Synthetic Control and Weighted Event Study Models with Staggered Adoption

Eli Ben-Michael, Avi Feller, and Jesse Rothstein

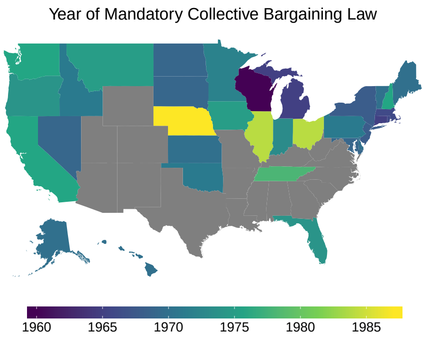
UC Berkeley

Berkeley-Stanford Econometrics Jamboree

November 2019

The impact of teacher unions

- *1960 – 1987:* 33 states grant collective bargaining rights to teachers
  - Long literature exploiting this timing [e.g., Hoxby, 1996; Lovenheim, 2009]
- Impact on teacher salaries, student spending
- Paglayan [2019] estimates precise zero
  - Uses ever-treated states
  - **We use all states**



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

- Event study requires parallel trends assumption, rests heavily on linearity
- 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

- Event study requires parallel trends assumption, rests heavily on linearity
- Synthetic Control Method (SCM) designed for single treated unit, poor fit for average

Our paper: One path forward

- *Generalize SCM:* Modify optimization problem to target overall and state-specific fit
- *Combined approach:* Combine event study modeling and SCM

Causal estimands

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

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

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]

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

Causal estimands

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

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

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]

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

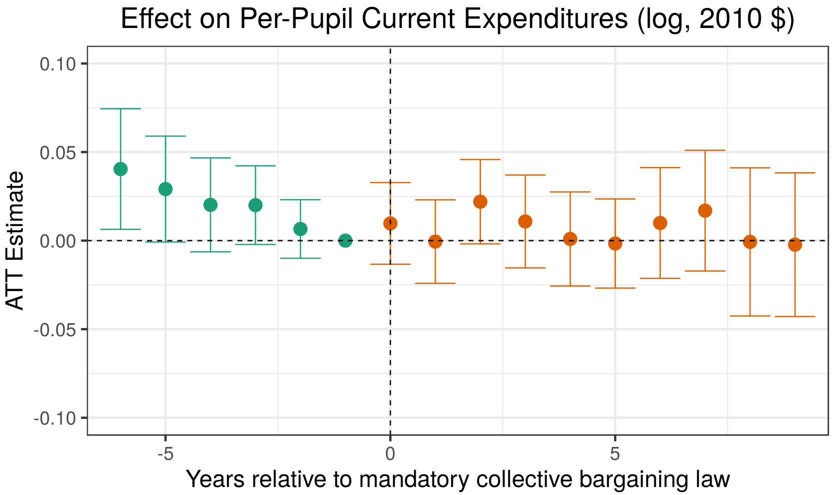
Basic building block: *Treatment effect for unit j*

$$\tau_{jk} = Y_{j,T_j+k}(1) - Y_{j,T_j+k}(0)$$

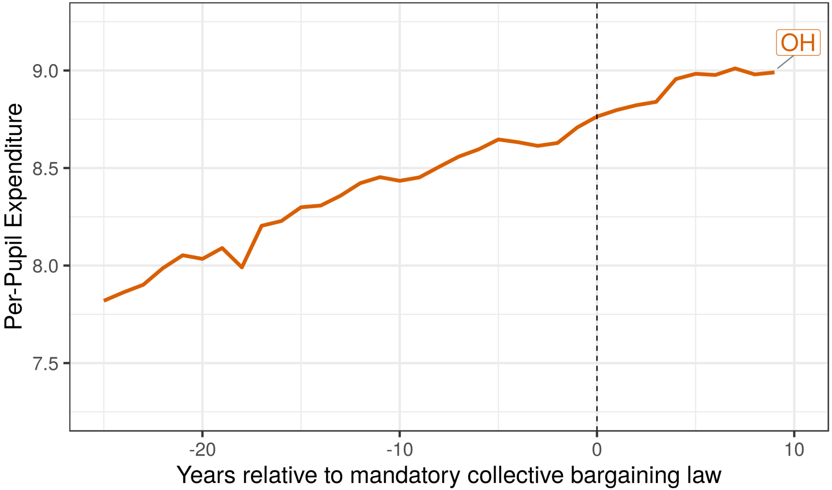
And other weighted averages [Dube and Zipperer, 2015]

Aggregate estimates:

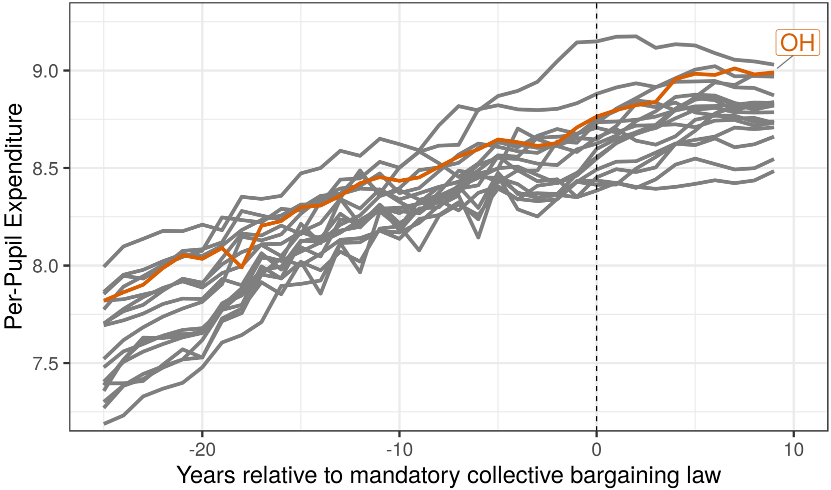
$$\text{ATT}_k = \frac{1}{J} \sum_{j=1}^{J} \tau_{jk}$$



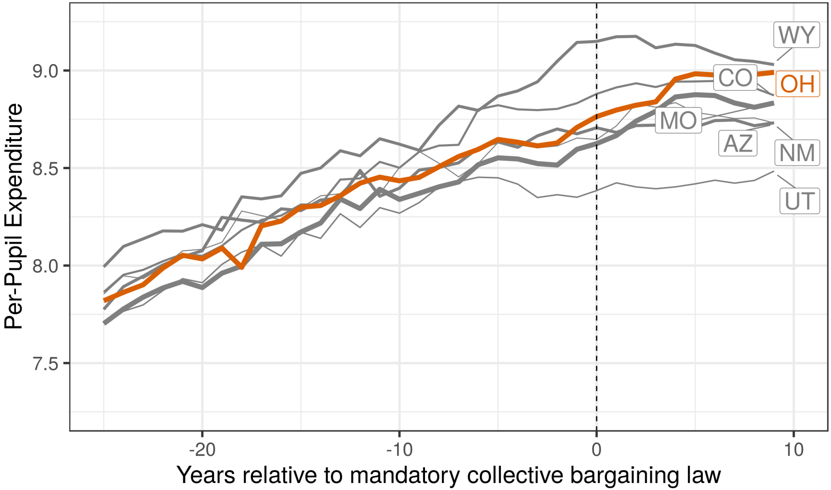
$$Y_{it} = \text{unit}_i + \text{time}_t + \sum_{\ell=2}^{L} \delta_\ell \mathbb{1}\{T_i = t - \ell\} + \sum_{k=0}^{K} \tau_k \mathbb{1}\{T_i = t + k\} + \varepsilon_{it},$$



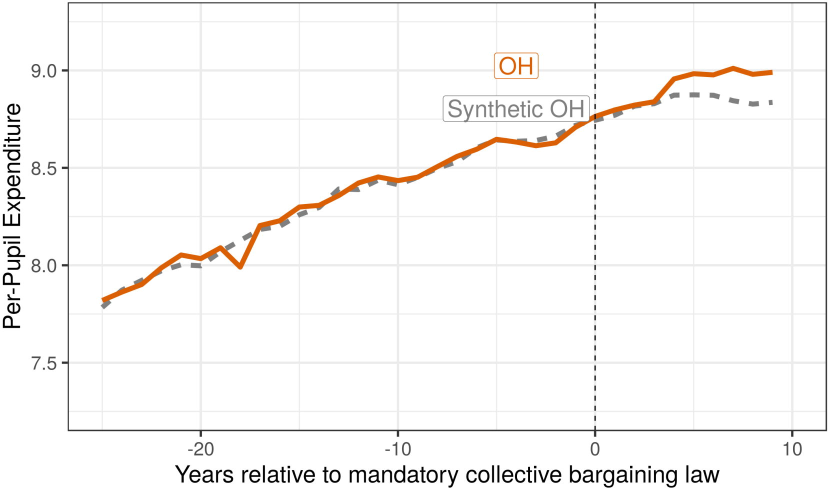
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \quad \|\text{State Balance}_j\|_2^2 + \lambda \sum_{i=1}^{N} f(\gamma_{ij})$$



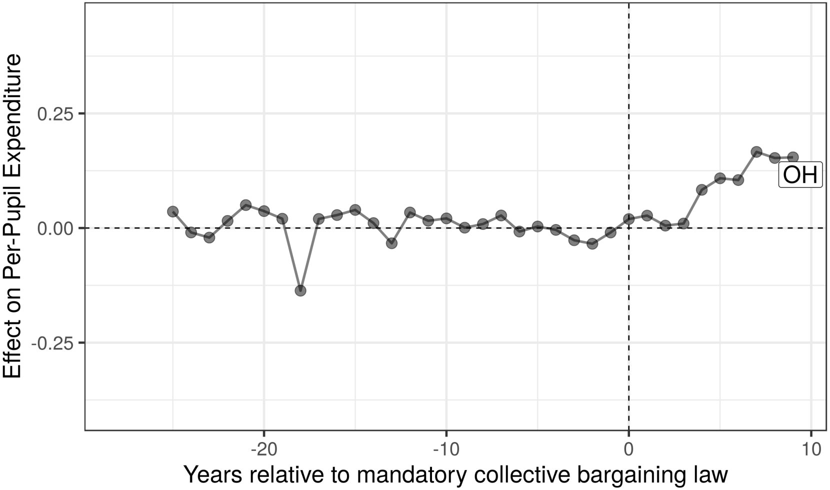
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \quad \|\text{State Balance}_j\|_2^2 + \lambda \sum_{i=1}^{N} f(\gamma_{ij})$$



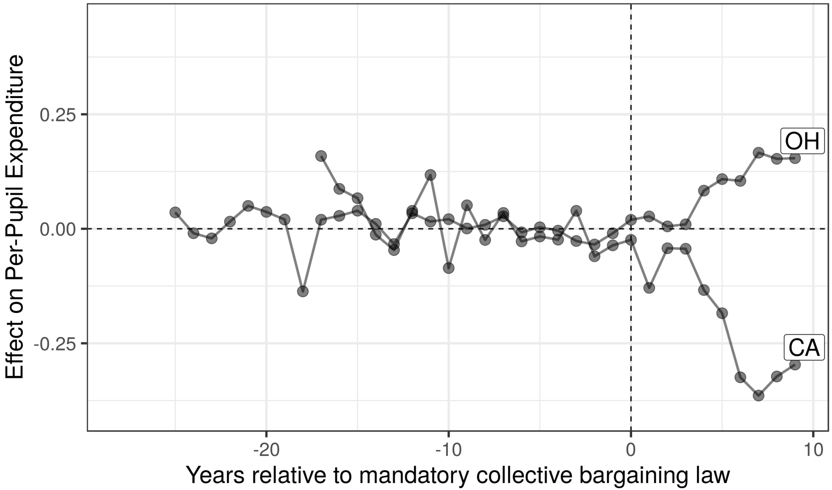
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \quad \|\text{State Balance}_j\|_2^2 + \lambda \sum_{i=1}^{N} f(\gamma_{ij})$$



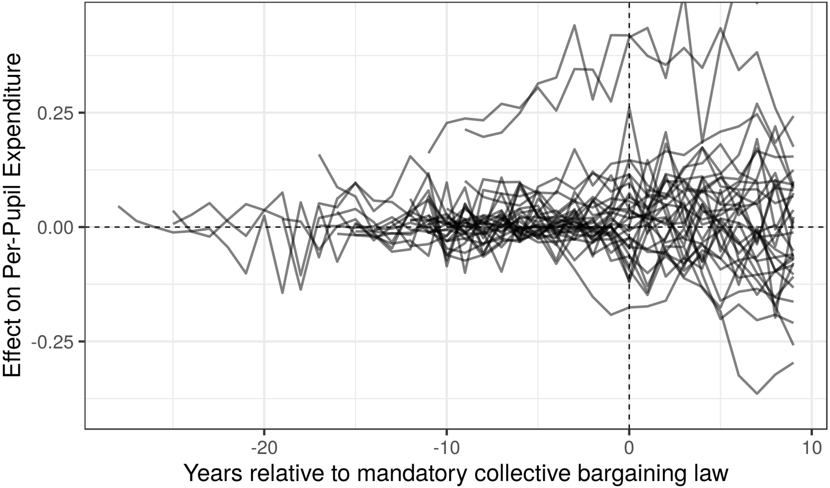
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \quad \|\text{State Balance}_j\|_2^2 + \lambda \sum_{i=1}^{N} f(\gamma_{ij})$$



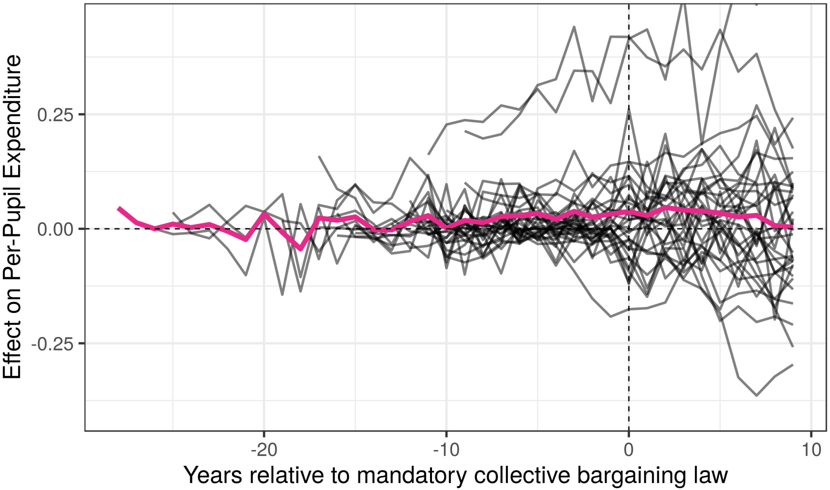
$$\min_{\gamma_j \in \Delta_j^{\text{scm}}} \quad \|\text{State Balance}_j\|_2^2 + \lambda \sum_{i=1}^{N} f(\gamma_{ij})$$



$$\min_{\gamma_1,\ldots,\gamma_J \in \Delta^{\text{scm}}} \frac{1}{J}\sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2 + \lambda \sum_{j=1}^{J}\sum_{i=1}^{n} f(\gamma_{ij})$$

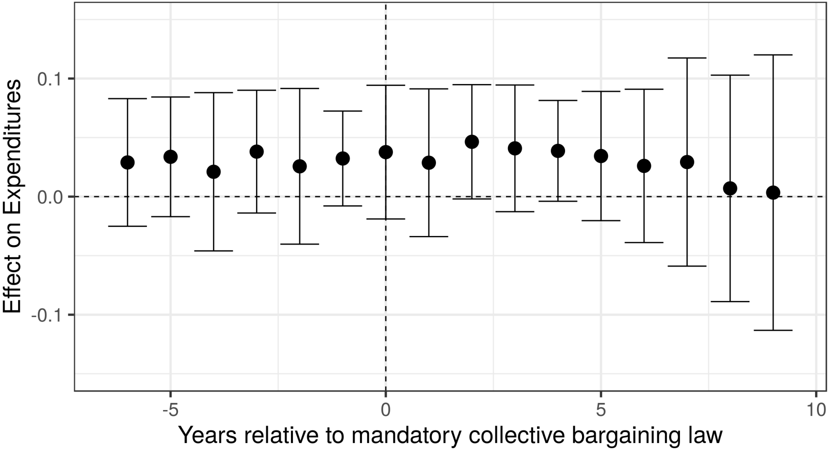


$$\min_{\gamma_1,\ldots,\gamma_J \in \Delta^{\text{scm}}} \frac{1}{J}\sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2 + \lambda \sum_{j=1}^{J}\sum_{i=1}^{n} f(\gamma_{ij})$$



$$\min_{\gamma_1,\ldots,\gamma_J \in \Delta^{\text{scm}}} \frac{1}{J} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2 + \lambda \sum_{j=1}^{J} \sum_{i=1}^{n} f(\gamma_{ij})$$

Separate SCM



$$\min_{\gamma_1,\ldots,\gamma_J \in \Delta^{\text{scm}}} \frac{1}{J} \sum_{j=1}^{J} \| \text{State Balance}_j \|_2^2 + \lambda \sum_{j=1}^{J} \sum_{i=1}^{n} f(\gamma_{ij})$$

Pre-treatment fit and bias

Generalization of parallel trends: Linear Factor Model

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

Pre-treatment fit and bias

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 $\rightarrow$ importance of Avg Balance
- S is the factor standard deviation $\rightarrow$ importance of State Balance
- Special case: unit fixed effects, only Avg Balance matters

Paritally pooled SCM: Control both imbalances

Can get gains from minimizing Avg Balance but State Balance still matters

Paritally pooled SCM: Control both imbalances

- Can get gains from minimizing Avg Balance but State Balance still matters
  - *¿Por que no los dos?*
- Relative weighting defined by ν:

$$\min_{\Gamma} \quad \frac{\nu}{L} \|\text{Avg Balance}\|_2^2 + \frac{1-\nu}{JL} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2 + \lambda \sum_{j=1}^{J} \sum_{i=1}^{n} f(\gamma_{ij})$$

  - Partial pooling in dual parameter space

Paritally pooled SCM: Control both imbalances

Can get gains from minimizing Avg Balance but State Balance still matters

- *¿Por que no los dos?*

Relative weighting defined by ν:

$$\min_{\Gamma} \quad \frac{\nu}{L} \|\text{Avg Balance}\|_2^2 + \frac{1-\nu}{JL} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2 + \lambda \sum_{j=1}^{J} \sum_{i=1}^{n} f(\gamma_{ij})$$

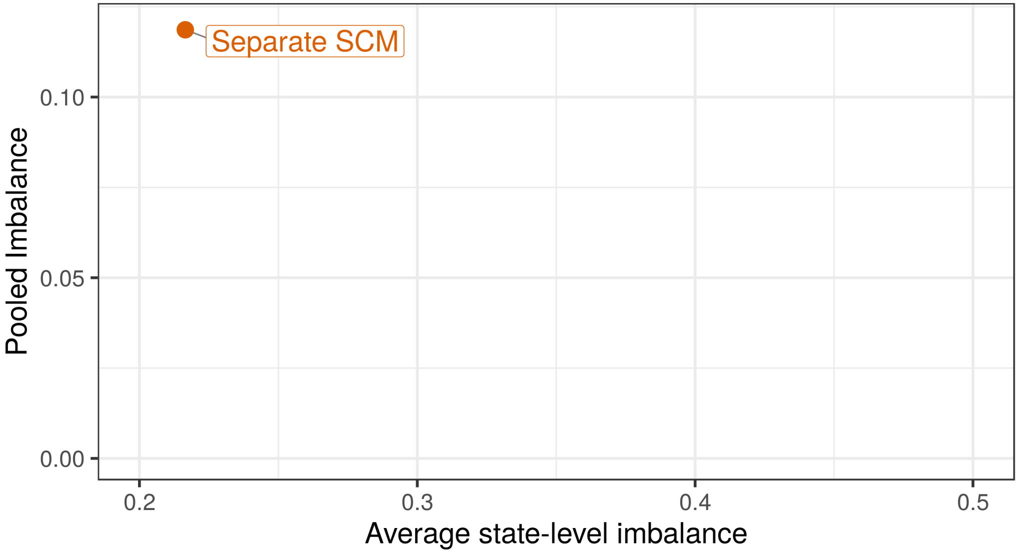
- Partial pooling in dual parameter space

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

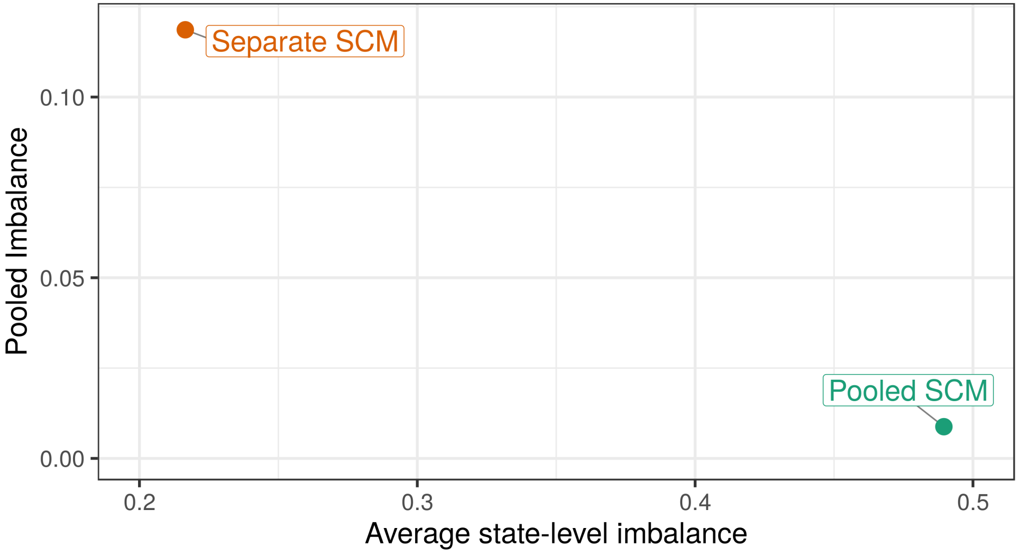
$$\hat{\nu} = \frac{\frac{1}{\sqrt{L}} \|\text{Avg Balance}\|_2}{\sqrt{\frac{1}{J} \sum_{j=1}^{J} \|\text{State Balance}_j\|_2^2}}$$

Pooled SCM $\rightarrow \nu = 1$

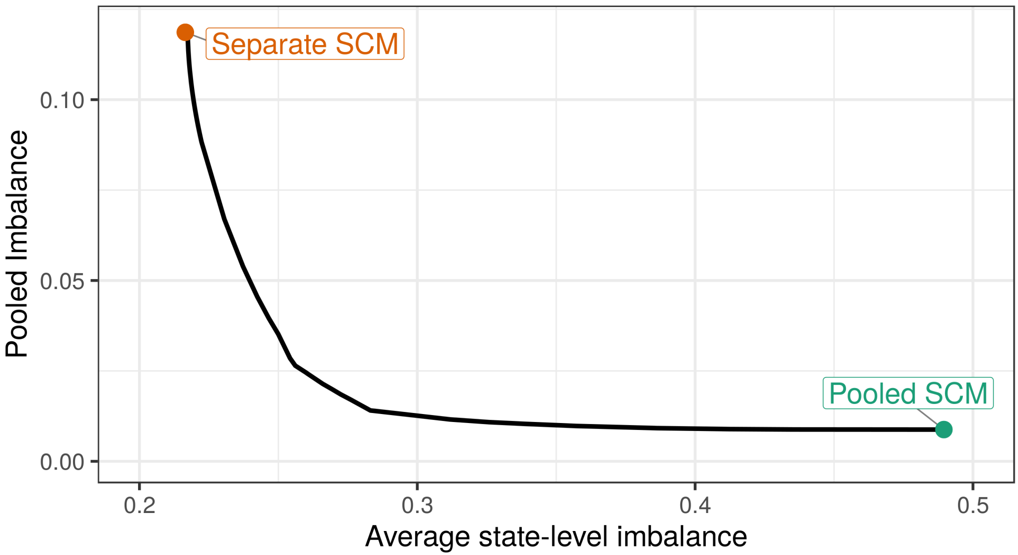
Balance possibility frontier



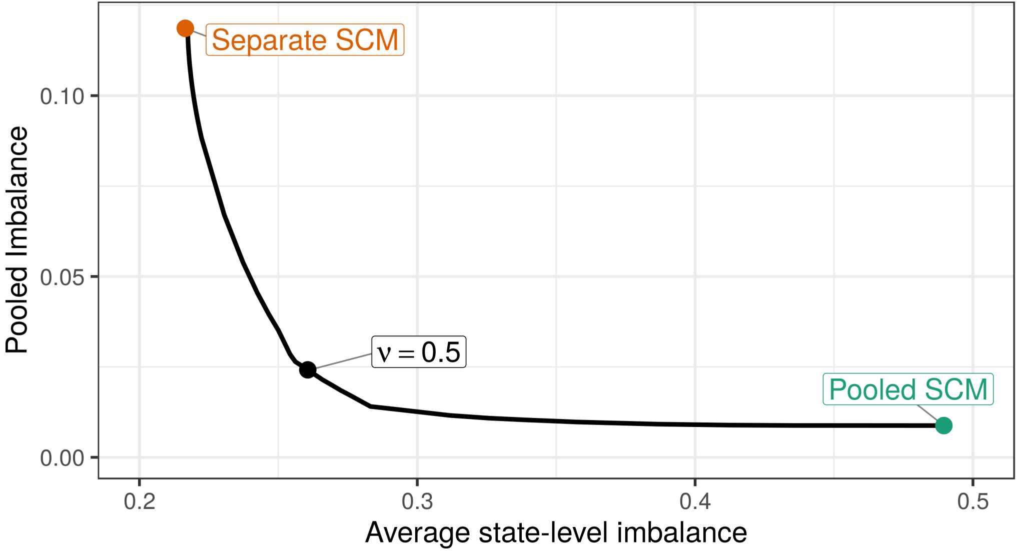
Balance possibility frontier



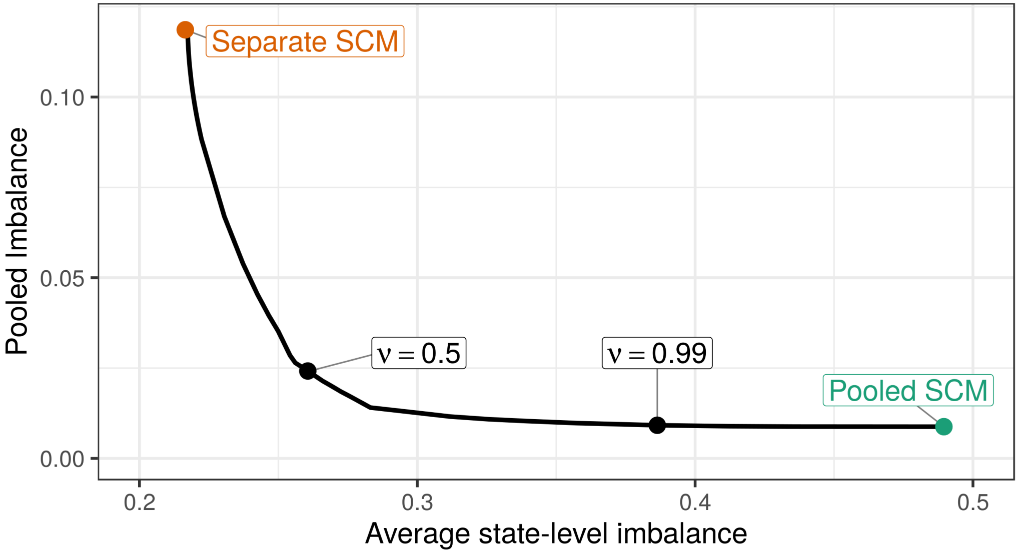
Balance possibility frontier



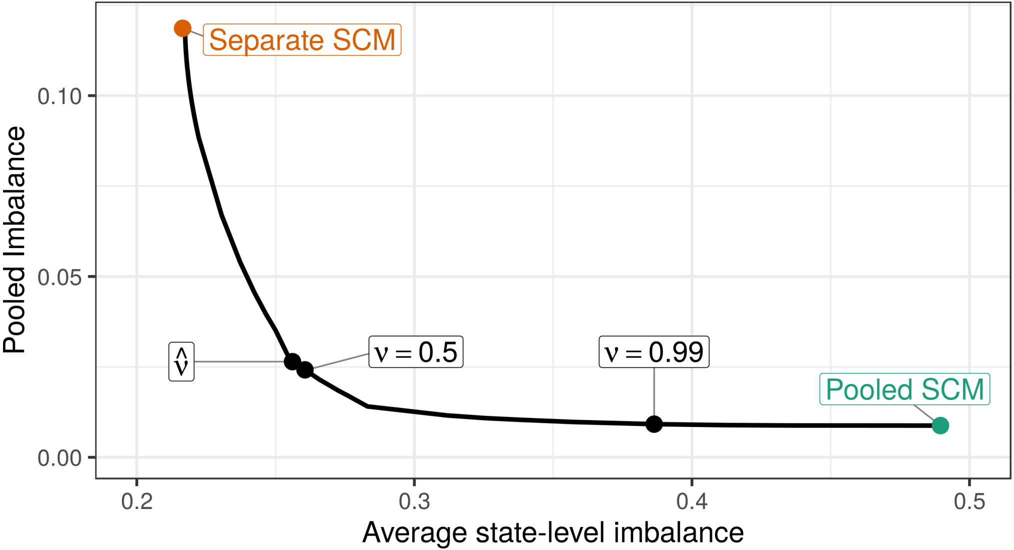
Balance possibility frontier



Balance possibility frontier



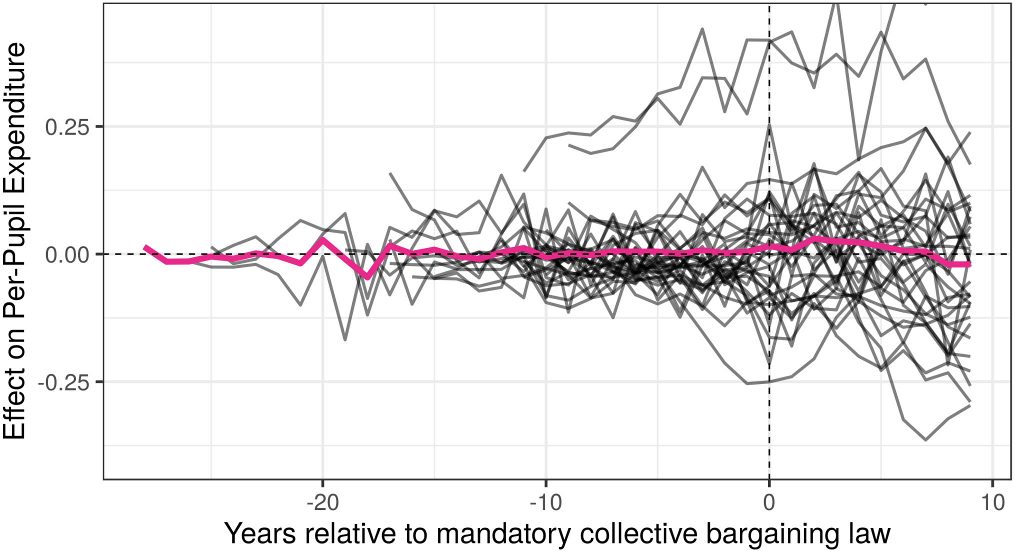
Balance possibility frontier



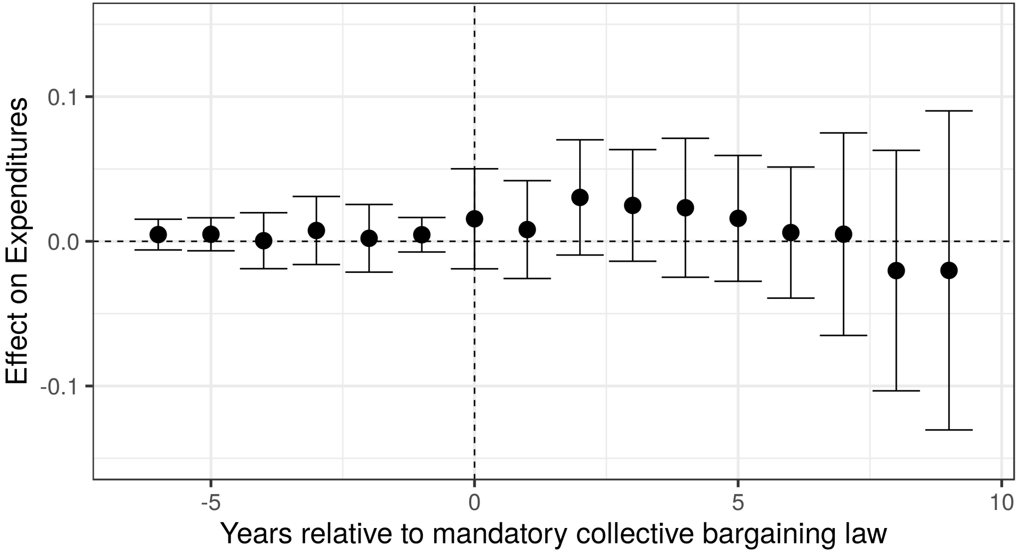
Pooled SCM


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



Weighted Event Study: FE + SCM

Combine outcome modeling and SCM weighting [Ben-Michael et al., 2018]

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

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

- Estimate SCM using **residuals**, equivalent to adding an intercept [Doudchenko and Imbens, 2017; Ferman and Pinto, 2018]

Weighted Event Study: FE + SCM

Combine outcome modeling and SCM weighting [Ben-Michael et al., 2018]

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

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

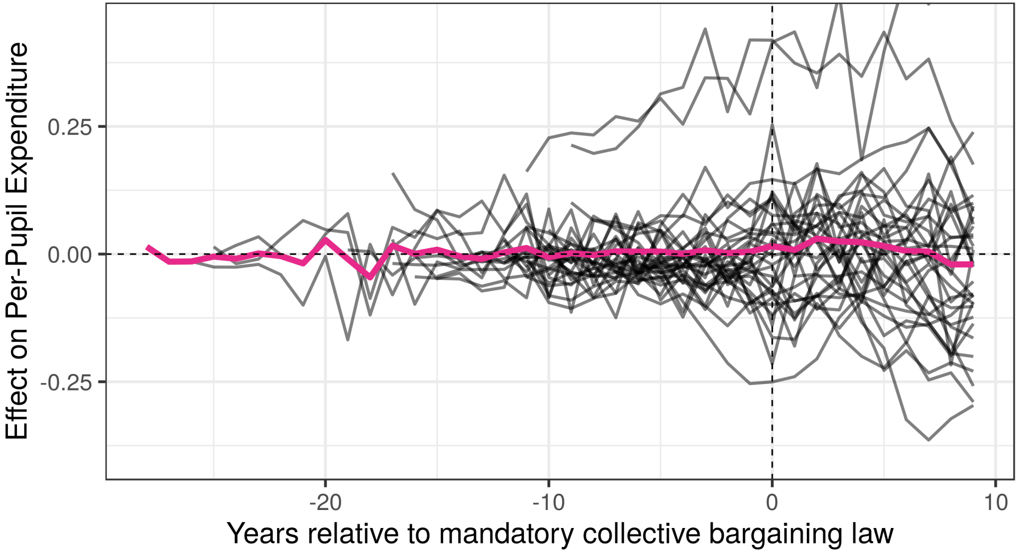
- Estimate SCM using **residuals**, equivalent to adding an intercept [Doudchenko and Imbens, 2017; Ferman and Pinto, 2018]

Treatment effect estimate is **weighted diff-in-diff**

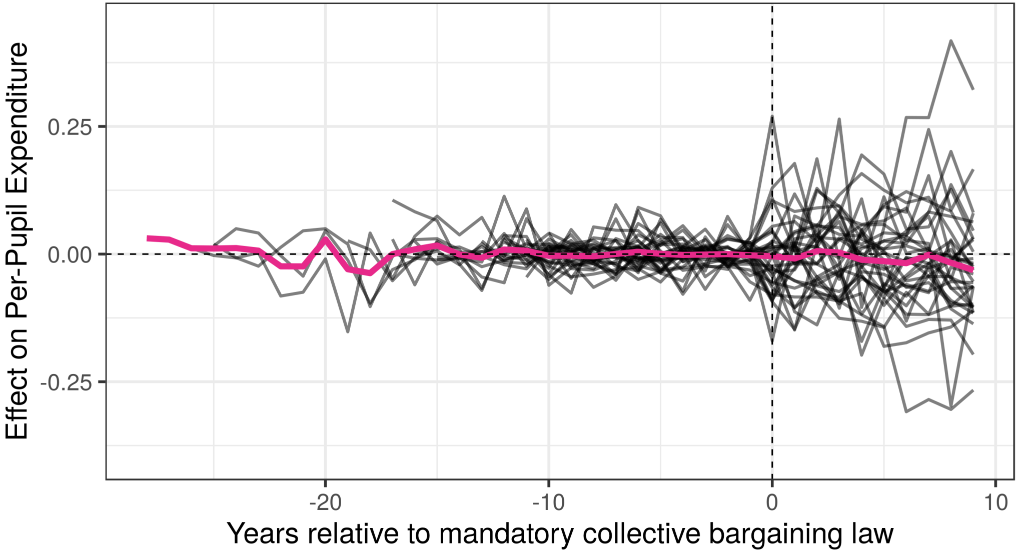
$$\hat{\tau}^{\text{aug}}_{jk} = \left(Y_{j,T_j+k} - \bar{Y}^{\text{pre}}_{j,T_j} \right) - \sum_{i=1}^{N} \hat{\gamma}_{ij} \left(Y_{i,T_j+k} - \bar{Y}^{\text{pre}}_{i,T_j} \right)$$

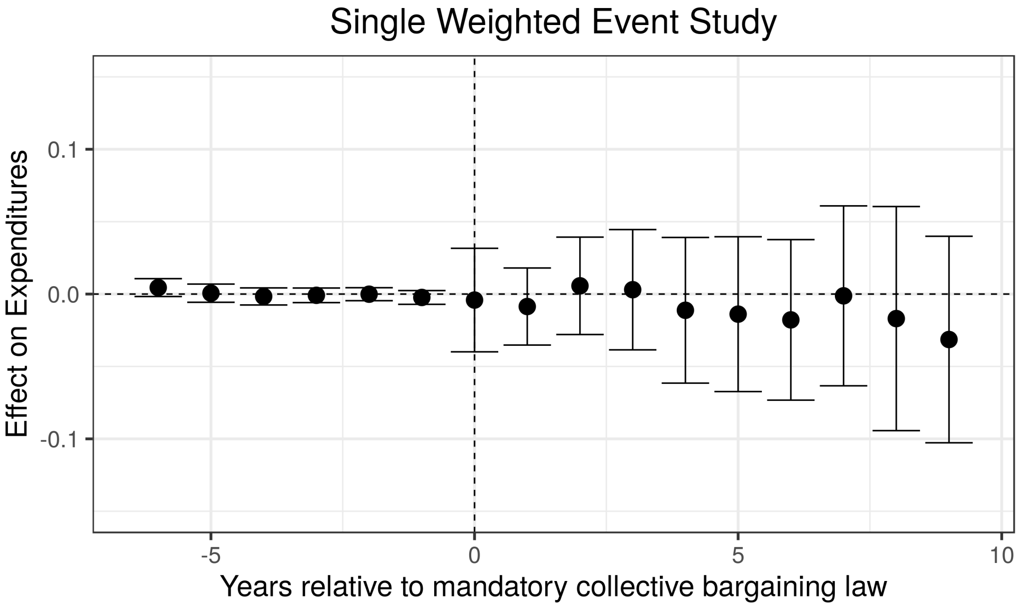
- Uniform weights recover direct estimate
- Connection to semiparametric DiD and conditional parallel trends [Abadie, 2005; Callaway and Sant'Anna, 2018]

Partially Pooled SCM



Weighted Event Study





Single Weighted Event Study
Effect on Expenditures
0.1
0.0
-0.1
-5
0
5
10
Years relative to mandatory collective bargaining law

Random effects AR simulation: level of pooling matters more

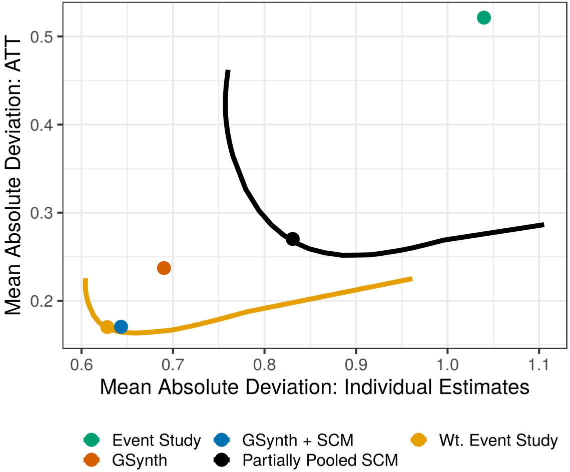
Calibrated sim study: Random Effects AR

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

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

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

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



Recap and next steps

Extending SCM to staggered adoption

- Find weights that control State Balance and Avg Balance
- Combine SCM with Event Study Modeling to improve over both

Recap and next steps

Extending SCM to staggered adoption

- Find weights that control State Balance and Avg Balance
- Combine SCM with Event Study Modeling to improve over both

Future: general approach for augmentation with staggered adoption

- Combining with other outcome models (e.g. matrix completion)
- Allowing for negative weights and include auxiliary covariates

Recap and next steps

Extending SCM to staggered adoption

- Find weights that control State Balance and Avg Balance
- Combine SCM with Event Study Modeling to improve over both

Future: general approach for augmentation with staggered adoption

- Combining with other outcome models (e.g. matrix completion)
- Allowing for negative weights and include auxiliary covariates

Thank you!

`ebenmichael.github.io`

Appendix

DGP is FE Model: Weighted event study performs well

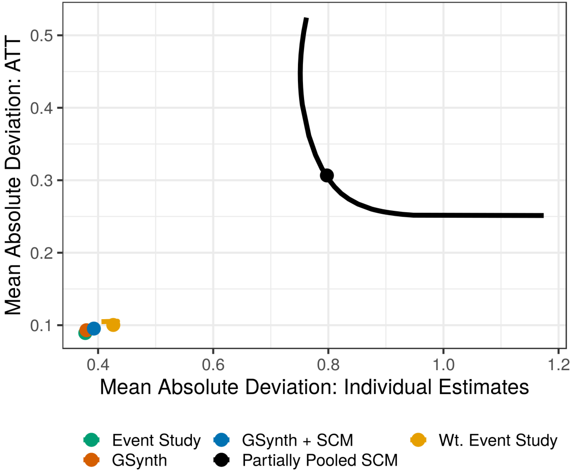
Calibrated sim study: FE

- 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



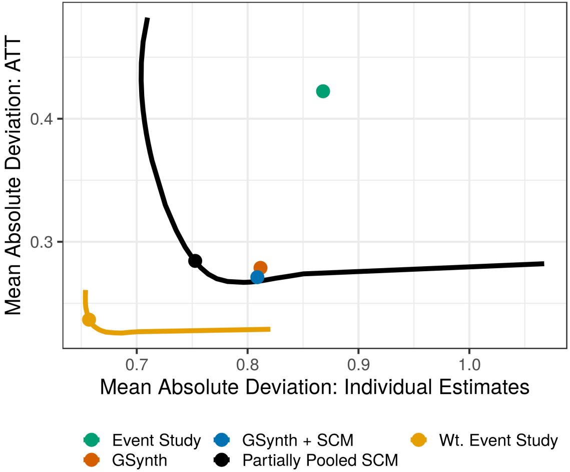
DGP is Factor Model: Weighted event study dominates

Calibrated sim study: Factor

- Fit gsynth [Xu, 2017]

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

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



References I

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

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. (2018). 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*.

Dube, A. and Zipperer, B. (2015). Pooling multiple case studies using synthetic controls: An application to minimum wage policies.

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.

Lovenheim, M. F. (2009). The effect of teachers unions on education production: Evidence from union election certifications in three midwestern states. *Journal of Labor Economics*, 27(4):525–587.

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.