

Synthetic Controls and Weighted Event Studies with Staggered Adoption

Eli Ben-Michael, Avi Feller, and Jesse Rothstein

UC Berkeley

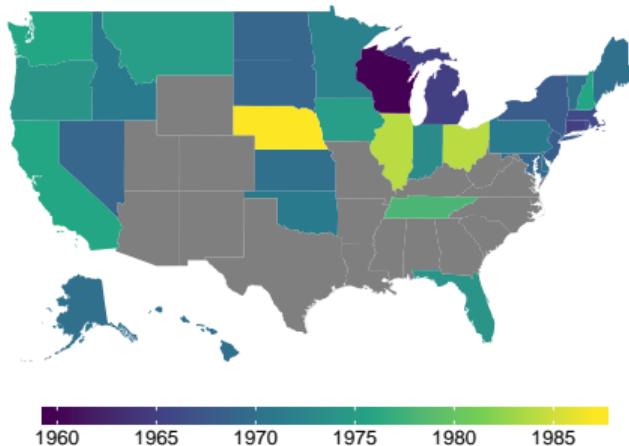
JSM 2020

August 2020

What is the impact of teacher unions on salaries and spending?

- 1964 - 1987: 33 states pass mandatory collective bargaining laws

Year of Mandatory Collective Bargaining Law

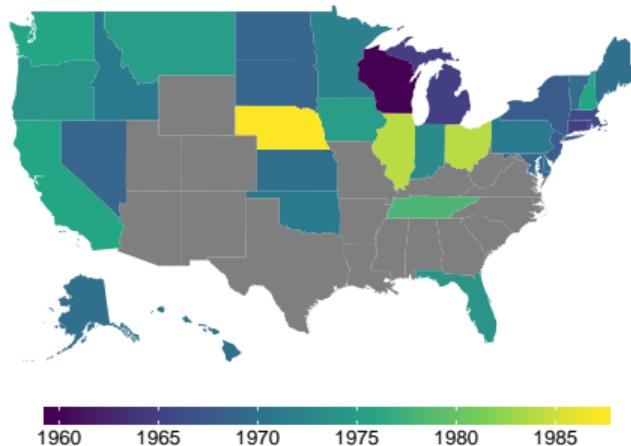


What is the impact of teacher unions on salaries and spending?

- 1964 - 1987: 33 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]

Year of Mandatory Collective Bargaining Law



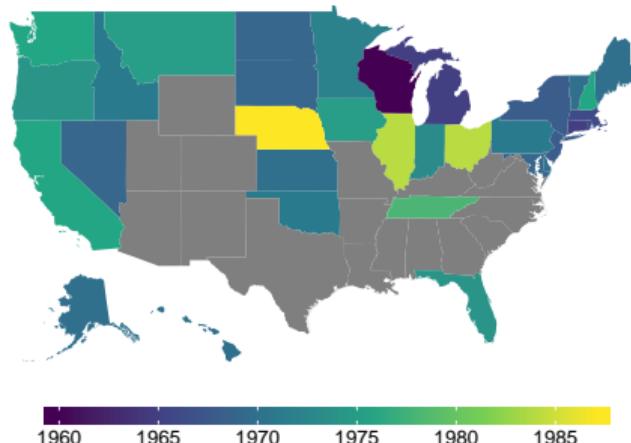
What is the impact of teacher unions on salaries and spending?

- 1964 - 1987: 33 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?

Year of Mandatory Collective Bargaining Law



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

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

Weighted Event Study

- Combine outcome modeling/DiD and SCM

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, ∞

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, ∞

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]

Basic building block:

$$\tau_{jk} = Y_{j,T_j+k}(1) - \underbrace{Y_{j,T_j+k}(0)}_{\sum \hat{\gamma}_{ij} Y_{i,T_j+k}}$$

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

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, ∞

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]

Basic building block:

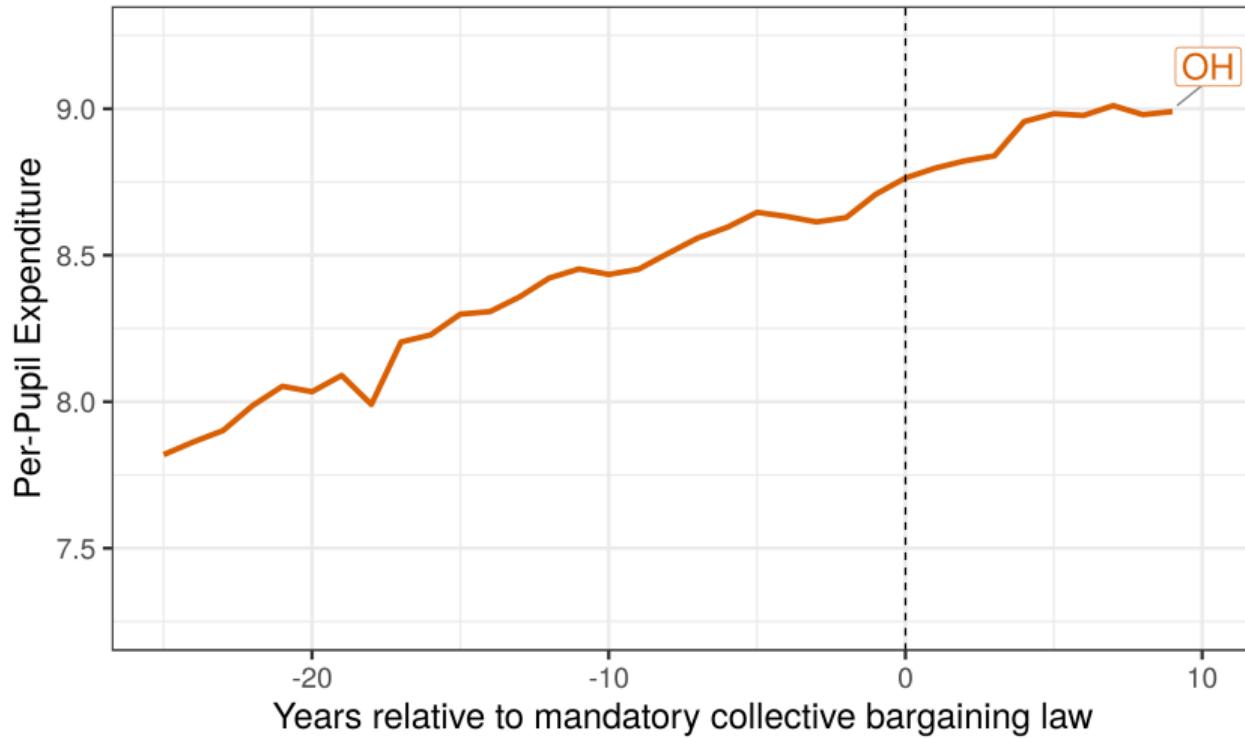
$$\tau_{jk} = Y_{j,T_j+k}(1) - \underbrace{Y_{j,T_j+k}(0)}_{\sum \hat{\gamma}_{ij} Y_{i,T_j+k}}$$

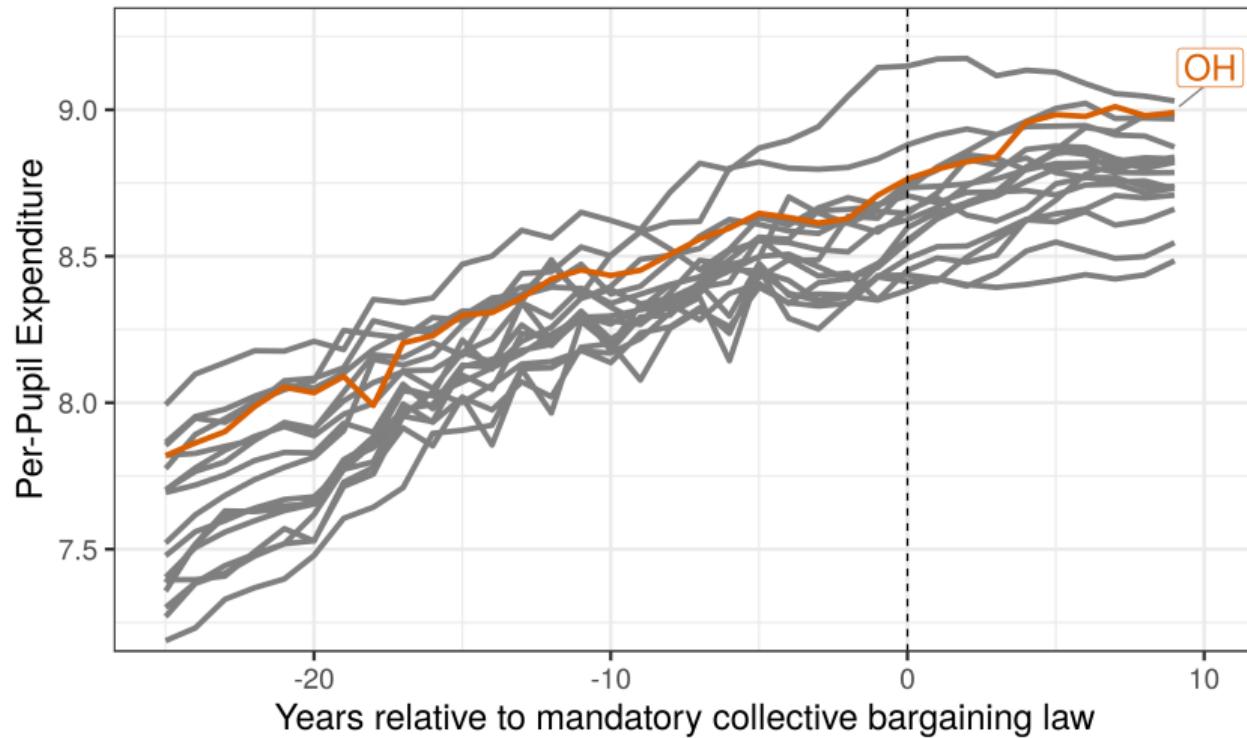
treat =
$$\begin{pmatrix} \checkmark & \checkmark & \checkmark \\ & \checkmark & \checkmark \\ & & \checkmark \end{pmatrix}$$

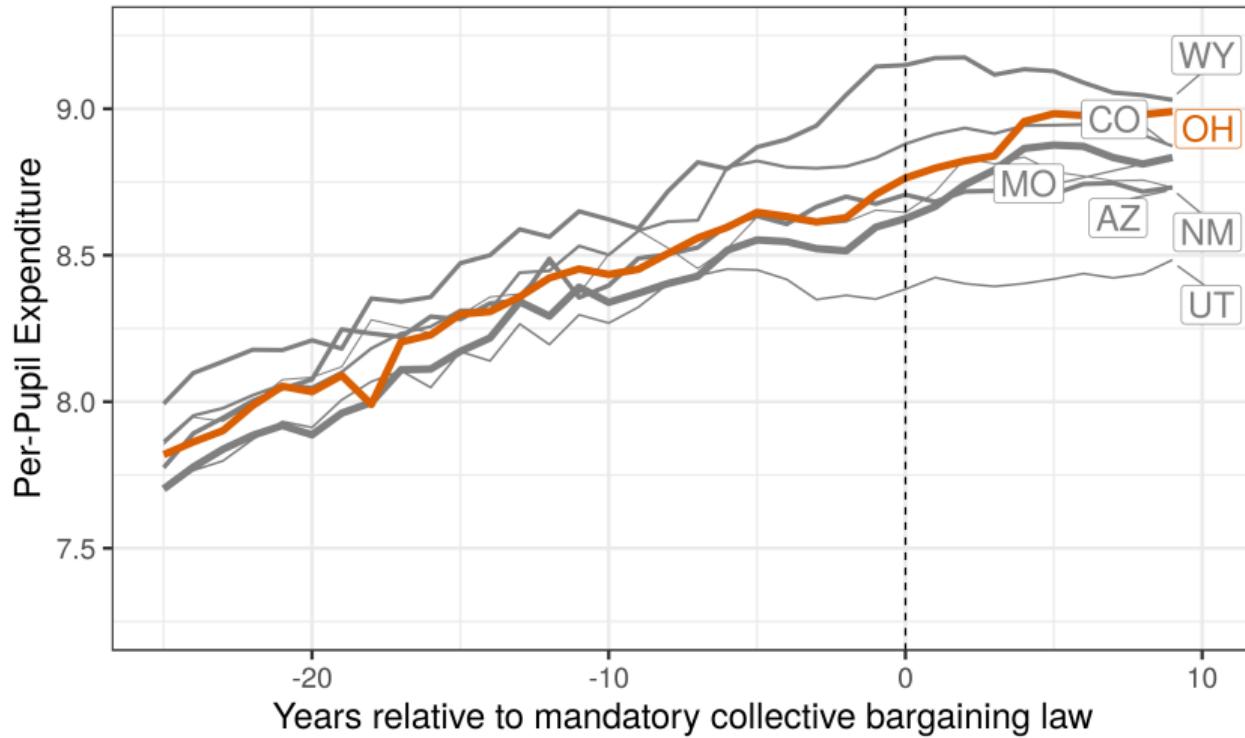
Average at event time k :

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

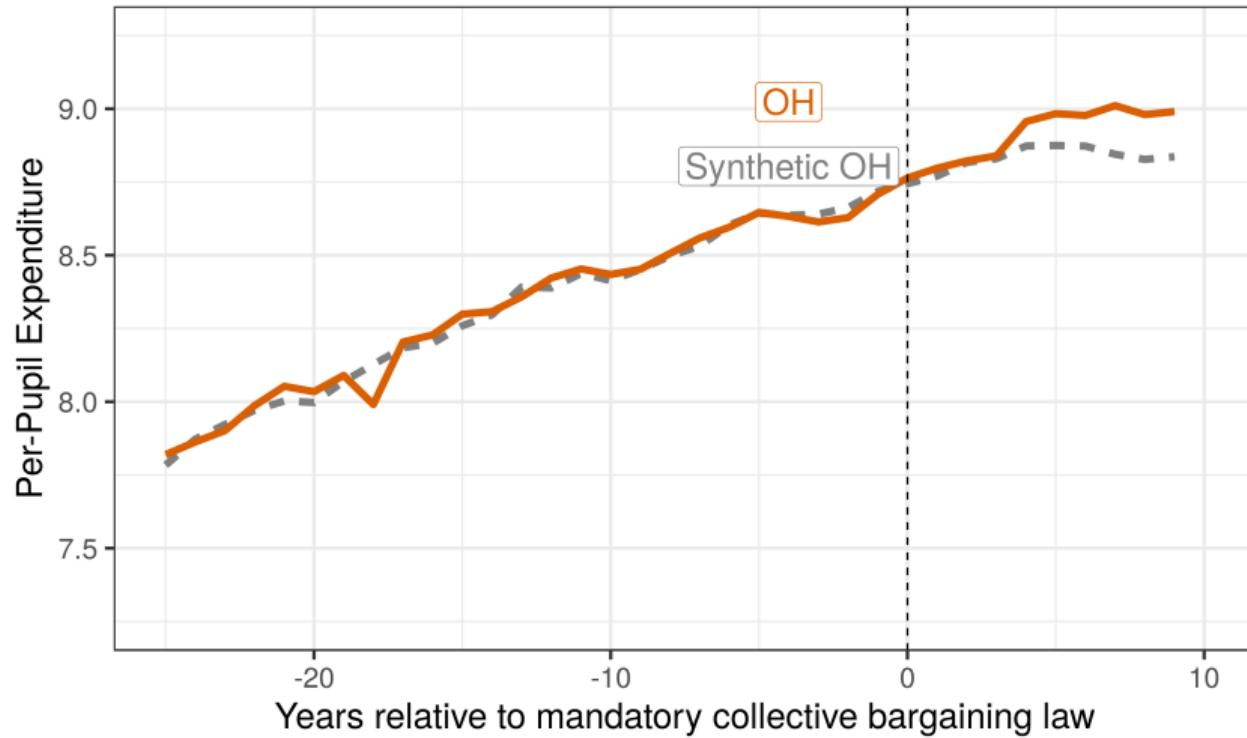
Separate
SCM



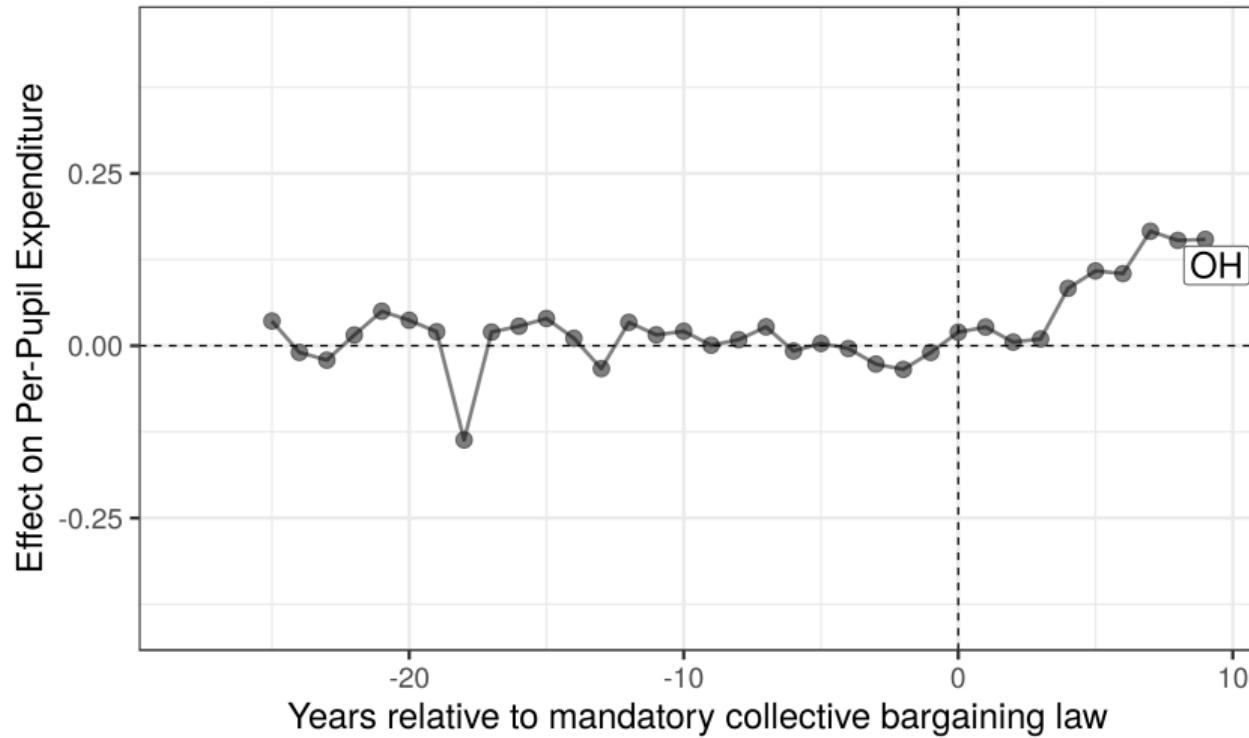




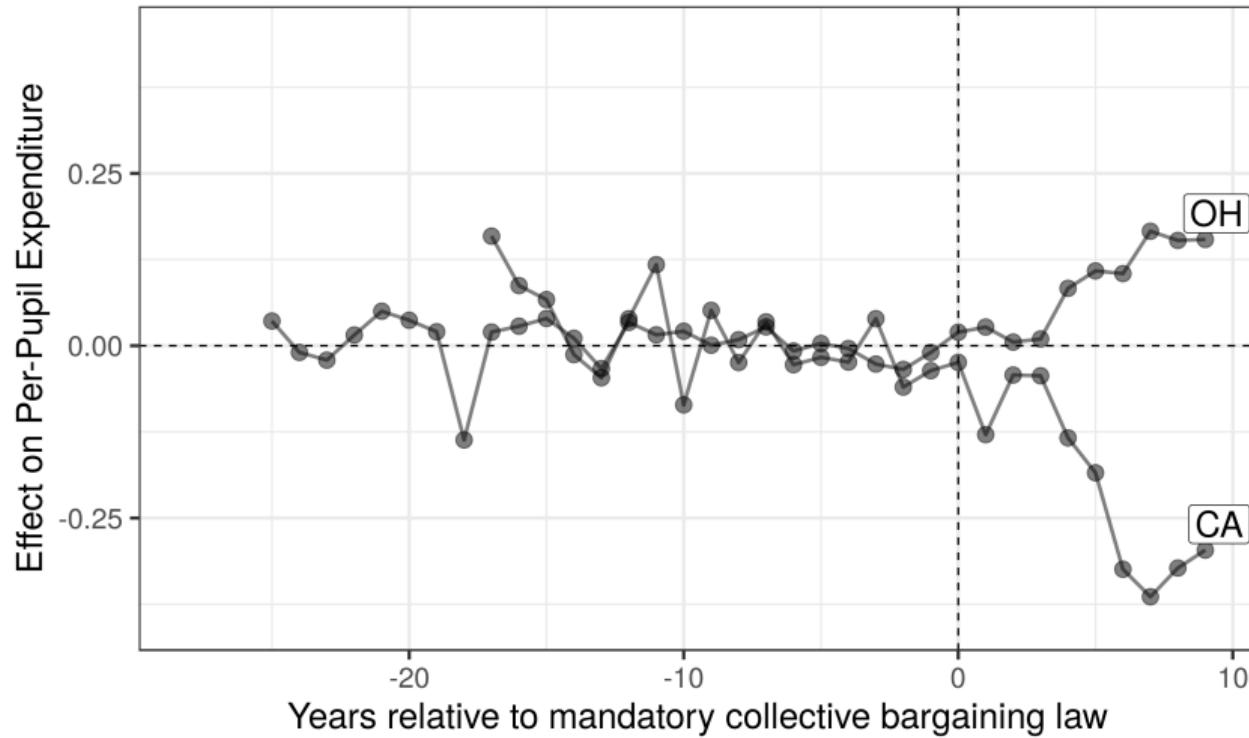
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \|\text{State Balance}_j\|_2^2$$



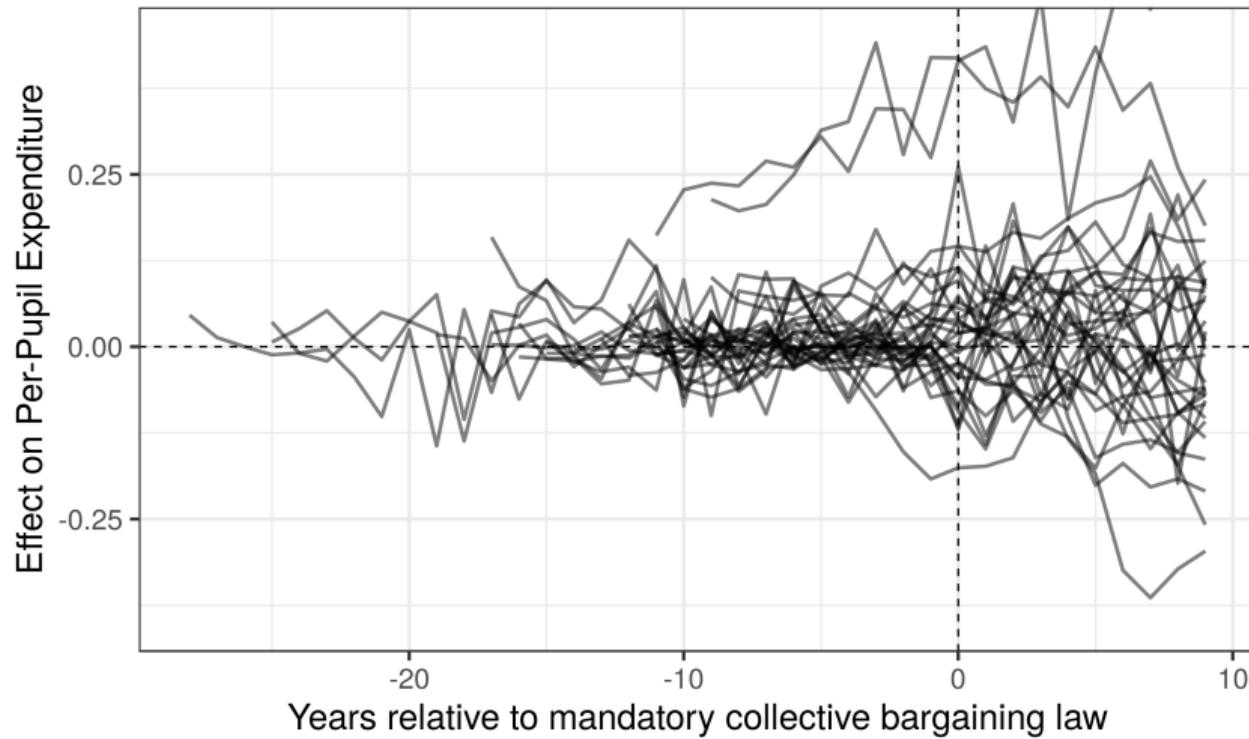
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \|\text{State Balance}_j\|_2^2$$



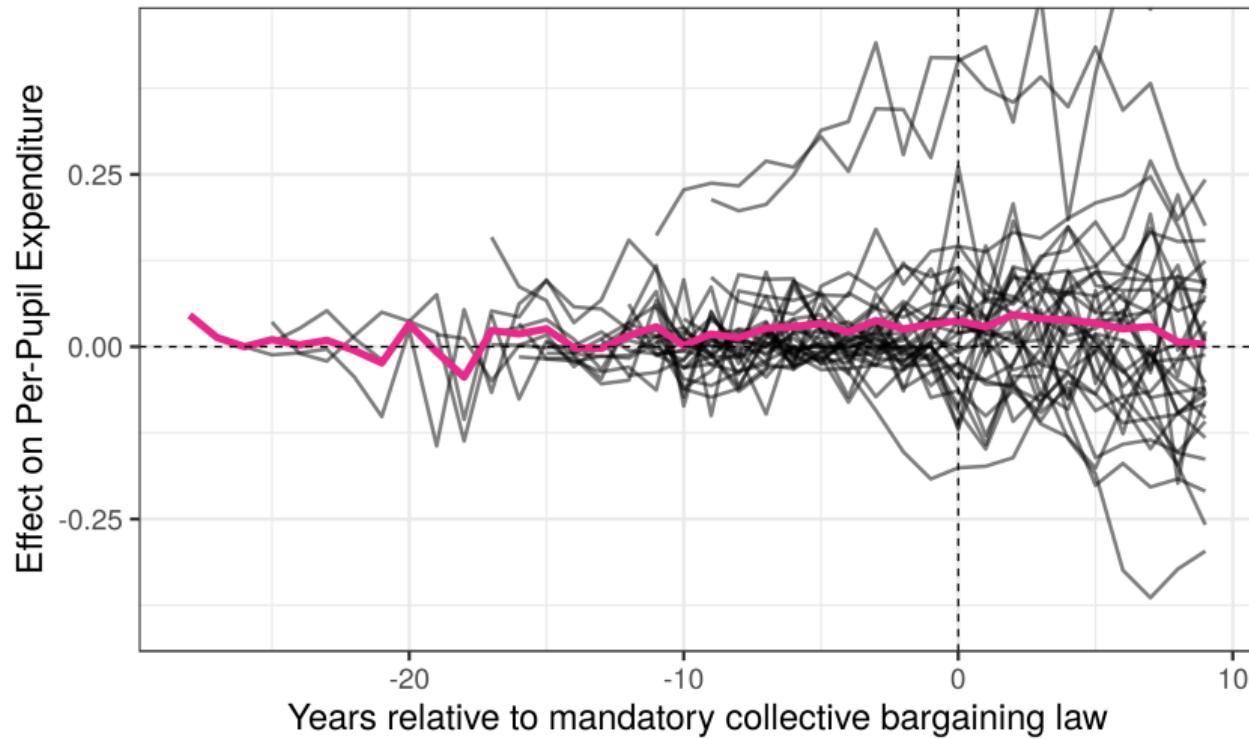
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \|\text{State Balance}_j\|_2^2$$



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

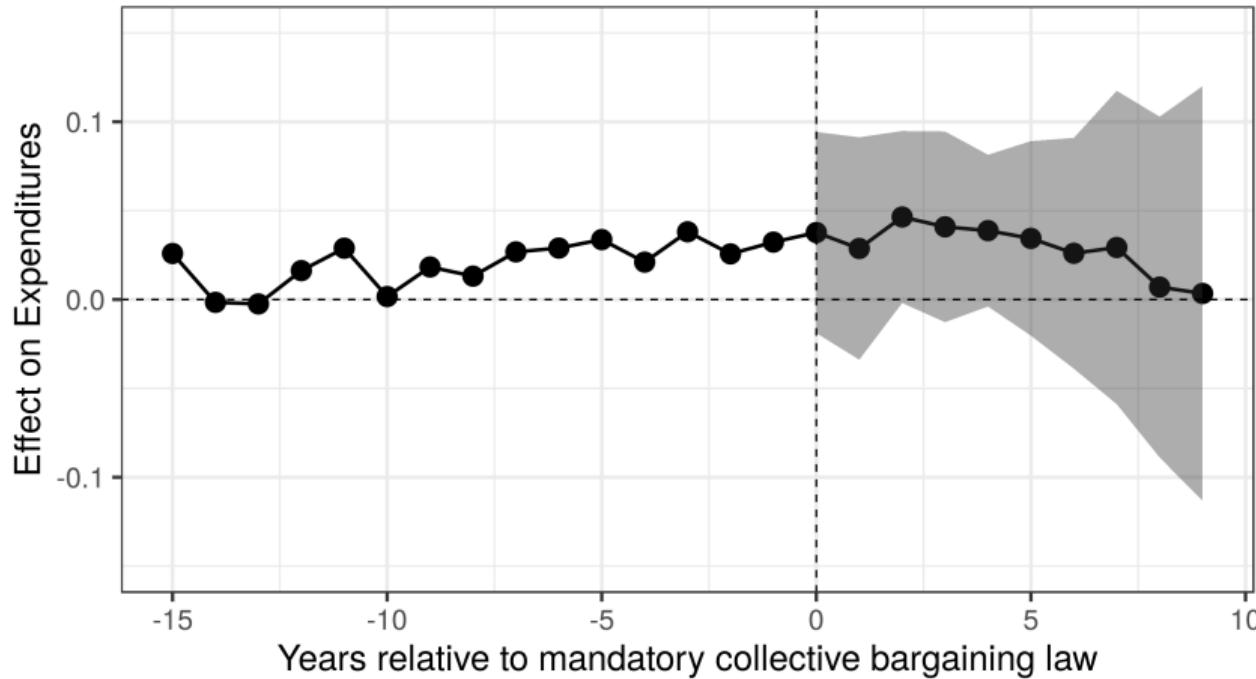


$$\min_{\gamma_1, \dots, \gamma_J \in \Delta^{\text{scm}}} \frac{1}{J} \sum_{j=1}^J \|\text{State Balance}_j\|_2^2$$



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

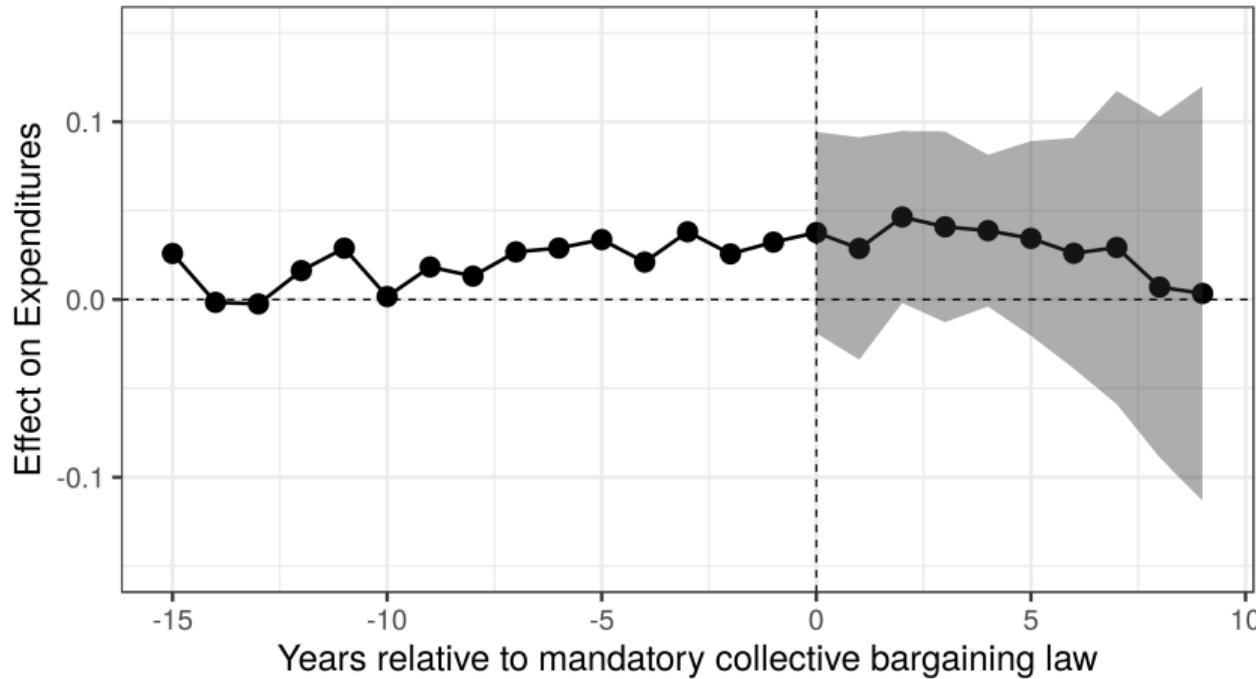
Separate SCM



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

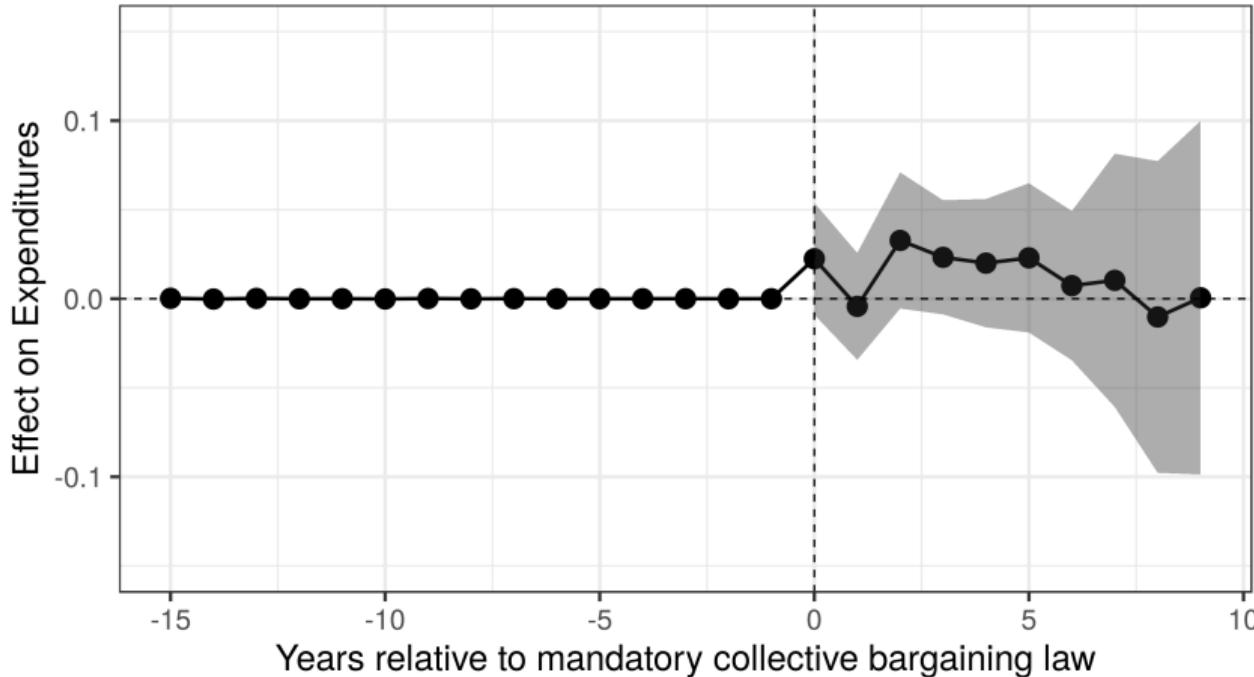
Moving Beyond Separate SCM

Separate SCM



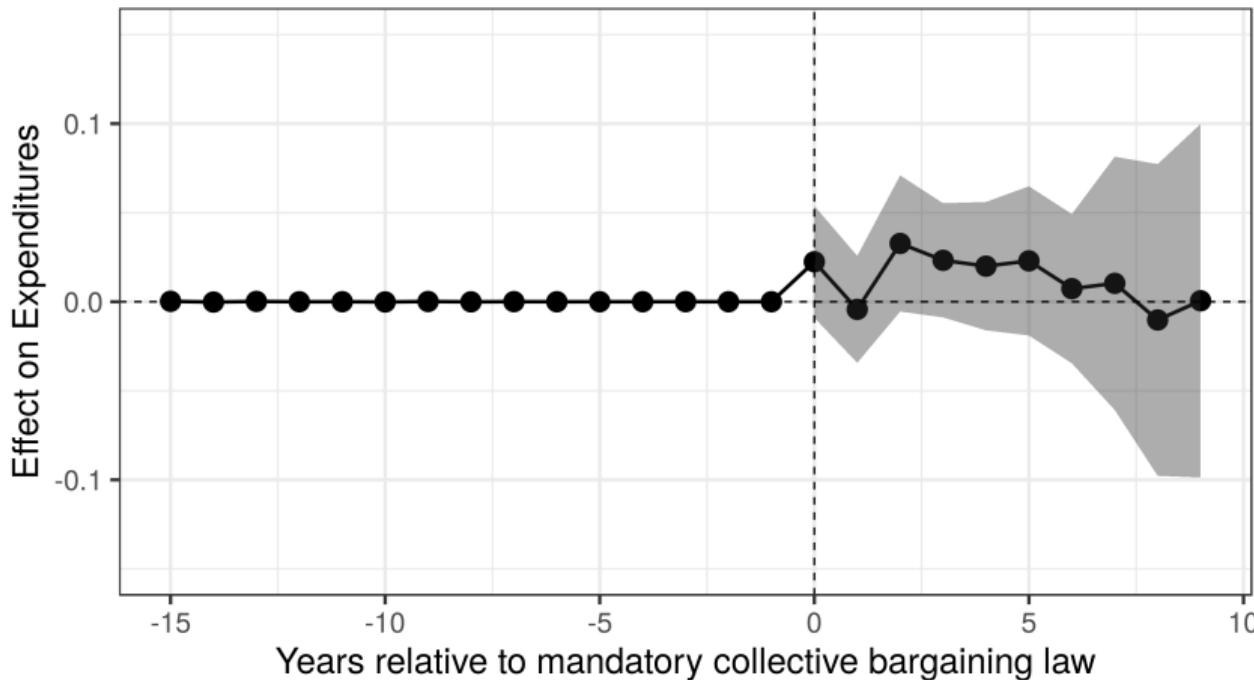
$$\min_{\Gamma} \frac{1}{J} \sum_{j=1}^J \|\text{State Balance}_j\|_2^2$$

Pooled SCM



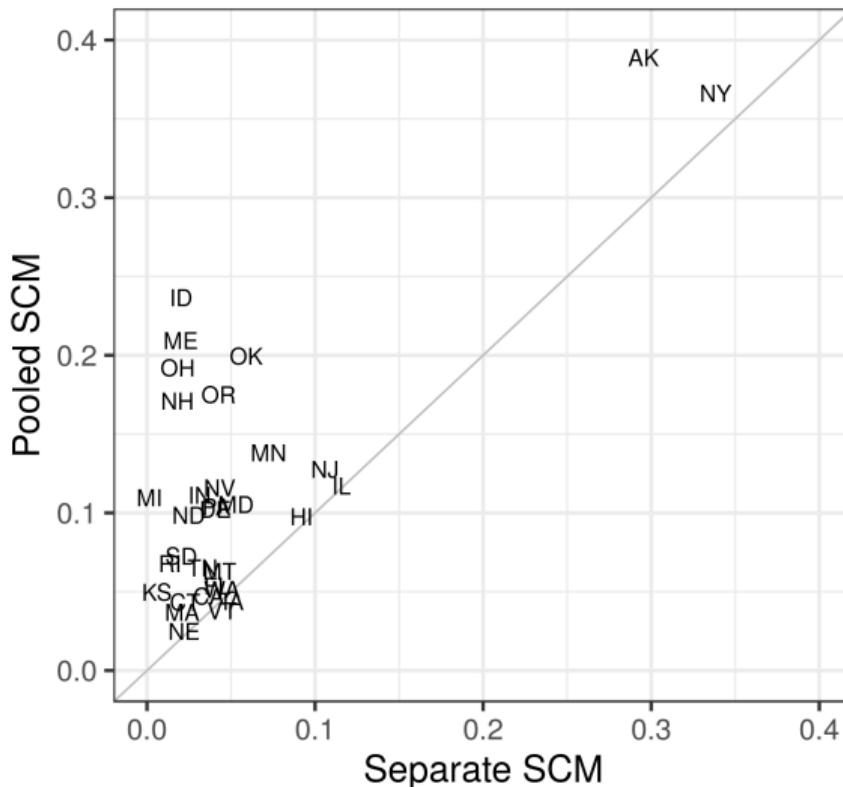
$$\min_{\Gamma} \left\| \frac{1}{J} \sum_{j=1}^J \text{State Balance}_j \right\|_2^2$$

Pooled SCM



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

SCM pre-treatment imbalance by state



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

Which matters more?

Generalization of parallel trends: Linear Factor Model

$$Y_{it}(0) = \phi_i' \mu_t + \varepsilon_{it}$$

Which matters more?

Generalization of parallel trends: Linear Factor Model

$$Y_{it}(0) = \phi'_i \mu_t + \varepsilon_{it}$$

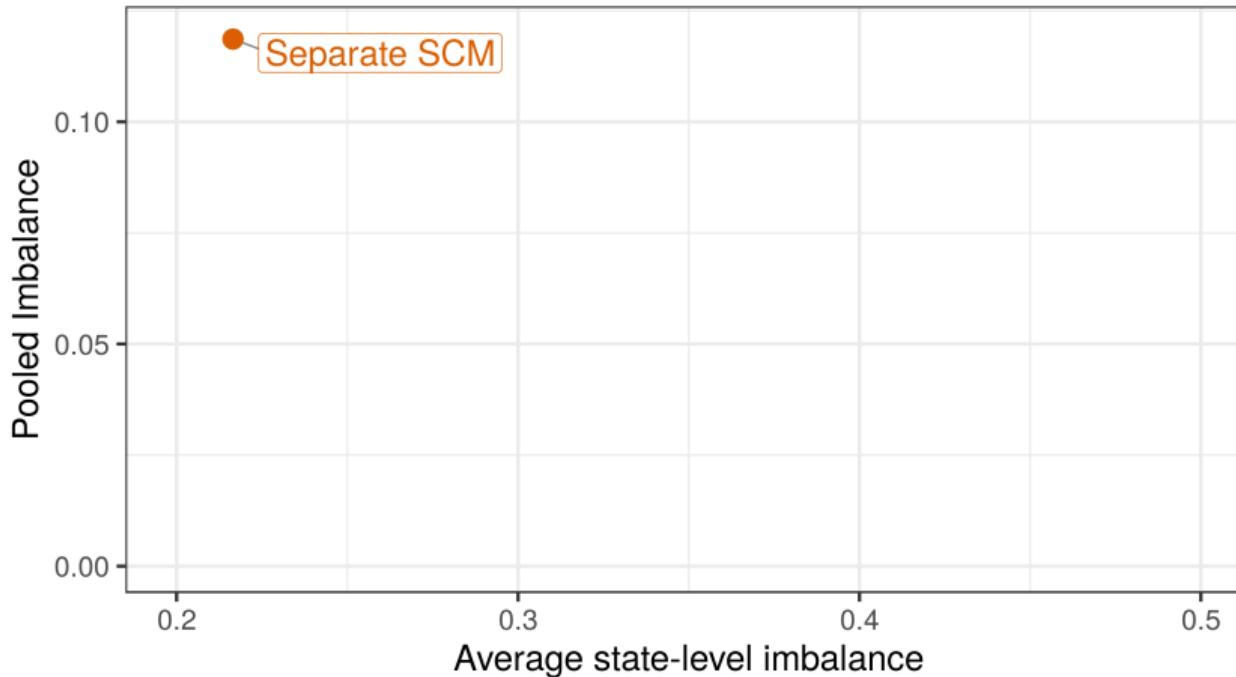
Error for ATT

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

Level of **heterogeneity over time** is important

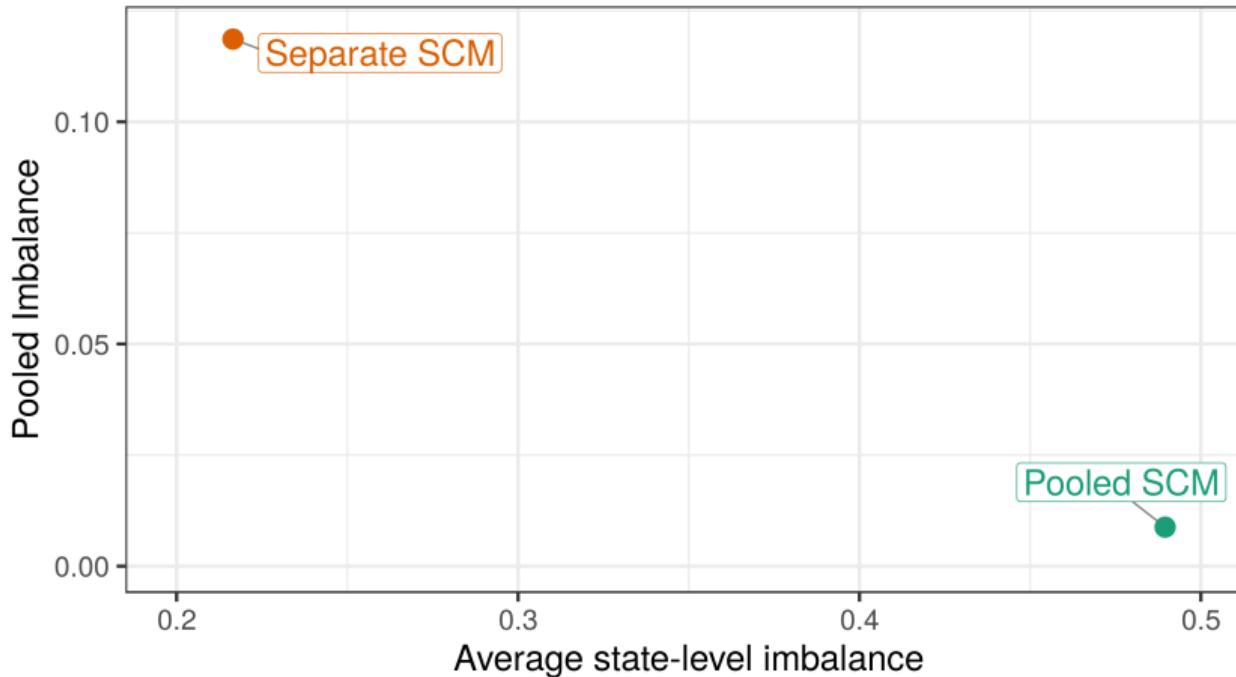
- $\bar{\mu}$ is the **average factor value** → importance of **Avg Balance**
- S is the **factor standard deviation** → importance of **State Balance**
- Special case: unit fixed effects, only **Avg Balance** matters

Balance possibility frontier



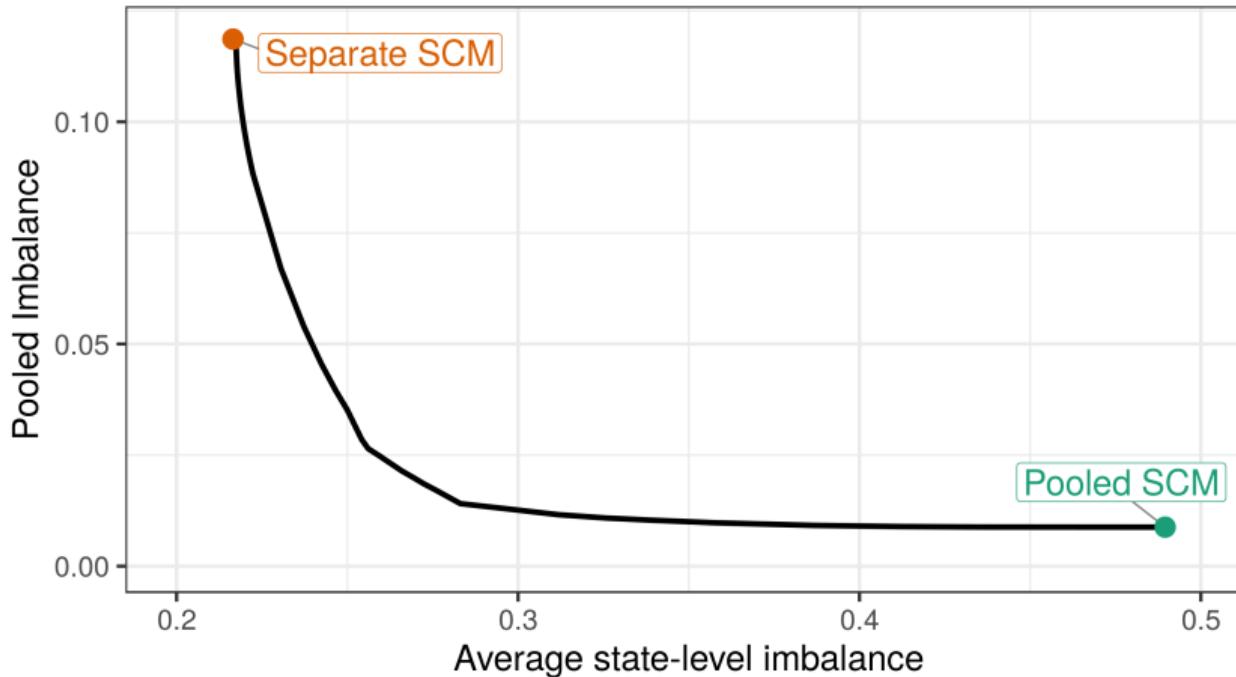
$$\min_{\Gamma} \frac{1}{J} \sum_{j=1}^J \|\text{State Balance}_j\|_2^2$$

Balance possibility frontier



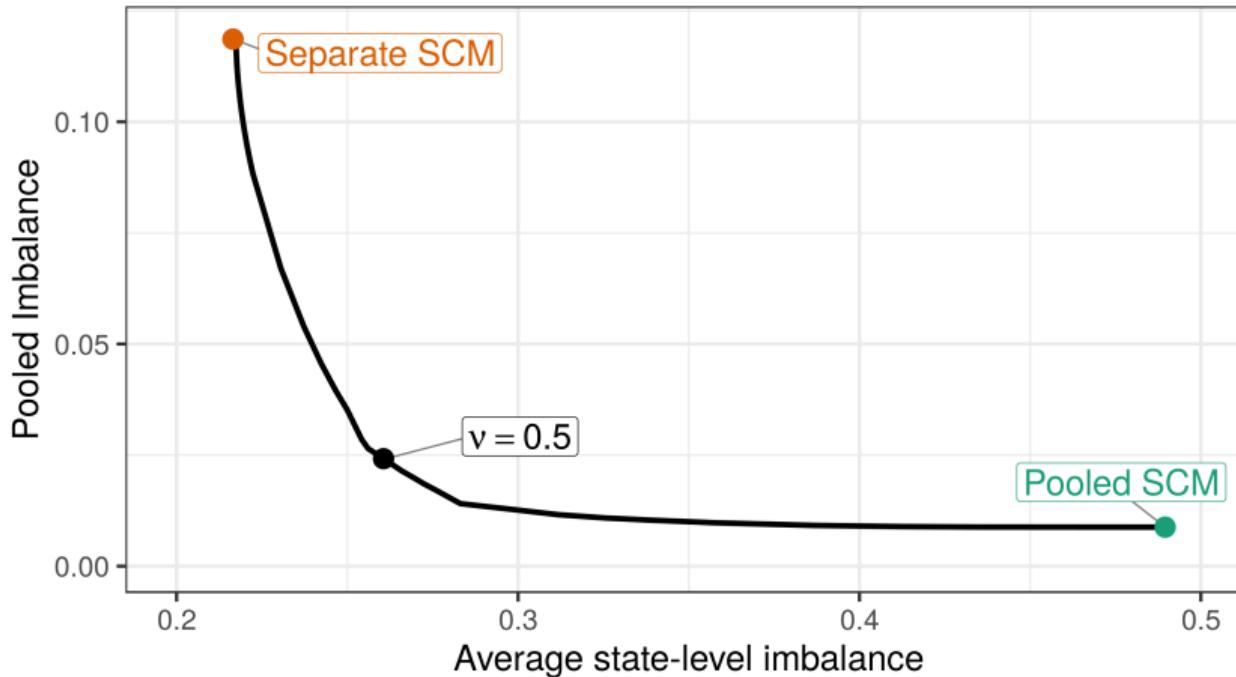
$$\min_{\Gamma} \|\text{Avg Balance}\|_2^2$$

Balance possibility frontier



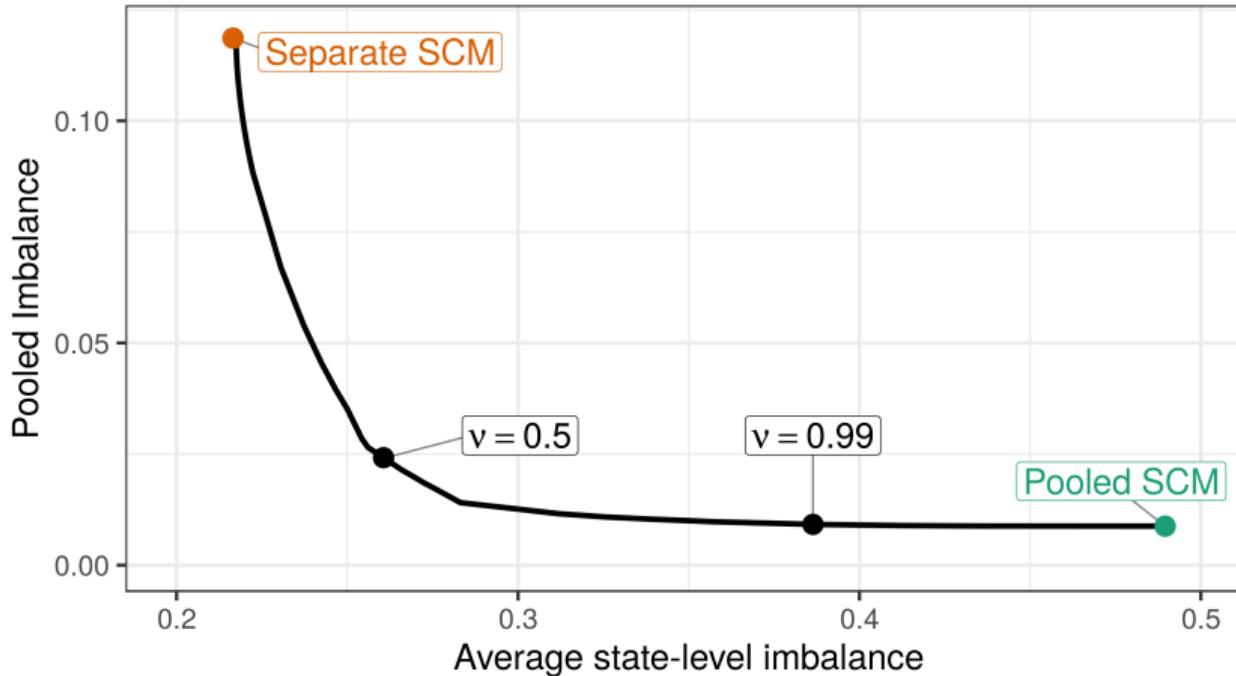
$$\min_{\Gamma} \quad \nu \|\text{Avg Balance}\|_2^2 + \frac{1-\nu}{J} \sum_{j=1}^J \|\text{State Balance}_j\|_2^2$$

Balance possibility frontier



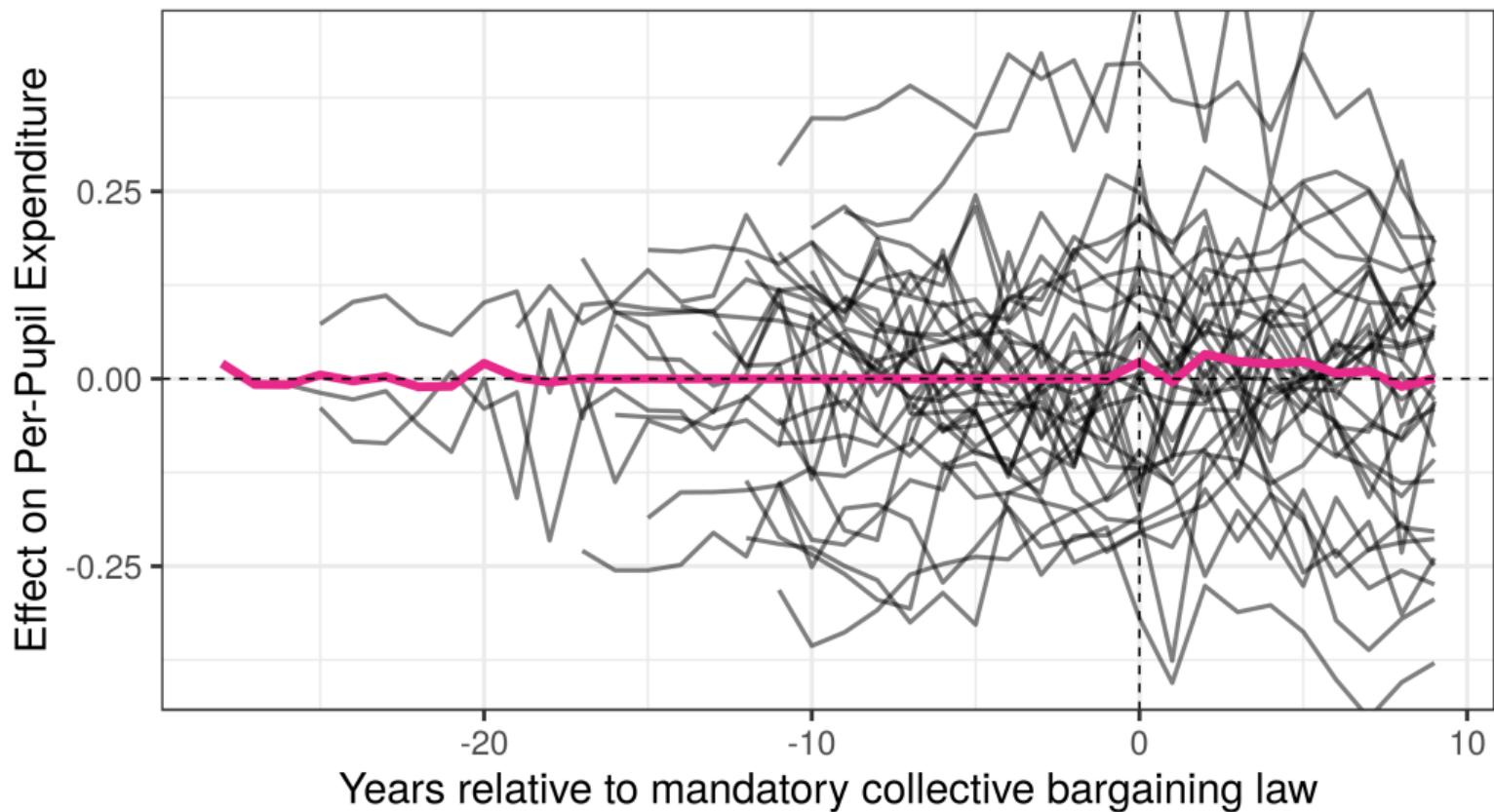
$$\min_{\Gamma} \nu \|\text{Avg Balance}\|_2^2 + \frac{1-\nu}{J} \sum_{j=1}^J \|\text{State Balance}_j\|_2^2$$

Balance possibility frontier

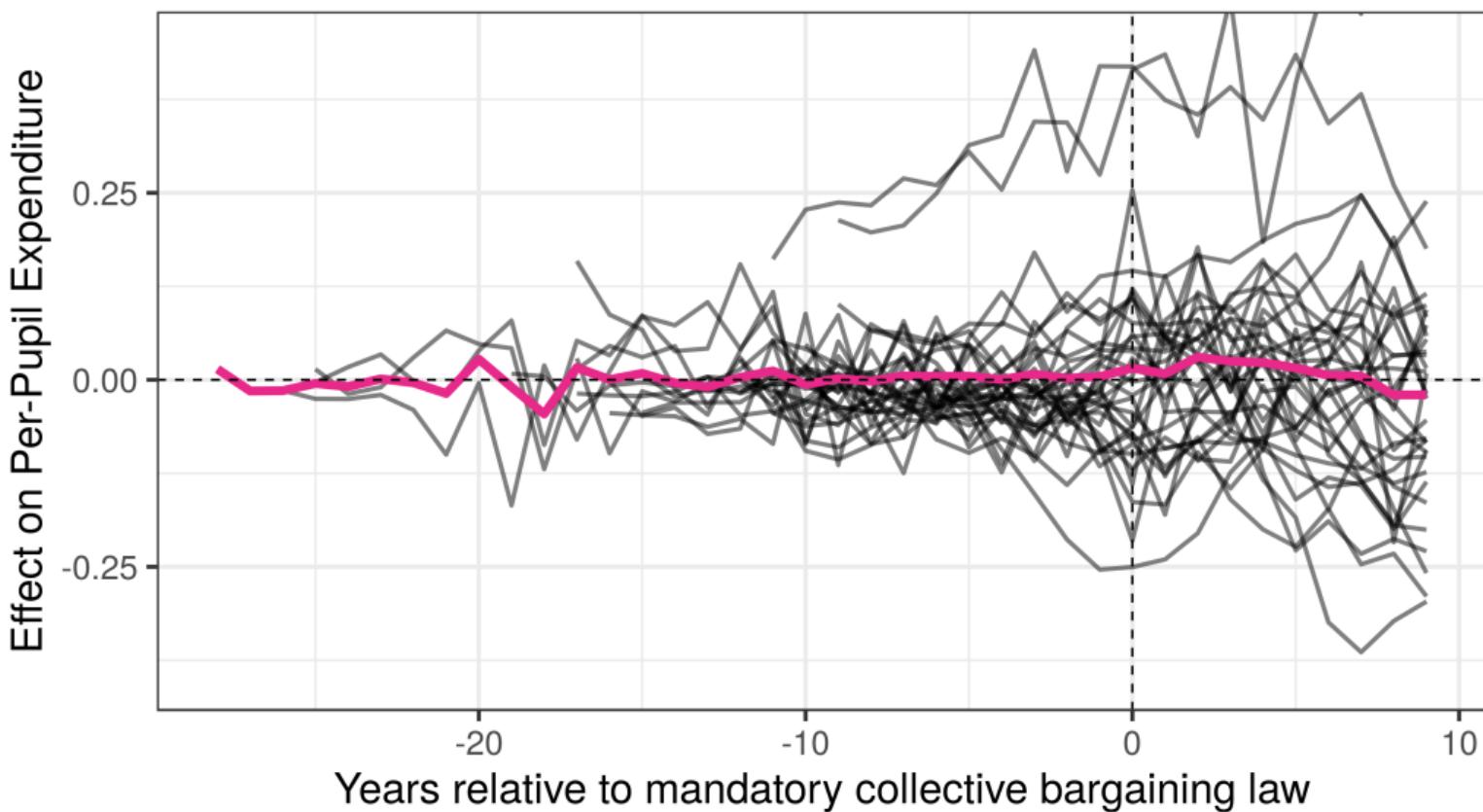


$$\min_{\Gamma} \quad \nu \|\text{Avg Balance}\|_2^2 + \frac{1-\nu}{J} \sum_{j=1}^J \|\text{State Balance}_j\|_2^2$$

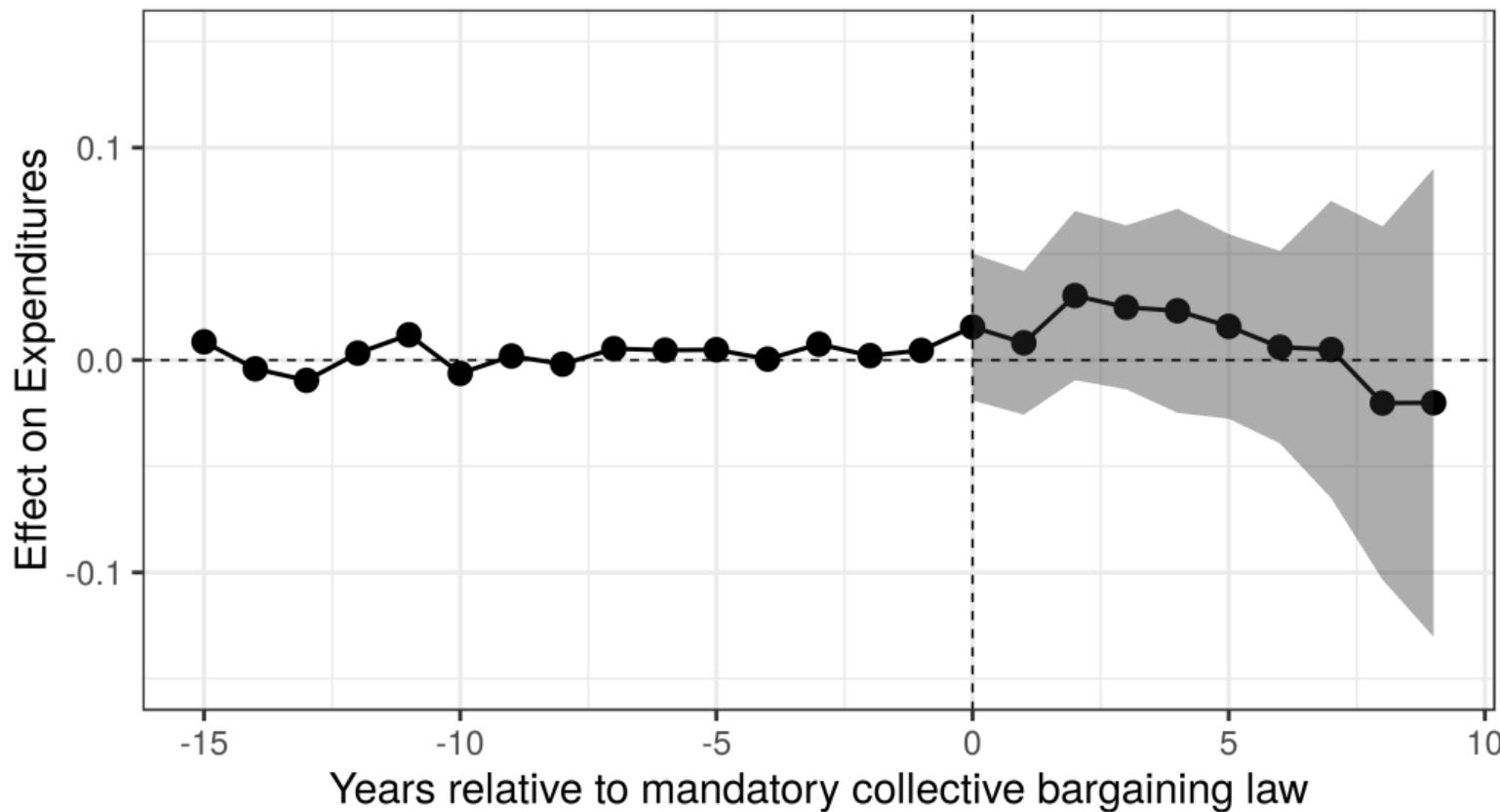
Pooled SCM



Partially Pooled SCM



Partially Pooled SCM



Combining SCM and Outcome Modeling

Weighted Event Study: Fixed Effects + SCM

- Estimate unit fixed effects via pre-treatment average $\bar{Y}_{i,T_j}^{\text{pre}}$

$$\hat{Y}_{j,T_j+k}^{\text{aug}}(0) = \bar{Y}_{j,T_j}^{\text{pre}} + \sum_{i=1}^N \hat{\gamma}_{ij}^* \left(Y_{i,T_j+k} - \bar{Y}_{i,T_j}^{\text{pre}} \right)$$

- Estimate SCM weights $\hat{\gamma}_{ij}^*$ using residuals, equivalent to adding an intercept
[Doudchenko and Imbens, 2017; Ferman and Pinto, 2018]

Weighted Event Study: Fixed Effects + SCM

- Estimate unit fixed effects via pre-treatment average $\bar{Y}_{i,T_j}^{\text{pre}}$

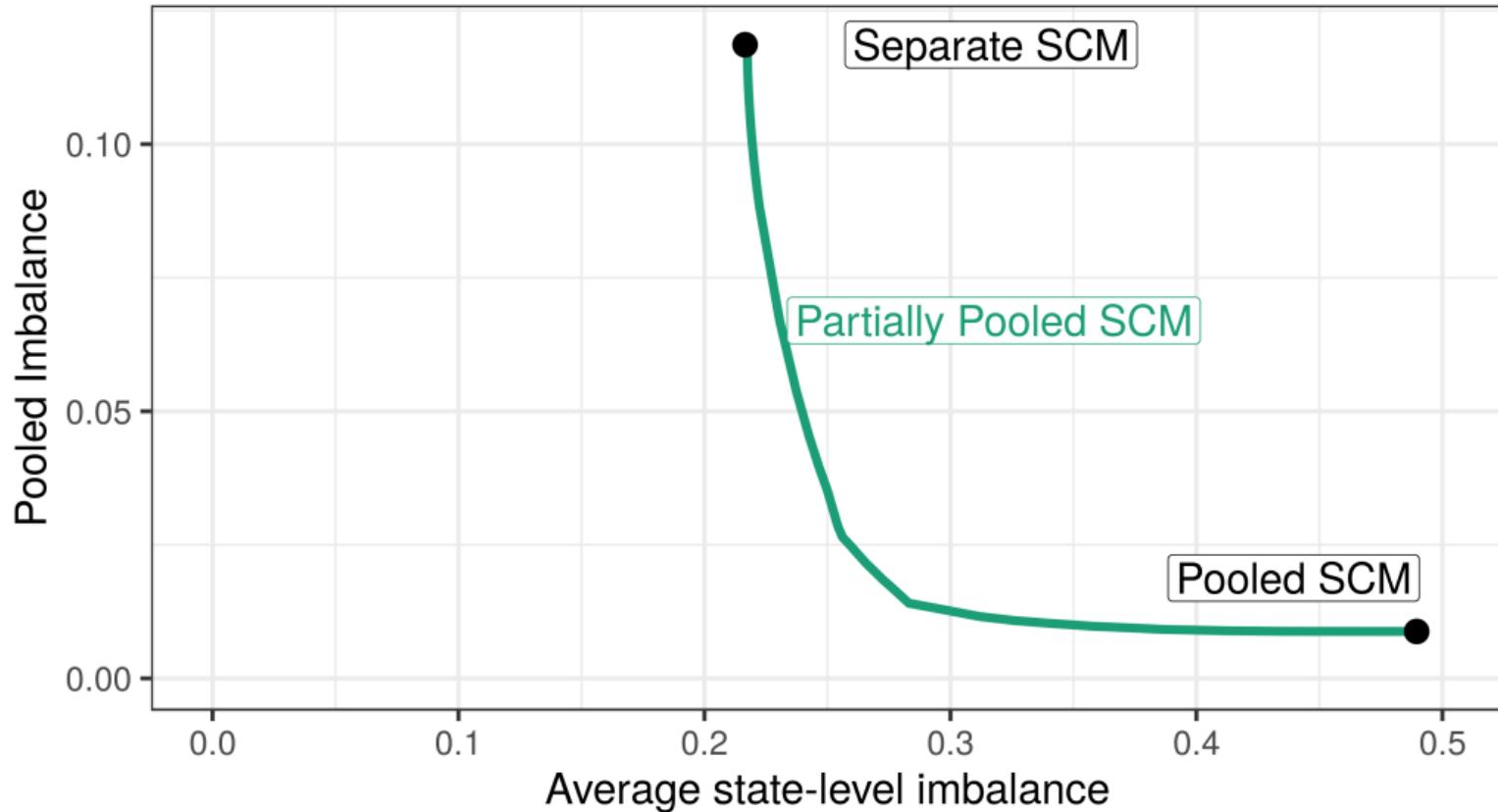
$$\hat{Y}_{j,T_j+k}^{\text{aug}}(0) = \bar{Y}_{j,T_j}^{\text{pre}} + \sum_{i=1}^N \hat{\gamma}_{ij}^* \left(Y_{i,T_j+k} - \bar{Y}_{i,T_j}^{\text{pre}} \right)$$

- Estimate SCM weights $\hat{\gamma}_{ij}^*$ using residuals, equivalent to adding an intercept
[Doudchenko and Imbens, 2017; Ferman and Pinto, 2018]
- Treatment effect estimate is weighted difference-in-differences

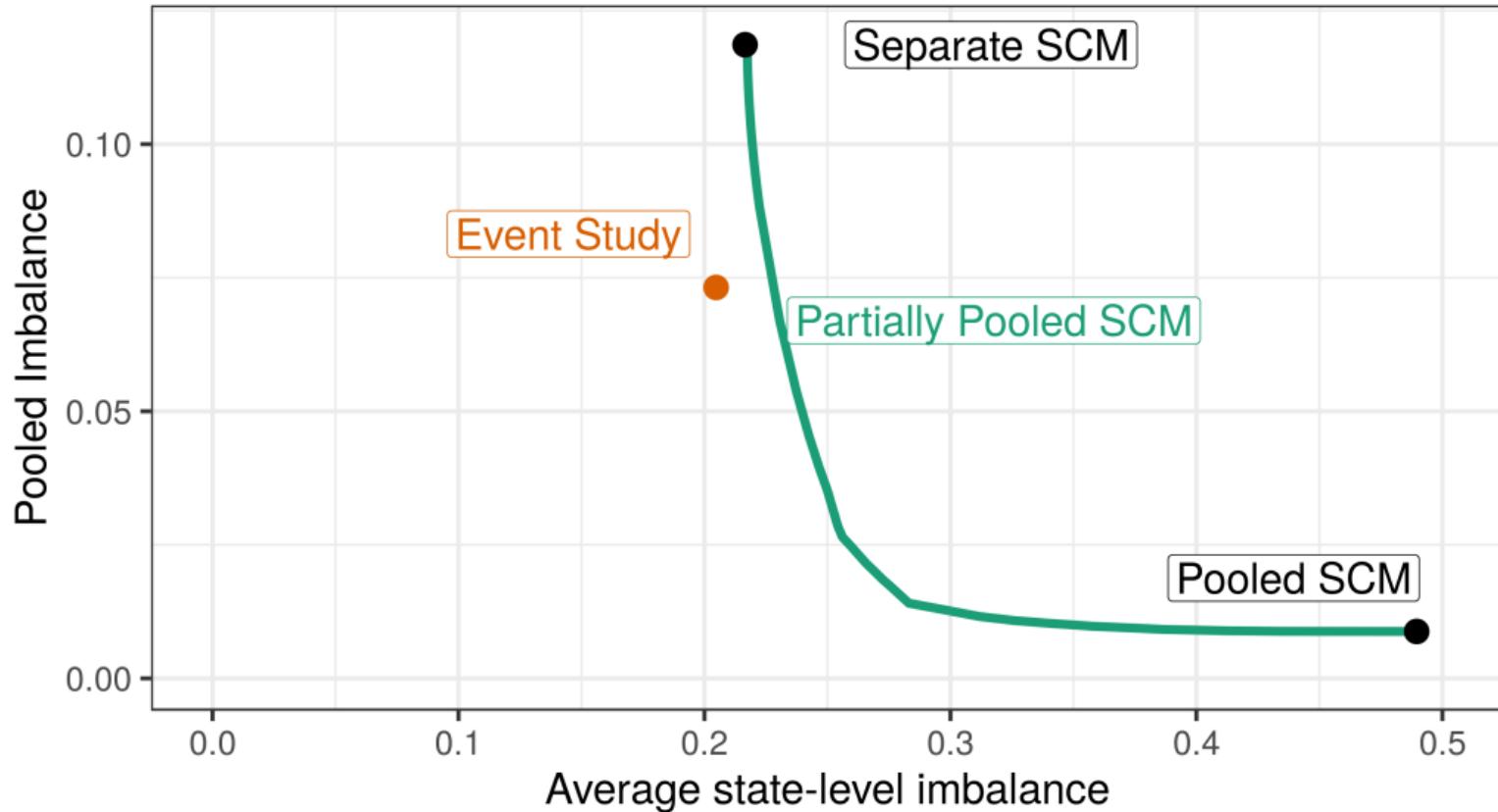
$$\hat{\tau}_{jk}^{\text{aug}} = \left(Y_{j,T_j+k} - \bar{Y}_{j,T_j}^{\text{pre}} \right) - \sum_{i=1}^N \hat{\gamma}_{ij}^* \left(Y_{i,T_j+k} - \bar{Y}_{i,T_j}^{\text{pre}} \right)$$

- Uniform weights recover “stacked” DiD [Abraham and Sun, 2018; Callaway and Sant’Anna, 2018]
- Extends to generic panel models [Ben-Michael et al., 2019]

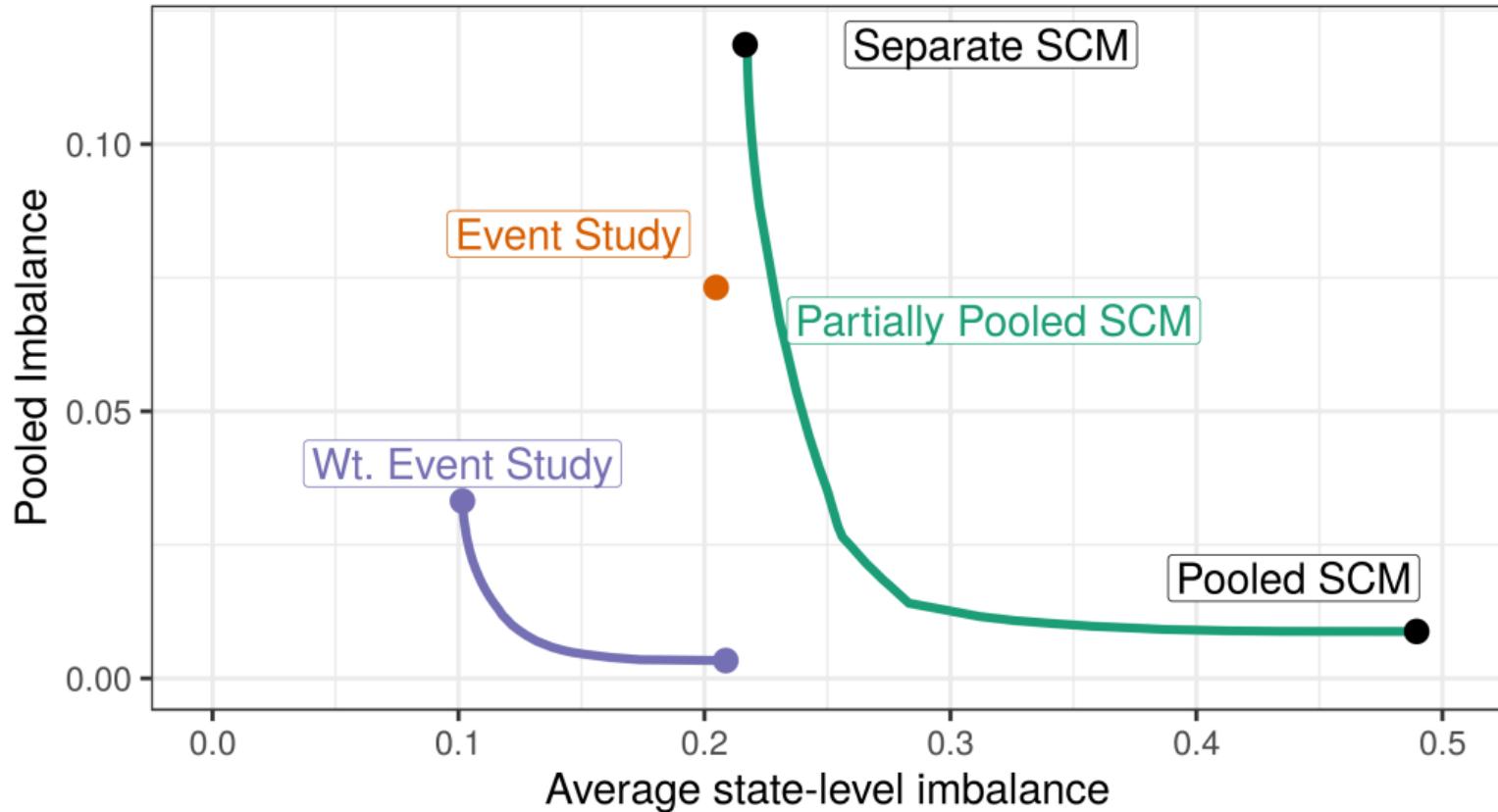
Balance possibility frontier



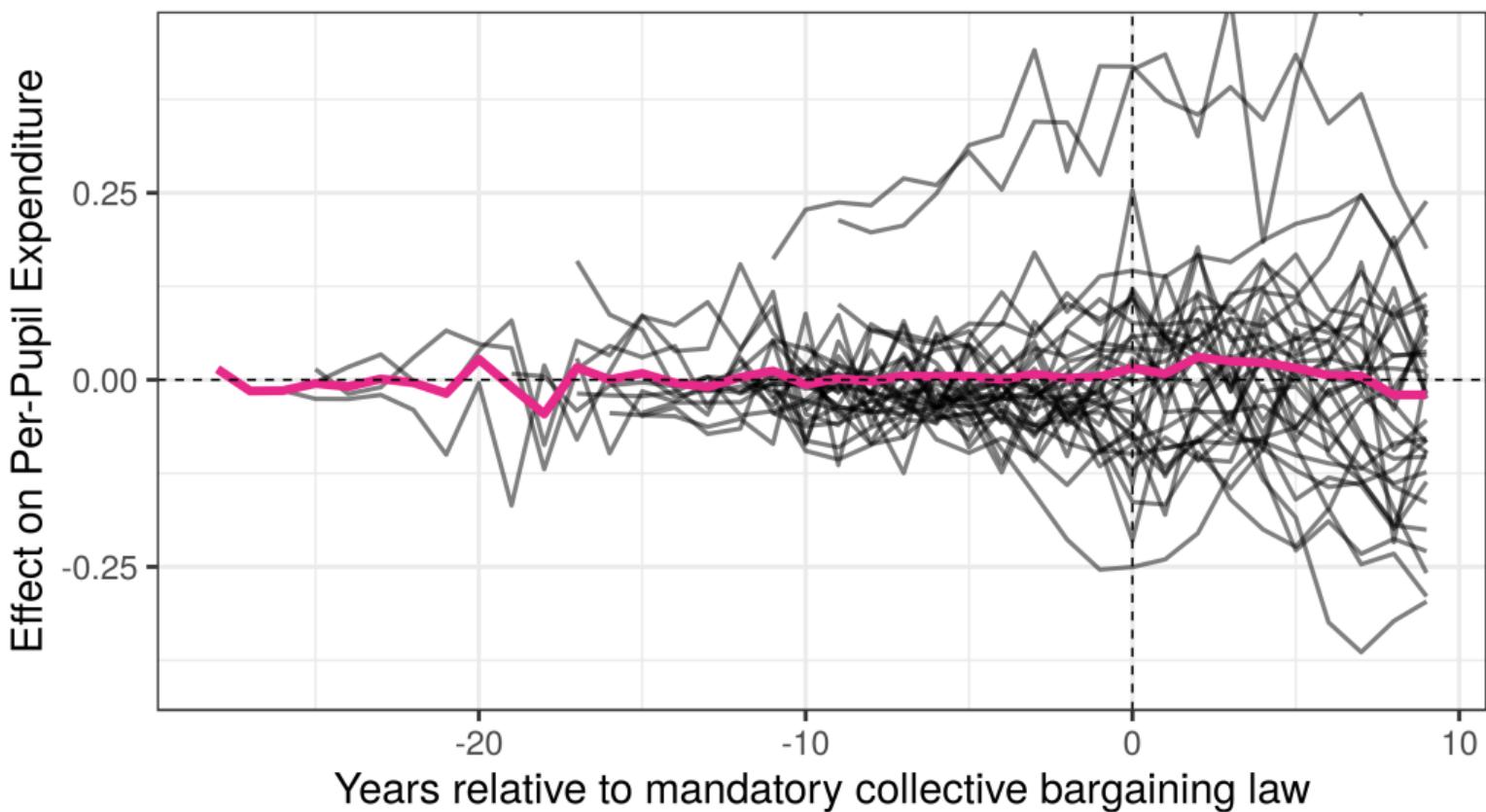
Balance possibility frontier



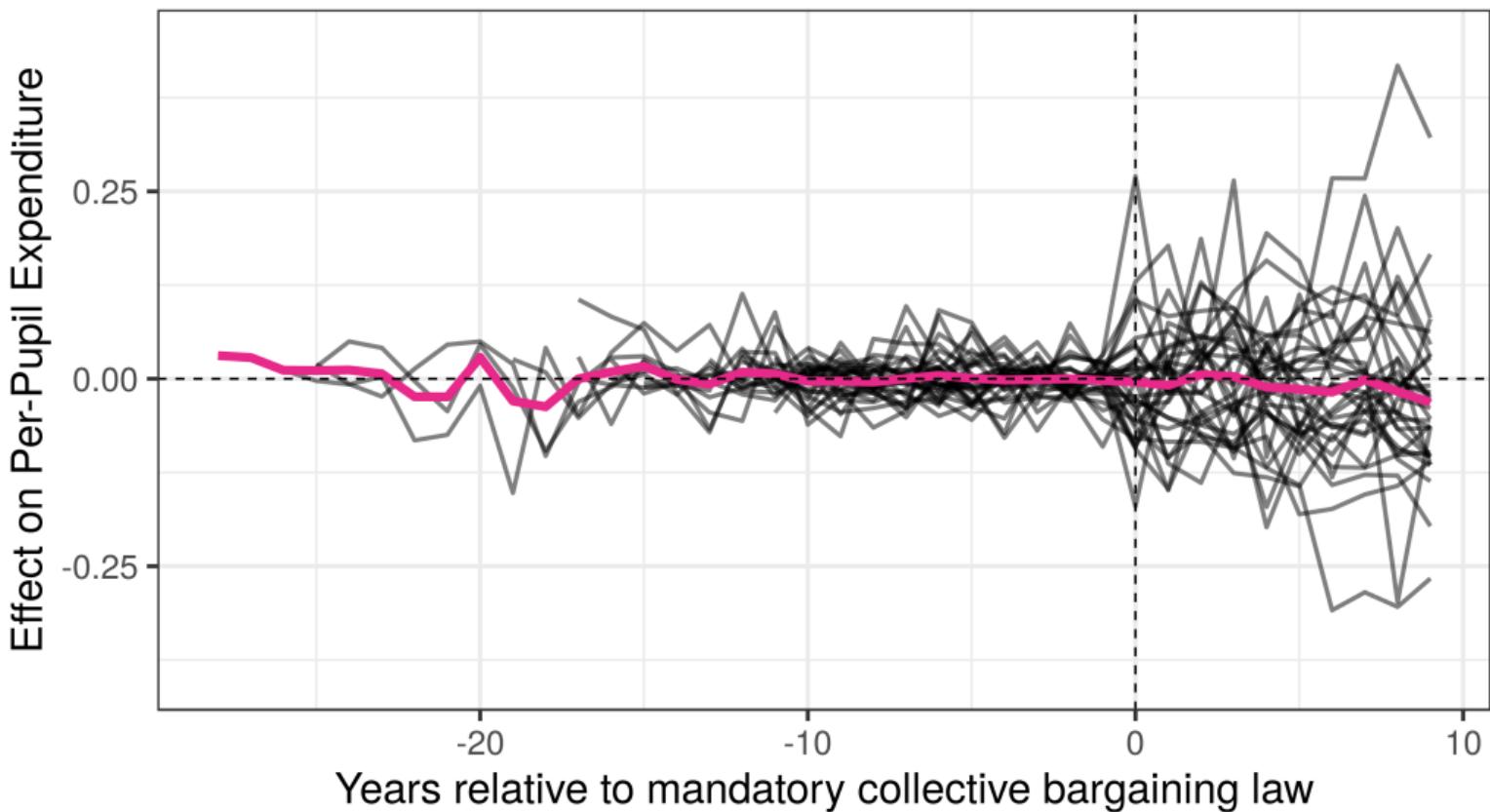
Balance possibility frontier



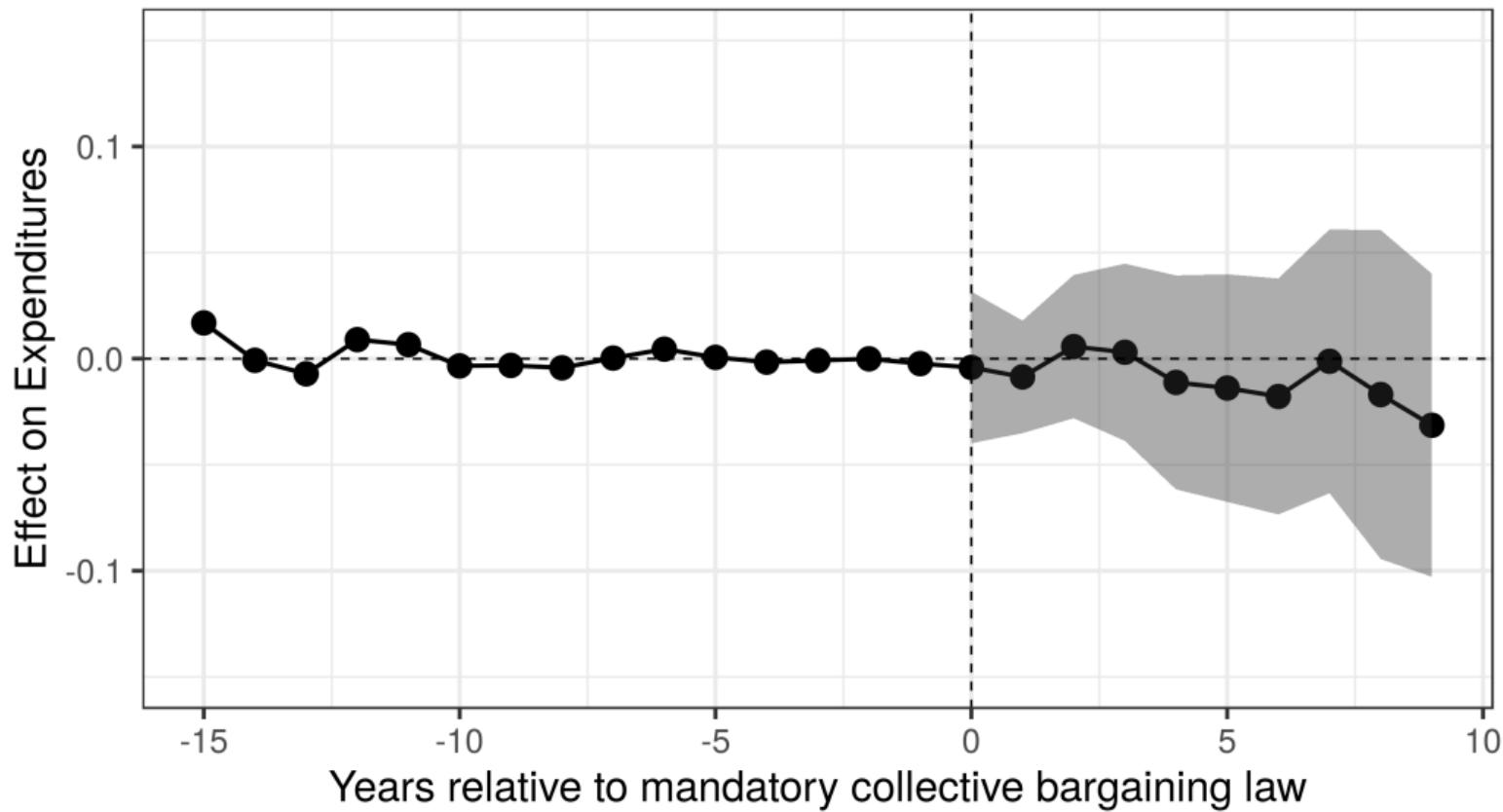
Partially Pooled SCM



Weighted Event Study



Weighted Event Study



Recap

This paper: Extend SCM to staggered adoption

- Find weights that control State Balance and Avg Balance
- Combine SCM and outcome modeling to improve over both
- *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
- Combine SCM and outcome modeling to improve over both
- *Under the hood:* Dual shrinkage; connection to (generalized) IPW

In progress: Generalizing the approach

- Include auxiliary covariates and multiple outcomes
- Extend to unbalanced panels
- Sensitivity analysis

Recap

This paper: Extend SCM to staggered adoption

- Find weights that control State Balance and Avg Balance
- Combine SCM and outcome modeling to improve over both
- *Under the hood:* Dual shrinkage; connection to (generalized) IPW

In progress: Generalizing the approach

- Include auxiliary covariates and multiple outcomes
- 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 matters more

Calibrated sim study: Random Effects AR

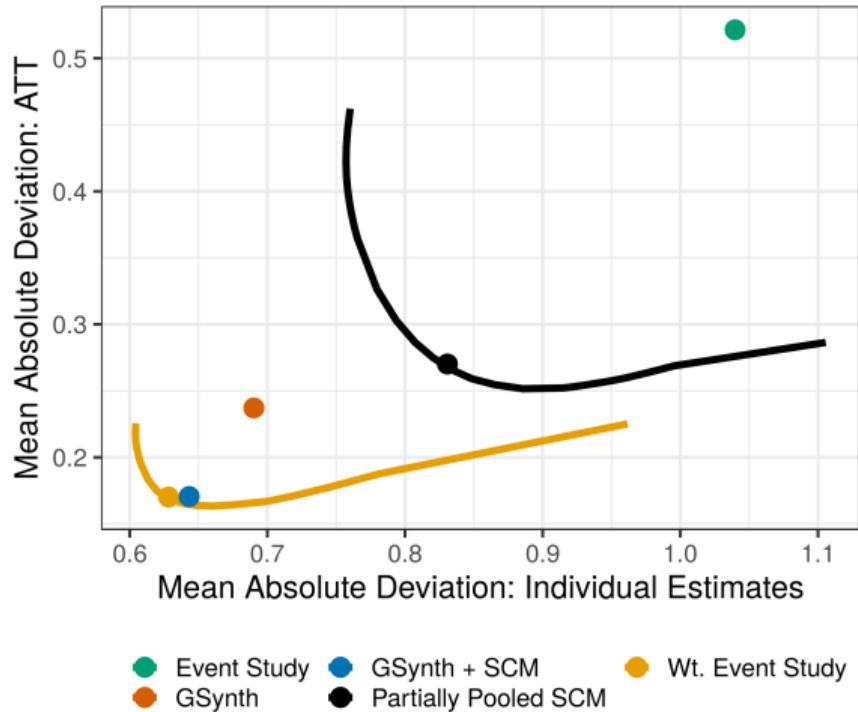
- Fit random effects model

[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: Weighted event study performs well

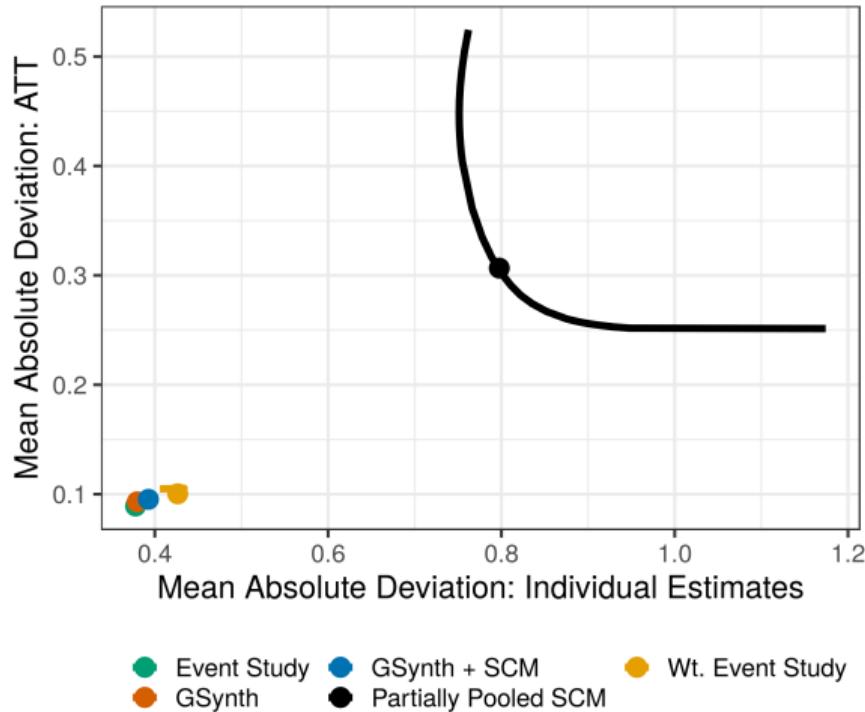
Calibrated sim study: FE

- 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



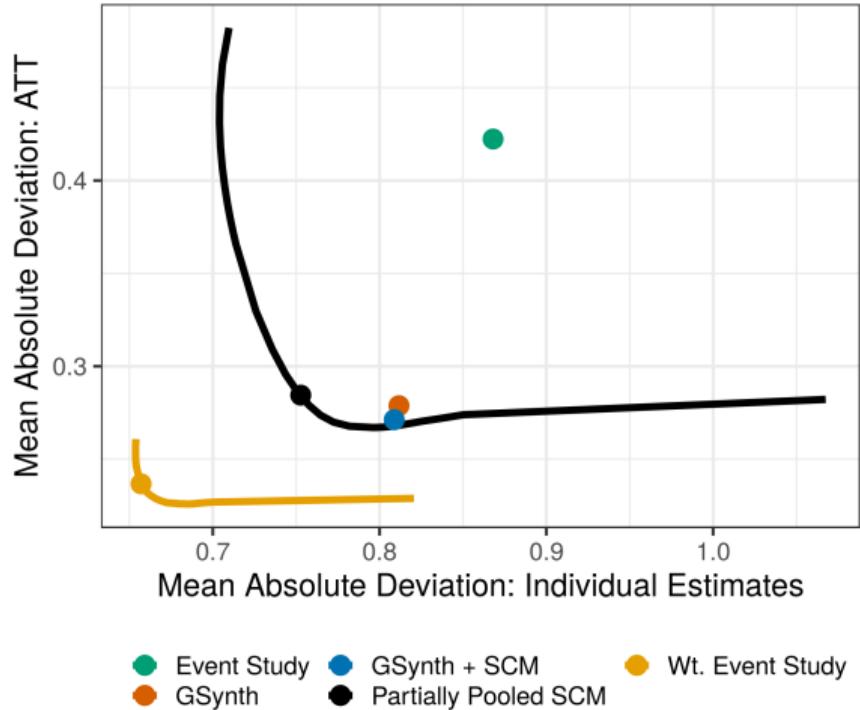
DGP is Factor Model: Weighted event study dominates

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

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

References I

- 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.
- Ben-Michael, E., Feller, A., and Rothstein, J. (2019). The Augmented Synthetic Control Method.
- 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/Hierarchical Models*.
- Hoxby, C. M. (1996). How teachers' unions affect education production. *The Quarterly Journal of Economics*, 111(3):671–718.

References II

- 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.