SUPPLEMENTARY MATERIAL TO "ON THE PROPERTIES OF THE SYNTHETIC CONTROL ESTIMATOR WITH MANY PERIODS AND MANY CONTROLS"

Bruno Ferman
Sao Paulo School of Economics - FGV

A Appendix

A.1 Proof of main results

For a generic $m \times n$ matrix $\mathbf{A}$, define $\|\mathbf{A}\|_2 = \sqrt{\sum_{p=1}^{m}\sum_{q=1}^{n}|a_{pq}|^2}$, $\|\mathbf{A}\|_\infty = \max_{p\in\{1,...,m\},q\in\{1,...,n\}}\{|a_{pq}|\}$, and $\|\mathbf{A}\|_1 = \sum_{p=1}^{m}\sum_{q=1}^{n}|a_{pq}|$.

A.1.1 Proof of Proposition 3.1

Proposition 3.1 *Let $\widehat{\boldsymbol{\mu}}_{SC}$ be defined as $\mathbf{M}_J{}'\widehat{\mathbf{w}}_{SC}$, where $\widehat{\mathbf{w}}_{SC}$ is defined in equation* (5)*, and suppose Assumptions 2.1, 2.2, 3.1, 3.2, 3.3, and 3.4(a) hold. Then, as $T_0, J \rightarrow \infty$, (i) for all $t \in \mathcal{T}_1$, $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{SC} \xrightarrow{p} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$, and (ii) $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(y_{0t} - \widehat{\mathbf{w}}'_{SC}\mathbf{y}_t)^2 \xrightarrow{p} \sigma_0^2$. Moreover, if we add Assumption 3.4(b), then (iii) $\|\widehat{\mathbf{w}}_{SC}\|_2 \xrightarrow{p} 0$.*

Proof.

We start by extending the function $\mathcal{H}_J(\boldsymbol{\mu})$ to the domain $\mathcal{M} = \mathrm{cl}(\cup_{J\in\mathbb{N}}\mathcal{M}_J)$, where $\mathrm{cl}(\mathcal{A})$ is the closure of set $\mathcal{A}$. For $\boldsymbol{\mu} \in \mathcal{M}$, we define

$$\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}) = \min_{\widetilde{\boldsymbol{\mu}}\in\mathcal{M}_J}\left\{\mathcal{H}_J(\widetilde{\boldsymbol{\mu}}) + K_J\left\|\boldsymbol{\mu} - \widetilde{\boldsymbol{\mu}}\right\|_2\right\}, \tag{1}$$

where we define later in the proof what the random variable K_J is. For now, it suffices to consider that $K_J > 0$ almost surely and that $K_J = O_p(1)$. Since for any $\boldsymbol{\mu} \in \mathcal{M}\setminus\mathcal{M}_J$ there is an $\boldsymbol{\mu}' \in \mathcal{M}_J$ such that $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}') < \widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$, it follows that $\underset{\boldsymbol{\mu}\in\mathcal{M}_J}{\mathrm{argmin}}\mathcal{H}_J(\boldsymbol{\mu}) = \underset{\boldsymbol{\mu}\in\mathcal{M}}{\mathrm{argmin}}\widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$. Therefore, we can analyze the behavior of the implied estimator for the factor loadings of the treated unit considering the objective function $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$. Let $\widehat{\boldsymbol{\mu}}_{\text{SC}}$ be a minimizer of $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$ subject to $\boldsymbol{\mu} \in \mathcal{M}$. Recall that considering $\widehat{\boldsymbol{\mu}}_{\text{SC}}$ is just an intermediate step for analyzing the asymptotic behavior of $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{SC}}$. Since we are ultimately interested in this product, our results are invariant to different normalizations in the factor structure. Finally, note that there may be multiple solutions to $\min\limits_{\boldsymbol{\mu} \in \mathcal{M}_J} \widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$. In this case, we consider $\widehat{\boldsymbol{\mu}}_{\text{SC}}$ as being one of the solutions.

Define the function $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}) \equiv (\boldsymbol{\mu}_0 - \boldsymbol{\mu})'\boldsymbol{\Omega}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}) + \sigma^2_0$, where $\boldsymbol{\Omega}$ is positive definite by Assumption 3.3. Therefore, $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$ is uniquely minimized at $\boldsymbol{\mu}_0$. We first show that $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}_0)$ is bounded from above by a term that converges in probability to $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0)$. Consider the $\mathbf{w}^*_J$ defined in Assumption 3.2, and let $\boldsymbol{\mu}^*_J \equiv \mathbf{M}_J{}'\mathbf{w}^*_J$. Then, since $\boldsymbol{\mu}^*_J \in \mathcal{M}_J$, and since $\mathbf{w}^*_J$ is a candidate solution for the minimization problem defined in $\mathcal{H}_J(\boldsymbol{\mu}^*_J)$, it follows that

$$\begin{aligned}
\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}_0) &\leq \mathcal{H}_J(\boldsymbol{\mu}^*_J) + K_J \left\|\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J\right\|_2 \leq \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\bar{\lambda}_t(\boldsymbol{\mu}^*_J) - (\boldsymbol{\epsilon}_t'\mathbf{w}^*_J)\right)^2 + K_J\left\|\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J\right\|_2 \quad (2)\\
&= \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\bar{\lambda}_t(\boldsymbol{\mu}^*_J)^2 + \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\epsilon}_t'\mathbf{w}^*_J)^2 - 2\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\bar{\lambda}_t(\boldsymbol{\mu}^*_J)(\boldsymbol{\epsilon}_t'\mathbf{w}^*_J) + K_J\left\|\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J\right\|_2\\
&\equiv \widetilde{\mathcal{H}}^{UB}_J(\boldsymbol{\mu}_0).
\end{aligned}$$

The first term of $\widetilde{\mathcal{H}}^{UB}_J(\boldsymbol{\mu}_0)$ equals to

$$\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\bar{\lambda}_t(\boldsymbol{\mu}^*_J)^2 = (\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J)'\left(\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t\right)(\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J) + 2(\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J)'\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\epsilon_{0t} + \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon^2_{0t}, \quad (3)$$

where $(\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J)'\left(\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t\right)(\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J) = o_p(1)$ because $(\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J) = o(1)$ and $\left(\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t\right) = O(1)$, $\left|(\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J)'\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\epsilon_{0t}\right| \leq \left\|\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J\right\|_1\left\|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\epsilon_{0t}\right\|_\infty$, where $\left\|\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J\right\|_1 = o(1)$ from Assumption 3.2 and from the fact that the dimension of $\boldsymbol{\mu}$ is fixed, and $\left\|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\epsilon_{0t}\right\|_\infty = o_p(1)$ from Assumption 3.4(a), and $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon^2_{0t} \xrightarrow{p} \sigma^2_0$. Therefore, $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\bar{\lambda}_t(\boldsymbol{\mu}^*_J)^2 \xrightarrow{p} \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0)$.

For the term $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\epsilon}_t'\mathbf{w}^*_J)^2$, note that $var^*(\boldsymbol{\epsilon}_t'\mathbf{w}^*_J) \leq \bar{\gamma}\left\|\mathbf{w}^*_J\right\|^2_2$, where $\bar{\gamma} = \sup_{j,t}\{var^*(\epsilon_{jt})\} < \infty$ since ϵ_{jt} has uniformly bounded fourth moments. Therefore,

$$\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\epsilon}_t'\mathbf{w}^*_J)^2 \leq \bar{\gamma}\left\|\mathbf{w}^*_J\right\|^2_2\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\frac{\boldsymbol{\epsilon}_t'\mathbf{w}^*_J}{\sqrt{var^*(\boldsymbol{\epsilon}_t'\mathbf{w}^*_J)}}\right)^2. \quad (4)$$

If we define $z_t \equiv \left(\frac{\boldsymbol{\epsilon}_t'\mathbf{w}_J^*}{\sqrt{var^*(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)}}\right)^2 - 1$, then $\mathbb{E}^*[z_t] = 0$. Moreover,

$$var^*(z_t) = \mathbb{E}^*\left[\left(\frac{\boldsymbol{\epsilon}_t'\mathbf{w}_J^*}{\sqrt{var^*(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)}}\right)^4\right] - 2\mathbb{E}^*\left[\left(\frac{\boldsymbol{\epsilon}_t'\mathbf{w}_J^*}{\sqrt{var^*(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)}}\right)^2\right] + 1 = \frac{\mathbb{E}^*[(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)^4]}{(var^*(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*))^2} - 1.$$

Now note that $\mathbb{E}^*[(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)^4] = \sum_{p,q,r,s}\mathbb{E}^*[\epsilon_{pt}\epsilon_{qt}\epsilon_{rt}\epsilon_{st}]w_p^*w_q^*w_r^*w_s^* = \sum_{p,q}\mathbb{E}^*[\epsilon_{pt}^2\epsilon_{qt}^2](w_p^*)^2(w_q^*)^2$, where the last equality follows from ϵ_{it} independent across i, and $\mathbb{E}^*[\epsilon_{it}] = 0$. Now given that ϵ_{it} has uniformly bounded fourth moments across i and t, we can define $\bar{\xi} = \sup_{i,t}\{\mathbb{E}^*[\epsilon_{it}^4]\} < \infty$. It follows that $\mathbb{E}^*[(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)^4] \leq \max\left\{\bar{\xi}, \bar{\gamma}^2\right\}\sum_{p,q}(w_p^*)^2(w_q^*)^2 = \max\left\{\bar{\xi}, \bar{\gamma}^2\right\}\|\mathbf{w}_J^*\|_2^4$. For the denominator, if we define $\underline{\gamma} = \inf_{j,t}\{var^*(\epsilon_{jt})\} > 0$, then $(var^*(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*))^2 \geq \underline{\gamma}^2\|\mathbf{w}_J^*\|_2^4$. Combining these two results, we have that $var(z_t)$ is uniformly bounded. It follows from Andrews (1988) that $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0} z_t \xrightarrow{p} 0$, which implies that $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\frac{\boldsymbol{\epsilon}_t'\mathbf{w}_J^*}{\sqrt{var^*(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)}}\right)^2 \xrightarrow{p} 1$. Since $\|\mathbf{w}_J^*\|_2^2 \to 0$ by Assumption 3.2, it follows that $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)^2$ is $o_p(1)$.

The term $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\bar{\lambda}_t(\boldsymbol{\mu}_J^*)(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)$ is given by $(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)'\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*) + \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{0t}(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)$. Note that

$$\left\|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)\right\|_1 = \sum_{f=1}^F\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\lambda_{ft}(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)\right| \leq \sum_{f=1}^F\left\|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\lambda_{ft}\boldsymbol{\epsilon}_t\right\|_\infty\|\mathbf{w}_J^*\|_1 \leq \sum_{f=1}^F\left\|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\lambda_{ft}\boldsymbol{\epsilon}_t\right\|_\infty, \quad (5)$$

where the first inequality follows from Hölder's inequality, and the second one follows from $\mathbf{w}_J^* \in \Delta^{J-1}$. From Assumption 3.4(a), we have $\left\|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\lambda_{ft}\boldsymbol{\epsilon}_t\right\|_\infty = o_p(1)$. Since $\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*$ is bounded, it follows that $(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)'\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*) = o_p(1)$. Also, note that $\epsilon_{0t}(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)$ has zero mean and uniformly bounded variance, given that ϵ_{0t} and $(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)$ have mean zero and uniformly bounded variance, and they are independent. Since we also have that $\epsilon_{0t}(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)$ is α-mixing, it follows from Andrews (1988) that $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{0t}(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*) = o_p(1)$. Finally, the fourth term of $\widetilde{\mathcal{H}}_J^{UB}(\boldsymbol{\mu}_0)$ is $o_p(1)$ because $K = O_p(1)$ and $\|\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*\|_2 = o(1)$. Therefore, $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}_0) \leq \widetilde{\mathcal{H}}_J^{UB}(\boldsymbol{\mu}_0) \xrightarrow{p} \sigma_{\bar{\lambda}}^2(\boldsymbol{\mu}_0)$.

We show next that $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$ is bounded from below by a function that converges uniformly in probability to $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$. We have that

$$\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}) \geq \min_{\widetilde{\boldsymbol{\mu}} \in \mathcal{M}} \left\{ \min_{\mathbf{b} \in \mathcal{W}} \left\{ \frac{1}{T_0} \sum_{t \in \mathcal{T}_0} (\bar{\lambda}_t(\widetilde{\boldsymbol{\mu}}) - \mathbf{b}'\boldsymbol{\epsilon}_t)^2 \right\} + K_J \left\| \boldsymbol{\mu} - \widetilde{\boldsymbol{\mu}} \right\|_2 \right\} \equiv \widetilde{\mathcal{H}}_J^{LB}(\boldsymbol{\mu}), \tag{6}$$

where $\mathcal{W} = \{\mathbf{w} \in \mathbb{R}^J | \left\| \mathbf{w} \right\|_1 \leq 1\}$. This inequality holds because we are relaxing three constraints in the minimization problems from $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$. We consider $\widetilde{\boldsymbol{\mu}} \in \mathcal{M} \supset \mathcal{M}_J$, we relax the condition $\mathbf{M}_J{}'\mathbf{w} = \widetilde{\boldsymbol{\mu}}$, and we consider a set $\mathcal{W} \supset \Delta^{J-1}$.

We first show that the function $Q_J(\boldsymbol{\mu}) \equiv \min_{\mathbf{b} \in \mathcal{W}} \left\{ \frac{1}{T_0} \sum_{t \in \mathcal{T}_0} (\bar{\lambda}_t(\boldsymbol{\mu}) - \mathbf{b}'\boldsymbol{\epsilon}_t)^2 \right\}$ is Lipschitz with a constant that is $O_p(1)$. Let $\widehat{\mathbf{b}}(\boldsymbol{\mu})$ be the solution to this minimization problem for a given $\boldsymbol{\mu}$. Since $\mathcal{M}$ is convex, from the mean value and the envelope theorems, it follows that, for any $\boldsymbol{\mu}$ and $\boldsymbol{\mu}' \in \mathcal{M}$, there is a $\widetilde{\boldsymbol{\mu}} \in \mathcal{M}$ such that

$$\begin{aligned} |Q_J(\boldsymbol{\mu}) - Q_J(\boldsymbol{\mu}')| &= \left| \frac{2}{T_0} \sum_{t \in \mathcal{T}_0} \boldsymbol{\lambda}_t'[\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0 - \widetilde{\boldsymbol{\mu}}) + \epsilon_{0t} - \widehat{\mathbf{b}}(\widetilde{\boldsymbol{\mu}})'\boldsymbol{\epsilon}_t] \cdot (\boldsymbol{\mu} - \boldsymbol{\mu}') \right| \\ &\leq \left\| \frac{2}{T_0} \sum_{t \in \mathcal{T}_0} \boldsymbol{\lambda}_t'[\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0 - \widetilde{\boldsymbol{\mu}}) + \epsilon_{0t} - \widehat{\mathbf{b}}(\widetilde{\boldsymbol{\mu}})'\boldsymbol{\epsilon}_t] \right\|_2 \times \left\| \boldsymbol{\mu} - \boldsymbol{\mu}' \right\|_2 \\ &\leq \left(\left\| \sum_{t \in \mathcal{T}_0} \frac{\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t}{T_0} \right\|_2 \left\| \boldsymbol{\mu}_0 - \widetilde{\boldsymbol{\mu}} \right\|_2 + \left\| \sum_{t \in \mathcal{T}_0} \frac{\boldsymbol{\lambda}_t'\epsilon_{0t}}{T_0} \right\|_2 + \left\| \sum_{t \in \mathcal{T}_0} \frac{\boldsymbol{\lambda}_t'\boldsymbol{\epsilon}_t'}{T_0} \widehat{\mathbf{b}}(\widetilde{\boldsymbol{\mu}}) \right\|_2 \right) \times \left\| \boldsymbol{\mu} - \boldsymbol{\mu}' \right\|_2 . \end{aligned} \tag{7}$$

Since $\left\| \boldsymbol{\mu}_0 - \widetilde{\boldsymbol{\mu}} \right\|_2 \leq C$ for some constant C (Assumption 3.2), and $\sum_{t \in \mathcal{T}_0} \frac{\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t}{T_0} \to \boldsymbol{\Omega}$ (Assumption 3.3), we have that $\left\| \sum_{t \in \mathcal{T}_0} \frac{\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t}{T_0} \right\|_2 \left\| \boldsymbol{\mu}_0 - \widetilde{\boldsymbol{\mu}} \right\|_2 \leq C \left\| \sum_{t \in \mathcal{T}_0} \frac{\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t}{T_0} \right\|_2 = O_p(1)$. From Assumption 3.4(a), $\left\| \sum_{t \in \mathcal{T}_0} \frac{\boldsymbol{\lambda}_t'\epsilon_{0t}}{T_0} \right\|_2 = o_p(1)$. Finally, $\left\| \sum_{t \in \mathcal{T}_0} \frac{\boldsymbol{\lambda}_t'\boldsymbol{\epsilon}_t'}{T_0} \widehat{\mathbf{b}}(\widetilde{\boldsymbol{\mu}}) \right\|_2 \leq \sum_{f=1}^F \left| \sum_{t \in \mathcal{T}_0} \frac{\lambda_{ft}\boldsymbol{\epsilon}_t'}{T_0} \widehat{\mathbf{b}}(\widetilde{\boldsymbol{\mu}}) \right| \leq \sum_{f=1}^F \left\| \sum_{t \in \mathcal{T}_0} \frac{\lambda_{ft}\boldsymbol{\epsilon}_t}{T_0} \right\|_\infty \left\| \widehat{\mathbf{b}}(\widetilde{\boldsymbol{\mu}}) \right\|_1 \leq \sum_{f=1}^F \left\| \sum_{t \in \mathcal{T}_0} \frac{\lambda_{ft}\boldsymbol{\epsilon}_t}{T_0} \right\|_\infty = o_p(1)$ from Assumption 3.4(a). Therefore, $|Q_J(\boldsymbol{\mu}) - Q_J(\boldsymbol{\mu}')| \leq \tilde{K}_J \left\| \boldsymbol{\mu} - \boldsymbol{\mu}' \right\|_2$, where $\tilde{K} = O_p(1)$ and does not depend on $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$. We define K used in function $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$ as $K = 1 + \tilde{K}$, so $K > 0$ and $K = O_p(1)$. Given that K is greater than the Lipschitz constant of $Q_J(\boldsymbol{\mu})$, we have that $\widetilde{\mathcal{H}}_J^{LB}(\boldsymbol{\mu}) = Q_J(\boldsymbol{\mu})$ for all $\boldsymbol{\mu} \in \mathcal{M}$. Therefore, $\widetilde{\mathcal{H}}_J^{LB}(\boldsymbol{\mu})$ is Lipschitz with a constant $O_p(1)$.

Now we show that $\widetilde{\mathcal{H}}_J^{LB}(\boldsymbol{\mu}) \xrightarrow{p} \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$ pointwise. Note that $\frac{1}{T_0} \sum_{t \in \mathcal{T}_0} (\bar{\lambda}_t(\boldsymbol{\mu}) - \mathbf{b}'\boldsymbol{\epsilon}_t)^2 = \frac{1}{T_0} \sum_{t \in \mathcal{T}_0} (\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0 - \boldsymbol{\mu}))^2 + \frac{2}{T_0} \sum_{t \in \mathcal{T}_0} (\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0 - \boldsymbol{\mu}))(\epsilon_{0t} - \mathbf{b}'\boldsymbol{\epsilon}_t) + \frac{1}{T_0} \sum_{t \in \mathcal{T}_0} (\epsilon_{0t} - \mathbf{b}'\boldsymbol{\epsilon}_t)^2$. We define $\widehat{\mathbf{c}} \in \underset{\mathbf{b}\in\mathcal{W}}{\operatorname{argmin}} \left\{ \frac{1}{T_0} \sum_{t\in\mathcal{T}_0} (\epsilon_{0t} - \mathbf{b}'\boldsymbol{\epsilon}_t)^2 \right\}$. Since $\mathcal{W}$ is compact, we can also define $\widehat{\mathbf{d}} \in \underset{\mathbf{b}\in\mathcal{W}}{\operatorname{argmin}} \left\{ \frac{2}{T_0} \sum_{t\in\mathcal{T}_0} (\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0 - \boldsymbol{\mu}))(\epsilon_{0t} - \mathbf{b}'\boldsymbol{\epsilon}_t) \right\}$. Therefore,

$$\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\boldsymbol{\mu}))^2 + \frac{2}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\boldsymbol{\mu}))(\epsilon_{0t}-\widehat{\mathbf{d}}'\boldsymbol{\epsilon}_t) + \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\epsilon_{0t}-\widehat{\mathbf{c}}'\boldsymbol{\epsilon}_t)^2 \leq Q_J(\boldsymbol{\mu}) \leq \tag{8}$$

$$\leq \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\boldsymbol{\mu}))^2 + \frac{2}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\boldsymbol{\mu}))(\epsilon_{0t}-\widehat{\mathbf{c}}'\boldsymbol{\epsilon}_t) + \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\epsilon_{0t}-\widehat{\mathbf{c}}'\boldsymbol{\epsilon}_t)^2, \tag{9}$$

where the first inequality holds because we give more flexibility in the minimization problem in $Q_J(\boldsymbol{\mu})$ by allowing different parameters to minimize the second and the third terms. The second inequality holds because $\widehat{\mathbf{c}}$ is a candidate solution to the minimization problem in $Q_J(\boldsymbol{\mu})$. Note that $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\boldsymbol{\mu}))^2 \rightarrow (\boldsymbol{\mu}_0-\boldsymbol{\mu})'\boldsymbol{\Omega}(\boldsymbol{\mu}_0-\boldsymbol{\mu})$. Moreover,

$$\begin{aligned}
\left|\frac{2}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\boldsymbol{\mu}))(\epsilon_{0t}-\mathbf{b}'\boldsymbol{\epsilon}_t)\right| &\leq \left|(\boldsymbol{\mu}_0-\boldsymbol{\mu})'\frac{2}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\epsilon_{0t}\right| + \left|(\boldsymbol{\mu}_0-\boldsymbol{\mu})'\frac{2}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\boldsymbol{\epsilon}_t'\mathbf{b}\right| \\
&\leq c_1\left\|\frac{2}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\epsilon_{0t}\right\|_\infty + c_1\left\|\frac{2}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\epsilon_t'\right\|_\infty \|\mathbf{b}\|_1 \qquad (10) \\
&\leq c_1\left\|\frac{2}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\epsilon_{0t}\right\|_\infty + c_1\left\|\frac{2}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\epsilon_t'\right\|_\infty = o_p(1), \qquad (11)
\end{aligned}$$

implying that $\frac{2}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\boldsymbol{\mu}))(\epsilon_{0t}-\mathbf{b}'\boldsymbol{\epsilon}_t) \xrightarrow{p} 0$ uniformly in $\mathbf{b}$. Therefore, we only have to show that $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\epsilon_{0t}-\widehat{\mathbf{c}}'\boldsymbol{\epsilon}_t)^2 \xrightarrow{p} \sigma_0^2$ to conclude that $Q_J(\boldsymbol{\mu}) \xrightarrow{p} \sigma^2_{\tilde{\lambda}}(\boldsymbol{\mu})$.

We essentially apply Lemma 1 from Chernozhukov et al. (2021). Since $\widehat{\mathbf{c}} \in \mathcal{W}$ is the argmin of $\left\{\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\epsilon_{0t}-\mathbf{b}'\boldsymbol{\epsilon}_t)^2\right\}$, and $0 \in \mathcal{W}$, it follows that $\|\boldsymbol{\epsilon}_0 - \mathbf{E}\widehat{\mathbf{c}}\|_2^2 \leq \|\boldsymbol{\epsilon}_0\|_2^2$, where $\mathbf{E}$ is the $T_0 \times J$ matrix with information on ϵ_{jt} for all $j = 1, ..., J$ and $t \in \mathcal{T}_0$, and $\boldsymbol{\epsilon}_0$ is the T_0 vector with information on ϵ_{0t} for all $t \in \mathcal{T}_0$. Therefore, equivalently to equation H.17 from Chernozhukov et al. (2021), we have the inequality

$$\frac{1}{T_0}\|\mathbf{E}\widehat{\mathbf{c}}\|_2^2 \leq \frac{1}{T_0}2\boldsymbol{\epsilon}_0'\mathbf{E}\widehat{\mathbf{c}} \leq \frac{1}{T_0}2\|\mathbf{E}'\boldsymbol{\epsilon}_0\|_\infty\|\widehat{\mathbf{c}}\|_1 \leq \frac{1}{T_0}2\|\mathbf{E}'\boldsymbol{\epsilon}_0\|_\infty. \tag{12}$$

From Assumption 3.4(a), $\frac{1}{T_0}2\|\mathbf{E}'\boldsymbol{\epsilon}_0\|_\infty = o_p(1)$, which implies that $\frac{1}{T_0}\|\mathbf{E}\widehat{\mathbf{c}}\|_2^2 = o_p(1)$ and

$\frac{1}{T_0}2\boldsymbol{\epsilon}_0'\mathbf{E}\widehat{\mathbf{c}} = o_p(1)$. Therefore, it follows that $\widetilde{\mathcal{H}}^{LB}_{T_0}(\boldsymbol{\mu}) \xrightarrow{p} \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$. Since $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$ is continuous and $\mathcal{M}$ is compact, then, based on Corollary 2.2 of Newey (1991), we have that $\mathcal{H}^{LB}_{T_0}(\boldsymbol{\mu})$ converges uniformly in probability to $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$.

Combining the results above that (i) $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}_0) \leq \widetilde{\mathcal{H}}^{UB}_J(\boldsymbol{\mu}_0) \xrightarrow{p} \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0)$, and (ii) $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}) \geq \widetilde{\mathcal{H}}^{LB}_J(\boldsymbol{\mu}) \xrightarrow{p} \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$ uniformly in $\boldsymbol{\mu} \in \mathcal{M}$, we show that $\widehat{\boldsymbol{\mu}}_{\text{SC}} \xrightarrow{p} \boldsymbol{\mu}_0$. This is a simple extension of Theorem 2.1 from Newey and McFadden (1994). For a given $\eta > 0$, since $\widehat{\boldsymbol{\mu}}_{\text{SC}} = \underset{\boldsymbol{\mu}\in\mathcal{M}}{\text{argmin}}\widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$, $\widetilde{\mathcal{H}}_J(\widehat{\boldsymbol{\mu}}_{\text{SC}}) < \widetilde{\mathcal{H}}_J(\boldsymbol{\mu}_0) + \frac{\eta}{3}$. From (ii), we have that $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}) < \widetilde{\mathcal{H}}^{LB}_J(\boldsymbol{\mu}) + \frac{\eta}{3} \leq \widetilde{\mathcal{H}}_J(\boldsymbol{\mu}) + \frac{\eta}{3}$ for all $\boldsymbol{\mu}$ with probability approaching one (wpa1). Combining (i) and (ii), we have that $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}_0) \xrightarrow{p} \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0)$, which implies that $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}_0) < \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0) + \frac{\eta}{3}$ wpa1. Combining these three inequalities, we have that $\sigma^2_{\bar{\lambda}}(\widehat{\boldsymbol{\mu}}_{\text{SC}}) < \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0) + \eta$ wpa1. Now let $\mathcal{V}$ be any open subset of $\mathcal{M}$ containing $\boldsymbol{\mu}_0$. Since $\mathcal{M} \cap \mathcal{V}^C$ is compact, $\boldsymbol{\mu}_0 = \underset{\boldsymbol{\mu}\in\mathcal{M}}{\text{argmin}}\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$, and $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$ is continuous, then $\inf_{\boldsymbol{\mu}\in\mathcal{M}\cap\mathcal{V}^C}\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}) = \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}^*) > \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0)$ for some $\boldsymbol{\mu}^* \in \mathcal{M} \cap \mathcal{V}^C$. Let $\eta = \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}^*) - \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0)$. Then, wpa1, $\sigma^2_{\bar{\lambda}}(\widehat{\boldsymbol{\mu}}_{\text{SC}}) < \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}^*)$, which implies that $\widehat{\boldsymbol{\mu}}_{\text{SC}} \in \mathcal{V}$. Therefore, $\widehat{\boldsymbol{\mu}}_{\text{SC}} \xrightarrow{p} \boldsymbol{\mu}_0$. Since we are considering a fixed sequence of $\boldsymbol{\lambda}_t$, it immediately follows that $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{SC}} \xrightarrow{p} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ for all $t \in \mathcal{T}_1$.

Now we prove the second result from Proposition 3.1, that $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(y_{0t} - \widehat{\mathbf{w}}'_{\text{SC}}\mathbf{y}_t\right)^2 \xrightarrow{p} \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0) = \sigma^2_0$. Since $\boldsymbol{\mu}^*_J \in \mathcal{M}_J$, and since $\mathbf{w}^*_J$ is a candidate solution for the minimization problem defined in $\mathcal{H}_J(\boldsymbol{\mu}^*_J)$, it follows that

$$\widetilde{\mathcal{H}}_J(\widehat{\boldsymbol{\mu}}_{\text{SC}}) \;\leq\; \mathcal{H}_J(\boldsymbol{\mu}^*_J) + K_J\left\|\widehat{\boldsymbol{\mu}}_{\text{SC}} - \boldsymbol{\mu}^*_J\right\|_2 \leq \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\bar{\lambda}_t(\boldsymbol{\mu}^*_J) - (\boldsymbol{\epsilon}'_t\mathbf{w}^*_J)\right)^2 + K_J\left\|\widehat{\boldsymbol{\mu}}_{\text{SC}} - \boldsymbol{\mu}^*_J\right\|_2.$$

Following the same arguments as above, $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\bar{\lambda}_t(\boldsymbol{\mu}^*_J) - (\boldsymbol{\epsilon}'_t\mathbf{w}^*_J)\right)^2 \xrightarrow{p} \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0)$. Moreover, since $\widehat{\boldsymbol{\mu}}_{\text{SC}} \xrightarrow{p} \boldsymbol{\mu}_0$ and $\boldsymbol{\mu}^*_J \to \boldsymbol{\mu}_0$, it follows that $K_J\left\|\widehat{\boldsymbol{\mu}}_{\text{SC}} - \boldsymbol{\mu}^*_J\right\|_2 \xrightarrow{p} 0$. Therefore, $\widetilde{\mathcal{H}}_J(\widehat{\boldsymbol{\mu}}_{\text{SC}})$ is bounded from above by a term that converges in probability to $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0)$. Since we also have that $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$ is bounded from below by a function that converges uniformly in $\boldsymbol{\mu} \in \mathcal{M}$ to the continuous function $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$, it follows that $\widetilde{\mathcal{H}}_J(\widehat{\boldsymbol{\mu}}_{\text{SC}}) = \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(y_{0t} - \widehat{\mathbf{w}}'_{\text{SC}}\mathbf{y}_t\right)^2 \xrightarrow{p} \sigma^2_{\bar{\lambda}}(\boldsymbol{\mu}_0) = \sigma^2_0$.

Finally, we consider the third result, that $\left\|\widehat{\mathbf{w}}_{\text{SC}}\right\|_2 \xrightarrow{p} 0$. We first show that $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\boldsymbol{\epsilon}'_t\widehat{\mathbf{w}}_{\text{SC}}\right)^2 \xrightarrow{p} 0$. Let $\boldsymbol{\Lambda}$ be the $(T_0\times F)$ matrix with rows $\boldsymbol{\lambda}_t$. Since $\widehat{\mathbf{w}}_{\text{SC}}$ is one of the solutions that minimize equation 6, it follows that $\left\|\boldsymbol{\epsilon}_0 + \boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}_{\text{SC}}) - \mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}\right\|_2^2 \;\leq\; \left\|\boldsymbol{\epsilon}_0 + \boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}^*_J) - \mathbf{E}\mathbf{w}^*_J\right\|_2^2$, which implies

$$
\begin{aligned}
\|\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}\|_2^2 \leq\ & \|\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}\|_2^2 + \|\boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}_{\text{SC}})\|_2^2 \leq 2\,|\boldsymbol{\epsilon}_0'\boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}_{\text{SC}})| + 2\,|\boldsymbol{\epsilon}_0'\boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)| \qquad (13)\\
& +2\,|(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}_{\text{SC}})'\boldsymbol{\Lambda}'\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}| + 2\,|(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)'\boldsymbol{\Lambda}'\mathbf{E}\mathbf{w}_J^*| + 2\,|\boldsymbol{\epsilon}_0'\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}| + \qquad (14)\\
& +2\,|\boldsymbol{\epsilon}_0'\mathbf{E}\mathbf{w}_J^*| + \|\mathbf{E}\mathbf{w}_J^*\|_2^2 + \|\boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)\|_2^2. \qquad (15)
\end{aligned}
$$

We show that all the terms on the right hand side of the above equation are $o_p(1)$ when divided by T_0. For any $\boldsymbol{\mu}$, $2\,|\boldsymbol{\epsilon}_0'\boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \boldsymbol{\mu})| \leq \|\boldsymbol{\epsilon}_0'\boldsymbol{\Lambda}\|_\infty \|\boldsymbol{\mu}_0 - \boldsymbol{\mu}\|_1 \leq c\,\|\boldsymbol{\epsilon}_0'\boldsymbol{\lambda}\|_\infty$, where $\frac{1}{T_0}\|\boldsymbol{\epsilon}_0'\boldsymbol{\lambda}\|_\infty = o_p(1)$ from Assumption 3.4(a). We also have that, for any $\boldsymbol{\mu}$ and $\mathbf{w}$, $|(\boldsymbol{\mu}_0 - \boldsymbol{\mu})'\boldsymbol{\Lambda}'\mathbf{E}\mathbf{w}| \leq \|\boldsymbol{\mu}_0 - \boldsymbol{\mu}\|_2 \|\boldsymbol{\Lambda}'\mathbf{E}\mathbf{w}\|_2 \leq c\,\|\boldsymbol{\Lambda}'\mathbf{E}\mathbf{w}\|_1 \leq c\,\|\boldsymbol{\Lambda}'\mathbf{E}\|_\infty \|\mathbf{w}\|_1 \leq c\,\|\boldsymbol{\Lambda}'\mathbf{E}\|_\infty$, where $\frac{1}{T_0}\|\boldsymbol{\Lambda}'\mathbf{E}\|_\infty = o_p(1)$ from Assumption 3.4(a). Likewise, $\frac{1}{T_0}|\boldsymbol{\epsilon}_0'\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}| = o_p(1)$ because $\frac{1}{T_0}\|\boldsymbol{\epsilon}_0'\mathbf{E}\|_\infty = o_p(1)$ from Assumption 3.4(a). Moreover, from equation 4, $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)^2 = o_p(1)$. Finally,

$$
\frac{1}{T_0}\|\boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)\|_2^2 = (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)'\left(\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t\right)(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*) = o(1), \qquad (16)
$$

since the $\|(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)\|_2 = o(1)$ and $\left\|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t\right\|_2 = O(1)$.

Combining all these results, we have $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}}\right)^2 \xrightarrow{p} 0$. Now suppose $\|\widehat{\mathbf{w}}_{\text{SC}}\|_2$ does not converge in probability to zero. Since $\mathbf{w} \in \Delta^{J-1}$, this would imply that there is a constant b such that $P\left(\max_i\{\widehat{w}_i^2\} > b\right)$ does not converge to zero. Note that this would not be true if we considered $\mathbf{w} \in \mathbb{R}^J$. However, given the restriction $\mathbf{w} \in \Delta^{J-1}$, we have that $\|\mathbf{w}\|_2^2 \leq \max_i\{w_i\}\sum_i w_i \leq \max_i\{w_i\}$. In this case, for infinitely many T_0, we have with probability greater than some $\xi > 0$,

$$
\begin{aligned}
\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}}\right)^2 &= \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\sum_{i=1}^{J}\epsilon_{it}^2\widehat{w}_i^2 + \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\sum_{i\neq j}\epsilon_{it}\epsilon_{jt}\widehat{w}_i\widehat{w}_j \geq \qquad (17)\\
&\geq b\min_{1\leq i\leq J}\left\{\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{it}^2\right\} + \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\sum_{i\neq j}\epsilon_{it}\epsilon_{jt}\widehat{w}_i\widehat{w}_j. \qquad (18)
\end{aligned}
$$

From Assumption 3.4(b), $b\min_{1\leq i\leq J}\left\{\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{it}^2\right\} > bc$ with probability $1 - o(1)$. Finally, note that $\widehat{w}_i\widehat{w}_j > 0$ and $\sum_{i\neq j}\widehat{w}_i\widehat{w}_j < 1$. Therefore,

$$\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\sum_{i\neq j}\epsilon_{it}\epsilon_{jt}\widehat{w}_i\widehat{w}_j\right| \leq \max_{1\leq i,j\leq J, i\neq j}\left\{\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{it}\epsilon_{jt}\right|\right\} = o_p(1), \tag{19}$$

from Assumption 3.4(b). Combining these results, this contradicts $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}}\right)^2 = o_p(1)$, which implies that $\|\widehat{\mathbf{w}}_{\text{SC}}\|_2 \xrightarrow{p} 0$. ■

A.1.2 Proof of Corollary 3.1

Corollary 3.1 *Suppose all assumptions for Proposition 3.1 are satisfied, and that, for all $t\in\mathcal{T}_1$, ϵ_{it} is independent from $\{\epsilon_{i\tau}\}_{\tau\in\mathcal{T}_0}$. Then, for any $t\in\mathcal{T}_1$, $\mathbf{y}_t'\widehat{\mathbf{w}}_{SC} \xrightarrow{p} \boldsymbol{\lambda_t\mu}_0$ when $T_0, J\to\infty$, implying that $\hat{\alpha}_{0t}^{SC} \xrightarrow{p} \alpha_{0t}+\epsilon_{0t}$. Moreover, $\mathbb{E}^*\left[\hat{\alpha}_{0t}^{SC}-\alpha_{0t}\right]\to 0$.*

Proof. Note that, for $t\in\mathcal{T}_1$, $\hat{\alpha}_{0t}^{\text{SC}} = \alpha_{0t}+\epsilon_{0t}+\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\widehat{\boldsymbol{\mu}}_{\text{SC}})-\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}}$. Since we consider $\boldsymbol{\lambda}_t$ fixed, it follows from Proposition 3.1(i) that $\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\widehat{\boldsymbol{\mu}}_{\text{SC}}) \xrightarrow{p} 0$. Now under the assumptions from Corollary 3.1, it follows that $\boldsymbol{\epsilon}_t$ for $t\in\mathcal{T}_1$ is independent of $\widehat{\mathbf{w}}_{\text{SC}}$. In this case, $\mathbb{E}^*\left[(\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}})^2\right] \leq \left[\sup_{1\leq i\leq J}\mathbb{E}^*\left[\epsilon_{it}^2\right]\right]\mathbb{E}^*\left[\|\widehat{\mathbf{w}}_{\text{SC}}\|_2^2\right]$, where the first term is $O(1)$ given Assumption 3.1, and $\mathbb{E}^*\left[\|\widehat{\mathbf{w}}_{\text{SC}}\|_2^2\right]\to 0$ from Proposition 3.1(iii) and $\|\widehat{\mathbf{w}}\|_2^2$ bounded. Combining these results, $\hat{\alpha}_{0t}^{\text{SC}} \xrightarrow{p} \alpha_{0t}+\epsilon_{0t}$ when $T_0\to\infty$. Moreover, we have that

$$\mathbb{E}^*\left[\hat{\alpha}_{0t}^{\text{SC}}-\alpha_{0t}\right] = \boldsymbol{\lambda}_t\mathbb{E}^*[(\boldsymbol{\mu}_0-\widehat{\boldsymbol{\mu}}_{\text{SC}})]-\mathbb{E}^*[\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}}]\to 0, \tag{20}$$

where the first term is $o(1)$ because $\widehat{\boldsymbol{\mu}}_{\text{SC}} \xrightarrow{p} \boldsymbol{\mu}_0$ and $\widehat{\boldsymbol{\mu}}_{\text{SC}}$ is bounded, while the second term is equal to zero under the assumption that ϵ_{it} is independent from $\{\epsilon_{i\tau}\}_{\tau\in\mathcal{T}_0}$ for all $i=0,...,J$. ■

A.1.3 Proof of Proposition 4.1

Proposition 4.1 *Let $\widehat{\boldsymbol{\mu}}_{OLS}$ be defined as $\mathbf{M}_J{}'\widehat{\mathbf{b}}_{OLS}$, where $\widehat{\mathbf{b}}_{OLS}$ is defined in equation* (9)*. Assume that $J/T_0\to c\in[0,1)$, and that Assumptions 2.1, 2.2, 3.1, 4.1, and 4.2 hold. Then, when $T_0, J\to\infty$, $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{OLS} \xrightarrow{p} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ for all $t\in\mathcal{T}_1$.*

Proof. Given Assumption 4.1, we can label the first RF control units so that each block of F control units is such that the $F\times F$ matrix of factor loadings for each of those blocks, $\boldsymbol{\mu}(p)$ for $p=1,...,R$, are invertible with uniformly bounded $\|\boldsymbol{\mu}(p)\|_2$ and $\|(\boldsymbol{\mu}(p))^{-1}\|_2$. The

$(J-RF)\times F$ matrix with the factor loadings of the remaining $J-RF$ control units is defined as $\boldsymbol{\mu}(R+1)$. Therefore, $\mathbf{M}_J = [\boldsymbol{\mu}(1)' \ \dots \ \boldsymbol{\mu}(R)' \ \boldsymbol{\mu}(R+1)']'$. Likewise, let $\mathbf{y}_t(p)$ ($\boldsymbol{\epsilon}_t(p)$) be the $F\times 1$ vector of outcomes (errors) for the F control units in block $p \in \{1,...,R\}$ at time t, while $\mathbf{y}_t(R+1)$ ($\boldsymbol{\epsilon}_t(R+1)$) is the same information for the remaining $J-RF$ control units. We also define $\mathbf{Y}(p)$ ($\mathbf{E}(p)$) as the $T_0 \times F$ matrix with the pre-treatment periods observations for the outcomes (errors) of control units in group p. These terms without the index in the parenthesis will refer to information on all J control units. Finally, we define $\mathbf{y}_0$ ($\boldsymbol{\epsilon}_0$) as the $T_0 \times 1$ vector of pre-treatment outcomes (errors) of the treated unit.

Under Assumption 4.1, we can construct a $\boldsymbol{\beta}^*_J = (\boldsymbol{\beta}^*_J(1)', \ \dots \ , \boldsymbol{\beta}^*_J(R)', \boldsymbol{\beta}^*_J(R+1)')' \in \mathbb{R}^J$ such that $\mathbf{M}_J{}'\boldsymbol{\beta}^*_J = \sum_{p=1}^{R+1}\boldsymbol{\mu}(p)'\boldsymbol{\beta}^*_J(p) = \boldsymbol{\mu}_0$. For each $p=1,...,R$, let $\boldsymbol{\beta}^*_J(p) = \frac{1}{R}(\boldsymbol{\mu}(p)')^{-1}\boldsymbol{\mu}_0$, and $\boldsymbol{\beta}^*_J(R+1)=0$. In this case, $\boldsymbol{\beta}^*_J$ satisfies $\boldsymbol{\mu}_0 = \mathbf{M}_J{}'\boldsymbol{\beta}^*_J$, and we have that $\|\boldsymbol{\beta}^*_J\|_2 = o(1)$. It follows that $\mathbf{y}_0 = \mathbf{Y}\boldsymbol{\beta}^*_J + \boldsymbol{\epsilon}_0 - \mathbf{E}\boldsymbol{\beta}^*_J$. We consider a change in variables so that we can focus on $\boldsymbol{\mu}_0$. Since $\boldsymbol{\mu}(1)$ is invertible, we have that $\mathbf{y}_0 = \mathbf{Y}\mathbf{H}\mathbf{H}^{-1}\boldsymbol{\beta}^*_J + \boldsymbol{\epsilon}_0 - \mathbf{E}\boldsymbol{\beta}^*_J$, where

$$\mathbf{H} = \begin{bmatrix} (\boldsymbol{\mu}(1)')^{-1} & -(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(2)' & -(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(3)' & \dots & -(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(R+1)' \\ 0 & \mathbb{I}_F & 0 & ... & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & 0 & \dots & \mathbb{I}_{J-RF} \end{bmatrix} \tag{21}$$

and

$$\mathbf{H}^{-1} = \begin{bmatrix} \boldsymbol{\mu}(1)' & \boldsymbol{\mu}(2)' & \boldsymbol{\mu}(3)' & ... & \boldsymbol{\mu}(R+1)' \\ 0 & \mathbb{I}_F & 0 & ... & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & 0 & \dots & \mathbb{I}_{J-RF} \end{bmatrix}, \tag{22}$$

where $\mathbb{I}_q$ is a $q\times q$ identity matrix. Therefore, we have

$$\mathbf{y}_0 = \left[\mathbf{Y}(1)(\boldsymbol{\mu}(1)')^{-1}\right]\boldsymbol{\mu}_0 + \sum_{p=2}^{R+1}\left[\mathbf{Y}(p) - \mathbf{Y}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(p)'\right]\boldsymbol{\beta}^*_J(p) + \boldsymbol{\epsilon}_0 - \mathbf{E}\boldsymbol{\beta}^*_J. \tag{23}$$

Note that $\mathbf{Y}(p) - \mathbf{Y}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(p) = \boldsymbol{\epsilon}(p) - \boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(p)'$, which implies that

$$\mathbf{y}_0 = \left[\mathbf{Y}(1)(\boldsymbol{\mu}(1)')^{-1}\right]\boldsymbol{\mu}_0 + \sum_{p=2}^{R+1}\left[\boldsymbol{\epsilon}(p) - \boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(p)'\right]\boldsymbol{\beta}^*_J(p) + \boldsymbol{\epsilon}_0 - \mathbf{E}\boldsymbol{\beta}^*_J. \tag{24}$$

Now let $\widehat{\mathbf{b}}_{\text{OLS}}$ be the OLS estimator of $\mathbf{y}_0$ on $\mathbf{Y}$. Doing the same changes in variables as above, we have that

$$\mathbf{y}_0 = \left[\mathbf{Y}(1)(\boldsymbol{\mu}(1)')^{-1}\right]\widehat{\boldsymbol{\mu}}_{\text{OLS}} + \sum_{p=2}^{R+1}\left[\boldsymbol{\epsilon}(p) - \boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(p)'\right]\widehat{\mathbf{b}}_{\text{OLS}}(p) + \widehat{\mathbf{u}}, \tag{25}$$

where $\widehat{\boldsymbol{\mu}}_{\text{OLS}} = \mathbf{M}_J{}'\widehat{\mathbf{b}}_{\text{OLS}}$, and $\widehat{\mathbf{u}}$ is the OLS residual from $\mathbf{y}_0$ on $\mathbf{Y}$.

Using Frisch-Waugh-Lovell theorem, we have that

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}_{\text{OLS}} &= \left((\boldsymbol{\mu}(1))^{-1}\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)(\boldsymbol{\mu}(1)')^{-1}\right)^{-1}\left((\boldsymbol{\mu}(1))^{-1}\mathbf{Y}(1)'\mathbf{Q}\mathbf{y}_0\right) && (26)\\
&= \boldsymbol{\mu}(1)'\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{y}_0\right) && (27)\\
&= \boldsymbol{\mu}_0 + \boldsymbol{\mu}(1)'\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\boldsymbol{\epsilon}_{\mathbf{0}}\right) && (28)\\
&\quad -\boldsymbol{\mu}(1)'\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{E}\boldsymbol{\beta}^*_J\right), && (29)
\end{aligned}$$

where $\mathbf{Q}$ is the $(T_0{\times}T_0)$ residual-maker matrix for a regression on $\{\boldsymbol{\epsilon}(p) - \boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(p)'\}_{p=2}^{R+1}$.

We want to show that $\widehat{\boldsymbol{\mu}}_{\text{OLS}} \xrightarrow{p} \boldsymbol{\mu}_0$. Consider first the term $\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1) = \boldsymbol{\mu}(1)\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\Lambda}\boldsymbol{\mu}(1)' + 2\boldsymbol{\mu}(1)\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\epsilon}(1) + \boldsymbol{\epsilon}(1)'\mathbf{Q}\boldsymbol{\epsilon}(1)$. Let $K = T_0 - J + F$. Since $J/T_0 \to c \in [0,1)$, we have that $K \to \infty$. Also, $rank(\mathbf{Q}) = K$. From Assumption 4.2, we have that $\left\|\frac{1}{\sqrt{K}}\boldsymbol{\Lambda}'\mathbf{Q}\right\|_2^2 = \text{tr}\left(\frac{1}{K}\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\Lambda}\right) = O_p(1)$, which implies that $\left\|\frac{1}{\sqrt{K}}\boldsymbol{\mu}(1)\boldsymbol{\Lambda}'\mathbf{Q}\right\|_2^2 = \text{tr}\left(\frac{1}{K}\boldsymbol{\mu}(1)\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\Lambda}\boldsymbol{\mu}(1)'\right) = O_p(1)$. Now consider the term $\mathbf{Q}\boldsymbol{\epsilon}(1)$. By definition of $\mathbf{Q}$, we have that $\mathbf{Q}(\boldsymbol{\epsilon}(p) - \boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(p)') = 0$, which implies $\mathbf{Q}\boldsymbol{\epsilon}(1) = \mathbf{Q}\boldsymbol{\epsilon}(p)(\boldsymbol{\mu}(p)'))^{-1}\boldsymbol{\mu}(1)'$ for all $p = 1,..,R$. Therefore, $\mathbf{Q}\boldsymbol{\epsilon}(1) = \mathbf{Q}\frac{1}{R}\sum_{p=1}^{R}\boldsymbol{\epsilon}(p)(\boldsymbol{\mu}(p)')^{-1}\boldsymbol{\mu}(1)'$. Now define the $T_0 \times F$ matrix $\widetilde{\boldsymbol{\epsilon}}(p) \equiv \boldsymbol{\epsilon}(p)(\boldsymbol{\mu}(p)')^{-1}\boldsymbol{\mu}(1)'$, with elements $\tilde{\epsilon}_{ft}(p) = \mathbf{a}_f(p)'\boldsymbol{\epsilon}_t(p)$, where $\mathbf{a}_f(p)$ is an $F \times 1$ given by the f-th column of $(\boldsymbol{\mu}(p)')^{-1}\boldsymbol{\mu}(1)'$. Given that $var(\epsilon_{it})$ is uniformly bounded by $\bar{\gamma}$, it follows that $var(\tilde{\epsilon}_{ft}(p)) \leq \bar{\gamma}\left\|\mathbf{a}_f(p)\right\|_2^2$. Given Assumption 4.1, $\left\|\mathbf{a}_f(p)\right\|_2^2$ is uniformly bounded by an $\bar{\mathbf{a}}$, which implies that $var(\tilde{\epsilon}_{ft}(p)) \leq \bar{\gamma}\bar{\mathbf{a}}$. Now note that $\left\|\mathbf{Q}\frac{1}{R}\sum_{p=1}^{R}\widetilde{\boldsymbol{\epsilon}}(p)\right\|_2^2 = \sum_{f=1}^{F}\left\|\left[\mathbf{Q}\frac{1}{R}\sum_{p=1}^{R}\widetilde{\boldsymbol{\epsilon}}(p)\right]_f\right\|_2^2$, where, for a generic matrix $\mathbf{A}$, we define $[\mathbf{A}]_f$ as the f-th column of $\mathbf{A}$. Note that $\left\|\left[\mathbf{Q}\frac{1}{R}\sum_{p=1}^{R}\widetilde{\boldsymbol{\epsilon}}(p)\right]_f\right\|_2^2$ is the sum of squared residual of the OLS regression of $\left[\frac{1}{R}\sum_{p=1}^{R}\widetilde{\boldsymbol{\epsilon}}(p)\right]_f$ on $\{\boldsymbol{\epsilon}(p) - \boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(p)'\}_{p=2}^{R+1}$. Since $\mathbf{b} = 0 \in \mathbb{R}^{J-F}$ is a candidate solution for the

OLS, we have

$$
\begin{aligned}
\left\|\frac{1}{\sqrt{K}}\left(\left[\mathbf{Q}\frac{1}{R}\sum_{p=1}^{R}\widetilde{\boldsymbol{\epsilon}}(p)\right]_f\right)\right\|_2^2 \;&\leq\; \frac{1}{K}\sum_{t\in\mathcal{T}_0}\left(\frac{1}{R}\sum_{p=1}^{R}\tilde{\epsilon}_{ft}(p)\right)^2 = \frac{1}{R}\frac{T_0}{K}\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\frac{1}{\sqrt{R}}\sum_{p=1}^{R}\tilde{\epsilon}_{ft}(p)\right)^2 \\
&\leq \frac{1}{R}\bar{\gamma}\bar{\mathbf{a}}\frac{T_0}{K}\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\frac{\frac{1}{\sqrt{R}}\sum_{p=1}^{R}\tilde{\epsilon}_{ft}(p)}{\sqrt{var^*\left(\frac{1}{\sqrt{R}}\sum_{p=1}^{R}\tilde{\epsilon}_{ft}(p)\right)}}\right)^2. \qquad (30)
\end{aligned}
$$

Let $z_t = \left(\frac{\frac{1}{\sqrt{R}}\sum_{p=1}^{R}\tilde{\epsilon}_{ft}(p)}{\sqrt{var^*\left(\frac{1}{\sqrt{R}}\sum_{p=1}^{R}\tilde{\epsilon}_{ft}(p)\right)}}\right)^2$. By construction $\mathbb{E}^*[z_t]=1$. Since $\left(var^*\left(\frac{1}{\sqrt{R}}\sum_{p=1}^{R}\tilde{\epsilon}_{ft}(p)\right)\right)^2 \geq \underline{\gamma}^2\left(\frac{1}{R}\sum_{p=1}^{R}\|\mathbf{a}_f(p)\|_2^2\right)^2$, and $\mathbb{E}^*\left[\left(\frac{1}{\sqrt{R}}\sum_{p=1}^{R}\tilde{\epsilon}_{ft}(p)\right)^4\right] \leq \max\{\bar{\gamma}^2,\bar{\xi}\}\left(\frac{1}{R}\sum_{p=1}^{R}\|\mathbf{a}_f(p)\|_2^2\right)^2$, then $var^*(z_t)$ is uniformly bounded. Therefore, $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0} z_t \xrightarrow{p} 1$. Since $\frac{T_0}{K}\to\frac{1}{1-c}$, and $R\to\infty$, it follows that $\frac{1}{\sqrt{K}}\mathbf{Q}\boldsymbol{\epsilon}(1) = o_p(1)$. Combining the results above, we have that $\frac{1}{K}\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1) = \boldsymbol{\mu}(1)\left[\frac{1}{K}\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\Lambda}\right]\boldsymbol{\mu}(1)' + o_p(1)$. From Assumption 4.2, we have

$$
\left(\boldsymbol{\mu}(1)\left[\frac{1}{K}\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\Lambda}\right]\boldsymbol{\mu}(1)'\right)^{-1} = (\boldsymbol{\mu}(1)')^{-1}\left[\frac{1}{K}\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\Lambda}\right]^{-1}(\boldsymbol{\mu}(1))^{-1} = O_p(1), \qquad (31)
$$

which implies that $(\frac{1}{K}\mathbf{Y}'\mathbf{Q}\mathbf{Y})^{-1} = O_p(1)$.

Consider now $\mathbf{Y}(1)'\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^*$. From the definition of $\mathbf{Q}$, we have that

$$
\begin{aligned}
\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^*(p) &= \mathbf{Q}\sum_{p=1}^{R+1}\boldsymbol{\epsilon}(p)\boldsymbol{\beta}_J^*(p) = \sum_{p=1}^{R+1}\mathbf{Q}\boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(p)'\boldsymbol{\beta}_J^*(p) \\
&= \mathbf{Q}\boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\sum_{p=1}^{R+1}\boldsymbol{\mu}(p)'\boldsymbol{\beta}_J^*(p) = \mathbf{Q}\boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}_0, \qquad (32)
\end{aligned}
$$

which implies that $\frac{1}{\sqrt{K}}\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^* = o_p(1)$. Since $\frac{1}{\sqrt{K}}\mathbf{Q}\mathbf{Y}(1) = O_p(1)$, it follows that $\frac{1}{K}\mathbf{Y}(1)'\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^* = o_p(1)$.

Consider now $\mathbf{Y}(1)'\mathbf{Q}\boldsymbol{\epsilon}_0 = \boldsymbol{\mu}(1)\boldsymbol{\lambda}'\mathbf{Q}\boldsymbol{\epsilon}_0 + \boldsymbol{\epsilon}(1)'\mathbf{Q}\boldsymbol{\epsilon}_0$. Following the same arguments as above $\left\|\frac{1}{\sqrt{K}}\mathbf{Q}\boldsymbol{\epsilon}_0\right\|_2^2 \leq \frac{1}{K}\sum_{t\in\mathcal{T}_0}\epsilon_{0t}^2 = O_p(1)$, which implies $\frac{1}{K}\boldsymbol{\epsilon}(1)'\mathbf{Q}\boldsymbol{\epsilon}_0 = o_p(1)$. From Assumption 4.2, we have that $\frac{1}{K}\boldsymbol{\lambda}'\mathbf{Q}\boldsymbol{\epsilon}_0 = o_p(1)$, which implies $\frac{1}{K}\mathbf{Y}(1)'\mathbf{Q}\boldsymbol{\epsilon}_0 = o_p(1)$.

Combining all these results into equation 33, we have that $\widehat{\boldsymbol{\mu}}_{\text{OLS}} \xrightarrow{p} \boldsymbol{\mu}_0$. Since we consider a fixed sequence of $\boldsymbol{\lambda}_t$, it follows that $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{OLS}} \xrightarrow{p} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ for all $t\in\mathcal{T}_1$. ■

A.1.4 Proof of Proposition 4.2

Proposition 4.2 *Let $\widehat{\boldsymbol{\mu}}_{OLS}$ be defined as $\mathbf{M}_J{}'\widehat{\mathbf{b}}_{OLS}$, where $\widehat{\mathbf{b}}_{OLS}$ is defined in equation* (9). *Assume that $T_0 \geq J$, and that Assumptions 4.1, 4.3 and 4.4 hold. Then, when $T_0, J\to\infty$, $\mathbb{E}^{**}[\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{OLS} - \boldsymbol{\lambda}_t\boldsymbol{\mu}_0] \to 0$ and $\mathbb{E}^{**}\left[\hat{\alpha}_{0t}^{OLS} - \alpha_{0t}\right] \to 0$ for $t\in\mathcal{T}_1$.*

Proof. Following the same steps as the proof of Proposition 4.1, we have that

$$\widehat{\boldsymbol{\mu}}_{\text{OLS}} = \boldsymbol{\mu}_0 + \boldsymbol{\mu}(1)'\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\boldsymbol{\epsilon}_\mathbf{0}\right) - \boldsymbol{\mu}(1)'\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^*\right),$$

where $\mathbf{Q}$ is the $(T_0\times T_0)$ residual-maker matrix for a regression on $\{\boldsymbol{\epsilon}(p) - \boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}(p)'\}_{p=2}^{R+1}$.

We want to show that $\mathbb{E}^{**}\left[\widehat{\boldsymbol{\mu}}_{\text{OLS}} - \boldsymbol{\mu}_0\right] \to 0$. First, note that

$$\mathbb{E}^{**}\left[\boldsymbol{\mu}(1)'\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\mathbf{Y}(1)'\mathbf{Q}\boldsymbol{\epsilon}_\mathbf{0}|\mathbf{Y}\right] = \boldsymbol{\mu}(1)'\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\mathbf{Y}(1)'\mathbf{Q}\mathbb{E}^{**}\left[\boldsymbol{\epsilon}_\mathbf{0}|\mathbf{Y}\right] = 0.$$

Consider now the term $\boldsymbol{\mu}(1)'\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^*\right)$. Note that $\mathbb{E}^{**}[\mathbf{E}|\mathbf{Y}] \neq 0$. Therefore, with finite J, the estimator is biased, which is consistent with the results from Ferman and Pinto (2021). We show that, in a setting in which $J\to\infty$, this bias goes to zero. From equation 32, we have $\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^* = \mathbf{Q}\boldsymbol{\epsilon}(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}_0$. Therefore, we need to consider the conditional expectation $\mathbb{E}^{**}[\boldsymbol{\epsilon}(1)|\mathbf{Y}]$.

Given that $(\boldsymbol{\lambda}_t, \epsilon_{0t}, \boldsymbol{\epsilon}_t)$ is iid multivariate normal (Assumption 4.4), it follows that $\mathbb{E}^{**}[\boldsymbol{\epsilon}_t(1)|\mathbf{Y}] = \mathbb{E}^{**}[\boldsymbol{\epsilon}_t(1)|\mathbf{y}_t]$, where this conditional expectation is linear in $\mathbf{y}_t$. Using a change in variables, we can re-write this linear conditional expectation as $\mathbb{E}^{**}[\boldsymbol{\epsilon}_t(1)|\mathbf{y}_t] = \mathbf{B}^*\widetilde{\mathbf{y}}_t$, where $\widetilde{\mathbf{y}}_t = (\mathbf{y}_t(1)',\ \mathbf{y}_t(2)' - (\boldsymbol{\mu}(2)(\boldsymbol{\mu}(1))^{-1}\mathbf{y}_t(1))', \ldots, \mathbf{y}_t(R{+}1)' - (\boldsymbol{\mu}(R{+}1)(\boldsymbol{\mu}(1))^{-1}\mathbf{y}_t(1))')'$, and $\mathbf{B}^*$ is an $F\times J$ matrix. The j-th row of matrix $\mathbf{B}^*$ is given by $\mathbf{b}_j^* = \underset{\mathbf{b}\in\mathbb{R}^J}{\text{argmin}}\mathbb{E}^{**}\left[(\epsilon_{jt} - \mathbf{b}'\widetilde{\mathbf{y}}_t)^2\right]$. We show that the parameters associated with $\mathbf{y}_t(1)$, $\mathbf{b}_j^*(1)$, converge to zero when $J\to\infty$. For any $\widetilde{\mathbf{b}}_j$ such that $\widetilde{\mathbf{b}}_j(1)\neq 0$, note that $\mathbb{E}^{**}\left[\epsilon_{jt} - \widetilde{\mathbf{b}}_j'\widetilde{\mathbf{y}}_t\right]^2 \geq \widetilde{\mathbf{b}}_j(1)'\boldsymbol{\mu}(1)\mathbb{E}^{**}[\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t]\boldsymbol{\mu}(1)'\widetilde{\mathbf{b}}_j(1) >$

0, since $\boldsymbol{\mu}(1)\mathbb{E}^{**}[\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t]\boldsymbol{\mu}(1)'$ is positive definite. Now note that

$$\left[-\frac{1}{R}\boldsymbol{\mu}(1)(\boldsymbol{\mu}(p))^{-1}\right]\left[\mathbf{y}_t(p)-\boldsymbol{\mu}(p)(\boldsymbol{\mu}(1))^{-1}\mathbf{y}_t(1)\right]=\frac{1}{R}\left[\boldsymbol{\epsilon}_t(1)-\boldsymbol{\mu}(1)(\boldsymbol{\mu}(p))^{-1}\boldsymbol{\epsilon}_t(p)\right]. \quad (33)$$

Let $\widetilde{\mathbf{a}}_j(p)$ be the j-th row of $-\boldsymbol{\mu}(1)(\boldsymbol{\mu}(p))^{-1}$. Then the j-th column of $\left[-\frac{1}{R}\boldsymbol{\mu}(1)(\boldsymbol{\mu}(p))^{-1}\right][\mathbf{y}_t(p)-\boldsymbol{\mu}(p)(\boldsymbol{\mu}(1))^{-1}\mathbf{y}_t(1)]$ is given by $\frac{1}{R}\left(\epsilon_{jt}-\widetilde{\mathbf{a}}_j(p)\boldsymbol{\epsilon}_t(p)\right)$. Given Assumption 4.1, $\|(\boldsymbol{\mu}(p))^{-1}\|_2$ is uniformly bounded, for $p=1,...,R$. Therefore, $\|\widetilde{\mathbf{a}}_j(p)\|_2^2$ is uniformly bounded, which implies that $var(\widetilde{\mathbf{a}}_j(p)\boldsymbol{\epsilon}_t(p))$ is uniformly bounded. Therefore, if we choose $\mathbf{b}$ with $-\frac{1}{R}\widetilde{\mathbf{a}}_j(p)$ in the j-entry of block p, and zero otherwise, we have $\mathbb{E}^{**}\left[\epsilon_{jt}-\mathbf{b}'\widetilde{\mathbf{y}}_t\right]^2\to 0$ when $R\to\infty$. Therefore, it must be that $\mathbf{b}_j^*(1)\to 0$.

Back to equation 32, we have

$$\begin{aligned}\mathbf{Q}\mathbb{E}[\mathbf{E}|\mathbf{Y}]\boldsymbol{\beta}_J^* &= \mathbf{Q}\mathbb{E}^{**}[\boldsymbol{\epsilon}(1)|\mathbf{Y}](\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}_0=\mathbf{Q}\widetilde{\mathbf{Y}}(\mathbf{B}^*)'(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}_0 && (34)\\ &= \mathbf{Q}\mathbf{Y}(1)\mathbf{B}^*(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}_0, && (35)\end{aligned}$$

where $\mathbf{B}^*(1)$ is the first F columns of $\mathbf{B}^*$. The last equality follows from the definition of matrix $\mathbf{Q}$. Therefore,

$$(\boldsymbol{\mu}(1)')\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\mathbb{E}^{**}[\mathbf{E}|\mathbf{Y}]\boldsymbol{\beta}_J^*\right)=(\boldsymbol{\mu}(1)')\mathbf{B}^*(1)(\boldsymbol{\mu}(1)')^{-1}\boldsymbol{\mu}_0\to 0. \quad (36)$$

Combining the results above, we have that $\mathbb{E}^{**}[\widehat{\boldsymbol{\mu}}_{\text{OLS}}-\boldsymbol{\mu}_0]\to 0$. Therefore, for $t\in\mathcal{T}_1$, $\mathbb{E}^{**}[\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{OLS}}-\boldsymbol{\lambda}_t\boldsymbol{\mu}_0]=\mathbb{E}^{**}[\boldsymbol{\lambda}_t]\mathbb{E}^{**}[\widehat{\boldsymbol{\mu}}_{\text{OLS}}-\boldsymbol{\mu}_0]\to 0$.

Finally, note that

$$\mathbb{E}^{**}\left[\hat{\alpha}_{0t}^{\text{OLS}}-\alpha_{0t}\right]=\mathbb{E}^{**}[\boldsymbol{\lambda}_t\boldsymbol{\mu}_0-\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{OLS}}]+\mathbb{E}^{**}[\epsilon_{0t}-\boldsymbol{\epsilon}_t'\widehat{\mathbf{b}}_{\text{OLS}}]=\mathbb{E}^{**}[\boldsymbol{\lambda}_t\boldsymbol{\mu}_0-\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{OLS}}]\to 0. \quad (37)$$

■

Remark A.1 In the proof of Proposition 4.2 we have to consider the expectation of the term $\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^*\right)$ conditional on $\widetilde{\mathbf{Y}}$, because this term involves a non-linear function of these variables. We show that $\mathbb{E}^{**}\left[\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^*\right)|\widetilde{\mathbf{Y}}\right]$ equals a term that does not depend on $\widetilde{\mathbf{Y}}$, and converges to zero. Therefore, we also have that the unconditional expectation $\mathbb{E}^{**}\left[\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^*\right)\right]\to 0$. If we consider a fixed sequence of $\boldsymbol{\Lambda}$, then this proof would not work, because $\mathbb{E}^{**}\left[\mathbf{E}|\widetilde{\mathbf{Y}},\boldsymbol{\Lambda}\right]=\mathbf{E}$. Therefore, we cannot guarantee that $\mathbb{E}^{**}\left[\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{Y}(1)\right)^{-1}\left(\mathbf{Y}(1)'\mathbf{Q}\mathbf{E}\boldsymbol{\beta}_J^*\right)|\boldsymbol{\Lambda}\right]\to 0$.

A.2 Other results

A.2.1 Conditions for Assumption 3.2

Suppose the underlying distribution of $\boldsymbol{\mu}_i$ has finite support $\{\mathbf{m}_1, ..., \mathbf{m}_{\bar{q}}\}$, with $Pr(\boldsymbol{\mu}_i = \mathbf{m}_q) = p_q > 0$ independent across i. Fix a $\boldsymbol{\mu}_0 = \mathbf{m}_q$. If we let J_q be the number of observations with $\boldsymbol{\mu}_i = \mathbf{m}_q$, then by the Strong Law of Large Numbers, we have that $P\left(\frac{J_q}{J} \rightarrow p_q\right) = 1$. Therefore, with probability one, there is a $\widetilde{J} \in \mathbb{N}$ such that $J > \widetilde{J}$ implies $J_q/J > \frac{p_q}{2}$, which implies $J_q > cJ$ for a constant c. Now consider a $\mathbf{w}_J^*$ that assigns weights $\frac{1}{\lfloor cJ \rfloor}$ for $\lfloor cJ \rfloor$ control units with $\boldsymbol{\mu}_0 = \mathbf{m}_q$. By construction $\mathbf{M}_J{}'\mathbf{w}_J^* = \mathbf{m}_q$. Moreover, we have that $\|\mathbf{w}_J^*\|_2^2 = \frac{1}{\lfloor cJ \rfloor}$, which implies that $\|\mathbf{w}_J^*\|_2^2 \rightarrow 0$.

A.2.2 Conditions for Assumption 3.4

We show a very simple example in which Assumption 3.4 is satisfied. Suppose ϵ_{it} is independent across i and t, and has uniformly bounded k-th moments across i and t, for an even k. In this case, for any $\eta > 0$,

$$
\begin{aligned}
P\left(\max_{1\leq j\leq J}\left\{\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{0t}\epsilon_{jt}\right|\right\} > \eta\right) &\leq \sum_{j=1}^{J} P\left(\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{0t}\epsilon_{jt}\right| > \eta\right) && (38)\\
&\leq \sum_{j=1}^{J} \frac{\mathbb{E}\left[\left(\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{0t}\epsilon_{jt}\right)^k\right]}{\eta^k} && (39)\\
&\leq \frac{J}{T_0^k}\left(\sum_{h=1}^{\frac{k}{2}} C_h T_0^h\right), && (40)
\end{aligned}
$$

for constants $C_1, ..., C_{k/2}$. Therefore, $\max_{1\leq j\leq J}\left\{\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{0t}\epsilon_{jt}\right|\right\} = o_p(1)$ if $\frac{J}{T_0^{k/2}} \rightarrow 0$, which implies that this condition can be valid even when J grows at a faster rate than T_0 if we assume enough uniformly bounded moments for the idiosyncratic shocks. We can check the other conditions considered in Assumption 3.4 following the same idea. Note that the term $\max_{1\leq i,j\leq J, i\neq j}\left\{\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{it}\epsilon_{jt}\right|\right\}$ would be bounded by the sum of $J(J-1)/2$ terms, which implies that we would require a larger k to guarantee this condition.

A.2.3 Alternative for Corollary 3.1

We consider a different set of assumptions in which we can derive $\hat{\alpha}_{0t}^{\text{SC}} \xrightarrow{p} \alpha_{0t} + \epsilon_{0t}$ when $T_0 \to \infty$, allowing time-dependency for the idiosyncratic shocks. We consider the following set of assumptions, which are similar to the ones considered by Chernozhukov et al. (2021) for their Lemma 1.

Assumption A.1 *(Number of control units and pre-treatment periods)* $\log(J) = o(T_0^{\frac{\tau}{3\tau+1}})$, where τ is a constant defined in Assumption A.2.

Assumption A.2 *(idiosyncratic shocks)* (a) $\mathbb{E}^*[\epsilon_{it}] = 0$ for all i and t; (b) $\{\epsilon_{it}\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$ are independent across i; (c) $\{\epsilon_{0t}, ..., \epsilon_{Jt}\}_{t\in\mathcal{T}_0}$ is β-mixing, with coefficients satisfying $\beta(t) \leq D_1\exp(-D_2 t^{\tau})$, where $D_1, D_2, \tau > 0$ are constants; (d) ϵ_{it} have uniformly bounded fourth moments across i and t, and $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\mathbb{E}^*[\epsilon_{0t}^2] \to \sigma_0^2$; (e) $\exists \underline{\gamma} > 0$ such that $\mathbb{E}^*[\epsilon_{it}^2] \geq \underline{\gamma}$ across i and t.

Assumption A.3 *(factor loadings)* (a) As $J \to \infty$, there is a sequence $\mathbf{w}_J^* \in \Delta^{J-1}$ such that $\|\mathbf{M}_J{}'\mathbf{w}_J^* - \boldsymbol{\mu}_0\|_2 \to 0$, and $\|\mathbf{w}_J^*\|_2 \to 0$, and (b) the sequence $\boldsymbol{\mu}_i$ is uniformly bounded.

Assumption A.4 *(common factors)* (a) $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t \to \boldsymbol{\Omega}$ positive definite; (b) let $m = \lfloor[4D_2^{-1}\log(JT_0)]^{\frac{1}{\tau}}\rfloor$ and $k = \lfloor T_0/m\rfloor$, and define the sets $H_p = \{-p, -m-p, -2m-p, \ldots, -(k-1)m-p\}$ for $p = 1, ..., m$. We assume that there are positive constants b_1 and b_2 such that $\liminf_{T_0\to\infty}\left(\min_{p=1,...,m}\left\{\frac{1}{k}\sum_{t\in H_p}|\lambda_{ft}|^2\right\}\right) > b_1$ and $\limsup_{T_0\to\infty}\left(\max_{p=1,...,m}\left\{\frac{1}{k}\sum_{t\in H_p}|\lambda_{ft}|^3\right\}\right) < b_2$.

Assumption A.5 *(other assumptions)* (a) $\exists\ c > 0$ such that $\max_{1\leq j\leq J}\sum_{t\in\mathcal{T}_0}\left|\epsilon_{0t}^2\epsilon_{jt}^2\right| \leq c^2 T_0$ and $\max_{1\leq j\leq J}\sum_{t\in\mathcal{T}_0}\left|\lambda_{ft}^2\epsilon_{jt}^2\right| \leq c^2 T_0$ for all $f \in \{1, ..., F\}$ with probability $1 - o(1)$; (b) there is a sequence $l_J > 0$ such that $l_J[\log(T_0 \vee J)]^{\frac{1+\tau}{2\tau}} T_0^{-1/2} \to 0$ that satisfies (i) for any $t \in \mathcal{T}_1$, $(\boldsymbol{\epsilon}_t'\delta)^2 \leq l_J\frac{1}{T_0}\sum_{q\in\mathcal{T}_0}(\boldsymbol{\epsilon}_q'\delta)^2$ for all $\delta \in \Delta^{J-1}$ with probability $1 - o(1)$, and (ii) for $\mathbf{w}_J^*$ defined in Assumption A.3, $l_J^{1/2}\|\mathbf{M}_J{}'\mathbf{w}_J^* - \boldsymbol{\mu}_0\|_2 \to 0$, and $l_J^{1/2}\|\mathbf{w}_J^*\|_2 \to 0$.

We first prove the following lemma, which is based on Lemma H.8 from Chernozhukov et al. (2021).

Lemma A.1 Under Assumptions A.1, A.2, A.4, and A.5(a), we have that $\frac{1}{T_0}\left\|\sum_{t\in\mathcal{T}_0}\boldsymbol{\epsilon}_t\epsilon_{0t}\right\|_\infty = o_p(1)$, $\frac{1}{T_0}\left\|\sum_{t\in\mathcal{T}_0}\boldsymbol{\epsilon}_t\lambda_{ft}\right\|_\infty = o_p(1)$, and $\frac{1}{T_0}\left\|\sum_{t\in\mathcal{T}_0}\epsilon_{0t}\lambda_{ft}\right\|_\infty = o_p(1)$. Moreover, $l_J\frac{1}{T_0}\left\|\sum_{t\in\mathcal{T}_0}\boldsymbol{\epsilon}_t\epsilon_{0t}\right\|_\infty = o_p(1)$, $l_J\frac{1}{T_0}\left\|\sum_{t\in\mathcal{T}_0}\boldsymbol{\epsilon}_t\lambda_{ft}\right\|_\infty = o_p(1)$, and $l_J\frac{1}{T_0}\left\|\sum_{t\in\mathcal{T}_0}\epsilon_{0t}\lambda_{ft}\right\|_\infty = o_p(1)$, for l_t defined in Assumption A.5(b).

Proof. The result $l_J \frac{1}{T_0} \left\| \sum_{t \in \mathcal{T}_0} \boldsymbol{\epsilon}_t \epsilon_{0t} \right\|_\infty = o_p(1)$ follows simply from checking that the assumptions for Lemma H.8 from Chernozhukov et al. (2021) are valid given our assumptions. The results $l_J \frac{1}{T_0} \left\| \sum_{t \in \mathcal{T}_0} \boldsymbol{\epsilon}_t \lambda_{ft} \right\|_\infty = o_p(1)$, and $l_J \frac{1}{T_0} \left\| \sum_{t \in \mathcal{T}_0} \epsilon_{0t} \lambda_{ft} \right\|_\infty = o_p(1)$ follow from a minor adjustment on Lemma H.8 from Chernozhukov et al. (2021) to allow for a fixed sequence of $\boldsymbol{\lambda}_t$. We re-write their proof for our setting, considering these adjustments to allow for a fixed sequence of $\boldsymbol{\lambda}_t$.

Let $m = \lfloor [4D_2^{-1} \log(JT_0)]^{\frac{1}{\tau}} \rfloor$ and $k = \lfloor T_0/m \rfloor$, and define the sets $H_p = \{-p, -m - p, -2m - p, \ldots, -(k-1)m - p\}$ for $p = 1, ..., m$. We assume for now that T_0/m is an integer. From Berbee's coupling, there exist a sequence of random variables $\{\tilde{\epsilon}_{it}\}_{t \in H_p}$ such that (1) $\{\tilde{\epsilon}_{it}\}_{t \in H_p}$ is independent across t, (2) $\tilde{\epsilon}_{it}$ has the same distribution as ϵ_{it} for $t \in H_p$, and (3) $P\left(\cup_{t \in H_p} \{\tilde{\epsilon}_{it} \neq \epsilon_{it}\}\right) \leq k\beta(m)$. Now note that

$$\frac{1}{k} \sum_{t \in H_p} \mathbb{E}^* \left|\lambda_{ft} \tilde{\epsilon}_{it}\right|^2 = \frac{1}{k} \sum_{t \in H_p} \left|\lambda_{ft}\right|^2 \mathbb{E}^* \left|\tilde{\epsilon}_{it}\right|^2 \geq \underline{\gamma} \frac{1}{k} \sum_{t \in H_p} \left|\lambda_{ft}\right|^2, \tag{41}$$

where the last inequality follows from Assumption A.2. From Assumption A.4, there is a T_1^* such that, for $T_0 > T_1^*$, $\frac{1}{k} \sum_{t \in H_p} \left|\lambda_{ft}\right|^2 > b_1$, implying that $\frac{1}{k} \sum_{t \in H_p} \mathbb{E}^* \left|\lambda_{ft} \tilde{\epsilon}_{it}\right|^2 > \underline{\gamma} b_1$. Likewise, there is a T_2^* such that $\frac{1}{k} \sum_{t \in H_p} \mathbb{E}^* \left|\lambda_{ft} \tilde{\epsilon}_{it}\right|^3 < \bar{\gamma} b_2$ for $T_0 > T_2^*$. Therefore, for $T_0 > \max\{T_1^*, T_2^*\}$, we have

$$\frac{\left(\sum_{t \in H_p} \mathbb{E}^* \left|\lambda_{ft} \tilde{\epsilon}_{it}\right|^2\right)^{\frac{1}{2}}}{\left(\sum_{t \in H_p} \mathbb{E}^* \left|\lambda_{ft} \tilde{\epsilon}_{it}\right|^3\right)^{\frac{1}{3}}} = \frac{k^{\frac{1}{2}} \left(\frac{1}{k} \sum_{t \in H_p} \mathbb{E}^* \left|\lambda_{ft} \tilde{\epsilon}_{it}\right|^2\right)^{\frac{1}{2}}}{k^{\frac{1}{3}} \left(\frac{1}{k} \sum_{t \in H_p} \mathbb{E}^* \left|\lambda_{ft} \tilde{\epsilon}_{it}\right|^3\right)^{\frac{1}{3}}} > k^{\frac{1}{6}} C_0, \tag{42}$$

for a constant C_0. Now let $W_{it} = \lambda_{ft} \epsilon_{it}$ and $\widetilde{W}_{it} = \lambda_{ft} \tilde{\epsilon}_{it}$. Since $\mathbb{E}^*[\lambda_{ft} \tilde{\epsilon}_{it}] = 0$, by Theorem 7.4 of de la Peña et al. (2004), there is a constant C_1 such that, for any $0 \leq x \leq C_0 k^{\frac{1}{6}}$,

$$P\left(\left|\frac{\sum_{t \in H_p} \widetilde{W}_{it}}{\sqrt{\sum_{t \in H_p} \widetilde{W}_{it}^2}}\right| > x\right) \leq C_1(1 - \Phi(x)), \tag{43}$$

where $\Phi(.)$ is the cdf of a $N(0,1)$. Therefore, for any $0 \leq x \leq C_0 k^{\frac{1}{6}}$,

$$P\left(\left|\frac{\sum_{t\in H_p} W_{it}}{\sqrt{\sum_{t\in H_p} W_{it}^2}}\right| > x\right) \leq P\left(\left|\frac{\sum_{t\in H_p} \widetilde{W}_{it}}{\sqrt{\sum_{t\in H_p} \widetilde{W}_{it}^2}}\right| > x\right) + P\left(\cup_{t\in H_p}\{\tilde{\epsilon}_{it} \neq \epsilon_{it}\}\right) \quad (44)$$

$$\leq C_1(1-\Phi(x)) + k\beta(m). \quad (45)$$

Following exactly the same steps as the proof of Lemma H.8 from Chernozhukov et al. (2021), we have that, for any $0 \leq x \leq C_0 k^{1/6}\sqrt{m}$,

$$P\left(\max_{1\leq i\leq J}\left|\frac{\sum_{t\in \mathcal{T}_0} W_{it}}{\sqrt{\sum_{t\in \mathcal{T}_0} W_{it}^2}}\right| > x\right) \leq C_1 J m^{3/2} x^{-1} \exp\left(-\frac{x^2}{2m}\right) + D_1 J T_0 \exp\left(-D_2 m^{\tau}\right) (46)$$

Setting $x = 2\sqrt{m\log(Jm^{3/2})}$, and using that $\log(J) = o(T_0^{\frac{\tau}{3\tau+1}})$ (Assumption A.1), we have that, for large enough T_0, $x < C_0 k^{1/6}\sqrt{m}$. Moreover, we can show that the right-hand-side of equation 46 is $o(1)$. Therefore, for some constant κ,

$$\max_{1\leq i\leq J}\left|\sum_{t\in\mathcal{T}_0}\lambda_{ft}\epsilon_{it}\right| < \kappa[\log(T_0 \vee J)]^{(1+\tau)/(2\tau)} \max_{1\leq i\leq J}\sqrt{\sum_{t\in\mathcal{T}_0}\lambda_{ft}^2\epsilon_{it}^2} \quad (47)$$

with probability $1-o(1)$. Under Assumption A.5(a), we have that, with probability $1-o(1)$,

$$\max_{1\leq i\leq J}\frac{1}{T_0}\left|\sum_{t\in\mathcal{T}_0}\lambda_{ft}\epsilon_{it}\right| < \kappa[\log(T_0 \vee J)]^{(1+\tau)/(2\tau)} T_0^{-1/2} \max_{1\leq i\leq J}\sqrt{\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\lambda_{ft}^2\epsilon_{it}^2} \quad (48)$$

$$< \kappa[\log(T_0 \vee J)]^{(1+\tau)/(2\tau)} T_0^{-1/2} c. \quad (49)$$

It follows that $\max_{1\leq i\leq J}\frac{1}{T_0}\left|\sum_{t\in\mathcal{T}_0}\lambda_{ft}\epsilon_{it}\right| \xrightarrow{p} 0$ and, for l_J defined in Assumption A.5(b), $l_J \max_{1\leq i\leq J}\frac{1}{T_0}\left|\sum_{t\in\mathcal{T}_0}\lambda_{ft}\epsilon_{it}\right| \xrightarrow{p} 0$.

The proof that $\frac{1}{T_0}\left\|\sum_{t\in\mathcal{T}_0}\lambda_{ft}\epsilon_{0t}\right\|_\infty = o_p(1)$ and $l_J\frac{1}{T_0}\left\|\sum_{t\in\mathcal{T}_0}\lambda_{ft}\epsilon_{0t}\right\|_\infty = o_p(1)$ follows the same steps as above, by setting $J=1$ and considering only $i=0$. The proof that $\frac{1}{T_0}\left\|\sum_{t\in\mathcal{T}_0}\boldsymbol{\epsilon}_t\epsilon_{0t}\right\|_\infty = o_p(1)$ and $l_J\frac{1}{T_0}\left\|\sum_{t\in\mathcal{T}_0}\boldsymbol{\epsilon}_t\epsilon_{0t}\right\|_\infty = o_p(1)$ follows from noting that, since ϵ_{0t} and ϵ_{it} are β-mixing and independent, $\epsilon_{0t}\epsilon_{it}$ is also β-mixing (Theorem 5.2 from Bradley

(2005)). Moreover, from Assumption A.2, we have, similar to equation 42,

$$\frac{\left(\sum_{t\in H_p} \mathbb{E}^* \left|\tilde{\epsilon}_{0t}\tilde{\epsilon}_{it}\right|^2\right)^{\frac{1}{2}}}{\left(\sum_{t\in H_p} \mathbb{E}^* \left|\tilde{\epsilon}_{0t}\tilde{\epsilon}_{it}\right|^3\right)^{\frac{1}{3}}} > k^{\frac{1}{6}} C_0, \tag{50}$$

for some constant C_0. This inequality is valid for all T_0. Then we just follow the same steps of the proof setting $W_{it} = \epsilon_{0t}\epsilon_{it}$ and $\widetilde{W}_{it} = \tilde{\epsilon}_{0t}\tilde{\epsilon}_{it}$. ■

Given Lemma A.1, note that Assumptions A.1, A.2, A.3 A.4, and A.5(a) imply Assumptions 3.1, 3.2, 3.3, and 3.4(a). Therefore, the results (i) and (ii) from Proposition 3.1 remain valid under the assumptions considered in this appendix.

We now present the following conditions in which $\hat{\alpha}_{0t}^{\text{SC}} \xrightarrow{p} \alpha_{0t} + \epsilon_{0t}$.

Corollary A.1 *Suppose Assumptions A.1, A.2, A.3, A.4, and A.5 hold. Then, for any* $t \in \mathcal{T}_1$, $\hat{\alpha}_{0t}^{SC} \xrightarrow{p} \alpha_{0t} + \epsilon_{0t}$ *when* $T_0 \to \infty$.

Proof.

The proof follows exactly the same steps as the proof of Proposition 3.1 for results (i) and (ii). Now note that $\hat{\alpha}_{0t}^{\text{SC}} = \alpha_{0t} + \epsilon_{0t} + \boldsymbol{\lambda}_t(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}_{\text{SC}}) - \boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}}$. Since we are considering a fixed sequence of $\boldsymbol{\lambda}_t$, and $\widehat{\boldsymbol{\mu}}_{\text{SC}} \xrightarrow{p} \boldsymbol{\mu}_0$, it follows that $\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}_{\text{SC}}) \xrightarrow{p} 0$. It remains to show that $\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}} \xrightarrow{p} 0$.

From the proof of result (iii) from Proposition 3.1, we have that

$$\begin{aligned}
\|\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}\|_2^2 \;\leq\; & \|\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}\|_2^2 + \|\boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}_{\text{SC}})\|_2^2 \leq 2\left|\boldsymbol{\epsilon}_0'\boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}_{\text{SC}})\right| + 2\left|\boldsymbol{\epsilon}_0'\boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)\right| & (51)\\
& +2\left|(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}_{\text{SC}})'\boldsymbol{\Lambda}'\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}\right| + 2\left|(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)'\boldsymbol{\Lambda}'\mathbf{E}\mathbf{w}_J^*\right| + 2\left|\boldsymbol{\epsilon}_0'\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}\right| + & (52)\\
& +2\left|\boldsymbol{\epsilon}_0'\mathbf{E}\mathbf{w}_J^*\right| + \|\mathbf{E}\mathbf{w}_J^*\|_2^2 + \|\boldsymbol{\Lambda}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*)\|_2^2. & (53)
\end{aligned}$$

Now differently from what we do in the proof of Proposition 3.1, we can show using the assumptions considered in the appendix and Lemma A.1 that all the terms in the right hand side of the equation above are $o_p(1)$ when multiplied by l_J/T_0, for l_J defined in Assumption A.5(b). Therefore, $\frac{l_J}{T_0}\|\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}\|_2^2 = o_p(1)$. Now using Assumption A.5(b), we have that, for $t \in \mathcal{T}_1$, with probability $1 - o(1)$, $(\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}})^2 \leq \frac{l_J}{T_0}\|\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}}\|_2^2 = o_p(1)$, which completes the proof. ■

A.2.4 Demeaned SC estimator

Consider the demeaned SC estimator, where the weights are estimated by

$$\widehat{\mathbf{w}}_{\mathrm{SC}'} = \underset{\mathbf{w}\in\Delta^{J-1}}{\operatorname{argmin}} \left\{ \frac{1}{T_0} \sum_{t\in\mathcal{T}_0} \left(y_{0t} - \mathbf{w}'\mathbf{y}_t - (\bar{y}_0 - \bar{\mathbf{y}}'\mathbf{w})\right)^2 \right\} = \underset{\mathbf{w}\in\Delta^{J-1}}{\operatorname{argmin}} \left\{ \frac{1}{T_0} \sum_{t\in\mathcal{T}_0} \left(y_{0t} - \mathbf{w}'\mathbf{y}_t\right)^2 - \left(\bar{y}_0 - \bar{\mathbf{y}}'\mathbf{w}\right)^2 \right\}, \quad (54)$$

where $\bar{y}_0$ is the pre-treatment average of y_{0t}, and $\bar{\mathbf{y}}$ is the pre-treatment average of $\mathbf{y}_t$. We show that the results from Proposition 3.1 remain valid for the demeaned SC estimator with minor adjustments in the proof.

We start adjusting the objective function in equation 7 to

$$\ddot{\mathcal{H}}_J(\boldsymbol{\mu}) = \min_{\mathbf{w}\in\Delta^{J-1}:\ \mathbf{M}_J{}'\mathbf{w}=\boldsymbol{\mu}} \left\{ \frac{1}{T_0} \sum_{t\in\mathcal{T}_0} (\bar{\lambda}_t(\boldsymbol{\mu}) - \mathbf{w}'\boldsymbol{\epsilon}_t)^2 - \left(\bar{\boldsymbol{\lambda}}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}) + \bar{\epsilon}_0 - \bar{\boldsymbol{\epsilon}}'\mathbf{w}\right)^2 \right\}, \quad (55)$$

where $\bar{\boldsymbol{\lambda}}$, $\bar{\epsilon}_0$, and $\bar{\boldsymbol{\epsilon}}$ are pre-treatment averages of, respectively, $\boldsymbol{\lambda}_t$, ϵ_{0t}, and $\boldsymbol{\epsilon}_t$. We add to Assumption 3.3 that $\bar{\boldsymbol{\lambda}} \rightarrow \boldsymbol{\omega}_0$, and that $\boldsymbol{\Omega} - \boldsymbol{\omega}_0'\boldsymbol{\omega}_0$ is positive definite. The matrix $\boldsymbol{\Omega} - \boldsymbol{\omega}_0'\boldsymbol{\omega}_0$ will not be positive definite if there is a time-invariant common factor. However, we can redefine $\boldsymbol{\lambda}_t$ so that it does not include time-invariant common factors. Since time-invariant common factors are eliminated in the demeaning process, this will not lead to any problem in the analysis. We define $\tilde{\ddot{\mathcal{H}}}_{T_0}(\boldsymbol{\mu})$ as we do in equation 1. We also redefine the function $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$ to $\ddot{\sigma}^2_{\bar{\lambda}}(\boldsymbol{\mu}) = (\boldsymbol{\mu}_0 - \boldsymbol{\mu})'(\boldsymbol{\Omega} - \boldsymbol{\omega}_0'\boldsymbol{\omega}_0)(\boldsymbol{\mu}_0 - \boldsymbol{\mu}) + \sigma^2_\epsilon$, which is uniquely minimized at $\boldsymbol{\mu}_0$.

Following similar steps to the proof of Proposition 3.1, we can construct an upper bound to $\ddot{\mathcal{H}}_J(\boldsymbol{\mu}_0)$,

$$\tilde{\ddot{\mathcal{H}}}_{T_0}(\boldsymbol{\mu}_0) \leq \tilde{\ddot{\mathcal{H}}}_J^{UB}(\boldsymbol{\mu}_0) \equiv \frac{1}{T_0}\sum_{t\in\mathcal{T}_0} \bar{\lambda}_t(\boldsymbol{\mu}_J^*)^2 + \frac{1}{T_0}\sum_{t\in\mathcal{T}_0} (\boldsymbol{\epsilon}_t'\mathbf{w}_J^*)^2 - 2\frac{1}{T_0}\sum_{t\in\mathcal{T}_0} \bar{\lambda}_t(\boldsymbol{\mu}_J^*)(\boldsymbol{\epsilon}_t'\mathbf{w}_J^*) + \quad (56)$$

$$- \left(\bar{\lambda}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*) + \bar{\epsilon}_0 - \bar{\boldsymbol{\epsilon}}'\mathbf{w}_J^*\right)^2 + K_J \left\|\boldsymbol{\mu}_0 - \boldsymbol{\mu}_J^*\right\|_2 \xrightarrow{p} \sigma^2_\epsilon. \quad (57)$$

We can also define $\tilde{\ddot{\mathcal{H}}}_J^{LB}(\boldsymbol{\mu}_0)$ such that

$$\tilde{\ddot{\mathcal{H}}}_{T_0}(\boldsymbol{\mu}) \geq \tilde{\ddot{\mathcal{H}}}_J^{LB}(\boldsymbol{\mu}) = \min_{\mathbf{w}\in\mathcal{W}} \left\{ \frac{1}{T_0}\sum_{t\in\mathcal{T}_0} (\bar{\lambda}_t(\boldsymbol{\mu}) - \mathbf{w}'\boldsymbol{\epsilon}_t)^2 - \left(\bar{\lambda}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}) + \bar{\epsilon}_0 - \bar{\boldsymbol{\epsilon}}'\mathbf{w}\right)^2 \right\}, \quad (58)$$

and show that $\ddot{\tilde{\mathcal{H}}}_J^{LB}(\boldsymbol{\mu})$ converges uniformly in probability to $\ddot{\sigma}^2_{\bar{\lambda}}(\boldsymbol{\mu})$. We just have to show that the function $\min_{\mathbf{w}\in\mathcal{W}}\left\{\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\bar{\lambda}_t(\boldsymbol{\mu})-\mathbf{w}'\boldsymbol{\epsilon}_t)^2-\left(\bar{\lambda}(\boldsymbol{\mu}_0-\boldsymbol{\mu})+\bar{\epsilon}_0-\bar{\boldsymbol{\epsilon}}'\mathbf{w}\right)^2\right\}$ is Lipschitz as we do in equation 7, and then use the same strategy as we do in equation 8 to show pointwise convergence.

For the third result, we use that $\left\|\boldsymbol{\epsilon}_0+\boldsymbol{\lambda}(\boldsymbol{\mu}_0-\widehat{\boldsymbol{\mu}}_{\text{SC}'})-\mathbf{E}\widehat{\mathbf{w}}_{\text{SC}'}-(\bar{\epsilon}_0+\bar{\boldsymbol{\lambda}}(\boldsymbol{\mu}_0-\widehat{\boldsymbol{\mu}}_{\text{SC}'})-\bar{\boldsymbol{\epsilon}}\widehat{\mathbf{w}}_{\text{SC}'})\right\|_2^2\leq \left\|\boldsymbol{\epsilon}_0+\boldsymbol{\lambda}(\boldsymbol{\mu}_0-\boldsymbol{\mu}_J^*)-\mathbf{E}\mathbf{w}_J^*-(\bar{\epsilon}_0+\bar{\boldsymbol{\lambda}}(\boldsymbol{\mu}_0-\boldsymbol{\mu}_J^*)-\bar{\boldsymbol{\epsilon}}\mathbf{w}_J^*)\right\|_2^2$. Then we can follow the same steps as in the proof of part (iii) of Proposition 3.1.

A.2.5 SC specifications with covariates

Theoretical results

Consider now the case with covariates. The model for potential outcomes are now given by

$$\begin{cases} y_{it}^N=\boldsymbol{\lambda}_t\boldsymbol{\mu}_i+\boldsymbol{\theta}_t\mathbf{z}_i+\epsilon_{it}\\ y_{it}^I=\alpha_{it}+y_{it}^N,\end{cases}\tag{59}$$

where $\mathbf{z}_i$ is a $q\times 1$ vector of observed time-invariant covariates, and $\boldsymbol{\theta}_t$ are unobserved time-varying effects.

Define $\boldsymbol{\rho}_i\equiv(\boldsymbol{\mu}_i',\mathbf{z}_i')'$ and $\boldsymbol{\gamma}_t\equiv(\boldsymbol{\lambda}_t,\boldsymbol{\theta}_t)$, and assume that Assumptions 3.2, 3.3, and 3.4 are valid for $\boldsymbol{\rho}_i$ and $\boldsymbol{\gamma}_t$ instead of $\boldsymbol{\mu}_i$ and $\boldsymbol{\lambda}_t$. Botosaru and Ferman (2019) discuss cases in which the sequence $\boldsymbol{\gamma}_t$ might be multicolinear. In this case, Assumption 3.3 would not be valid for $\boldsymbol{\gamma}_t$, but the SC estimator would still control for the effects of these observed and unobserved covariates. While, for simplicity, we focus on the case in which Assumption 3.3 is valid for $\boldsymbol{\gamma}_t$, the same conclusions from Botosaru and Ferman (2019) apply here.

In this case, if we also have Assumption 3.1, then Proposition 3.1 is valid for the model $y_{it}^N=\boldsymbol{\gamma}_t\boldsymbol{\rho}_i+\epsilon_{it}$, where we treat the observed variables $\mathbf{z}_i$ as unobserved factor loadings, as Botosaru and Ferman (2019) do. Therefore, the SC weights using all pre-treatment outcomes as predictors will be such that $\widehat{\boldsymbol{\mu}}_{\text{SC}}\xrightarrow{p}\boldsymbol{\mu}_0$ and $\widehat{\mathbf{z}}_{\text{SC}}\equiv\mathbf{Z}_J'\widehat{\mathbf{w}}_{\text{SC}}\xrightarrow{p}\mathbf{z}_0$, where $\mathbf{Z}_J$ is the $J\times q$ matrix with information on the covariates $\mathbf{z}_i$ for all the controls.

Now consider an alternative SC specification, where the SC weights are estimated using both pre-treatment outcomes and the covariates as predictors. Abadie and Gardeazabal (2003) suggest a nested minimization problem where in the first step we select an $(R\times 1)$ vector of predictors for the treated unit, $\mathbf{x}_0$, and the corresponding $(R\times J)$ matrix of predictors of the control units, $\mathbf{X}_1$. The rows of these matrices may include functions of pre-treatment outcomes, and observed covariates. In the first step, they propose choosing weights that minimize the distance $\|\mathbf{x}_0 - \mathbf{X}_1\mathbf{w}\|_{\mathbf{V}}$ for a given positive semi-definite matrix $\mathbf{V}$. Then the matrix $\mathbf{V}$ is chosen to minimize mean squared errors of the difference between the pre-treatment outcomes and the weighted average of the control outcomes with weights $\mathbf{w}(\mathbf{V}) \in \text{argmin}_{\mathbf{w}\in\Delta^{J-1}} \|\mathbf{x}_0 - \mathbf{X}_1\mathbf{w}\|_{\mathbf{V}}$. Therefore, we can re-write this problem as $\widehat{\mathbf{w}}_{cov} = \underset{\mathbf{w}\in\Theta_J}{\text{argmin}} \left\{ \frac{1}{T_0}\sum_{t\in\mathcal{T}_0} \left(y_{0t} - \mathbf{w}'\mathbf{y}_t\right)^2 \right\}$, where $\Theta_J = \{\widetilde{\mathbf{w}} | \widetilde{\mathbf{w}} \in \text{argmin}_{\mathbf{w}\in\Delta^{J-1}} \|\mathbf{x}_0 - \mathbf{X}_1\mathbf{w}\|_{\mathbf{V}}$ for some $\mathbf{V}\}$. We can also consider a time frame for the minimization problem that defines $\widehat{\mathbf{w}}_{cov}$ different from the time frame for the minimization $\|\mathbf{x}_0 - \mathbf{X}_1\mathbf{w}\|_{\mathbf{V}}$, as suggested by Abadie et al. (2015). Our results still apply, as long as the number of periods in both minimization problems goes to infinity. As we do in Section 3, we define

$$\widetilde{\mathcal{H}}_J^{cov}(\boldsymbol{\rho}) = \min_{\widetilde{\boldsymbol{\rho}}\in\Gamma_J} \left\{ \min_{\mathbf{w}\in\Theta_J:\ [\mathbf{M}_J{}'\ \mathbf{Z}'_J]\mathbf{w}=\widetilde{\boldsymbol{\rho}}} \left\{ \frac{1}{T_0} \sum_{t\in\mathcal{T}_0} (\bar{\gamma}_t(\widetilde{\boldsymbol{\rho}}) - \mathbf{w}'\boldsymbol{\epsilon}_t)^2 \right\} + K_J \left\|\boldsymbol{\rho} - \widetilde{\boldsymbol{\rho}}\right\|_2 \right\}, \tag{60}$$

where Γ_J is the set of $\boldsymbol{\rho}$ that are attainable with the specification that includes covariates when there are J control units, and $\bar{\gamma}_t(\boldsymbol{\rho}) = \boldsymbol{\gamma}_t(\boldsymbol{\rho}_0 - \boldsymbol{\rho}) + \epsilon_{0t}$. As before, we have that $\widehat{\boldsymbol{\rho}}_{cov} = \text{argmin}_{\boldsymbol{\rho}\in\tilde{\Gamma}} \widetilde{\mathcal{H}}_J^{cov}(\boldsymbol{\rho})$, where $\tilde{\Gamma} = \text{cl}\left(\cup_{J\in\mathbb{N}}\Gamma_J\right)$.

Assume that the pre-treatment outcomes included as predictors when there are T_0 pre-treatment periods are such that Assumptions 3.3 and 3.4 are satisfied if we consider only the pre-treatment periods used as predictors. Therefore, it must be that the number of pre-treatment outcomes used as predictors goes to infinity when $T_0 \to \infty$. In this case, we show that the implied estimators for $\boldsymbol{\rho}_0 = (\boldsymbol{\mu}'_0, \mathbf{z}'_0)'$ are such that $\widehat{\boldsymbol{\mu}}_{cov} \xrightarrow{p} \boldsymbol{\mu}_0$ and $\widehat{\mathbf{z}}_{cov} \xrightarrow{p} \mathbf{z}_0$. First, consider a sequence of diagonal matrices $\mathbf{V}_J$ where the diagonal elements are equal to one for the pre-treatment outcomes, and zero otherwise. In this case, the minimization problem $\|\mathbf{x}_0 - \mathbf{X}_1\mathbf{w}\|_{\mathbf{V}_J}$ is equivalent to the problem analyzed in Section 3, and satisfies the conditions from Proposition 3.1. Therefore, $\widehat{\boldsymbol{\mu}}(\mathbf{V}_J) \equiv \mathbf{M}_J{}'\widehat{\mathbf{w}}(\mathbf{V}_J) \xrightarrow{p} \boldsymbol{\mu}_0$ and $\widehat{\mathbf{z}}(\mathbf{V}_J) \equiv \mathbf{Z}_J{}'\widehat{\mathbf{w}}(\mathbf{V}_J) \xrightarrow{p} \mathbf{z}_0$. Since these are candidate solutions for the minimization problem 60, it follows that

$$\widetilde{\mathcal{H}}_J^{cov}(\boldsymbol{\rho}_0) \leq \left\{ \frac{1}{T_0} \sum_{t\in\mathcal{T}_0} (\bar{\gamma}_t(\widehat{\boldsymbol{\rho}}_J(\mathbf{V}_J)) - \widehat{\mathbf{w}}(\mathbf{V}_J)'\boldsymbol{\epsilon}_t)^2 \right\} + K \left\|\boldsymbol{\rho}_0 - \widehat{\boldsymbol{\rho}}(\mathbf{V}_J)\right\|_2 \xrightarrow{p} \text{plim} \frac{1}{T_0} \sum_{t\in\mathcal{T}_0} \bar{\gamma}_t^2(\boldsymbol{\rho}_0). \tag{61}$$

Also, similarly to the proof of Proposition 3.1, $\widetilde{\mathcal{H}}_J^{cov}(\boldsymbol{\rho})$ is bounded from below by a function that converges uniformly to $\text{plim}\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\bar{\gamma}_t^2(\boldsymbol{\rho})$, which is uniquely minimized at $\boldsymbol{\rho}_0$. Following the same steps of the proof of Proposition 3.1, it follows that $\widehat{\boldsymbol{\mu}}_{cov}\xrightarrow{p}\boldsymbol{\mu}_0$ and $\widehat{\mathbf{z}}_{cov}\xrightarrow{p}\mathbf{z}_0$.

Monte Carlo simulations

We present an MC exercise similar to the one presented in Section 5, but with the inclusion of observed covariates. We continue to consider a setting with $\boldsymbol{\lambda}_t=[\lambda_{1t}\ \lambda_{2t}]$, but now we also add two observed covariates $\mathbf{z}_i=[z_{1i}\ z_{2i}]'$, with effects that vary with time, $\theta_{1t}\sim N(0,1)$ and $\theta_{2t}\sim N(0,1)$. The treated unit has $\boldsymbol{\mu}_0=(1,0)$, and $\mathbf{z}_0=(1,0)$. The control units are divided in four groups of equal size, with a combination of $\boldsymbol{\mu}_i\in\{(1,0),(0,1)\}$ and $\mathbf{z}_i\in\{(1,0),(0,1)\}$. Therefore, the goal of the SC estimator is to allocate positive weights only to the control units with $\boldsymbol{\mu}_i=(1,0)$, and $\mathbf{z}_i=(1,0)$.

We consider three specifications for the SC estimator. The first one uses all pre-treatment periods as predictors, the second one includes the first half of the pre-treatment outcomes and the two covariates as predictors, and the third one includes the pre-treatment outcome average and the two covariates as predictors. We present in Appendix Table A.1 the implied estimators for μ_{10} and z_{10} (given the adding-up constraint, $\hat{\mu}_{20}=1-\hat{\mu}_{10}$ and $\hat{z}_{20}=1-\hat{z}_{10}$). Overall, we see that the first two specifications perform very similarly. Moreover, when J and T_0 gets large, the SC unit using either of these two specifications control well both for the unobserved factor loading $\boldsymbol{\mu}$ and the observed covariates $\mathbf{z}$. The third specification, which does not satisfy the properties considered in the theory presented in this section, does a better job in matching the observed covariates, but does a poor job in matching the unobserved factor loadings. Even when $J,T_0\to\infty$, the estimator for μ_{10} remains roughly constant around 0.67, which suggests that the implied estimator for the factor loadings will not generally be consistent if we consider such specification to estimate the SC weights.

[Appendix Table A.1 here]

A.2.6 Simple example with $F=1$

Consider a simple example in which $y_{it}=\lambda_t+\epsilon_{it}$ for all $i=0,...,J$. Assume $\lambda_t\overset{iid}{\sim}N(0,\sigma_\lambda^2)$, $\epsilon_{jt}\overset{iid}{\sim}N(0,\sigma^2)$, and that these variables are all independent from each other. Finally, assume that $J/T_0\to c\in[0,1)$. Let $\mathbf{y}_i$ and $\boldsymbol{\epsilon}_i$ be the $T_0\times 1$ vectors with the outcomes and errors of unit i. Then $\mathbf{y}_0=\sum_{i=1}^J\frac{1}{J}\mathbf{y}_i+\boldsymbol{\epsilon}_0-\sum_{i=1}^J\frac{1}{J}\boldsymbol{\epsilon}_i$. Using Using Frisch-Waugh-Lovell theorem, we have that the OLS estimator associated to unit i, b_i, is given by

$$b_i = \left(\mathbf{y}_i'\mathbf{Q}_{(i)}\mathbf{y}_i\right)^{-1}\left(\mathbf{y}_i'\mathbf{Q}_{(i)}\mathbf{y}_0\right) = \frac{1}{J} + \left(\mathbf{y}_i'\mathbf{Q}_{(i)}\mathbf{y}_i\right)^{-1}\left(\mathbf{y}_i'\mathbf{Q}_{(i)}\left(\boldsymbol{\epsilon}_0 - \sum_{i'=1}^{J}\frac{1}{J}\boldsymbol{\epsilon}_{i'}\right)\right) \quad (62)$$

$$= \frac{1}{J} + \frac{\boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_0 + \boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_0 - \boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\sum_{i'=1}^{J}\frac{1}{J}\boldsymbol{\epsilon}_{i'} - \boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\sum_{i'=1}^{J}\frac{1}{J}\boldsymbol{\epsilon}_{i'}}{\boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\boldsymbol{\Lambda} + 2\boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i + \boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i}, \quad (63)$$

where $\mathbf{Q}_{(i)}$ is the residual-maker matrix of a regression on $\{\mathbf{y}_1,\dots,\mathbf{y}_J\}\setminus\{\mathbf{y}_i\}$. Let $A_i = \boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_0 + \boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_0 - \boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\sum_{i'=1}^{J}\frac{1}{J}\boldsymbol{\epsilon}_{i'} - \boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\sum_{i'=1}^{J}\frac{1}{J}\boldsymbol{\epsilon}_{i'}$ and $B_i = \boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\boldsymbol{\Lambda} + 2\boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i + \boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i$.

From Proposition 4.1, we know that $\sum_{i=1}^{J} b_i \xrightarrow{p} 1$, which implies that $\sum_{i=1}^{J}\frac{A_i}{B_i} \xrightarrow{p} 0$. We want to derive the probability limit of $\mathbf{b}'\mathbf{b} = \sum_{i=1}^{J} b_i^2$, which is given by

$$\sum_{i=1}^{J} b_i^2 = \frac{1}{J} + 2\frac{1}{J}\sum_{i=1}^{J}\frac{A_i}{B_i} + \sum_{i=1}^{J}\left(\frac{A_i}{B_i}\right)^2 = \sum_{i=1}^{J}\left(\frac{A_i}{B_i}\right)^2 + o_p(1). \quad (64)$$

Now note that

$$\frac{1}{\max_{i=1,\dots,J}\left\{\left(\frac{1}{K}B_i\right)^2\right\}}\sum_{i=1}^{J}\left(\frac{1}{K}A_i\right)^2 \leq \sum_{i=1}^{J}\left(\frac{A_i}{B_i}\right)^2 \leq \frac{1}{\min_{i=1,\dots,J}\left\{\left(\frac{1}{K}B_i\right)^2\right\}}\sum_{i=1}^{J}\left(\frac{1}{K}A_i\right)^2, \quad (65)$$

where $K = T_0 - J + 1$.

We first show that $\min_{i=1,\dots,J}\left\{\left(\frac{1}{K}B_i\right)^2\right\}$ and $\max_{i=1,\dots,J}\left\{\left(\frac{1}{K}B_i\right)^2\right\}$ converge in probability to σ^4. We start with the term $\frac{1}{K}\boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\boldsymbol{\Lambda}$. We can write $\boldsymbol{\Lambda} = \mathbf{Y}_{(i)}\boldsymbol{\phi} + \mathbf{u}_{(i)}$, where $\mathbf{Y}_{(i)}$ is a matrix with information on $\{\mathbf{y}_1,\dots,\mathbf{y}_J\}\setminus\{\mathbf{y}_i\}$, and $\boldsymbol{\phi}$ is the population OLS parameters of a regression of λ_t on $\{y_{1t},\dots,y_{Jt}\}\setminus\{y_{it}\}$. Given that the data is iid normal, we have that $\mathbf{u}_{(i)}|\mathbf{Y}_{(i)} \sim N(0,\sigma_u^2\mathbb{I}_{T_0})$. Moreover, it is easy to show that $\sigma_u^2 = o(1)$ and $J\sigma_u^2 = O(1)$. The intuition is that with the average of many observations y_{it} across i we become close to λ_t, so the variance of the error in this population OLS regression goes to zero when $J\to\infty$.

Therefore, conditional on $\mathbf{Y}_{(i)}$, $\frac{\boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\boldsymbol{\Lambda}}{\sigma_u^2} = \frac{\mathbf{u}_{(i)}'\mathbf{Q}_{(i)}\mathbf{u}_{(i)}}{\sigma_u^2} \sim \chi_K^2$, which implies that $\mathbb{E}^{**}\left[\frac{1}{K}\frac{\boldsymbol{\Lambda}'\mathbf{Q}_{(i)}\boldsymbol{\Lambda}}{\sigma_u^2}\right] =$

1 and $\mathbb{E}^{**}\left[\left(\frac{1}{K}\frac{\mathbf{\Lambda}'\mathbf{Q}_{(i)}\mathbf{\Lambda}}{\sigma_u^2}\right)^2\right]=O(1)$. Given that, for any $e>0$,

$$P\left(\max_{i=1,\ldots,J}\left|\frac{1}{K}\mathbf{\Lambda}'\mathbf{Q}_{(i)}\mathbf{\Lambda}\right|>e\right)\leq JP\left(\left|\frac{1}{K}\frac{\mathbf{\Lambda}'\mathbf{Q}_{(i)}\mathbf{\Lambda}}{\sigma_u^2}\right|>\frac{e}{\sigma_u^2}\right)\leq\sigma_u^2(J\sigma_u^2)\frac{\mathbb{E}^{**}\left[\left(\frac{1}{K}\frac{\mathbf{\Lambda}'\mathbf{Q}_{(i)}\mathbf{\Lambda}}{\sigma_u^2}\right)^2\right]}{e^2},\tag{66}$$

where $\sigma_u^2=o(1)$ and the other two terms are $O(1)$, which implies that $\max\limits_{i=1,\ldots,J}\left|\frac{1}{K}\mathbf{\Lambda}'\mathbf{Q}_{(i)}\mathbf{\Lambda}\right|\xrightarrow{p}0$. Now since $\boldsymbol{\epsilon}_i$ is independent from $\{\mathbf{y}_1,\ldots,\mathbf{y}_J\}\backslash\{\mathbf{y}_i\}$, we have $\frac{\boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i}{\sigma^2}\sim\chi_K^2$, which implies that $\frac{1}{K^2}\mathbb{E}\left[\left(\boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i-K\sigma^2\right)^4\right]=O(1)$. Therefore, for any $e>0$,

$$\begin{aligned}P\left(\max_{i=1,\ldots,J}\left|\frac{1}{K}\boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i-\sigma^2\right|>e\right) &\leq JP\left(\left|\frac{1}{K}\boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i-\sigma^2\right|>e\right) \\ &\leq \frac{1}{K}\frac{J}{K}\frac{\frac{1}{K^2}\mathbb{E}^{**}\left[\left|\boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i-K\sigma^2\right|^4\right]}{e^4}=o(1)O(1)O(1),\end{aligned}\tag{67}$$

which implies that $\max\limits_{i=1,\ldots,J}\left|\frac{1}{K}\boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i\right|\xrightarrow{p}\sigma^2$.

Finally, consider the term $\mathbf{\Lambda}'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_i=\sum_{1=1}^K\tilde{u}_{q(i)}\tilde{\epsilon}_{iq}$, where $\tilde{u}_{q(i)}\overset{iid}{\sim}N(0,\sigma_u^2)$, and $\tilde{\epsilon}_{iq}\overset{iid}{\sim}N(0,\sigma^2)$. Moreover, $\tilde{u}_{q(i)}$ and $\tilde{\epsilon}_q$ are independent. Therefore, $\mathbb{E}[\tilde{u}_{q(i)}\tilde{\epsilon}_{iq}]=0$, and $var[\tilde{u}_{q(i)}\tilde{\epsilon}_{iq}]=\sigma_u^2\sigma^2$. Therefore,

$$P\left(\max_{i=1,\ldots,J}\left|\frac{1}{K}\sum_{q=1}^K\tilde{u}_{q(i)}\tilde{\epsilon}_{iq}\right|>e\right)\leq JP\left(\left|\frac{1}{K}\sum_{q=1}^K\tilde{u}_{q(i)}\tilde{\epsilon}_{iq}\right|>e\right)\leq\frac{J}{K}\frac{\sigma_u^2\sigma^2}{e^2}=o(1),\tag{68}$$

because $J/K=O(1)$ and $\sigma_u^2=o(1)$.

Likewise, we can do the same calculations for $\min_{i=1,\ldots,J}\left\{\left(\frac{1}{K}B_i\right)^2\right\}$. Combining all these results, we have that $\min_{i=1,\ldots,J}\left\{\left(\frac{1}{K}B_i\right)^2\right\}$ and $\max_{i=1,\ldots,J}\left\{\left(\frac{1}{K}B_i\right)^2\right\}$ converge in probability to σ^4.

Now we consider $\sum_{i=1}^J\left(\frac{1}{K}A_i\right)^2$. Consider first $\sum_{i=1}^J\left(\frac{1}{K}\boldsymbol{\epsilon}_i'\mathbf{Q}_{(i)}\boldsymbol{\epsilon}_0\right)^2=\sum_{i=1}^J\left(\frac{1}{K}\sum_{q=1}^K\tilde{\epsilon}_{iq}\tilde{\epsilon}_{0(i)q}\right)^2=\frac{J}{K}\frac{1}{J}\sum_{i=1}^J\left(\frac{1}{\sqrt{K}}\sum_{q=1}^K\tilde{\epsilon}_{iq}\tilde{\epsilon}_{0(i)q}\right)^2$, where $\tilde{\epsilon}_{iq}\overset{iid}{\sim}N(0,\sigma^2)$, and $\tilde{\epsilon}_{0(i)q}\overset{iid}{\sim}N(0,\sigma^2)$. Note that

$\mathbb{E}^{**}\left[\left(\frac{1}{\sqrt{K}}\sum_{q=1}^{K}\tilde{\epsilon}_{iq}\tilde{\epsilon}_{0(i)q}\right)^2\right] = var^{**}\left[\left(\frac{1}{\sqrt{K}}\sum_{q=1}^{K}\tilde{\epsilon}_{iq}\tilde{\epsilon}_{0(i)q}\right)\right] = \sigma^4$. Moreover, we also have that $var^{**}\left(\frac{1}{J}\sum_{i=1}^{J}\left(\frac{1}{\sqrt{K}}\sum_{q=1}^{K}\tilde{\epsilon}_{iq}\tilde{\epsilon}_{0(i)q}\right)^2\right) \rightarrow 0$, which implies that $\frac{J}{K}\frac{1}{J}\sum_{i=1}^{J}\left(\frac{1}{\sqrt{K}}\sum_{q=1}^{K}\tilde{\epsilon}_{iq}\tilde{\epsilon}_{0(i)q}\right)^2 \xrightarrow{p} \frac{c}{1-c}\sigma^4$. Using similar calculations, all the other terms in the numerator converge in probability to zero.

Therefore, both the upper and the lower bounds from equation 65 converge in probability to $\frac{c}{1-c}$, which implies that $\mathbf{b}'\mathbf{b} \xrightarrow{p} \frac{c}{1-c}$. Now note that, for any $t \in \mathcal{T}_1$, $\hat{\alpha}_{0t} = \alpha_{0t} + \lambda_t(1 - \widehat{\mu}_{\text{OLS}}) + \epsilon_{0t} - \boldsymbol{\epsilon}_t'\mathbf{b}$. From Proposition 4.1, $\lambda_t(1 - \widehat{\mu}_{\text{OLS}}) \xrightarrow{p} 0$. Since data is iid normal across time, we have that $\epsilon_{0t} - \boldsymbol{\epsilon}_t'\mathbf{b}|\mathbf{b} \sim N\left(0, \sigma^2\left(1 + \mathbf{b}'\mathbf{b}\right)\right)$. Since $\mathbf{b}'\mathbf{b} \xrightarrow{p} \frac{c}{1-c}$, it follows that $\epsilon_{0t} - \boldsymbol{\epsilon}_t'\mathbf{b} \xrightarrow{d} N\left(0, \frac{\sigma^2}{1-c}\right)$, which implies that $\hat{\alpha}_{0t} \xrightarrow{d} N\left(\alpha_{0t}, \frac{\sigma^2}{1-c}\right)$.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, 93(1):113–132.

Andrews, D. W. K. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory*, 4(3):458–467.

Botosaru, I. and Ferman, B. (2019). On the role of covariates in the synthetic control method. *The Econometrics Journal*, 22(2):117–130.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probab. Surveys*, 2:107–144.

Chernozhukov, V., Wuthrich, K., and Zhu, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association*, 0(0):1–16.

de la Peña, V. H., Klass, M. J., and Leung Lai, T. (2004). Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *Ann. Probab.*, 32(3):1902–1933.

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pre-treatment fit. *Quantitative Economics, forthcoming*.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, 59(4):1161–1167.

Newey, W. K. and McFadden, D. (1994). Chapter 36 large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111 – 2245. Elsevier.

Table A.1: **Monte Carlo Simulations with Covariates**

	All pre-treatment outcome lags				Half of the pre-treatment outcome lags + covariates				Average of pre-treatment outcomes + covariates			
J	4	12	40	100	4	12	40	100	4	12	40	100
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: $T_0 = J + 5$												
$\mathbb{E}^{**}[\hat{\mu}_{01}]$	0.732	0.814	0.885	0.925	0.731	0.811	0.889	0.927	0.675	0.686	0.659	0.673
$se^{**}[\hat{\mu}_{01}]$	0.222	0.151	0.089	0.058	0.241	0.164	0.094	0.063	0.340	0.262	0.228	0.197
$\mathbb{E}^{**}[\hat{z}_{01}]$	0.733	0.820	0.880	0.921	0.770	0.840	0.890	0.925	0.858	0.956	0.989	0.992
$se^{**}[\hat{z}_{01}]$	0.200	0.148	0.090	0.055	0.202	0.147	0.091	0.060	0.188	0.122	0.048	0.038
$se^{**}(\hat{\alpha})$	1.408	1.275	1.132	1.063	1.430	1.277	1.142	1.070	1.496	1.353	1.222	1.184
Panel B: $T_0 = 2 \times J$												
$\mathbb{E}^{**}[\hat{\mu}_{01}]$	0.728	0.832	0.902	0.938	0.726	0.836	0.905	0.942	0.688	0.692	0.674	0.666
$se^{**}[\hat{\mu}_{01}]$	0.219	0.126	0.069	0.039	0.230	0.131	0.073	0.042	0.342	0.264	0.231	0.192
$\mathbb{E}^{**}[\hat{z}_{01}]$	0.738	0.827	0.908	0.938	0.772	0.840	0.912	0.941	0.865	0.962	0.986	0.995
$se^{**}[\hat{z}_{01}]$	0.230	0.128	0.066	0.042	0.229	0.129	0.067	0.043	0.204	0.099	0.055	0.024
$se^{**}(\hat{\alpha})$	1.406	1.186	1.098	1.058	1.407	1.203	1.104	1.069	1.566	1.294	1.225	1.223

Notes: this table presents the expected value and the standard error of the estimators for μ_{01} and z_{01} using the specification that includes all pre-treatment outcomes lags as predictors (columns 1 to 4), the first half of the pre-treatment outcome lags and the covariates as predictors (columns 5 to 8), and the average of the pre-treatment outcomes and the covariates as predictors (columns 9 to 12). It also presents the standard error of $\hat{\alpha}$. Since $\mathbb{E}^{**}[\lambda_t] = 0$, $\mathbb{E}^{**}[\hat{\alpha}_{01}] = 0$, which is the true treatment effect. Panel A presents results with $T_0 = J + 5$, while Panel B presents results with $T_0 = 2 \times J$. Results based on 500 simulations. The DGP is described in detail in Appendix Section A.2.5.