Staggered-Adoption Difference-in-Differences

ECON 2400
Applied Econometrics II

Andres Mena

Spring 2024

BROWN

Table of Contents

Content

Staggered adoption

	i=Z	i=A	i=B	i=C	i=D
t=1					
t=2					
t=3					
t=4					
t=5					
t=6					

- Binary treatment D_{it}, for i units and t periods.

Staggered adoption

	i=Z	i=A	i=B	i=C	i=D
t=1					
t=2					
t=3					
t=4					
t=5					
t=6					

- Binary treatment D_{it}, for i units and t periods.
- Event E_i triggers the treatment forever. $D_{it} = 1[t \geq E_i]$.
  - $E_i = min\{t : D_{it} = 1\}$

Staggered adoption

	i=Z	i=A	i=B	i=C	i=D
t=1					
t=2					
t=3					
t=4					
t=5					
t=6					

- Binary treatment D_{it}, for i units and t periods.
- Event E_i triggers the treatment forever. $D_{it} = 1[t \geq E_i]$.
  - $E_i = min\{t : D_{it} = 1\}$
- The treatment is staggered, i.e., the event date E_i varies across units.
  - $E_i = e$, $e \in (1, 2, ..., \infty)$ 'Cohort" indicator
  - $E_Z = 1$ indicate "Always treated" units
  - $E_D = \infty$ indicate "Never treated" units

Setting

- Balanced panel of oberservations $\{Y_{it}, D_{it}\}$, $it \in \Omega$, $|\Omega| = N$.
  - $\Omega_0 = \{it \in \Omega : D_{it} = 0\}$, $|\Omega_0| = N_0$
  - $\Omega_1 = \{it \in \Omega : D_{it} = 1\}$, $|\Omega_1| = N_1$

Setting

- Balanced panel of oberservations $\{Y_{it}, D_{it}\}$, $it \in \Omega$, $|\Omega| = N$.
  - $\Omega_0 = \{it \in \Omega : D_{it} = 0\}$, $|\Omega_0| = N_0$
  - $\Omega_1 = \{it \in \Omega : D_{it} = 1\}$, $|\Omega_1| = N_1$
- Potential Outcomes $(Y_{it}(0), Y_{it}(1))$

Setting

- Balanced panel of oberservations $\{Y_{it}, D_{it}\}$, $it \in \Omega$, $|\Omega| = N$.
  - $\Omega_0 = \{it \in \Omega : D_{it} = 0\}$, $|\Omega_0| = N_0$
  - $\Omega_1 = \{it \in \Omega : D_{it} = 1\}$, $|\Omega_1| = N_1$
- Potential Outcomes $(Y_{it}(0), Y_{it}(1))$
- Observed Outcome $Y_{it} = Y_{it}(D_{it})$

Setting

- Balanced panel of oberservations $\{Y_{it}, D_{it}\}$, $it \in \Omega$, $|\Omega| = N$.
  - $\Omega_0 = \{it \in \Omega : D_{it} = 0\}$, $|\Omega_0| = N_0$
  - $\Omega_1 = \{it \in \Omega : D_{it} = 1\}$, $|\Omega_1| = N_1$
- Potential Outcomes $(Y_{it}(0), Y_{it}(1))$
- Observed Outcome $Y_{it} = Y_{it}(D_{it})$
- Treatment-effect: $\tau_{it} = \mathbb{E}[Y_{it}(1) - Y_{it}(0)]$

Setting

- Balanced panel of oberservations $\{Y_{it}, D_{it}\}$, $it \in \Omega$, $|\Omega| = N$.
  - $\Omega_0 = \{it \in \Omega : D_{it} = 0\}$, $|\Omega_0| = N_0$
  - $\Omega_1 = \{it \in \Omega : D_{it} = 1\}$, $|\Omega_1| = N_1$
- Potential Outcomes $(Y_{it}(0), Y_{it}(1))$
- Observed Outcome $Y_{it} = Y_{it}(D_{it})$
- Treatment-effect: $\tau_{it} = \mathbb{E}[Y_{it}(1) - Y_{it}(0)]$
- Assumptioms (Borusyak et al., 2021):
  - A1: Parallel trends

$$\mathbb{E}[Y_{it}(0)] = \alpha_i + \beta_t \quad \forall it \in \Omega$$

  - A2: No Anticipation

$$Y_{it} = Y_{it}(0) \quad \forall it \in \Omega_0$$

Estimands

Researchers are usually interested in summary statistics in the form of some weighted average: $\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

	i=A	i=B	i=C	i=D
t=1				
t=2				
t=3				
t=4				
t=5				
t=6				

Estimands

Researchers are usually interested in summary statistics in the form of some weighted average: $\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

	i=A	i=B	i=C	i=D
t=1				
t=2				
t=3				
t=4				
t=5				
t=6				

- Overall τ_{ATT}: $\sum_{it \in \Omega_1} \frac{1}{N_1} \tau_{it}$

Estimands

Let h be the number of periods after the event E_i,
$\Omega_{1h} = \{it \in \Omega_1 : t - E_i = h\}$ and $N_{1h} = |\Omega_{1h}|$

	i=A	i=B	i=C	i=D
t=1				
t=2				
t=3	h=0			
t=4	h=1	h=0		
t=5	h=2	h=1	h=0	
t=6	h=3	h=2	h=1	

Estimands

Let h be the number of periods after the event E_i,
$\Omega_{1h} = \{it \in \Omega_1 : t - E_i = h\}$ and $N_{1h} = |\Omega_{1h}|$

	i=A	i=B	i=C	i=D
t=1				
t=2				
t=3	h=0			
t=4	h=1	h=0		
t=5	h=2	h=1	h=0	
t=6	h=3	h=2	h=1	

- Event Study τ_{ATT_h}: $\sum_{it \in \Omega_{1h}} \frac{1}{N_{1h}} \tau_{it}$

Content

Two-Way Fixed Effects Estimation

- Static TWFE Specification

$$Y_{it} = \alpha_i + \beta_t + \tau_{static} D_{it} + \epsilon_{it}$$

Two-Way Fixed Effects Estimation

- Static TWFE Specification

$$Y_{it} = \alpha_i + \beta_t + \tau_{static} D_{it} + \epsilon_{it}$$

- We can reduce the dimensionality of the regression by using a two-step estimator

Two-Way Fixed Effects Estimation

- Static TWFE Specification

$$Y_{it} = \alpha_i + \beta_t + \tau_{static} D_{it} + \epsilon_{it}$$

- We can reduce the dimensionality of the regression by using a two-step estimator
- Estimation:

Two-Way Fixed Effects Estimation

- Static TWFE Specification

$$Y_{it} = \alpha_i + \beta_t + \tau_{static} D_{it} + \epsilon_{it}$$

- We can reduce the dimensionality of the regression by using a two-step estimator
- Estimation:
  - First stage:

$$D_{it} = \gamma + \alpha_i + \beta_t + \varepsilon_{it}$$

Two-Way Fixed Effects Estimation

- Static TWFE Specification

$$Y_{it} = \alpha_i + \beta_t + \tau_{static} D_{it} + \epsilon_{it}$$

- We can reduce the dimensionality of the regression by using a two-step estimator
- Estimation:
  - First stage:

$$D_{it} = \gamma + \alpha_i + \beta_t + \varepsilon_{it}$$

  - Predicted Propensity Score:

$$\hat{D}_{it} = \hat{\gamma} + \hat{\alpha}_i + \hat{\beta}_t = -\bar{D} + \bar{D}_i + \bar{D}_t$$

Two-Way Fixed Effects Estimation

- Static TWFE Specification

$$Y_{it} = \alpha_i + \beta_t + \tau_{static} D_{it} + \epsilon_{it}$$

- We can reduce the dimensionality of the regression by using a two-step estimator
- Estimation:
  - First stage:

$$D_{it} = \gamma + \alpha_i + \beta_t + \varepsilon_{it}$$

  - Predicted Propensity Score:

$$\hat{D}_{it} = \hat{\gamma} + \hat{\alpha}_i + \hat{\beta}_t = -\bar{D} + \bar{D}_i + \bar{D}_t$$

  - Residual:

$$\tilde{D}_{it} = D_{it} - \hat{D}_{it} = D_{it} - \bar{D}_i - \bar{D}_t + \bar{D}$$

Two-Way Fixed Effects Estimation

- Static TWFE Specification

$$Y_{it} = \alpha_i + \beta_t + \tau_{static} D_{it} + \epsilon_{it}$$

- We can reduce the dimensionality of the regression by using a two-step estimator
- Estimation:
  - First stage:

$$D_{it} = \gamma + \alpha_i + \beta_t + \varepsilon_{it}$$

  - Predicted Propensity Score:

$$\hat{D}_{it} = \hat{\gamma} + \hat{\alpha}_i + \hat{\beta}_t = -\bar{D} + \bar{D}_i + \bar{D}_t$$

  - Residual:

$$\tilde{D}_{it} = D_{it} - \hat{D}_{it} = D_{it} - \bar{D}_i - \bar{D}_t + \bar{D}$$

We'll come back to the residual in a moment!

Single Variable OLS Estimator

- Second stage:

$$Y_{it} = \tau_{static}\tilde{D}_{it} + \nu_{it}$$

Single Variable OLS Estimator

- Second stage:

$$Y_{it} = \tau_{static} \tilde{D}_{it} + \nu_{it}$$

- Estimator:

$$\hat{\tau}_{static} = \frac{\hat{C}(Y_{it}, \tilde{D}_{it})}{\hat{V}(\tilde{D}_{it})} = \frac{\sum_{it} Y_{it} \tilde{D}_{it}}{\sum_{it} \tilde{D}_{it}^2}$$

Single Variable OLS Estimator

- Second stage:

$$Y_{it} = \tau_{static} \tilde{D}_{it} + \nu_{it}$$

- Estimator:

$$\hat{\tau}_{static} = \frac{\hat{C}(Y_{it}, \tilde{D}_{it})}{\hat{V}(\tilde{D}_{it})} = \frac{\sum_{it} Y_{it} \tilde{D}_{it}}{\sum_{it} \tilde{D}_{it}^2}$$

  - Define $\hat{w}_{it}^{fe} = \frac{\tilde{D}_{it}}{\sum_{it} \tilde{D}_{it}^2}$

$$\hat{\tau}_{static} = \sum_{it} \hat{w}_{it}^{fe} Y_{it}$$

Is $\hat{\tau}_{static}$ an unbiased estimator of τ_{ATT}?

TWFE weighting: The econometric alchemy at play

- Note that $\sum_{it} \hat{w}_{it}^{fe} = 0$ while $\sum_{it} \hat{w}_{it}^{fe} D_{it} = 1$ and $\sum_{it} \hat{w}_{it}^{fe}(1 - D_{it}) = -1$.

Why?

TWFE weighting: The econometric alchemy at play

- Note that $\sum_{it} \hat{w}_{it}^{fe} = 0$ while $\sum_{it} \hat{w}_{it}^{fe} D_{it} = 1$ and $\sum_{it} \hat{w}_{it}^{fe}(1 - D_{it}) = -1$.

  $\hat{\tau}_{static}$ *is a weighted average of the outcomes in the treated group* **minus** *a weighted average of the outcomes in the control group.*

TWFE weighting: The econometric alchemy at play

- Note that $\sum_{it} \hat{w}_{it}^{fe} = 0$ while $\sum_{it} \hat{w}_{it}^{fe} D_{it} = 1$ and $\sum_{it} \hat{w}_{it}^{fe}(1 - D_{it}) = -1$.

  $\hat{\tau}_{static}$ *is a weighted average of the outcomes in the treated group* **minus** *a weighted average of the outcomes in the control group.*

- Potential Outcome model $Y_{it} = Y_{it}(0) + D_{it}\tau_{it}$

TWFE weighting: The econometric alchemy at play

- Note that $\sum_{it} \hat{w}_{it}^{fe} = 0$ while $\sum_{it} \hat{w}_{it}^{fe} D_{it} = 1$ and $\sum_{it} \hat{w}_{it}^{fe}(1 - D_{it}) = -1$.

  $\hat{\tau}_{static}$ *is a weighted average of the outcomes in the treated group* **minus** *a weighted average of the outcomes in the control group.*

- Potential Outcome model $Y_{it} = Y_{it}(0) + D_{it}\tau_{it}$

$$\hat{\tau}_{static} = \sum_{it} \hat{w}_{it}^{fe} [Y_{it}(0) + D_{it}\tau_{it}]$$

TWFE weighting: Unveiling the Econometric Alchemy

- Note that $\sum_{it} \hat{w}^{fe} it = 0$ while $\sum it \hat{w}^{fe} it Dit = 1$ and $\sum_{it} \hat{w}^{fe} it(1 - Dit) = -1$.

  $\hat{\tau}_{static}$ represents a weighted average of outcomes in the treated group **minus** *a weighted average of outcomes in the control group.*

- Utilizing the Potential Outcome Model: $Y_{it} = Y_{it}(0) + D_{it}\tau_{it}$

$$
\begin{aligned}
\hat{\tau}_{static} &= \sum_{it} \hat{w}^{fe}_{it}[Y_{it}(0) + D_{it}\tau_{it}] \\
&= \sum_{it} \hat{w}^{fe}_{it} Y_{it}(0) + \sum_{it} \hat{w}^{fe}_{it} D_{it}\tau_{it}
\end{aligned}
$$

TWFE weighting: The econometric alchemy at play

- Note that $\sum_{it} \hat{w}_{it}^{fe} = 0$ while $\sum_{it} \hat{w}_{it}^{fe} D_{it} = 1$ and $\sum_{it} \hat{w}_{it}^{fe}(1 - D_{it}) = -1$.

  $\hat{\tau}_{static}$ *is a weighted average of the outcomes in the treated group* **minus** *a weighted average of the outcomes in the control group.*

- Potential Outcome model $Y_{it} = Y_{it}(0) + D_{it}\tau_{it}$

$$
\begin{aligned}
\hat{\tau}_{static} &= \sum_{it} \hat{w}_{it}^{fe}[Y_{it}(0) + D_{it}\tau_{it}] \\
&= \sum_{it} \hat{w}_{it}^{fe} Y_{it}(0) + \sum_{it} \hat{w}_{it}^{fe} D_{it}\tau_{it} \\
&= \sum_{it} \hat{w}_{it}^{fe} D_{it}\tau_{it} \qquad \text{(if Parallel trends holds)}
\end{aligned}
$$

TWFE weighting: The econometric alchemy at play

- Note that $\sum_{it} \hat{w}_{it}^{fe} = 0$ while $\sum_{it} \hat{w}_{it}^{fe} D_{it} = 1$ and $\sum_{it} \hat{w}_{it}^{fe}(1 - D_{it}) = -1$.

  $\hat{\tau}_{static}$ is a weighted average of the outcomes in the treated group **minus** *a weighted average of the outcomes in the control group.*

- Potential Outcome model $Y_{it} = Y_{it}(0) + D_{it}\tau_{it}$

$$
\begin{aligned}
\hat{\tau}_{static} &= \sum_{it} \hat{w}_{it}^{fe} Y_{it}(0) + D_{it}\tau_{it} \\
&= \sum_{it} \hat{w}_{it}^{fe} Y_{it}(0) + \sum_{it} \hat{w}_{it}^{fe} D_{it}\tau_{it} \\
&= \sum_{it} \hat{w}_{it}^{fe} D_{it}\tau_{it} \qquad \text{(if Parallel trends holds)}
\end{aligned}
$$

$$
\mathbb{E}\left[\hat{\tau}_{static}\right] = \mathbb{E}\left[\sum_{it\in\Omega_1} \hat{w}_{it}^{fe}\tau_{it}\right] \neq \sum_{it\in\Omega_1} \frac{1}{N_1}\tau_{it} = \tau_{ATT} \quad (\text{if } \tau_{it} \neq \tau)
$$

Content

Negative weights mechanics

- Under parallel trends, the TWFE estimator $\hat{\tau}_{static}$ is a weighted average of treatment effects.

Negative weights mechanics

- Under parallel trends, the TWFE estimator $\hat{\tau}_{static}$ is a weighted average of treatment effects.
- The weights *add up to 1* among the treated group

$$\sum_{it \in \Omega_1} \hat{w}_{it}^{fe} = 1$$

Negative weights mechanics

- Under parallel trends, the TWFE estimator $\hat{\tau}_{static}$ is a weighted average of treatment effects.
- The weights *add up to 1* among the treated group

$$\sum_{it \in \Omega_1} \hat{w}_{it}^{fe} = 1$$

- But some weights *may be* negative.

$$\hat{w}_{it}^{fe} = \frac{\tilde{D}_{it}}{\hat{V}(\tilde{D}_{it})} < 0$$

Negative weights mechanics

- Under parallel trends, the TWFE estimator $\hat{\tau}_{static}$ is a weighted average of treatment effects.
- The weights *add up to 1* among the treated group

$$\sum_{it \in \Omega_1} \hat{w}_{it}^{fe} = 1$$

- But some weights *may be* negative.

$$\hat{w}_{it}^{fe} = \frac{\tilde{D}_{it}}{\hat{V}(\tilde{D}_{it})} < 0$$

- The predicted propensity score is estimated using a linear probability model that *may be* out of the unit interval

$$\hat{D}_{it} = \bar{D}_i + \bar{D}_t - \bar{D} > 1$$

Negative weights mechanics

- Under parallel trends, the TWFE estimator $\hat{\tau}_{static}$ is a weighted average of treatment effects.
- The weights *add up to 1* among the treated group

$$\sum_{it \in \Omega_1} \hat{w}_{it}^{fe} = 1$$

- But some weights *may be* negative.

$$\hat{w}_{it}^{fe} = \frac{\tilde{D}_{it}}{\hat{V}(\tilde{D}_{it})} < 0$$

- The predicted propensity score is estimated using a linear probability model that *may be* out of the unit interval

$$\hat{D}_{it} = \bar{D}_i + \bar{D}_t - \bar{D} > 1$$

- The residual *may be* negative

$$\tilde{D}_{it} = D_{it} - \hat{D}_{it} < 0 \quad D_{it} \in \{0, 1\}$$

Toy Example: Later periods may be an issue

- Suppose unit A is treated at a very early date, then $\bar{D}_A \approx 1$

Toy Example: Later periods may be an issue

- Suppose unit A is treated at a very early date, then $\bar{D}_A \approx 1$
- Suppose at the last period T, most of the units have been treated, then $\bar{D}_T \approx 1$

Toy Example: Later periods may be an issue

- Suppose unit A is treated at a very early date, then $\bar{D}_A \approx 1$
- Suppose at the last period T, most of the units have been treated, then $\bar{D}_T \approx 1$
- Suppose treatment adoption so that on average units have been treated nearly half of the time, then $\bar{D} \approx 0.5$

Toy Example: Later periods may be an issue

- Suppose unit A is treated at a very early date, then $\bar{D}_A \approx 1$
- Suppose at the last period T, most of the units have been treated, then $\bar{D}_T \approx 1$
- Suppose treatment adoption so that on average units have been treated nearly half of the time, then $\bar{D} \approx 0.5$

$$\hat{D}_{AT} \approx 1 + 1 - 0.5 \approx 1.5$$

Toy Example: Later periods may be an issue

- Suppose unit A is treated at a very early date, then $\bar{D}_A \approx 1$
- Suppose at the last period T, most of the units have been treated, then $\bar{D}_T \approx 1$
- Suppose treatment adoption so that on average units have been treated nearly half of the time, then $\bar{D} \approx 0.5$

$$\hat{D}_{AT} \approx 1 + 1 - 0.5 \approx 1.5$$

$$\tilde{D}_{AT} \approx 1 - 1.5 = -0.5$$

Toy Example: Later periods may be an issue

- Suppose unit A is treated at a very early date, then $\bar{D}_A \approx 1$
- Suppose at the last period T, most of the units have been treated, then $\bar{D}_T \approx 1$
- Suppose treatment adoption so that on average units have been treated nearly half of the time, then $\bar{D} \approx 0.5$

$$\hat{D}_{AT} \approx 1 + 1 - 0.5 \approx 1.5$$

$$\tilde{D}_{AT} \approx 1 - 1.5 = -0.5$$

$$\hat{w}_{AT}^{fe} < 0$$

Empirical Example

- Data: 15 sub-Saharan African countries between 1981-2015
- Treatment: Elimination of Primary School Tuition Fees

Panel A: Dependent Variable: Gross Enrollment in Primary School

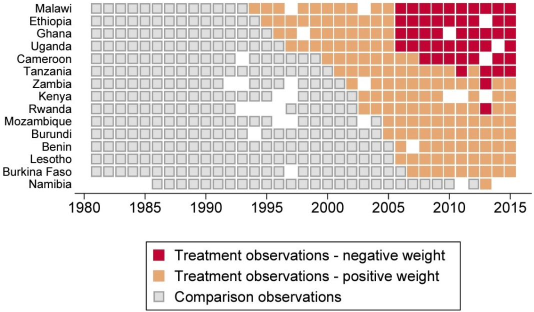
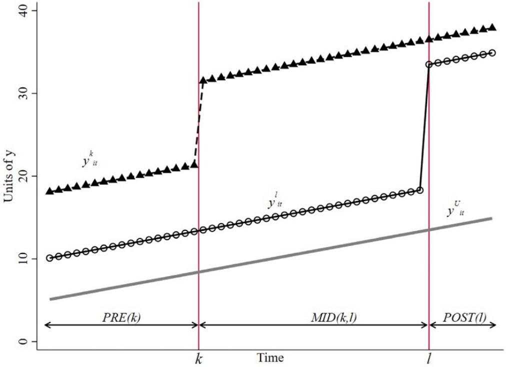


Figure: Jakiela (2021)

Content

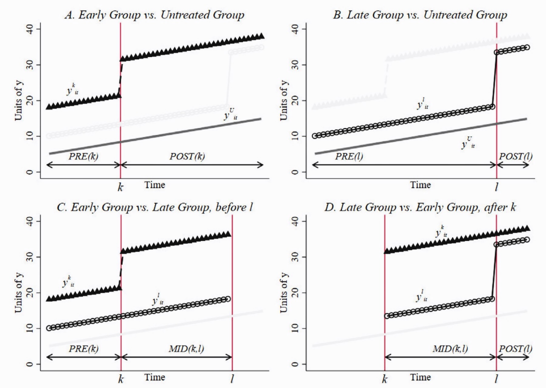
Forbidden comparison

- $\hat{\tau}_{static}$ can be decomposed as the convex weighted average of several two-by-two DiD comparison.(Goodman-Bacon, 2021)
- $g = \{1, 2, ..., k, ..., l, ..., \infty\}$ be the group indicators and $k < l$.



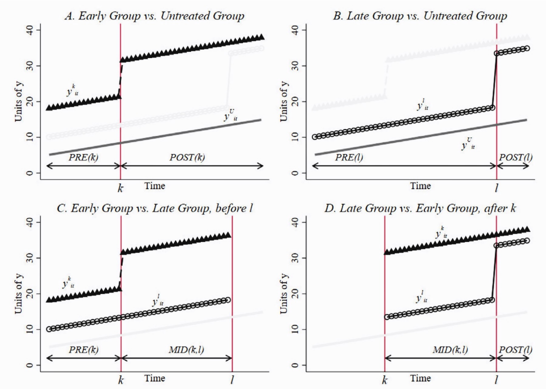
Forbidden comparison

- $DD_k = (\bar{Y}_k^{\geq k} - \bar{Y}_k^{<k}) - (\bar{Y}_\infty^{\geq k} - \bar{Y}_\infty^{<k})$ (good)



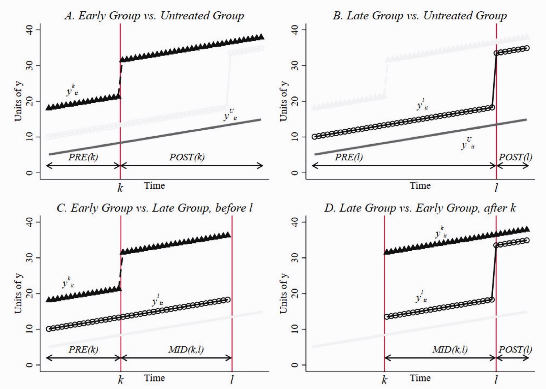
Forbidden comparison

- $DD_k = (\bar{Y}_k^{\geq k} - \bar{Y}_k^{<k}) - (\bar{Y}_\infty^{\geq k} - \bar{Y}_\infty^{<k})$ (good)
- $DD_{kl}^{early} = (\bar{Y}_k^{k \geq t > l} - \bar{Y}_k^{t<k}) - (\bar{Y}_l^{k \geq t > l} - \bar{Y}_l^{<k})$ (good)



Forbidden comparison

- $DD_k = (\bar{Y}_k^{\geq k} - \bar{Y}_k^{<k}) - (\bar{Y}_\infty^{\geq k} - \bar{Y}_\infty^{<k})$ (good)
- $DD_{kl}^{early} = (\bar{Y}_k^{k \geq t > l} - \bar{Y}_k^{t<k}) - (\bar{Y}_l^{k \geq t > l} - \bar{Y}_l^{<k})$ (good)
- $DD_{kl}^{late} = (\bar{Y}_l^{\geq l} - \bar{Y}_l^{k \geq t > l}) - (\bar{Y}_k^{\geq l} - \bar{Y}_k^{k \geq t > l})$ (Forbidden)



2x2 DiD Decomposition

- Goodman-Bacon (2021) show that $\hat{\tau}_{\text{static}}$ can be decomposed as the convex weighted average of all the possible two-by-two DiD comparisons in this format.

2x2 DiD Decomposition

- Goodman-Bacon (2021) show that $\hat{\tau}_{\text{static}}$ can be decomposed as the convex weighted average of all the possible two-by-two DiD comparisons in this format.
- $$\hat{\tau}_{\text{static}} = \sum_{k \neq U} s_k DD_k + \sum_{k \neq U} \sum_{l > k} s_{kl} DD_{kl}^{early} + \sum_{k \neq U} \sum_{l > k} s_{lk} DD_{kl}^{late}$$

2x2 DiD Decomposition

- Goodman-Bacon (2021) show that $\hat{\tau}_{\text{static}}$ can be decomposed as the convex weighted average of all the possible two-by-two DiD comparisons in this format.
- $$\hat{\tau}_{\text{static}} = \sum_{k \neq U} s_k DD_k + \sum_{k \neq U} \sum_{l > k} s_{kl} DD_{kl}^{early} + \sum_{k \neq U} \sum_{l > k} s_{lk} DD_{kl}^{late}$$
- s_k, s_{kl}, and s_{lk} are positive weights that add up to 1:

$$\sum_{k \neq U} s_k + \sum_{k \neq U} \sum_{l > k} (s_{kl} + s_{lk}) = 1$$

Toy Example: 2 Groups and 3 Periods

$$Y_{it} = \alpha_i + \beta_t + \tau_{static} D_{it} + \epsilon_{it}$$

	i=A	i=B
t=1	α_A	α_B
t=2	$\alpha_A + \beta_2 + \tau_{A2}$	$\alpha_B + \beta_2$
t=3	$\alpha_A + \beta_3 + \tau_{A3}$	$\alpha_B + \beta_3 + \tau_{B3}$

Here $\hat{\tau}_{static} = \frac{DD_{AB}^{early} + DD_{AB}^{late}}{2}$

$$DD_{AB}^{early} = (Y_{A2} - Y_{A1}) - (Y_{B2} - Y_{B1}) = \tau_{A2}$$

Toy Example: 2 Groups and 3 Periods

$$Y_{it} = \alpha_i + \beta_t + \tau_{static} D_{it} + \epsilon_{it}$$

	i=A	i=B
t=1	α_A	α_B
t=2	$\alpha_A + \beta_2 + \tau_{A2}$	$\alpha_B + \beta_2$
t=3	$\alpha_A + \beta_3 + \tau_{A3}$	$\alpha_B + \beta_3 + \tau_{B3}$

Here $\hat{\tau}_{static} = \frac{DD_{AB}^{early} + DD_{AB}^{late}}{2}$

$$DD_{AB}^{early} = (Y_{A2} - Y_{A1}) - (Y_{B2} - Y_{B1}) = \tau_{A2}$$
$$DD_{AB}^{late} = (Y_{B3} - Y_{B2}) - (Y_{A3} - Y_{A2}) = \tau_{B3} - (\tau_{A3} - \tau_{A2})$$

Toy Example: 2 Groups and 3 Periods

$$Y_{it} = \alpha_i + \beta_t + \tau_{static} D_{it} + \epsilon_{it}$$

	i=A	i=B
t=1	α_A	α_B
t=2	$\alpha_A + \beta_2 + \tau_{A2}$	$\alpha_B + \beta_2$
t=3	$\alpha_A + \beta_3 + \tau_{A3}$	$\alpha_B + \beta_3 + \tau_{B3}$

Here $\hat{\tau}_{static} = \frac{DD^{early}_{AB} + DD^{late}_{AB}}{2}$

$$DD^{early}_{AB} = (Y_{A2} - Y_{A1}) - (Y_{B2} - Y_{B1}) = \tau_{A2}$$

$$DD^{late}_{AB} = (Y_{B3} - Y_{B2}) - (Y_{A3} - Y_{A2}) = \tau_{B3} - (\tau_{A3} - \tau_{A2})$$

$$\hat{\tau}_{static} = \tau_{A2} - \frac{1}{2}\tau_{A3} + \frac{1}{2}\tau_{B3}$$

Semi-Dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{h \geq 0} \tau_h 1[t = E_i + h] + \epsilon_{it}$$

	i=A	i=B
t=1	α_A	α_B
t=2	$\alpha_A + \beta_2 + \tau_{A2}$	$\alpha_B + \beta_2$
t=3	$\alpha_A + \beta_3 + \tau_{A3}$	$\alpha_B + \beta_3 + \tau_{B3}$

Here $\hat{\tau}_1 = (Y_{A3} - Y_{B3}) + (Y_{A2} - Y_{B2}) - 2(Y_{A1} - Y_{B1})$

- *Without never-treated units long-run effects are not identified and semi-dynamic specification produces spurious estimates* (Borusyak et al., 2021)

Semi-Dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{h \geq 0} \tau_h 1[t = E_i + h] + \epsilon_{it}$$

	i=A	i=B
t=1	α_A	α_B
t=2	$\alpha_A + \beta_2 + \tau_{A2}$	$\alpha_B + \beta_2$
t=3	$\alpha_A + \beta_3 + \tau_{A3}$	$\alpha_B + \beta_3 + \tau_{B3}$

Here $\hat{\tau}_1 = (Y_{A3} - Y_{B3}) + (Y_{A2} - Y_{B2}) - 2(Y_{A1} - Y_{B1})$

- *Without never-treated units long-run effects are not identified and semi-dynamic specification produces spurious estimates* (Borusyak et al., 2021)
-
$$\hat{\tau}_1 = \tau_{A3} + (\tau_{A2} - \tau_{B3})$$

Semi-Dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{h \geq 0} \tau_h 1[t = E_i + h] + \epsilon_{it}$$

	i=A	i=B
t=1	α_A	α_B
t=2	$\alpha_A + \beta_2 + \tau_{A2}$	$\alpha_B + \beta_2$
t=3	$\alpha_A + \beta_3 + \tau_{A3}$	$\alpha_B + \beta_3 + \tau_{B3}$

Here $\hat{\tau}_1 = (Y_{A3} - Y_{B3}) + (Y_{A2} - Y_{B2}) - 2(Y_{A1} - Y_{B1})$

- *Without never-treated units long-run effects are not identified and semi-dynamic specification produces spurious estimates* (Borusyak et al., 2021)
- $$\hat{\tau}_1 = \tau_{A3} + (\tau_{A2} - \tau_{B3})$$

  Extrapolations are invalid with heterogeneous treatment effects

Dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{h \neq -1} \tau_h 1[t = E_i + h] + \epsilon_{it}$$

- Dynamic specification is robust only to heterogeneity across h. Under cohort-specific parallel trend assumptions:

$$\hat{\tau}_h = \tau_{ATT_h} \quad \text{if } \tau_{it} = \tau \quad \forall i, t \in \Omega_{1h}$$

Dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{h \neq -1} \tau_h 1[t = E_i + h] + \epsilon_{it}$$

- Dynamic specification is robust only to heterogeneity across h. Under cohort-specific parallel trend assumptions:

$$\hat{\tau}_h = \tau_{ATT_h} \quad \text{if } \tau_{it} = \tau \quad \forall i, t \in \Omega_{1h}$$

- $\hat{\tau}_h$ may put negative weight on the treatment effects for some units.

Dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{h \neq -1} \tau_h 1[t = E_i + h] + \epsilon_{it}$$

- Dynamic specification is robust only to heterogeneity across h. Under cohort-specific parallel trend assumptions:

$$\hat{\tau}_h = \tau_{ATT_h} \quad \text{if } \tau_{it} = \tau \quad \forall i, t \in \Omega_{1h}$$

- $\hat{\tau}_h$ may put negative weight on the treatment effects for some units.
- Contamination-Bias between lags (Goldsmith-Pinkham et al., 2022)

Dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{h \neq -1} \tau_h 1[t = E_i + h] + \epsilon_{it}$$

- Dynamic specification is robust only to heterogeneity across h. Under cohort-specific parallel trend assumptions:

$$\hat{\tau}_h = \tau_{ATT_h} \quad \text{if } \tau_{it} = \tau \quad \forall i, t \in \Omega_{1h}$$

- $\hat{\tau}_h$ may put negative weight on the treatment effects for some units.
- Contamination-Bias between lags (Goldsmith-Pinkham et al., 2022)
- *Without never-treated units* $\{t_h\}_{h \neq -1}$ *is not point identified* (Borusyak et al., 2021)

Dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{h \neq -1} \tau_h 1[t = E_i + h] + \epsilon_{it}$$

- Dynamic specification is robust only to heterogeneity across h. Under cohort-specific parallel trend assumptions:

$$\hat{\tau}_h = \tau_{ATT_h} \quad \text{if } \tau_{it} = \tau \quad \forall i, t \in \Omega_{1h}$$

- $\hat{\tau}_h$ may put negative weight on the treatment effects for some units.
- Contamination-Bias between lags (Goldsmith-Pinkham et al., 2022)
- *Without never-treated units* $\{t_h\}_{h \neq -1}$ *is not point identified* (Borusyak et al., 2021)
- Pre-trend test fails: coefficients are contaminated by treatment effect heterogeneity (Roth, 2022)

Content

Simple Diagnostic TWFE

- Jakiela (2021) suggest two simple diagnostics

Simple Diagnostic TWFE

- Jakiela (2021) suggest two simple diagnostics
  - **First**: Compute the share of units receiving negative weights

Panel A: Dependent Variable: Gross Enrollment in Primary School

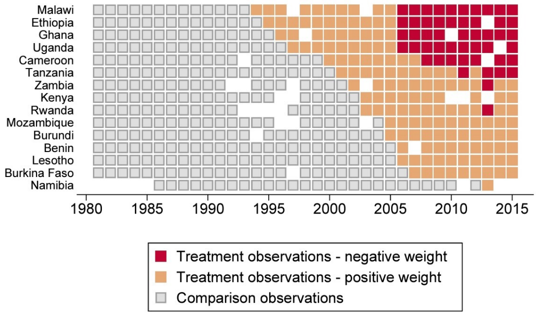


Figure: 26% of the treated units receive negative weights

Simple Diagnostic TWFE

- Jakiela (2021) suggest two simple diagnostics
  - **First**: Compute the share of units receiving negative weights

Panel A: Dependent Variable: Gross Enrollment in Primary School

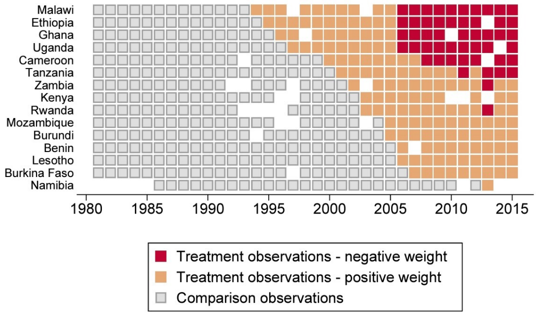


Figure: 26% of the treated units receive negative weights

Don't provide any criteria to decide when the share is big/small!

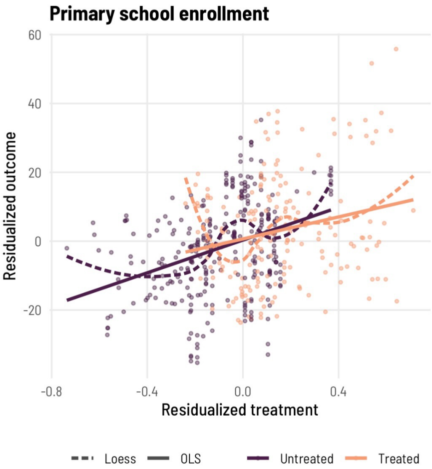
Simple Diagnostic TWFE

- **Second**: Test for heterogeneity of treatment effects

Simple Diagnostic TWFE

- **Second**: Test for heterogeneity of treatment effects

$$\tilde{Y}_{it} = \beta \tilde{D}_{it} + \gamma_0 D_{it} + \gamma_1 D_{it} * \tilde{D}_{it} + \epsilon_{it}$$



Robustness of $\hat{\tau}_{\text{static}}$ to heterogeneity

- De Chaisemartin and d'Haultfoeuille (2020) suggests reporting the degree of heterogeneity in treatment effects that would be necessary for the estimated treatment effect to have the "wrong sign."

Robustness of $\hat{\tau}_{\text{static}}$ to heterogeneity

- De Chaisemartin and d'Haultfoeuille (2020) suggests reporting the degree of heterogeneity in treatment effects that would be necessary for the estimated treatment effect to have the "wrong sign."
- Minimal heterogeneity compatible with opposite signs in coefficients:

$$\sigma_{fe} = \frac{\left|\hat{\tau}_{\text{static}}\right|}{\sigma\left(\hat{w}_{fe}\right)}$$

Robustness of $\hat{\tau}_{\text{static}}$ to heterogeneity

- De Chaisemartin and d'Haultfoeuille (2020) suggests reporting the degree of heterogeneity in treatment effects that would be necessary for the estimated treatment effect to have the "wrong sign."
- Minimal heterogeneity compatible with opposite signs in coefficients:

$$\sigma_{fe} = \frac{|\hat{\tau}_{\text{static}}|}{\sigma(\hat{w}_{fe})}$$

- σ_{fe} close to 0 then τ_{ATT} can be of opposite signs even under small treatment effect heterogeneity.

Robustness of $\hat{\tau}_{\text{static}}$ to heterogeneity

- De Chaisemartin and d'Haultfoeuille (2020) suggests reporting the degree of heterogeneity in treatment effects that would be necessary for the estimated treatment effect to have the "wrong sign."
- Minimal heterogeneity compatible with opposite signs in coefficients:

$$\sigma_{fe} = \frac{\left|\hat{\tau}_{\text{static}}\right|}{\sigma(\hat{w}_{fe})}$$

- σ_{fe} close to 0 then τ_{ATT} can be of opposite signs even under small treatment effect heterogeneity.
- Big σ_{fe} implies τ_{ATT} can only be of opposite signs under a very large and implausible amount of treatment effect heterogeneity.

Robustness of $\hat{\tau}_{\text{static}}$ to heterogeneity

- De Chaisemartin and d'Haultfoeuille (2020) suggests reporting the degree of heterogeneity in treatment effects that would be necessary for the estimated treatment effect to have the "wrong sign."
- Minimal heterogeneity compatible with opposite signs in coefficients:

$$\sigma_{fe} = \frac{|\hat{\tau}_{\text{static}}|}{\sigma(\hat{w}_{fe})}$$

- σ_{fe} close to 0 then τ_{ATT} can be of opposite signs even under small treatment effect heterogeneity.
- Big σ_{fe} implies τ_{ATT} can only be of opposite signs under a very large and implausible amount of treatment effect heterogeneity.
- When the number of cohorts goes to infinity, $\sigma(\hat{w}_{fe})$ goes to 0 and heterogeneity is not a problem.

Share of forbidden comparisons

- Goodman-Bacon (2021) proposes to evaluate how much weight is put on "forbidden comparisons".

Share of forbidden comparisons

- Goodman-Bacon (2021) proposes to evaluate how much weight is put on "forbidden comparisons".
  - Under parallel trends:

$$DD_{kl}^{late} = ATT_l^{post} + (ATT_l^{post} - ATT_k^{mid})$$

Share of forbidden comparisons

- Goodman-Bacon (2021) proposes to evaluate how much weight is put on "forbidden comparisons".
  - Under parallel trends:

$$DD_{kl}^{late} = ATT_l^{post} + (ATT_l^{post} - ATT_k^{mid})$$

  - Weights:

$$S_{\text{forbidden}} = \sum_{k \neq U} \sum_{l > k} s_{lk}$$

Share of forbidden comparisons

- Goodman-Bacon (2021) proposes to evaluate how much weight is put on "forbidden comparisons".
  - Under parallel trends:

$$DD_{kl}^{late} = ATT_l^{post} + (ATT_l^{post} - ATT_k^{mid})$$

  - Weights:

$$S_{\text{forbidden}} = \sum_{k \neq U} \sum_{l > k} s_{lk}$$

$$s_{lk} = \frac{(n_k + n_l)\bar{D}_l(\alpha_{kl}(1-\alpha_{kl})\frac{\bar{D}_k - \bar{D}_l}{1-\bar{D}_l}\frac{1-\bar{D}_k}{1-\bar{D}_l})}{\hat{V}(D)} > 0 \quad \text{for all } l, k$$

Share of forbidden comparisons

- Goodman-Bacon (2021) proposes to evaluate how much weight is put on "forbidden comparisons".
  - Under parallel trends:

$$DD_{kl}^{late} = ATT_l^{post} + (ATT_l^{post} - ATT_k^{mid})$$

  - Weights:

$$S_{\text{forbidden}} = \sum_{k \neq U} \sum_{l > k} s_{lk}$$

$$s_{lk} = \frac{(n_k + n_l)\bar{D}_l(\alpha_{kl}(1-\alpha_{kl})\frac{\bar{D}_k - \bar{D}_l}{1-\bar{D}_l}\frac{1-\bar{D}_k}{1-\bar{D}_l})}{\hat{V}(D)} > 0 \quad \text{for all } l, k$$

- $0 < S_{\text{forbidden}} < 1$ is the total weight put on forbidden comparisons

Content

Imputation Methods

- Borusyak et al. (2021) (`did_imputation`) and Gardner (2022) (`did2s`) propose a numerically equivalent two-step estimator that is robust to treatment effect heterogeneity.

Imputation Methods

- Borusyak et al. (2021) (`did_imputation`) and Gardner (2022) (`did2s`) propose a numerically equivalent two-step estimator that is robust to treatment effect heterogeneity.
  1. **First Stage**: Estimate $\hat{Y}(0)_{it}$ by OLS using only the non-treated units:

$$Y(0)_{it} = \alpha_i + \beta_t + \epsilon_{it} \quad it \in \Omega_0$$

Imputation Methods

- Borusyak et al. (2021) (`did_imputation`) and Gardner (2022) (`did2s`) propose a numerically equivalent two-step estimator that is robust to treatment effect heterogeneity.
  1. **First Stage**: Estimate $\hat{Y}(0)_{it}$ by OLS using only the non-treated units:

     $$Y(0)_{it} = \alpha_i + \beta_t + \epsilon_{it} \quad it \in \Omega_0$$

  2. **Second Stage**: Estimate $\hat{\tau}_i t$ for the treated units:

     $$\hat{\tau}_{it} = Y_{it} - \hat{Y}(0)_{it}$$

Imputation Methods

- Borusyak et al. (2021) (`did_imputation`) and Gardner (2022) (`did2s`) propose a numerically equivalent two-step estimator that is robust to treatment effect heterogeneity.
  1. **First Stage**: Estimate $\hat{Y}(0)_{it}$ by OLS using only the non-treated units:
  $$Y(0)_{it} = \alpha_i + \beta_t + \epsilon_{it} \quad it \in \Omega_0$$
  2. **Second Stage**: Estimate $\hat{\tau}_i t$ for the treated units:
  $$\hat{\tau}_{it} = Y_{it} - \hat{Y}(0)_{it}$$
  3. Compute the weighted average using population weights:
  $$\hat{\tau}_{imputation} = \sum_{it \in \Omega_1} w_{it} \hat{\tau}_{it}$$

Imputation Methods

- Borusyak et al. (2021) (`did_imputation`) and Gardner (2022) (`did2s`) propose a numerically equivalent two-step estimator that is robust to treatment effect heterogeneity.
  1. **First Stage**: Estimate $\hat{Y}(0)_{it}$ by OLS using only the non-treated units:
  $$Y(0)_{it} = \alpha_i + \beta_t + \epsilon_{it} \quad it \in \Omega_0$$
  2. **Second Stage**: Estimate $\hat{\tau}_i t$ for the treated units:
  $$\hat{\tau}_{it} = Y_{it} - \hat{Y}(0)_{it}$$
  3. Compute the weighted average using population weights:
  $$\hat{\tau}_{imputation} = \sum_{it \in \Omega_1} w_{it} \hat{\tau}_{it}$$

Alternatively, BJS proposes choosing w_{it} in an efficient way to improve inference.

Inference

- Exact inference becomes infeasible when treatment effects are heterogeneous, but conservative inference is possible.

Inference

- Exact inference becomes infeasible when treatment effects are heterogeneous, but conservative inference is possible.

$$\hat{\sigma}_w^2 = \sum_i \left(\sum_{t, it \in \Omega_1} w_{it} \tilde{\epsilon}_{it} \right)$$

Inference

- Exact inference becomes infeasible when treatment effects are heterogeneous, but conservative inference is possible.

$$\hat{\sigma}_w^2 = \sum_i \left(\sum_{t, it \in \Omega_1} w_{it} \tilde{\epsilon}_{it} \right)$$

- Conservative inference leads to over-rejection for the null hypothesis.

Inference

- Exact inference becomes infeasible when treatment effects are heterogeneous, but conservative inference is possible.

$$\hat{\sigma}_w^2 = \sum_i \left(\sum_{t, it \in \Omega_1} w_{it} \tilde{\epsilon}_{it} \right)$$

- Conservative inference leads to over-rejection for the null hypothesis.
- Gardner (2022) proposes a *GMM* based inference that accounts the the two-steps in the estimation process.

Inference

- Exact inference becomes infeasible when treatment effects are heterogeneous, but conservative inference is possible.

$$\hat{\sigma}_w^2 = \sum_i \left(\sum_{t, it \in \Omega_1} w_{it} \tilde{\epsilon}_{it} \right)$$

- Conservative inference leads to over-rejection for the null hypothesis.
- Gardner (2022) proposes a *GMM* based inference that accounts the the two-steps in the estimation process.

$$\mathbb{E} \begin{bmatrix} (1 - D_{it})(Y_{it} - \alpha_i - \beta_t) \\ \sum_{t=1}^{T} D_{it}(Y_{it} - \alpha_i - \beta_t - \tau D_{it}) \end{bmatrix} = 0$$

Inference

- Exact inference becomes infeasible when treatment effects are heterogeneous, but conservative inference is possible.

$$\hat{\sigma}_w^2 = \sum_i \left(\sum_{t, it \in \Omega_1} w_{it} \tilde{\epsilon}_{it} \right)$$

- Conservative inference leads to over-rejection for the null hypothesis.
- Gardner (2022) proposes a *GMM* based inference that accounts the the two-steps in the estimation process.

$$\mathbb{E} \begin{bmatrix} (1 - D_{it})(Y_{it} - \alpha_i - \beta_t) \\ \sum_{t=1}^{T} D_{it}(Y_{it} - \alpha_i - \beta_t - \tau D_{it}) \end{bmatrix} = 0$$

- The asymptotic variance of $\hat{\tau}$ results from the quadratic form of the moment conditions.

Manual averaging

- Callaway and Sant'Anna (2021), De Chaisemartin and d'Haultfoeuille (2020) and Sun and Abraham (2021) propose a manual averaging of building blocks in the form:

$$ATT(g,t) = \mathbb{E}[Y_{i,t}(g) - Y_{i,g-1}(\infty)|E_i = g]$$

Manual averaging

- Callaway and Sant'Anna (2021), De Chaisemartin and d'Haultfoeuille (2020) and Sun and Abraham (2021) propose a manual averaging of building blocks in the form:

$$ATT(g,t) = \mathbb{E}[Y_{i,t}(g) - Y_{i,g-1}(\infty) | E_i = g]$$

- Identification of the previous estimand requires a less restrictive assumption of the parallel trends than the imputation method (Post-treatment only):

$$\mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty) | E_i = g] = \mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty) | E_i = g']$$

Manual averaging

- Callaway and Sant'Anna (2021), De Chaisemartin and d'Haultfoeuille (2020) and Sun and Abraham (2021) propose a manual averaging of building blocks in the form:

$$ATT(g,t) = \mathbb{E}[Y_{i,t}(g) - Y_{i,g-1}(\infty)|E_i = g]$$

- Identification of the previous estimand requires a less restrictive assumption of the parallel trends than the imputation method (Post-treatment only):

$$\mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty)|E_i = g] = \mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty)|E_i = g']$$

- $ATT(g,t)$ can be estimated as the 2x2 *DID* between cohort g and some subset from the not treated cohorts denoted as $\Omega_{0,Cg}$ and cardinality $|\Omega_{0,Cg'}| = N_{Cg'}$.

$$ATT(g,t) = \frac{1}{N_g}\sum_{t=1}^{N_g}(Y_{i,t} - Y_{i,g-1}) - \frac{1}{N_{Cg'}}\sum_{t=1}^{N_{Cg'}}(Y_{i,t} - Y_{i,g-1})$$

Manual averaging: Callaway and Sant'Anna

- Callaway and Sant'Anna (2021) proposes two options for Cg:

Manual averaging: Callaway and Sant'Anna

- Callaway and Sant'Anna (2021) proposes two options for Cg:
  - $Cg' = \{\infty\}$: Using only *never-treated* units as comparison group

Manual averaging: Callaway and Sant'Anna

- Callaway and Sant'Anna (2021) proposes two options for Cg:
  - $Cg' = \{\infty\}$: Using only *never-treated* units as comparison group
  - $Cg' = \{g' : g' > t\}$: Using all the *not-yet-treated* units as comparison group

Manual averaging: Callaway and Sant'Anna

- Callaway and Sant'Anna (2021) proposes two options for Cg:
  - $Cg' = \{\infty\}$: Using only *never-treated* units as comparison group
  - $Cg' = \{g' : g' > t\}$: Using all the *not-yet-treated* units as comparison group
- Decision between the two options is based on the parallel-trend assumption one is willing to make.

Manual averaging: Callaway and Sant'Anna

- Callaway and Sant'Anna (2021) proposes two options for Cg:
  - $Cg' = \{\infty\}$: Using only *never-treated* units as comparison group
  - $Cg' = \{g' : g' > t\}$: Using all the *not-yet-treated* units as comparison group
- Decision between the two options is based on the parallel-trend assumption one is willing to make.
- After all cohorts have been treated, only the never treated units are valid as comparison group.

Manual averaging: Callaway and Sant'Anna

- Callaway and Sant'Anna (2021) proposes two options for Cg:
  - $Cg' = \{\infty\}$: Using only *never-treated* units as comparison group
  - $Cg' = \{g' : g' > t\}$: Using all the *not-yet-treated* units as comparison group
- Decision between the two options is based on the parallel-trend assumption one is willing to make.
- After all cohorts have been treated, only the never treated units are valid as comparison group.
- The last step is a computing the average of the $ATT(g,t)$ using weights according to the estimated paramter:

$$t\hat{a}u_{ATT} = \sum_g tw_{gt} A\hat{T}T(g,t) \quad t\hat{a}u_h = \sum_g w_g A\hat{T}T(g, g+h)$$

Manual averaging: Callaway and Sant'Anna

- Callaway and Sant'Anna (2021) proposes two options for Cg:
  - $Cg' = \{\infty\}$: Using only *never-treated* units as comparison group
  - $Cg' = \{g' : g' > t\}$: Using all the *not-yet-treated* units as comparison group
- Decision between the two options is based on the parallel-trend assumption one is willing to make.
- After all cohorts have been treated, only the never treated units are valid as comparison group.
- The last step is a computing the average of the $ATT(g, t)$ using weights according to the estimated paramter:

$$t\hat{a}u_{ATT} = \sum_g tw_{gt} A\hat{T}T(g, t) \quad t\hat{a}u_h = \sum_g w_g A\hat{T}T(g, g + h)$$

- Some criteria is provided on how to choose weights

Imputation vs Manual Averaging

	i=A	i=B	i=Z
t=1	α_A	α_B	α_Z
t=2	$\alpha_A + \beta_2^g$	$\alpha_B + \beta_2^g$	$\alpha_Z + \beta_2^z$
t=3	$\alpha_A + \beta_3^g + \tau_{A3}$	$\alpha_B + \beta_3^g + \tau_{B3}$	$\alpha_Z + \beta_3^z$

Note that $g = 3$ is the only cohort that has been treated

$$\hat{\tau}_{CS}(3,3) = \left(\frac{Y_{A3}+Y_{B3}}{2} - Y_{Z3}\right) - \left(\frac{Y_{A2}+Y_{B2}}{2} - Y_{Z2}\right)$$

Imputation vs Manual Averaging

	i=A	i=B	i=Z
t=1	α_A	α_B	α_Z
t=2	$\alpha_A + \beta_2^g$	$\alpha_B + \beta_2^g$	$\alpha_Z + \beta_2^z$
t=3	$\alpha_A + \beta_3^g + \tau_{A3}$	$\alpha_B + \beta_3^g + \tau_{B3}$	$\alpha_Z + \beta_3^z$

Note that $g = 3$ is the only cohort that has been treated

$$\hat{\tau}_{CS}(3,3) = \left(\frac{Y_{A3}+Y_{B3}}{2} - Y_{Z3}\right) - \left(\frac{Y_{A2}+Y_{B2}}{2} - Y_{Z2}\right)$$

$$\hat{\tau}_{BJ}(3,3) = \left(\frac{Y_{A3}+Y_{B3}}{2} - Y_{Z3}\right) - \left(\frac{Y_{A1}+Y_{B1}+Y_{A2}+Y_{B2}}{4} - \frac{Y_{Z2}+Y_{Z1}}{2}\right)$$

Imputation vs Manual Averaging

	i=A	i=B	i=Z
t=1	α_A	α_B	α_Z
t=2	$\alpha_A + \beta_2^g$	$\alpha_B + \beta_2^g$	$\alpha_Z + \beta_2^z$
t=3	$\alpha_A + \beta_3^g + \tau_{A3}$	$\alpha_B + \beta_3^g + \tau_{B3}$	$\alpha_Z + \beta_3^z$

Note that $g = 3$ is the only cohort that has been treated

$$\hat{\tau}_{CS}(3,3) = \left(\frac{Y_{A3}+Y_{B3}}{2} - Y_{Z3}\right) - \left(\frac{Y_{A2}+Y_{B2}}{2} - Y_{Z2}\right)$$

$$\hat{\tau}_{BJ}(3,3) = \left(\frac{Y_{A3}+Y_{B3}}{2} - Y_{Z3}\right) - \left(\frac{Y_{A1}+Y_{B1}+Y_{A2}+Y_{B2}}{4} - \frac{Y_{Z2}+Y_{Z1}}{2}\right)$$

Suppose only group-parallel-trends assumptions in CS holds, then:

$$\hat{\tau}_{CS}(3,3) = \hat{\tau}_{BJ}(3,3) + \frac{\beta_2^g - \beta_2^z}{2}$$

Imputation vs Manual Averaging

	i=A	i=B	i=Z
t=1	α_A	α_B	α_Z
t=2	$\alpha_A + \beta_2^g$	$\alpha_B + \beta_2^g$	$\alpha_Z + \beta_2^z$
t=3	$\alpha_A + \beta_3^g + \tau_{A3}$	$\alpha_B + \beta_3^g + \tau_{B3}$	$\alpha_Z + \beta_3^z$

Note that $g = 3$ is the only cohort that has been treated

$$\hat{\tau}_{CS}(3,3) = \left(\frac{Y_{A3}+Y_{B3}}{2} - Y_{Z3}\right) - \left(\frac{Y_{A2}+Y_{B2}}{2} - Y_{Z2}\right)$$

$$\hat{\tau}_{BJ}(3,3) = \left(\frac{Y_{A3}+Y_{B3}}{2} - Y_{Z3}\right) - \left(\frac{Y_{A1}+Y_{B1}+Y_{A2}+Y_{B2}}{4} - \frac{Y_{Z2}+Y_{Z1}}{2}\right)$$

Suppose only group-parallel-trends assumptions in CS holds, then:

$$\hat{\tau}_{CS}(3,3) = \hat{\tau}_{BJ}(3,3) + \frac{\beta_2^g - \beta_2^z}{2}$$

Trade-off is between efficiency (BJ) vs strength of assumptions (CS)

Simulation design

- Similarly to Bertrand et al. (2004), Gardner (2022) simulates random placebo interventions and compares the performance of the different estimators.

Simulation design

- Similarly to Bertrand et al. (2004), Gardner (2022) simulates random placebo interventions and compares the performance of the different estimators.
- **Data**: Current population survey (CPS) from 1979 to 2020. 50 states by 41 years balanced panel of average wages. 2100 state-year cells.

Simulation design

- Similarly to Bertrand et al. (2004), Gardner (2022) simulates random placebo interventions and compares the performance of the different estimators.
- **Data**: Current population survey (CPS) from 1979 to 2020. 50 states by 41 years balanced panel of average wages. 2100 state-year cells.
- **Output**: state-year average of log wage of women between ages 25 and 50 in the CPS. ($n > 1 million$)

Simulation design

- Similarly to Bertrand et al. (2004), Gardner (2022) simulates random placebo interventions and compares the performance of the different estimators.
- **Data**: Current population survey (CPS) from 1979 to 2020. 50 states by 41 years balanced panel of average wages. 2100 state-year cells.
- **Output**: state-year average of log wage of women between ages 25 and 50 in the CPS. ($n > 1 million$)
- **Treatment**: Staggered random assignment. Create a vector of groups $E = \{g_1, g_2, ..., g_23\}$ for years $1982 < g < 2005$ and then randomly assign 40 states ensuring at least one to each group. Treatment lasts exactly 15 years for each group.

Simulation results

Table 5: Simulations (CPS wage data, heterogeneous treatment effects): 40 states treated over 15 years (at least 1 per year)

Method	Period	Rejection rate	S.E.	Bias	RMSE	Speed (secs)
2SDD	0	5.2	0.6278	0.0316	0.6675	0.1409
	1	4.6	0.6393	-0.0167	0.6299	
	2	5.6	0.6320	0.0307	0.6738	
	3	4.2	0.6368	0.0462	0.6221	
	4	6.4	0.6395	-0.0026	0.6184	
BJS	0	12.7	0.5176	0.0316	0.6675	0.9318
	1	10.9	0.5267	-0.0167	0.6299	
	2	15.0	0.5188	0.0307	0.6738	
	3	10.9	0.5175	0.0462	0.6221	
	4	10.5	0.5329	-0.0026	0.6184	
CS	0	5.9	0.8952	0.0486	0.8910	26.2
	1	6.4	0.9086	0.0076	0.9216	
	2	6.4	0.8989	0.0675	0.9158	
	3	3.6	0.8970	0.0789	0.8529	
	4	4.5	0.9006	0.0437	0.8693	
SA	0	1.8	1.0175	0.0475	0.8802	28.4
	1	2.3	1.0339	0.0071	0.9069	
	2	3.2	1.0223	0.0664	0.9011	
	3	2.3	1.0204	0.0781	0.8466	
	4	2.7	1.0269	0.0419	0.8433	

Note: The table reports results from 500 simulations of 40 treated states over 15 years, with at least one treated state in each of those years. See the note accompanying Table 1 for further information.

References I

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Gardner, J. (2022). Two-stage differences in differences. *arXiv preprint arXiv:2207.05943*.

References II

Goldsmith-Pinkham, P., Hull, P., and Kolesár, M. (2022). Contamination bias in linear regressions. Technical report, National Bureau of Economic Research.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Jakiela, P. (2021). Simple diagnostics for two-way fixed effects. *arXiv preprint arXiv:2103.13229*.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–22.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.