A Negative Correlation Strategy for Bracketing in Difference-in-Differences

Ting Ye

Department of Biostatistics, University of Washington
tingye1@uw.edu
Slides at https://ting-ye.com/talks

January 5, 2024 @ ASSA

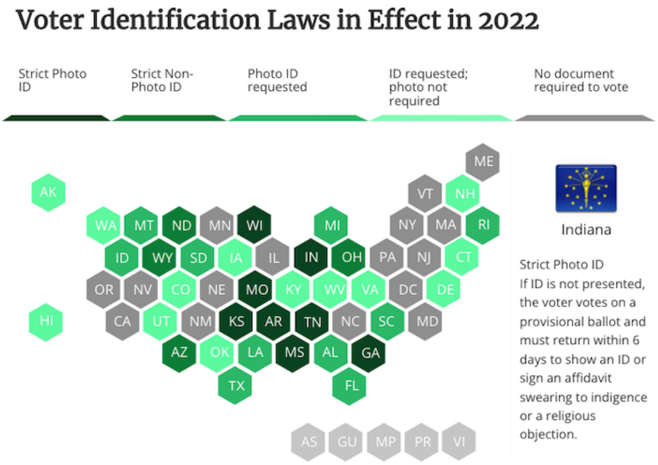
Acknowledgments

Dylan Small
UPenn

Luke Keele
UPenn

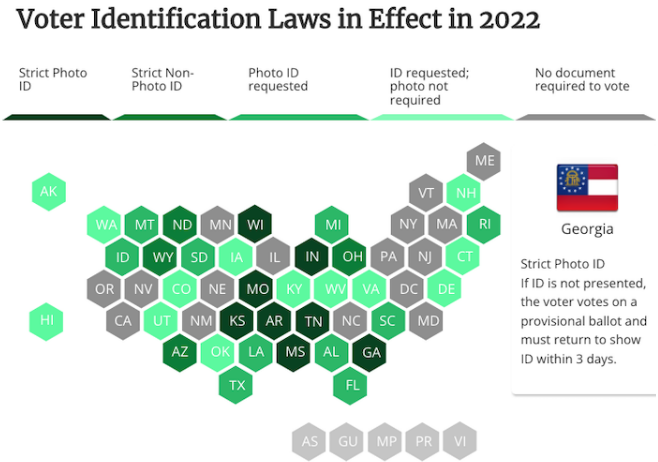
Raiden Hasegawa
Google

Strict Voter ID Laws



(National Conference of State Legislatures)

Strict Voter ID Laws



(National Conference of State Legislatures)

Do Strict Voter ID Laws Change Turnout?

Challenges

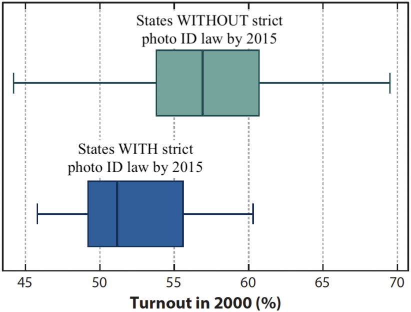
> *The fact that states with strict identification laws differ from states without them in other ways that may be related to turnout complicates the casual inference process.* (Highton, 2017[1])

[1] Highton (2017). Voter Identification Laws and Turnout in the United States. *Annu. Rev. Polit. Sci.*

Do Strict Voter ID Laws Change Turnout?

Challenges

The fact that states with strict identification laws differ from states without them in other ways that may be related to turnout complicates the casual inference process. (Highton, 2017[1])



(Highton, 2017. Data source: United States Elections Project.)

[1] Highton (2017). Voter Identification Laws and Turnout in the United States. *Annu. Rev. Polit. Sci.*

DID for Causal Effects

- Difference-in-Differences (DID) is commonly used for estimating causal effects.

DID for Causal Effects

- Difference-in-Differences (DID) is commonly used for estimating causal effects.
- Prototypical DID application: how do changes in state policies affect individual.

DID for Causal Effects

- Difference-in-Differences (DID) is commonly used for estimating causal effects.
- Prototypical DID application: how do changes in state policies affect individual.
- Did Missouri's handgun purchaser licensing law affects firearm homicide rates?

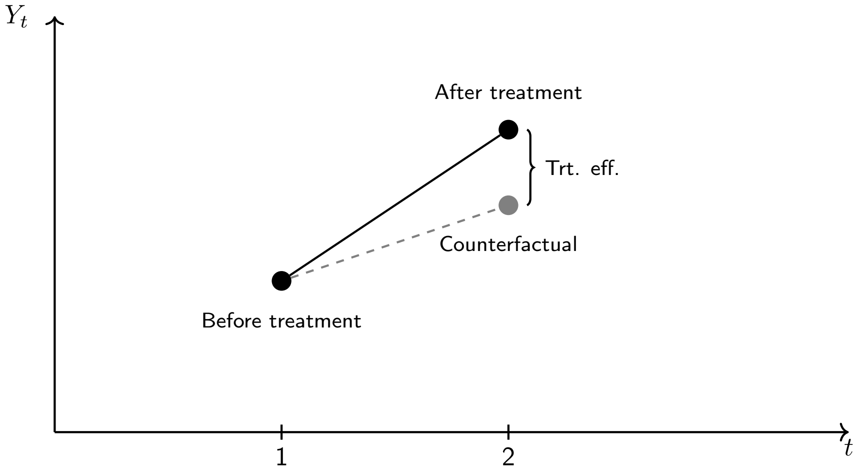
DID for Causal Effects

- Difference-in-Differences (DID) is commonly used for estimating causal effects.
- Prototypical DID application: how do changes in state policies affect individual.
- Did Missouri's handgun purchaser licensing law affects firearm homicide rates?
- Did minimum wage laws change drug overdose death?

DID for Causal Effects

- Difference-in-Differences (DID) is commonly used for estimating causal effects.
- Prototypical DID application: how do changes in state policies affect individual.
- Did Missouri's handgun purchaser licensing law affects firearm homicide rates?
- Did minimum wage laws change drug overdose death?
- Motivating application: Did strict voter ID laws change turnout?

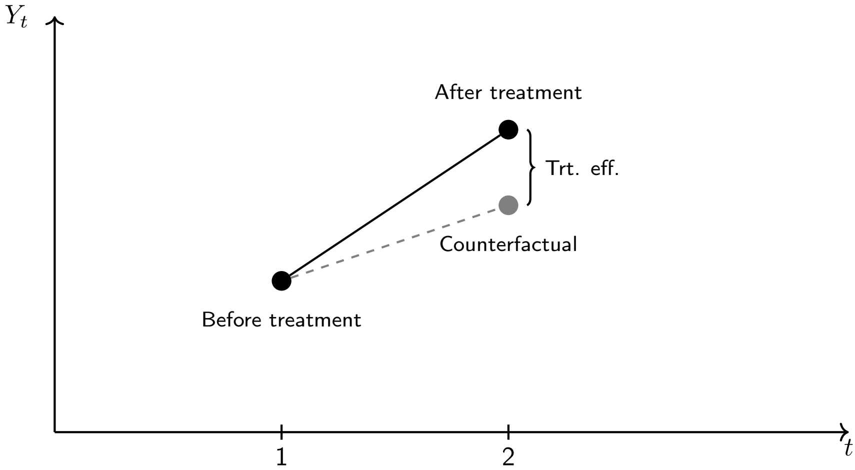
DID for Causal Effects



DID for Causal Effects

Identify the counterfactual $\Leftrightarrow$ Identify the treatment effect

(Note: Identification = express quantities of interest using observed variables)

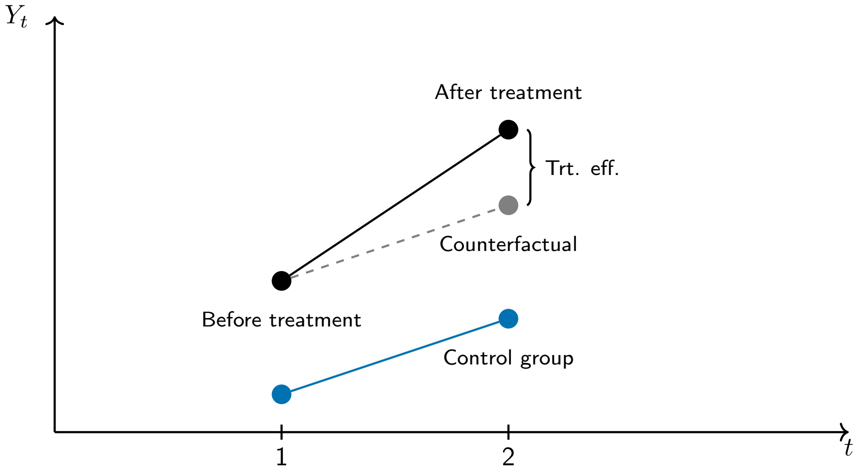


DID for Causal Effects

Identify the counterfactual ⇔ Identify the treatment effect

(Note: Identification = express quantities of interest using observed variables)

Parallel Trends: Absent treatment, treated and control would evolve over time in the same way.

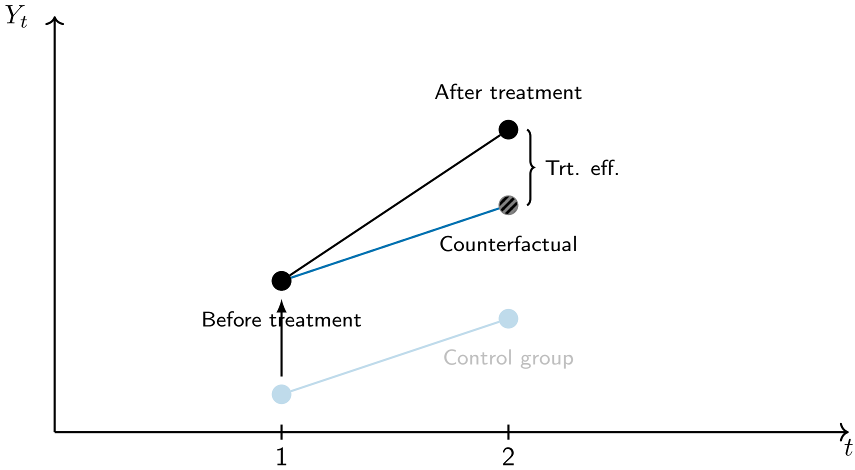


DID for Causal Effects

Identify the counterfactual ⇔ Identify the treatment effect

(Note: Identification = express quantities of interest using observed variables)

Parallel Trends: Absent treatment, treated and control would evolve over time in the same way.

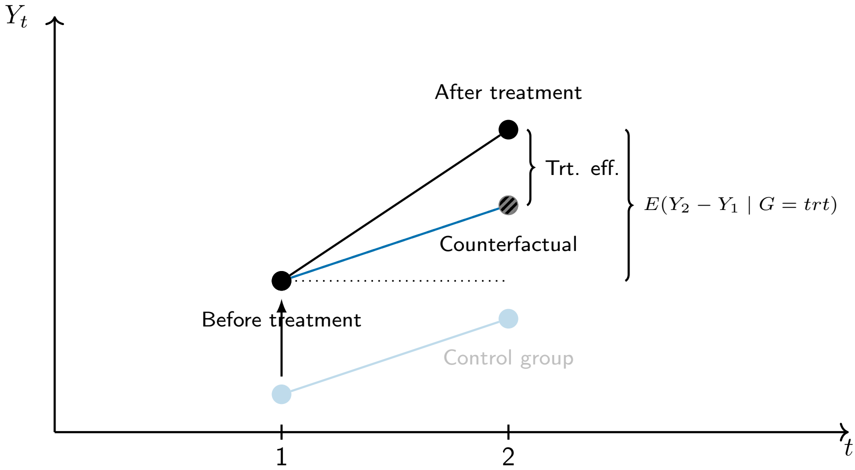


DID for Causal Effects

Identify the counterfactual ⇔ Identify the treatment effect

(Note: Identification = express quantities of interest using observed variables)

Parallel Trends: Absent treatment, treated and control would evolve over time in the same way.

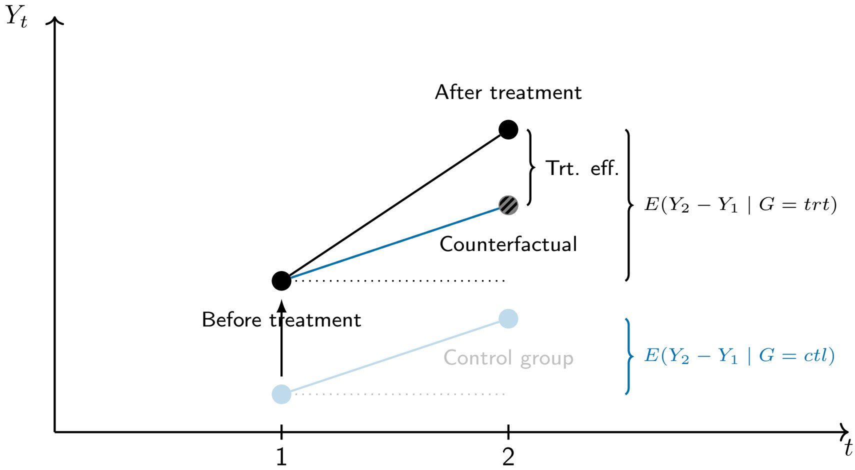


DID for Causal Effects

Identify the counterfactual ⇔ Identify the treatment effect

(Note: Identification = express quantities of interest using observed variables)

Parallel Trends: Absent treatment, treated and control would evolve over time in the same way.

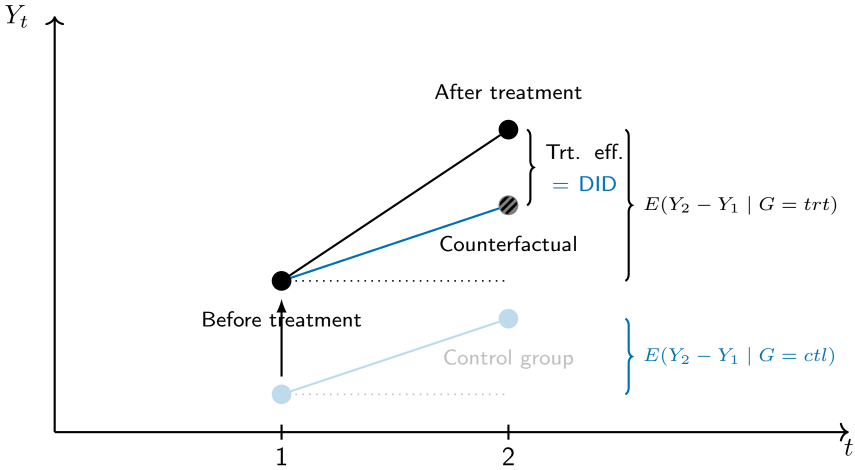


DID for Causal Effects

Identify the counterfactual ⇔ Identify the treatment effect

(Note: Identification = express quantities of interest using observed variables)

Parallel Trends: Absent treatment, treated and control would evolve over time in the same way.



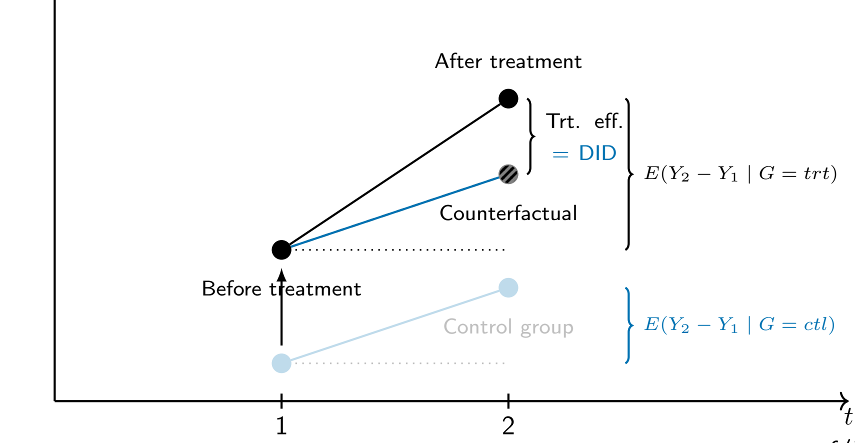
DID for Causal Effects

Identify the counterfactual $\Leftrightarrow$ Identify the treatment effect

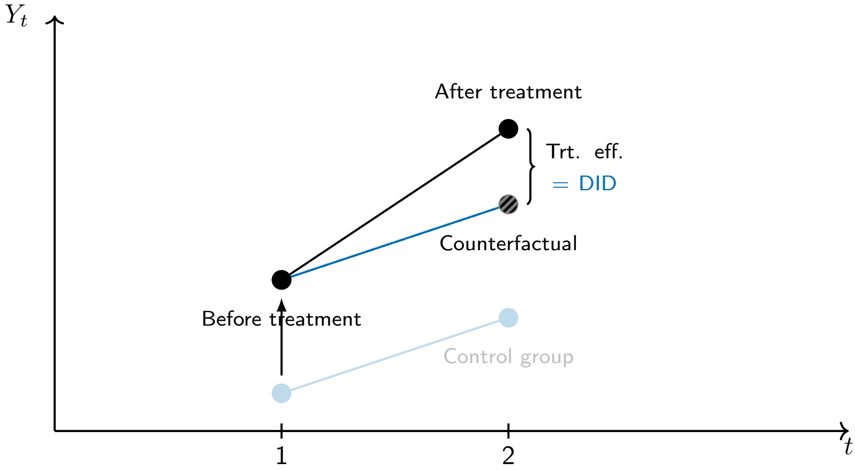
(Note: Identification = express quantities of interest using observed variables)

Parallel Trends: Absent treatment, treated and control would evolve over time in the same way.

Alternatively: $Y_{igt} = \lambda_t + \alpha_g + \text{DID} \cdot D_{gt} + \varepsilon_{igt}$

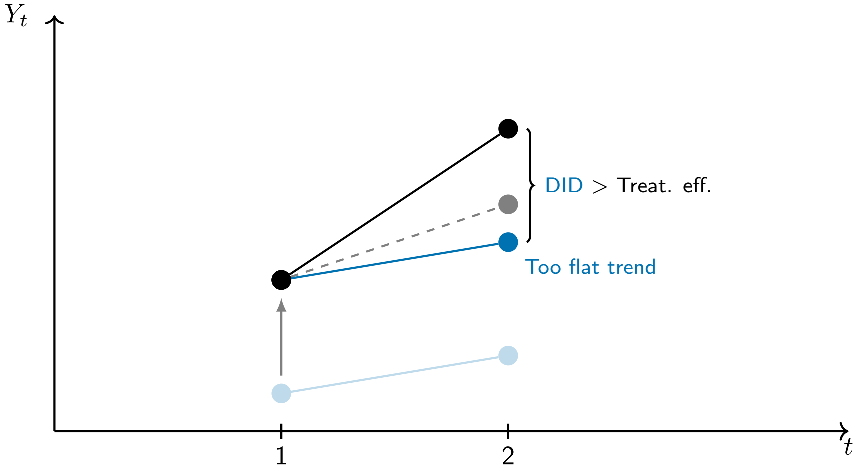


Is Parallel Trends Plausible?



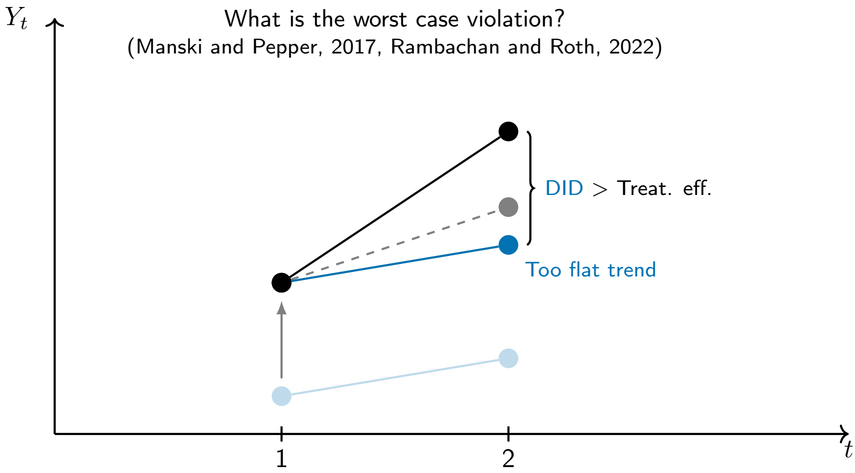
Is Parallel Trends Plausible?

The parallel trends can be easily violated due to state-specific events (e.g., the treated group was targeted by both major-party campaigns in the post-treatment period), resulting in biased DID estimates.



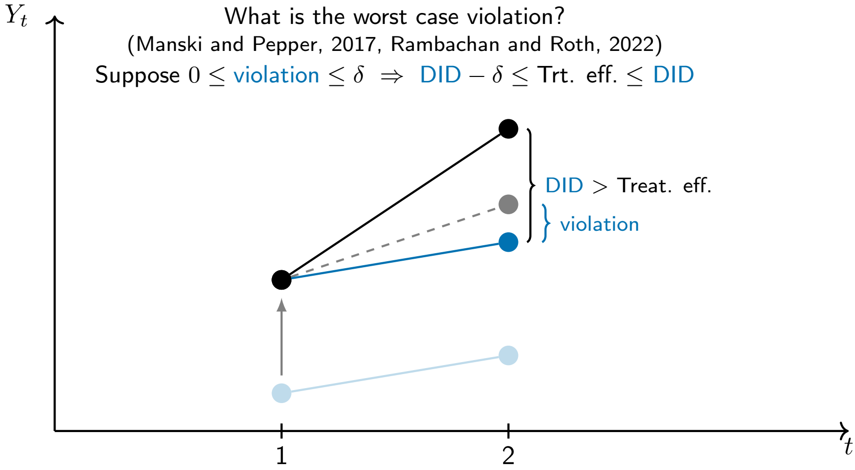
Is Parallel Trends Plausible?

The parallel trends can be easily violated due to state-specific events (e.g., the treated group was targeted by both major-party campaigns in the post-treatment period), resulting in biased DID estimates.



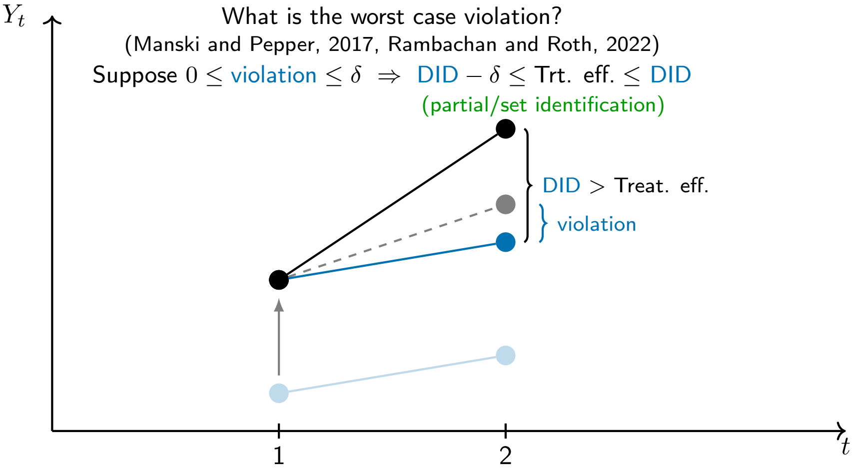
Is Parallel Trends Plausible?

The parallel trends can be easily violated due to state-specific events (e.g., the treated group was targeted by both major-party campaigns in the post-treatment period), resulting in biased DID estimates.



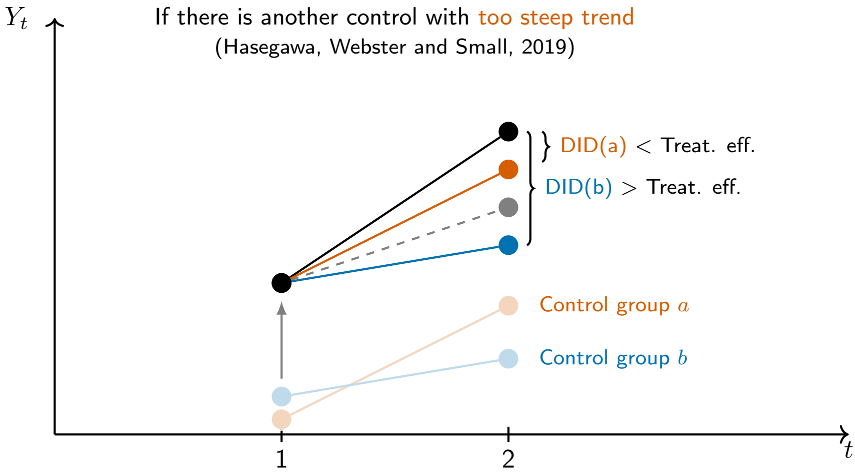
Is Parallel Trends Plausible?

The parallel trends can be easily violated due to state-specific events (e.g., the treated group was targeted by both major-party campaigns in the post-treatment period), resulting in biased DID estimates.



Is Parallel Trends Plausible?

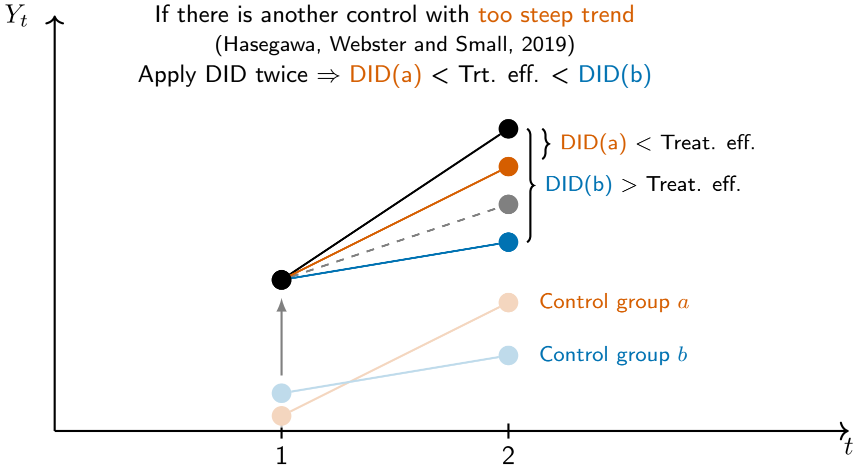
The parallel trends can be easily violated due to state-specific events (e.g., the treated group was targeted by both major-party campaigns in the post-treatment period), resulting in biased DID estimates.



Is Parallel Trends Plausible?

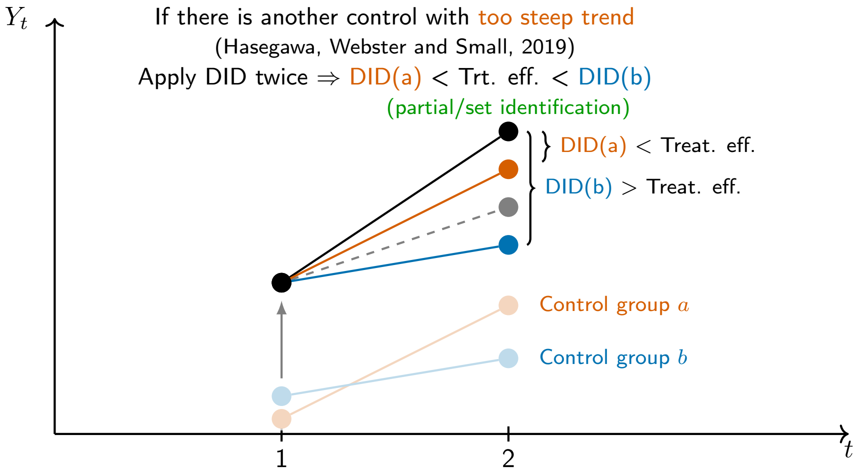
The parallel trends can be easily violated due to state-specific events (e.g., the treated group was targeted by both major-party campaigns in the post-treatment period), resulting in biased DID estimates.

If there is another control with too steep trend
(Hasegawa, Webster and Small, 2019)
Apply DID twice ⇒ DID(a) $<$ Trt. eff. $<$ DID(b)



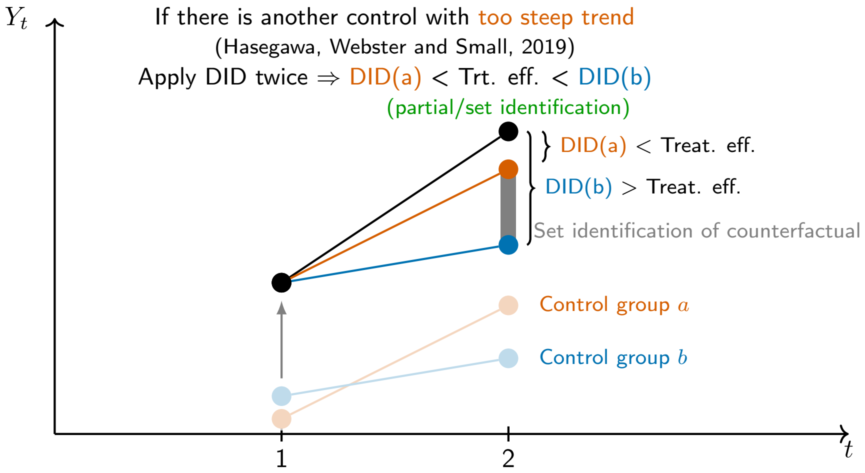
Is Parallel Trends Plausible?

The parallel trends can be easily violated due to state-specific events (e.g., the treated group was targeted by both major-party campaigns in the post-treatment period), resulting in biased DID estimates.



Is Parallel Trends Plausible?

The parallel trends can be easily violated due to state-specific events (e.g., the treated group was targeted by both major-party campaigns in the post-treatment period), resulting in biased DID estimates.



Outline for DID Bracket

Motivation: multiple timepoints

- Data are available for multiple post-treatment time periods.
- We are also interested in how treatment effect changes over time.

Outline for DID Bracket

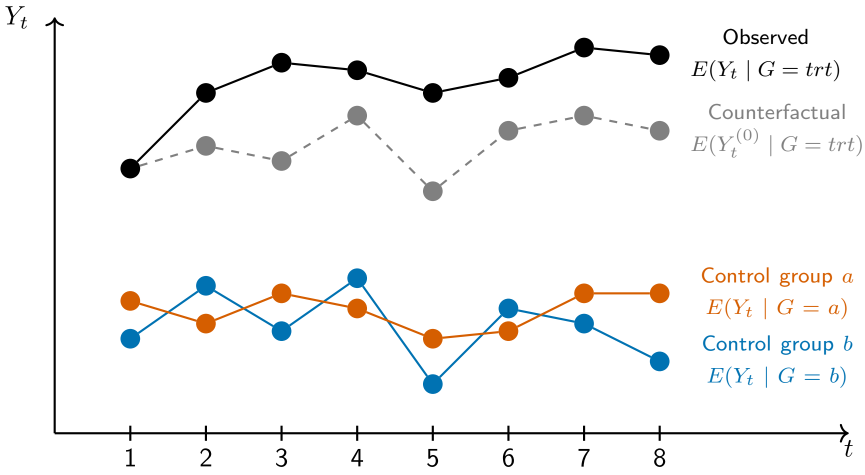
Motivation: multiple timepoints

- Data are available for multiple post-treatment time periods.
- We are also interested in how treatment effect changes over time.

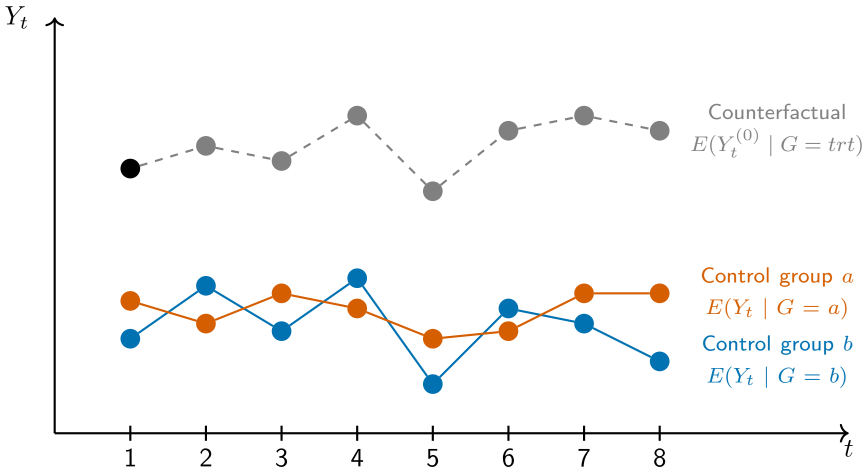
A framework for more flexible DID analysis ~~(parallel trends)~~

- Partial/set identification of the treatment effect
  - Bracketed trends assumption
  - Negative correlation strategy
  - Bracket the treatment effect
- Novel bootstrap inference

Bracketed Trends Assumption

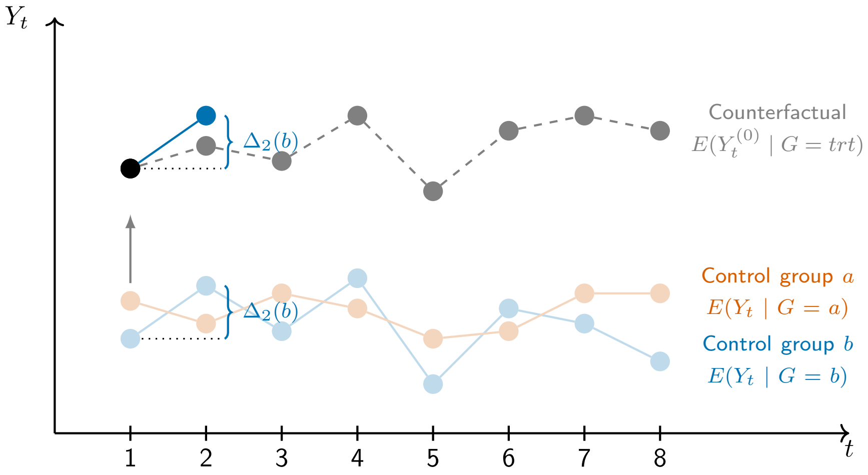


Bracketed Trends Assumption



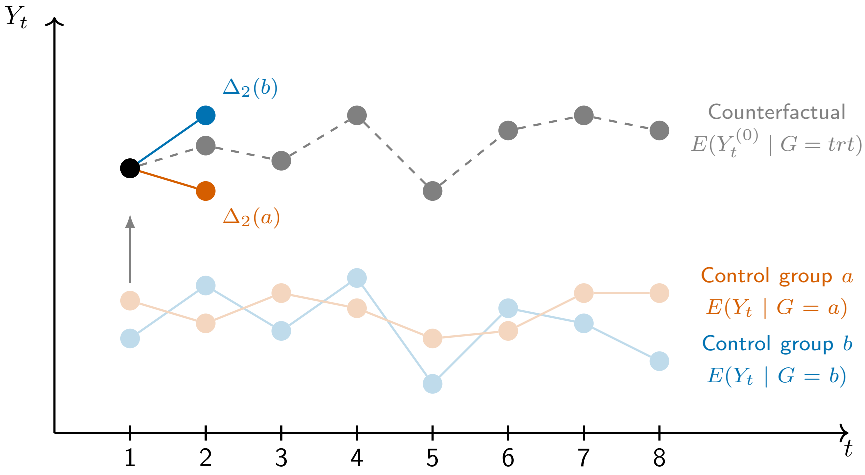
Bracketed Trends Assumption

Let $\Delta_t(g) = E(Y_t - Y_{t-1} | G = g)$, $g \in \{a, b\}$ (change in outcome)



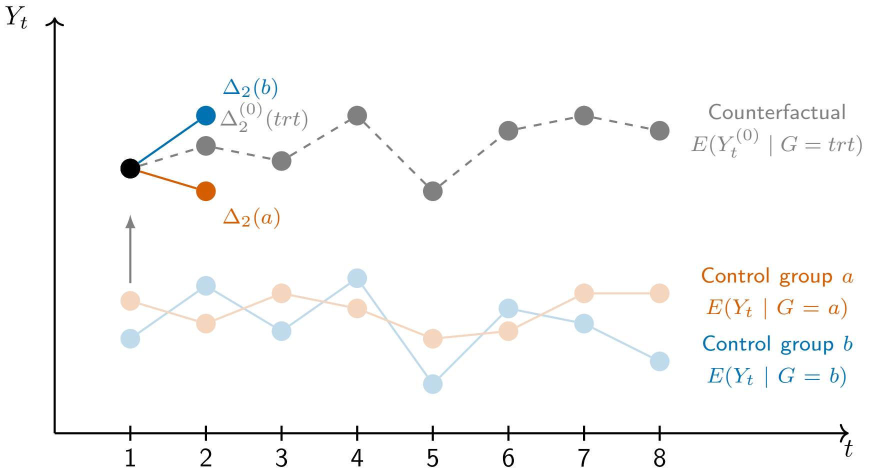
Bracketed Trends Assumption

Let $\Delta_t(g) = E(Y_t - Y_{t-1}|G = g)$, $g \in \{a, b\}$ (change in outcome)



Bracketed Trends Assumption

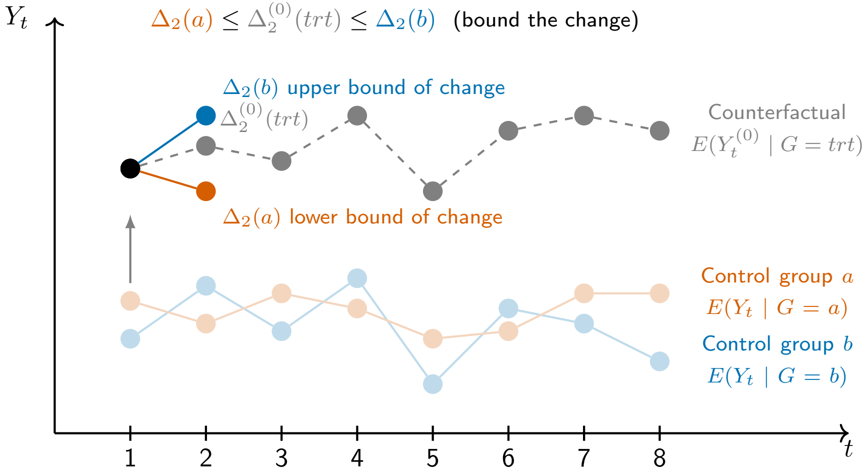
Let $\Delta_t(g) = E(Y_t - Y_{t-1}|G = g)$, $g \in \{a, b\}$ (change in outcome)



Bracketed Trends Assumption

Let $\Delta_t(g) = E(Y_t - Y_{t-1} | G = g)$, $g \in \{a, b\}$ (change in outcome)

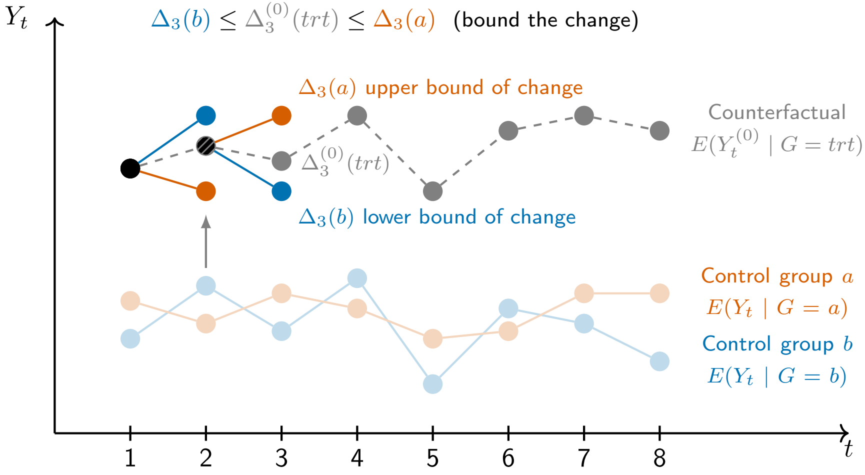
$\Delta_2(a) \leq \Delta_2^{(0)}(trt) \leq \Delta_2(b)$ (bound the change)



Bracketed Trends Assumption

Let $\Delta_t(g) = E(Y_t - Y_{t-1} | G = g)$, $g \in \{a, b\}$ (change in outcome)

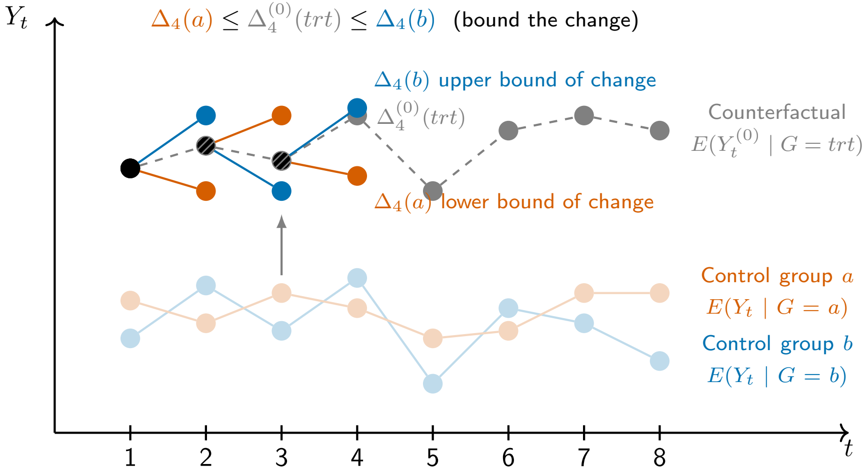
$\Delta_3(b) \leq \Delta_3^{(0)}(trt) \leq \Delta_3(a)$ (bound the change)



Bracketed Trends Assumption

Let $\Delta_t(g) = E(Y_t - Y_{t-1}|G = g)$, $g \in \{a, b\}$ (change in outcome)

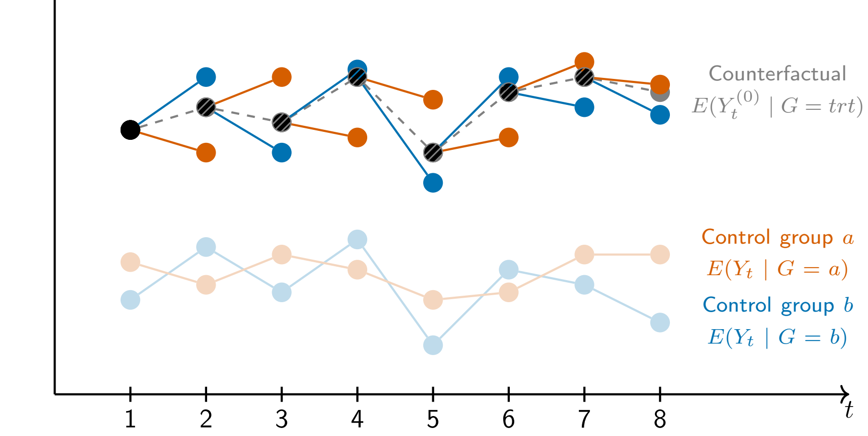
$\Delta_4(a) \leq \Delta_4^{(0)}(trt) \leq \Delta_4(b)$ (bound the change)



Bracketed Trends Assumption

Let $\Delta_t(g) = E(Y_t - Y_{t-1} | G = g)$, $g \in \{a, b\}$ (change in outcome)

$\min\{\Delta_t(a), \Delta_t(b)\} \leq \Delta_t^{(0)}(trt) \leq \max\{\Delta_t(a), \Delta_t(b)\}$ (bound the change)



Bracketed Trends Assumption

Bracketed trends: Let $\Delta_t(g) = E(Y_t - Y_{t-1}|G = g)$. For every $t \geq 2$,

$$\min\{\Delta_t(a), \Delta_t(b)\} \leq \Delta_t^{(0)}(trt) \leq \max\{\Delta_t(a), \Delta_t(b)\}$$

Connections to existing models:

- Parallel trends: $\Delta_t(a) = \Delta_t(b) = \Delta_t^{(0)}(trt)$

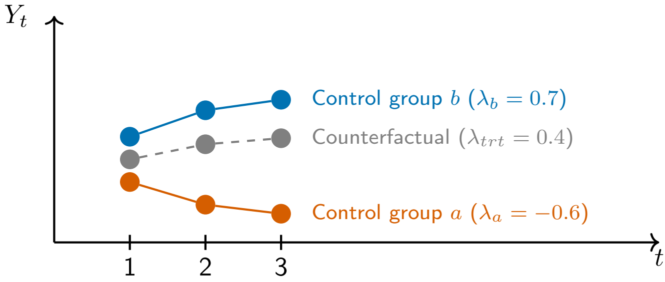
Bracketed Trends Assumption

Bracketed trends: Let $\Delta_t(g) = E(Y_t - Y_{t-1} | G = g)$. For every $t \geq 2$,

$$\min\{\Delta_t(a), \Delta_t(b)\} \leq \Delta_t^{(0)}(trt) \leq \max\{\Delta_t(a), \Delta_t(b)\}$$

Connections to existing models:

- Parallel trends: $\Delta_t(a) = \Delta_t(b) = \Delta_t^{(0)}(trt)$
- Interactive fixed-effect model $Y_{gt}^{(0)} = \theta_t + \eta_g + \lambda_g F_t + \varepsilon_{gt}$ (Bai, 2009)



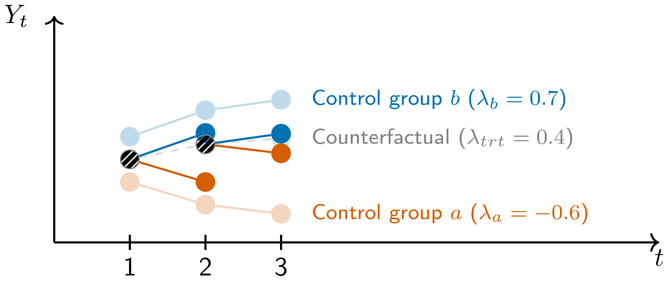
Bracketed Trends Assumption

Bracketed trends: Let $\Delta_t(g) = E(Y_t - Y_{t-1} | G = g)$. For every $t \geq 2$,

$$\min\{\Delta_t(a), \Delta_t(b)\} \leq \Delta_t^{(0)}(trt) \leq \max\{\Delta_t(a), \Delta_t(b)\}$$

Connections to existing models:

- Parallel trends: $\Delta_t(a) = \Delta_t(b) = \Delta_t^{(0)}(trt)$
- Interactive fixed-effect model $Y_{gt}^{(0)} = \theta_t + \eta_g + \lambda_g F_t + \varepsilon_{gt}$ (Bai, 2009)



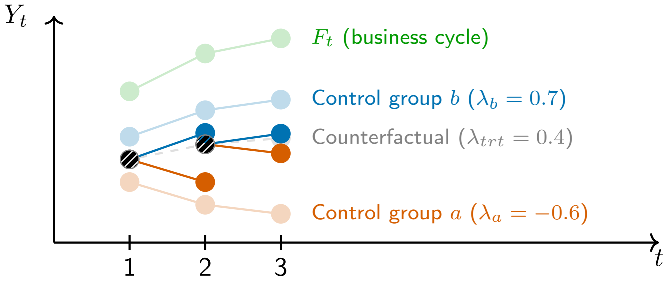
Bracketed Trends Assumption

Bracketed trends: Let $\Delta_t(g) = E(Y_t - Y_{t-1} | G = g)$. For every $t \geq 2$,

$$\min\{\Delta_t(a), \Delta_t(b)\} \leq \Delta_t^{(0)}(trt) \leq \max\{\Delta_t(a), \Delta_t(b)\}$$

Connections to existing models:

- Parallel trends: $\Delta_t(a) = \Delta_t(b) = \Delta_t^{(0)}(trt)$
- Interactive fixed-effect model $Y_{gt}^{(0)} = \theta_t + \eta_g + \lambda_g F_t + \varepsilon_{gt}$ (Bai, 2009)



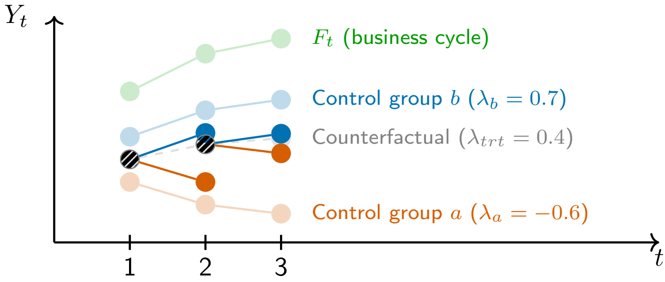
Bracketed Trends Assumption

Bracketed trends: Let $\Delta_t(g) = E(Y_t - Y_{t-1}|G = g)$. For every $t \geq 2$,

$$\min\{\Delta_t(a), \Delta_t(b)\} \leq \Delta_t^{(0)}(trt) \leq \max\{\Delta_t(a), \Delta_t(b)\}$$

Connections to existing models:

- Parallel trends: $\Delta_t(a) = \Delta_t(b) = \Delta_t^{(0)}(trt)$
- Interactive fixed-effect model $Y_{gt}^{(0)} = \theta_t + \eta_g + \lambda_g F_t + \varepsilon_{gt}$ (Bai, 2009)
- Parallel growth (Mora and Reggio, 2012)
- Changes-in-Changes (Athey and Imbens, 2006)



Control Groups Selection

- Eg 1: Effect of minimum wage on employment (Derenoncourt and Montialoux, 2021; Berman and Pfleeger, 1997)
- Macro-economic shocks can have different effects on sectors
- Some industries are cyclical while some are countercyclical

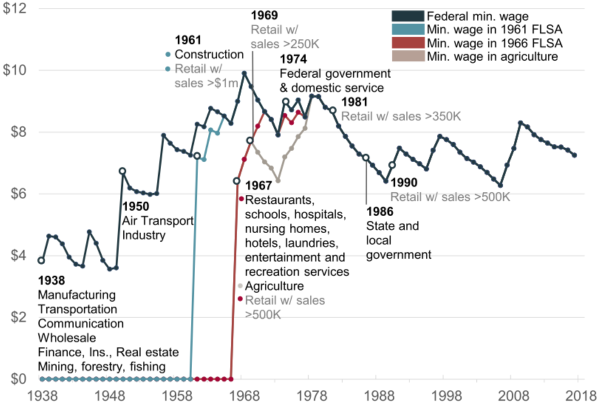


FIGURE III

Expansions in Minimum Wage Coverage, and Real Values of the Minimum Wage, 1938–2018 ($2017)

Control Groups Selection

- Eg 1: Effect of minimum wage on employment (Derenoncourt and Montialoux, 2021; Berman and Pfleeger, 1997)
- Macro-economic shocks can have different effects on sectors
- Some industries are cyclical while some are countercyclical

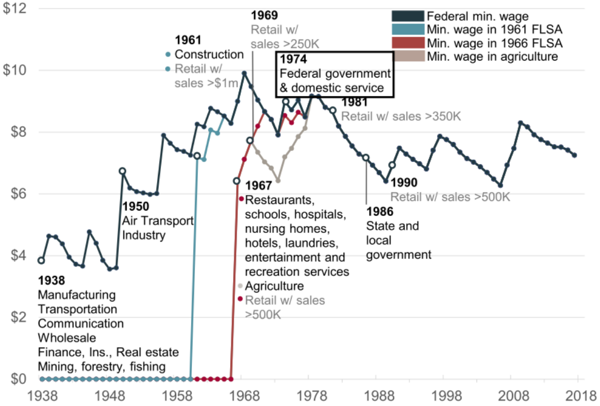


FIGURE III

Expansions in Minimum Wage Coverage, and Real Values of the Minimum Wage, 1938–2018 ($2017)

Control Groups Selection

- Eg 1: Effect of minimum wage on employment (Derenoncourt and Montialoux, 2021; Berman and Pfleeger, 1997)
- Macro-economic shocks can have different effects on sectors
- Some industries are cyclical while some are countercyclical

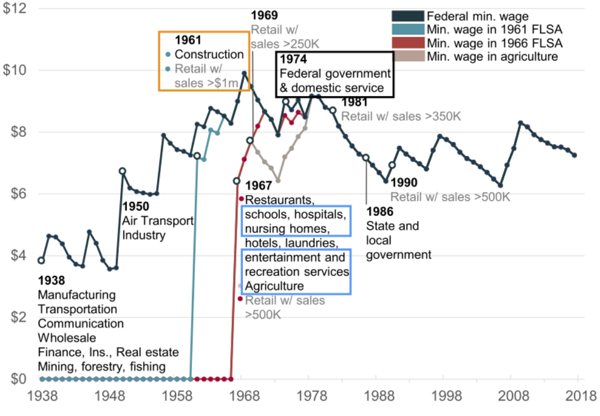
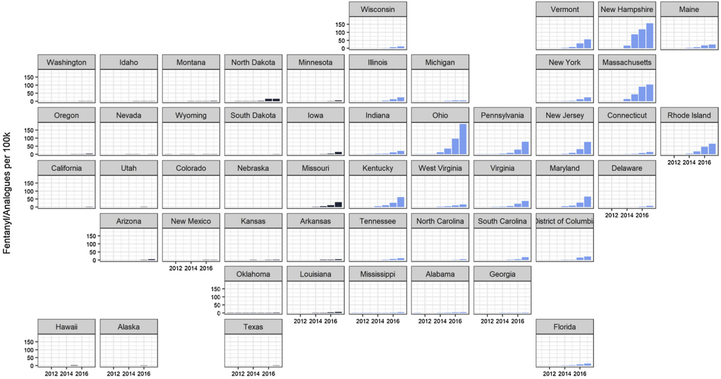


FIGURE III

Expansions in Minimum Wage Coverage, and Real Values of the Minimum Wage, 1938–2018 ($2017)

Control Groups Selection

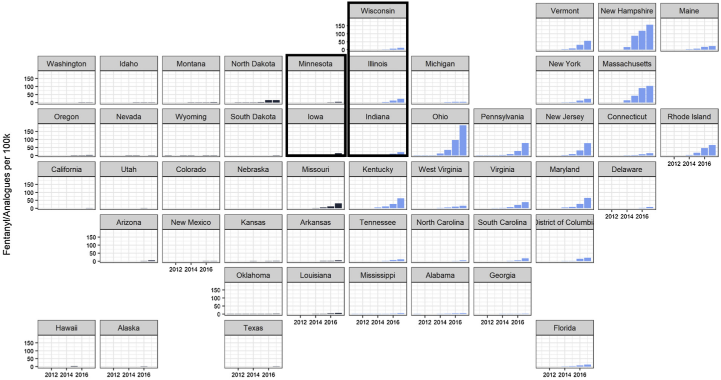
- Eg 2: Effect of income support policies on drug overdose deaths
- US fentanyl supply can have different effects on states (Zoorob, 2019)
- Some states were hit harder while some were relatively insulated from fentanyl



Source: National Forensic Laboratory Information System (NFLIS)

Control Groups Selection

- Eg 2: Effect of income support policies on drug overdose deaths
- US fentanyl supply can have different effects on states (Zoorob, 2019)
- Some states were hit harder while some were relatively insulated from fentanyl



Source: National Forensic Laboratory Information System (NFLIS)

Control Groups Selection

- Eg 2: Effect of income support policies on drug overdose deaths
- US fentanyl supply can have different effects on states (Zoorob, 2019)
- Some states were hit harder while some were relatively insulated from fentanyl

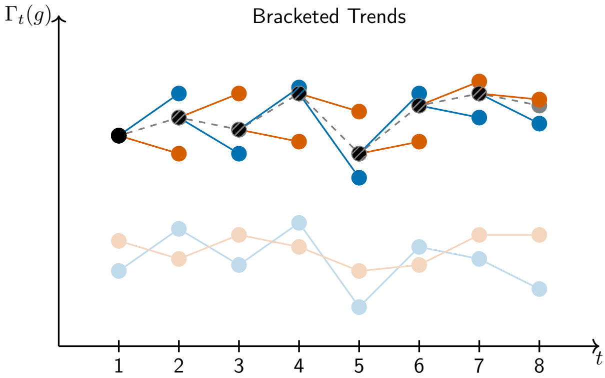
Source: National Forensic Laboratory Information System (NFLIS)

A Negative Correlation Strategy

Lemma (Bracketed Trends $\Leftrightarrow$ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

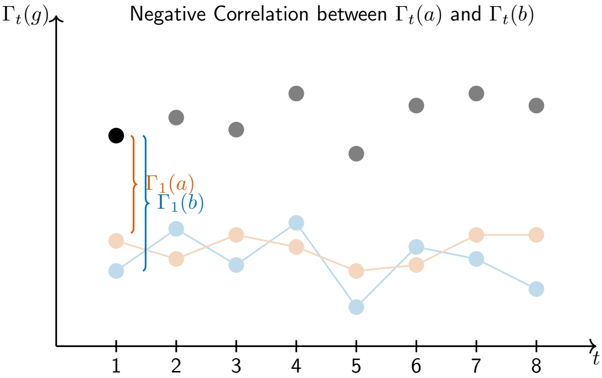


A Negative Correlation Strategy

Lemma (Bracketed Trends ⇔ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

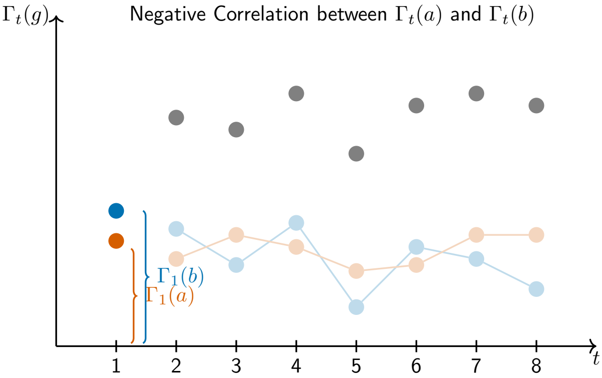


A Negative Correlation Strategy

Lemma (Bracketed Trends ⇔ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

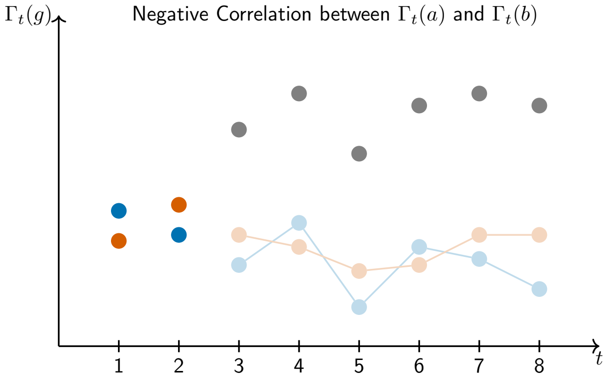


A Negative Correlation Strategy

Lemma (Bracketed Trends $\Leftrightarrow$ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

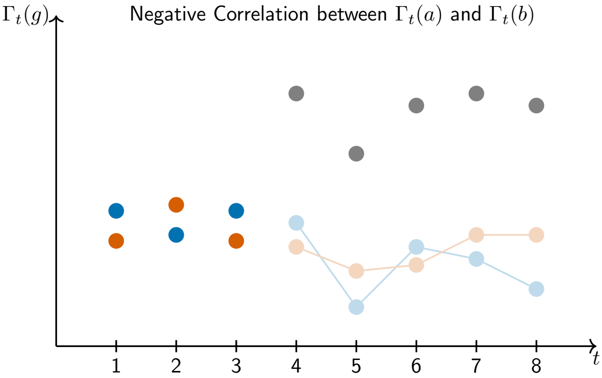


A Negative Correlation Strategy

Lemma (Bracketed Trends $\Leftrightarrow$ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

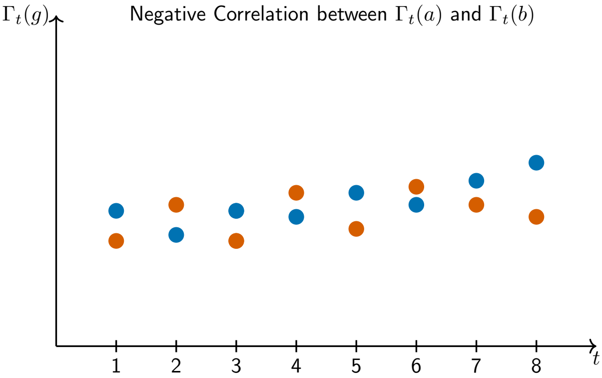


A Negative Correlation Strategy

Lemma (Bracketed Trends ⇔ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

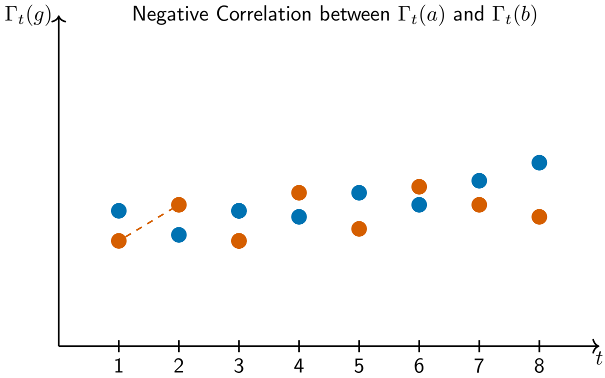


A Negative Correlation Strategy

Lemma (Bracketed Trends ⇔ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

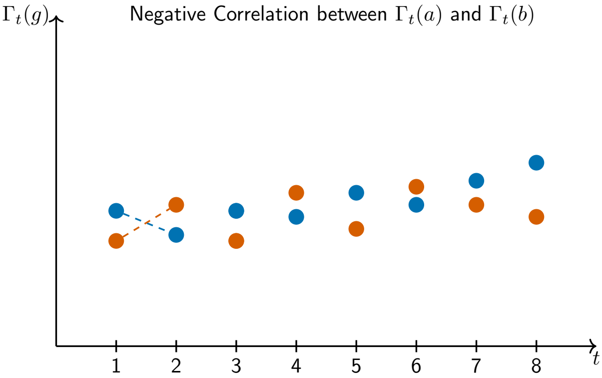


A Negative Correlation Strategy

Lemma (Bracketed Trends ⇔ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

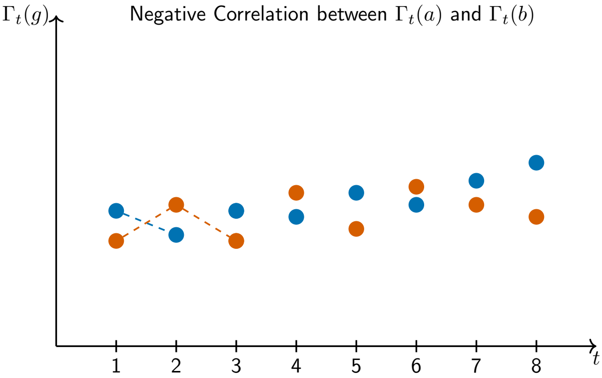


A Negative Correlation Strategy

Lemma (Bracketed Trends $\Leftrightarrow$ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

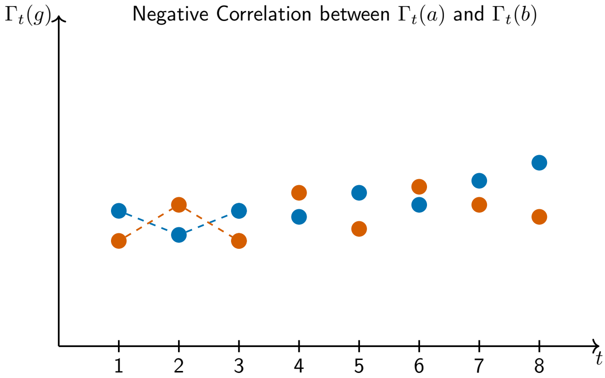


A Negative Correlation Strategy

Lemma (Bracketed Trends $\Leftrightarrow$ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

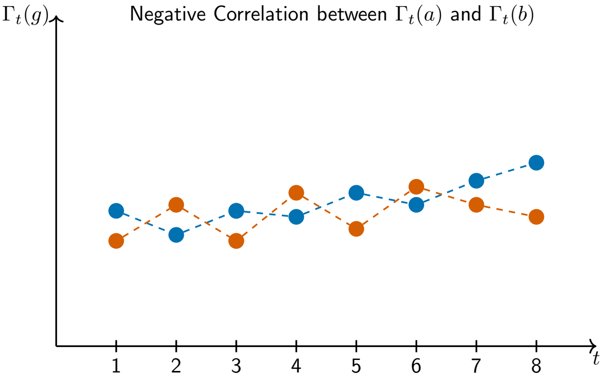


A Negative Correlation Strategy

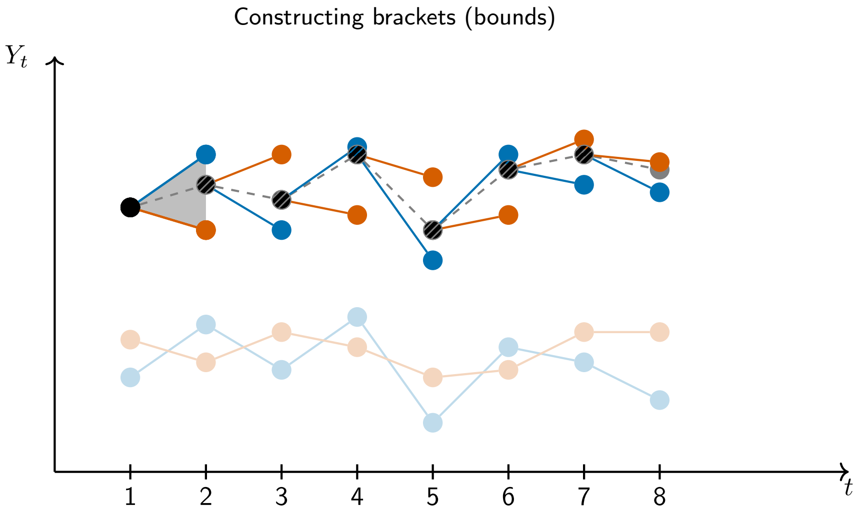
Lemma (Bracketed Trends ⇔ Negative Correlation)

Let $\Gamma_t(g) = E(Y_t^{(0)}|G = trt) - E(Y_t|G = g)$, $g \in \{a, b\}$. Bracketed trends is equivalent with

$$\{\Gamma_t(a) - \Gamma_{t-1}(a)\}\{\Gamma_t(b) - \Gamma_{t-1}(b)\} \leq 0 \quad \text{(negative correlation)}$$

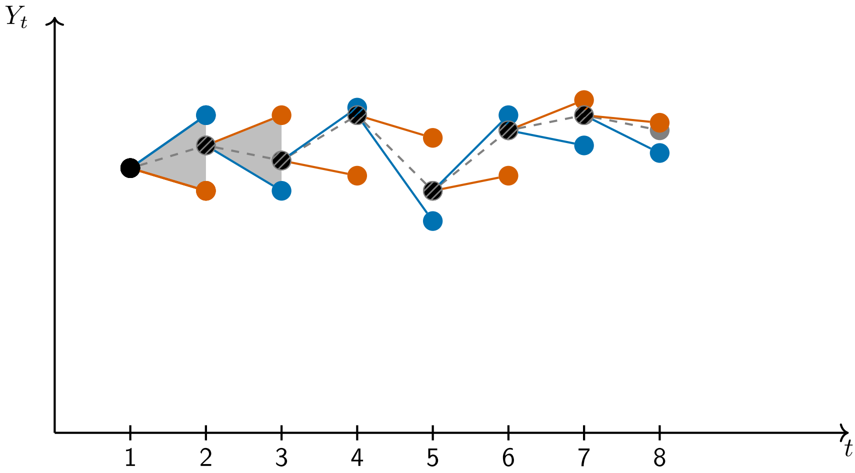


Bracket the Counterfactual Outcome Path for the Treated

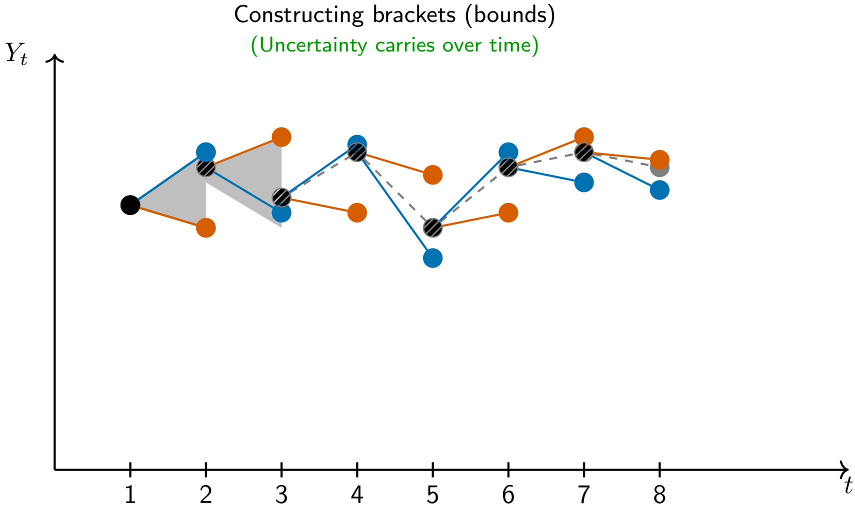


Bracket the Counterfactual Outcome Path for the Treated

Constructing brackets (bounds)



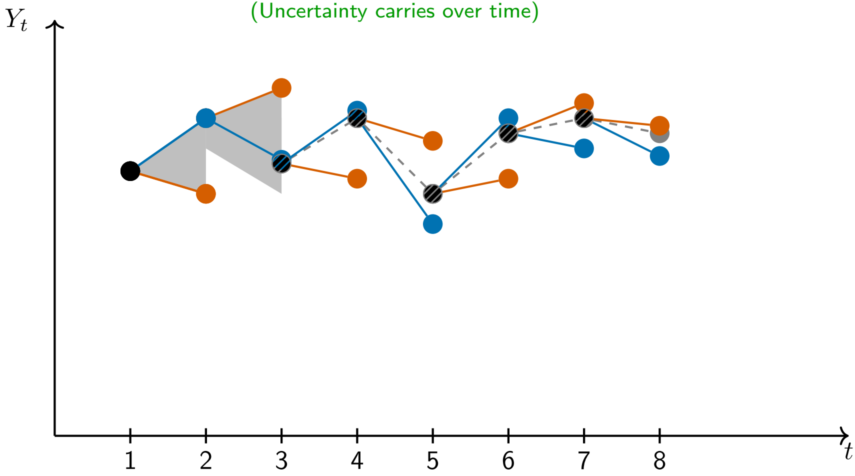
Bracket the Counterfactual Outcome Path for the Treated



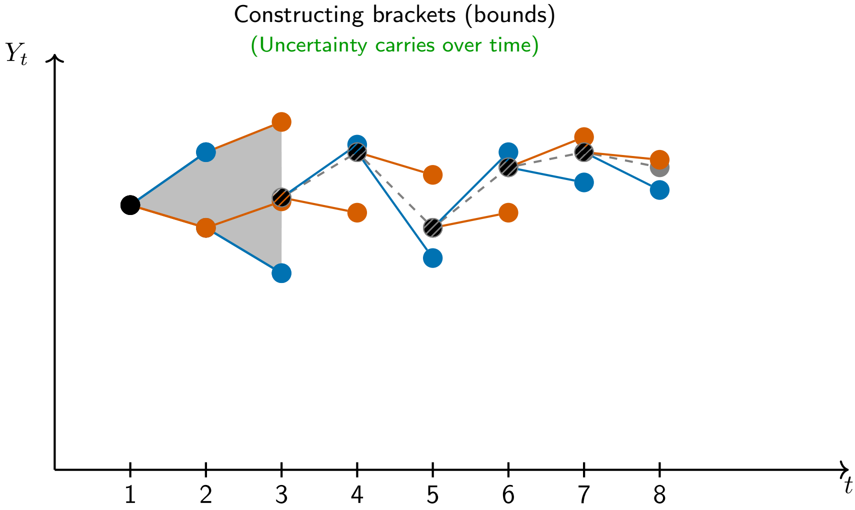
Bracket the Counterfactual Outcome Path for the Treated

Constructing brackets (bounds)

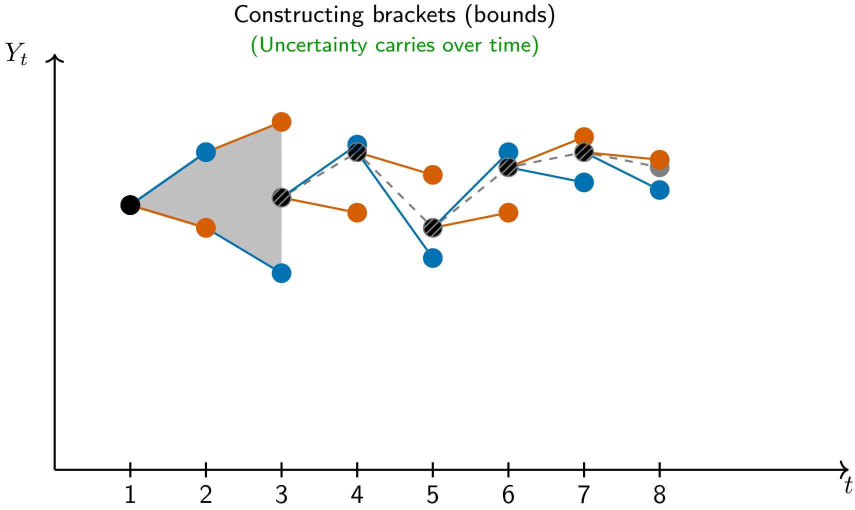
(Uncertainty carries over time)



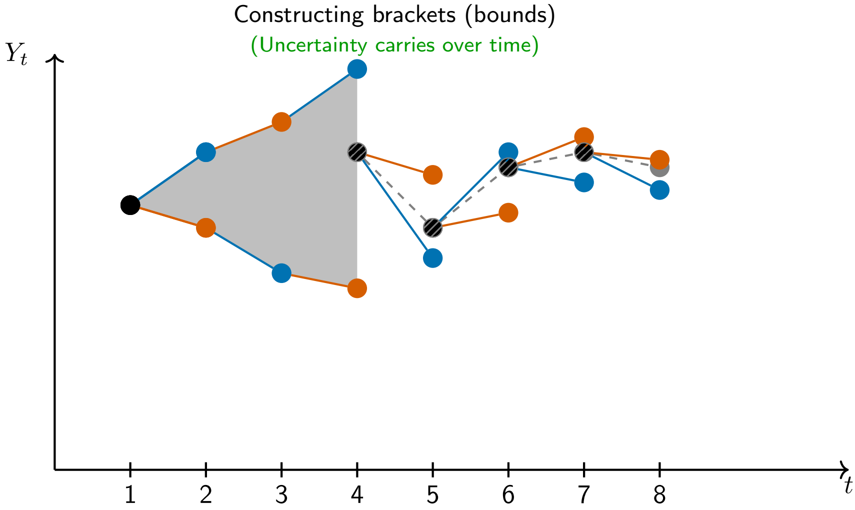
Bracket the Counterfactual Outcome Path for the Treated



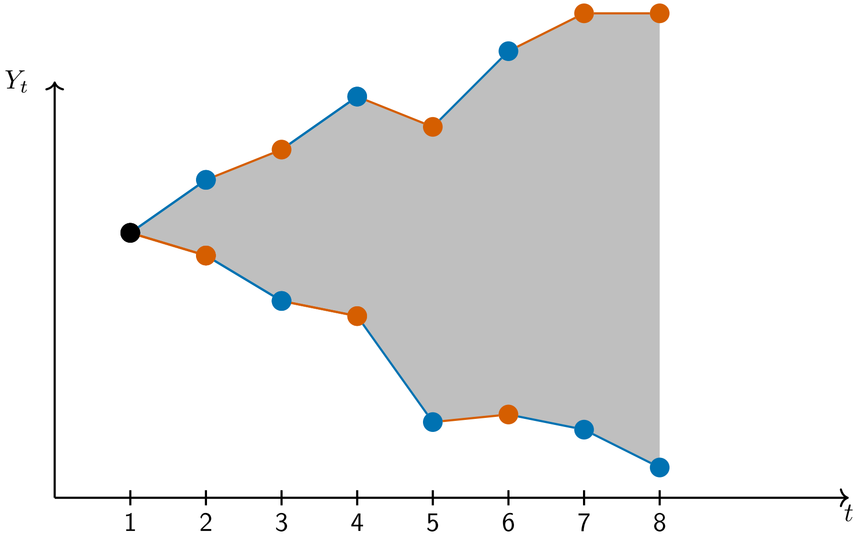
Bracket the Counterfactual Outcome Path for the Treated



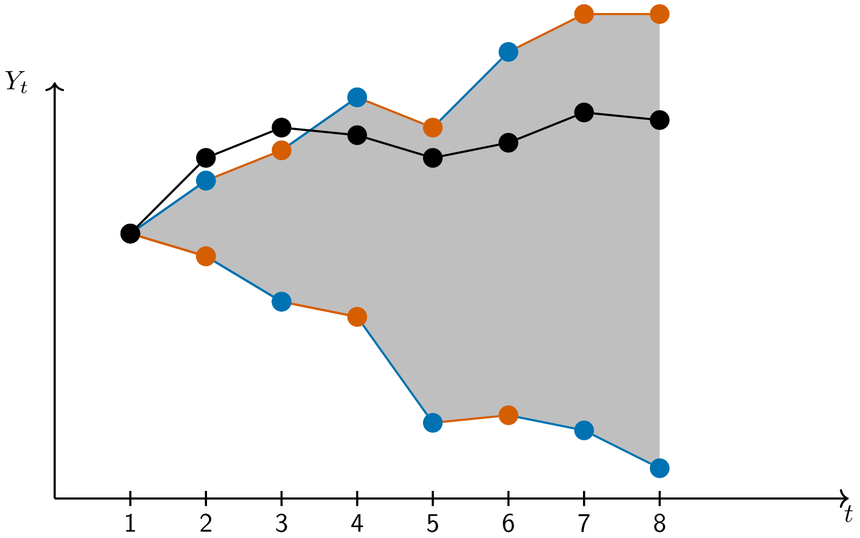
Bracket the Counterfactual Outcome Path for the Treated



Bracket the Counterfactual Outcome Path for the Treated

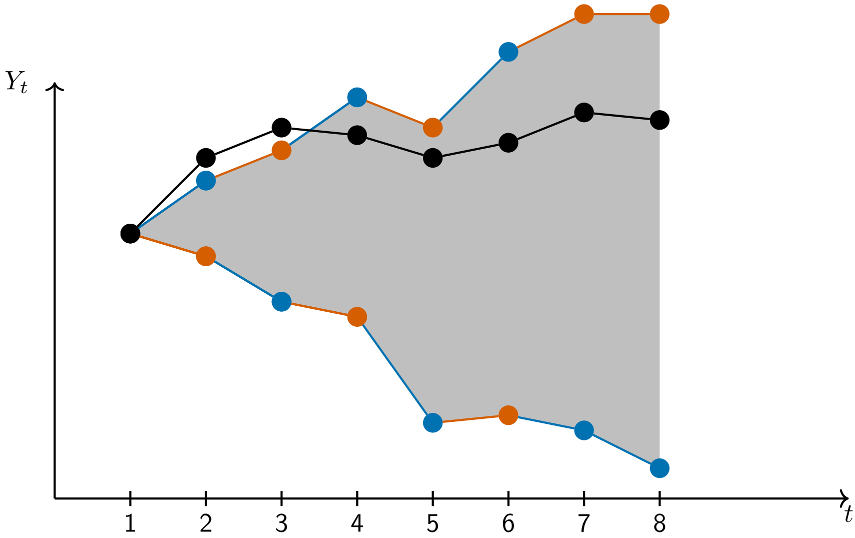


Bracket the Treatment Effect



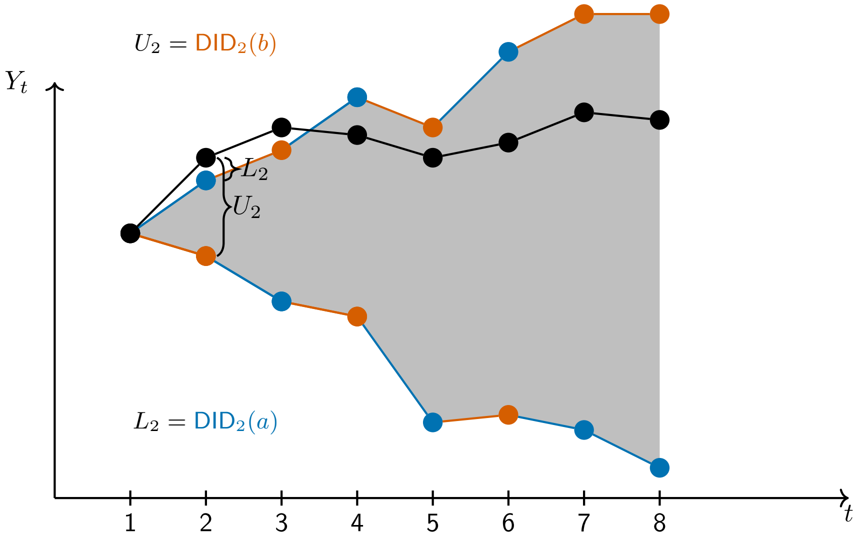
Bracket the Treatment Effect

Let $\text{DID}_s(g) = E(Y_s - Y_{s-1} \mid G = trt) - E(Y_s - Y_{s-1} \mid G = g)$, $g \in \{a, b\}$.



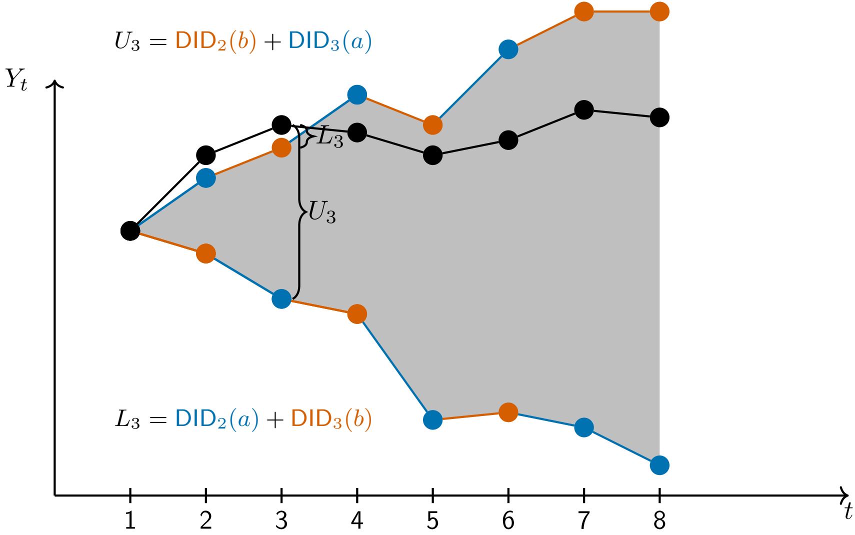
Bracket the Treatment Effect

Let $\mathsf{DID}_s(g) = E(Y_s - Y_{s-1} \mid G = trt) - E(Y_s - Y_{s-1} \mid G = g)$, $g \in \{a, b\}$.



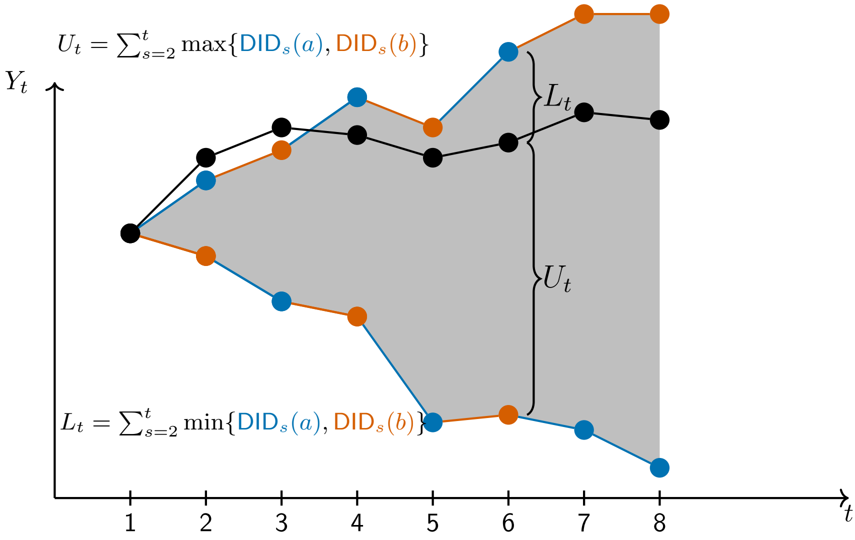
Bracket the Treatment Effect

Let $\mathsf{DID}_s(g) = E(Y_s - Y_{s-1} \mid G = trt) - E(Y_s - Y_{s-1} \mid G = g)$, $g \in \{a, b\}$.



Bracket the Treatment Effect

Let $\text{DID}_s(g) = E(Y_s - Y_{s-1} \mid G = trt) - E(Y_s - Y_{s-1} \mid G = g)$, $g \in \{a, b\}$.



Partial Identification of ATT_t

Theorem

Under the bracketed trends, $ATT_t, t \geq 2$ can be partially identified via

$$\sum_{s=2}^{t} \min\{\mathrm{DID}_s(a), \mathrm{DID}_s(b)\} \leq ATT_t \leq \sum_{s=2}^{t} \max\{\mathrm{DID}_s(a), \mathrm{DID}_s(b)\}$$

where $\mathrm{DID}_s(g) = E(Y_s - Y_{s-1} \mid G = trt) - E(Y_s - Y_{s-1} \mid G = g)$, $g \in \{a, b\}$.

Partial Identification of ATT_t

Theorem

Under the bracketed trends, $ATT_t, t \geq 2$ can be partially identified via

$$\sum_{s=2}^{t} \min\{\mathrm{DID}_s(a), \mathrm{DID}_s(b)\} \leq ATT_t \leq \sum_{s=2}^{t} \max\{\mathrm{DID}_s(a), \mathrm{DID}_s(b)\}$$

where $\mathrm{DID}_s(g) = E(Y_s - Y_{s-1} \mid G = trt) - E(Y_s - Y_{s-1} \mid G = g)$*,* $g \in \{a, b\}$*.*

- Parallel trends $\Rightarrow$ $\mathrm{DID}_s(a) = \mathrm{DID}_s(b)$ $\Rightarrow$ point identification.

Partial Identification of ATT_t

Theorem

Under the bracketed trends, $ATT_t, t \geq 2$ can be partially identified via

$$\sum_{s=2}^{t} \min\{\mathrm{DID}_s(a), \mathrm{DID}_s(b)\} \leq ATT_t \leq \sum_{s=2}^{t} \max\{\mathrm{DID}_s(a), \mathrm{DID}_s(b)\}$$

where $\mathrm{DID}_s(g) = E(Y_s - Y_{s-1} \mid G = trt) - E(Y_s - Y_{s-1} \mid G = g)$, $g \in \{a, b\}$.

- Parallel trends $\Rightarrow$ $\mathrm{DID}_s(a) = \mathrm{DID}_s(b)$ $\Rightarrow$ point identification.
- The identified set can be equivalently written as

$$\min_{g_s \in \{a,b\}} \left\{ \sum_{s=2}^{t} \mathrm{DID}_s(g_s) \right\} \leq \mathrm{ATT}_t \leq \max_{g_s \in \{a,b\}} \left\{ \sum_{s=2}^{t} \mathrm{DID}_s(g_s) \right\}$$

For example, when $t = 3$, the lower end $\sum_{s=2}^{3} \min\{\mathrm{DID}_s(a), \mathrm{DID}_s(b)\} =$ $\min\{\mathrm{DID}_2(a) + \mathrm{DID}_3(a), \mathrm{DID}_2(a) + \mathrm{DID}_3(b), \mathrm{DID}_2(b) + \mathrm{DID}_3(a), \mathrm{DID}_2(b) + \mathrm{DID}_3(b)\}$, which is the minimum of 4 parameters.

Partial Identification of ATT_t

Theorem

Under the bracketed trends, $ATT_t, t \geq 2$ can be partially identified via

$$\sum_{s=2}^{t} \min\{\text{DID}_s(a), \text{DID}_s(b)\} \leq \mathsf{ATT}_t \leq \sum_{s=2}^{t} \max\{\text{DID}_s(a), \text{DID}_s(b)\}$$

where $\text{DID}_s(g) = E(Y_s - Y_{s-1} \mid G = trt) - E(Y_s - Y_{s-1} \mid G = g)$, $g \in \{a, b\}$.

- Parallel trends $\Rightarrow$ $\mathsf{DID}_s(a) = \mathsf{DID}_s(b)$ $\Rightarrow$ point identification.
- The identified set can be equivalently written as

$$\min_{g_s \in \{a,b\}} \left\{ \sum_{s=2}^{t} \mathsf{DID}_s(g_s) \right\} \leq \mathsf{ATT}_t \leq \max_{g_s \in \{a,b\}} \left\{ \sum_{s=2}^{t} \mathsf{DID}_s(g_s) \right\}$$

  For example, when $t = 3$, the lower end $\sum_{s=2}^{3} \min\{\mathsf{DID}_s(a), \mathsf{DID}_s(b)\} = \min\{\mathsf{DID}_2(a) + \mathsf{DID}_3(a), \mathsf{DID}_2(a) + \mathsf{DID}_3(b), \mathsf{DID}_2(b) + \mathsf{DID}_3(a), \mathsf{DID}_2(b) + \mathsf{DID}_3(b)\}$, which is the minimum of 4 parameters.

- Define $\{\theta_j\}_t = \{\sum_{s=2}^{t} \mathsf{DID}_s(g_s) : g_s \in \{a, b\}\}$, which contains $2^{t-1} < \infty$ elements. For notation simplicity, the subscript t will be omitted.

Inference for union bounds

- In general, the goal is to construct valid CIs[2] for the identified set $[\min_j \theta_j, \max_j \theta_j]$ (union bounds) and the parameter of interest $\mathsf{ATT}_t \in [\min_j \theta_j, \max_j \theta_j]$.

[2]Differences between CIs for the identified set and for the parameter of interest within that set have been well-addressed in the literature (Imbens and Manski, 2004; Stoye, 2009)

[3]Hirano and Porter (2012) demonstrated that unbiased estimators are nonexistent.

[4]Shao, 1994, Romano and Shaikh, 2008.

Inference for union bounds

- In general, the goal is to construct valid CIs[2] for the identified set $[\min_j \theta_j, \max_j \theta_j]$ (union bounds) and the parameter of interest $\mathsf{ATT}_t \in [\min_j \theta_j, \max_j \theta_j]$.
- Challenges: **min and max operators.** (i) simple sample analogous are biased[3]; (ii) canonical bootstrap is inconsistent[4].

[2] Differences between CIs for the identified set and for the parameter of interest within that set have been well-addressed in the literature (Imbens and Manski, 2004; Stoye, 2009)

[3] Hirano and Porter (2012) demonstrated that unbiased estimators are nonexistent.

[4] Shao, 1994, Romano and Shaikh, 2008.

Inference for union bounds

- In general, the goal is to construct valid CIs[2] for the identified set $[\min_j \theta_j, \max_j \theta_j]$ (union bounds) and the parameter of interest $\text{ATT}_t \in [\min_j \theta_j, \max_j \theta_j]$.
- Challenges: **min and max operators.** (i) simple sample analogous are biased[3]; (ii) canonical bootstrap is inconsistent[4].
- Existing approaches can be overly conservative:
  1. Intersection-union method (Berger and Hsu, 1996)

$$\big[\min_j(\hat{\theta}_j - z_{1-\alpha/2}\hat{\sigma}_j), \max_j(\hat{\theta}_j + z_{1-\alpha/2}\hat{\sigma}_j)\big]$$

  2. Percentile bootstrap $\big[Q_{\alpha/2}\big(\{\min_j \hat{\theta}_j^*\}\big), Q_{1-\alpha/2}\big(\{\max_j \hat{\theta}_j^*\}\big)\big]$

---

[2] Differences between CIs for the identified set and for the parameter of interest within that set have been well-addressed in the literature (Imbens and Manski, 2004; Stoye, 2009)

[3] Hirano and Porter (2012) demonstrated that unbiased estimators are nonexistent.

[4] Shao, 1994, Romano and Shaikh, 2008.

Inference for union bounds

- In general, the goal is to construct valid CIs[2] for the identified set $[\min_j \theta_j, \max_j \theta_j]$ (union bounds) and the parameter of interest $\text{ATT}_t \in [\min_j \theta_j, \max_j \theta_j]$.
- Challenges: **min and max operators.** (i) simple sample analogous are biased[3]; (ii) canonical bootstrap is inconsistent[4].
- Existing approaches can be overly conservative:
  1. Intersection-union method (Berger and Hsu, 1996)

     $$\big[\min_j(\hat{\theta}_j - z_{1-\alpha/2}\hat{\sigma}_j), \max_j(\hat{\theta}_j + z_{1-\alpha/2}\hat{\sigma}_j)\big]$$

  2. Percentile bootstrap $\big[Q_{\alpha/2}\big(\{\min_j \hat{\theta}^*_j\}\big), Q_{1-\alpha/2}\big(\{\max_j \hat{\theta}^*_j\}\big)\big]$
- **Our proposal**:

$$\big[2\min_j \hat{\theta}_j - Q_{1-\alpha/2}\big(\{\min_j \hat{\theta}^*_j\}\big), 2\max_j \hat{\theta}_j - Q_{\alpha/2}\big(\{\max_j \hat{\theta}^*_j\}\big)\big]$$

---

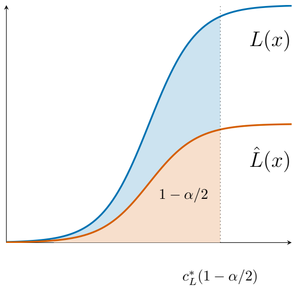
[2] Differences between CIs for the identified set and for the parameter of interest within that set have been well-addressed in the literature (Imbens and Manski, 2004; Stoye, 2009)

[3] Hirano and Porter (2012) demonstrated that unbiased estimators are nonexistent.

[4] Shao, 1994, Romano and Shaikh, 2008.

Intuition

Consider the lower end. Let $L(x) = P\{\sqrt{N}(\min_j \hat{\theta}_j - \min_j \theta_j) \leq x\}$ be the true distribution, $\hat{L}(x) = P\{\sqrt{N}(\min_j \hat{\theta}^*_j - \min_j \hat{\theta}_j) \leq x\}$ be the bootstrap estimate. When $\{\theta_j\}$ does not have a unique minimum, $\hat{L}(x)$ is not consistently to $L(x)$.



By construction, $L(c^*_L(1 - \alpha/2)) \geq 1 - \alpha/2$, and equivalently,

$$P(\underbrace{\min_j \hat{\theta}_j - N^{-1/2} c^*_L(1 - \alpha/2)} \leq \min_j \theta_j) \geq 1 - \alpha/2$$

$$C_L = \min_j \hat{\theta}_j - Q_{1-\alpha/2}(\{\min_j \hat{\theta}^*_j - \min_j \hat{\theta}_j\}) = 2 \min_j \hat{\theta}_j - Q_{1-\alpha/2}(\{\min_j \hat{\theta}^*_j\})$$

Some Remarks

- For inference of $\text{ATT}_t \in [\min_j \theta_j, \max_j \theta_j]$, we can set $\hat{p} = 1 - \alpha$ when the bounds are "wide enough" and set $\hat{p} = 1 - \alpha/2$ otherwise.
- When parallel trends holds, our CI is similar to that from standard DID. When bracketed trends holds but parallel trends is violated, our CI is valid while that from standard DID is not.
- The theorem also provides half-median-unbiased estimators $\min_j \hat{\theta}_j - N^{-1/2} c_L^*(1/2)$ and $\max_j \hat{\theta}_j - N^{-1/2} c_U^*(1/2)$ for $\min_j \theta_j$ and $\max_j \theta_j$.

Simulations

Case I: parallel trends:
$E[Y_1^{(0)}|G=trt]=3, E[Y_1^{(0)}|G=a]=10, E[Y_1^{(0)}|G=b]=4,$
$\Delta_t(trt)=\Delta_t(a)=\Delta_t(b)\equiv\Delta_t$ for every t, where $\Delta_2=1, \Delta_3=-2, \Delta_4=-1.$
Case II: partially parallel trends:
$E[Y_1^{(0)}|G=trt]=3, E[Y_1^{(0)}|G=a]=10, E[Y_1^{(0)}|G=b]=4,$
$\Delta_2(trt)=1, \Delta_3(trt)=-4, \Delta_4(trt)=1, \Delta_2(a)=1, \Delta_3(a)=-1, \Delta_4(a)=1, \Delta_2(b)=2, \Delta_3(b)=-4, \Delta_4(b)=1.$

Table: Simulations with 1000 runs, $N=1000$ and $B=300$. The truth are $\text{ATT}_2=2, \text{ATT}_3=3, \text{ATT}_4=1.$

	$\hat\theta_{\min}$	$\hat\theta_{\max}$	Modified bootstrap $\hat\theta_{\min}^{\text{med}}$	$\hat\theta_{\max}^{\text{med}}$	CI (Set)		CI (ATT_t)		Intersec.-Union CI (Set)		Perc. Boot. CI (Set)	
	Mean	Mean	Mean	Mean	Length	CP	Length	CP	Length	CP	Length	CP
Case I												
$t=2$	1.952	2.047	1.970	2.030	0.483	96.7	0.478	96.7	0.553	99.0	0.581	99.3
$t=3$	2.905	3.098	2.941	3.063	0.583	97.7	0.575	97.7	0.730	99.8	0.771	99.9
$t=4$	0.859	1.144	0.913	1.090	0.672	98.4	0.661	98.4	0.893	100.0	0.955	100.0
Case II												
$t=2$	1.003	1.997	1.003	1.997	1.455	97.9	1.404	96.8	1.443	97.7	1.455	97.8
$t=3$	-0.994	2.997	-0.994	2.998	4.562	98.1	4.472	96.1	4.547	97.8	4.563	98.1
$t=4$	-3.041	1.043	-3.021	1.025	4.633	98.5	4.542	96.7	4.693	98.7	4.728	99.3

Simulations

Case I: parallel trends:
$E[Y_1^{(0)}|G = trt] = 3, E[Y_1^{(0)}|G = a] = 10, E[Y_1^{(0)}|G = b] = 4,$
$\Delta_t(trt) = \Delta_t(a) = \Delta_t(b) \equiv \Delta_t$ for every t, where $\Delta_2 = 1, \Delta_3 = -2, \Delta_4 = -1.$
Case II: partially parallel trends:
$E[Y_1^{(0)}|G = trt] = 3, E[Y_1^{(0)}|G = a] = 10, E[Y_1^{(0)}|G = b] = 4,$
$\Delta_2(trt) = 1, \Delta_3(trt) = -4, \Delta_4(trt) = 1, \Delta_2(a) = 1, \Delta_3(a) = -1, \Delta_4(a) = 1, \Delta_2(b) = 2, \Delta_3(b) = -4, \Delta_4(b) = 1.$

Table: Simulations with 1000 runs, $N = 1000$ and $B = 300$. The truth are $\text{ATT}_2 = 2, \text{ATT}_3 = 3, \text{ATT}_4 = 1$.

			Modified bootstrap						Intersec.-Union		Perc. Boot.	
	$\hat{\theta}_{\min}$	$\hat{\theta}_{\max}$	$\hat{\theta}^{\text{med}}_{\min}$	$\hat{\theta}^{\text{med}}_{\max}$	CI (Set)		CI (ATT_t)		CI (Set)		CI (Set)	
	Mean	Mean	Mean	Mean	Length	CP	Length	CP	Length	CP	Length	CP
Case I												
$t = 2$	1.952	2.047	1.970	2.030	0.483	96.7	0.478	96.7	0.553	99.0	0.581	99.3
$t = 3$	2.905	3.098	2.941	3.063	0.583	97.7	0.575	97.7	0.730	99.8	0.771	99.9
$t = 4$	0.859	1.144	0.913	1.090	0.672	98.4	0.661	98.4	0.893	100.0	0.955	100.0
Case II												
$t = 2$	1.003	1.997	1.003	1.997	1.455	97.9	1.404	96.8	1.443	97.7	1.455	97.8
$t = 3$	-0.994	2.997	-0.994	2.998	4.562	98.1	4.472	96.1	4.547	97.8	4.563	98.1
$t = 4$	-3.041	1.043	-3.021	1.025	4.633	98.5	4.542	96.7	4.693	98.7	4.728	99.3

Simulations

Case I: parallel trends:
$E[Y_1^{(0)}|G = trt] = 3, E[Y_1^{(0)}|G = a] = 10, E[Y_1^{(0)}|G = b] = 4,$
$\Delta_t(trt) = \Delta_t(a) = \Delta_t(b) \equiv \Delta_t$ for every t, where $\Delta_2 = 1, \Delta_3 = -2, \Delta_4 = -1.$
Case II: partially parallel trends:
$E[Y_1^{(0)}|G = trt] = 3, E[Y_1^{(0)}|G = a] = 10, E[Y_1^{(0)}|G = b] = 4,$
$\Delta_2(trt) = 1, \Delta_3(trt) = -4, \Delta_4(trt) = 1, \Delta_2(a) = 1, \Delta_3(a) = -1, \Delta_4(a) = 1, \Delta_2(b) = 2, \Delta_3(b) = -4, \Delta_4(b) = 1.$

Table: Simulations with 1000 runs, $N = 1000$ and $B = 300$. The truth are $\text{ATT}_2 = 2, \text{ATT}_3 = 3, \text{ATT}_4 = 1$.

			Modified bootstrap						Intersec.-Union		Perc. Boot.	
	$\hat{\theta}_{\min}$	$\hat{\theta}_{\max}$	$\hat{\theta}^{\text{med}}_{\min}$	$\hat{\theta}^{\text{med}}_{\max}$	CI (Set)		CI (ATT_t)		CI (Set)		CI (Set)	
	Mean	Mean	Mean	Mean	Length	CP	Length	CP	Length	CP	Length	CP
Case I												
$t = 2$	1.952	2.047	1.970	2.030	0.483	96.7	0.478	96.7	0.553	99.0	0.581	99.3
$t = 3$	2.905	3.098	2.941	3.063	0.583	97.7	0.575	97.7	0.730	99.8	0.771	99.9
$t = 4$	0.859	1.144	0.913	1.090	0.672	98.4	0.661	98.4	0.893	100.0	0.955	100.0
Case II												
$t = 2$	1.003	1.997	1.003	1.997	1.455	97.9	1.404	96.8	1.443	97.7	1.455	97.8
$t = 3$	-0.994	2.997	-0.994	2.998	4.562	98.1	4.472	96.1	4.547	97.8	4.563	98.1
$t = 4$	-3.041	1.043	-3.021	1.025	4.633	98.5	4.542	96.7	4.693	98.7	4.728	99.3

Simulations

Case I: parallel trends:
$E[Y_1^{(0)}|G = trt] = 3, E[Y_1^{(0)}|G = a] = 10, E[Y_1^{(0)}|G = b] = 4,$
$\Delta_t(trt) = \Delta_t(a) = \Delta_t(b) \equiv \Delta_t$ for every t, where $\Delta_2 = 1, \Delta_3 = -2, \Delta_4 = -1.$
Case II: partially parallel trends:
$E[Y_1^{(0)}|G = trt] = 3, E[Y_1^{(0)}|G = a] = 10, E[Y_1^{(0)}|G = b] = 4,$
$\Delta_2(trt) = 1, \Delta_3(trt) = -4, \Delta_4(trt) = 1, \Delta_2(a) = 1, \Delta_3(a) = -1, \Delta_4(a) = 1, \Delta_2(b) = 2, \Delta_3(b) = -4, \Delta_4(b) = 1.$

Table: Simulations with 1000 runs, $N = 1000$ and $B = 300$. The truth are $\text{ATT}_2 = 2, \text{ATT}_3 = 3, \text{ATT}_4 = 1.$

			Modified bootstrap						Intersec.-Union		Perc. Boot.	
	$\hat{\theta}_{\min}$	$\hat{\theta}_{\max}$	$\hat{\theta}^{\text{med}}_{\min}$	$\hat{\theta}^{\text{med}}_{\max}$	CI (Set)		CI (ATT_t)		CI (Set)		CI (Set)	
	Mean	Mean	Mean	Mean	Length	CP	Length	CP	Length	CP	Length	CP
Case I												
$t = 2$	1.952	2.047	1.970	2.030	0.483	96.7	0.478	96.7	0.553	99.0	0.581	99.3
$t = 3$	2.905	3.098	2.941	3.063	0.583	97.7	0.575	97.7	0.730	99.8	0.771	99.9
$t = 4$	0.859	1.144	0.913	1.090	0.672	98.4	0.661	98.4	0.893	100.0	0.955	100.0
Case II												
$t = 2$	1.003	1.997	1.003	1.997	1.455	97.9	1.404	96.8	1.443	97.7	1.455	97.8
$t = 3$	-0.994	2.997	-0.994	2.998	4.562	98.1	4.472	96.1	4.547	97.8	4.563	98.1
$t = 4$	-3.041	1.043	-3.021	1.025	4.633	98.5	4.542	96.7	4.693	98.7	4.728	99.3

Summary

A framework for more flexible DID analysis ~~(parallel trends)~~

- Bracket the treatment effect in DID with multiple time points.
- Bootstrap inference with statistical guarantee.
- Application to the effect of minimum wage laws on employment and wages (not covered).
- Falsification test, and sensitivity analysis (not covered).

References

- Ye, T., Keele, L., Hasegawa, R., & Small, D. S. (2023). A negative correlation strategy for bracketing in difference-in-differences. *Journal of the American Statistical Association*.
- R package DIDBracket is available on GitHub (`https://github.com/tye27/DIDBracket`)

Thank You!

Questions?

tingye1@uw.edu

Statistical Properties

Let $L(x) = P\{\sqrt{N}(\min_j \hat{\theta}_j - \min_j \theta_j) \leq x\}$ be the true distribution, $\hat{L}(x) = P\{\sqrt{N}(\min_j \hat{\theta}^*_j - \min_j \hat{\theta}_j) \leq x\}$ be the bootstrap estimate. Define $R(x)$ and $\hat{R}(x)$ with $\min$ replaced by $\max$.

Theorem

Suppose $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ are i.i.d, $E\|\boldsymbol{X}_1\|^2 < \infty$, $\hat{\theta}_j = \theta_j(\bar{\boldsymbol{X}})$, $\theta_j(\cdot)$ is continuously differentiable at $\boldsymbol{\mu} = E(\boldsymbol{X}_1)$ with $\nabla\theta_j(\boldsymbol{\mu}) \neq 0$, and $\theta_j = \theta_j(\boldsymbol{\mu})$.

(a) $\lim_{N\to\infty} \sup_{x\in\mathbb{R}}\{\hat{L}(x) - L(x)\} \leq 0, \lim_{N\to\infty} \sup_{x\in\mathbb{R}}\{\hat{R}(x) - R(x)\} \geq 0.$

(b) *Let* $c^*_L(p) = \inf\{x \in \mathbb{R} : \hat{L}(x) \geq p\}$, $c^*_U(p) = \sup\{x \in \mathbb{R} : \hat{R}(x) \leq p\}$,

$$CI_{1-\alpha} := \left[\min_j \hat{\theta}_j - N^{-1/2} c^*_L(1-\alpha/2), \max_j \hat{\theta}_j - N^{-1/2} c^*_U(\alpha/2)\right]$$

satisfies $\lim_{N\to\infty} P([\min_j \theta_j, \max_j \theta_j] \subset CI_{1-\alpha}) \geq 1-\alpha$.

(c) *Let* $\hat{w}^+ = \hat{w} I(\hat{w} > 0)$, *where* $\hat{w} = \{\max_j \hat{\theta}_j - N^{-1/2} c^*_U(1/2)\} - \{\min_j \hat{\theta}_j - N^{-1/2} c^*_L(1/2)\}$, *and* $\hat{p} = 1 - \Phi(\rho\hat{w}^+)\alpha$, *with* $\rho \to \infty$ *and* $N^{-1/2}\rho \to 0$ *and* $\rho|\hat{w}^+ - (\max_j \theta_j - \min_j \theta_j)| = o_p(1)$, *then*

$$CI^{ATT}_{1-\alpha} := \left[\min_j \hat{\theta}_j - N^{-1/2} c^*_L(\hat{p}), \max_j \hat{\theta}_j - N^{-1/2} c^*_U(1-\hat{p})\right]$$

satisfies $\lim_{N\to\infty} \inf_{ATT_t \in [\min_j \theta_j, \max_j \theta_j]} P(ATT_t \in CI^{ATT}_{1-\alpha}) \geq 1-\alpha$.