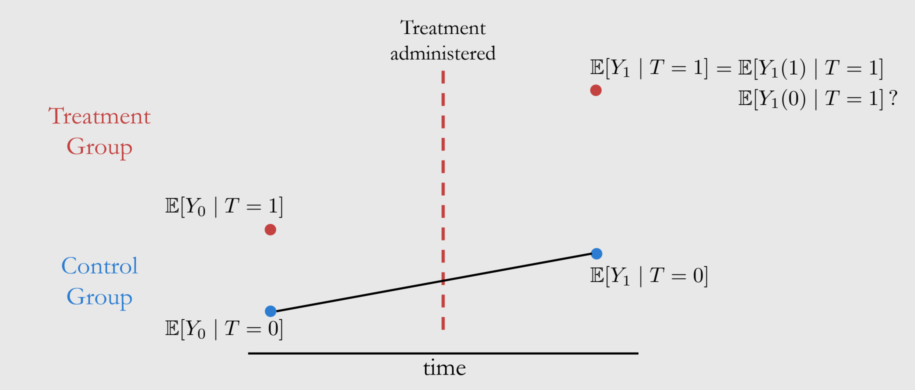
ATT estimand without time: $\mathbb{E}[Y(1) - Y(0) \mid T = 1]$
ATT estimand with time: $\mathbb{E}[\underline{Y_1(1)} - Y_1(0) \mid T = 1]$

Introducing Time



ATT estimand without time: $\mathbb{E}[Y(1) - Y(0) \mid T = 1]$
ATT estimand with time: $\mathbb{E}[Y_1(1) - \underline{Y_1(0)} \mid T = 1]$

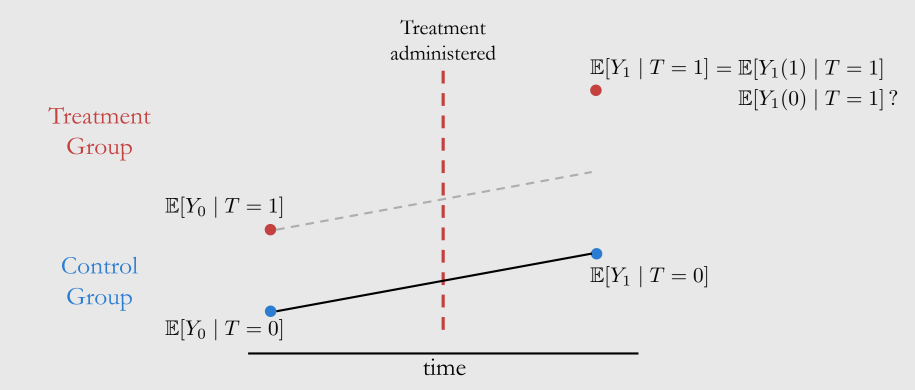
Introducing Time



ATT estimand without time: $\mathbb{E}[Y(1) - Y(0) \mid T = 1]$

ATT estimand with time: $\mathbb{E}[Y_1(1) - \underline{Y_1(0)} \mid T = 1]$

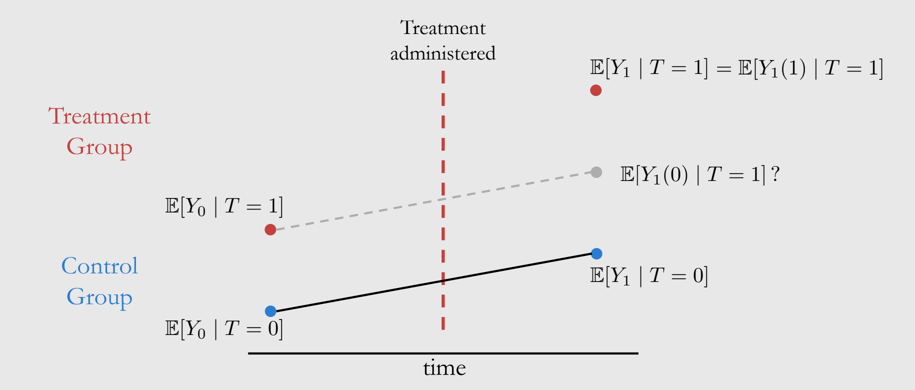
Introducing Time



ATT estimand without time: $\mathbb{E}[Y(1) - Y(0) \mid T = 1]$

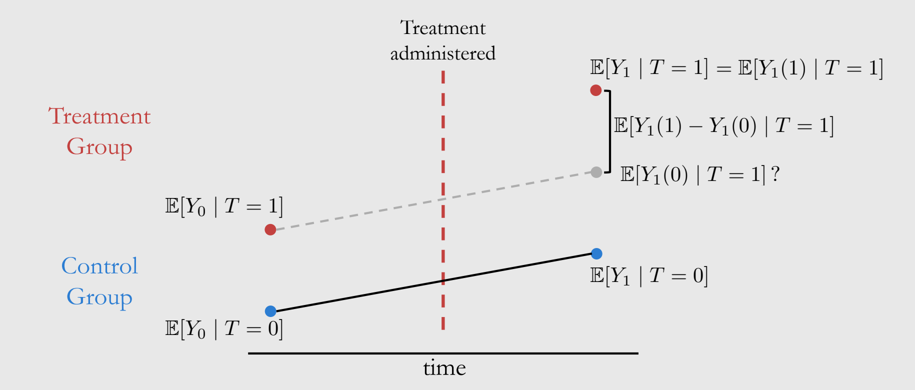
ATT estimand with time: $\mathbb{E}[Y_1(1) - \underline{Y_1(0)} \mid T = 1]$

Introducing Time



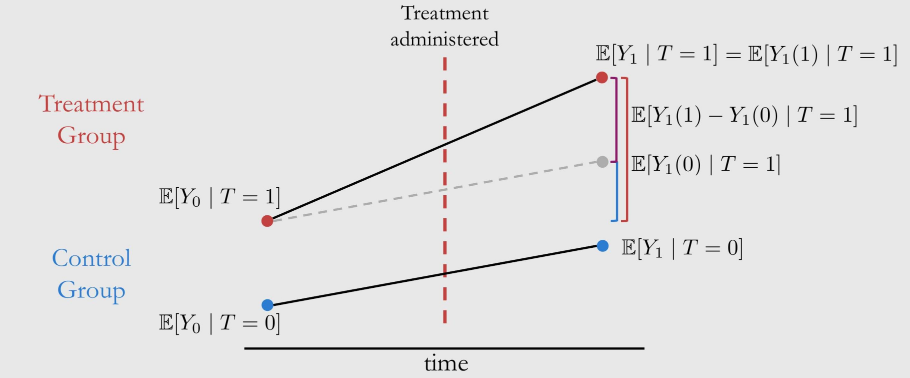
ATT estimand without time: $\mathbb{E}[Y(1) - Y(0) \mid T = 1]$
ATT estimand with time: $\mathbb{E}[Y_1(1) - \underline{Y_1(0)} \mid T = 1]$

Introducing Time



ATT estimand without time: $\mathbb{E}[Y(1) - Y(0) \mid T = 1]$
ATT estimand with time: $\mathbb{E}[Y_1(1) - \underline{Y_1(0)} \mid T = 1]$

Difference-in-Differences



$$\underline{\mathbb{E}[Y_1(1) - Y_1(0) \mid T=1]} = \underline{(\mathbb{E}[Y_1 \mid T=1] - \mathbb{E}[Y_0 \mid T=1])} - \underline{(\mathbb{E}[Y_1 \mid T=0] - \mathbb{E}[Y_0 \mid T=0])}$$

Tolerates Time-Invariant Unobserved Confounding

Unobserved confounders that are constant with time are no problem, since they'll cancel out in the time differences

$$\mathbb{E}[Y_1(1) - Y_1(0) \mid T = 1]$$

$$= \underbrace{(\mathbb{E}[Y_1 \mid T = 1] - \mathbb{E}[Y_0 \mid T = 1])}_{\text{Time difference in treatment group}} - \underbrace{(\mathbb{E}[Y_1 \mid T = 0] - \mathbb{E}[Y_0 \mid T = 0])}_{\text{Time difference in control group}}$$

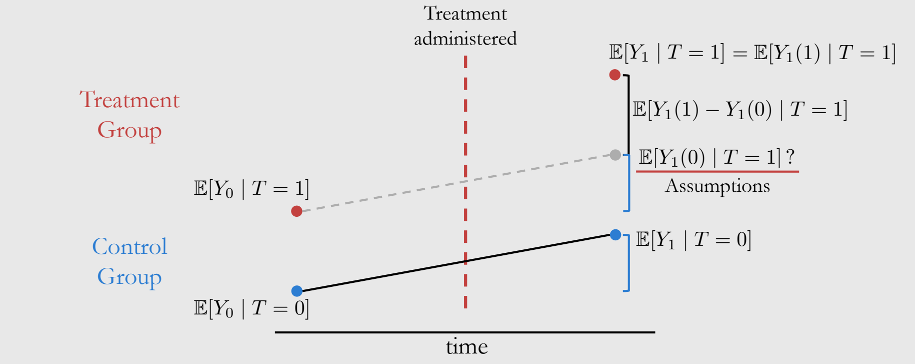
Outline

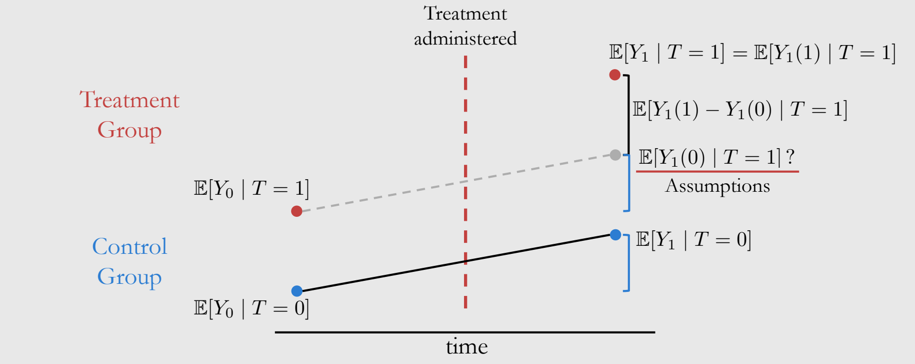
Consistency Assumption Extended to Time

$$\forall \tau, \quad T = t \Longrightarrow Y_\tau = Y_\tau(t)$$

Causal Estimand	*Statistical Estimand*	*Causal Estimand*	*Statistical Estimand*
$\mathbb{E}[Y_\tau(1) \mid T = 1] =$	$\mathbb{E}[Y_\tau \mid T = 1]$	$\mathbb{E}[Y_\tau(0) \mid T = 0] =$	$\mathbb{E}[Y_\tau \mid T = 0]$
$\mathbb{E}[Y_\tau(1) \mid T = 0]$		$\mathbb{E}[Y_\tau(0) \mid T = 1]$	

Parallel Trends Assumption

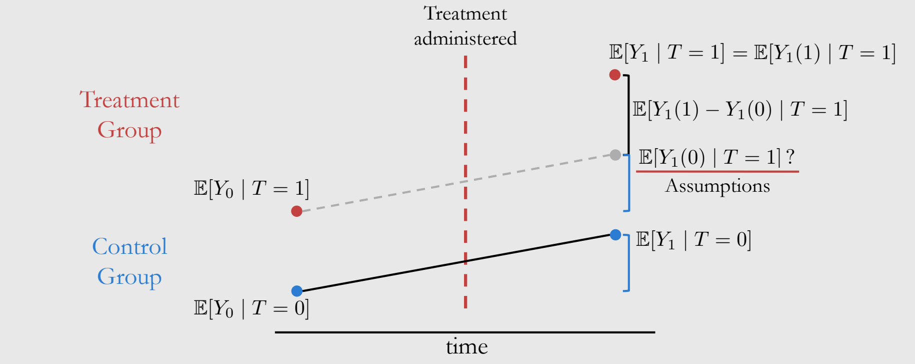




Parallel Trends Assumption

$$\mathbb{E}\left[Y_1(0) - Y_0(0) \mid T = 1\right] = \mathbb{E}\left[Y_1(0) - Y_0(0) \mid T = 0\right]$$

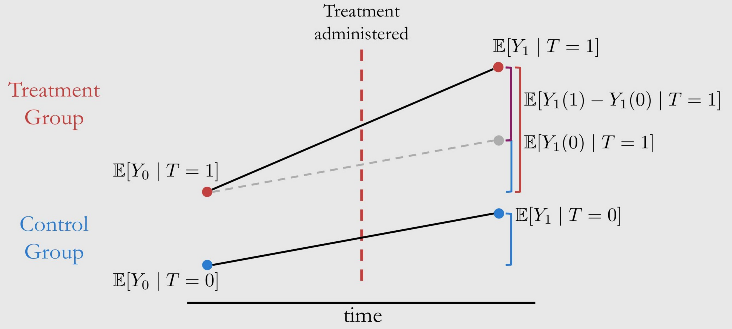
No Pretreatment Effect



No Pretreatment Effect Assumption

$$\mathbb{E}\left[Y_0(1) \mid T = 1\right] - \mathbb{E}\left[Y_0(0) \mid T = 1\right] = 0$$

Difference-in-Differences



Proposition

$$
\begin{aligned}
&\mathbb{E}[Y_1(1) - Y_1(0) \mid T = 1] \\
&= (\mathbb{E}[Y_1 \mid T = 1] - \mathbb{E}[Y_0 \mid T = 1]) - (\mathbb{E}[Y_1 \mid T = 0] - \mathbb{E}[Y_0 \mid T = 0])
\end{aligned}
$$

Proof

$$
\begin{aligned}
\mathbb{E}[Y_1(1) - Y_1(0) \mid T = 1] &= \mathbb{E}[Y_1(1) \mid T = 1] - \mathbb{E}[Y_1(0) \mid T = 1] \\
&= \mathbb{E}[Y_1 \mid T = 1] - \mathbb{E}[Y_1(0) \mid T = 1]
\end{aligned}
$$

Common trends:
$\mathbb{E}[Y_1(0) \mid T = 1] - \mathbb{E}[Y_0(0) \mid T = 1] = \mathbb{E}[Y_1(0) \mid T = 0] - \mathbb{E}[Y_0(0) \mid T = 0]$

$$
\begin{aligned}
\mathbb{E}[Y_1(0) \mid T = 1] &= \mathbb{E}[Y_0(0) \mid T = 1] + \mathbb{E}[Y_1(0) \mid T = 0] - \mathbb{E}[Y_0(0) \mid T = 0] \\
&= \mathbb{E}[Y_0(0) \mid T = 1] + \mathbb{E}[Y_1 \mid T = 0] - \mathbb{E}[Y_0 \mid T = 0] \\
&= \mathbb{E}[Y_0(1) \mid T = 1] + \mathbb{E}[Y_1 \mid T = 0] - \mathbb{E}[Y_0 \mid T = 0] \\
&= \mathbb{E}[Y_0 \mid T = 1] + \mathbb{E}[Y_1 \mid T = 0] - \mathbb{E}[Y_0 \mid T = 0]
\end{aligned}
$$

No pretreatment effect:
$\mathbb{E}[Y_0(1) \mid T = 1] - \mathbb{E}[Y_0(0) \mid T = 1] = 0$

$$
\begin{aligned}
\mathbb{E}[Y_1(1) - Y_1(0) \mid T = 1] &= \mathbb{E}[Y_1 \mid T = 1] - (\mathbb{E}[Y_0 \mid T = 1] + \mathbb{E}[Y_1 \mid T = 0] - \mathbb{E}[Y_0 \mid T = 0]) \\
&= (\mathbb{E}[Y_1 \mid T = 1] - \mathbb{E}[Y_0 \mid T = 1]) - (\mathbb{E}[Y_1 \mid T = 0] - \mathbb{E}[Y_0 \mid T = 0])
\end{aligned}
$$

Outline

Violations of Parallel Trends

Violation:

$$\mathbb{E}\left[Y_1(0) - Y_0(0) \mid T = 1\right] \neq \mathbb{E}\left[Y_1(0) - Y_0(0) \mid T = 0\right]$$

Control for relevant confounders:

$$\mathbb{E}\left[Y_1(0) - Y_0(0) \mid T = 1, W\right] = \mathbb{E}\left[Y_1(0) - Y_0(0) \mid T = 0, W\right]$$

Violation example: whenever there is an interaction term between treatment and time in the following model:

$$Y_\tau(0) = \ldots + T\tau \quad \Longrightarrow \quad \text{Parallel trends violation}$$

Parallel Trends is Scale-Specific

$$\mathbb{E}\left[Y_1(0) \mid T = 1\right] - \mathbb{E}\left[Y_0(0) \mid T = 1\right] = \mathbb{E}\left[Y_1(0) \mid T = 0\right] - \mathbb{E}\left[Y_0(0) \mid T = 0\right]$$

Does not imply (and is not implied by)

$$\mathbb{E}\left[\log Y_1(0) \mid T = 1\right] - \mathbb{E}\left[\log Y_0(0) \mid T = 1\right] = \mathbb{E}\left[\log Y_1(0) \mid T = 0\right] - \mathbb{E}\left[\log Y_0(0) \mid T = 0\right]$$

Outline

DDD

- Suppose there are two states, A and B.
- State B introduces **a new health care policy** targeting those aged 65 or older, but does not apply to other age groups.
- We would like to examine the impact of this policy on health status.
- The health status of different age groups (the experimental and control groups) change over time in inconsistent trends.

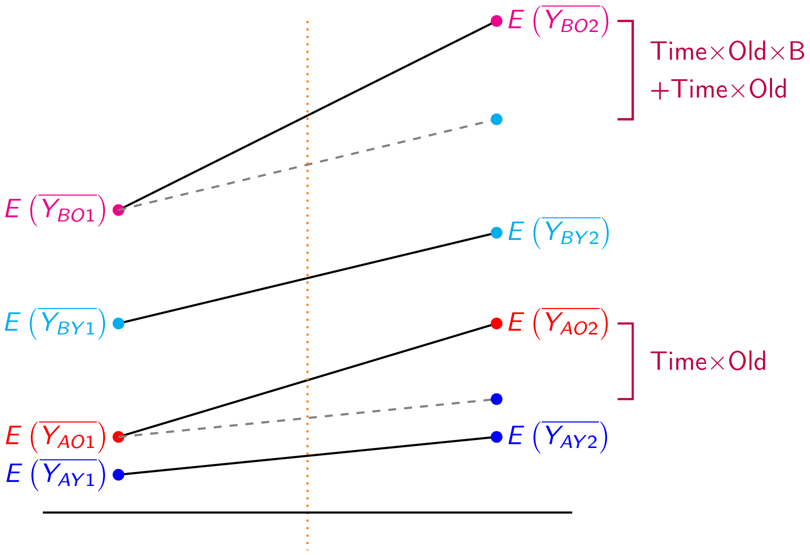
  $\Rightarrow$ violations of the parallel trend assumption
- This inconsistent trend can be captured by calculating the difference between older and younger age groups in neighboring State A (equivalent to using the DID again).

DDD

- Each individual is identified by two indices: i, j, with two different outcomes at different time t
- $B_i = \mathbb{1}$ (the person is in State B)
- $Old_j = \mathbb{1}$ (Age is above 65)
- $Time_t = \mathbb{1}$ (Time = 2)
- The outcome of individual (i, j) at time t is denoted as Y_{ijt}

$$
\begin{aligned}
Y_{ijt} = & \beta_0 + \beta_1 B_i \times \text{ old }_j \times \text{ time }_t \\
& + \beta_2 B_i \times \text{ old }_j + \beta_3 B_i \times \text{ time }_t + \beta_4 \text{ old }_j \times \text{ time }_t \\
& + \gamma_1 \times B_i + \gamma_2 \times \text{ old }_j + \gamma_3 \times \text{ time }_t + \varepsilon_{ijt}
\end{aligned}
$$

DDD



$E\left(\overline{Y_{BO2}}\right)$
Time×Old×B
+Time×Old
$E\left(\overline{Y_{BO1}}\right)$
$E\left(\overline{Y_{BY2}}\right)$
$E\left(\overline{Y_{BY1}}\right)$
$E\left(\overline{Y_{AO2}}\right)$
Time×Old
$E\left(\overline{Y_{AO1}}\right)$
$E\left(\overline{Y_{AY2}}\right)$
$E\left(\overline{Y_{AY1}}\right)$

DDD

$$
\begin{aligned}
E\left(\overline{Y_{BO2}}\right) &= \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \gamma_1 + \gamma_2 + \gamma_3 \\
E\left(\overline{Y_{BO1}}\right) &= \beta_0 + \beta_2 + \gamma_1 + \gamma_2 \\
E\left(\overline{Y_{BY2}}\right) &= \beta_0 + \beta_3 + \gamma_1 + \gamma_3 \\
E\left(\overline{Y_{BY1}}\right) &= \beta_0 + \gamma_1 \\
E\left(\overline{Y_{AO2}}\right) &= \beta_0 + \beta_4 + \gamma_2 + \gamma_3 \\
E\left(\overline{Y_{AO1}}\right) &= \beta_0 + \gamma_2 \\
E\left(\overline{Y_{AY2}}\right) &= \beta_0 + \gamma_3 \\
E\left(\overline{Y_{AY1}}\right) &= \beta_0
\end{aligned}
$$

To obtain β_1, consider the difference of the following two

$$
\left[E\left(\overline{Y_{BO2}}\right) - E\left(\overline{Y_{BO1}}\right)\right] - \left[E\left(\overline{Y_{BY2}}\right) - E\left(\overline{Y_{BY1}}\right)\right] = \beta_1 + \beta_4
$$
$$
\left[E\left(\overline{Y_{AO2}}\right) - E\left(\overline{Y_{AO1}}\right)\right] - \left[E\left(\overline{Y_{AY2}}\right) - E\left(\overline{Y_{AY1}}\right)\right] = \beta_4
$$