Time-Weighted Difference-in-Differences: Accounting for Common Factors in Short T Panels

Timo Schenk (U Amsterdam & Tinbergen Institute)

2nd International Econometrics PhD Conference
Rotterdam 2022

Motivation

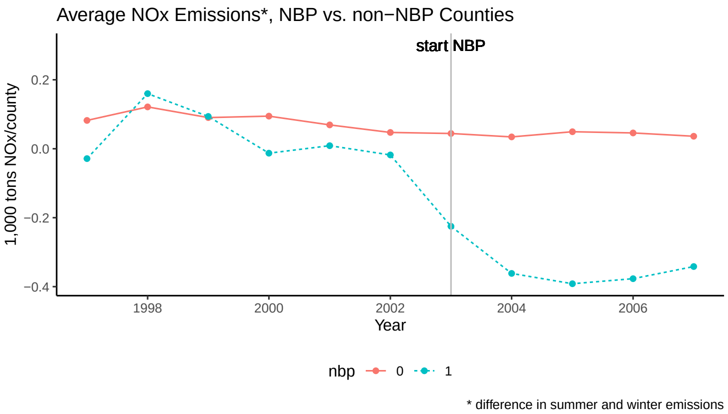
Setting: Binary treatment with **sharp** timing

- Observe outcome y_{it} for units $i = 1, \dots, N$ (large), periods $t = 1, \dots, T \geq 3$ (**small**),
- Two groups of units: $D_i = \begin{cases} 0 & \text{never treated} \\ 1 & \text{treated in } t = T_0 + 1, \dots, T \end{cases}$

Problem: Diff-in-diff (DID) estimation biased in presence of interactive fixed effects.

Solution: Equip the DID estimator with time weights.

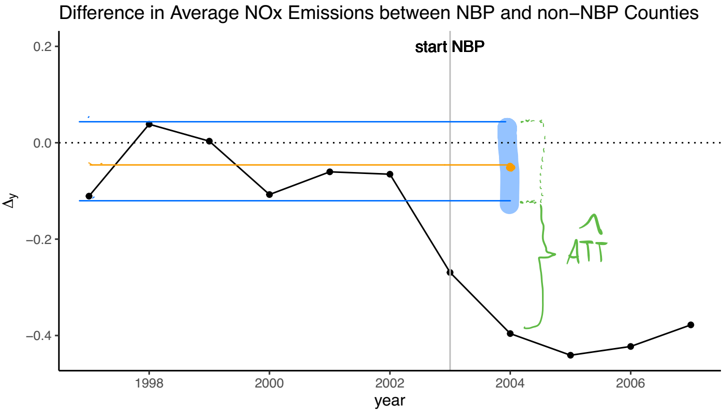
Empirical example: Deschenes et al. (2017) AER



NOx Bugdet Program (NBP) 2003-2008

- y_{it}: county/year level NOx emissions
- $N = 2539$ counties, of which ca. 50% are treated
- $T_0 = 6$ pre-treatment periods

Idea: use time weights



$$\widehat{ATT} = \Delta_T - \hat{\Delta}^{(0)}, \qquad \Delta_t = \bar{y}_t^{(1)} - \bar{y}_t^{(0)}$$

Benchmark: $\hat{\Delta}^{(0)}_{did} = \frac{1}{T_0} \sum_{t \leq T_0} \Delta_t$

This paper: $\hat{\Delta}^{(0)}(\boldsymbol{v}) = \sum_{t \leq T_0} v_t \Delta_t$, with $\sum_t v_t = 1,\ v_t \geq 0$

Related work

Synthetic Control & Synthetic DID

Abadie et al. (2015), Ferman and Pinto (2016), Arkhangelsky et al. (2021)

- SC: unit weights, no time weights; small N, large T
- SDID: unit weights and time weights; large N, large T

Panel Data with Interactive Fixed Effects (IFE)

Pesaran (2006), Bai (2009), Moon and Weidner (2015), Gobillon and Magnac (2016)

- large N, large T

Treatment Effects with IFE

Callaway and Karami, (2022)

- large N, small T
- requires time-invariant covariate Z_i with constant effect on y_{it}

This paper: only time weights, no covariate Z_i; large N, small T

Potential outcomes framework

Object of interest:

$$\tau := ATT = \mathrm{E}[y_{i,T}(1) - y_{i,T}(0)|D_i = 1]$$

- Potential outcomes $y_{it}(0), y_{it}(1)$
- Observed outcome $y_{it} = y_{it}(1)D_i + y_{it}(0)(1 - D_i)$,
- In the paper: multiple treated periods and staggered adoption.

Assumption

$y_{it}(1) = y_{it}(0)$ for all $t \leq T_0$ (no anticipation)

Interactive fixed effects model

$$y_{it}(0) = \beta_i + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + \varepsilon_{it}$$

- $\mathrm{E}[\boldsymbol{\varepsilon}_i | D_i, \boldsymbol{\lambda}_i] = 0$ and $\mathrm{E}[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' | D_i, \boldsymbol{\lambda}_i] = \boldsymbol{\Sigma}_{\varepsilon,i} > 0$
- $\boldsymbol{f}_t$: unobserved common factors with loadings $\boldsymbol{\lambda}_i$
- $\mathrm{Var}[\boldsymbol{\lambda}_i | D_i = j] = \boldsymbol{\Sigma}_{\lambda}^{(j)} > 0$: variation in how units are affected by $\boldsymbol{f}_t$
- $\mathrm{E}[\boldsymbol{\lambda}_i | D_i = 1] - \mathrm{E}[\boldsymbol{\lambda}_i | D_i = 0] = \boldsymbol{\xi}_{\lambda}$:
  treated and untreated units differ in how they are (on average) affected by $\boldsymbol{f}_t$.

Convex Hull Condition (CHC)

Assumption

The post-treatment factor $\boldsymbol{f}_T$ can be written as weighted average of pre-treatment factors $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{T_0}$:

$$\exists \boldsymbol{v} \in \mathbb{V}\colon \boldsymbol{f}_T = \boldsymbol{F}'_{\mathrm{pre}} \boldsymbol{v} \; (= \sum_{t \le T_0} v_t \boldsymbol{f}_t)$$

with $\mathbb{V} = \{\boldsymbol{v} \in \mathbb{R}^{T_0}\colon v_t \geq 0, \; \sum_{t=1}^{T_0} v_t = 1\}$ the set of non-negative weights that sum to one.

Examples

- $\boldsymbol{f} = (1, 2; 3)'$ CHC violated
- $\boldsymbol{f} = (1, 2; 1.5)'$ CHC met for $\boldsymbol{v} = (0.5, 0.5)'$
- $\boldsymbol{f} = (1, 2, 3; 2)'$ CHC met for $\boldsymbol{v} = (\alpha, 1 - 2\alpha, \alpha)'$ for all $\alpha \in [0, 0.5]$

Decomposing the ATT

$$
\begin{aligned}
ATT &= \mathrm{E}[y_{i,T}(1)|D_i = 1] - \mathrm{E}[y_{i,T}(0)|D_i = 0] \\
&\quad - \sum_{t \le T_0} v_t (\mathrm{E}[y_{i,t}(0)|D_i = 1] - \mathrm{E}[y_{i,t}(0)|D_i = 0]) \\
&\quad - \boldsymbol{\xi}_\lambda' (\boldsymbol{f}_T - \sum_{t \le T_0} v_t \boldsymbol{f}_t)
\end{aligned}
$$

with factor imbalance $\boldsymbol{\xi}_f(\boldsymbol{v}) = \boldsymbol{f}_T - \sum_{t \le T_0} v_t \boldsymbol{f}_t$.

- DID: requires $\boldsymbol{\xi}_\lambda = 0$ (parallel trends), use $v_t = \frac{1}{T_0}$.
- TWDID: requires $\boldsymbol{\xi}_f(\boldsymbol{v}) = 0$ (no trends) for some $\boldsymbol{v}$.

Balancing common shocks with time weights

Time weighted DID estimator for given weights $\boldsymbol{v}$:

$$\hat{\tau}(\boldsymbol{v}) = \bar{y}_T^{(1)} - \bar{y}_T^{(0)} - \sum_{t \leq T_0} v_t(\bar{y}_t^{(1)} - \bar{y}_t^{(0)})$$

which is implemented by a weighted 2wfe regression.

Theorem

1. $\mathrm{E}[\hat{\tau}(\boldsymbol{v}) - \tau] = \boldsymbol{\xi}_\lambda' \boldsymbol{\xi}_f(\boldsymbol{v})$
2. $\mathrm{Var}[\hat{\tau}(\boldsymbol{v})] = \boldsymbol{\xi}_f(\boldsymbol{v})' \boldsymbol{\Sigma}_\lambda \boldsymbol{\xi}_f(\boldsymbol{v}) + V_z(\boldsymbol{v})$

Factor imbalance causes bias and amplifies the variance.

Estimating the weights from control units

Which weighted average of pre-treatment outcomes predicts best the post-treatment outcome?
Estimate

$$y_{i,T} = \alpha + \sum_{t \leq T_0} v_t y_{it} + \eta_i, \quad i \in \mathcal{N}_0 \text{ (control units)}$$

s.t. $\sum_{t \leq T_0} v_t = 1$ and $v_t \geq 0$ by restricted least-squares.

$$\min_{\boldsymbol{v} \in \mathbb{V}, \alpha} \sum_{i \in \mathcal{N}_0} \left(y_{i,T} - \alpha - \sum_{t \leq T_0} v_t y_{it} \right)^2$$

Properties of the estimated time weights

Does $\hat{\boldsymbol{v}}$ converge to something desireable?

Theorem

$$\hat{\boldsymbol{v}} \xrightarrow{p} \boldsymbol{v}^* := \arg\min_{\boldsymbol{v} \in \mathbb{V}} \underbrace{\left\{ \boldsymbol{\xi}_f(\boldsymbol{v})' \boldsymbol{\Sigma}_\lambda^{(0)} \boldsymbol{\xi}_f(\boldsymbol{v}) + V_z^{(0)}(\boldsymbol{v}) \right\}}_{\mathrm{Var}[\hat{\tau}(\boldsymbol{v})]}$$

and

$$\sqrt{N}(\hat{\boldsymbol{v}} - \boldsymbol{v}^*) \xrightarrow{d} \mathcal{N}\left[0, \boldsymbol{\Sigma}_{\hat{v}}\right]$$

Take-away

- The weights minimize the limiting variance of $\hat{\tau}(\boldsymbol{v})$,
- ...but may not balance the factors perfectly ($\boldsymbol{\xi}_f(\boldsymbol{v}^*) \neq 0$)
- ...so some bias $b(\boldsymbol{v}^*) = \boldsymbol{\xi}_\lambda' \boldsymbol{\xi}_f(\boldsymbol{v}^*)$ remains

On the remaining bias

Define oracle weights

$$\boldsymbol{v}_0 = \arg\min_{\boldsymbol{v} \in \mathbb{V}} V_z^{(0)}(\boldsymbol{v}) \quad s.t.\ \boldsymbol{\xi}_f(\boldsymbol{v}) = 0$$

and variance minimizing weights

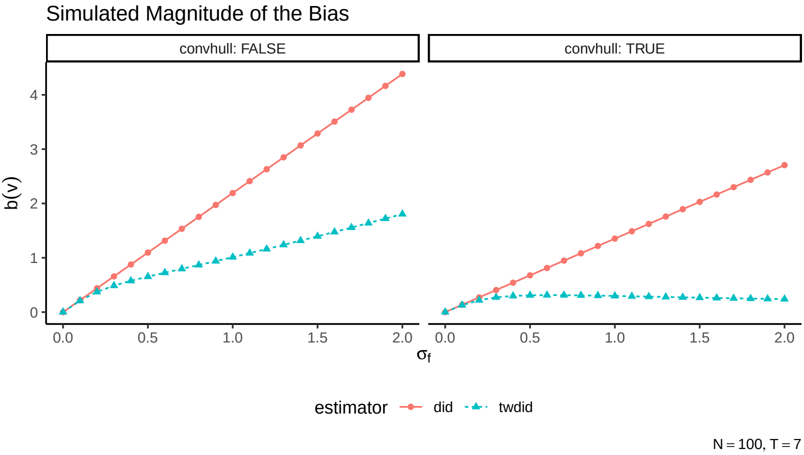
$$\boldsymbol{v}_\varepsilon = \arg\min_{\boldsymbol{v} \in \mathbb{V}} V_z^{(0)}(\boldsymbol{v})$$

Then

$$\boldsymbol{\xi}_f(\boldsymbol{v}^*) = \boldsymbol{F}'_{\mathrm{pre}}(\boldsymbol{I} + \boldsymbol{A}_{\mathrm{snr}})^{-1}(\boldsymbol{v}_0 - \boldsymbol{v}_\varepsilon)$$

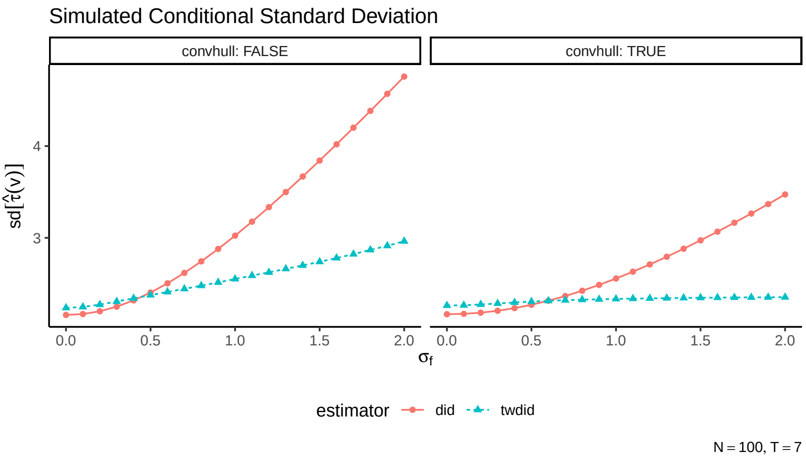
with $\boldsymbol{A}_{\mathrm{snr}}$ measuring the signal-to-noise ratio.

Monte Carlo: Bias



- DGP: $y_{it} = \sigma_f \lambda_i f_t + \varepsilon_{it}$, $\lambda_i | D_i \sim \mathcal{N}[1 + 0.2D_i, 1]$, $\varepsilon_{it} \sim \mathcal{N}[0, 1]$
- $f_t \sim \mathcal{N}[0, 1]\ \forall t$ (left) vs. $f_t \sim \mathcal{N}[0, 1]\ \forall t \leq T_0$, $f_T \sim \mathcal{TN}[f_{(1)}, f_{(T_0)}]$ (right)

Monte Carlo: Standard Deviation



- DGP: $y_{it} = \sigma_f \lambda_i f_t + \varepsilon_{it}$, $\lambda_i | D_i \sim \mathcal{N}[1 + 0.2D_i, 1]$, $\varepsilon_{it} \sim \mathcal{N}[0, 1]$
- $f_t \sim \mathcal{N}[0, 1]\ \forall t$ (left) vs. $f_t \sim \mathcal{N}[0, 1]\ \forall t \leq T_0$, $f_T \sim \mathcal{TN}[f_{(1)}, f_{(T_0)}]$ (right)

Inference

Asymptotic normality

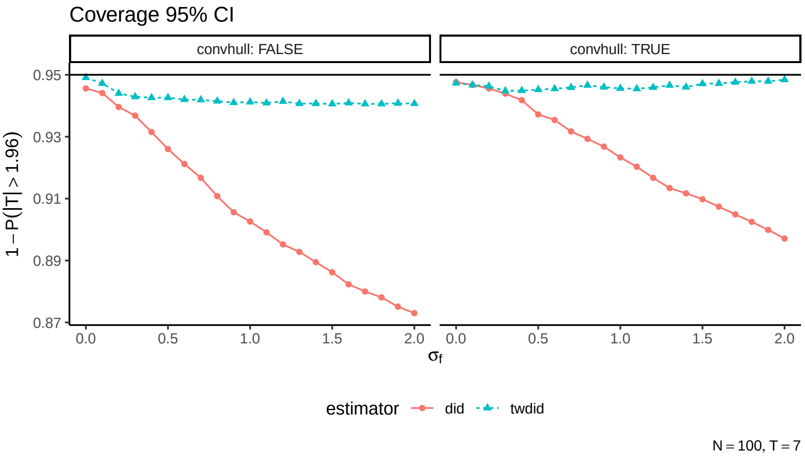
$$\sqrt{N}(\hat{\tau}(\hat{\boldsymbol{v}}) - \tau - b(\boldsymbol{v}^*)) \xrightarrow{d} \mathcal{N}[0, V_{\hat{\tau}}]; \quad V_{\hat{\tau}} = \text{var}[\hat{\tau}(\boldsymbol{v}^*)] + \xi_\lambda' \boldsymbol{F}_{pre}' \boldsymbol{\Sigma}_{\hat{v}} \boldsymbol{F}_{pre} \xi_\lambda$$

Standard errors accounting for weight estimation uncertainty:

$$\widehat{V}_{\hat{\tau}} = \widehat{V}_{\mathrm{ccm}} + \dot{\boldsymbol{\Delta}}_{pre}' \widehat{\boldsymbol{\Sigma}}_{\hat{v}} \dot{\boldsymbol{\Delta}}_{pre}$$

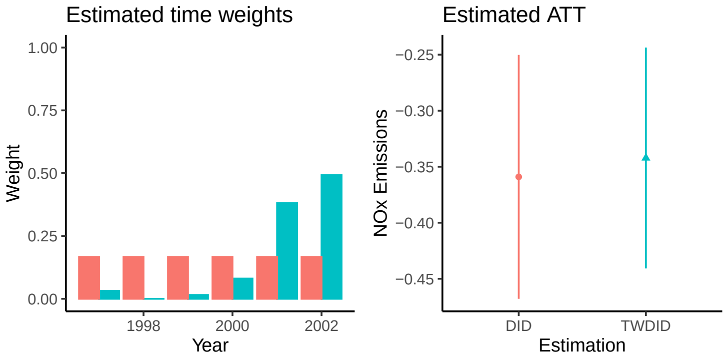
- $\widehat{V}_{\mathrm{ccm}}$ weighted cluster covariance matrix (CCM) estimator
- $\widehat{\boldsymbol{\Sigma}}_{\hat{v}}$ the estimated time weight covariance matrix
- $\dot{\boldsymbol{\Delta}}_{pre}$ the demeaned pre-treatment differences in outcomes

DID vs. TWDID: Coverage of CI



What difference does time weighting make?

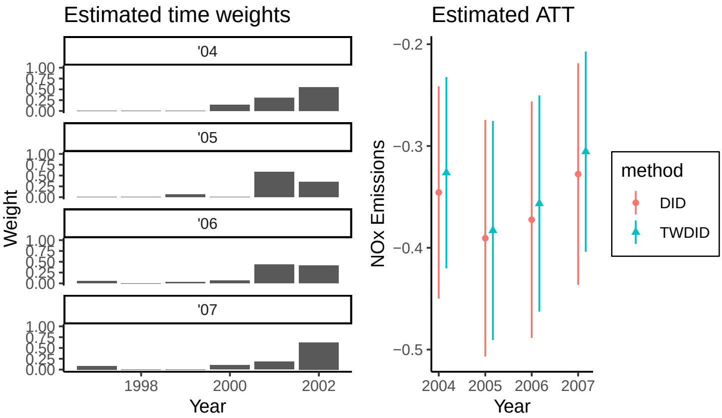
Average ATT



- 95% confidence interval: $\left[\hat{\tau}(\hat{\boldsymbol{v}}) \pm 1.96\sqrt{\widehat{V}_{\hat{\tau}}}\right]$
- TWDID standard error 10% smaller, point estimate similar.

What difference does time weighting make?

Dynamic ATT



Summary

Problem: Diff-in-diff (DID) estimation biased in presence of interactive fixed effects.

Solution: Equip the DID estimator with time weights!

- Substantial bias and variance reduction
- Standard errors need to be adjusted for weight estimation uncertainty
- NOx application: TWDID yields similar point estimates but 10% smaller standard errors

Thank you!

✉ t.d.schenk@uva.nl
🌍 timoschenk.github.io

Time weight estimation

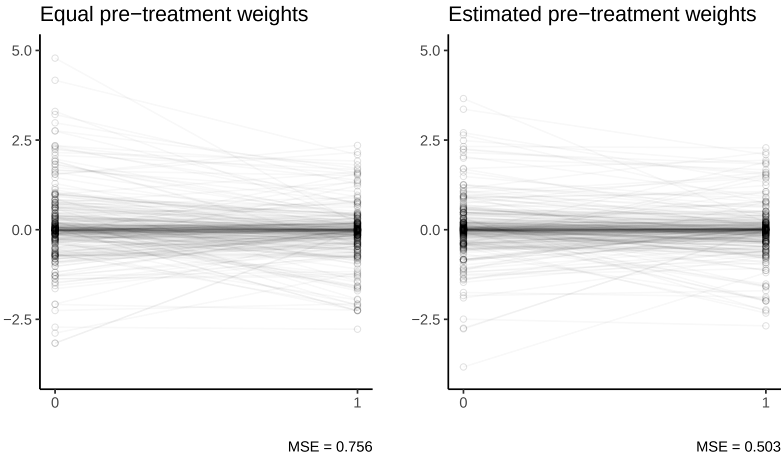
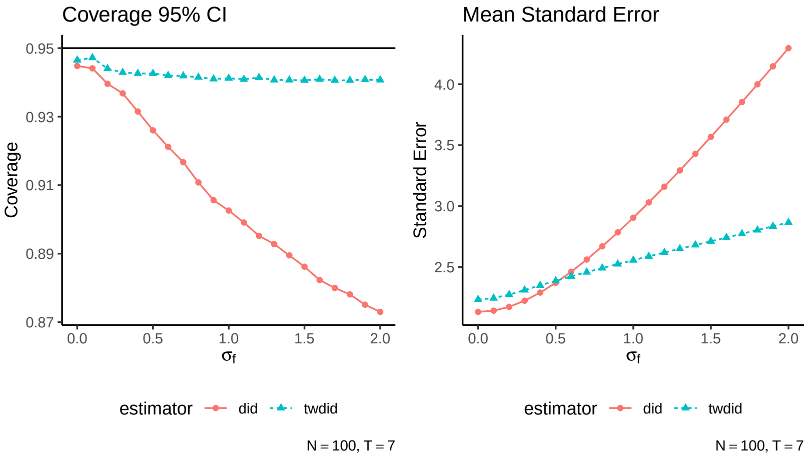


Figure: $\sum_{t \leq T_0} v_t y_{it}$ (0) vs. $\bar{y}_{i,post}$ (1) for control unit $i \in \mathcal{N}_0$. Left: equal weights $v_t = \frac{1}{T_0}$, Right: estimated weights $\hat{v}_t$.

DID vs. TWDID: Coverage and length of CI



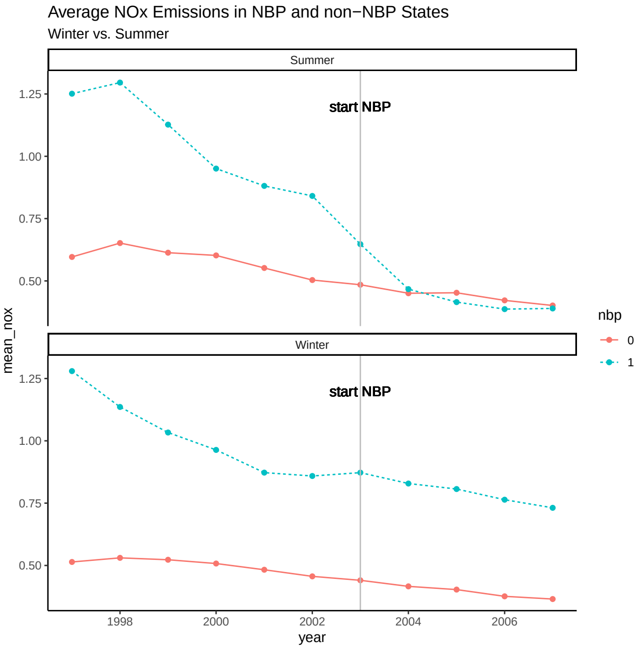
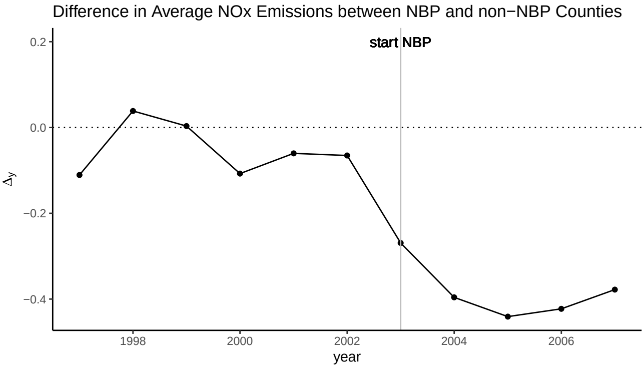


Figure: Deschenes et al. (2017)

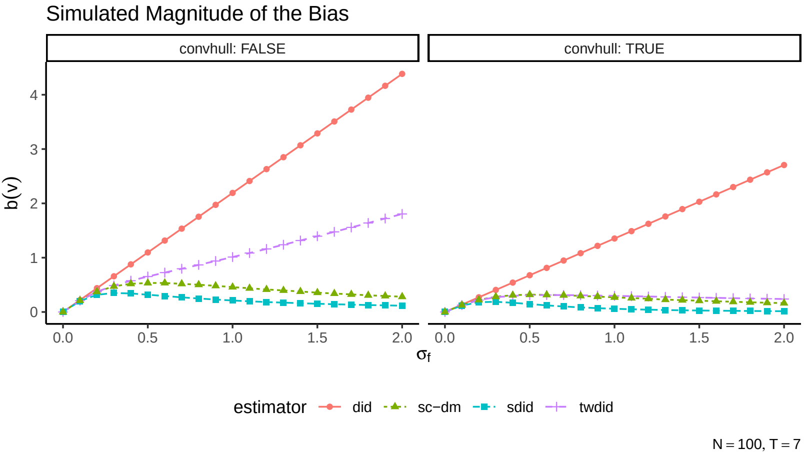
Evidence of the factor structure



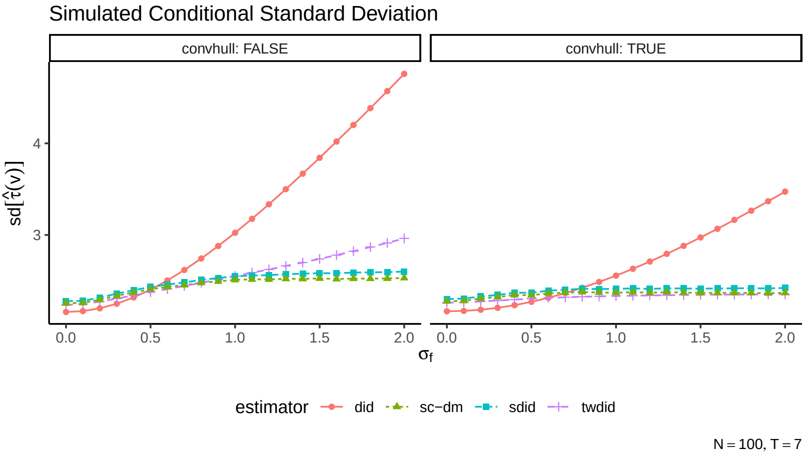
$$\Delta_t = \bar{\beta}^{(1)} - \bar{\beta}^{(0)} + \boldsymbol{\xi}_\lambda' \boldsymbol{f}_t + \tau \mathrm{I}(t > T_0) + O_p(\frac{1}{\sqrt{N}})$$

with loading imbalance $\boldsymbol{\xi}_\lambda = \bar{\boldsymbol{\lambda}}^{(1)} - \bar{\boldsymbol{\lambda}}^{(0)}$

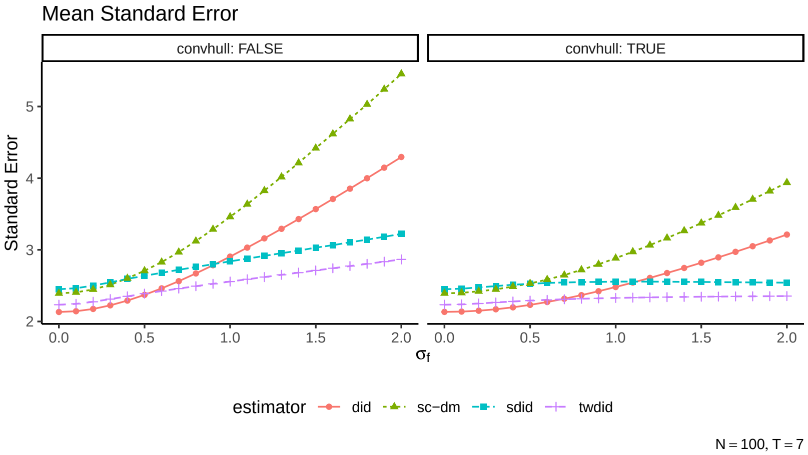
Additional Simulations



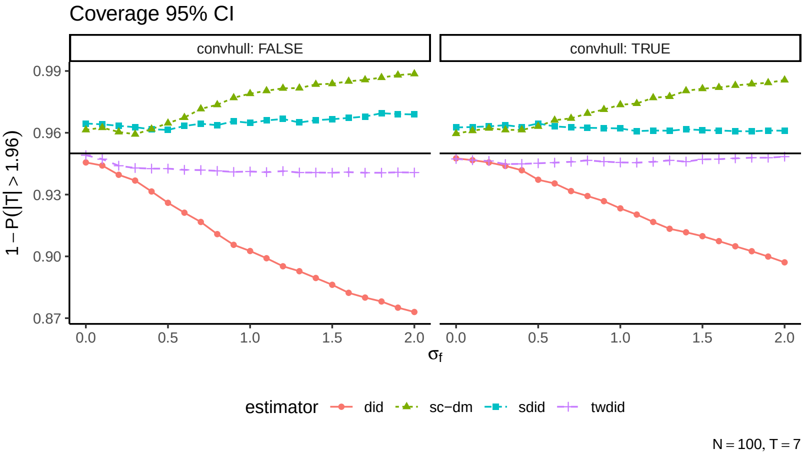
Additional Simulations



Additional Simulations

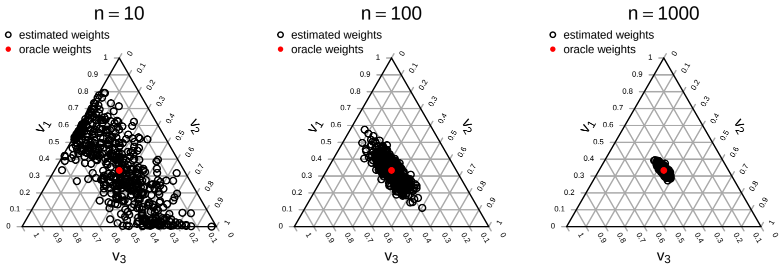


Additional Simulations



Convergence of the time weights

Simulate $\hat{\boldsymbol{v}}$ with $T_0 = 3$, $\boldsymbol{v}_0 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})'$, $\boldsymbol{\Sigma}_\varepsilon = \sigma_\varepsilon^2 \boldsymbol{I}_T$.



Difference-in-Differences in Environmental Economics

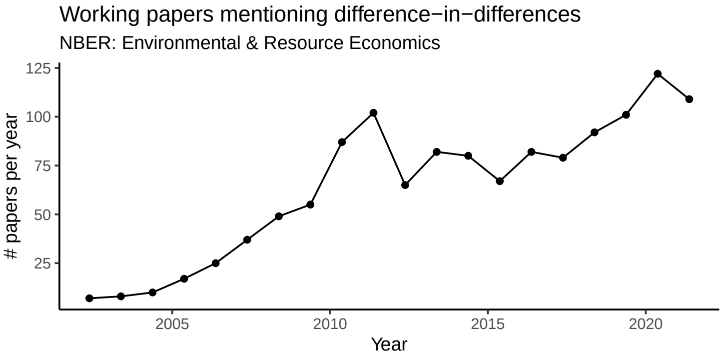


Figure: Papers contain the phrase "difference-in-differences", manually obtained from `https://www.nber.org/`.