A Supplemental Appendix: Revisiting the Synthetic Control Estimator (For Online Publication)

A.1 Proof of the Main Results

A.1.1 Proposition 1

Proof.

The SC weights $\widehat{\mathbf{w}}^{\text{SC}} \in \mathbb{R}^J$ are given by[28]

$$\widehat{\mathbf{w}}^{\text{SC}} = \underset{\mathbf{w}\in\Delta^{J-1}}{\text{argmin}} \frac{1}{T_0} \sum_{t\in\mathcal{T}_0} \left(y_{0t} - \mathbf{y}_t'\mathbf{w}\right)^2. \quad (14)$$

Under Assumptions 1, 2 and 4, the objective function $\widehat{Q}_{T_0}(\mathbf{w}) \equiv \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(y_{0t} - \mathbf{y}_t'\mathbf{w}\right)^2$ converges pointwise in probability to

$$Q_0(\mathbf{w}) \equiv \sigma_\epsilon^2(1+\mathbf{w}'\mathbf{w}) + \left[\left(\mu_0 - \boldsymbol{\mu}'\mathbf{w}\right)'\Omega_0\left(\mu_0 - \boldsymbol{\mu}'\mathbf{w}\right) + \left(c_0 - \mathbf{c}'\mathbf{w}\right)^2\right] \quad (15)$$

which is a continuous and strictly convex function. Therefore, $Q_0(\mathbf{w})$ is uniquely minimized over Δ^{J-1}, and we define its minimum as $\bar{\mathbf{w}}^{\text{SC}} \in \Delta^{J-1}$.

We show that this convergence in probability is uniform over $\mathbf{w} \in \Delta^{J-1}$. Define $\tilde{y}_{0t} = y_{0t} - \delta_t$ and $\tilde{\mathbf{y}}_t = \mathbf{y}_t - \delta_t\mathbf{i}$, where $\mathbf{i}$ is a $J\times 1$ vector of ones. For any $\mathbf{w}', \mathbf{w} \in \Delta^{J-1}$, using the mean value theorem, we can find a $\widetilde{\mathbf{w}} \in \Delta^{J-1}$ such that

$$\begin{aligned} \left|\widehat{Q}_{T_0}(\mathbf{w}') - \widehat{Q}_{T_0}(\mathbf{w})\right| &= \left|2\left(\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\tilde{\mathbf{y}}_t\tilde{y}_{0t} - \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t'\widetilde{\mathbf{w}}\right)\cdot(\mathbf{w}'-\mathbf{w})\right| \\ &\leq \left[\left(2\left|\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\tilde{\mathbf{y}}_t\tilde{y}_{0t}\right|\right| + \left|\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t'\right|\right| \times ||\widetilde{\mathbf{w}}||\right)||\mathbf{w}'-\mathbf{w}||\right]. \quad (16) \end{aligned}$$

Define $B_{T_0} = 2\left|\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\tilde{\mathbf{y}}_t\tilde{y}_{0t}\right|\right| + \left|\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t'\right|\right| \times C$. Since Δ^{J-1} is compact, $||\widetilde{\mathbf{w}}||$ is bounded, so we can find a constant C such that $\left|\widehat{Q}_{T_0}(\mathbf{w}') - \widehat{Q}_{T_0}(\mathbf{w})\right| \leq B_{T_0}\left(||\mathbf{w}'-\mathbf{w}||\right)^{\frac{1}{2}}$. Since $\tilde{y}_{0t}\tilde{\mathbf{y}}_t$ and $\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t'$ are linear combinations of cross products of λ_t and ϵ_{it}, from Assumptions 1, 2, and 4, we have that B_{T_0} converges in probability to a positive constant, so $B_{T_0} = O_p(1)$. Note also that $Q_0(\mathbf{w})$ is uniformly continuous on Δ^{J-1}. Therefore, from Corollary 2.2 of Newey (1991), we have that $\widehat{Q}_{T_0}$ converges uniformly in probability to Q_0. Since Q_0 is uniquely minimized at $\bar{\mathbf{w}}^{\text{SC}}$, Δ^{J-1} is a compact space, Q_0 is continuous and $\widehat{Q}_{T_0}$ converges uniformly to Q_0, from Theorem 2.1

[28] If the number of control units is greater than the number of pre-treatment periods, then the solution to this minimization problem might not be unique. However, since we consider the asymptotics with $T_0 \to \infty$, then we guarantee that, for large enough T_0, the solution will be unique.

of Newey and McFadden (1994), $\widehat{\mathbf{w}}^{\text{SC}}$ exists with probability approaching one, and $\widehat{\mathbf{w}}^{\text{SC}} \xrightarrow{p} \bar{\mathbf{w}}^{\text{SC}}$.

Now we show that $\bar{\mathbf{w}}^{\text{SC}}$ does not generally reconstruct the factor loadings. Note that Q_0 has two parts. The first one reflects that different choices of weights will generate different weighted averages of the idiosyncratic shocks ϵ_{it}. In this simpler case, this part would be minimized when we set all weights equal to $\frac{1}{J}$. Let the $J \times 1$ vector $\mathbf{j_J} = \left(\frac{1}{J}, ..., \frac{1}{J}\right)' \in \Delta^{J-1}$. The second part reflects the presence of common factors λ_t and of the unit fixed effects that would remain after we choose the weights to construct the SC unit. This part is minimized if we choose a $\mathbf{w}^* \in \widetilde{\Phi}$. Suppose that we start at $\mathbf{w}^* \in \Phi$ and move in the direction of $\mathbf{j_J}$, with $\mathbf{w}(\Delta) = \mathbf{w}^* + \Delta(\mathbf{j_J} - \mathbf{w}^*)$. Note that, for all $\Delta \in [0, 1]$, these weights will continue to satisfy the constraints of the minimization problem. If we consider the derivative of function 15 with respect to Δ at $\Delta = 0$, we have that

$$\Gamma'(\mathbf{w}^*) = 2\sigma_\epsilon^2 \left(\frac{1}{J} - \mathbf{w}^{*\prime}\mathbf{w}^*\right) < 0 \text{ unless } \mathbf{w}^* = \mathbf{j_J} \text{ or } \sigma_\epsilon^2 = 0,$$

where we used the fact that $\mathbf{j_J}'\mathbf{w}^* = \frac{1}{J}$, because weights are restricted to sum one.

Therefore, $\mathbf{w}^*$ will not, in general, minimize Q_0. This implies that, when $T_0 \to \infty$, the SC weights will converge in probability to weights $\bar{\mathbf{w}}^{\text{SC}}$ that does not reconstruct the factor loadings of the treated unit, unless it turns out that $\mathbf{w}^*$ also minimizes the variance of this linear combination of the idiosyncratic errors or if $\sigma_\epsilon^2 = 0$.

Now considering the SC estimator,

$$\hat{\alpha}_{0t}^{\text{SC}} = y_{0t} - \mathbf{y}_t\widehat{\mathbf{w}}^{\text{SC}} \xrightarrow{p} \alpha_{0t} + \left(\epsilon_{0t} - \boldsymbol{\epsilon}_t'\bar{\mathbf{w}}^{\text{SC}}\right) + \lambda_t\left(\mu_0 - \boldsymbol{\mu}'\bar{\mathbf{w}}^{\text{SC}}\right) + \left(c_0 - \mathbf{c}'\bar{\mathbf{w}}^{\text{SC}}\right). \quad (17)$$

■

A.1.2 Proposition 2

Proof.

The demeaned SC estimator is given by $\widehat{\mathbf{w}}^{\text{SC}'} = \underset{\mathbf{w}\in\Delta^{J-1}}{\text{argmin}}\ \widehat{Q}'_{T_0}(\mathbf{w})$, where

$$\begin{aligned}\widehat{Q}'_{T_0}(\mathbf{w}) &= \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(y_{0t} - \mathbf{y}_t'\mathbf{w} - \left(\frac{1}{T_0}\sum_{t'\in\mathcal{T}_0} y_{0t'} - \frac{1}{T_0}\sum_{t'\in\mathcal{T}_0}\mathbf{y}_{t'}'\mathbf{w}\right)\right)^2 \\ &= \widehat{Q}_{T_0}(\mathbf{w}) - \left(\frac{1}{T_0}\sum_{t\in\mathcal{T}_0} y_{0t} - \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\mathbf{y}_t'\mathbf{w}\right)^2. \end{aligned} \quad (18)$$

$\widehat{Q}'_{T_0}(\mathbf{w})$ converges pointwise in probability to

$$Q_0'(\mathbf{w}) \equiv \sigma_\epsilon^2(1 + \mathbf{w}'\mathbf{w}) + \left(\mu_0 - \boldsymbol{\mu}'\mathbf{w}\right)'\Omega\left(\mu_0 - \boldsymbol{\mu}'\mathbf{w}\right) \quad (19)$$

where $\Omega_0 - \omega_0'\omega_0$ is positive semi-definite, so $Q_0'(\mathbf{w})$ is a continuous and convex function.

The proof that $\widehat{\mathbf{w}}^{SC'} \xrightarrow{p} \bar{\mathbf{w}}^{SC'}$ where $\bar{\mathbf{w}}^{SC'}$ will generally not reconstruct the factor loadings of the treated unit follows exactly the same steps as the proof of Proposition 1. Therefore

$$\begin{aligned} \hat{\alpha}_{0t}^{SC'} &= y_{0t} - \mathbf{y}_t \widehat{\mathbf{w}}^{SC'} - \left[\frac{1}{T_0} \sum_{t' \in \mathcal{T}_0} y_{0t} - \frac{1}{T_0} \sum_{t' \in \mathcal{T}_0} \mathbf{y}_t' \widehat{\mathbf{w}}^{SC'} \right] && (20) \\ &\xrightarrow{p} \alpha_{0t} + \left(\epsilon_{0t} - \boldsymbol{\epsilon}_t' \bar{\mathbf{w}}^{SC'} \right) + \lambda_t \left(\mu_0 - \boldsymbol{\mu}' \bar{\mathbf{w}}^{SC'} \right). && (21) \end{aligned}$$

■

A.1.3 Proposition 3

Proof.

For any estimator $\hat{\alpha}_{0t}(\widetilde{\mathbf{w}}) = y_{0t} - \mathbf{y}_t \widetilde{\mathbf{w}} - \left[\frac{1}{T_0} \sum_{t' \in \mathcal{T}_0} y_{0t} - \frac{1}{T_0} \sum_{t' \in \mathcal{T}_0} \mathbf{y}_t' \widetilde{\mathbf{w}} \right]$ such that $\widetilde{\mathbf{w}} \xrightarrow{p} \mathbf{w}$, we have that, under Assumptions 1 to 5,

$$a.var(\hat{\alpha}_{0t}(\widetilde{\mathbf{w}})) = \sigma_\epsilon^2 (1 + \mathbf{w}'\mathbf{w}) + (\mu_0 - \boldsymbol{\mu}\mathbf{w})' \, \Omega \, (\mu_0 - \boldsymbol{\mu}\mathbf{w}) = Q_0'(\mathbf{w}), \tag{22}$$

which implies that $a.var(\hat{\alpha}_{0t}^{SC'}) = Q_0'(\bar{\mathbf{w}}^{SC'})$, and $a.var(\hat{\alpha}_{0t}^{DID}) = Q_0'(\bar{\mathbf{w}}^{DID})$. By definition of $\bar{\mathbf{w}}^{SC'}$, it must be that $Q_0'(\bar{\mathbf{w}}^{SC'}) \leq Q_0'(\bar{\mathbf{w}}^{DID})$. ■

A.1.4 Proposition 4

Proof. Consider the trivial identity

$$0 = \left(\bar{\mathbf{w}}^{SC'} - \frac{1}{J}\mathbf{i} \right)' (\mathbf{y}_t - \omega) - \left(\bar{\mathbf{w}}^{SC'} - \frac{1}{J}\mathbf{i} \right)' (\mathbf{y}_t - \omega), \tag{23}$$

where the demeaned SC weights converge to $\bar{\mathbf{w}}^{SC'}$, and $\omega = \mathbb{E}[\mathbf{y}_t - \mathbf{i}\delta_t]$.[29] Note that these two vectors are well defined given the assumption that λ_t and ϵ_{jt} are stationary.

Following the notation from Chernozhukov et al. (2017), we define $P_t^N = \left(\bar{\mathbf{w}}^{SC'} - \frac{1}{J}\mathbf{i} \right)' (\mathbf{y}_t - \omega)$ and $u_t = -\left(\bar{\mathbf{w}}^{SC'} - \frac{1}{J}\mathbf{i} \right)' (\mathbf{y}_t - \omega)$. Note that

$$u_t = -\left(\bar{\mathbf{w}}^{SC'} - \frac{1}{J}\mathbf{i} \right)' (\boldsymbol{\mu}\lambda_t' + \boldsymbol{\epsilon}_t), \tag{24}$$

where we use the fact that $(\bar{\mathbf{w}}^{SC'})'\mathbf{i} = \frac{1}{J}\mathbf{i}'\mathbf{i} = 1$ to eliminate δ_t. Since λ_t and $\boldsymbol{\epsilon}_t$ are weakly dependent stationary with mean zero, we have that u_t is weakly dependent stationary with mean zero.

---

[29] Although we estimate the weights using all treatment periods instead of only the pre-treatment periods, these weights will converge in probability to $\bar{\mathbf{w}}^{SC'}$ because we consider a setting in which $T_0 \to \infty$ while T_1 is fixed.

Now consider

$$\hat{P}_t^N = \left(\widetilde{\mathbf{w}} - \frac{1}{J}\mathbf{i}\right)' \left(\mathbf{y}_t - \frac{1}{T_0+T_1}\sum_{\tau\in\mathcal{T}_0\cup\mathcal{T}_1}\mathbf{y}_\tau\right) = -\hat{u}_t. \tag{25}$$

Note that

$$\hat{P}_t^N - P_t^N = \left(\widetilde{\mathbf{w}} - \frac{1}{J}\mathbf{i}\right)' \left(\frac{1}{T_0+T_1}\sum_{\tau\in\mathcal{T}_0\cup\mathcal{T}_1}\left(\boldsymbol{\mu}\lambda_\tau' + \boldsymbol{\epsilon}_\tau\right)\right) + \left(\widetilde{\mathbf{w}} - \bar{\mathbf{w}}^{\text{SC}'}\right)'\left(\boldsymbol{\mu}\lambda_t' + \boldsymbol{\epsilon}_t\right), \tag{26}$$

where the first term on the RHS of the previous equation is $O_p(1)o_p(1)$, while the second one is $o_p(1)O_p(1)$. Therefore, the model considered in equation 23 satisfies all conditions for Theorem 1 from Chernozhukov et al. (2017). ■

A.2 Case with finite T_0

We consider here the case with T_0 fixed. For weights $\mathbf{w}^* \in \widetilde{\Phi}$, note that:

$$y_{0t} = \mathbf{y}_t'\mathbf{w}^* + \eta_t, \text{ for } t\in\mathcal{I}_0, \text{ where } \eta_t = \epsilon_{0t} - \boldsymbol{\epsilon}_t'\mathbf{w}^* \tag{27}$$

Since $\sum_{j=1}^{J} w_j^* = 1$, we can write:

$$\dot{y}_{0t} = \dot{\mathbf{y}}_t'\dot{\mathbf{w}}^* + \eta_t \tag{28}$$

where $\dot{y}_{jt} = y_{jt} - y_{Jt}$, $\dot{\mathbf{y}}_t = (\dot{y}_{1t}, ..., \dot{y}_{J-1,t})'$, and $\dot{\mathbf{w}}^*$ is the $J-1$ vector excluding the last entry of $\mathbf{w}^*$. The SC weights will be given by the OLS regression in 28 with the non-negativity constraints, and with the constraint that the sum of the $J-1$ weights in $\widehat{\dot{\mathbf{w}}}^*$ must be smaller than 1. We ignore for now these constraints. Then we have that

$$\widehat{\dot{\mathbf{w}}}^* = \left(\sum_{t\in\mathcal{T}_0}\dot{\mathbf{y}}_t\dot{\mathbf{y}}_t'\right)^{-1}\sum_{t\in\mathcal{T}_0}\dot{\mathbf{y}}_t\dot{y}_{0t}. \tag{29}$$

We assume that T_0 is large enough so that $\left(\sum_{t\in\mathcal{T}_0}\dot{\mathbf{y}}_t\dot{\mathbf{y}}_t'\right)$ has full rank. Therefore:

$$\mathbb{E}\left[\widehat{\dot{\mathbf{w}}}^*|\{\dot{\mathbf{y}}_t\}_{t\in\mathcal{T}_0}\right] = \dot{\mathbf{w}}^* + \left(\sum_{t\in\mathcal{T}_0}\dot{\mathbf{y}}_t\dot{\mathbf{y}}_t'\right)^{-1}\sum_{t\in\mathcal{T}_0}\dot{\mathbf{y}}_t\mathbb{E}[\eta_t|\{\dot{\mathbf{y}}_t\}_{t\in\mathcal{T}_0}] \tag{30}$$

By definition of η_t, we have that $\mathbb{E}[\eta_t|\{\dot{\mathbf{y}}_t\}_{t\in\mathcal{T}_0}] \neq 0$ for $t\in\mathcal{I}_0$, which implies that $\widehat{\dot{\mathbf{w}}}^*$ is a biased estimator of $\dot{\mathbf{w}}^*$. Intuitively, the outcomes of the control units work as a proxy to the factor loadings of the treated unit. However, such proxy is imperfect, because the idiosyncratic shocks behave as a measurement error.

If we consider the case without the non-negativity constraints, and assume that λ_t and ϵ_{jt} are i.i.d. normal, then the conditional expectation function of y_{0t} given $\mathbf{y}_t$ would be linear. As a consequence, the expected value of the SC weights would be exactly the same for any T_0, which, in turn, would be the same as the asymptotic value when $T_0 \to \infty$. If we relax the i.i.d. normality assumption and/or include the non-negativity constraints, then $\mathbb{E}\left[\widehat{\dot{\mathbf{w}}}^* | \{\dot{\mathbf{y}}_t\}_{t \in \mathcal{T}_0}\right]$ would not be constant irrespectively of $\{\dot{\mathbf{y}}_t\}_{t \in \mathcal{T}_0}$. However, the $\mathbb{E}\left[\widehat{\dot{\mathbf{w}}}^*\right]$ would be the integral of $\mathbb{E}\left[\widehat{\dot{\mathbf{w}}}^* | \{\dot{\mathbf{y}}_t\}_{t \in \mathcal{T}_0}\right]$ over the distribution of $\{\dot{\mathbf{y}}_t\}_{t \in \mathcal{T}_0}$. Therefore, we have no reason to believe that the distortion in the SC weights would be ameliorated if we consider a finite T_0 setting in comparison to the asymptotic distortion when $T_0 \to \infty$.

Considering the non-negativity constraints would also affect the distribution of $\widehat{\dot{\mathbf{w}}}^*$ because, with finite T_0, there will be a positive probability that the solution to the unrestricted OLS problem will not satisfy the non-negativity constraints. However, this would not change the conclusion that $\widehat{\dot{\mathbf{w}}}^*$ is a biased estimator of $\dot{\mathbf{w}}^*$. In Section 4 we show MC simulations in which the distortion in the SC weights is aggravated when T_0 is small and we consider the non-negativity constraints.

The larger bias of the SC weights when T_0 is smaller is discussed in detail for a particular set of linear factor models considered in the a previous version of our paper (see Ferman and Pinto (2019)). We present there a justification why we should expect (in that particular model) a larger bias for the SC weights when T_0 is finite.

A.3 Setting with diverging common factors

A.3.1 Main Results with diverging common factors

While the assumptions considered in Sections 3.1 and 3.2 allow for outcomes with divergent pre-treatment averages (which would be the case when we consider, for example, GDP or average wages), we restrict to settings in which such diverging common shocks affect all units in the same way. We now consider that case in which we may have diverging common shocks that may have heterogeneous effects across unit. We modify Assumption 1 to include both common shocks that are non-diverging and diverging.

Assumption 1′ (potential outcomes) Potential outcomes are given by

$$\begin{cases} y_{jt}^N = c_j + \delta_t + \gamma_t \theta_j + \lambda_t \mu_j + \epsilon_{jt} \\ y_{jt}^I = \alpha_{jt} + y_{jt}^N. \end{cases} \tag{31}$$

We now separate the factor structure in two parts. One part, $\lambda_t \mu_j$, that has the same properties as considered in Sections 3.1 and 3.2, and another one, $\gamma_t \theta_j$, which are "diverging", in the sense that pre-treatment averages of γ_t diverge.

Assumption 2′ (sampling) We observe a realization of $\{y_{0t}, ..., y_{Jt}\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$, where $y_{jt} = d_{jt} y_{jt}^I + (1 - d_{jt}) y_{jt}^N$, while $d_{jt} = 1$ if $j = 0$ and $t \in \mathcal{T}_1$, and zero otherwise. Potential outcomes are determined by equation (31). We treat $\{\mu_j\}_{j=0}^J$, $\{\theta_j\}_{j=0}^J$, and $\{\gamma_t\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$ as fixed, and $\{\lambda_t\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$ and $\{\epsilon_{jt}\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$ for $j=0,...,J$ as stochastic.

An important difference relative to the setting considered in Sections 3.1 and 3.2 is that we can consider a fixed sequence of γ_t. The idea is that, in this setting, we can find conditions in which the estimator is asymptotically unbiased even conditional on the realization of γ_t.[30] Since in this setting we expect γ_t to diverge as $T_0\rightarrow\infty$, we have to consider the possibility that, for $\tau\in\mathcal{T}_1$, $\gamma_\tau\rightarrow\infty$ when $T_0\rightarrow\infty$.[31] The assumption below imposes restrictions on the sequence of γ_t and on the other common and idiosyncratic shocks. Let $\tilde{\gamma}_t=[1\ \gamma_t]$, $\eta_j=\lambda_t\mu_j+\epsilon_{jt}$, and $\boldsymbol{\eta}_t=(\eta_{1t},...,\eta_{Jt})$, and consider $A=\text{diag}(T_0^{f_1},...,T_0^{f_{F_1}})$ and $\tilde{A}=\text{diag}(1,T_0^{f_1},...,T_0^{f_{F_1}})$ for constants $(f_1,...,f_{F_1})\in\mathbb{R}_+^{F_1}$.

Assumption 4′ (Common and idiosyncratic shocks) $\exists(f_1,...,f_{F_1})\in\mathbb{R}_+^{F_1}$ such that, (i) $T_0^{-1}\sum_{t\in\mathcal{T}_0}[\eta_{0t}\ \boldsymbol{\eta}_t']\rightarrow 0$, (ii) $T_0^{-1}\sum_{t\in\mathcal{T}_0}[\eta_{0t}\ \boldsymbol{\eta}_t']'[\eta_{0t}\ \boldsymbol{\eta}_t']\rightarrow\Sigma$ positive definite, (iii) $T_0^{-1}\sum_{t\in\mathcal{T}_0}\tilde{A}^{-1}\tilde{\gamma}_t'\tilde{\gamma}_t\tilde{A}^{-1}\rightarrow\Omega$ positive definite, (iv) $T_0^{-1}\sum_{t\in\mathcal{T}_0}\tilde{A}^{-1}\tilde{\gamma}_t'\eta_{jt}\rightarrow 0$ for all $j=0,...,J$, and (v) $A^{-1}\gamma_t=O(1)$.

Assumptions 4′(i) and 4′(ii) are equivalent to the assumptions we consider in Sections 3.1 and 3.2 for the "non-diverging" shocks. Assumptions 4′(iii), 4′(iv) and 4′(v) determine the rates in which the components of γ_t diverge. Note that these assumptions would be satisfied if γ_t is a polynomial trend. Moreover, we also show in a previous version of the paper that we can instead assume that γ_t is a combination of $I(1)$ and polynomial trend factors (see Ferman and Pinto (2019)).

We also consider an additional assumption on the factor loadings associated with the non-stationary common trends. Let Θ be the $J\times F_1$ matrix with information on the factor loadings θ_j of the controls.

Assumption 6 (factor loadings) (i) $rank(\Theta)=F_1$, and (ii) $\exists\ \mathbf{w}^*\in W$ such that $\theta_0=\Theta'\mathbf{w}^*$, where W is the set of possible weights given the constrains on the weights the researcher is willing to consider.

The first part of Assumption 6 guarantees that the each diverging common shock generates enough independent variation on the outcomes of the controls. The second part of the assumption assumes existence of weights that reconstruct the factor loadings of unit 0 associated with the non-stationary common trends. If this condition does not hold, then the asymptotic distribution of the SC estimators would trivially depend on the factor structure $\gamma_t\theta_j$. Importantly, we do *not* need to assume existence of weights that satisfy Assumption 6 and also reconstruct μ_0. Let Φ be the set of weights that reconstruct the factor loadings of both the diverging and non-diverging common shocks.

We focus first on the demeaned SC estimator, and then we consider the original SC estimator.

---

[30] In contrast, the conditions for asymptotic unbiasedness considered in Sections 3.1 and 3.2 were valid over the distribution of λ_t.

[31] We can think of that as a triangular array, where we fix a post-treatment periods τ, and γ_τ potentially changes once we increase T_0.

Proposition 5 Under Assumptions 1′, 2′, 3, 4′, and 6, for $\tau \in \mathcal{T}_1$,

$$\hat{\alpha}_{0\tau}^{\text{SC}'} \overset{p}{\to} \alpha_{0\tau} + \left(\epsilon_{0\tau} - \bar{\mathbf{w}}'\boldsymbol{\epsilon}_{\tau}\right) + \lambda_{\tau}\left(\mu_0 - \boldsymbol{\mu}'\bar{\mathbf{w}}\right) \text{ when } T_0 \to \infty \tag{32}$$

where $\mu_0 \neq \boldsymbol{\mu}'\bar{\mathbf{w}}$, unless $\sigma_\epsilon^2 = 0$ or $\Phi \cap \underset{\mathbf{w}\in W}{\text{argmin}}\{\mathbf{w}'\mathbf{w}\} \neq \varnothing$.

We present the proof in Appendix A.3.2. Proposition 5 has two important implications. First, if Assumption 6 is valid, then the asymptotic distribution of the *demeaned* SC estimator does not depend on the diverging common trends. The intuition of this result is the following. As $T_0 \to \infty$ minimizing the variance of a linear combination of the idiosyncratic shocks becomes irrelevant relative to the cost of failing to recover the factor loadings associated with the diverging common shocks. Therefore, we do not have the distortion on the SC weights we find in Section 3.1 when we consider the diverging shocks. Interestingly, while $\widehat{\mathbf{w}}^{\text{SC}'}$ will generally be only $\sqrt{T_0}-$consistent when $\Phi_1 \equiv \{\mathbf{w} \in W | \theta_0 = \Theta'\mathbf{w}^*\}$ is not a singleton, we show in the proof that there are linear combinations of $\widehat{\mathbf{w}}^{\text{SC}'}$ that will converge at a faster rate, implying that $\gamma_t(\theta_0 - \sum_{j\neq 0}\hat{w}_j^{\text{SC}'}\theta_j) \overset{p}{\to} 0$, despite the fact that γ_t explodes when $T_0 \to \infty$. Therefore, such diverging common trends will not lead to asymptotic bias in the SC estimator.

Second, the demeaned SC estimator will be biased if there is correlation between treatment assignment and the non-diverging common factors λ_t. The intuition is that the demeaned SC weights will converge in probability to weights that minimize the asymptotic variance of $u_t = y_{0t} - \mathbf{w}'\mathbf{y}_t = \lambda_t(\mu_0 - \boldsymbol{\mu}'\mathbf{w}) + (\epsilon_{0t} - \mathbf{w}'\boldsymbol{\epsilon}_t)$, restricting to the weights that satisfy Assumption 6. Following the same arguments as in Proposition 1, $\widehat{\mathbf{w}}^{\text{SC}'}$ will not eliminate these non-diverging common factors, unless we have that $\sigma_\epsilon^2 = 0$ or it coincides that there is a $\mathbf{w} \in \Phi$ that also minimizes the linear combination of idiosyncratic shocks.

The result that the asymptotic distribution of the SC estimator does not depend on the non-stationary common trends depends crucially on Assumption 6. If there were no linear combination of the control units that reconstruct the factor loadings of the treated unit associated to the diverging common trends, then the asymptotic distribution of the SC estimator would trivially depend on these common trends, which might lead to bias in the SC estimator if treatment assignment is correlated with such diverging trends.

Proposition 5 remains valid when we relax the adding-up and/or the non-negativity constraints, with minor variations in the conditions for unbiasedness. However, these results are not valid when we consider the no-intercept constraint, as the original SC estimator does. When the intercept is not included, it remains true that $\widehat{\mathbf{w}}^{\text{SC}}$ converges in probability to weights in Φ_1. However, in this case, the weights will not converge fast enough to compensate the fact that γ_t explodes, implying that the result from Proposition 5 that the asymptotic distribution of the estimator does not depend on the diverging common factor does not hold if we consider the estimator with no intercept. We present a counter-example in Appendix A.3.2.

A.3.2 Technical results with diverging common factors

Proof of Proposition 5 without constraints

We show this result for the case without the adding-up, non-negativity, and no intercept constraints. In this case, the time fixed effects δ_t may enter either in the γ_t or in the λ_t vectors. Let $\widehat{\mathbf{w}}$ be the estimator for the weights in this case. We then extend these results for the cases with the adding-up and/or non-negativity constraints. After that, we show a counterexample in which this result is not valid when we use the no intercept constraint.

First, let Θ_a^b contain the rows a to b of matrix Θ. If we set $a = 0$, then the first row of Θ_a^b is given by θ_0'. Since $rank(\Theta) = F_1$, we can assume, without loss of generality, that $rank(\Theta_{J-F_1+1}^J)$ (that is, the last F_1 control units have θ_j that form a basis of $\mathbb{R}^{F_1}$. Therefore, we have

$$\begin{bmatrix} y_{0,t} \\ y_{1,t} \\ \vdots \\ y_{J-F_1,t} \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{J-F_1} \end{bmatrix} - \Theta_0^{J-F_1}(\Theta_{J-F_1+1}^J)^{-1} \begin{bmatrix} c_{J-F_1+1} \\ \vdots \\ c_J \end{bmatrix} + \Theta_0^{J-F_1}(\Theta_{J-F_1+1}^J)^{-1} \begin{bmatrix} y_{J-F_1+1,t} \\ \vdots \\ y_{J,t} \end{bmatrix}$$
$$+ \begin{bmatrix} \eta_{0,t} \\ \eta_{1,t} \\ \vdots \\ \eta_{J-F_1,t} \end{bmatrix} - \Theta_0^{J-F_1}(\Theta_{J-F_1+1}^J)^{-1} \begin{bmatrix} \eta_{J-F_1+1,t} \\ \vdots \\ \eta_{J,t} \end{bmatrix}, \quad (33)$$

which is similar to the triangular representation from Phillips (1991) for cointegrating relations.

We re-write this equation as

$$\begin{bmatrix} y_{0,t} \\ y_{1,t} \\ \vdots \\ y_{J-F_1,t} \end{bmatrix} = \begin{bmatrix} \bar{c}_0 \\ \bar{c}_1 \\ \vdots \\ \bar{c}_{J-F_1} \end{bmatrix} + \Theta_0^{J-F_1} \begin{bmatrix} \tilde{y}_{J-F_1+1,t} \\ \vdots \\ \tilde{y}_{J,t} \end{bmatrix} + \begin{bmatrix} \bar{\eta}_{0,t} \\ \bar{\eta}_{1,t} \\ \vdots \\ \bar{\eta}_{J-F_1,t} \end{bmatrix}. \quad (34)$$

Now define $\beta \in \mathbb{R}^{F_1}$ such that $u_t = \bar{\eta}_{0,t} - [\bar{\eta}_{1,t} \ \dots \ \bar{\eta}_{J-F_1}]'\beta \to_p 0$, and consider the OLS regression of $\bar{\eta}_{0,t}$ on $\bar{\boldsymbol{\eta}}_t \equiv (\bar{\eta}_{1,t}, \dots, \bar{\eta}_{J-F_1})$, a constant, and $\tilde{\mathbf{y}}_t \equiv (\tilde{y}_{J-F_1+1,t}, \dots, \tilde{y}_{J,t})$. The OLS estimators ($\hat{\beta}$, $\hat{\kappa}$. and $\hat{\phi}$) are given by

$$\begin{bmatrix} \hat{\beta} - \beta \\ \hat{\kappa} \\ A\hat{\phi} \end{bmatrix} = \begin{bmatrix} T_0^{-1}\sum_{t\in\mathcal{T}_0} \bar{\boldsymbol{\eta}}_t\bar{\boldsymbol{\eta}}_t' & T_0^{-1}\sum_{t\in\mathcal{T}_0} \bar{\boldsymbol{\eta}}_t & T_0^{-1}\sum_{t\in\mathcal{T}_0} \bar{\boldsymbol{\eta}}_t \left(A^{-1}\tilde{\mathbf{y}}_t\right)' \\ T_0^{-1}\sum_{t\in\mathcal{T}_0} \bar{\boldsymbol{\eta}}_t' & 1 & T_0^{-1}\sum_{t\in\mathcal{T}_0} \left(A^{-1}\tilde{\mathbf{y}}_t\right)' \\ T_0^{-1}\sum_{t\in\mathcal{T}_0} \left(A^{-1}\tilde{\mathbf{y}}_t\right)\bar{\boldsymbol{\eta}}_t' & T_0^{-1}\sum_{t\in\mathcal{T}_0} \left(A^{-1}\tilde{\mathbf{y}}_t\right) & T_0^{-1}\sum_{t\in\mathcal{T}_0} \left(A^{-1}\tilde{\mathbf{y}}_t\right)\left(A^{-1}\tilde{\mathbf{y}}_t\right)' \end{bmatrix}^{-1} \times$$
$$\times \begin{bmatrix} T_0^{-1}\sum_{t\in\mathcal{T}_0} \bar{\boldsymbol{\eta}}_t u_t \\ T_0^{-1}\sum_{t\in\mathcal{T}_0} u_t \\ T_0^{-1}\sum_{t\in\mathcal{T}_0} \tilde{\mathbf{y}}_t u_t \end{bmatrix}. \quad (35)$$

From Assumption 4′, we have that

$$\begin{bmatrix} T_0^{-1}\sum_{t\in\mathcal{T}_0}\bar{\boldsymbol{\eta}}_t\bar{\boldsymbol{\eta}}_t' & T_0^{-1}\sum_{t\in\mathcal{T}_0}\bar{\boldsymbol{\eta}}_t & T_0^{-1}\sum_{t\in\mathcal{T}_0}\bar{\boldsymbol{\eta}}_t\left(A^{-1}\tilde{\mathbf{y}}_t\right)' \\ T_0^{-1}\sum_{t\in\mathcal{T}_0}\bar{\boldsymbol{\eta}}_t' & 1 & T_0^{-1}\sum_{t\in\mathcal{T}_0}\left(A^{-1}\tilde{\mathbf{y}}_t\right)' \\ T_0^{-1}\sum_{t\in\mathcal{T}_0}\left(A^{-1}\tilde{\mathbf{y}}_t\right)\bar{\boldsymbol{\eta}}_t' & T_0^{-1}\sum_{t\in\mathcal{T}_0}\left(A^{-1}\tilde{\mathbf{y}}_t\right) & T_0^{-1}\sum_{t\in\mathcal{T}_0}\left(A^{-1}\tilde{\mathbf{y}}_t\right)\left(A^{-1}\tilde{\mathbf{y}}_t\right)' \end{bmatrix} \to_p \begin{bmatrix} \Sigma & 0 \\ 0 & \Omega \end{bmatrix}, \quad (36)$$

which is positive definite, and we also have that

$$\begin{bmatrix} T_0^{-1}\sum_{t\in\mathcal{T}_0}\bar{\boldsymbol{\eta}}_t u_t \\ T_0^{-1}\sum_{t\in\mathcal{T}_0}u_t \\ T_0^{-1}\sum_{t\in\mathcal{T}_0}\tilde{\mathbf{y}}_t u_t \end{bmatrix} \to_p 0. \quad (37)$$

Therefore,

$$\begin{bmatrix} \hat{\beta}-\beta \\ \hat{\kappa} \\ A\hat{\phi} \end{bmatrix} \to_p 0. \quad (38)$$

Now note that, from equation (34), we have that

$$y_{0,t} = \begin{bmatrix} 1 & -\hat{\beta}' \end{bmatrix} \begin{bmatrix} \bar{c}_0 \\ \bar{c}_1 \\ \vdots \\ \bar{c}_{J-F_1} \end{bmatrix} + \hat{\kappa} + \hat{\beta}' \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{J-F_1,t} \end{bmatrix} + \quad (39)$$

$$+ \left(\begin{bmatrix} 1 & -\hat{\beta}' \end{bmatrix} \Theta_0^{J-F_1}(\Theta_{J-F_1+1}^{J})^{-1} + \hat{\phi}(\Theta_{J-F_1+1}^{J})^{-1} \right) \begin{bmatrix} y_{J-F_1+1,t} \\ \vdots \\ y_{J,t} \end{bmatrix} + \hat{u}_t, \quad (40)$$

which implies that an OLS regression of $y_{0,t}$ on a constant, $(y_{1,t},\dots,y_{J-F_1})$, and $(y_{J-F_1+1,t},\dots,y_{J,t})$ yields estimators $\hat{c} = \begin{bmatrix} 1 & -\hat{\beta}' \end{bmatrix}\begin{bmatrix} \bar{c}_0 & \bar{c}_1 & \dots & \bar{c}_{J-F_1} \end{bmatrix}' + \hat{\kappa}$, $\hat{\beta}$, and $\left(\begin{bmatrix} 1 & -\hat{\beta}' \end{bmatrix} \Theta_0^{J-F_1}(\Theta_{J-F_1+1}^{J})^{-1} + \hat{\phi}(\Theta_{J-F_1+1}^{J})^{-1} \right)$.

We are interested in the limiting distribution of $\hat{\alpha}_{0\tau}$, for $\tau \in \mathcal{T}_1$:

$$\begin{aligned} \hat{\alpha}_{0\tau} &= y_{0\tau} - \mathbf{y}_\tau'\widehat{\mathbf{w}} = \alpha_{0\tau} + \lambda_\tau\left(\mu_0 - \boldsymbol{\mu}'\widehat{\mathbf{w}}\right) + \gamma_\tau\left(\theta_0 - \Theta'\widehat{\mathbf{w}}\right) + \left(\epsilon_{0\tau} - \boldsymbol{\epsilon}_\tau'\widehat{\mathbf{w}}\right) \\ &\quad + c_0 - [c_1 \ \dots \ c_J]\widehat{\mathbf{w}} - \hat{c}. \end{aligned} \quad (41)$$

With some algebra, we have that

$$\gamma_\tau\left(\theta_0 - \Theta'\widehat{\mathbf{w}}\right) = \gamma_\tau\hat{\phi} = (\gamma_\tau A^{-1})(A\hat{\phi}) = o_p(1). \quad (42)$$

Likewise, we have that

$$c_0 - [c_1 \ \dots \ c_J]\widehat{\mathbf{w}} - \hat{c} = \hat{\kappa} = o_p(1), \quad (43)$$

implying that

$$\hat{\alpha}_{0\tau} \to_p \alpha_{0\tau} + \lambda_\tau \left(\mu_0 - \boldsymbol{\mu}'\bar{\mathbf{w}}\right) + \left(\epsilon_{0\tau} - \boldsymbol{\epsilon}_\tau'\bar{\mathbf{w}}\right). \quad (44)$$

Finally, by definition of u_t, the OLS estimator converges to weights that minimize plim$[(y_{0t} - \mathbf{y}_t'\mathbf{w})^2]$ subject to $\mathbf{w} \in \Phi_1$. Therefore, the proof that $\widehat{\mathbf{w}} \xrightarrow{p} \bar{\mathbf{w}} \notin \Phi$ is essentially the same as the proof of Proposition 1.

Proof of Proposition 5 with adding-up and non-negativity constraints

To show that this result is also valid for the case with adding-up constraint we just have to consider the OLS regression of $y_{0t} - y_{1t}$ on a constant and $y_{2t} - y_{1t}, ..., y_{Jt} - y_{1t}$. Under Assumption 6, this transformed model is also cointegrated, so we can apply our previous result.

We now consider the case with the non-negative constraints. We prove the case $W = \{\mathbf{w} \in \mathbb{R}^J \mid w_j \geq 0\}$. Including an adding-up constraint then follows directly from a change in variables as we did for the case without non-negative constraints. Let $\widehat{\mathbf{w}}$ be such estimator for the weights.

We first show that $\widehat{\mathbf{w}} \xrightarrow{p} \bar{\mathbf{w}}$ where $\bar{\mathbf{w}}$ minimizes $\mathbb{E}[u_t^2]$ subject to $\mathbf{w} \in \Phi_1 \cap W$. Suppose that $\bar{\mathbf{w}} \in int(W)$. This implies that $\bar{\mathbf{w}} \in int(\Phi_1 \cap W)$ relative to Φ_1. By convexity of $E[u_t^2]$, $\bar{\mathbf{w}}$ also minimizes $E[u_t^2]$ subject to Φ_1. We know that OLS without the non-negativity constraints converges in probability to $\bar{\mathbf{w}}$. Let $\widehat{\mathbf{w}}_u$ be the OLS estimator without the non-negativity constraints and $\widehat{\mathbf{w}}_r$ be the OLS estimator with the non-negativity constraint. Since $\bar{\mathbf{w}} \in int(W)$, then it must be that, for all $\epsilon > 0$, $||\widehat{\mathbf{w}}_u - \bar{\mathbf{w}}|| < \epsilon$ with probability approaching to 1 (w.p.a.1). Since $\widehat{\mathbf{w}}_u = \widehat{\mathbf{w}}_r$ when $\widehat{\mathbf{w}}_u \in int(W)$ (due to convexity of the OLS objective function), these two estimators are asymptotically equivalent.

Consider now the case in which $\bar{\mathbf{w}}$ is on the boundary of W. This means that $\bar{w}_j = 0$ for at least one j. Let $A = \{j | w_j^* = 0\}$. Note first that $\bar{\mathbf{w}}$ also minimizes $E[u_t^2]$ subject to $\mathbf{w} \in \Phi_1 \cap \{\mathbf{w} | w_j = 0 \; \forall j \in A\}$. That is, if we impose the restriction $w_j = 0$ for all j such that $\bar{w}_j = 0$, then we would have the same minimizer, even if we ignore the other non-negative constraints. Suppose there is an $\tilde{\mathbf{w}} \neq \bar{\mathbf{w}}$ that minimizes $E[u_t^2]$ subject to $\mathbf{w} \in \Phi_1 \cap \{\mathbf{w} | w_j = 0 \; \forall j \in A\}$. By strict convexity of the objective function and the fact that $\bar{\mathbf{w}}$ is in the interior of $\Phi \cap W \cap \{\mathbf{w} | w_j = 0 \; \forall j \in A\}$ relative to $\Phi_1 \cap \{\mathbf{w} | w_j = 0 \; \forall j \in A\}$, there must be $\mathbf{w}' \in \Phi_1 \cap W \cap \{\mathbf{w} | w_j = 0 \; \forall j \in A\} \subset \Phi_1 \cap W$ that attains a lower value in the objective function than $\bar{\mathbf{w}}$. However, this contradicts the fact that $\bar{\mathbf{w}} \in \Phi_1 \cap W$ is the minimum.

Now let $\widehat{\mathbf{w}}'$ be the OLS estimator subject to $\{\mathbf{w} | w_j = 0 \; \forall j \in A\}$. We have that $\widehat{\mathbf{w}}'$ is consistent for $\bar{\mathbf{w}}$. Now we show that $\widehat{\mathbf{w}}'$ is asymptotically equivalent to $\widehat{\mathbf{w}}''$, the OLS estimator subject to $\{\mathbf{w} | w_j \geq 0 \; \forall j \in A\}$. We prove the case in which $A = \{j\}$ (there is only one restriction that binds). The general case follows by induction. Suppose these two estimators are not asymptotically equivalent. Then there is $\epsilon > 0$ such that $LimPr(|\widehat{\mathbf{w}}' - \widehat{\mathbf{w}}''| > \epsilon) \neq 0$. There are two possible cases.

First, suppose that $\text{Lim}Pr\left(|\widehat{w}_j''| > \epsilon'\right) = 0$ for all $\epsilon' > 0$ (that is, the OLS subject to $\{\mathbf{w} | w_j \geq 0 \; \forall j \in A\}$ converges in probability to $\bar{\mathbf{w}}$ such that $\bar{w}_j = 0$). However, since the two estimators are not asymptotically equivalent, for all T_0', we can always find a $T_0 > T_0'$ such that, with positive probability, $|\widehat{\mathbf{w}}' - \widehat{\mathbf{w}}''| > \epsilon$. Since $\{\mathbf{w}|w_j = 0 \ \forall j \in A\} \subset \{\mathbf{w}|w_j \geq 0 \ \forall j \in A\}$ and $\widehat{\mathbf{w}}' \neq \widehat{\mathbf{w}}''$, then $Q_{T_0}(\widehat{\mathbf{w}}'') < Q_{T_0}(\widehat{\mathbf{w}}')$, where $Q_{T_0}()$ is the OLS objective function. Now using the continuity of the OLS objective function and the fact that $\widehat{w}_j''$ converges in probability to zero, we can always find T_0' such that there will be a positive probability that $Q_{T_0}(\widehat{\mathbf{w}}'' - e_j\hat{w}_j'') < Q_{T_0}(\widehat{\mathbf{w}}')$. Since $\widehat{\mathbf{w}}'' - e_j\hat{w}_j'' \in \{\mathbf{w}|w_j = 0 \ \forall j \in A\}$, this contradicts $\widehat{\mathbf{w}}'$ being OLS subject to $\{\mathbf{w}|w_j = 0 \ \forall j \in A\}$.

Alternatively, suppose that there exists $\epsilon' > 0$ such that $\text{Lim}Pr\left(|\widehat{w}_j''| > \epsilon'\right) \neq 0$. This means that, for all T_0', we can find $T_0 > T_0'$ such that there is a positive probability that the solution to OLS on $\{\mathbf{w}|w_j \geq 0 \ \forall j \in A\}$ is in an interior point $\widehat{\mathbf{w}}''$ with $\hat{w}_j'' > \epsilon' > 0$. By convexity of $Q_{T_0}()$, this would imply that $\widehat{\mathbf{w}}''$ is also the solution to the OLS without any restriction. However, this contradicts the fact that OLS without non-negativity restriction is consistent (see proof of Proposition 5).

Finally, we show that $\widehat{\mathbf{w}}''$ and $\widehat{\mathbf{w}}_r$ are asymptotically equivalent. Note that $\bar{\mathbf{w}}$ is in the interior of W relative to $\{\mathbf{w}|w_j \geq 0 \ \forall j \in A\}$. Therefore, w.p.a.1, $\widehat{\mathbf{w}}'' \in W$, which implies that $\widehat{\mathbf{w}}'' = \widehat{\mathbf{w}}_r$.

We still need to show that linear combinations of $\widehat{\mathbf{w}}_r$ converge fast enough to reconstruct the factor loadings of the treated unit associated with the non-stationary common factors, so that $\gamma_t(\theta_0 - \sum_{j\neq 0}\hat{w}_j^r\theta_j) \xrightarrow{p} 0$. Let $Q_{T_0}()$ be the OLS objective function, and let $\widetilde{\mathcal{W}} = \{\widetilde{\mathbf{w}}_1, ..., \widetilde{\mathbf{w}}_{2^J}\}$ be the set of all possible OLS estimators when we consider some of the non-negative constraints as equality and ignore the other ones. Let $\widetilde{\mathcal{W}}' \subset \widetilde{\mathcal{W}}$ be the set of estimators in $\widetilde{\mathcal{W}}$ such that all non-negative constraints are satisfied. Then we know that $\widehat{\mathbf{w}}_r = argmin_{\mathbf{w}\in\widetilde{\mathcal{W}}'}Q_{T_0}(\mathbf{w})$.

Suppose first that, for each of the 2^J combinations of restrictions, there is at least one $\mathbf{w} \in \Phi_1$ that satisfy these restrictions. In this case, we know from the first part of the proof that $\gamma_t\left(\theta_0 - \sum_{j\neq 0}\widetilde{w}_j^h\theta_j\right) \xrightarrow{p} 0$ for all $h = 1, ..., 2^J$, where $\widetilde{\mathbf{w}}_h = (\widetilde{w}_1^h, ..., \widetilde{w}_J^h)'$. Moreover, since $\widetilde{\mathcal{W}}$ is finite, then this convergence is uniform in $\widetilde{\mathcal{W}}$. Therefore, it must be that $\gamma_t(\theta_0 - \sum_{j\neq 0}\hat{w}_j^r\theta_j) \xrightarrow{p} 0$. Suppose now that for the combination of restrictions considered for $\widetilde{\mathbf{w}}_h$, with $h \in \{1, ..., 2^J\}$, there is no $\mathbf{w} \in \Phi_1$ that satisfies these restrictions. Since the parameter space with this combination of restrictions is closed, then $\exists\eta > 0$ such that $||\theta_0 - \sum_{j\neq 0}w_j\theta_j|| > \eta$ for all $\mathbf{w}$ that satisfy this combinations of restrictions.[32] Therefore, $Q_{T_0}(\widetilde{\mathbf{w}}_h)$ diverge when $T_0 \to \infty$, implying that, w.p.a.1, $\widehat{\mathbf{w}}_r \neq \widetilde{\mathbf{w}}_h$.

Example with no intercept

We consider now a very simple example to show that it is not possible to guarantee that $\gamma_t\left(\theta_0 - \sum_{j\neq 0}\hat{w}_j\theta_j\right) \xrightarrow{p} 0$ if we do not include the intercept. Consider the case in which there are only one treated and one control unit, and $y_{0t} = \mu_0 + t + u_{0t}$ while $y_{1t} = \mu_1 + t + u_{1t}$. We consider a regression of y_{0t} on y_{1t} without the intercept. Note that $y_{0t} = (\mu_0 - \mu_1) + y_{1t} + u_{0t} - u_{1t} = \mu + y_{1t} + u_t$. Then we have that:

$$\hat{\beta} = \frac{\sum_{t=1}^{T_0} y_{1t}y_{0t}}{\sum_{t=1}^{T_0} y_{1t}^2} = 1 + \frac{\sum_{t=1}^{T_0}(\mu\mu_1 + \mu t + \mu u_{1t} + \mu_1 u_t + tu_t + u_t u_{1t})}{\sum_{t=1}^{T_0}(t^2 + \mu_1^2 + u_{1t}^2 + \text{“cross terms”})} \tag{45}$$

[32] Otherwise, there would be $\mathbf{w} \in \Phi_1$ that satisfies this combination of restrictions.

which implies that:

$$T(\hat{\beta}-1)=\frac{\frac{1}{T^2}\sum_{t=1}^{T_0}(\mu\mu_1+\mu t+\mu u_{1t}+\mu_1 u_t+tu_t+u_t u_{1t})}{\frac{1}{T^3}\sum_{t=1}^{T_0}(t^2+\mu_1^2+u_{1t}^2+\text{“cross terms”})}\overset{p}{\to}\frac{\frac{1}{2}\mu}{\frac{1}{3}} \tag{46}$$

Therefore, while $\hat{\beta}\overset{p}{\to}1$, it does not converge fast enough so that $T(\hat{\beta}-1)\overset{p}{\to}0$, except when $\mu_0=\mu_1$.

A.4 Example: SC Estimator vs DID Estimator

We provide an example in which the asymptotic bias of the SC estimator can be higher than the asymptotic bias of the DID estimator. Assume we have 1 treated and 4 control units in a model with 2 common factors. For simplicity, assume that there is no additive fixed effects and that $\mathbb{E}[\lambda_t]=0$. We have that the factor loadings are given by:

$$\mu_0=\begin{pmatrix}1\\1\end{pmatrix},\ \mu_2=\begin{pmatrix}0.5\\1\end{pmatrix},\ \mu_3=\begin{pmatrix}1.5\\1\end{pmatrix},\ \mu_4=\begin{pmatrix}0.5\\0\end{pmatrix},\ \mu_5=\begin{pmatrix}1.5\\1\end{pmatrix} \tag{47}$$

Note that any linear combination $0.5\mu_2+w_1^3\mu_3+w_1^5\mu_5$ with $w_1^3+w_1^5=0.5$ recovers μ_0. Note also that DID equal weights would set the first factor loading to 1, which is equal to μ_0^1, but the second factor loading would be equal to $0.75\neq\mu_0^2$. We want to show that the SC weights would improve the construction of the second factor loading but it will distort the combination for the first factor loading. If we set $\sigma_\epsilon^2=\mathbb{E}[(\lambda_t^1)^2]=\mathbb{E}[(\lambda_t^2)^2]=1$, then the factor loadings of the SC unit would be given by $(1.038, 0.8458)$. Therefore, there is small loss in the construction of the first factor loading and a gain in the construction of the second factor loading. Therefore, if selection into treatment is correlated with the common shock λ_t^1, then the SC estimator would be more asymptotically biased than the DID estimator.

A.5 Alternatives specifications and alternative estimators

A.5.1 Average of pre-intervention outcome as economic predictor

We consider now another very common specification in SC applications, which is to use the average pre-treatment outcome as the economic predictor. Note that if one uses only the average pre-treatment outcome as the economic predictor then the choice of matrix V would be irrelevant. In this case, the minimization problem would be given by:

$$\begin{aligned}\{\hat{w}_j\}_{j\neq 0} &= \text{argmin}_{w\in\Delta^{J-1}}\left[\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(y_{0t}-\sum_{j\neq 0}w_j y_{jt}\right)\right]^2\\ &= \text{argmin}_{w\in\Delta^{J-1}}\left[\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\epsilon_{0t}-\sum_{j\neq 0}w_j\epsilon_{jt}+\lambda_t\left(\mu_0-\sum_{j\neq 0}w_j\mu_j\right)+c_0-\sum_{j\neq 0}w_j c_j\right)\right]^2 \end{aligned} \tag{48}$$

Therefore, under Assumptions 2, 3 and 4, the objective function converges in probability to:

$$\Gamma(\mathbf{w}) = \left(c_0 - \sum_{j \neq 0} w_j c_j \right)^2 \tag{49}$$

Therefore, if there are weights that reconstruct the unit fixed effects without reconstructing the other factor loadings of the treated unit, then there is no guarantee that the SC control method will choose weights that are close to the correct ones. This result is consistent with the MC simulations by Ferman et al. (2020), who show that this specification performs particularly bad in allocating the weights correctly.

A.5.2 Adding other covariates as predictors

Most SC applications that use the average pre-intervention outcome value as economic predictor also consider other time invariant covariates as economic predictors. Let Z_i be a $(R \times 1)$ vector of observed covariates (not affected by the intervention). Assumption 1 changes to:

$$\begin{cases} y_{it}^N = \delta_t + c_i + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it} \\ y_{it}^I = \alpha_{it} + y_{it}^N \end{cases} \tag{50}$$

We redefine the set $\Phi = \{\mathbf{w} \in \Delta^{J-1} \mid c_0 = \sum_{j \neq 0} c_j w_j, \mu_0 = \sum_{j \neq 0} w_j \mu_j, \, Z_0 = \sum_{j \neq 0} w_j Z_j\}$. Let X_1 be an $((R+1) \times 1)$ vector that contains the average pre-intervention outcome and all covariates for unit 1, while X_0 is a $((R+1) \times J)$ matrix that contains the same information for the control units. For a given V, the first step of the nested optimization problem suggested in Abadie et al. (2010) would be given by:

$$\widehat{\mathbf{w}}(V) \in \text{argmin}_{\mathbf{w} \in \Delta^{J-1}} ||X_1 - X_0 \mathbf{w}||_V. \tag{51}$$

Considering again the assumptions from Section 3.1, the objective function of this minimization problem converges to $||\bar{X}_1 - \bar{X}_0 \mathbf{w}||_V$, where:

$$\bar{X}_1 - \bar{X}_0 \mathbf{w} = \begin{bmatrix} \bar{\theta} \left(Z_0 - \sum_{j \neq 0} w_j Z_j \right) + \left(c_0 - \sum_{j \neq 0} w_j c_j \right) \\ \left(Z_0^1 - \sum_{j \neq 0} w_j Z_j^1 \right) \\ \vdots \\ \left(Z_0^R - \sum_{j \neq 0} w_j Z_j^R \right) \end{bmatrix}, \tag{52}$$

where we assume $\frac{1}{T_0} \sum_{t \in \mathcal{T}_0} \theta_t \rightarrow_p \bar{\theta}$. Therefore, there is no guarantee that an estimator based on this minimization problem would converge to weights in Φ for any given matrix V, even if $\Phi \neq \varnothing$.

The second step in the nested optimization problem is to choose V such that $\widehat{\mathbf{w}}(V)$ minimizes the pre-intervention prediction error. Note that this problem is essentially given by:

$$\widehat{\mathbf{w}} = \text{argmin}_{w\in\widetilde{W}} \left[\frac{1}{T_0} \sum_{t\in\mathcal{T}_0} \left(y_{0t} - \sum_{j\neq 0} w_j y_{jt} \right) \right]^2 \tag{53}$$

where $\widetilde{W} \subseteq \Delta^{J-1}$ is the set of $\mathbf{w}$ such that $\mathbf{w}$ is the solution to problem 51 for some positive semidefinite matrix V. Similarly to the SC estimator that includes all pre-treatment outcomes, there is no guarantee that this minimization problem will choose weights in Φ, even when $T_0 \to \infty$. Therefore, it is not possible to guarantee that this SC estimator would be asymptotically unbiased. MC simulation presented by Ferman et al. (2020) confirm that this SC specification systematically misallocates more weight than alternatives that use a large number of pre-treatment outcome lags as predictors.

A.5.3 Relaxing constraints on the weights & other estimators

Our main result that the original and the demeaned SC estimators are generally asymptotically biased if there are unobserved time-varying confounders (Propositions 1 and 2) still applies if we also relax the non-negative and the adding-up constraints, which essentially leads to the panel data approach suggested by Hsiao et al. (2012), and further explored by Li and Bell (2017).[33] Our conditions for unbiasedness of the SC estimator also apply to the estimators proposed by Carvalho et al. (2018) and de Carvalho et al. (2016) when J is fixed.

These papers rely on assumptions that essentially imply no selection on unobservables to derive consistency results, which reconciles our results with theirs. Hsiao et al. (2012) and Li and Bell (2017) implicitly rely on stability in the linear projection of the potential outcomes of the treated unit on the outcomes of the control units, before and after the intervention, to show that their proposed estimators are unbiasedness and consistent. See, for example, equation A.4 from Li and Bell (2017). For simplicity, consider that $\lambda_t \mu_i$ includes the fixed effects c_i and δ_t. Then the linear projection of y_{0t}^N given $\mathbf{y}_t$ for any given t is given by $\delta_1(t) + \mathbf{y}_t'\delta(t)$, where

$$\begin{cases} \delta(t) = \left[\boldsymbol{\mu} var(\lambda_t)\boldsymbol{\mu}'\right]^{-1} \boldsymbol{\mu} var(\lambda_t)\mu_0, \text{ and} \\ \delta_1(t) = \mathbb{E}[\lambda_t](\mu_0 - \boldsymbol{\mu}'\delta(t)). \end{cases} \tag{54}$$

Therefore, in general, we will only have $(\delta_1(t), \delta(t))$ constant for all t if the distribution of λ_t is stable over time. However, the idea that treatment assignment is correlated with the factor model structure essentially means that the distribution of λ_t is different before and after the treatment assignment. In this case, it would not be reasonable to assume that the parameters of the linear projection of y_{0t}^N given $\mathbf{y}_t$ are the same for $t \in \mathcal{T}_0$ and $t \in \mathcal{T}_1$ if we consider that treatment assignment is correlated with the factor model structure. Chernozhukov et al. (2018) assume that y_{0t}^N and $\mathbf{y}_t$

[33] In this case, since we do not constraint the weights to sum 1, we need to adjust Assumption 4 so that it also includes convergence of the pre-treatment averages of the first and second moments of δ_t.

are covariance-stationary for all periods (see their Assumption 6), which implies that $(\delta_1(t), \delta(t))$ constant for all t. Therefore, they also implicitly imply that there is no selection on unobservables. Since they consider a setting with both large J and T, however, it is possible that their estimator is consistent when there is selection on unobservables under conditions similar to the ones considered by Ferman (2019).

Carvalho et al. (2018), de Carvalho et al. (2016), Masini and Medeiros (2019), and Zhou and Geng (2019) assume that the outcome of the control units are independent from treatment assignment. If we consider the linear factor model structure from Assumption 1, then this essentially means that there is no selection on unobservables. Given Assumption 3, if treatment assignment is correlated with the potential outcomes of the treated unit, then it must be correlated with $\lambda_t \mu_0$. However, if this is the case, then treatment assignment must also be correlated with at least some control units, implying that their assumption that the outcome of the control units are independent from treatment assignment would be violated. Note that Carvalho et al. (2018), Masini and Medeiros (2019), and Zhou and Geng (2019) encompass a setting with both large J and T. Therefore, it might be possible to consider a different set of assumptions, as the ones considered by Ferman (2019), so that their estimator is asymptotically unbiased when J also increases.

Overall, our results clarify what selection on unobservables means in this setting, and the conditions under which these estimators are asymptotically unbiased when J is fixed. These results also clarify that there is no contradiction between these papers and the literature on factor models, which shows that factor loadings can only be consistently estimated with fixed J under strong assumptions on the idiosyncratic shocks.

B Appendix Tables and Figures

Table A.1: **MC Results - Specification Test**

	No break				Break in λ_{1t}			
	$T_0 = 120$	$T_0 = 240$	$T_0 = 480$	$T_0 = 1200$	$T_0 = 120$	$T_0 = 240$	$T_0 = 480$	$T_0 = 1200$
μ_{10}	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
-2.6689	0.172	0.110	0.079	0.059	-2.669	0.619	0.510	0.430
-2.4079	0.179	0.117	0.084	0.063	-2.408	0.691	0.582	0.503
-1.5034	0.183	0.121	0.090	0.063	-1.503	0.670	0.562	0.477
-1.4303	0.170	0.115	0.091	0.064	-1.430	0.594	0.477	0.391
-1.1359	0.167	0.112	0.090	0.063	-1.136	0.560	0.429	0.350
-1.0772	0.168	0.117	0.091	0.068	-1.077	0.666	0.553	0.484
-1.0604	0.173	0.111	0.090	0.069	-1.060	0.626	0.526	0.448
-1.0173	0.165	0.111	0.085	0.061	-1.017	0.598	0.507	0.430
-1.0066	0.167	0.117	0.088	0.059	-1.007	0.576	0.481	0.385
-0.8201	0.150	0.114	0.080	0.065	-0.820	0.616	0.563	0.506
-0.8087	0.151	0.110	0.080	0.061	-0.809	0.577	0.489	0.431
-0.6899	0.170	0.125	0.095	0.064	-0.690	0.412	0.313	0.218
-0.6813	0.145	0.105	0.081	0.068	-0.681	0.534	0.476	0.447
-0.6594	0.158	0.116	0.098	0.061	-0.659	0.459	0.362	0.315
-0.6573	0.152	0.120	0.097	0.060	-0.657	0.479	0.382	0.298
-0.5299	0.155	0.109	0.085	0.063	-0.530	0.374	0.287	0.229
-0.4925	0.138	0.098	0.074	0.059	-0.493	0.412	0.345	0.326
-0.3721	0.156	0.113	0.092	0.063	-0.372	0.324	0.244	0.187
-0.3253	0.158	0.128	0.103	0.065	-0.325	0.291	0.223	0.163
-0.2952	0.126	0.101	0.088	0.060	-0.295	0.321	0.265	0.230
-0.1566	0.138	0.080	0.070	0.049	-0.157	0.270	0.183	0.144
-0.1291	0.136	0.116	0.086	0.060	-0.129	0.214	0.167	0.120
-0.1251	0.138	0.115	0.107	0.066	-0.125	0.233	0.178	0.141
-0.1190	0.153	0.121	0.097	0.062	-0.119	0.271	0.192	0.133
-0.1147	0.136	0.100	0.074	0.062	-0.115	0.243	0.170	0.121
-0.0297	0.145	0.120	0.103	0.066	-0.030	0.225	0.163	0.119
-0.0155	0.131	0.100	0.073	0.057	-0.015	0.202	0.139	0.098
0.1411	0.129	0.112	0.089	0.063	0.141	0.258	0.184	0.130
0.1616	0.126	0.105	0.087	0.059	0.162	0.261	0.202	0.160
0.1895	0.150	0.116	0.093	0.063	0.190	0.247	0.178	0.133
0.2039	0.152	0.125	0.104	0.066	0.204	0.233	0.169	0.127
0.2043	0.145	0.115	0.086	0.059	0.204	0.248	0.181	0.113
0.3557	0.135	0.115	0.100	0.064	0.356	0.408	0.359	0.288
0.3874	0.152	0.106	0.076	0.058	0.387	0.350	0.274	0.201
0.5107	0.152	0.102	0.081	0.057	0.511	0.383	0.297	0.248
0.6244	0.157	0.112	0.093	0.058	0.624	0.512	0.419	0.337
0.6743	0.153	0.120	0.096	0.057	0.674	0.536	0.439	0.345
0.6887	0.155	0.102	0.083	0.056	0.689	0.466	0.355	0.307
0.7582	0.148	0.105	0.080	0.067	0.758	0.504	0.421	0.381
0.7728	0.161	0.110	0.093	0.058	0.773	0.461	0.356	0.284
0.9193	0.160	0.108	0.082	0.067	0.919	0.593	0.486	0.429
0.9395	0.157	0.111	0.086	0.061	0.939	0.650	0.583	0.522
0.9810	0.182	0.111	0.080	0.061	0.981	0.621	0.514	0.451
1.1221	0.159	0.112	0.093	0.068	1.122	0.594	0.497	0.421
1.2940	0.173	0.117	0.092	0.056	1.294	0.629	0.527	0.450
1.3090	0.186	0.126	0.083	0.064	1.309	0.687	0.578	0.506
1.3762	0.187	0.128	0.095	0.063	1.376	0.719	0.609	0.519
1.3897	0.176	0.108	0.086	0.068	1.390	0.659	0.546	0.467
1.5060	0.168	0.119	0.084	0.068	1.506	0.601	0.494	0.413
1.6281	0.178	0.120	0.087	0.060	1.628	0.692	0.586	0.498
2.1912	0.189	0.119	0.086	0.065	2.191	0.712	0.598	0.513

Notes: this table presents rejection rates for the specification test presented in Section 3.2. In columns 1 to 4, there is no structural break, while in columns 5 to 8 the first common factor has expected value equal to two times its standard deviation in the post-treatment periods.

Table A.2: Estimated weights - Empirical Illustration

	Original SC	Demeaned SC	Abadie et al. (2003)
Andalucia	0.0000	0.0000	0.0000
Aragon	0.0000	0.0000	0.0000
Baleares (Islas)	0.3111	0.2539	0.0000
Canarias	0.0000	0.0000	0.0000
Cantabria	0.0000	0.0008	0.0000
Castilla Y Leon	0.0000	0.0002	0.0000
Castilla-La Mancha	0.0000	0.0000	0.0000
Cataluna	0.0000	0.0536	0.8508
Comunidad Valenciana	0.0000	0.0003	0.0000
Extremadura	0.0000	0.0000	0.0000
Galicia	0.0000	0.0000	0.0000
Madrid (Comunidad De)	0.4831	0.2879	0.1492
Murcia (Region de)	0.0000	0.1898	0.0000
Navarra	0.0000	0.0190	0.0000
Principado De Asturias	0.0000	0.0072	0.0000
Rioja (La)	0.2058	0.1873	0.0000