Synthetic Controls with Staggered Adoption

Eli Ben-Michael, Avi Feller, and Jesse Rothstein

UC Berkeley

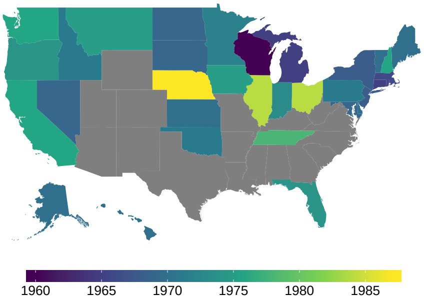
Online Causal Inference Seminar

September 2020

What is the impact of teacher unionization on education spending?

- *1960 – 1987:* 34 states pass mandatory collective bargaining laws

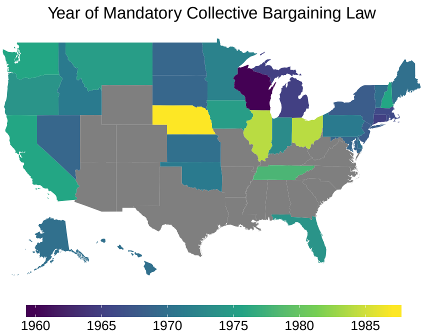
Year of Mandatory Collective Bargaining Law



What is the impact of teacher unionization on education spending?

- *1960 – 1987:* 34 states pass mandatory collective bargaining laws

- Impact of teachers unions unclear
  - ↑ Increase expenditures by **12%** [Hoxby, 1996]
  - ↔ Or really no effect at all? [Paglayan, 2019]

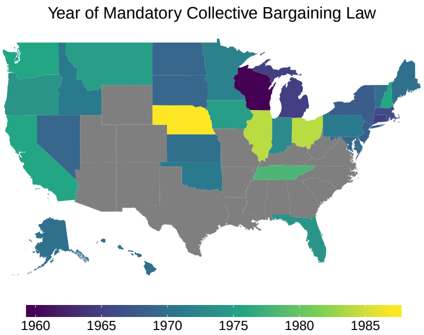


What is the impact of teacher unionization on education spending?

- *1960 – 1987:* 34 states pass mandatory collective bargaining laws

- Impact of teachers unions unclear
  - ↑ Increase expenditures by **12%** [Hoxby, 1996]
  - ↔ Or really no effect at all? [Paglayan, 2019]

- What should we believe?



Estimating effects under staggered adoption

Staggered adoption: Multiple units adopt treatment over time

Estimating effects under staggered adoption

Staggered adoption: Multiple units adopt treatment over time

Common approaches can fail: Little guidance when this happens

- Difference in Differences (DiD) requires parallel trends assumption
- Synthetic Control Method (SCM) designed for single treated unit, poor fit for average

Estimating effects under staggered adoption

Staggered adoption: Multiple units adopt treatment over time

Common approaches can fail: Little guidance when this happens

- Difference in Differences (DiD) requires parallel trends assumption
- Synthetic Control Method (SCM) designed for single treated unit, poor fit for average

Partially pooled SCM

- Modify optimization problem to target overall and state-specific fit
- Account for level differences with Intercept-Shifted SCM

What do we want to estimate?

Units: $i = 1, \ldots, N$, J total treated units

Time: $t = 1, \ldots, T$, treatment times $T_1, \ldots, T_J, \infty$

Outcome: at *event time* k, Y_{i,T_j+k}

– Some assumptions to write down potential outcomes
  [Athey and Imbens, 2018; Imai and Kim, 2019]

$$\text{treat} = \begin{pmatrix} & \checkmark & \checkmark & \checkmark \\ & & \checkmark & \checkmark \\ & & & \checkmark \\ & & & \end{pmatrix}$$

What do we want to estimate?

Units: $i = 1, \dots, N$, J total treated units

Time: $t = 1, \dots, T$, treatment times $T_1, \dots, T_J, \infty$

Outcome: at *event time* k, Y_{i,T_j+k}

- Some assumptions to write down potential outcomes
  [Athey and Imbens, 2018; Imai and Kim, 2019]

$$\text{treat} = \begin{pmatrix} & \checkmark & \checkmark & \checkmark \\ & & \checkmark & \checkmark \\ & & & \checkmark \\ & & & \end{pmatrix}$$

Basic building block:

$$\tau_{jk} = Y_{j,T_j+k}(T_j) \quad - \quad \underbrace{Y_{j,T_j+k}(\infty)}_{\sum \hat{\gamma}_{ij} Y_{i,T_j+k}}$$

What do we want to estimate?

Units: $i = 1, \ldots, N$, J total treated units

Time: $t = 1, \ldots, T$, treatment times $T_1, \ldots, T_J, \infty$

Outcome: at *event time* k, Y_{i,T_j+k}

- Some assumptions to write down potential outcomes
  [Athey and Imbens, 2018; Imai and Kim, 2019]

$$\text{treat} = \begin{pmatrix} & \checkmark & \checkmark & \checkmark \\ & & \checkmark & \checkmark \\ & & & \checkmark \\ & & & \end{pmatrix}$$

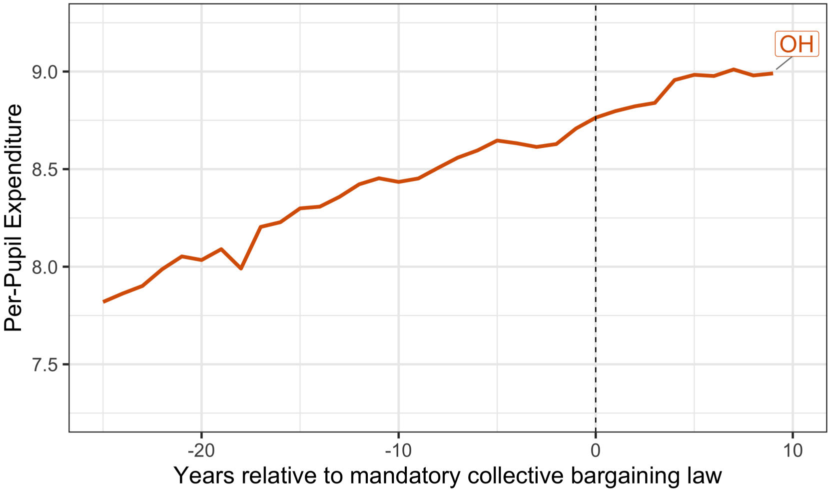
Basic building block:

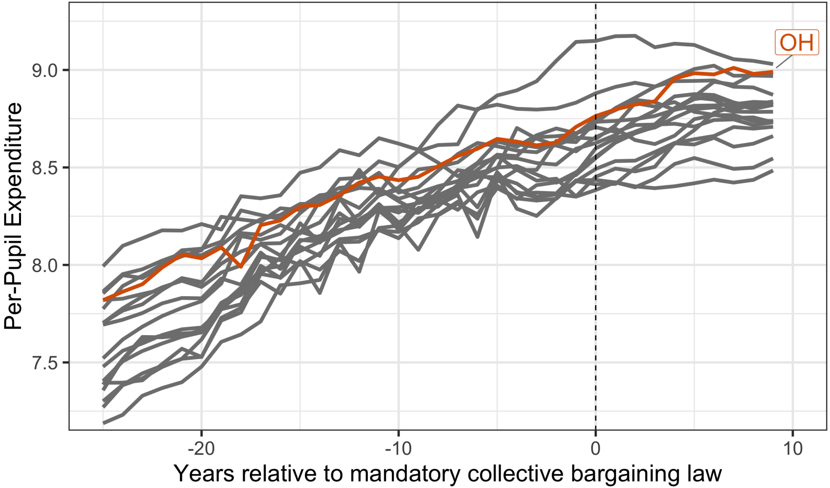
$$\tau_{jk} = Y_{j,T_j+k}(T_j) \quad - \quad \underbrace{Y_{j,T_j+k}(\infty)}_{\sum \hat{\gamma}_{ij} Y_{i,T_j+k}}$$

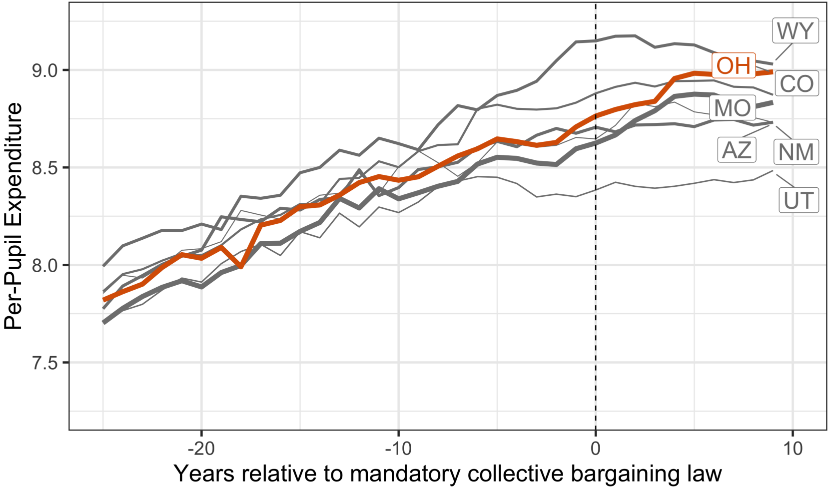
Average at event time k:

$$\text{ATT}_k = \frac{1}{J} \sum_{j=1}^{J} \tau_{jk}$$

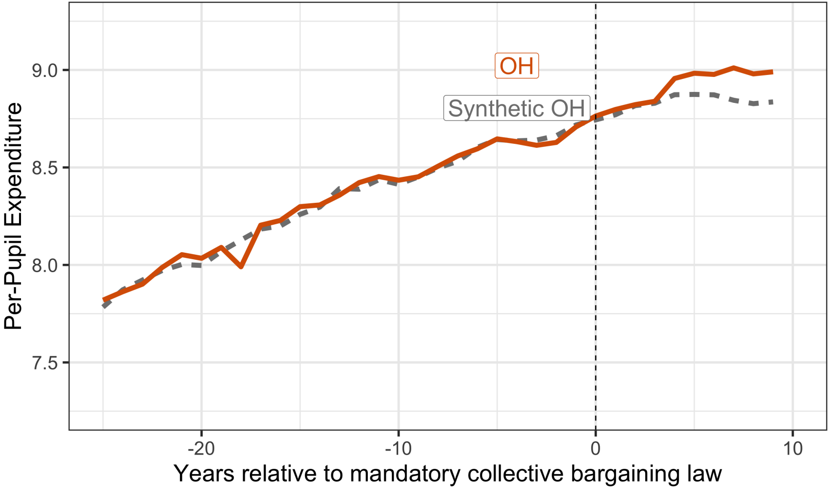
Separate SCM


OH
9.0
8.5
8.0
7.5
Per-Pupil Expenditure
-20
-10
0
10
Years relative to mandatory collective bargaining law

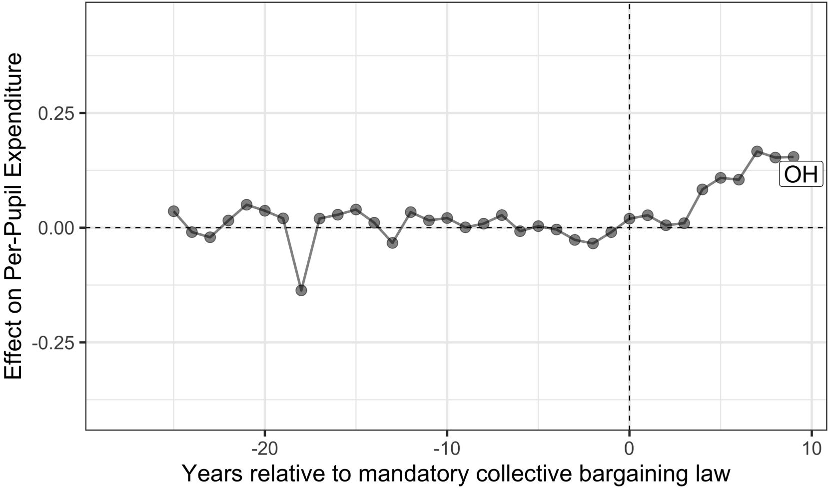

OH
9.0
8.5
8.0
7.5
Per-Pupil Expenditure
-20
-10
0
10
Years relative to mandatory collective bargaining law



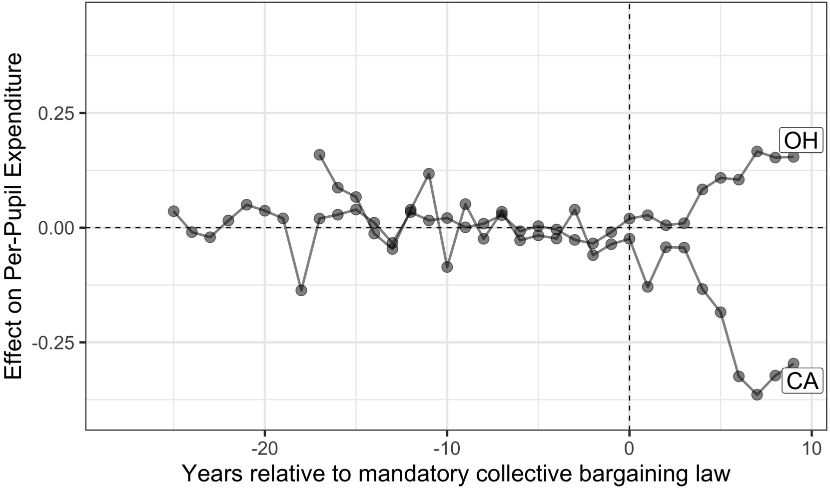
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \quad \|\text{State Balance}_j\|_2^2$$



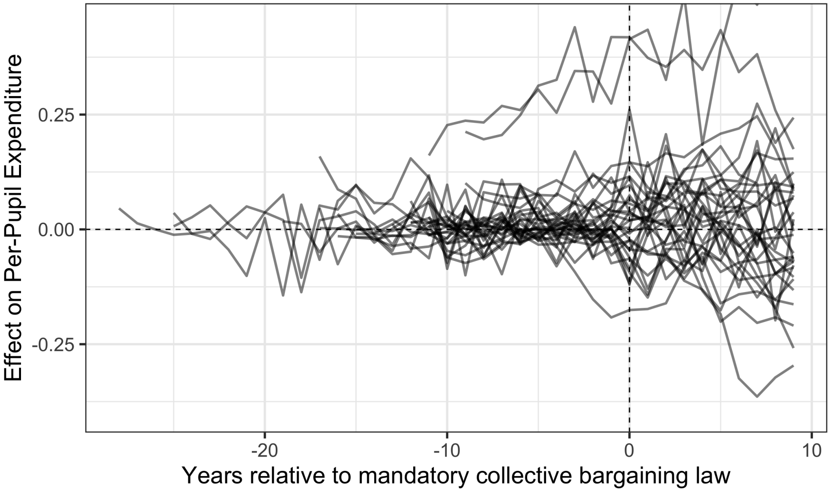
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \quad \|\text{State Balance}_j\|_2^2$$



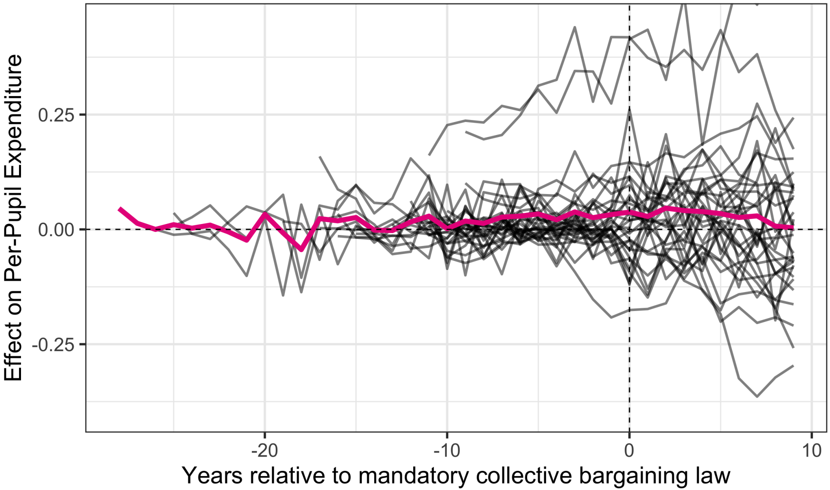
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \quad \|\text{State Balance}_j\|_2^2$$



$$\min_{\gamma_1,\ldots,\gamma_J \in \Delta^{\text{scm}}} \frac{1}{J} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2$$

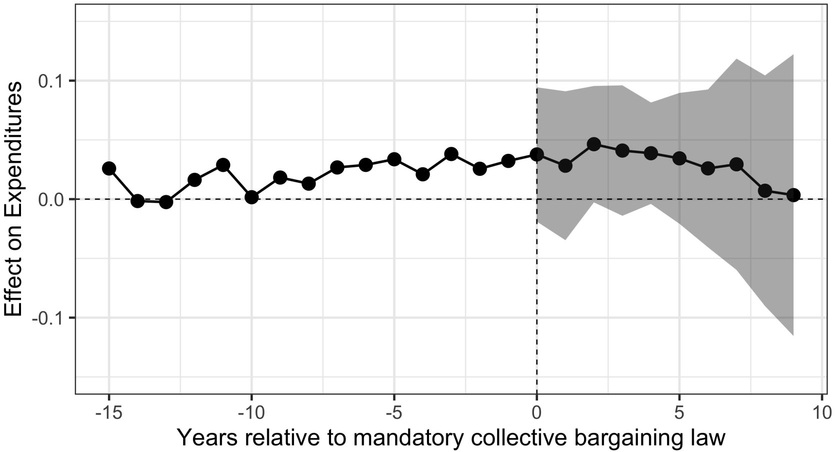


$$\min_{\gamma_1,\ldots,\gamma_J \in \Delta^{\text{scm}}} \frac{1}{J} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2$$



$$\min_{\gamma_1,\ldots,\gamma_J \in \Delta^{\text{scm}}} \frac{1}{J} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2$$

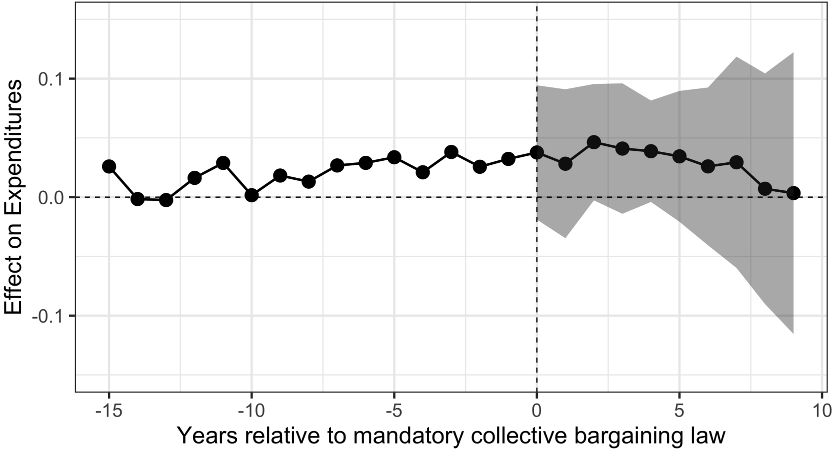
Separate SCM



$$\min_{\gamma_1,\ldots,\gamma_J \in \Delta^{\text{scm}}} \frac{1}{J} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2$$

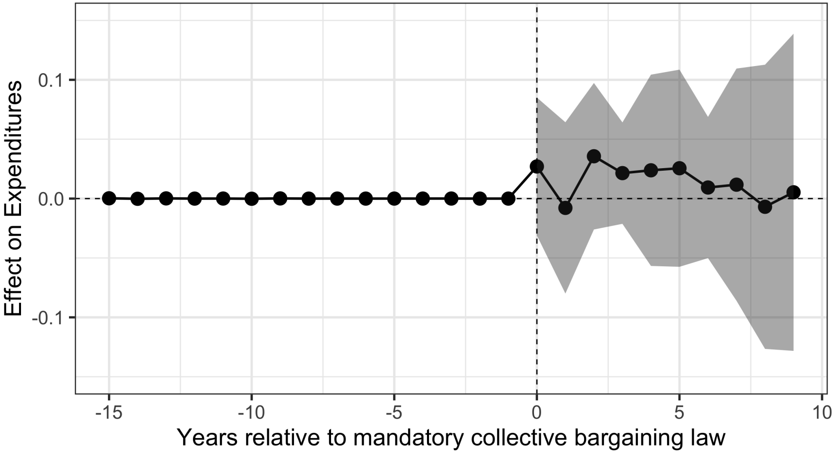
Moving Beyond Separate SCM

Separate SCM



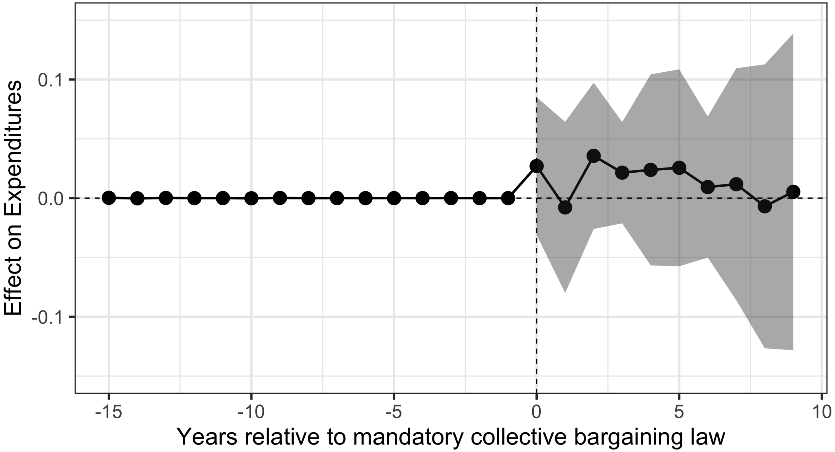
$$\min_{\Gamma} \frac{1}{J} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2$$

Pooled SCM

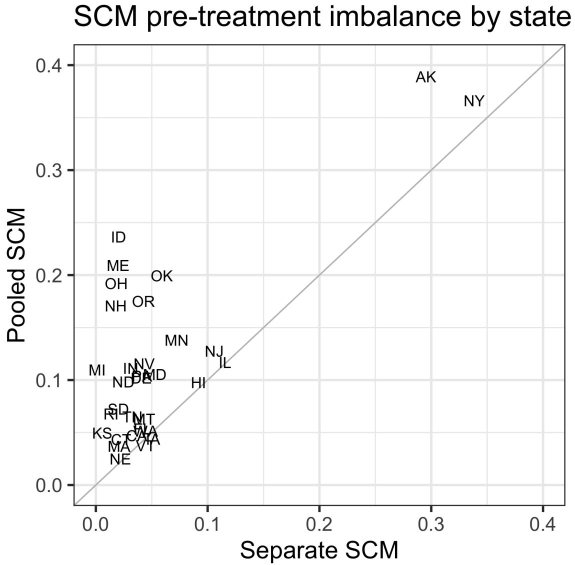


$$\min_{\Gamma} \left\| \frac{1}{J} \sum_{j=1}^{J} \text{State Balance}_j \right\|_2^2$$

Pooled SCM



$$\min_{\Gamma} \|\text{Avg Balance}\|_2^2$$



- Avg Balance is better
- but State Balance is worse.

Which matters more?

Generalization of parallel trends: Linear Factor Model

$$Y_{it}(\infty) = \phi_i' \mu_t + \varepsilon_{it}$$

Which matters more?

Generalization of parallel trends: Linear Factor Model

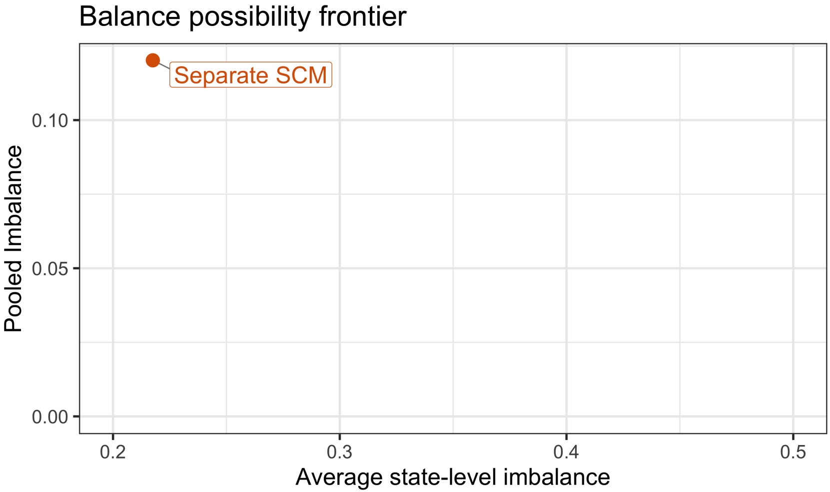
$$Y_{it}(\infty) = \phi_i' \mu_t + \varepsilon_{it}$$

Error for ATT

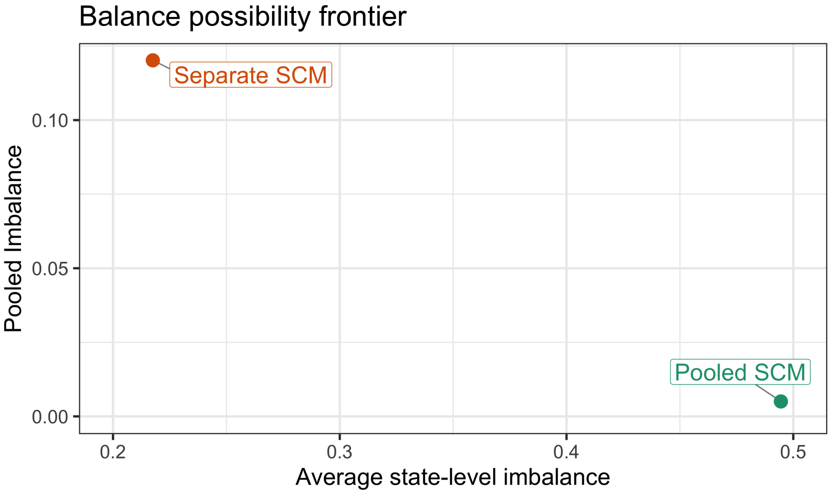
$$\left| \widehat{\text{ATT}}_0 - \text{ATT}_0 \right| \lesssim \|\bar{\mu}\|_2 \|\text{Avg Balance}\|_2 + S \sqrt{\sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2} + \sqrt{\frac{\log NJ}{T}}$$

Level of **heterogeneity over time** is important

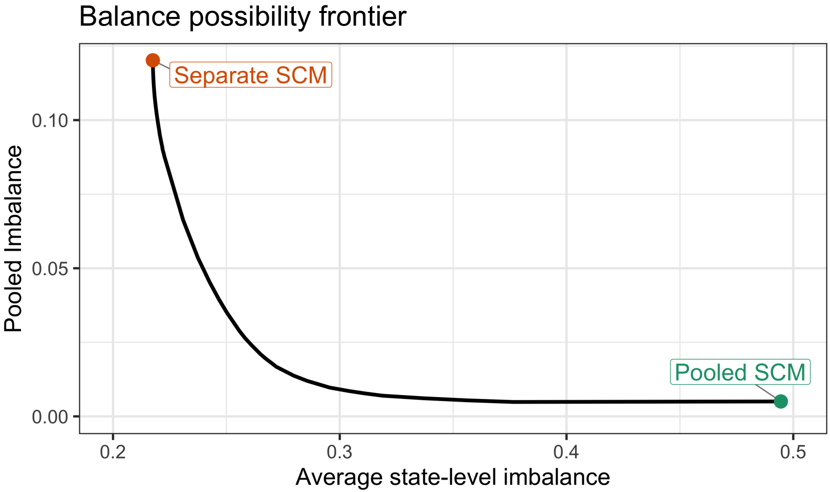
- $\bar{\mu}$ is the average factor value → importance of Avg Balance
- S is the factor standard deviation → importance of State Balance
- Special case: unit fixed effects, only Avg Balance matters



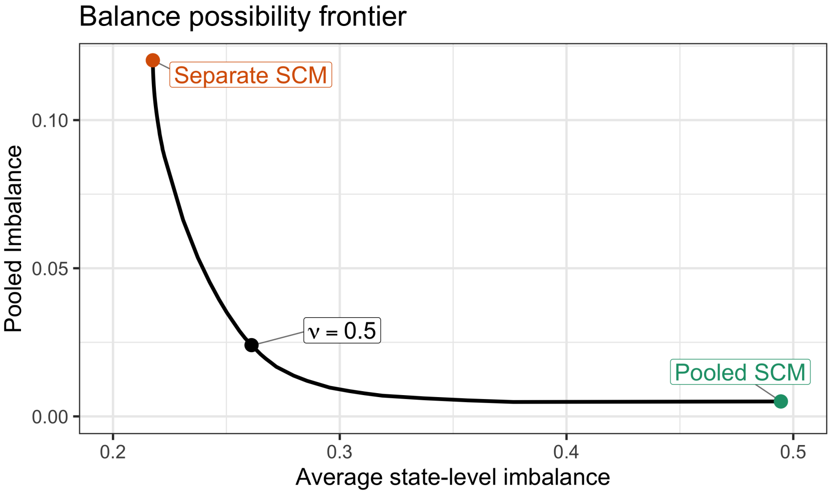
$$\min_{\Gamma} \frac{1}{J} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2$$



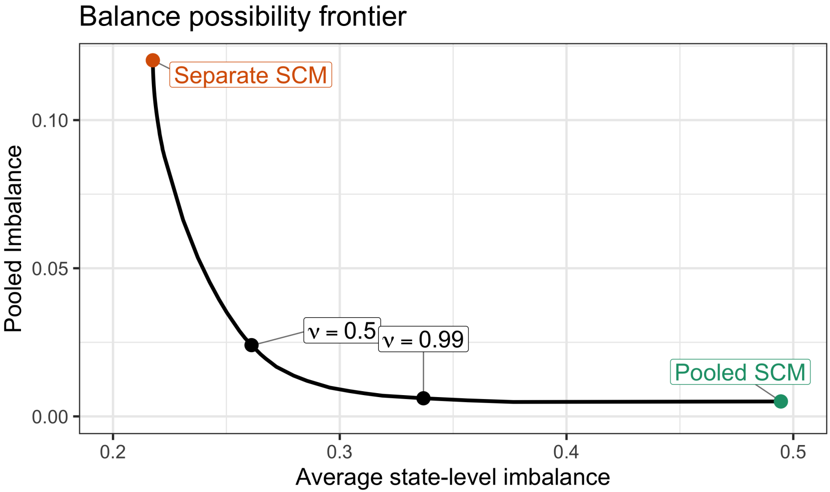
$$\min_{\Gamma} \|\text{Avg Balance}\|_2^2$$



$$\min_{\Gamma} \quad \nu \, \|\text{Avg Balance}\|_2^2 + \frac{1-\nu}{J} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2$$



$$\min_{\Gamma} \quad \nu \,\|\text{Avg Balance}\|_2^2 + \frac{1-\nu}{J}\sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2$$

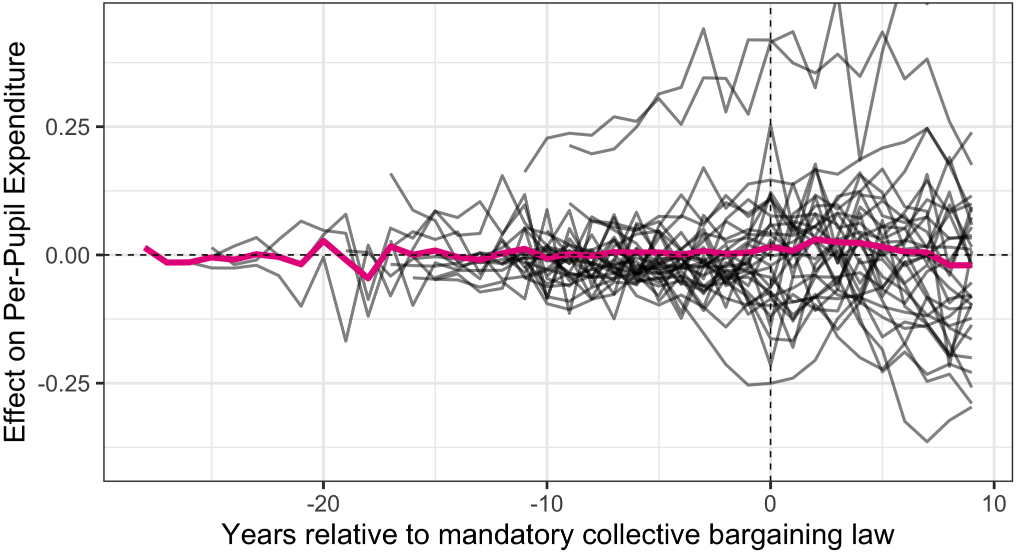


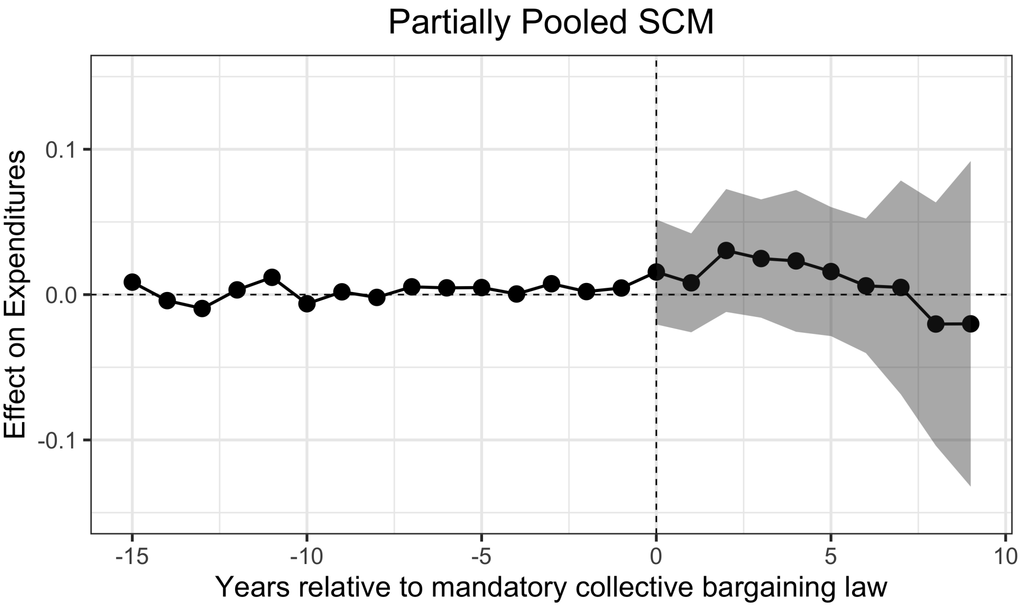
$$\min_{\Gamma} \quad \nu \left\| \text{Avg Balance} \right\|_2^2 + \frac{1-\nu}{J} \sum_{j=1}^{J} \left\| \text{State Balance}_j \right\|_2^2$$

Pooled SCM


Effect on Per-Pupil Expenditure
0.25
0.00
-0.25
-20
-10
0
10
Years relative to mandatory collective bargaining law

Partially Pooled SCM




Partially Pooled SCM
Effect on Expenditures
0.1
0.0
-0.1
-15
-10
-5
0
5
10
Years relative to mandatory collective bargaining law

Intercept Shifts

Intercept-Shifted SCM

Adjust for level differences by adding an intercept to the optimization problem

[Doudchenko and Imbens, 2017; Ferman and Pinto, 2018]

$$\hat{Y}^*_{j,T_j+k}(\infty) = \hat{\alpha}_j + \sum_i \hat{\gamma}^*_{ij} Y_{i,T_j+k}$$

Intercept-Shifted SCM

Adjust for level differences by adding an intercept to the optimization problem [Doudchenko and Imbens, 2017; Ferman and Pinto, 2018]

$$\hat{Y}^*_{j,T_j+k}(\infty) = \hat{\alpha}_j + \sum_i \hat{\gamma}^*_{ij} Y_{i,T_j+k}$$

Solution: De-meaning by pre-treatment average $\bar{Y}^{\text{pre}}_{i,T_j}$

Intercept-Shifted SCM

Adjust for level differences by adding an intercept to the optimization problem
[Doudchenko and Imbens, 2017; Ferman and Pinto, 2018]

$$\hat{Y}^*_{j,T_j+k}(\infty) = \hat{\alpha}_j + \sum_i \hat{\gamma}^*_{ij} Y_{i,T_j+k}$$

Solution: De-meaning by pre-treatment average $\bar{Y}^{\text{pre}}_{i,T_j}$

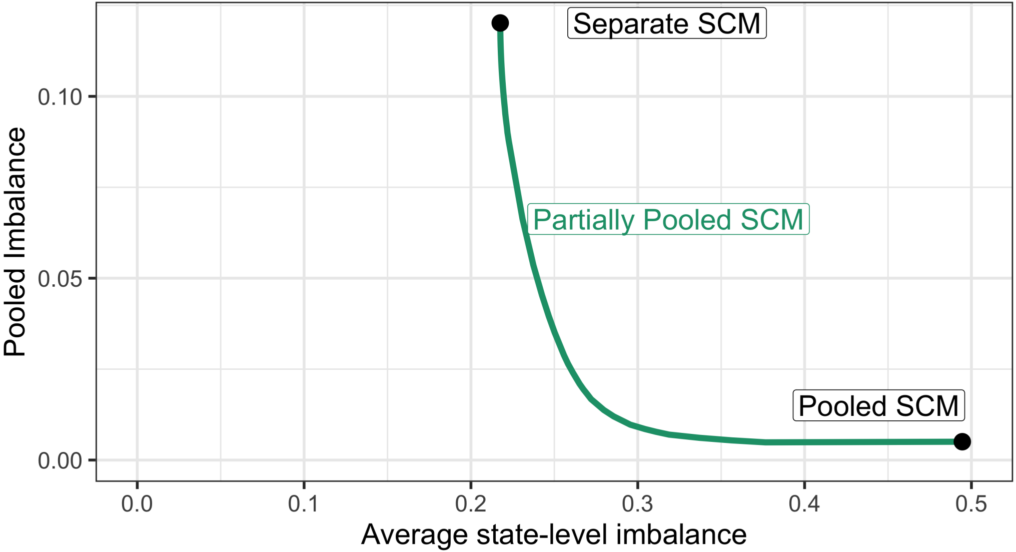
Treatment effect estimate is **weighted difference-in-differences**

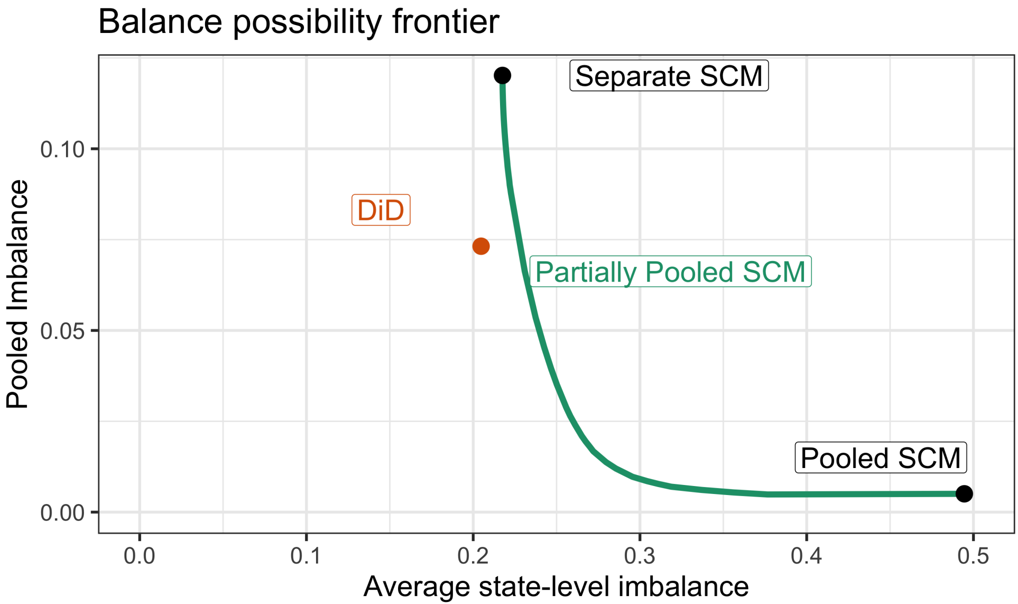
$$\hat{\tau}^{\text{aug}}_{jk} = \left(Y_{j,T_j+k} - \bar{Y}^{\text{pre}}_{j,T_j}\right) - \sum_{i=1}^{N} \hat{\gamma}^*_{ij} \left(Y_{i,T_j+k} - \bar{Y}^{\text{pre}}_{i,T_j}\right)$$

→ Uniform weights recover “stacked” DiD [Abraham and Sun, 2018]

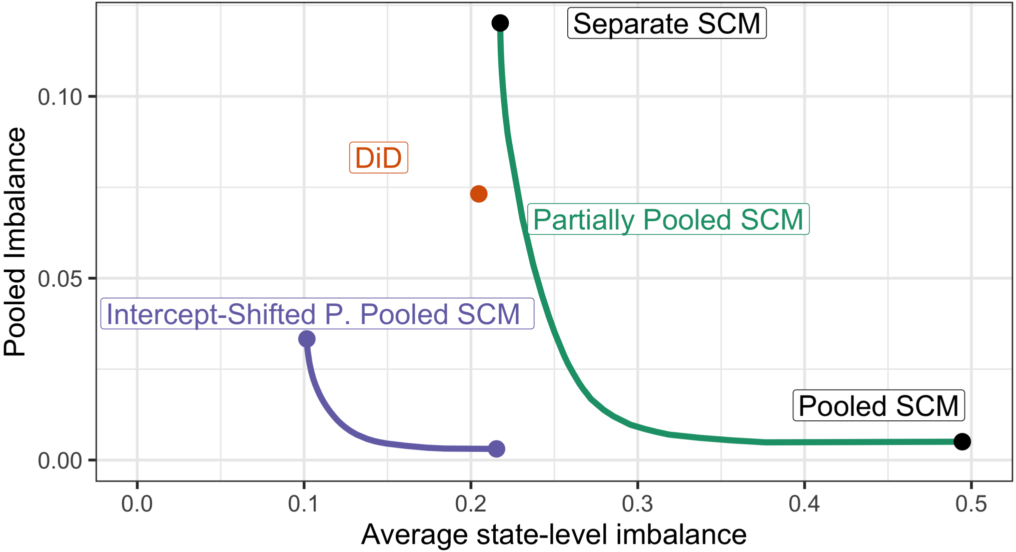
→ Similar in form to P-score weighted DiD [Abadie, 2005; Callaway and Sant’Anna, 2018]

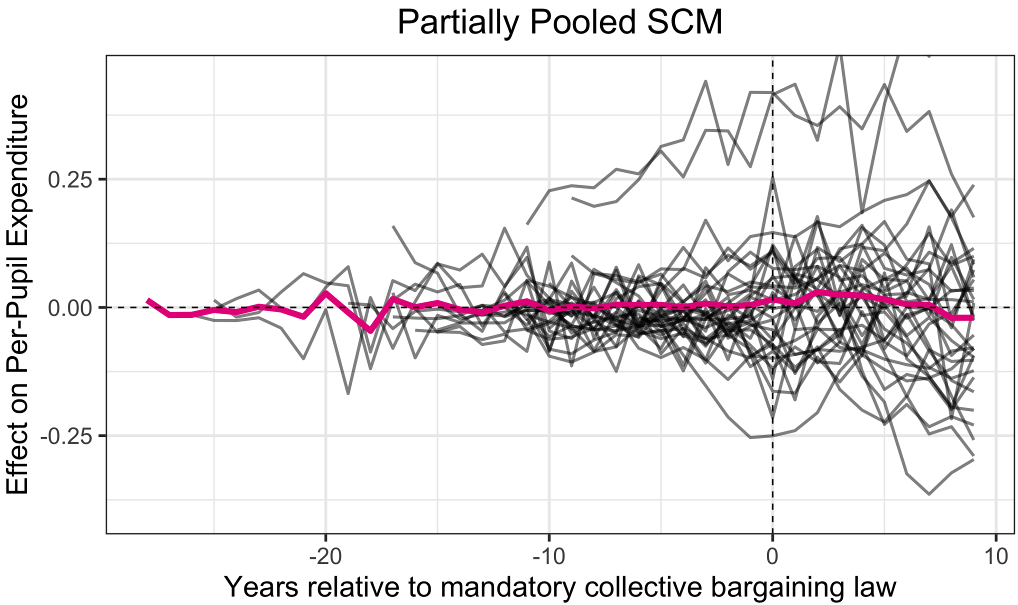
Balance possibility frontier




Balance possibility frontier
Separate SCM
DiD
Partially Pooled SCM
Pooled SCM
Pooled Imbalance
0.10
0.05
0.00
0.0
0.1
0.2
0.3
0.4
0.5
Average state-level imbalance

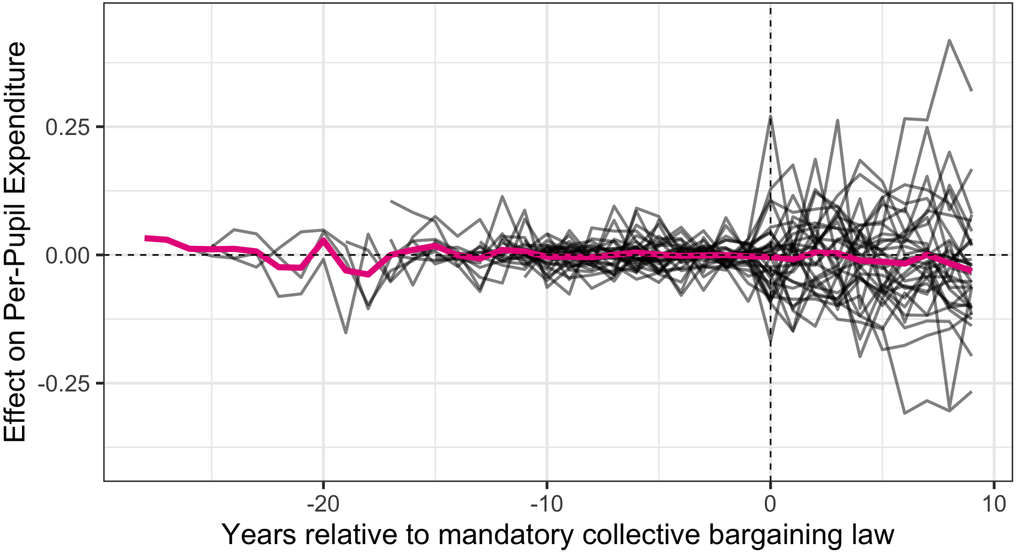
Balance possibility frontier

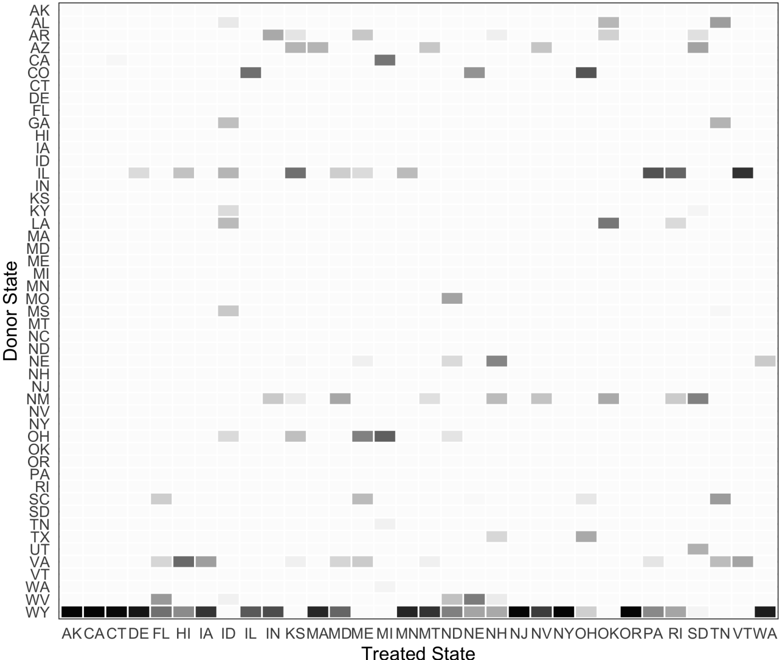


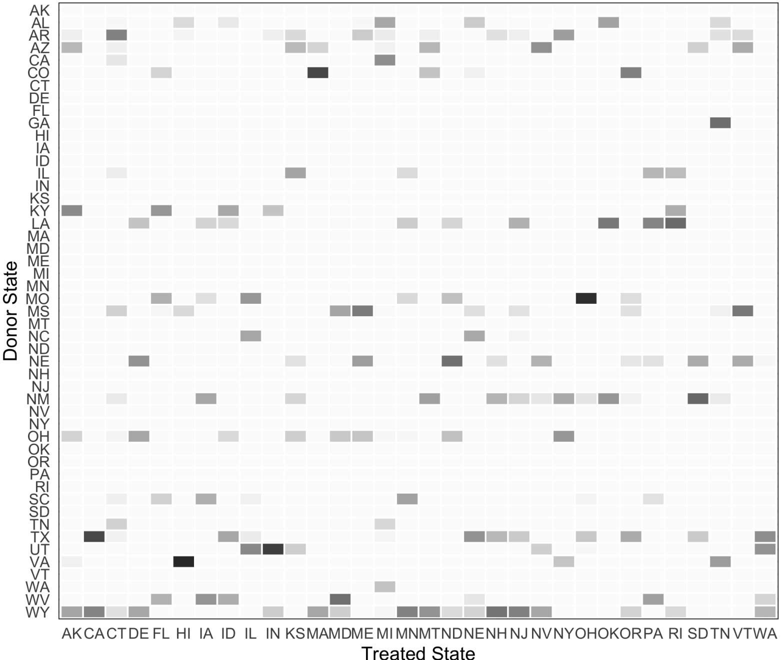


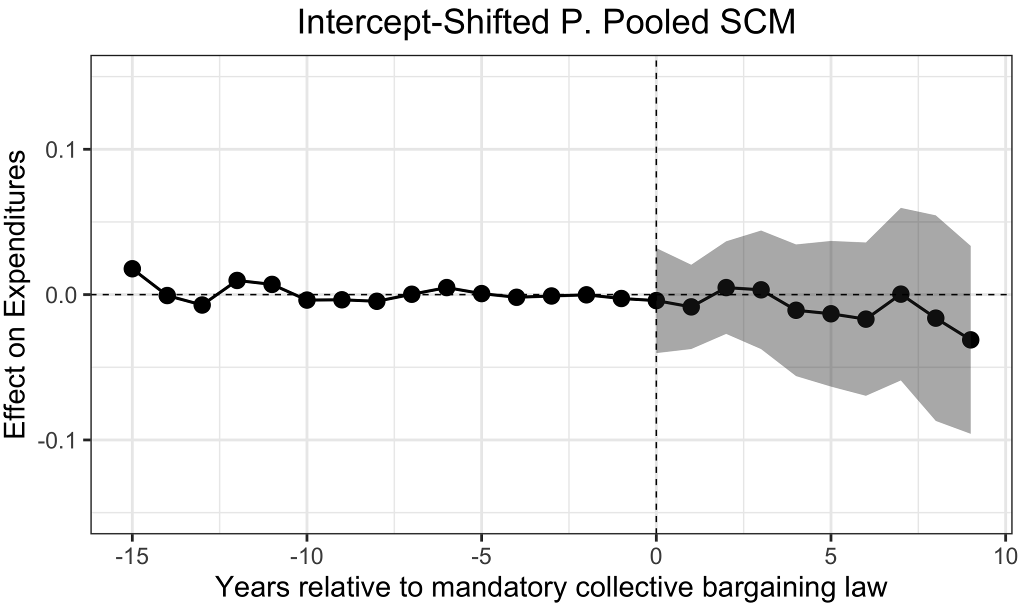
Partially Pooled SCM
Effect on Per-Pupil Expenditure
0.25
0.00
-0.25
-20
-10
0
10
Years relative to mandatory collective bargaining law

Intercept-Shifted P. Pooled SCM




Donor State
AK
AL
AR
AZ
CA
CO
CT
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
RI
SC
SD
TN
TX
UT
VA
VT
WA
WV
WY
AK CA CT DE FL HI IA ID IL IN KS MA MD ME MI MN MT ND NE NH NJ NV NY OH OK OR PA RI SD TN VT WA
Treated State


Donor State
AK AL AR AZ CA CO CT DE FL GA HI IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WV WY
Treated State
AK CA CT DE FL HI IA ID IL IN KS MA MD ME MI MN MT ND NE NH NJ NV NY OH OK OR PA RI SD TN VT WA


Intercept-Shifted P. Pooled SCM
Effect on Expenditures
0.1
0.0
-0.1
-15
-10
-5
0
5
10
Years relative to mandatory collective bargaining law

Recap

This paper: Extend SCM to staggered adoption

- Find weights that control State Balance and Avg Balance
- Include an intercept to adjust for level differences
- *Under the hood:* Dual shrinkage; connection to (generalized) IPW

Recap

This paper: Extend SCM to staggered adoption

- Find weights that control State Balance and Avg Balance
- Include an intercept to adjust for level differences
- *Under the hood:* Dual shrinkage; connection to (generalized) IPW

Extras:

- Incorporating auxiliary covariates
- Weighted bootstrap confidence intervals

In progress:

- Extend to unbalanced panels
- Sensitivity analysis

Recap

This paper: Extend SCM to staggered adoption

- Find weights that control State Balance and Avg Balance
- Include an intercept to adjust for level differences
- *Under the hood:* Dual shrinkage; connection to (generalized) IPW

Extras:

- Incorporating auxiliary covariates
- Weighted bootstrap confidence intervals

In progress:

- Extend to unbalanced panels
- Sensitivity analysis

Thank you!

https://arxiv.org/abs/1912.03290

https://github.com/ebenmichael/augsynth

Appendix

Random effects AR simulation: level of pooling really matters

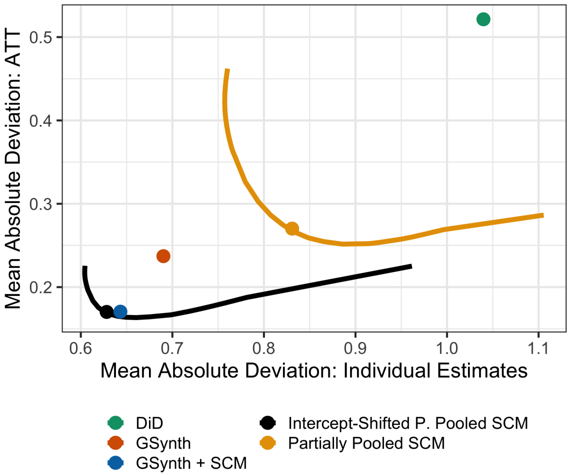
Calibrated sim study: Random Effects AR

- Fit random effects model
  [Gelman and Hill, 2007]

$$Y_{it} = \sum_{k=1}^{3} \rho_{tk} Y_{i(t-k)} + \varepsilon_{it}$$

$$\rho_t \sim N(\bar{\rho}, \Sigma)$$

- $\pi_i = \text{logit}\left(\theta_0 + \theta_1 \sum_{k=-3}^{1} Y_{i(t-k)}\right)$



DGP is FE Model: Intercept-Shifting + Partial Pooling performs well

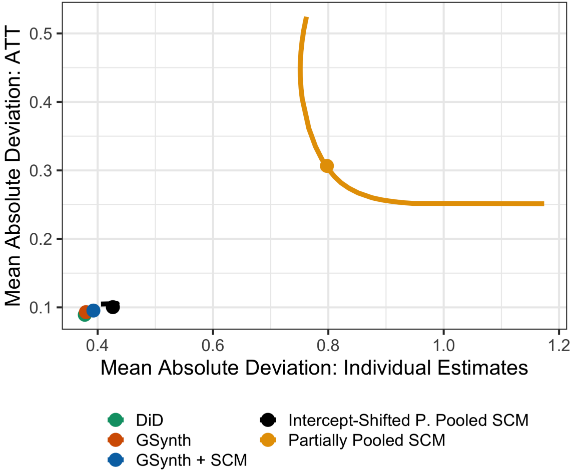
Calibrated sim study: FE

- Fit FE model

$$Y_{it} = \text{unit}_i + \text{time}_t + \varepsilon_{it}$$

- $\text{unit}_i \sim \widehat{\text{Normal}}$
- $\pi_i = \text{logit}(\theta_0 + \theta_1 \cdot \text{unit}_i)$

Event study is correct model



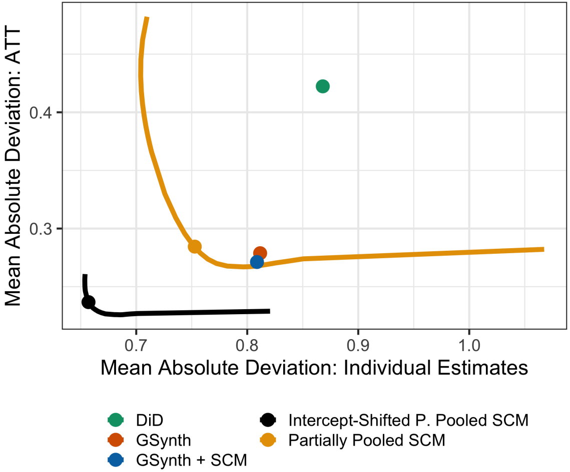
DGP is Factor Model: Intercept-Shifting + Partial Pooling does best

Calibrated sim study: Factor

- Fit gsynth [Xu, 2017]

$$Y_{it} = \text{unit}_i + \text{time}_t + \phi_i' \mu_t + \varepsilon_{it}$$

- $\{\text{unit}_i, \phi_i\} \sim \widehat{\text{MVN}}$
- $\pi_i = \text{logit}(\theta_0 + \theta_1(\text{unit}_i + \phi_{i1} + \phi_{i2}))$



Heuristic for ν: fit with $\nu = 0$ then choose

$$\hat{\nu} = \frac{\frac{1}{\sqrt{L}} \left\| \text{Avg Balance} \right\|_2}{\sqrt{\frac{1}{J} \sum_{j=1}^{J} \left\| \text{State Balance}_j \right\|_2^2}}$$

References I

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19.

Abraham, S. and Sun, L. (2018). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects.

Athey, S. and Imbens, G. W. (2018). Design-based analysis in difference-in-differences settings with staggered adoption. Technical report, National Bureau of Economic Research.

Callaway, B. and Sant'Anna, P. H. C. (2018). Difference-in-Differences With Multiple Time Periods and an Application on the Minimum Wage and Employment.

Doudchenko, N. and Imbens, G. W. (2017). Difference-In-Differences and Synthetic Control Methods: A Synthesis. *arxiv 1610.07748*.

Ferman, B. and Pinto, C. (2018). Synthetic controls with imperfect pre-treatment fit.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierachical Models*.

References II

Hoxby, C. M. (1996). How teachers' unions affect education production. *The Quarterly Journal of Economics*, 111(3):671–718.

Imai, K. and Kim, I. S. (2019). On the use of two-way fixed effects regression models for causal inference with panel data.

Paglayan, A. S. (2019). Public-sector unions and the size of government. *American Journal of Political Science*, 63(1):21–36.

Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Political Analysis*, 25:57–76.