Econometrics II

Lecture 9: Static Difference-in-Differences

David Schönholzer

Stockholm University

April 30, 2024

Plan for Today

Table of Contents

Definition of Treatment Structure

- Let $i \in \mathcal{N} = \{1, ..., N\}$ and $t \in \mathcal{T} = \{1, ..., T\}$
- Consider only balanced panel and $D_{it} \in \{0, 1\}$ throughout
- A unit's *treatment path* is $1 \times T$ vector $\mathbf{D}_i = (D_{i1}, ..., D_{iT})$
  - For example, $\mathbf{D}_i = (0, 1, 0, 1)$ or $(1, 1, 0, 0)$

- Say units i and i' are in the same *group* if $\mathbf{D}_i = \mathbf{D}_{i'}$
  - Let $G_i \in \mathcal{G} \subseteq \{1, ..., T, \infty\}$ denote the group i is in
  - For example, $\mathcal{G} = \{1, 2\}$
  - Let $\mathbf{D}\left(g\right)$ be the treatment path of group $g \in \mathcal{G}$

- Define the $|\mathcal{G}| \times T$ *grouped treatment structure* $\mathbf{D}$ as

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}\left(1\right) \\ \vdots \\ \mathbf{D}\left(G\right) \end{bmatrix} \overset{\text{e.g.}}{=} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Block Structure

- A *block structure* consists of only two groups:
  1. *Treatment group* $G_i = g$: $D_{it} = 0$ for $t < g$; $D_{it} = 1$ for $t \geq g$
  2. *Control group* $G_i = \infty$: $D_{it} = 0$ for all t
- *Example*: 2×3 block structure with $g = 2$:

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

- Define
  - Indicator for *ever-treated* units: $D_i \equiv \max_t D_{it}$
  - So if $D_i = 1$, then $G_i < \infty$ and if $D_i = 0$, then $G_i = \infty$
  - Indicator for *post-treated* periods: $P_t \equiv \max_i D_{it}$
- Typically reverse-sort $\mathbf{D}(g)$ by g, e.g. $\{\infty, 2\}$
- *Lemma*: Assume there is more than one group.
  Then: *Treatment structure is block structure iff* $D_{it} = D_i P_t$

Staggered Rollout Structure

- *Staggered rollout structure*: at least two groups:
  - Groups $G_i = g$ have path $D_{it} = 0$ for $t < g$; $D_{it} = 1$ for $t \geq g$
  - There may or may not be a control group $G_i = \infty$
- Could also be called *absorbing structure*:
  - $D_{is} \leq D_{it}$ for $s < t$
  - Can define group indices as $G_i = \arg\min_t D_{it}$
- *Example*: 3×4 staggered rollout structure:

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

- *Remark* 1: Block structure is special case with $G_i \in \{g, \infty\}$
- *Remark* 2: In SRS, $\mathbf{D}$ is just a function of $\mathcal{G}$ and T
  - In the example above, $T = 4$ and $G_i \in \{1, 2, 4\}$

Time Series Structures

- *Time series structure* has only one group

$$G_i = g \text{ with } 1 < g \leq T$$

- For example 1×4 with $G_i = 3$ for all i:

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix}$$

- Can we ever have cross-sectional variation?
- Sometimes: cohort panel with eligibility varying by cohort

- Example: Petra Persson (2020)
  - Social insurance reform in Sweden in 1988
  - Cohorts based on birth quarter of first child
  - National reform is staggered rollout for cohorts!

Assumptions for Static Causal Effects

- Start off with *static and homogeneous causal effects*
- Define potential outcomes $Y_{it}(d)$ with $d \in \{0, 1\}$

- Built-in assumptions by writing $Y_{it}(D_{it})$:
  - SUTVA in panels: no contamination across units
  - No anticipation/memory: no contamination across periods

- Individual causal effect: $Y_{it}(1) - Y_{it}(0)$
- ATE: $\mathbb{E}[Y_{it}(1) - Y_{it}(0)]$
- Estimand is ATT: $\tau \equiv \mathbb{E}[Y_{it}(1) - Y_{it}(0) | D_{it} = 1]$

Table of Contents

Basic Idea

- Origin: Snow (1849); prominence: Card and Krueger (1994)
- Effect of minimum wage increase D_{it} on employment Y_{it}
- New Jersey (NJ) increases minimum wage in mid-1992
    - Call early 1992 $t = 0$ and late 1992 $t = 1$
    - NJ gets $D_{i1} = 1$, while Pennsylvania (PA) keeps $D_{i1} = 0$
- Before enactment: NJ and PA averages: $\bar{Y}_{\text{NJ},0}$ and $\bar{Y}_{\text{PA},0}$
- After: $\bar{Y}_{\text{NJ},1}$ and $\bar{Y}_{\text{PA},1}$
- Intuitively:

$$\begin{aligned}\text{Treatment effect} &= (\bar{Y}_{\text{NJ},1} - \bar{Y}_{\text{PA},1}) - (\bar{Y}_{\text{NJ},0} - \bar{Y}_{\text{PA},0}) \\ &= (\bar{Y}_{\text{NJ},1} - \bar{Y}_{\text{NJ},0}) - (\bar{Y}_{\text{PA},1} - \bar{Y}_{\text{PA},0})\end{aligned}$$

- Plausible *even though D_{it} is clearly not randomized!* Why?
- We now study the mathematics that underlies this intuition

Setup

- Consider 2×2 block structure:

$$\mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

- Let $t \in \{0, 1\}$ (pre/post) and $D_i \in \{0, 1\}$ (control/treatment)
- Define the 2×2 *CEF matrix*:

$$\begin{bmatrix} \mathbb{E}\left[Y_{i0} | D_i = 0\right] & \mathbb{E}\left[Y_{i1} | D_i = 0\right] \\ \mathbb{E}\left[Y_{i0} | D_i = 1\right] & \mathbb{E}\left[Y_{i1} | D_i = 1\right] \end{bmatrix}$$

- The *potential outcomes matrix* (POM) for $d \in \{0, 1\}$ is

$$\begin{bmatrix} \mathbb{E}\left[Y_{i0}(d) | D_i = 0\right] & \mathbb{E}\left[Y_{i1}(d) | D_i = 0\right] \\ \mathbb{E}\left[Y_{i0}(d) | D_i = 1\right] & \mathbb{E}\left[Y_{i1}(d) | D_i = 1\right] \end{bmatrix}$$

Potential Outcomes Matrices

- Are elements of POM counterfactual or "factual"?
  - Counterfactuals are unobserved, factuals are observed
  - We can write $Y_{it} = [1 - D_i P_t] \, Y_{it}(0) + D_i P_t Y_{it}(1)$
  - Consider all eight POs in the 2×2 block structure POM:

$$Y_{it}(0) : \begin{bmatrix} \mathbb{E}[Y_{i0}(0)|D_i = 0] & \mathbb{E}[Y_{i1}(0)|D_i = 0] \\ \mathbb{E}[Y_{i0}(0)|D_i = 1] & \mathbb{E}[Y_{i1}(0)|D_i = 1] \end{bmatrix}$$

$$Y_{it}(1) : \begin{bmatrix} \mathbb{E}[Y_{i0}(1)|D_i = 0] & \mathbb{E}[Y_{i1}(1)|D_i = 0] \\ \mathbb{E}[Y_{i0}(1)|D_i = 1] & \mathbb{E}[Y_{i1}(1)|D_i = 1] \end{bmatrix}$$

- Notice observability mimics structure of **D**
- So the CEF matrix corresponds to following POs:

$$Y_{it} : \begin{bmatrix} \mathbb{E}[Y_{i0}(0)|D_i = 0] & \mathbb{E}[Y_{i1}(0)|D_i = 0] \\ \mathbb{E}[Y_{i0}(0)|D_i = 1] & \mathbb{E}[Y_{i1}(1)|D_i = 1] \end{bmatrix}$$

Identification of ATT

- We are interested in $\tau \equiv \mathbb{E}\left[Y_{it}(1) - Y_{it}(0) \,|D_{it} = 1\right]$
- Since $D_{it} = D_i P_t$ in block structures, we have

$$\begin{aligned}\tau &= \mathbb{E}\left[Y_{it}(1) - Y_{it}(0) \,|D_i = 1, P_t = 1\right] \\ &= \mathbb{E}\left[Y_{i1}(1) - Y_{i1}(0) \,|D_i = 1\right]\end{aligned}$$

  and hence

$$\begin{aligned}\tau &= \mathbb{E}\left[Y_{i1}(1) \,|D_i = 1\right] - \mathbb{E}\left[Y_{i1}(0) \,|D_i = 1\right] \\ &= \underbrace{\mathbb{E}\left[Y_{i1}|D_i = 1\right]}_{\text{observed}} - \underbrace{\mathbb{E}\left[Y_{i1}(0) \,|D_i = 1\right]}_{\text{need model}}\end{aligned}$$

- We model missing PO using *parallel trends assumption*:

$$\underbrace{\mathbb{E}\left[Y_{i1}(0) - Y_{i0}(0) \,|D_i = 1\right]}_{\text{counterfactual trend in } D_i = 1 \text{ group}} = \underbrace{\mathbb{E}\left[Y_{i1}(0) - Y_{i0}(0) \,|D_i = 0\right]}_{\text{"factual" trend in } D_i = 0 \text{ group}}$$

Modeling the Counterfactual through Parallel Trends

- Reorganizing parallel trends yields

$$
\begin{aligned}
\mathbb{E}\left[Y_{i1}(0)\,|D_i=1\right] &= \mathbb{E}\left[Y_{i0}(0)\,|D_i=1\right] \\
&\quad + \mathbb{E}\left[Y_{i1}(0)-Y_{i0}(0)\,|D_i=0\right] \\
&= \underbrace{\mathbb{E}\left[Y_{i0}|D_i=1\right]}_{\text{observed}} + \underbrace{\mathbb{E}\left[Y_{i1}-Y_{i0}|D_i=0\right]}_{\text{observed}}
\end{aligned}
$$

- From this, it follows that

$$
\begin{aligned}
\tau &= \mathbb{E}\left[Y_{i1}|D_i=1\right] - \mathbb{E}\left[Y_{i1}(0)\,|D_i=1\right] \\
&= \underbrace{\mathbb{E}\left[Y_{i1}-Y_{i0}|D_i=1\right]}_{\text{Treated pre/post diff}} - \underbrace{\mathbb{E}\left[Y_{i1}-Y_{i0}|D_i=0\right]}_{\text{Control pre/post diff}}
\end{aligned}
$$

- So DID simply fills in the missing PO using parallel trends

Table of Contents

Method of Moments in Simple Numerical Example

- Let $\bar{Y}_{gt} = \frac{1}{N_g} \sum_{i:G_i=g} Y_{it}$ be the group g mean in t
- Observable *group means matrix*:

$$\begin{bmatrix} \bar{Y}_{00} & \bar{Y}_{01} \\ \bar{Y}_{10} & \bar{Y}_{11} \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$

- Can use Method of Moments:

$$\begin{aligned} \hat{\tau}_{\mathrm{MM}} &= \widehat{\mathbb{E}}\left[Y_{i1} - Y_{i0} \middle| D_i = 1\right] - \widehat{\mathbb{E}}\left[Y_{i1} - Y_{i0} \middle| D_i = 0\right] \\ &= \left(\bar{Y}_{11} - \bar{Y}_{10}\right) - \left(\bar{Y}_{01} - \bar{Y}_{00}\right) \\ &= (5-3) - (3-2) = 2 - 1 = 1 \end{aligned}$$

- Note: equivalently, we estimated $\widehat{\mathbb{E}}\left[Y_{i1}(0) \middle| D_i = 1\right] = 4$

Additive Separability

- Parallel trends are equivalent to additive separability:

$$\mathbb{E}\left[Y_{it}\left(0\right)|D_i, P_t\right] = \mu + \alpha D_i + \gamma P_t$$

where $\mu \equiv \mathbb{E}\left[Y_{it}\left(0\right)|D_i = 0, P_t = 0\right]$ and

$$\alpha \equiv \mathbb{E}\left[Y_{it}\left(0\right)|D_i = 1, P_t = 0\right] - \mu$$
$$\gamma \equiv \mathbb{E}\left[Y_{it}\left(0\right)|D_i = 0, P_t = 1\right] - \mu$$

- Then, use observed outcome as function of POs:

$$Y_{it} = Y_{it}\left(0\right) + \left[Y_{it}\left(1\right) - Y_{it}\left(0\right)\right] D_i P_t$$

- Taking conditional expectations, we get

$$\begin{aligned}\mathbb{E}\left[Y_{it}|D_i, P_t\right] &= \mathbb{E}\left[Y_{it}\left(0\right)|D_i, P_t\right] \\ &\quad + \mathbb{E}\left[Y_{it}\left(1\right) - Y_{it}\left(0\right)|D_i, P_t\right] D_i P_t \\ &= \mu + \alpha D_i + \gamma P_t + \tau D_i P_t\end{aligned}$$

DID Estimation through Regression

- Recall estimating cell means – this is the same:

$$\mathbb{E}\left[Y_{it}|D_i, P_t\right] = \mu + \alpha D_i + \gamma P_t + \tau D_i P_t$$

- So this equals saturated OLS model:

$$Y_{it} = \mu + \alpha D_i + \gamma P_t + \tau D_i P_t + \varepsilon_{it}$$

- Hence, let $\beta = (\mu, \alpha, \gamma, \tau)$ and $\mathbf{X}_{it} = (1, D_i, P_t, D_i P_t)$:

$$\widehat{\beta}_{\text{OLS}} = \left(\sum_{i=1}^{N}\sum_{t=0}^{1}\mathbf{X}_{it}\mathbf{X}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=0}^{1}\mathbf{X}_{it}Y_{it}\right)$$

  and it can be shown that $\widehat{\tau}_{\text{OLS}} = \widehat{\tau}_{\text{MM}}$ if $N_0 = N_1$

Card and Krueger (1994) Example

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

Variable	Stores by state			Stores in New Jersey[a]			Differences within NJ[b]	
	PA (i)	NJ (ii)	Difference, NJ – PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low–high (vii)	Midrange–high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	–2.69 (1.37)	–2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	–2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores[c]	–2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	–2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0[d]	–2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	–2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

DID with Fixed Effects

- What about $Y_{it} = \alpha_i + \gamma_t + \tau D_i P_t + \varepsilon_{it}$? (see TWFE below)
- *Concerning* γ_t, in 2×2, full rank $\mathbf{X}_{it}\mathbf{X}_{it}'$ implies
    - Need to normalize one γ_t and include dummy for other one
    - This is equivalent to γP_t, just notational change
- *Concerning* α_i, apply FWL to Y_{it} and $D_{it} = D_i P_t$ to get

$$Y_{it} - \bar{Y}_i = \gamma_t + \tau \left(D_{it} - \bar{D}_i \right) + \varepsilon_{it} - \bar{\varepsilon}_i$$

    - But note that $D_{it} - \bar{D}_i = 0$ for $D_i = 0$
    - For treatment group, $D_{it} - \bar{D}_i = D_{it} - \frac{T-g-1}{T}$
    - Now let $\mu = -\tau \times \frac{T-g-1}{T}$ and $\alpha D_i + \bar{\varepsilon}_i = \bar{Y}_i$
- Then we have recovered $Y_{it} = \mu + \alpha D_i + \gamma P_t + \tau D_i P_t + \varepsilon_{it}$
- So algebraically equivalent!
- Only true in 2×2 block structures, strongly balanced panel

Table of Contents

Estimation $2 \times T$ DID

- Consider now $2 \times T$ block structures with $t = 1, ..., T$, e.g.

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

- 2×2 pre/post outcome matrix:

$$\begin{bmatrix} \mathbb{E}\left[Y_{it} | D_i = 0, P_t = 0\right] & \mathbb{E}\left[Y_{it} | D_i = 0, P_t = 1\right] \\ \mathbb{E}\left[Y_{it} | D_i = 1, P_t = 0\right] & \mathbb{E}\left[Y_{it} | D_i = 1, P_t = 1\right] \end{bmatrix}$$

- Corresponding 2×2 sample averages:

$$\begin{bmatrix} \bar{Y}_{0,\text{pre}} & \bar{Y}_{0,\text{post}} \\ \bar{Y}_{1,\text{pre}} & \bar{Y}_{1,\text{post}} \end{bmatrix}$$

where for $d \in \{0, 1\}$

$$\bar{Y}_{d,\text{pre}} = \frac{1}{N_d(g-1)} \sum_{i:D_i=d, t<g} Y_{it}; \quad \bar{Y}_{d,\text{post}} = \frac{1}{N_d(T-g+1)} \sum_{i:D_i=d, t\geq g} Y_{it}$$

Method of Moments and Regression Estimators

- MM: $\widehat{\tau}_{\text{MM}} = \left(\bar{Y}_{1,\text{post}} - \bar{Y}_{1,\text{pre}}\right) - \left(\bar{Y}_{0,\text{post}} - \bar{Y}_{0,\text{pre}}\right)$
- Regression: $\mathbb{E}^*\left[Y_{it}|1, D_i, P_t, D_iP_t\right]$ just like in 2×2
- $\widehat{\tau}_{\text{MM}} = \widehat{\tau}_{\text{OLS}}$ if $N_0 = N_1$ and $2(g-1) = T$

- Wide format illustration with $g = 4$, $T = 6$:

i	D_i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	Y_{i6}
1	0						
2	0		$\bar{Y}_{0,\text{pre}}$			$\bar{Y}_{0,\text{post}}$	
3	0						
4	1						
5	1		$\bar{Y}_{1,\text{pre}}$			$\bar{Y}_{1,\text{post}}$	
6	1						

- Run regression in long format

Table of Contents

Treatment-Timing Groups

- Let's turn to staggered rollout structures
- We now have at least two $G_i < \infty$, e.g.

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

- We might still have a never-treated group, e.g.

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

- No longer always $D_{it} = D_i P_t$ (since not block structure)
- Instead, we have $D_{it} = 1\left[t \geq G_i\right]$ for each $g \in \mathcal{G}$
  - Generalization; collapses to $D_{it} = D_i P_t$ in block structures

Identification and Estimation

- Identification of $\tau = \mathbb{E}\left[Y_{it}(1) - Y_{it}(0) \,|\, D_{it} = 1\right]$:
  - Analogous to $2 \times T$ case (see Appendix)
  - Again requires parallel trends: for $t \geq 2$ and any $g, g' \in \mathcal{G}$

$$\mathbb{E}\left[Y_{it}(0) - Y_{it-1}(0) \,|\, G_i = g\right] = \mathbb{E}\left[Y_{it}(0) - Y_{it-1}(0) \,|\, G_i = g'\right]$$

- Estimation with *two-way fixed effects* (TWFE) regression:

$$Y_{it} = \alpha_i + \gamma_t + D_{it}\tau_{\text{TWFE}} + \varepsilon_{it}$$

  - Called *staggered DID regression* when $D_{it} = 1\left[t \geq G_i\right]$
  - But even used if generic $D_{it} \in \{0, 1\}$ (i.e. D_{it} not absorbing)

- Two important recent insights in TWFE:
  1. $\tau_{\text{TWFE}} \neq \tau$ if effects are heterogeneous along group/time
  2. For staggered rollouts, τ_{TWFE} is weighted avg. of $2 \times T$ DIDs

Static Group-Time Treatment Effects

- Assume generic $\mathbf{D}$ with N_{gt} observations in cell (g, t)
  - Consider running $Y_{it} = \alpha_{G_i} + \gamma_t + \tau_{\text{TWFE}} D_{it} + \varepsilon_{it}$
  - How does τ_{TWFE} relate to τ?
- Write *static group-time treatment effects* as

$$\tau_{gt} = \frac{1}{N_{gt}} \sum_{i \in g} \left[Y_{it}(1) - Y_{it}(0) \right]$$

  which is called the ATE in cell (g, t)
- Define $N_1 = \sum_{i,t} D_{it}$ as the number of treated observations
- We can then write the ATT as

$$\tau = \mathbb{E} \left[\sum_{(g,t):D_{it}=1} \frac{N_{gt}}{N_1} \tau_{gt} \right]$$

  which says that τ is weighted average of group-time effects

Decomposing the TWFE Estimator

Theorem (de Chaisemartin & d'Haultfoeuille 2020)

TWFE regression is given by:

$$\tau_{TWFE} = \mathbb{E}\left[\sum_{(g,t):D_{it}=1} \frac{N_{gt}}{N_1} w_{gt} \tau_{gt}\right]$$

where

$$w_{gt} = \frac{\tilde{D}_{gt}}{\sum_{(g,t):D_{it}=1} \frac{N_{gt}}{N_1} \tilde{D}_{gt}}$$

where $\tilde{D}_{gt} = D_{it} - \mathbb{E}^*\left[D_{it} | \alpha_{G_i}, \gamma_t\right]$ *are the residuals of* D_{it} *after removing TWFE, which are constant within group (as is* D_{it}*).*

Understanding the Decomposition Result: Example

- What does this result say?
  - If τ_{gt} are constant, then $\tau_{\text{TWFE}} = \tau$ (as $\sum_{(g,t):D_{it}=1} \frac{N_{gt}}{N_1} w_{gt} = 1$)
  - But with varying τ_{gt}, in general $\tau_{\text{TWFE}} \neq \tau$
  - Hence τ_{TWFE} is generally a biased estimator of τ
- Consider staggered rollout with $T = 3$ and $\mathcal{G} = \{2, 3\}$

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}(3) \\ \mathbf{D}(2) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

- According to FWL, we have $\tilde{D}_{gt} = D_{it} - \bar{D}_g - \bar{D}_t + \bar{D}$
- This implies:

$$\tilde{D}_{33} = 1 - \frac{1}{3} - 1 + \frac{1}{2} = \frac{1}{6}$$
$$\tilde{D}_{22} = 1 - \frac{2}{3} - \frac{1}{2} + \frac{1}{2} = \frac{1}{3}$$
$$\tilde{D}_{23} = 1 - \frac{2}{3} - 1 + \frac{1}{2} = -\frac{1}{6}$$

Negative TWFE Although all (g, t)-ATEs are Positive

- Consider, for example, $\mathbb{E}[\tau_{33}] = \mathbb{E}[\tau_{22}] = 1$ and $\mathbb{E}[\tau_{23}] = 4$
- It follows then from the Theorem that

$$\tau_{\text{TWFE}} = \frac{1}{2}\mathbb{E}[\tau_{33}] + \mathbb{E}[\tau_{22}] - \frac{1}{2}\mathbb{E}[\tau_{23}] = -\frac{1}{2}$$

- So τ_{TWFE} is negative even though all $\mathbb{E}[\tau_{gt}]$ are positive!
- However, if e.g. $\mathbb{E}[\tau_{gt}] = 1$ (i.e. homogeneous), then

$$\tau_{\text{TWFE}} = \tau = 1$$

- The negative weight makes τ_{TWFE} very different from τ
- But why does this problem happen?

Where do Negative Weights Come From?

- It arises because it can be shown that in the example:

$$\tau_{\text{TWFE}} = \left(\text{DID}_1 + \text{DID}_2\right)/2$$

where

$$\begin{aligned}
\text{DID}_1 &= \mathbb{E}\left[\left(\bar{Y}_{22} - \bar{Y}_{21}\right) - \left(\bar{Y}_{32} - \bar{Y}_{31}\right)\right] \\
\text{DID}_2 &= \mathbb{E}\left[\left(\bar{Y}_{33} - \bar{Y}_{32}\right) - \left(\bar{Y}_{23} - \bar{Y}_{22}\right)\right]
\end{aligned}$$

- So it is the average of all 2×2 DIDs:

$$\text{DID}_1 : \left[\begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right] \text{ and } \text{DID}_2 : \left[\begin{array}{c|cc} 0 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right]$$

- $\text{DID}_1 = \mathbb{E}\left[\tau_{22}\right]$ as hoped, but $\text{DID}_2 = \mathbb{E}\left[\tau_{33}\right] - \left(\mathbb{E}\left[\tau_{23}\right] - \mathbb{E}\left[\tau_{22}\right]\right)$
- So DID_2 goes awry if τ_{23} is very different from τ_{22}
- Also see Goodman-Bacon (2020) for a similar result

Robustness to Heterogeneity

- de Chaisemartin and d'Haultfoeuille (2020) propose:
  - Diagnostic to assess sensitivity to heterogeneity
  - Alternative heterogeneity-robust estimator

- The target parameter for estimator is:

$$\tau_S \equiv \mathbb{E}\left[\frac{1}{N_S}\sum_{(g,t):t\geq 2, D_{it}\neq D_{it-1}} \left(Y_{it}(1) - Y_{it}(0)\right)\right]$$

  where $N_S = \sum_{(g,t):t\geq 2, D_{it}\neq D_{it-1}}$ are obs in *switching* cells

- This is the ATE of all switching cells $D_{it} \neq D_{it-1}$
- In staggered adoption, mean ATE at start of treatment

Heterogeneity-Robust Estimator

- The estimator is

$$\tau_{\text{dCdH}} = \sum_{t=2}^{T} \left(\omega_{+,t} \text{DID}_{+,t} + \omega_{-,t} \text{DID}_{-,t} \right)$$

  where $\omega_{+,t}$ and $\omega_{-,t}$ are sample weights
- Basic idea:
  - $\text{DID}_{+,t}$ measures "joiner-effect": $[0, 1]$ against $[0, 0]$
  - $\text{DID}_{-,t}$ "leaver-effect": $[1, 0]$ against $[1, 1]$
  - Ignore the rest – so use only DID_1 in example

- See `did_multiplegt`

Real-World Example: Gentzkow et al. (2011)

- $\hat{\beta}_{fe} = \widehat{\tau}_{\text{TWFE}}$
- $\text{DID}_\text{M} = \widehat{\tau}_{\text{dCdH}}$

TABLE 3—ESTIMATES OF THE EFFECT OF ONE ADDITIONAL NEWSPAPER ON TURNOUT

	Estimate	Standard error	Observations
$\hat{\beta}_{fd}$	0.0026	0.0009	15,627
$\hat{\beta}_{fe}$	−0.0011	0.0011	16,872
DID_M	0.0043	0.0014	16,872
$\text{DID}_\text{M}^\text{pl}$	−0.0009	0.0016	13,221
DID_M, on placebo subsample	0.0045	0.0019	13,221

Notes: This table reports estimates of the effect of one additional newspaper on turnout, as well as a placebo estimate of the common trends assumption underlying DID_M. Estimators are computed using the data of Gentzkow, Shapiro, and Sinkinson (2011), with state-year fixed effects as controls. Standard errors are clustered by county. To compute the DID_M estimators, the number of newspapers is grouped into 4 categories: 0, 1, 2, and more than 3.

Table of Contents

Serial Correlation in Y_{it} in Long Panels

- Consider staggered rollouts with moderaly large T (say 30)

- $Cov(Y_{it}, Y_{is})$ and $Cov(D_{it}, D_{is})$ are often high in this case
  - Makes it likely that $Cov(\varepsilon_{it}, \varepsilon_{is})$ is high as well
  - Ignoring serial correlation overstates precision of $\widehat{\beta}_{\text{OLS}}$

- Bertrand et al. (2004):
  - Robust standard errors reject too often
  - Treatment-unit clustered SEs work well $\rightarrow$ do this
  - Some "fancier" standard errors work, others do not

- Takeaway: always use G_i-clustered SEs as default

Table of Contents

$2 \times T$ Design Setup

- Consider now $2 \times T$ block structures with $t = 1, ..., T$, e.g.

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

- Two groups: $G_i \in \{g, \infty\}$ with $1 < g \leq T$
- Parameter of interest still: $\tau = \mathbb{E}\left[Y_{it}(1) - Y_{it}(0) | D_{it} = 1\right]$
- Now write $P_t \equiv \max_i D_{it} = 1\left[t \geq g\right]$

- Just like before, can split estimand:

$$\begin{aligned} \tau &= {\color{blue}\mathbb{E}\left[Y_{it}(1) | D_i = 1, P_t = 1\right]} - {\color{red}\mathbb{E}\left[Y_{it}(0) | D_i = 1, P_t = 1\right]} \\ &= \underbrace{\mathbb{E}\left[Y_{it} | D_i = 1, P_t = 1\right]}_{\text{observed}} - \underbrace{\mathbb{E}\left[Y_{it}(0) | D_i = 1, P_t = 1\right]}_{\text{need model}} \end{aligned}$$

Parallel Trends in $2 \times T$

- Parallel trends: for $t \geq 2$

$$\mathbb{E}\left[Y_{it}(0) - Y_{it-1}(0) \,|\, D_i = 1\right] = \mathbb{E}\left[Y_{it}(0) - Y_{it-1}(0) \,|\, D_i = 0\right]$$

so that specifically for $t \geq g$

$$\begin{aligned} \mathbb{E}\left[Y_{it}(0) \,|\, D_i = 1\right] &= \mathbb{E}\left[Y_{it}(0) \,|\, D_i = 1, P_t = 1\right] \\ &= \mathbb{E}\left[Y_{it-1}(0) \,|\, D_i = 1\right] \\ &\quad + \mathbb{E}\left[Y_{it}(0) - Y_{it-1}(0) \,|\, D_i = 0\right] \end{aligned}$$

- So we are already identified if $g = T$ (same as 2×2)

Multiple Post-Periods

- Recursively use $\mathbb{E}\left[Y_{it-1}(0)\,|D_i = 1\right]$ until it "turns blue"
  - If $t = g$: $\mathbb{E}\left[Y_{it-1}(0)\,|D_i = 1\right] = \mathbb{E}\left[Y_{it-1}|D_i = 1, P_{t-1} = 0\right]$
  - If $t = g + 1$ (i.e. second post-period), then

$$\begin{aligned}\mathbb{E}\left[Y_{it-1}(0)\,|D_i = 1\right] &= \mathbb{E}\left[Y_{it-2}|D_i = 1, P_{t-2} = 0\right] \\ &\quad + \mathbb{E}\left[Y_{it-1}(0) - Y_{it-2}(0)\,|D_i = 0\right]\end{aligned}$$

  - And so forth, with $s \geq 0$ for $t + s = g + 1$
- Hence we can show that τ is identified:

$$\begin{aligned}\tau = &\left(\mathbb{E}\left[Y_{it}|D_i = 1, P_t = 1\right] - \mathbb{E}\left[Y_{it}|D_i = 1, P_t = 0\right]\right) \\ &- \left(\mathbb{E}\left[Y_{it}|D_i = 0, P_t = 1\right] - \mathbb{E}\left[Y_{it}|D_i = 0, P_t = 0\right]\right)\end{aligned}$$