Supplementary Materials for:
"Synthetic Controls with Staggered Adoption"

A Additional theoretical results

A.1 Further discussion of inference

We now continue the discussion of inference from the main text in Section 5.3. Our goal here is to discuss the conditions under which the proposed estimator is asymptotically Normal. Since asymptotic theory is not the focus of our paper, we leave for future work a rigorous derivation of the validity of the wild bootstrap procedure, in particular, adapting the proof of the main theorem in Otsu and Rai (2017) and showing that the additional conditions in that proof are satisfied with our proposed procedure.

In order to discuss inferential procedures for partially pooled SCM with an intercept shift, we will consider a generalization of parallel trends. For each time period g, we assume that the expected differences between post-g and pre-g outcomes do not depend on whether unit i is treated at time g, conditional on auxiliary covariates X_i and the vector of pre-g residuals $\dot{Y}_i^g \equiv (Y_{ig-L}, \ldots, Y_{ig-1}) - \frac{1}{L}\sum_{\ell=1}^{L} Y_{ig-\ell}$.

Assumption A.1 (Conditional parallel trends)**.** With $L < T_1$, for all $k \geq 0$ and $\ell \geq 1$

$$\mathbb{E}[Y_{ig+k}(\infty) - Y_{ig-\ell}(\infty) \mid T_i = g, \dot{Y}_i^g, X_i] = \mathbb{E}[Y_{ig+k}(\infty) - Y_{ig-\ell}(\infty) \mid \dot{Y}_i^g, X_i] \equiv m_{gk\ell}(\dot{Y}_i^g, X_i)$$

Assumption A.1 is a generalization of the conditional parallel trends assumption in Abadie (2005) to the staggered adoption setting, including the pre-treatment residuals $\dot{Y}_i^g$. It loosens the usual parallel trends assumption by allowing trends to differ depending on the auxiliary covariates and the deviation of lagged outcomes from their baseline value. Thus, we are essentially conditioning on pre-treatment "dynamics," rather than pre-treatment levels. For instance, even if two states have very different levels of student expenditures, under conditional parallel trends we can compare them so long as they have similar pre-treatment trends and shocks. See Hazlett and Xu (2018) and Callaway and Sant'Anna (2020) for related conditional parallel trends assumptions. In addition, we will assume that the conditional expectation of the post- and pre-g differences is linear.

Assumption A.2.

$$m_{gk\ell}(\dot{Y}_i^g, X_i) = \beta_{gk\ell}^Y \cdot \dot{Y}_i^g + \beta_{gk\ell}^X \cdot X_i$$

We make two further assumptions that allow for asymptotic normality as the number of units grows while the number of lags L stays fixed. First, we assume that the synthetic controls have perfect fit when averaged within time-cohorts; second, we assume that the sum of the squared weights is bounded.

Assumption A.3 (Exact balance within treatment cohorts and bounded weights)**.** Assume that

$$\frac{1}{n_g}\sum_{T_i=g}\dot{Y}_i^g = \frac{1}{n_g}\sum_{i=1}^{N}\sum_{T_j=g}\hat{\gamma}_{ij}\dot{Y}_i^g \text{ and } \frac{1}{n_g}\sum_{T_i=g}X_i = \frac{1}{n_g}\sum_{i=1}^{N}\sum_{T_j=g}\hat{\gamma}_{ij}X_i,$$

for all $g = T_1, \ldots, T_J$. Furthermore, $\|\hat{\gamma}_j\|_2 \leq \frac{C}{\sqrt{N_0}}$ for all $j = 1, \ldots, J$ and some constant C.

Note that by transforming from the penalized optimization problem (7) to the constrained form, there is a choice of λ that guarantees that the constraint on the weights are satisfied, if there exists a feasible solution. Finally, we make two assumptions on the noise terms $\varepsilon_{igk} \equiv Y_{ig+k}(\infty) - \frac{1}{L}\sum_{\ell=1}^{L} Y_{ig-\ell}(\infty) - \frac{1}{L}\sum_{\ell=1}^{L} m_{k\ell}(g, \dot{Y}_i^g, X_i)$. First, we assume that they are independent across units; second, we assume that they are sufficiently regular so that their average satisfies a central limit theorem.

Assumption A.4. ε_{igk} are independent across units $i = 1, \ldots, N$, and for some $\delta > 0$, the $2+\delta^{\text{th}}$ moment exists, $\mathbb{E}\left[|\varepsilon_{igk}|^{2+\delta}\right] < \infty$, and furthermore

$$\lim_{N\to\infty} \frac{\sum_{T_i\neq\infty}\mathbb{E}\left[|\varepsilon_{iT_ik}|^{2+\delta}\right]}{\left(\sum_{T_i\neq\infty}\mathbb{E}\left[\varepsilon_{iT_ik}^2\right]\right)^{1+\frac{\delta}{2}}} = 0.$$

Under these assumptions, the estimate of the effect k periods after treatment, $\widehat{\text{ATT}}_k$, will be asymptotically normal as N grows with a fixed number of lags L, and where the number of control units N_0 grows more quickly than the number of treated units J.

Theorem A.1. Assume that $\frac{J}{N_0} \to 0$ as both $J, N_0 \to \infty$, with L fixed. Under Assumptions A.1, A.2, A.3, and A.4

$$\sqrt{J}\left(\widehat{\text{ATT}}_k - \text{ATT}\right) = \frac{1}{\sqrt{J}}\sum_{T_i\neq\infty}\varepsilon_{iT_j+k} + o_p(1).$$

Furthermore, $\frac{\widehat{\text{ATT}}_k - \text{ATT}}{\frac{1}{J}\sum_{T_i\neq\infty}\mathbb{E}\left[\varepsilon_{iT_ik}^2\right]} \xrightarrow{d} N(0,1)$.

Jackknife. Finally, we briefly discuss constructing confidence intervals via the leave-one-unit-out jackknife approach, which proceeds as follows. Fix hyperparameter values ν, ξ, and λ; for each unit $i = 1, \ldots, N$: drop unit i and re-fit the intercepts and the weights via Equation (11) to obtain $\hat{\alpha}^{(-i)}$, $\widehat{\Gamma}^{(-i)}$, and the synthetic control estimates $\hat{Y}_{jT_j+k}^{(-i)}$. Then compute the leave-one-unit-out estimate $\widehat{\text{ATT}}_k^{(-i)} = \frac{1}{J^{(-i)}}\sum_{j=1}^{J}\mathbb{1}_{j\neq i}\left\{Y_{jT_j+k} - \hat{Y}_{jT_j+k}^{(-i)}\right\}$, where $J^{(-i)} \equiv J - \mathbb{1}_{T_i<\infty}$. The jackknife estimate of the standard error is then:

$$\hat{V}_k = \frac{n-1}{n}\sum_{i=1}^{n}\left(\widehat{\text{ATT}}_k^{(-i)} - \frac{1}{n}\sum_{j=1}^{n}\widehat{\text{ATT}}_k^{(-j)}\right)^2, \tag{A.1}$$

with an approximate 95% confidence interval $\widehat{\text{ATT}}_k \pm 1.96\sqrt{\hat{V}_k}$. We include Monte Carlo estimates of the coverage under our simulation setup in Figures B.3 and B.4.

A.2 Fully pooling within time cohorts

As we discuss in Section 3, if all units are treated at the same time, $T_1 = \cdots = T_J$, our error bounds depend only on the pooled imbalance and do not include the unit-level imbalance. Thus, if units are treated in cohorts (i.e., several units treated at the same time), then the bounds suggest modeling variation in pre-treatment outcomes *between* treatment cohorts separately from the pooled average. This leads to a natural modification of our partially pooled estimator: We can fully pool within cohorts by applying the estimator to treatment cohorts rather than individual treated units, optimizing a weighted average of the overall imbalance and the average cohort-level imbalance. Concretely, let G be the number of distinct treatment times, which we denote $T(g)$, $g = 1, \ldots, G$, and let $n_g = \sum_{i=1}^N \mathbb{1}\{T_i = T(g)\}$ be the number of units treated in time $T(g)$. We can modify the optimization problem to find G sets of weights, where the individual objective for treatment cohort g is

$$q_g(\gamma_g)^{\text{cohort}} = \sqrt{\frac{1}{L_g}\sum_{\ell=1}^{L_g}\left(\sum_{i=1}^N \mathbb{1}\{T_j = T(g)\}Y_{iT(g)-\ell} - \sum_{i=1}^N \gamma_{ig} Y_{iT(g)-\ell}\right)^2}.$$

As before, we will restrict the set of donor units for cohort g to those not yet treated K periods after $T(g)$, $\mathcal{D}(g) \equiv \{i : T_i > T(g) + K\}$, and we will restrict the weights so that $\gamma_g \in \Delta^{\text{scm}}(g)$ satisfies $\gamma_{ig} \geq 0$ for all i, $\sum_i \gamma_{ig} = n_g$, and $\gamma_{ig} = 0$ if $i \notin D(g)$. We then define the relevant separate and pooled balance measures:

$$q^{\text{sep cohort}}(\Gamma) = \sqrt{\frac{1}{G}\sum_{g=1}^G \frac{1}{L_g}\sum_{\ell=1}^{L_g}\left(\sum_{i=1}^N \mathbb{1}\{T_j = T(g)\}Y_{iT(g)-\ell} - \sum_{i=1}^N \gamma_{ig} Y_{iT(g)-\ell}\right)^2},$$

and

$$q^{\text{pool cohort}}(\Gamma) = \sqrt{\frac{1}{\max_g L_g}\sum_{\ell=1}^{\max_g L_g}\left(\frac{1}{G}\sum_{g=1}^G\sum_{i=1}^N \mathbb{1}\{T_j = T(g)\}Y_{iT(g)-\ell} - \sum_{i=1}^N \gamma_{ig} Y_{iT(g)-\ell}\right)^2}.$$

We can then use these cohort-level measures of imbalance in the partially pooled SCM optimization problem (6), and similarly can include an intercept as in (7). More generally, if we do not want to fully pool within clusters, we can include three (or more) imbalance terms in our objective function to capture unit-level, pooled, and intermediate cluster-level imbalance.

A.3 Partially pooled SCM: Dual shrinkage

We now inspect the Lagrangian dual problem to the partially pooled SCM problem in Equation (6), showing that the optimization problem partially pools a set of unit-specific dual variables toward global dual variables. We focus on balancing the first $L_j = L \leq T_1 - 1$ lagged outcomes, which are observed for each treated unit.

For each treated unit j, the sum-to-one constraint induces a Lagrange multiplier $\alpha_j \in \mathbb{R}$, and the state-level balance measure induces a set of Lagrange multipliers $\beta_j \in \mathbb{R}^L$, with elements $\beta_{\ell j}$. We combine these dual parameters into a vector $\alpha = [\alpha_1, \ldots, \alpha_J] \in \mathbb{R}^J$ and a matrix $\beta = [\beta_1, \ldots, \beta_J] \in \mathbb{R}^{L \times J}$. In addition to the J sets of Lagrange multipliers — one for each treated unit — the pooled balance measure in the partially pooled SCM problem Equation (6) induces a set of global Lagrange multipliers $\mu_\beta \in \mathbb{R}^L$. As we see in the following proposition, in the dual problem the parameters $\beta_1, \ldots, \beta_J$ are regularized toward this set of pooled Lagrange multipliers, μ_β.

Proposition A.1. The Lagrangian dual to Equation (6) with un-normalized objevtices q^{sep} and q^{pool} with $L_j = L < T_1$ and $\lambda > 0$ is:

$$\min_{\alpha, \mu_\beta, \beta} \mathcal{L}(\alpha, \beta) + \frac{\lambda L}{2} \left(\frac{1}{(1-\nu)} \sum_{j=1}^{J} \|\beta_j - \mu_\beta\|_2^2 + \frac{J}{\nu} \|\mu_\beta\|_2^2 \right), \tag{A.2}$$

where the dual objective function is

$$\mathcal{L}(\alpha, \beta) \equiv \frac{1}{J} \sum_{j=1}^{J} \left[\sum_{i \in \mathcal{D}_j} \left[\alpha_j + \sum_{\ell=1}^{L} \beta_{\ell j} Y_{iT_j - \ell} \right]_+^2 - \left(\alpha_j + \sum_{\ell=1}^{L} \beta_{\ell j} Y_{jT_1 - \ell} \right) \right], \tag{A.3}$$

where $[x]_+ = \max\{0, x\}$. For treated unit j, the synthetic control weight on unit i is $\hat{\gamma}_{ij} = \left[\hat{\alpha}_j + \sum_{\ell=1}^{L} \hat{\beta}_{\ell j} Y_{jT_j - \ell} \right]_+$.

Proposition A.1 highlights that the estimator partially pools the individual synthetic controls to the pooled synthetic control *in the dual parameter space*, with ν controlling the level of pooling. When $\nu = 0$ in the separate SCM problem, the parameters $\beta_1, \ldots \beta_J$ are shrunk towards zero rather than a set of global parameters. By contrast, when $\nu = 1$, $\beta_1, \ldots, \beta_J$ are constrained to be equal to μ_β, fitting a single pooled synthetic control in the dual parameter space. By choosing $\nu \in (0, 1)$, we move continuously between the two extremes of J separate Lagrangian dual problems and a single dual problem, regularizing the individual β_js toward the pooled μ_β, allowing for some limited differences between the J dual parameters.

B Additional figures

B.1 Additional simulation results

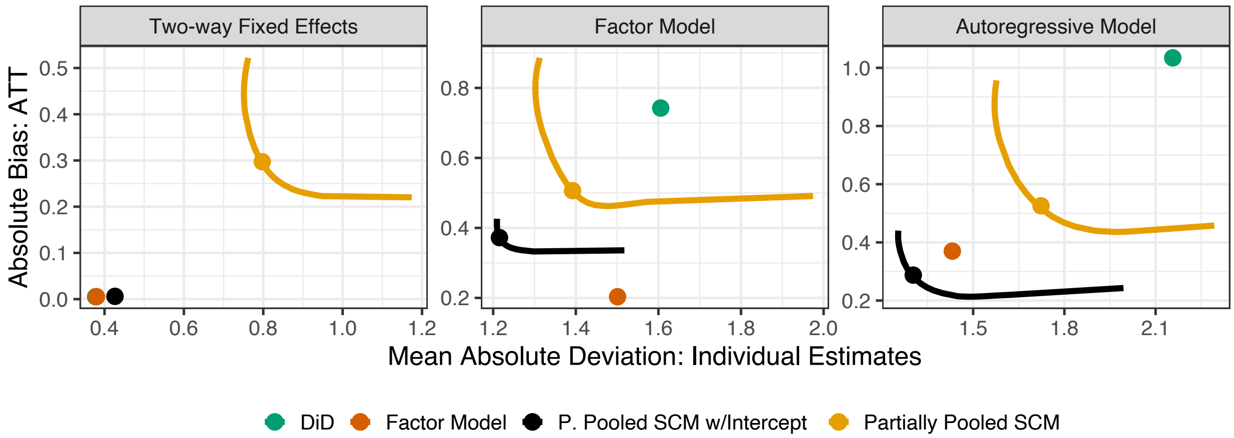

Figure B.1: Monte Carlo estimates of the bias for the overall ATT vs the MAD for the individual ATT estimates.

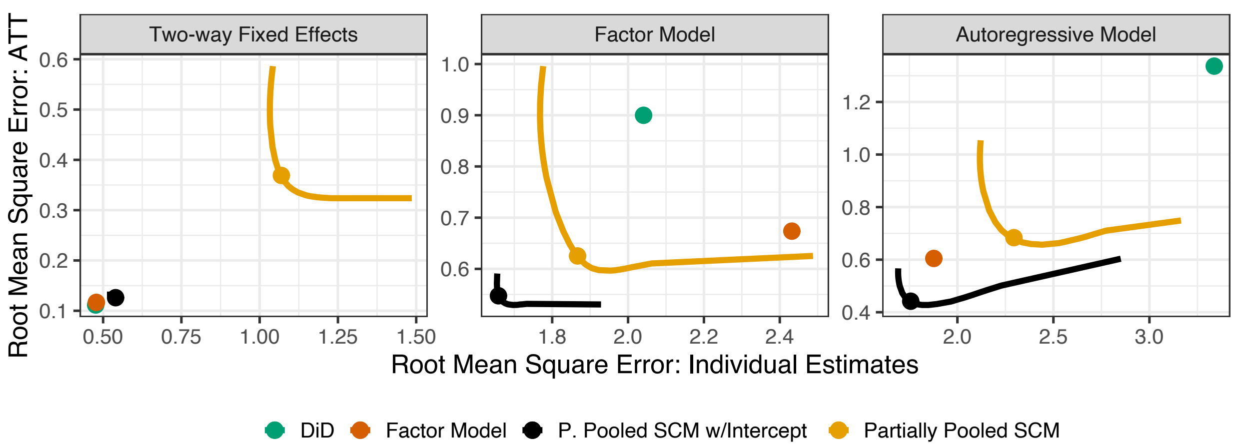

Figure B.2: Monte Carlo estimates of the RMSE for the overall ATT vs the RMSE of the individual ATT estimates.

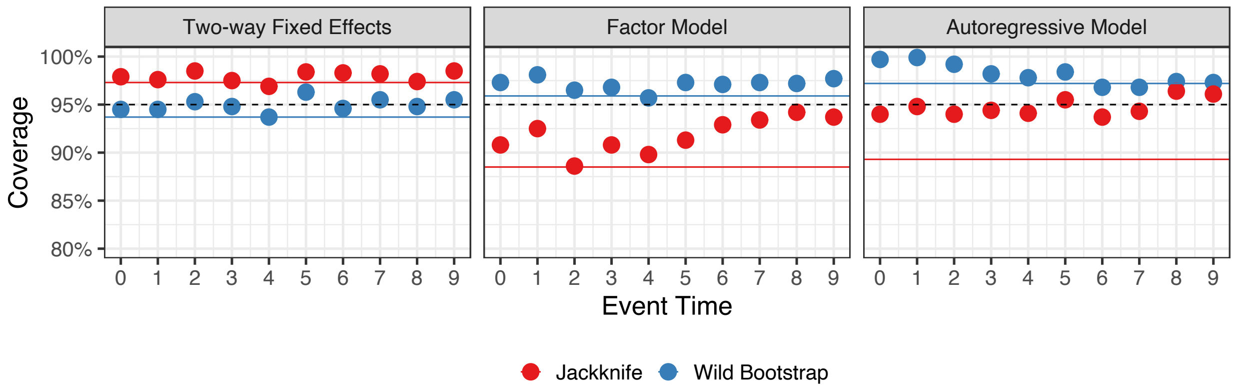

Figure B.3: Monte Carlo estimates of the coverage of approximate 95% confidence intervals $k = 0, \ldots, 9$ periods after treatment using partially pooled SCM with an intercept. The solid line indicates the coverage for the overall ATT estimate averaged across all post-treatment periods.

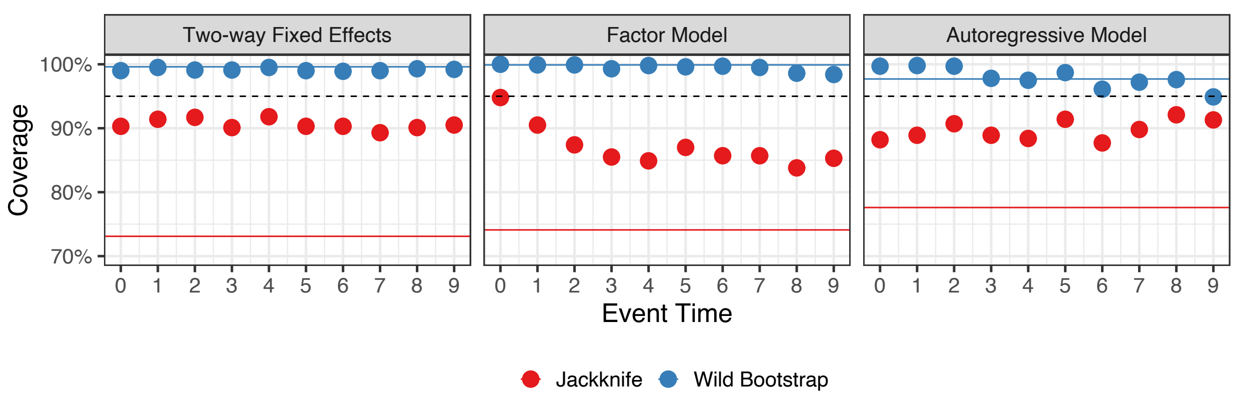

Figure B.4: Monte Carlo estimates of the coverage of approximate 95% confidence intervals $k = 0, \ldots, 9$ periods after treatment using partially pooled SCM *without* an intercept. The solid line indicates the coverage for the overall ATT estimate averaged across all post-treatment periods.

B.2 Additional results for the mandatory collective bargaining application

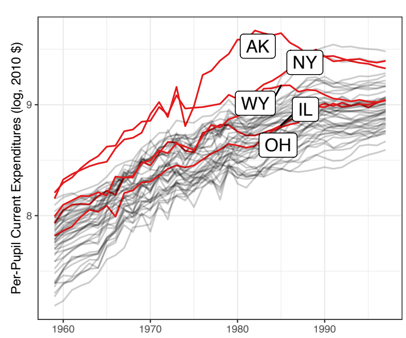

Figure B.5: Per-pupil expenditures for US states over the study period.

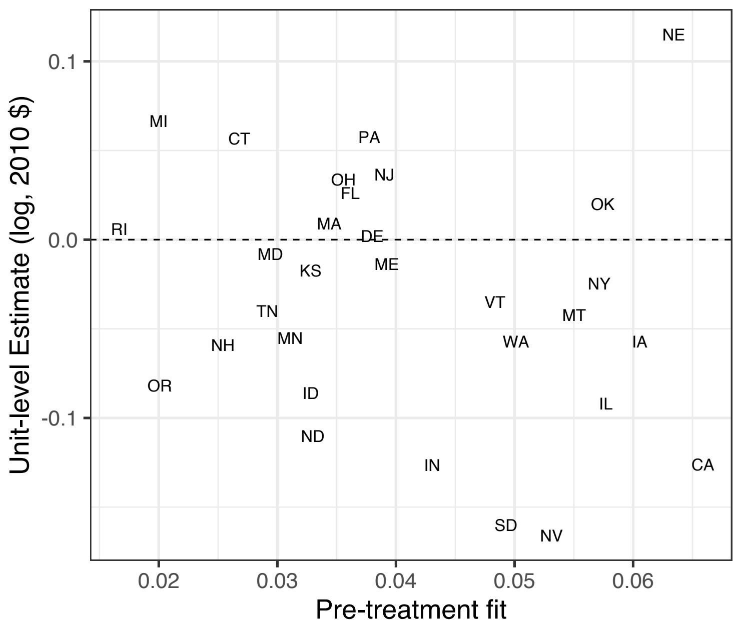

Figure B.6: Average post-treatment effect estimates $\frac{1}{K+1}\sum_{k=0}^{K}\hat{\tau}_{jk}$ for the treated states, plotted against the root-mean square pre-treatment fit $q_j(\hat{\gamma}_j)$.

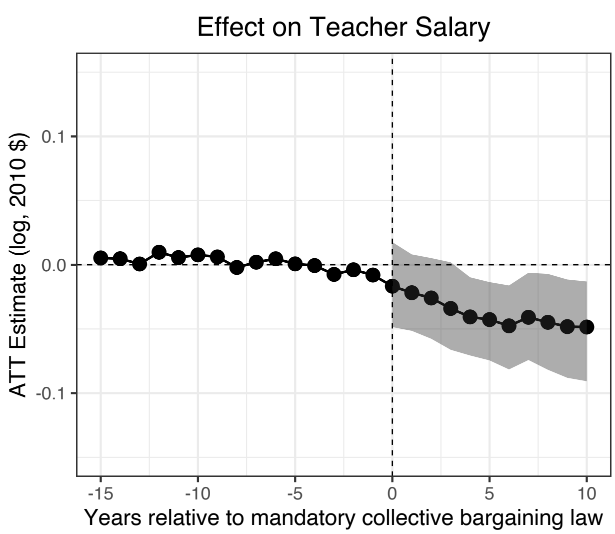

Figure B.7: Partially-pooled SCM with intercept shifts and covariates ($\hat{\nu} = 0.26$), estimates of the impact of mandatory collective bargaining laws on average teacher salary (log, 2010 $).

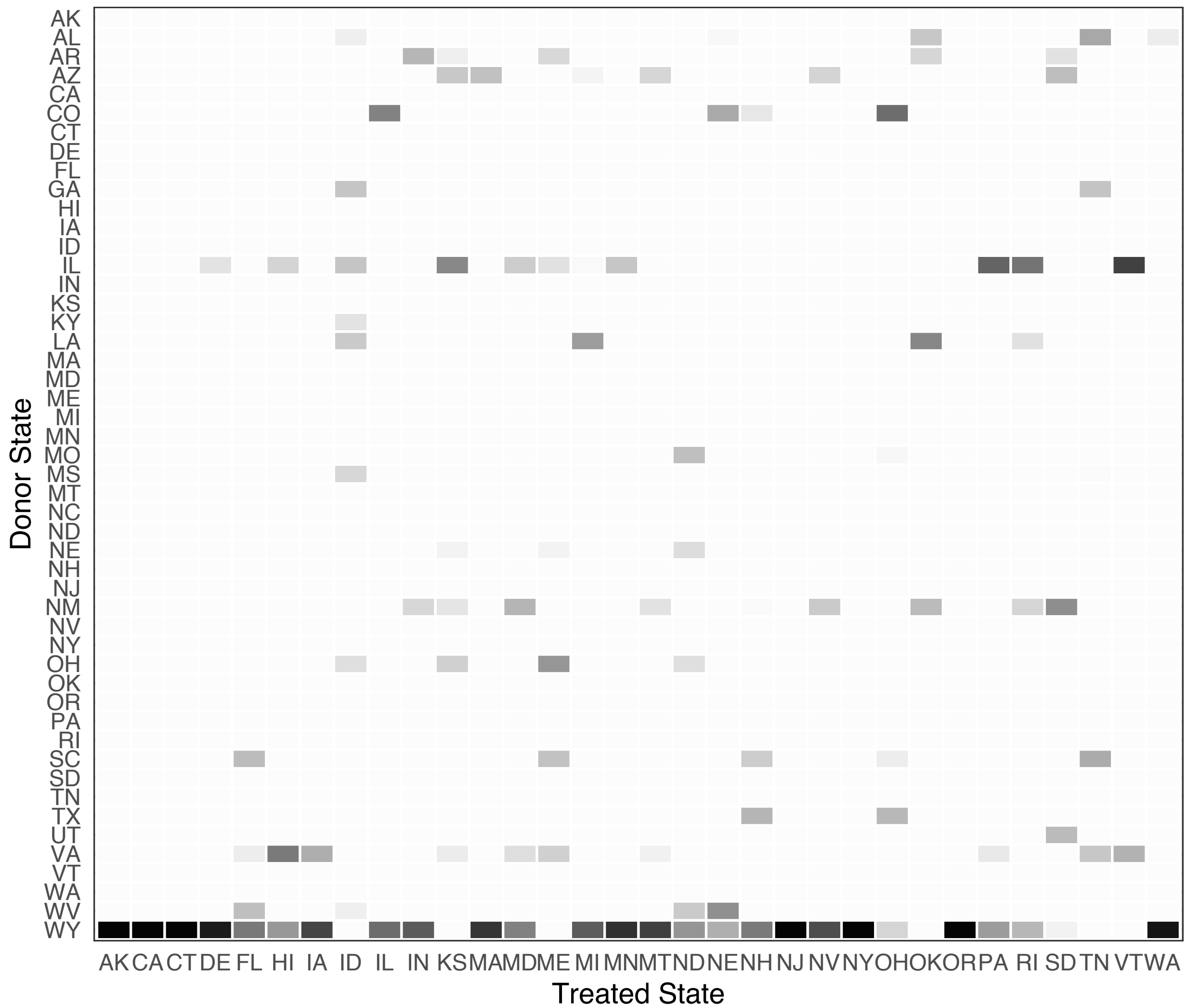

Figure B.8: Partially pooled SCM weights. White cells indicate zero weight, black cells indicate a weight of 1.

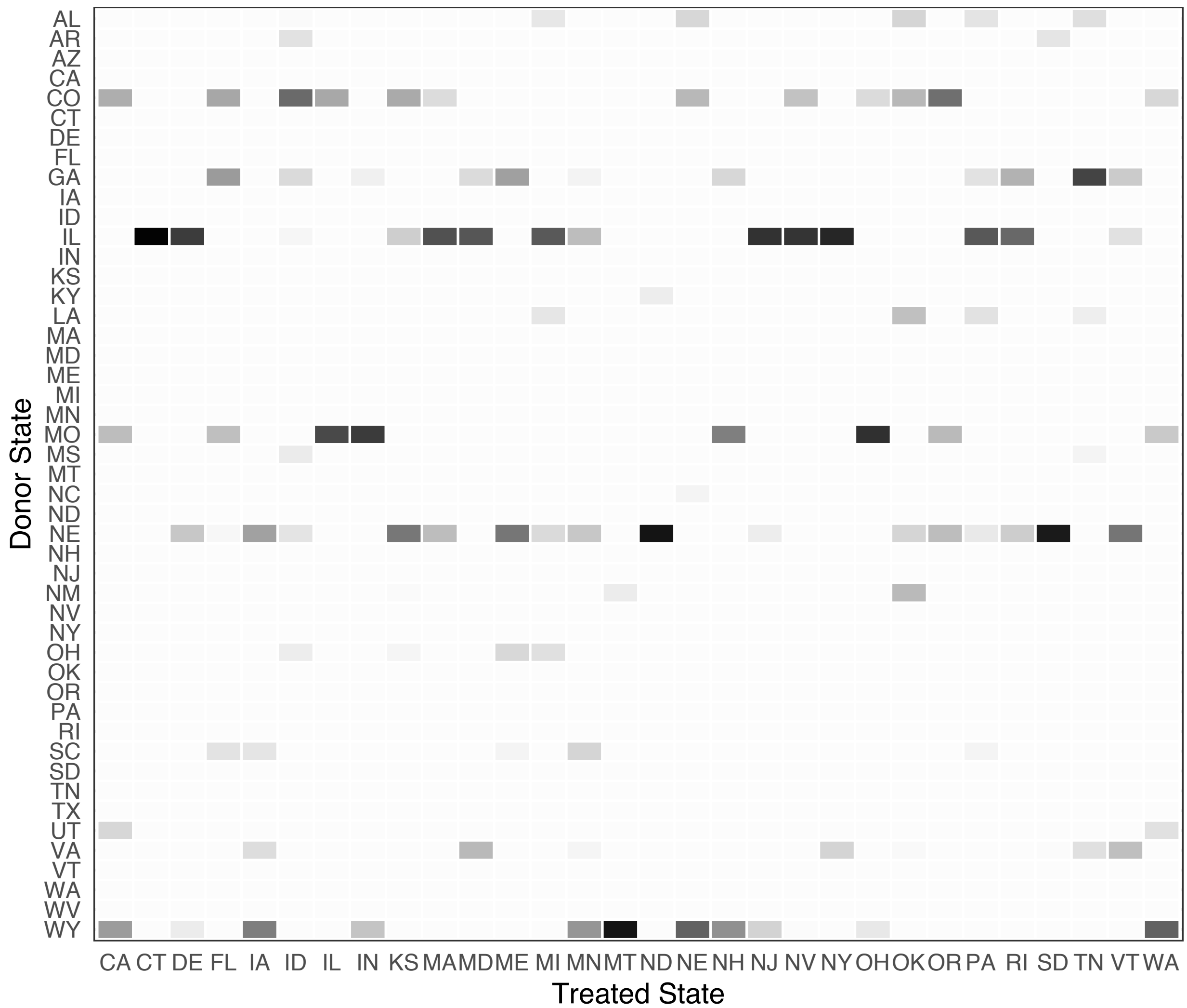

Figure B.9: Partially pooled SCM weights when including an intercept. White cells indicate zero weight, black cells indicate a weight of 1.

C Proofs

C.1 Error bounds

Proof of Theorem 1. Defining $\xi_t = \rho_t - \bar{\rho}$, the error is

$$\hat{\tau}_{j0} - \tau_{j0} = \sum_{\ell=1}^{L} (\bar{\rho} + \xi_{T_j}) \left(Y_{jT_j-\ell} - \sum_{i \in \mathcal{D}_j} \hat{\gamma}_{ij} Y_{iT_j-\ell} \right) + \left(\varepsilon_{jT_j} - \sum_{i \in \mathcal{D}_j} \hat{\gamma}_{ij} \varepsilon_{iT_j} \right)$$

So by the triangle and Cauchy-Schwarz inequalities,

$$|\hat{\tau}_{j0} - \tau_{j0}| \leq \|\bar{\rho} + \xi_{T_j}\|_2 \sqrt{\sum_{\ell=1}^{L} \left(Y_{jT_j-\ell} - \sum_{i \in \mathcal{D}_j} \gamma_{ij} Y_{iT_j-\ell} \right)^2} + \left| \varepsilon_{jT_j} - \sum_{i \in \mathcal{D}_j} \gamma_{ij} \varepsilon_{iT_j} \right|$$

Since $\hat{\gamma}_j$ is fit on pre-T_j outcomes, the weights are independent of ε_{T_j}, and so the second term above is sub-Gaussian with scale parameter $\sigma\sqrt{1 + \|\hat{\gamma}_j\|_2^2} \leq \sigma(1 + \|\hat{\gamma}_j\|_2)$. This implies that

$$P\left(\left| \varepsilon_{jT_j} - \sum_{i \in \mathcal{D}_j} \hat{\gamma}_{ij} \varepsilon_{iT_j} \right| \geq \delta\sigma \left(1 + \|\hat{\gamma}_j\|_2\right) \right) \leq 2 \exp\left(-\frac{\delta^2}{2} \right)$$

For the bound on $\widehat{\text{ATT}}_0$, notice that

$$\begin{aligned}
\widehat{\text{ATT}}_0 - \text{ATT}_0 = \frac{1}{J} \sum_{j=1}^{J} \hat{\tau}_{j0} - \tau_{j0} &= \frac{1}{J} \sum_{j=1}^{J} \left[\sum_{\ell=1}^{L} (\bar{\rho}_\ell + \xi_{T_j \ell}) \left(Y_{jT_j-\ell} - \sum_{i \in \mathcal{D}_j} \hat{\gamma}_{ij} Y_{iT_j-\ell} \right) + \left(\varepsilon_{jT_j} - \sum_{i \in \mathcal{D}_j} \hat{\gamma}_{ij} \varepsilon_{iT_j} \right) \right] \\
&= \sum_{\ell=1}^{L} \bar{\rho}_\ell \frac{1}{J} \sum_{j=1}^{J} \left(Y_{jT_j-\ell} - \sum_{i \in \mathcal{D}_j} \hat{\gamma}_{ij} Y_{iT_j-\ell} \right) \\
&+ \frac{1}{J} \sum_{j=1}^{J} \sum_{\ell=1}^{L} \xi_{T_j \ell} \left(Y_{jT_j-\ell} - \sum_{i \in \mathcal{D}_j} \hat{\gamma}_{ij} Y_{iT_j-\ell} \right) \\
&+ \frac{1}{J} \sum_{j=1}^{J} \left(\varepsilon_{jT_j} - \sum_{i \in \mathcal{D}_j} \hat{\gamma}_{ij} \varepsilon_{iT_j} \right)
\end{aligned} \tag{A.4}$$

By Cauchy-Schwarz the absolute value of the first term is

$$\left| \sum_{\ell=1}^{L} \bar{\rho}_\ell \frac{1}{J} \sum_{j=1}^{J} \left(Y_{jT_j-\ell} - \sum_{i \in \mathcal{D}_j} \hat{\gamma}_{ij} Y_{iT_j-\ell} \right) \right| \leq \|\bar{\rho}\|_2 \sqrt{\sum_{\ell=1}^{L} \left(\frac{1}{J} \sum_{j=1}^{J} \left[Y_{jT_j-\ell} - \sum_{i \in \mathcal{D}_j} \hat{\gamma}_{ij} Y_{iT_j-\ell} \right] \right)^2}.$$

Similarly, the absolute value of the second term is

$$\left|\frac{1}{J}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\xi_{T_j\ell}\left(Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right)\right| \leq \frac{1}{J}\sum_{j=1}^{J}\|\xi_{T_j}\|_2\sqrt{\sum_{\ell=1}^{L}\left(Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right)^2}$$
$$\leq S_\rho\sqrt{\frac{1}{J}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\left(Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right)^2}$$

Finally, notice that $\frac{1}{J}\sum_{j=1}^{J}\varepsilon_{jT_j}$ is the average of J independent sub-Gaussian random variables and so is itself sub-Gaussian with scale parameter $\frac{\sigma}{\sqrt{J}}$. However, $\frac{1}{J}\sum_{j=1}^{J}\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}\varepsilon_{iT_j}$ is the weighted average of sub-Gaussian variables that are independent over i but not necessarily independent over j, and so the weighted average is sub-Gaussian with scale parameter $\frac{\sigma}{\sqrt{J}}\|\Gamma\|_F$. The two averages are independent of each other, so

$$P\left(\frac{1}{J}\sum_{j=1}^{J}\left(\varepsilon_{jT_j}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}\varepsilon_{iT_j}\right)\geq\frac{\delta\sigma}{\sqrt{J}}\left(1+\|\hat{\Gamma}\|_F\right)\right)\leq 2\exp\left(-\frac{\delta^2}{2}\right)$$

Putting together the pieces completes the proof. □

Proof of Theorem 2. Following Abadie et al. (2010), we can re-write ϕ_i in terms of the lagged outcomes as

$$\begin{aligned}\phi_i &= (\Omega_j'\Omega_j)^{-1}\sum_{\ell=1}^{L}\mu_{T_j-\ell}(Y_{iT_j-\ell}-\varepsilon_{iT_j-\ell})\\ &= \frac{1}{\sqrt{L}}\sum_{\ell=1}^{L}P_\ell^{(j)}(Y_{iT_j-\ell}-\varepsilon_{iT_j-\ell})\end{aligned}\tag{A.5}$$

where $\Omega_j\in\mathbb{R}^{L\times F}$ is the matrix of factors from time $t=T_j-L,\ldots,T_j-1$, $\frac{1}{\sqrt{L}}P_\ell^{(j)}=(\Omega_j'\Omega_j)^{-1}\mu_{T_j-\ell}\in\mathbb{R}^F$, and $\frac{1}{\sqrt{L}}P^{(j)}=\frac{1}{\sqrt{L}}[P_1^{(j)},\ldots,P_J^{(j)}]\in\mathbb{R}^{F\times L}$. Using Equation (A.5), we can write the error for the ATT as

$$\begin{aligned}\widehat{\text{ATT}}_k-\text{ATT}_k=\frac{1}{J}\sum_{j=1}^{J}\hat{\tau}_{jk}-\tau_{jk} &= \frac{1}{J\sqrt{L}}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\mu_{T_j+k}'P_\ell^{(j)}\left(Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right)\\ &-\frac{1}{J\sqrt{L}}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\mu_{T_j+k}'P_\ell^{(j)}\left(\varepsilon_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}\varepsilon_{iT_j-\ell}\right)\\ &+\frac{1}{J}\sum_{j=1}^{J}\left(\varepsilon_{jT_j+k}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}\varepsilon_{iT_j+k}\right).\end{aligned}\tag{A.6}$$

From the proof of Theorem 1, we can bound the final term in Equation (A.6). We now bound the first two terms. First, as in the proof of Theorem 1, we decompose the first term into a time constant, and a time varying component:

$$\underbrace{\frac{1}{J\sqrt{L}}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\mu_{T_j+k}' P_\ell^{(j)}\left(Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right)}_{(*)} = \frac{1}{J\sqrt{L}}\sum_{\ell=1}^{L}\bar{\mu}_{k\ell}\sum_{j=1}^{J}\left(Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right)$$

$$+\frac{1}{J\sqrt{L}}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\xi_{(T_j+k)\ell}\left(Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right),$$

where $\bar{\mu}_{k\ell}\equiv\frac{1}{J}\sum_{j=1}^{J}P_\ell^{(j)\prime}\mu_{T_j+k}$, and $\xi_{(T_j+k)\ell}\equiv P_\ell^{(j)\prime}\mu_{T_j+k}-\bar{\mu}_{k\ell}$. Now by Cauchy-Schwarz, we get that

$$|(*)|\leq\|\bar{\mu}_k\|_2\sqrt{\frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{1}{J}\sum_{j=1}^{J}Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right)^2}+\frac{1}{J}\sum_{j=1}^{J}\|\xi_{T_j+k}\|_2\sqrt{\frac{1}{L}\sum_{\ell=1}^{L}\left(Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right)^2}$$

$$\leq\|\bar{\mu}_k\|_2\sqrt{\frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{1}{J}\sum_{j=1}^{J}Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right)^2}+S_k\sqrt{\frac{1}{JL}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\left(Y_{jT_j-\ell}-\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}Y_{iT_j-\ell}\right)^2}$$

We now turn to the second term in Equation (A.6). Since ε_{it} are independent sub-Gaussian random variables and $\frac{1}{\sqrt{L}}\|\mu_{T_j+k}'P^{(j)}\|_2\leq\frac{M^2F}{\sqrt{L}}$,

$$P\left(\frac{1}{\sqrt{L}}\left|\frac{1}{J}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\mu_{T_j+k}'P_\ell^{(j)}\varepsilon_{jT_j-\ell}\right|\geq\frac{\delta\sigma M^2F}{\sqrt{JL}}\right)\leq 2\exp\left(-\frac{\delta^2}{2}\right)$$

Next, since $\hat{\gamma}_1,\ldots,\hat{\gamma}_J\in\Delta^{\text{scm}}$, $\frac{1}{J}\sum_{j=1}^{J}\|\hat{\gamma}_j\|_1=1$, by Hölder's inequality

$$\left|\frac{1}{J\sqrt{L}}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\mu_{T_j+k}'P_\ell^{(j)}\sum_{i\in\mathcal{D}_j}\hat{\gamma}_{ij}\varepsilon_{iT_j-\ell}\right|\leq\max_{j\in\{1,\ldots,J\},i\in\mathcal{D}_j}\left|\frac{1}{\sqrt{L}}\sum_{\ell=1}^{L}\mu_{T_j+k}'P_\ell^{(j)}\varepsilon_{iT_j-\ell}\right|\leq 2\frac{\sigma M^2F}{\sqrt{L}}\left(\sqrt{\log NJ}+\delta\right)$$

where the final inequality holds with probability at least $1-2\exp\left(-\frac{\delta^2}{2}\right)$ by the standard tail bound on the maximum of sub-Gaussian random variables. Putting together the pieces with a union bound completes the proof.

□

C.2 Asymptotic normality

Proof of Theorem A.1. Define $\bar{\beta}_{gk}^Y=\frac{1}{L}\sum_{\ell=1}^{L}\beta_{gk\ell}^Y$ and $\bar{\beta}_{gk}^X=\frac{1}{L}\sum_{\ell=1}^{L}\beta_{gk\ell}^X$. Note that under linearity in Assumption A.2,

$$Y_{ig+k}(\infty)-\frac{1}{L}\sum_{\ell=1}^{L}Y_{ig-\ell}(\infty)=\bar{\beta}_{gk}^Y\cdot\dot{Y}_i^g+\bar{\beta}_{gk}^X\cdot X_i+\varepsilon_{igk}.$$

So the estimation error for the treatment effect for unit j at time k is

$$
\begin{aligned}
\hat{\tau}_{jk} - \tau_{jk} &= Y_{jT_j+k}(\infty) - \frac{1}{L}\sum_{\ell=1}^{L} Y_{iT_j-\ell}(\infty) - \sum_i \hat{\gamma}_{ij}\left(Y_{iT_j+k} - \frac{1}{L}\sum_{\ell=1}^{L} Y_{iT_j-\ell}\right) \\
&= \bar{\beta}^{Y}_{T_jk} \cdot \left(\dot{Y}_j^{T_j} - \sum_i \hat{\gamma}_{ij}\dot{Y}_i^{T_j}\right) + \bar{\beta}^{X}_{T_jk} \cdot \left(X_j - \sum_i \hat{\gamma}_{ij} X_i\right) + \varepsilon_{jT_jk} - \sum_i \hat{\gamma}_{ij}\varepsilon_{iT_jk}
\end{aligned}
$$

Aggregating across treated units we see that

$$
\begin{aligned}
\widehat{\mathrm{ATT}}_k - \mathrm{ATT} &= \frac{1}{J}\sum_{j=1}^{J} \hat{\tau}_{jk} - \tau_{jk} \\
&= \frac{1}{J}\sum_{g=1}^{T_J} n_g \bar{\beta}^{Y}_{gk} \cdot \left(\frac{1}{n_g}\sum_{T_i=g} \dot{Y}_i^g - \frac{1}{n_g}\sum_{i=1}^{N}\sum_{T_j=g} \hat{\gamma}_{ij}\dot{Y}_i^g\right) + n_g \bar{\beta}^{X}_{gk} \cdot \left(\frac{1}{n_g}\sum_{T_i=g} X_i - \frac{1}{n_g}\sum_{i=1}^{N}\sum_{T_j=g} \hat{\gamma}_{ij} X_i\right) \\
&\quad + \frac{1}{J}\sum_{j=1}^{J} \varepsilon_{jT_jk} - \sum_i \hat{\gamma}_{ij}\varepsilon_{iT_jk},
\end{aligned}
$$

where n_g is the number of units treated at time g. Now from Assumption A.3, we have exact balance within each cohort, so this reduces to $\widehat{\mathrm{ATT}}_k - \mathrm{ATT} = \frac{1}{J}\sum_{j=1}^{j} \varepsilon_{jT_jk} - \sum_i \hat{\gamma}_{ij}\varepsilon_{iT_jk}$. We now show that the second term is $o_p(J^{-1/2})$. Denote $\sigma^2_{\max} = \max_{igk} \mathrm{Var}(\varepsilon_{igk})$. Since the noise terms $\varepsilon_{i\ell k}$ are independent across units i,

$$
\begin{aligned}
\mathrm{Var}\left(\frac{1}{J}\sum_{j=1}^{J}\sum_i \varepsilon_{iT_jk}\hat{\gamma}_{ij}\right) &= \mathbb{E}\left[\mathrm{Var}\left(\frac{1}{J}\sum_{j=1}^{J}\sum_i \varepsilon_{iT_jk}\hat{\gamma}_{ij} \mid \Gamma\right)\right] + \mathrm{Var}\left(\mathbb{E}\left[\frac{1}{J}\sum_{j=1}^{J}\sum_i \varepsilon_{igk}\hat{\gamma}_{ij} \mid \Gamma\right]\right) \\
&= \mathbb{E}\left[\frac{1}{J^2}\sum_i \mathrm{Var}\left(\sum_{j=1}^{J} \varepsilon_{iT_jk}\hat{\gamma}_{ij} \mid \Gamma\right)\right] \\
&\leq \mathbb{E}\left[\frac{1}{J^2}\sigma^2_{\max}\sum_{j,j'}\sum_i \hat{\gamma}_{ij}\hat{\gamma}_{ij'}\right] \\
&\leq \mathbb{E}\left[\frac{1}{J^2}\sum_i \sigma^2_{\max}\sum_{j,j'} \|\hat{\gamma}_j\|_2 \|\hat{\gamma}_{j'}\|_2\right] \\
&\leq \frac{C^2\sigma^2_{\max}}{N_0}
\end{aligned}
$$

By Chebyshev's inequality, $P\left(\left|\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\sum_i \varepsilon_{iT_jk}\hat{\gamma}_{ij}\right| \geq \delta\right) \leq \frac{\sigma^2_{\max}C^2 J}{\delta^2 N_0}$. Now since $\frac{J}{N_0} \to 0$, this implies that $\sqrt{J}\left(\widehat{\mathrm{ATT}}_k - \mathrm{ATT}_k\right) = \frac{1}{\sqrt{J}}\sum_{T_i \neq \infty} \varepsilon_{iT_ik} + o_p(1)$. Applying the Lyapunov central limit theorem to the first term and Slutsky's theorem shows asymptotic normality. $\square$

C.3 Partial pooling of dual parameters

Lemma A.1. The Lagrangian dual to Equation (6) with $\nu = 0$, $\lambda > 0$, and $L_j = L < T_1$ is

$$\min_{\alpha,\beta} \underbrace{\frac{1}{J}\sum_{j=1}^{J}\left[\sum_{i\in\mathcal{D}_j}\left[\alpha_j + \sum_{\ell=1}^{L}\beta_{\ell j}Y_{iT_j-\ell}\right]_+^2 - \left(\alpha_j + \sum_{\ell=1}^{L}\beta_{\ell j}Y_{jT_1-\ell}\right)\right]}_{\mathcal{L}(\alpha,\beta)} + \sum_{j=1}^{J}\frac{\lambda L}{2}\|\beta_j\|_2^2, \tag{A.7}$$

The resulting donor weights are $\hat{\gamma}_{ij} = \left[\hat{\alpha}_j - \sum_{\ell=1}^{L}\hat{\beta}_{\ell j}Y_{iT_j-\ell}\right]_+$.

Proof of Lemma A.1. Notice that the separate synth problem separates into J optimization problems:

$$\begin{aligned}
&\min_{\gamma_1,\ldots,\gamma_J\in\Delta_j^{\text{scm}}} \quad \frac{1}{2}q^{\text{sep}}(\Gamma) \;+\; \frac{\lambda}{2}\sum_{j=1}^{J}\sum_{i=1}^{N}\gamma_{ij}^2 \\
=&\sum_{j=1}^{J}\min_{\gamma_j\in\Delta_j^{\text{scm}}}\left\{\left[\frac{1}{2JL}\sum_{\ell=1}^{L}\left(Y_{jT_j-\ell} \;-\; \sum_{i=1}^{N}\gamma_{ij}Y_{iT_j-\ell}\right)^2\right] \;+\; \frac{\lambda}{2}\sum_{i=1}^{N}\gamma_{ij}^2\right\}
\end{aligned} \tag{A.8}$$

Thus the Lagrangian dual objective is the sum of the Langrangian dual objectives of the individual objectives in Equation (A.8). Inserting the dual objectives derived by Ben-Michael et al. (2021) and scaling by $\frac{1}{J}$ yields the result. $\square$

Proof of Proposition A.1. We start be defining auxiliary variables, $\mathcal{E}_0, \mathcal{E}_1, \ldots, \mathcal{E}_J \in \mathbb{R}^L$ where $\mathcal{E}_{j\ell} = Y_{jT_j-\ell} - \sum_{i=1}^{N}\gamma_{ij}Y_{iT_j-\ell}$ for $j \geq 1$ and $\mathcal{E}_{0\ell} = \sum_{T_j>\ell}\left(Y_{jT_j-\ell} - \sum_{i=1}^{N}\gamma_{ij}Y_{iT_j-\ell}\right)$. Additionally we rescale by $\frac{1}{\lambda}$. Then we can write the partially pooled SCM problem (6) as

$$\begin{aligned}
\min_{\gamma_1,\ldots,\gamma_J,\mathcal{E}_0,\ldots,\mathcal{E}_J} \quad & \frac{\nu}{2J^2L\lambda}\sum_{\ell=1}^{L}\mathcal{E}_{0\ell}^2 + \frac{1-\nu}{2J\lambda}\sum_{j=1}^{J}\frac{1}{L}\mathcal{E}_{j\ell}^2 + \sum_{j=1}^{J}\sum_{i=1}^{N}\frac{1}{2}\gamma_{ij}^2 \\
\text{subject to} \quad & \mathcal{E}_{j\ell} = Y_{jT_j-\ell} - \sum_{i=1}^{N}\gamma_{ij}Y_{iT_j-\ell} \\
& \mathcal{E}_{0\ell} = \sum_{T_j>\ell}\left(Y_{jT_j-\ell} - \sum_{i=1}^{N}\gamma_{ij}Y_{iT_j-\ell}\right) \\
& \gamma_j \in \Delta_j^{\text{scm}}
\end{aligned} \tag{A.9}$$

With Lagrange multipliers $\mu_\beta, \zeta_1, \ldots, \zeta_J \in \mathbb{R}^L$ and $\alpha_1, \ldots, \alpha_J \in \mathbb{R}$, the Lagrangian to Equation (A.9) is

$$
\begin{aligned}
&\mathcal{L}(\Gamma, \mathcal{E}_0, \ldots, \mathcal{E}_J, \alpha_1, \ldots, \alpha_J, \mu_\beta, \zeta_1, \ldots, \zeta_J) = \\
&\qquad \sum_{\ell=1}^{L} \left[\frac{\nu}{2LJ^2\lambda} \mathcal{E}_{0\ell}^2 - \mu_{\beta\ell} \left(\sum_{j=1}^{J} Y_{jT_j-\ell} - \sum_{i \in \mathcal{D}_j} \gamma_{ij} Y_{iT_j-\ell} \right) - \mathcal{E}_{0\ell}\mu_{\beta\ell} \right] \\
&\qquad + \quad \sum_{j=1}^{J} \sum_{\ell=1}^{L} \left[\frac{1-\nu}{2JL\lambda} \mathcal{E}_{j\ell}^2 - \zeta_{\ell j} \left(Y_{jT_j-\ell} - \sum_{i \in \mathcal{D}_j} \gamma_{ij} Y_{iT_j-\ell} \right) - \zeta_{\ell j}\mathcal{E}_{j\ell} \right] \\
&\qquad + \quad \sum_{j=1}^{J} \sum_{i \in \mathcal{D}_j} \frac{1}{2}\gamma_{ij}^2 - \alpha_j \gamma_{ij} - \alpha_j
\end{aligned}
$$

Defining $\beta_j = \mu_\beta + \zeta_j$, the dual problem is:

$$
\begin{aligned}
-\min_{\Gamma, \mathcal{E}_0, \mathcal{E}_1, \ldots, \mathcal{E}_J} L(\cdot) = &-\sum_{j=1}^{J} \sum_{i \in \mathcal{D}_j} \min_{\gamma_{ij}} \left\{ \frac{1}{2}\gamma_{ij}^2 - \left(\alpha_j - \sum_{\ell=1}^{L} \beta_{\ell j} Y_{iT_j-\ell} \right) \gamma_{ij} \right\} + \sum_{j=1}^{J} \alpha_j + \sum_{\ell=1}^{L} \beta_{\ell j} Y_{jT_j-\ell} \\
&- \quad \sum_{\ell=1}^{L} \min_{\mathcal{E}_{j\ell}} \left\{ \frac{1-\nu}{2JL\lambda} \mathcal{E}_{j\ell}^2 - \mathcal{E}_{j\ell}(\beta_{\ell j} - \mu_{\beta\ell}) \right\} \\
&- \quad \sum_{\ell=1}^{L} \min_{\mathcal{E}_{0\ell}} \left\{ \frac{\nu}{2J^2L\lambda} \mathcal{E}_{0\ell}^2 - \mathcal{E}_{0\ell}\mu_{\beta\ell} \right\}
\end{aligned}
$$

From Lemma A.1, we see that the first term is $\mathcal{L}(\alpha, \beta)$ and we have the same form for the implied weights. The next two terms are the convex conjugates of a scaled L^2 norm. Using the computation that the convex conjugate of $\frac{a}{2}\|x\|_2^2$ is $\frac{1}{2a}\|x\|_2^2$. We then scale the whole dual problem by $\frac{1}{J}$. Finally, the primal problem (6) is still convex and a primal feasible point exists, so by Slater's condition strong duality holds. □

References

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

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association 105*(490), 493–505.

Ben-Michael, E., A. Feller, and J. Rothstein (2021). The augmented synthetic control method. *Journal of the American Statistical Association* (just-accepted), 1–34.

Callaway, B. and P. H. Sant'Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.

Hazlett, C. and Y. Xu (2018). Trajectory balancing: A general reweighting approach to causal inference with time-series cross-sectional data.

Otsu, T. and Y. Rai (2017). Bootstrap Inference of Matching Estimators for Average Treatment Effects. *Journal of the American Statistical Association 112*, 1720–1732.