Inference for Factor Model with Synthetic Control under Fixed Number of Control-Units

Dennis Lim*

WORKING PAPER

Abstract

Current estimators are generally biased if treatment assignment is correlated with unobserved confounders, even when the number of pre-treatment periods goes to infinity. Ferman and Pinto (2021) show that a demeaned version of the SC method can substantially improve in terms of bias and variance relative to the difference-in-difference estimator; however, their proposed method assumes that (1) the number of control-unit increases and (2) error term and common-factors are asymptotically independent. Most commonly in empirical settings, (1) may not be sufficiently satisfied, leading to finite sample bias. This paper proposes a test that can consistently estimate the correct null when (1) control-units are fixed and (2) error terms and common-factors are dependent.

Keywords: Synthetic control, factor model

JEL Classification: C13, C21, C23

1 Introduction

In estimating treatment effects when the number of treated units are few, usual methods generally tend to fail due to the lack of asymptotic approximation. Abadie, Diamond, and Hainmueller (2010) proposed the synthetic control (SC) method, which works by estimating a weighted-average of control-units in the pre-treatment period and reconstructing the counterfactual treatment effect of the treated unit during the treatment period. A key requirement of this approach is that there exist weights such that a weighted average of the control-units can perfectly reconstruct the outcomes of the treated unit for a set of pre-treatment periods, called "perfect pre-treatment fit".

An important contribution by Ferman and Pinto (2021) is the introduction of "imperfect-fit", where such weights may not exist. In particular, they introduce a demeaned-version of the SC method that eliminates this problem, allowing consistent estimation/inference. However, a key requirement is that the number of control-units diverge to infinity. In fact, Ferman (2021) showed

*Singapore Management University. E-mail address: dennis.lim.2019@phdecons.smu.edu.sg.

that even when the number of control-units is larger than the number of pre-treatment periods, well-known estimators are generally consistent. To under the difficulty of estimation under a fixed number of control-units, according to Ferman and Pinto (2021), “If potential outcomes follow a linear factor model structure, then it would be possible to construct a counterfactual for the treated unit if we could consistently estimate the factor loadings. However, with fixed control units, it is only possible to estimate factor loadings consistently under strong assumptions on the idiosyncratic shocks (e.g. Bai (2003).” The main reason for the number of control-units increasing is so that the variance of the error becomes negligible. This error variance is given as σ_ε^2 in (2.3), which disrupts the recovery of our factor loading for our treated-unit from the pre-treatment period. Our paper contributes to the literature by providing a consistent test that allow the number of control-units to be fixed.

Structure of Paper: Section 2 provides the motivation and model setup of our paper. Section 3 provides the theoretical results as well as details of our proposed test in a heuristic manner. The proofs of the result in the main text are contained in the Appendix.

2 Model Setup

We are interested in testing

$$H_0 : \alpha_{0t} = \alpha \quad \text{versus} \quad H_1 : \alpha_{0t} \neq \alpha$$

at some time point t. We assume that we observe a balanced panel with $J+1$ individuals, from time $t = 1$ to $T := T_0 + T_1$, where T_0 is the number of periods that no individuals are treated; T_1 is the number of periods that individual $j = 0$ is treated, with the remaining individuals $j = 1, ..., J$ still untreated. Following Chernozhukov, Wuthrich, and Zhu (2021) and Ferman and Pinto (2021), we make the general assumptions for our model.

Assumption 1. *(potential outcome) The potential outcome for unit j at time t for the treated (y_{jt}^I) and non-treated (y_{jt}^N) are given by*

$$\begin{aligned} y_{jt}^N &= c_j + \delta_t + \lambda_t' \mu_j + \varepsilon_{jt} \\ y_{jt}^I &= \alpha_{jt} + y_{jt}^N \end{aligned} \tag{2.1}$$

where δ_t is an unknown common factor with constant factor loadings across units, c_j is an unknown time-invariant fixed effect, λ_t is a $(F \times 1)$ vector of common factors, μ_j is a $(F \times 1)$ vector of unknown factor loadings, and the error terms ε_{jt} are unobserved idiosyncratic shocks

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

To motivate the problem, consider the synthetic control weights in Abadie et al. (2010) given as

$$\widehat{W}^{SC} := \underset{W\in\Delta_\eta^J}{\arg\min} \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(y_{0t} - y_t'W)^2.$$

Fixing an $W \in \Delta_\eta^J$, let

$$\begin{aligned}\widehat{Q}_{T_0}(W) &:= \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(y_{0t} - y_t'W)^2 \\ &= \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left\{c_0 + \delta_t + \lambda_t'\mu_0 + \varepsilon_{0t} - (c'W + \delta_t\iota'W + \lambda_t'\mu W + \varepsilon_t'W)\right\}^2 \\ &\overset{(i)}{=} \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left\{(c_0 - c'W) + \lambda_t'(\mu_0 - \mu W) + (\varepsilon_{0t} - \varepsilon_t'W)\right)\right\}^2 \qquad (2.2)\end{aligned}$$

where $\mu := (\mu_1, ..., \mu_J)'$, and (i) follows from $\iota'W = 1$. Under some mild assumptions (see Ferman and Pinto (2021)[assumption 4]),

$$\widehat{Q}_{T_0}(W) \xrightarrow{p} Q_0(W) := \sigma_\varepsilon^2(1 + W'W) + \left[(c_0 - c'W)^2 + (\mu_0 - \mu W)'\Omega_0(\mu_0 - \mu W)\right], \qquad (2.3)$$

where it is assumed that $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\widetilde{\varepsilon}_t\widetilde{\varepsilon}_t' \xrightarrow{p} \sigma_\varepsilon^2 I_{J+1}$ for $\widetilde{\varepsilon}_t := (\varepsilon_{0t}, \varepsilon_t')'$. Then it can be argued that

$$\widetilde{W}^{SC} \xrightarrow{p} \overline{W} := \underset{W\in\Delta_\eta^J}{\arg\min}\, Q_0(W)$$

We see that the σ_ε^2 given in (2.3) prevents us from recovering the pre-treatment weights, i.e. the variance of the error coming from $\varepsilon_{0t} - \varepsilon_t'W$ given in (2.2). Ferman and Pinto (2021) explains that the only way to fully recover the pre-treatment weights is for $\sigma_\varepsilon^2 = 0$ or for the existence of some $W \in \widetilde{\Phi}|W \in \arg\min_{W:||W||=1}\{W'W\}$, which may not always hold. In view of this short-coming, Ferman (2021) suggests that "when the number of control units increases, the importance of the variance of this weighted average of the idiosyncratic shocks vanishes if it is possible to recover the factor-loadings of the treated unit with weights that are diluted among an increasing number of control units". However, the two assumptions needed are

(1) $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\lambda_t\varepsilon_t \xrightarrow{p} 0$

(2) the number of control-units increases

Another way to remove the error σ_ε^2 is to break the non-treated sample $\mathcal{T}_0$ into sub-samples and run a block synthetic control weight, since this will allow the error terms to be negligible by their mean-zero property; we call this new estimator $\widetilde{Q}_{t_0}(W)$, i.e. we can obtain

$$\arg\min_W \widetilde{Q}_{T_0}(W) \xrightarrow{p} \arg\min_W Q_0(W) \tag{2.4}$$

for $Q_0(W) = (c_0 - c'W)^2 + (\mu_0 - \mu W)'\Omega_0(\mu_0 - \mu W)$, which generally holds under mild regularity conditions and Ω_0 being positive-definite. This is due to the strict convexity of $Q_0(W)$ leading to a unique solution of $\arg\min Q_0(W)$. Therefore, intuitively, the block synthetic weights approach can remove the need for the number of control units to increase. However, when Ω_0 is only positive semi-definite, the solution set of $\arg\min Q_0(W)$ may not be unique, preventing the type of logic used in (2.4) to hold. We overcome this by adding a ridge term. Our paper contributes to the SC literature in obtaining consistent confidence intervals around α_{0t} under the true null, while relaxing both assumptions (1) and (2), i.e. we allow Ω_0 to be only positive semi-definite and the number of controls to be fixed.

3 Theory

Suppose first that we have an estimator $\widetilde{W}^{SC}$ of some sort, such that $\widetilde{W}^{SC} \xrightarrow{P} \overline{W}$, with $\overline{W} \in \widetilde{\Phi} := \{W \in \Delta_\eta^J : W'c = c_0 \text{ and } W'\mu = \mu_0\}$ and $\Delta_\eta^J := \{W \in \mathbb{R}^J : ||W|| \leq \eta\}$ for some $\eta > 0$ that will be specified later. Define $\widetilde{\Phi}^* := \{W \in \Delta_\eta^J : c'W = c_0\}$. Then we have the following corollary, which is useful when $\frac{1}{T_0}\sum_{s=1}^r \overline{\lambda}^s(\overline{\lambda}^s)' \xrightarrow{p} 0$ (see later theorem).

Corollary 3.1. *Suppose $\widetilde{W}^{SC} \xrightarrow{P} \overline{W} \in \widetilde{\Phi}^*$. Then as $T_0 \to \infty$, for any fixed $t \in \mathcal{T}_1$,*

$$\widehat{\alpha}_{0t} := y_{0t}^I - \boldsymbol{y_t'}\widehat{W}^{SC} \xrightarrow{p} \alpha_{0t} + \lambda_t\left(\mu_0 - \boldsymbol{\mu}'\overline{W}\right) + \left(\varepsilon_{0t} - \boldsymbol{\varepsilon}_t'\overline{W}\right)$$

If furthermore, we have that $\widetilde{W}^{SC} \xrightarrow{P} \overline{W} \in \widetilde{\Phi}$, then

$$\widehat{\alpha}_{0t} \xrightarrow{p} \alpha_{0t} + \lambda_t\left(\mu_0 - \boldsymbol{\mu}'\overline{W}\right) + \left(c_0 - \mathbf{c}'\overline{W}\right) + \left(\varepsilon_{0t} - \boldsymbol{\varepsilon}_t'\overline{W}\right) = \alpha_{0t} + \left(\varepsilon_{0t} - \boldsymbol{\varepsilon}_t'\overline{W}\right)$$

We want to apply Theorem 1 of Chernozhukov et al. (2021) to both cases of $\overline{W}$ given in corollary 3.1, in order to obtain a conformal inference under the null. For any $t \in \mathcal{T}_1$, we define $\widehat{u}_t := -\widehat{P}_t^N$, where $\widehat{P}_t^N := \alpha_{0t} - \widehat{\alpha}_{0t}$. For $t \in \mathcal{T}_0$, define $\widehat{u}_t := y_{0t}^N - \mathbf{y_t'}\widehat{W}^{SC}$. For any $t \in \mathcal{T}_0 \cap \mathcal{T}_1$, define $P_t^N := \overline{W}'y_t^N - y_{0t}^N$.

Denote $\widehat{S}(\widehat{u}) := T_1^{-1/2}\left|\sum_{t\in\mathcal{T}_1}\widehat{u}_t\right|$. Let $\mathcal{T}_0 = \{1, ..., T_0\}$ and $\mathcal{T}_1 = \{T_0 + 1, ..., T\}$. We define the moving permutation for $m \in \{0, 1, ..., T-1\}$ as $\Pi := \{\pi_m\}_{m=1}^{T-1}$, where

$$\pi_m(i) \begin{cases} i+m & \text{if } \; i+m \leq T \\ i+m-T & \text{otherwise} \end{cases}$$

Then define the p-value as

$$\widehat{p} := \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{u}_\pi) \geq S(\widehat{u})\} \tag{3.1}$$

For notational simplicity, define $\varepsilon_t := (\varepsilon_{1t}, ..., \varepsilon_{Jt})'$ and $y_t := (y_{1t}, ..., y_{Jt})'$. Throughout the rest of the paper, We assume T_1 is fixed, and $T_0 \to \infty$.

Theorem 1. *Suppose* $\widetilde{W}^{SC} \xrightarrow{p} \overline{W}$. *Assume* $\sup_t \mathbb{E}||\lambda_t||$, $\sup_t \mathbb{E}||\varepsilon_t|| \leq C < \infty$ *for some constant* C. *If any of the cases hold,*

$$\begin{aligned} &(i) \quad \overline{W} \in \widetilde{\Phi} \\ &(ii) \quad \overline{W} \in \widetilde{\Phi}^* \quad \textit{and} \quad \mathbb{E}\lambda_t = 0 \\ &(iii) \quad \overline{W} \in \widetilde{\Phi}^* \quad \textit{and} \quad \frac{1}{T}\sum_{t=1}^{T} \lambda_t \lambda_t' = o_p(1) \\ &(iv) \quad \overline{W} \in \widetilde{\Phi}^* \quad \textit{and} \quad \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}||\lambda_t||^2 = o(1) \end{aligned}$$

then under the correct null of α_{0t}, *for any* $\theta \in (0, 1)$,

$$|\mathbb{P}(\widehat{p} \leq \theta) - \theta| = o_p(1)$$

3.1 Block Synthetic control for fixed Λ

Consider a r-fold cross-fitting procedure, where we fix $r \in \mathbb{N}$ and define $\Delta := \lfloor \frac{T_0}{r} \rfloor$. Then Δ can be seen as the number of elements we want to fit in a single block (in the general case we can take Δ to be the ceiling of $\frac{T_0}{r}$). Then for any $j \in \{0, 1, ..., J\}$, and $s \in \{1, ..., r\}$ and $q \in \{1, ..., \Delta\}$ define $\varepsilon^s_{jq} := \varepsilon_{j, s\Delta+q}$ and $\overline{\varepsilon}^s := \frac{1}{\Delta}(\varepsilon_{s\Delta+1} + \varepsilon_{s\Delta+2} + ... + \varepsilon_{(s+1)\Delta})$. Furthermore, write $\overline{y}^s := \frac{1}{\Delta}(y_{s\Delta+1} + y_{s\Delta+2} + ... + y_{(s+1)\Delta})$. Then we can define the block-synthetic weight

$$\widetilde{W}^{SC}_T(\Lambda) := \underset{W \in \Delta^J_\eta}{\arg\min} \frac{1}{r} \sum_{s=1}^{r} \left\{ \overline{y}^s_0 - (\overline{y}^s)'W \right\}^2 + \Lambda ||W||^2$$

where $\Lambda > 0$ is some fixed value.

Assumption 3. *Suppose T_1 is fixed, $T_0 \to \infty$, $\Delta \to \infty$, $\mathbb{E}\varepsilon_{jt} = 0$ and $\mathbb{E}||\lambda_t||^2, \mathbb{E}\delta_t^2, \mathbb{E}(\varepsilon_{jt})^2 \leq \sigma^2 < \infty$ for every $j \in \{0, 1, ..., J\}$ and $t \in \{1, ..., T_0\}$. Furthermore, suppose $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0} \widetilde{\varepsilon}_t\widetilde{\varepsilon}_t' \xrightarrow{p} \sigma_\varepsilon^2 I_{J+1}$, $\frac{1}{T_0}\sum_{t=1}^{T_0} \lambda_t \xrightarrow{p} 0$ and $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0} \lambda_t\lambda_t' \xrightarrow{p} \Omega_0$, a positive semi-definite matrix.*

Unlike Ferman and Pinto (2021)[Assumption 4], we do not require $\frac{1}{T_0}\sum_{t\in\mathcal{T}_1} \lambda_t\varepsilon_t = o_p(1)$ - these assumptions could be satisfied under stronger conditions such as α-mixing with exponential speed, but may be hard to ascertain - to derive our asymptotic results. Rather, assumption 3 implies that this term is $O_p(1)$.

Assumption 4. *(Solution existence) Suppose at least one solution to $GW = G_0$ exists, where*

$$G_0 := \begin{pmatrix} c_0 \\ \mu_0 \end{pmatrix}$$

and $G := (c, \mu)'$. Furthermore, assume the chosen η is such that there exists some $W^\diamond$ being the solution with $||W^\diamond|| \leq \eta$.

Empirically, we can always choose η to be very large to ensure that some solutions fall within the prescribed Δ_η^J. In the event that G has full column rank, $W = (G'G)^{-1}G'(c_0, \mu_0')'$ is the solution to $GW = G_0$. The only time when assumption 4 fails is when G_0 is not a linear combination of G. When the number of covariates is large, this scenario will be unlikely.

Theorem 2. *Suppose assumption 1, 2, 3 and 4 holds. Then for any fixed $\gamma > 0$,*

$$\sup_{\Lambda\in[\gamma,1]} |\widetilde{W}_T^{SC}(\Lambda) - \overline{W}(\Lambda)| = o_p(1)$$

where $\overline{W}(\Lambda) := \arg\min_{W\in\Delta_\eta^J} \mathcal{A}(W, \Lambda)$ and $\mathcal{A}(W, \Lambda) := (c_0 - c'W)^2 + (\mu_0 - \mu W)'\Omega_0(\mu_0 - \mu W) + \Lambda||W||^2$. Furthermore, as $0 < \Lambda \downarrow 0$,

$$\overline{W}(\Lambda) \to W^*$$

where W^* is the minimum-norm vector among the set of vectors that solves $GW = G_0$.

Remark 1. *If we can strengthen Ω_0 in assumption 3, then Theorem 2 holds for $\gamma = 0$. In this case, we can have that for **any** sequence of $\Lambda_T \downarrow 0$,*

$$\widetilde{W}_T^{SC}(\Lambda_T) = W^* + o_p(1).$$

Then an application of Theorem 1 yields exact asymptotic size control under the correct null.

Note that W^* is unique by Lemma A.1. The difficulty in taking $\gamma = 0$ stems from the assumption that $\Omega_0 = Plim_{T_0\to\infty}\frac{1}{T_0}\sum_{t\in\mathcal{T}_0} \lambda_t\lambda_t'$ is only positive semi-definite, which implies that the solution set of the probability limit of $\widetilde{W}_T^{SC}(0)$ may not be unique. The implication is that $\widetilde{W}_T^{SC}(0)$ can potentially converge in probability to any $W \in \widetilde{\Phi}$ due to this lack of identification. To be specific, note that for any $\Lambda > 0$,

$$\overline{W}(\Lambda) = (\Lambda I + cc' + \mu'\Omega_0\mu)^{-1}(\mu'\Omega_0\mu_0 + c_0 c).$$

As $\Lambda \downarrow 0$, the inverse diverges to infinity. This prevents $\overline{W}(\Lambda)$ from being equi-continuous in $\Lambda \in (0,1]$; consequently we are not able select a finite number of points $\Lambda_i \in (0,1]$ such that the union of balls around Λ_i covers the interval $(0,1]$ and the probability that any $\Lambda \in (0,1]$ is covered by one of the balls lipschitz continuous. However, the next corollary shows us that we can have an "almost-consistent" estimator of the solution set $GW = G_0$ if we take $\Lambda > 0$ to be arbitrarily small.

Corollary 3.2. *Suppose assumption 1, 2, 3 and 4 holds. Then for any $\xi > 0$, there exists a $\Lambda(\xi) > 0$ such that for any fixed $0 < \Lambda \leq \Lambda(\xi)$,*

$$|\widetilde{W}^{SC}(\Lambda) - W^*| \leq \xi + o_p(1)$$

Furthermore, Theorem 2 assures us that there is some sequence of Λ_T such that $\widetilde{W}_T^{SC}(\Lambda_T)$ consistently estimates W^*. This is formalized below.

Corollary 3.3. *Suppose assumption 1, 2, 3 and 4 holds. Then there exists a sequence $0 < \Lambda_T \downarrow 0$ such that*

$$\widetilde{W}_T^{SC}(\Lambda_T) = W^* + o_p(1)$$

By an application of corollary 3.3 and Theorem 1, we see that under the null, we can apply Theorem 1 for $\widehat{p}$ in (3.1) and obtain conformal inference if we have an idea of the rate of Λ_T is. Despite this difficulty, we can obtain an "almost exact" size control for any given $\Lambda \in (0,1]$. Intuitively, by choosing a small enough Λ, $\widetilde{W}_T^{SC}(\Lambda)$ should be approximately W^* asymptotically. We should therefore have an "almost-exact" size control in the sense of theorem 1 for small enough Λ. This is formalized in Theorem 3 below.

For any $t \in \mathcal{T}_1$, define

$$\begin{aligned}\widehat{\alpha}_{0t}(\Lambda) &:= y_{0t}^I - \mathbf{y_t'}\widetilde{W}_T^{SC}(\Lambda) \\ \widehat{P}_t^N(\Lambda) &:= \alpha_{0t} - \widehat{\alpha}_{0t}(\Lambda) =: -\widehat{u}_t(\Lambda)\end{aligned}$$

$$
\begin{aligned}
P_t^N &:= W^{*\prime}\mathbf{y}_t^N - y_{0t}^N =: -u_t \\
\widehat{S}(\widehat{u}(\Lambda)) &:= T_0^{-1/2}|\sum_{t\in\mathcal{T}_0}\widehat{u}_t(\Lambda)| \\
\widehat{p}(\Lambda) &:= \frac{1}{\Pi}\sum_{\pi\in\Pi}\mathbb{1}\{S(\widehat{u}_\pi(\Lambda)) \geq S(\widehat{u}(\Lambda))\}
\end{aligned}
$$

Theorem 3 (Main result)**.** *Suppose assumption 1-4 holds. Assume $S(u)$ has pdf bounded above by D, and $\{u_t\}_{t=1}^T$ is stationary and strong-mixing. Under the correct null of α_{0t}, fixing any $\Lambda \in (0,1]$ and $\theta \in (0,1)$, we have*

$$
\begin{aligned}
|\mathbb{P}(\widehat{p}(\Lambda) \leq \theta) - \theta| \leq \overline{C}\Big\{&(T_1/T_0)^{1/4} logT_0 + \max(C_1, C_2)\cdot||\overline{W}(\Lambda) - W^*|| \\
&+ \max(C_1^{1/2}, C_2^{1/2})\cdot||\overline{W}(\Lambda) - W^*||^{1/2}\Big\} + \gamma_T
\end{aligned}
$$

where $C_1 := 2\sqrt{(||c||^2 + 2||\mu||^2\sigma^2)}$, $C_2 := ||c|| + \sigma^2||\mu|| + \sigma^2$, $\overline{C}$ is some universal constant, and $\gamma_T = o(1)$ is such that with probability at least $1-\gamma_T$,

$$
||\widehat{P}^N(\Lambda) - P^N||_2/\sqrt{T} \leq C_1||\overline{W}(\Lambda) - W^*||,
$$

i.e. the speed of convergence of $\widetilde{W}_T^{SC}$ to $\overline{W}(\Lambda)$.

Remark 2. *y Theorem 2, as $0 < \Lambda \downarrow 0$ then $\overline{W}(\Lambda_T) \to W^*$. Combining with Theorem 3, there exists some $0 < \Lambda_T \downarrow 0$ such that $|\mathbb{P}(\widehat{p}(\Lambda_T) \leq \theta) - \theta| = o(1)$. This rate is generally unknown unless we impose more assumptions on the structure on the model. In practice, we can simply employ a very small $\Lambda_T \approx 0$ so that the estimation error is negligible.*

3.2 Block synthetic control is efficient under mis-specification

In this section we explore the asymptotic variance of different estimators of α_{0t}, i.e. the treatment effect. For every $\Lambda > 0$ and $t \in \mathcal{T}_1$, recall that we defined our block estimator as

$$
\widetilde{\alpha}_t^{SC}(\Lambda) := y_{0t}^I - y_t'\widetilde{W}_t^{SC}(\Lambda).
$$

By Theorem 2 we see that

$$
\begin{aligned}
\widetilde{\alpha}_t^{SC}(\Lambda) &\xrightarrow{p} \alpha_{0t} + y_{0t}^N - y_t'W^* - y_t'\left(\overline{W}(\Lambda) - W^*\right) \\
&= \alpha_{0t} + y_t'(\overline{W}(\Lambda) - W^*) \\
&= \alpha_{0t} + (c + \iota\delta_t + \mu\lambda_t + \varepsilon_t)'(\overline{W}(\Lambda) - W^*) \qquad (3.2)
\end{aligned}
$$

so that

$$\begin{aligned}
avar(\widetilde{\alpha}_t^{SC}(\Lambda)) &= (\overline{W}(\Lambda) - W^*)'(Var(\delta_t)\iota\iota' + \mu\, Var(\lambda_t)\mu' + \sigma^2 I)(\overline{W}(\Lambda) - W^*) \\
&\leq \sigma^2(\overline{W}(\Lambda) - W^*)'(\iota\iota' + \mu\mu' + I)(\overline{W}(\Lambda) - W^*) \\
&\leq \sigma^2 ||\overline{W}(\Lambda) - W^*||^2 \lambda_{max}(\Sigma)
\end{aligned}$$

where $\Sigma := \iota\iota' + \mu\mu' + I$

Next, we show that $\widetilde{\alpha}_t^{SC}(\Lambda)$ is more efficient than the class of estimators

$$\widehat{\alpha}_t(W) := [y_{0t} - y_t'W] + \frac{1}{T_0}\sum_{\tau\in\mathcal{T}_0}[y_{0\tau} - y_\tau'W] \tag{3.3}$$

for any $W \in \mathbb{R}^J$, when $0 < \Lambda \leq \overline{\Lambda}$, for some $\overline{\Lambda}$. Note that the difference-in-difference (DID) estimator is a special case of (3.3) with $W = W_* := (\frac{1}{J}, ..., \frac{1}{J})'$ (e.g. Doudchenko and Imbens (2016)), i.e.

$$\widehat{\alpha}_t^{DID} := \widehat{\alpha}_t(W_*)$$

We assume $\sigma_\varepsilon^2 > 0$ for simplicity. As shown in Ferman and Pinto (2021)[proposition 3], under assumptions 1-3, as $T_0 \to \infty$, then for any $t \in \mathcal{T}_1$, for any $\widetilde{W} \xrightarrow{p} \underline{W}$, we have

$$avar(\widehat{\alpha}_t(\widetilde{W})) = \sigma_\varepsilon^2(1 + \underline{W}'\underline{W}) + (c_0 - c'\underline{W})^2 + (\mu_0 - \mu'\underline{W})'\Omega_0(\mu_0 - \mu\underline{W}) \geq \sigma_\varepsilon^2$$

By Theorem 2 we know that there exists some $\overline{\Lambda} > 0$ such that for every $0 < \Lambda \leq \overline{\Lambda}$, then $||\overline{W}(\Lambda) - W^*|| \leq \frac{\sigma_\varepsilon^2}{\sigma^2\lambda_{max}(\Sigma)}$. In particular, for every $0 < \Lambda \leq \overline{\Lambda}$, $\widetilde{\alpha}_{0t}^{SC}$ is more efficient than the DID estimator.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association 105*(490), 493–505.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Chernozhukov, V., K. Wuthrich, and Y. Zhu (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association 116*(536), 1849–1864.

Doudchenko, N. and G. W. Imbens (2016, October). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. (22791).

Ferman, B. (2021). On the properties of the synthetic control estimator with many periods and many controls. *Journal of the American Statistical Association 116*(536), 1764–1772.

Ferman, B. and C. Pinto (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics 12*(4), 1197–1221.

Newey, W. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica 59*(4), 1161–67.

Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engel and D. McFadden (Eds.), *Handbook of Econometrics*. Elsvier Science, Amsterdam: Elsvier Science.

A Auxiliary Lemmas

Lemma A.1. *Let $A \in \mathbb{R}^{m\times n}$, so that by singular value decomposition we can write $A = U\Sigma V'$, where $\Sigma \in \mathbb{R}^{m\times n}$ has non-zero elements only on its diagonal, with these values equal $\sigma_1, ..., \sigma_r$. The minimum-norm least squares solution to the linear equation $AX = b$, that is, the shortest vector X that achieves*

$$\min_X ||AX - b||^2 \equiv \sum_{i=r+1}^{n} (U_i' b)^2$$

is unique, given by

$$\widehat{X} = V\Sigma^{\dagger} U' b$$

where

$$\Sigma^{\dagger} = \begin{pmatrix} 1/\sigma_1 & & & & 0 & \cdots & 0 \\ & 1/\sigma_2 & & & & & \\ & & \ddots & & \vdots & & \vdots \\ & & & 1/\sigma_r & \vdots & & \vdots \\ & & & & 0 & & \\ & & & & \vdots & \ddots & \\ & & & & 0 & \cdots & 0 \end{pmatrix}$$

Also, $||\widehat{X}||^2 = \sum_{i=1}^{r} (U_i' b/\sigma_i)^2$

Lemma A.1:
The least square solution to $AX = b$ can be written as

$$\min_X ||U\Sigma V'X - b|| = \min_X ||U(\Sigma V'X - U'b)|| \overset{(i)}{=} \min_X ||(\Sigma V'X - U'b)|| \overset{(ii)}{=} \min_y ||(\Sigma y - c)||$$

where (i) follows from the fact that U is orthogonally-normalized so that the euclidean-norm remains unchanged; (ii) follows by defining $y := V'X$ and $c := U'b$. We want to minimize the vector

$$\begin{pmatrix} \sigma_1 & 0 & & \cdots & 0 & 0 \\ & \ddots & & & & 0 \\ & & \sigma_r & & & \vdots \\ & & & 0 & & \vdots \\ & & & & \ddots & \vdots \\ & & & & & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_r \\ y_{r+1} \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} c_1 \\ \vdots \\ \vdots \\ c_r \\ c_{r+1} \\ \vdots \\ c_n \end{pmatrix}$$

which leads to the solution

$$y_i = \frac{c_i}{\sigma_i} \quad for\ i \in 1, ..., r$$

with the choice of y_i to be any number for $i \in r+1, ..., n$. However, note that by $VV' = I$, we have $||X|| = ||V'X|| = ||y||$. In order to minimize $||X||$ we have to minimize $||y||$, which forces us to choose $y_i := 0$ for $i \in r+1, ..., n$, i.e. $y = \Sigma^{\dagger} c$ is the unique solution to the minimum-norm least square problem. Solving for X yields

$$\widehat{X} = Vy = V\Sigma^{\dagger} c = V\Sigma^{\dagger} U' b$$

It is clearly unique. Furthermore, since $y_i \equiv 0$ for $i = r+1, ..., n$

$$\min_X ||AX - b|| = ||A\widehat{X} - b|| = ||(-c_{r+1}, \cdots, -c_n)|| = \sum_{i=r+1}^{n} (U_i' b)^2.$$

Finally,

$$||\widehat{X}||^2 = ||Vy||^2 = ||y||^2 = \sum_{i=1}^{r} (c_i/\sigma_i)^2 = \sum_{i=1}^{r} (U_i' b/\sigma_i)^2$$

□

B Proof of Theorem 1

Define $y_t := (y_{1t}, ..., y_{Jt})'$. Using the notations of Chernozhukov et al. (2021), consider the respective models:

Case (i),(ii)

$$P_t^N := \overline{W}' y_t^N - y_{0t}^N =: -u_t$$

Case (iii)

$$P_t^N := \overline{W}' y_t^N - y_{0t}^N - \overline{W}' \mu \lambda_t =: -u_t$$

Case (iv)

$$P_t^N := \overline{W}' y_t^N - y_{0t}^N - \overline{W}' \mu \mathbb{E} \lambda_t =: -u_t$$

Then note that $\mathbb{E} u_t = 0$ for all cases; defining $\gamma_t \equiv 0$ for case (i) and (ii), $\gamma_t := \overline{W}' \mu \lambda_t$ in case (iii), $\gamma_t := \overline{W}' \mu \mathbb{E} \lambda_t$ in case (iv), we have

$$\begin{aligned}
\widehat{P}_t^N - P_t^N &= -(y_{0t}^I - y_t' \widetilde{W}^{SC} - \alpha_{0t}) - \overline{W}' y_t^N + y_{0t}^N - \gamma_t \\
&= -(y_{0t}^N - y_t' \widetilde{W}^{SC}) - \overline{W}' y_t^N + y_{0t}^N - \gamma_t \\
&= (\widetilde{W}^{SC} - \overline{W})' y_t^N - \gamma_t \\
&= (\widetilde{W}^{SC} - \overline{W})' c + (\widetilde{W}^{SC} - \overline{W})' \mu \lambda_t + (\widehat{W}^{SC} - \overline{W})' \varepsilon_t - \gamma_t
\end{aligned}$$

where $c := (c_1, ..., c_J)'$, $\iota \in \mathbb{R}^J$ is the vector of ones, and $(\widetilde{W}^{SC} - \overline{W}) \iota \delta_t = 0$ since $\widetilde{W}^{SC} \iota = 1 = \overline{W} \iota$, as $\widehat{W}^{SC}, \overline{W} \in \Delta_\eta^J$. Therefore, by noting the simple inequality of $(a+b+c+d)^2 \leq 8a^2 + 8b^2 + 8c^2$

$+\ 8d^2$,

$$||\widehat{P}^N - P^N||_2^2/T = \frac{1}{T}\sum_{t=1}^{T}(\widehat{P}_t^N - P_t^N)^2$$
$$\leq 8\left\{(\widetilde{W}^{SC} - \overline{W})'c\right\}^2 + 8||(\widetilde{W}^{SC} - \overline{W})'\mu||^2\left\{\frac{1}{T}\sum_{t=1}^{T}||\lambda_t||^2\right\}$$
$$+ 8||(\widetilde{W}^{SC} - \overline{W})'\mu||^2\left\{\frac{1}{T}\sum_{t=1}^{T}||\varepsilon_t||^2\right\} + 8\frac{1}{T}\sum_{t=1}^{T}\gamma_t^2 = o_p(1)$$

under our assumptions. This satisfies assumption 3.1 of Chernozhukov et al. (2021). Furthermore, for any $t \in \mathcal{T}_1$,

$$|\widehat{P}_t^N - P_t^N| = o_p(1) \quad as \quad T_0 \to \infty$$

since $\mathbb{E}||\lambda_t||_2, \mathbb{E}||\varepsilon_t||_2 < \sigma^2 < \infty$ by assumption, Markov inequality, and $\widetilde{W}^{SC} \xrightarrow{p} \overline{W}$. This satisfies assumption 3.2 of Chernozhukov et al. (2021), so that an application of Theorem 1 of Chernozhukov et al. (2021) yields the result.

C Proof of Theorem 2

Step 1: We show that for any $W \in \Delta_\eta^J$,

$$\frac{1}{r\Delta^2}\sum_{s=1}^{r}\left\{\sum_{q=1}^{\Delta}(\varepsilon_{0q}^s - (\varepsilon_q^s)'W)\right\}^2 = o_p(1) \tag{C.1}$$

Fix any $W \in \Delta_\eta^J$ and observe

$$\frac{1}{r\Delta^2}\sum_{s=1}^{r}\left\{\sum_{q=1}^{\Delta}(\varepsilon_{0q}^s - (\varepsilon_q^s)'W)\right\}^2$$
$$= \frac{1}{r\Delta^2}\sum_{s=1}^{r}\sum_{q=1}^{\Delta}(\varepsilon_{0q}^s - (\varepsilon_q^s)'W)^2 + 2\frac{1}{r\Delta^2}\sum_{s=1}^{r}\sum_{\ell=1}^{\Delta}\sum_{q=1}^{\ell-1}(\varepsilon_{0\ell}^s - (\varepsilon_\ell^s)'W)(\varepsilon_{0q}^s - (\varepsilon_q^s)'W)$$
$$= \frac{1}{T_0\Delta}\sum_{t=1}^{T_0}(\varepsilon_{0t} - \varepsilon_t'W)^2 + +2\frac{1}{r\Delta^2}\sum_{s=1}^{r}\sum_{\ell=1}^{\Delta}\sum_{q=1}^{\ell-1}(\varepsilon_{0\ell}^s - (\varepsilon_\ell^s)'W)(\varepsilon_{0q}^s - (\varepsilon_q^s)'W)$$
$$\equiv A_1 + A_2.$$

We will show that $A_1, A_2 = o_p(1)$. Noting the simple inequality of $(a+b)^2 \leq 2a^2 + 2b^2$,

$$A_1 \leq \frac{1}{T_0\Delta}\sum_{t=1}^{T_0}\varepsilon_{0t}^2 + \frac{1}{\Delta}W'(\frac{1}{T_0}\sum_{t=1}^{T_0}\varepsilon_t\varepsilon_t')W$$

$$= \frac{1}{\Delta}(\sigma_\varepsilon^2 + o_p(1)) + \frac{1}{\Delta}W'(\sigma_\varepsilon^2 I_J + o_p(1))W = o_p(1)$$

as $\Delta \to \infty$. To deal with A_2, for notational simplicity, define $X_{0,\ell}^s := \sum_{q=1}^{\ell-1}(\varepsilon_{0\ell}^s - (\varepsilon_\ell^s)'W)(\varepsilon_{0q}^s - (\varepsilon_q^s)'W)$. Then we have

$$\begin{aligned}
var(A_2) &= \frac{4}{r^2\Delta^4} var(\sum_{s=1}^r \sum_{\ell=1}^\Delta X_{0,\ell}^s) \overset{(i)}{=} \frac{4}{r^2\Delta^4}\sum_{s=1}^r var(\sum_{\ell=1}^\Delta X_{0,\ell}^s) = \frac{4}{r^2\Delta^4}\sum_{s=1}^r \mathbb{E}\left\{\sum_{\ell=1}^\Delta X_{0,\ell}^s l\right\}^2 \\
&= \frac{4}{r^2\Delta^4}\sum_{s=1}^r\sum_{\ell=1}^\Delta\sum_{m=1}^\Delta \mathbb{E}(X_{0,\ell}^s X_{0,m}^s) \overset{(ii)}{=} \frac{4}{r^2\Delta^4}\sum_{s=1}^r\sum_{\ell=1}^\Delta \mathbb{E}(X_{0,\ell}^s)^2 \\
&\overset{(iii)}{\le} \frac{8\sigma^2}{r^2\Delta^4}\sum_{s=1}^r\sum_{\ell=1}^\Delta(\ell-1) = \frac{8\sigma^2}{T_0\Delta^2}\sum_{\ell=1}^\Delta(\ell-1) \le \frac{8\sigma^2}{T_0} = o(1)
\end{aligned}$$

where (i) follows from independence between blocks and $\mathbb{E}\left(\sum_{\ell=1}^\Delta X_{0,\ell}^s\right) = 0$; (ii) follows from the observation that, for any $\ell \neq m$ (we can w.l.o.g. assume $\ell < m$),

$$\begin{aligned}
\mathbb{E}(X_{0,\ell}^s X_{0,m}^s) &= \mathbb{E}\left(\sum_{q=1}^{\ell-1}\sum_{h=1}^{m-1}\{\varepsilon_{0\ell}^s - (\varepsilon_\ell^s)'W\}\{\varepsilon_{0q}^s - (\varepsilon_q^s)'W\}\{\varepsilon_{0m}^s - (\varepsilon_m^s)'W\}\{\varepsilon_{0h}^s - (\varepsilon_h^s)'W\}\right) \\
&= \sum_{q=1}^{\ell-1}\sum_{h=1}^{m-1}\mathbb{E}\left(\varepsilon_{0,m}^s - (\varepsilon_m^s)'W\right)\cdot\mathbb{E}\left(\{\varepsilon_{0\ell}^s - (\varepsilon_\ell^s)'W\}\{\varepsilon_{0q}^s - (\varepsilon_q^s)'W\}\{\varepsilon_{0h}^s - (\varepsilon_h^s)'W\}\right) = 0;
\end{aligned}$$

(iii) follows from

$$\begin{aligned}
\mathbb{E}(X_{0,\ell}^s)^2 &= \sum_{q=1}^{\ell-1}\sum_{h=1}^{\ell-1}\mathbb{E}\left(\varepsilon_{0\ell}^s - (\varepsilon_\ell^s)'W\right)^2\cdot\mathbb{E}\left((\varepsilon_{0q}^s - (\varepsilon_q^s)'W)\cdot(\varepsilon_{0h}^s - (\varepsilon_h^s)'W)\right) \\
&= \sum_{q=1}^{\ell-1}\mathbb{E}\left(\varepsilon_{0\ell}^s - (\varepsilon_\ell^s)'W\right)^2\cdot\mathbb{E}\left(\varepsilon_{0q}^s - (\varepsilon_q^s)'W\right)^2 \overset{(iv)}{\le} 2(\ell-1)\sigma^2
\end{aligned}$$

where (iv) follows from

$$\mathbb{E}\left(\varepsilon_{0q}^s - (\varepsilon_q^s)'W\right)^2 = \mathbb{E}(\varepsilon_{0q}^s)^2 + \sum_{j=1}^J W_j^2\mathbb{E}(\varepsilon_{i,q}^s)^2 \le \sigma^2 + \sigma^2\sum_{j=1}^J W_j = 2\sigma^2$$

so $A_2 = o_p(1)$ by Markov-inequality and the fact that $\mathbb{E}A_2 = 0$. Therefore (C.1) is shown.

step 2: Define

$$\widetilde{\mathcal{A}}_{T_0}(W,\Lambda) := \frac{1}{r}\sum_{s=1}^r\{\overline{y}_0^s - (\overline{y}^s)'W\}^2 + \Lambda||W||^2$$

and

$$\mathcal{A}(W,\Lambda) := (c_0 - c'W)^2 + (\mu_0 - \mu W)'\Omega_0(\mu_0 - \mu W) + \Lambda||W||^2,$$

we want to show that

$$\sup_{(W,\Lambda)\in\Delta_\eta^J\times[0,1]} \left|\widetilde{\mathcal{A}}_{T_0}(W,\Lambda) - \mathcal{A}(W,\Lambda)\right| = o_p(1) \tag{C.2}$$

First we require a lemma:

Lemma C.1. *(Corollary 2.2 of Newey (1991)) Assume (1) $\Delta_\eta^J\times[0,1]$ is compact, (2) $\widetilde{\mathcal{A}}_{T_0}(W,\Lambda) \xrightarrow{p} \mathcal{A}(W,\Lambda)$ for every $(W,\Lambda) \in \Delta_\eta^J \times [0,1]$, (3) $\Delta_\eta^J \times [0,1]$ is a metric space and (4) there is a B_{T_0} such that $B_{T_0} = O_p(1)$ and for all $(W_1,\Lambda_1),(W_2,\Lambda_2) \in \Delta_\eta^J$, $|\widetilde{\mathcal{A}}_{T_0}(W_1,\Lambda_1) - \widetilde{\mathcal{A}}_{T_0}(W_2,\Lambda_2)| \leq B_{T_0}||(W_1,\Lambda_1) - (W_2,\Lambda_2)||$ and (5) $\{\mathcal{A}(W,\Lambda)\}_{(W,\Lambda)\in\Delta_\eta^J\times[0,1]}$ is equi-continuous. Then $\widetilde{\mathcal{A}}_{T_0}(W,\Lambda) \xrightarrow{p} \mathcal{A}(W,\Lambda)$ uniformly over $(W,\Lambda) \in \Delta_\eta^J \times [0,1]$*

Fixing any $(W,\Lambda) \in \Delta_\eta^J \times [0,1]$, we have

$$\begin{aligned}
\widetilde{\mathcal{A}}_{T_0}(W,\lambda) &= \frac{1}{r}\sum_{s=1}^{r}\Big[(c_0 - c'W) + (\overline{\lambda}^s)'(\mu_0 - \mu'W) + (\overline{\varepsilon}_0^s - (\overline{\varepsilon}^s)'W)\Big]^2 + \Lambda||W||^2 \\
&= (c_0 - c'W)^2 + (\mu_0 - \mu'W)'\left(\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\lambda}^s)'\right)(\mu_0 - \mu'W) + \frac{1}{r}\sum_{s=1}^{r}(\overline{\varepsilon}_0^s - (\overline{\varepsilon}^s)'W)^2 \\
&+ 2(c_0 - c'W)\left(\frac{1}{r}\sum_{s=1}^{r}(\overline{\lambda^s})'\right)(\mu_0 - \mu'W) + 2(c_0 - c'W)\left(\frac{1}{r}\sum_{s=1}^{r}(\overline{\varepsilon}_0^s - (\overline{\varepsilon}^s)'W)\right) \\
&+ 2(\mu_0 - \mu'W)'\left(\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\varepsilon}_0^s - (\overline{\varepsilon}^s)'W)\right) + \Lambda||W||^2,
\end{aligned}$$

so that by

$$(a) \quad \frac{1}{r}\sum_{s=1}^{r}(\overline{\varepsilon}_0^s - (\overline{\varepsilon}^s)'W)^2 = \frac{1}{r\Delta^2}\sum_{s=1}^{r}\left\{\sum_{q=1}^{\Delta}(\varepsilon_{0q}^s - (\varepsilon_q^s)'W)\right\}^2 = o_p(1) \quad \text{by (C.1)}$$

$$(b) \quad \frac{1}{r}\sum_{s=1}^{r}(\overline{\lambda^s})' = \frac{1}{T_0}\sum_{t=1}^{T_0}(\lambda_t)' = o_p(1) \quad \text{by assumption}$$

$$(c) \quad \frac{1}{r}\sum_{s=1}^{r}(\overline{\varepsilon}_0^s - (\overline{\varepsilon}^s)'W) = \frac{1}{T_0}\sum_{t=1}^{T_0}\varepsilon_{0t} - \frac{1}{T_0}\sum_{t=1}^{T_0}\varepsilon_t'W = o_p(1) \quad \text{by assumption}$$

$$(d) \quad \frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\varepsilon}_0^s - (\overline{\varepsilon}^s)'W) = \frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s\overline{\varepsilon}_0^s - \frac{1}{T_0}\sum_{t=1}^{T_0}(\varepsilon_t)'W = o_p(1)$$

where the last equality in (d) follows from Markov-inequality, the simple inequality that $\mathbb{E}(\lambda_k^s\lambda_{j,K'}^s) \leq$

$2\mathbb{E}(\lambda^s_{j,k})^2 + 2\mathbb{E}(\lambda^s_{k'})^2 \leq 4C$ by assumption, and

$$\begin{aligned}
\mathbb{E}\left|\left|\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s\overline{\varepsilon}_0^s\right|\right|_2^2 &= \sum_{j=1}^{J}\left(\frac{1}{r^2}\sum_{s=1}^{r}\sum_{m=1}^{r}\mathbb{E}(\overline{\lambda}_j^s\overline{\lambda}_j^m\overline{\varepsilon}_0^s\overline{\varepsilon}_0^m)\right) = \sum_{j=1}^{J}\left(\frac{1}{r^2}\sum_{s=1}^{r}\mathbb{E}(\overline{\lambda}_j^s)^2\mathbb{E}(\overline{\varepsilon}_0^s)^2\right) \\
&= \sum_{j=1}^{J}\left(\frac{1}{r^2}\sum_{s=1}^{r}\mathbb{E}(\frac{1}{\Delta}\sum_{k=1}^{\Delta}\lambda_{j,k}^s)^2\mathbb{E}(\frac{1}{\Delta}\sum_{k=1}^{\Delta}\varepsilon_{0,k}^s)^2\right) \\
&\leq \sum_{j=1}^{J}\left(\frac{1}{r^2}\sum_{s=1}^{r}\frac{1}{\Delta^2}(\sum_{k=1}^{\Delta}\sum_{k'=1}^{\Delta}4C)\cdot\frac{1}{\Delta^2}\sum_{k=1}^{\Delta}\mathbb{E}(\varepsilon_{0,k}^s)^2\right) \\
&\leq \frac{4C^2J}{r^2\Delta} = \frac{4C^2J}{T_0 r} \to 0
\end{aligned}$$

it follows by our assumption that

$$\widetilde{\mathcal{A}}_{T_0}(W,\Lambda) \xrightarrow{p} \mathcal{A}(W,\Lambda) \tag{C.3}$$

It is clear that $\Delta_\eta^J \times [0,1]$ is compact, so condition (1) of Lemma C.1 is satisfied. It is clear that condition (3) and (5) also holds, i.e. $\{\mathcal{A}(W,\Lambda)\}_{(W,\Lambda)\in\Delta_\eta^J\times[0,1]}$ is equi-continuous. Condition (2) follows from (C.3). To show (C.2), it remains to prove that condition (4) of Lemma C.1 holds, which is what we do now.

We can remove the common time-effect from $\widetilde{\mathcal{A}}_{T_0}(W)$ by defining $\widetilde{y}_0^s := \overline{y}_0^s - \overline{\delta}^s$ and $\widetilde{y}^s := \overline{y}^s - \iota\overline{\delta}^s$, so that

$$\widetilde{\mathcal{A}}_{T_0}(W,\Lambda) = \frac{1}{r}\sum_{s=1}^{r}\{\widetilde{y}_0^s - (\widetilde{y}^s)'W + \overline{\delta}^s(1-\iota'W)\}^2 + \Lambda||W||^2$$

Then using mean value theorem, for any $(W_1,\Lambda_1),(W_2,\Lambda_2) \in \Delta_\eta^J \times [0,1]$, there exists a $(W_3,\Lambda_3) \in \Delta_\eta^J \times [0,1]$ such that

$$\begin{aligned}
&\left|\widetilde{\mathcal{A}}_{T_0}(W_1,\Lambda_1) - \widetilde{\mathcal{A}}_{T_0}(W_2,\Lambda_2)\right| \\
&= \left|\left(\frac{2}{r}\sum_{s=1}^{r}\{\widetilde{y}_0^s - (\widetilde{y}^s)'W_3\}(-\widetilde{y}^s - \overline{\delta}^s\iota) + ||W_3||^2 + 2\Lambda_3 W_3\right)\cdot||(W_1,\Lambda_1)-(W_2,\Lambda_2)||\right| \\
&= B_{T_0}||(W_1,\Lambda_1)-(W_2,\Lambda_2)||
\end{aligned}$$

with

$$\begin{aligned}
B_{T_0} &:= \left|\left|\left(\frac{2}{r}\sum_{s=1}^{r}\{\widetilde{y}_0^s - (\widetilde{y}^s)'W_3\}(-\widetilde{y}^s - \overline{\delta}^s\iota) + ||W_3||^2 + 2\Lambda_3 W_3\right)\right|\right| \\
&\leq \left|\left|\left(\frac{2}{r}\sum_{s=1}^{r}\{\widetilde{y}_0^s - (\widetilde{y}^s)'W_3\}(\widetilde{y}^s)\right)\right|\right| + \left|\left|\left(\frac{2}{r}\sum_{s=1}^{r}\{\widetilde{y}_0^s - (\widetilde{y}^s)'W_3\}\overline{\delta}^s\right)\right|\right|\cdot||\iota|| + ||W_3||^2 + 2\Lambda_3||W_3||
\end{aligned}$$

$$\leq \left|\left|\frac{2}{r}\sum_{s=1}^{r}\widetilde{y}_0^s\widetilde{y}^s\right|\right| + \left|\left|\frac{2}{r}\sum_{s=1}^{r}\widetilde{y}^s(\widetilde{y}^s)'\right|\right| \times ||W_3|| + \left|\left|\frac{2}{r}\sum_{s=1}^{r}\widetilde{y}_0^s\overline{\delta}^s\right|\right| + \left|\left|\frac{2}{r}\sum_{s=1}^{r}\widetilde{y}^s\overline{\delta}^s\right|\right| \times ||W_3|| \times ||\iota|| + \eta^2 + 2\eta$$
$$= ||2A_1|| + ||2A_2|| \times \eta + ||2A_3|| + ||2A_4|| \times \sqrt{J}\eta + \eta^2 + 2\eta$$

where the second last inequality follows from $||W_3|| \leq \eta$ and $\Lambda \leq 1$. We will show that $B_{T_0} = O_p(1)$ by showing that each term $A_1, ..., A_4$ is $O_p(1)$. Observe first that

$$(a)\quad \frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\lambda}^s)' = O_p(1)$$
$$(b)\quad \frac{1}{r}\sum_{s=1}^{r}\overline{\varepsilon}^s(\overline{\varepsilon}^s)' = O_p(1)$$
$$(c)\quad \frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s = \frac{1}{T_0}\sum_{t=1}^{T_0}\lambda_t = o_p(1)$$
$$(d)\quad \frac{1}{r}\sum_{s=1}^{r}(\overline{\varepsilon}^s)' = \frac{1}{T_0}\sum_{t=1}^{T_0}(\varepsilon_t)' = o_p(1)$$
$$(e)\quad \frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\varepsilon}^s)' = O_p(1)$$
$$(f)\quad \frac{1}{r}\sum_{s=1}^{r}\overline{\varepsilon}_0^s(\overline{\varepsilon}^s)' = O_p(1)$$
$$(g)\quad \frac{1}{r}\sum_{s=1}^{r}\overline{\delta}^s = \frac{1}{T_0}\sum_{t=1}^{T_0}\delta_t = O_p(1)$$
$$(h)\quad \frac{1}{r}\sum_{s=1}^{r}\overline{\delta}^s(\overline{\lambda}^s)' = O_p(1)$$
$$(i)\quad \frac{1}{r}\sum_{s=1}^{r}\overline{\delta}^s(\overline{\varepsilon}^s)' = O_p(1)$$

where (c) and (d) follows from the assumptions, (a) follows from Markov-inequality and

$$\mathbb{E}\left|\left|\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\lambda}^s)'\right|\right| \leq \frac{1}{r\Delta^2}\sum_{s=1}^{r}\mathbb{E}\left|\left|\sum_{m=1}^{\Delta}\lambda_m^s\right|\right|^2 \leq \frac{1}{r\Delta^2}\sum_{s=1}^{r}\sum_{m=1}^{\Delta}\sum_{\ell=1}^{\Delta}\mathbb{E}\left(||\lambda_m^s||\cdot||\lambda_\ell^s||\right)$$
$$\leq \frac{2}{r\Delta^2}\sum_{s=1}^{r}\sum_{m=1}^{\Delta}\sum_{\ell=1}^{\Delta}\left(\mathbb{E}||\lambda_m^s||^2 + \mathbb{E}||\lambda_\ell^s||^2\right) \leq 2\sigma^2,$$

(b) follows in an analogous manner, (e) follows from Markov inequality and

$$\mathbb{E}\left|\left|\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\varepsilon}^s)'\right|\right| \leq \frac{1}{r}\sum_{s=1}^{r}\mathbb{E}\left(||\frac{1}{\Delta}\sum_{m=1}^{\Delta}\lambda_m^s||\cdot||\frac{1}{\Delta}\sum_{m=1}^{\Delta}\varepsilon_m^s||\right)$$

$$\leq \frac{1}{r\Delta^2}\sum_{s=1}^{r}\sum_{m=1}^{\Delta}\sum_{\ell=1}^{\Delta}\mathbb{E}\left(||\lambda_m^s||\cdot||\varepsilon_m^s||\right) \leq 2\sigma^2,$$

(f) follows from Markov inequality and

$$\mathbb{E}\left|\left|\frac{1}{r}\sum_{s=1}^{r}\overline{\varepsilon}_0^s(\overline{\varepsilon}^s)'\right|\right| \leq \frac{1}{r}\sum_{s=1}^{r}\mathbb{E}\left(||\frac{1}{\Delta}\sum_{m=1}^{\Delta}\varepsilon_{0,m}^s||\cdot||\frac{1}{\Delta}\sum_{m=1}^{\Delta}\varepsilon_m^s||\right)$$

$$\leq \frac{1}{r\Delta^2}\sum_{s=1}^{r}\sum_{m=1}^{\Delta}\sum_{\ell=1}^{\Delta}\mathbb{E}\left(||\varepsilon_{0,m}^s||\cdot||\varepsilon_m^s||\right) \leq 2\sigma^2,$$

(g) follows from Markov inequality and bounded second moment of δ_t, both (h) and (i) follows in the same way as (e).

We can show $A_1, ..., A_4 = O_p(1)$ by writing

$$\begin{aligned}
A_1 &= \frac{1}{r}\sum_{s=1}^{r}(c_0 + \mu_0'\overline{\lambda}^s + \overline{\varepsilon}_0^s)(c + \mu'\overline{\lambda}^s + \overline{\varepsilon}^s)' \\
&= c_oc + c_0\mu\cdot\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s + c_0(\frac{1}{r}\sum_{s=1}^{r}\overline{\varepsilon}^s) + c\mu_0'(\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s) + \mu_0'(\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\lambda}^s)')\mu \\
&+ \mu_0'(\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\varepsilon}^s)') + (\frac{1}{r}\sum_{s=1}^{r}\overline{\varepsilon}_0^s)c' + (\frac{1}{r}\sum_{s=1}^{r}\overline{\varepsilon}_0^s(\overline{\lambda}^s)')\mu + \frac{1}{r}\sum_{s=1}^{r}\overline{\varepsilon}_0^s(\overline{\varepsilon}^s)', \\
A_2 &= \frac{1}{r}\sum_{s=1}^{r}(c + \mu'\overline{\lambda}^s + \overline{\varepsilon}^s)\cdot(c + \mu'\overline{\lambda}^s + \overline{\varepsilon}^s)' \\
&= cc' + \mu'\left(\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\lambda}^s)\right)\mu + \frac{1}{r}\sum_{s=1}^{r}\overline{\varepsilon}^s(\overline{\varepsilon}^s)' \\
&+ 2c\left(\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s\right)\mu + 2c\frac{1}{r}\sum_{s=1}^{r}(\overline{\varepsilon}^s)' + 2\mu'\left(\frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\varepsilon}^s)'\right), \\
A_3 &= \frac{1}{r}\sum_{s=1}^{r}\overline{\delta}^s(c + \mu'\overline{\lambda}^s + \overline{\varepsilon}^s)' = (\frac{1}{r}\sum_{s=1}^{r}\overline{\delta}^s)c' + (\frac{1}{r}\sum_{s=1}^{r}\overline{\delta}^s(\overline{\lambda}^s)')\mu + \frac{1}{r}\sum_{s=1}^{r}\overline{\delta}^s(\overline{\varepsilon})' \\
A_4 &= \frac{1}{r}\sum_{s=1}^{r}(c + \mu'\overline{\lambda}^s + \overline{\varepsilon}^s)\overline{\delta}^s = c(\sum_{s=1}^{r}\overline{\delta}^s) + \mu'(\sum_{s=1}^{r}\overline{\lambda}^s\overline{\delta}^s) + \sum_{s=1}^{r}\overline{\varepsilon}^s\overline{\delta}^s
\end{aligned}$$

and then applying $(a) - (i)$. Therefore condition (4) of Lemma C.1 is shown, implying (C.2) and (**??**)

Step 3: We show that for any fixed $\gamma > 0$,

$$\sup_{\Lambda \in [\gamma,1]} \left| \widetilde{W}^{SC}(\Lambda) - \overline{W}(\Lambda) \right| = o_p(1), \tag{C.4}$$

where $\overline{W}(\Lambda) = \arg\min_{W \in \Delta^J} \mathcal{A}^{(i)}(W, \Lambda)$ in case (i), or $\overline{W} = \arg\min_{W \in \Delta^J} \mathcal{A}^{(ii)}(W, \Lambda)$ in case (ii).

Lemma C.2. *(Newey and McFadden (1994)[Theorem 2.1]) Suppose there is a function $Q_0(\theta)$ such that it is (i) uniquely minimized at θ_0; (ii) Θ is compact, where $\theta \in \Theta$; (iii) $Q_0(\theta)$ is continuous and (iv) $\sup_{\theta \in \Theta} |\widehat{Q}(\theta) - Q_0(\theta)| = o_p(1)$. Then for $\widehat{\theta} := \arg\min \widehat{Q}(\theta)$, we have $\widehat{\theta} \xrightarrow{p} \theta_0$*

We first show point-wise convergence of $\widetilde{W}^{SC}(\Lambda)$ for every $\Lambda \in (0, 1]$. Replace Θ by Δ_η^J, which is compact. We fix any Λ and replace $Q_0(\theta)$ by $\mathcal{A}(\Lambda, W)$. Since $\mathcal{A}(\Lambda, W)$ is strictly convex, it has a uniquely-minimized solution. Clearly $\mathcal{A}(W, \Lambda)$ is continuous in Θ and condition (iv) of Lemma C.2 follows from equation (C.2), with $\widehat{Q}(\theta)$ as $\widetilde{\mathcal{A}}_{T_0}(W, \Lambda)$. Therefore we have

$$\widetilde{W}^{SC}(\Lambda) \xrightarrow{p} \overline{W}(\Lambda) \tag{C.5}$$

for every fixed $\Lambda \in (0, 1]$. This satisfies condition (2) of Lemma C.1. Since $[\gamma, 1]$ is compact, condition (1) is satisfied. Condition (3) is clear. To show that $\overline{W}(\Lambda)_{\Lambda \in [\gamma,1]}$ is equi-continuous, first observe that

$$\overline{W}(\Lambda) = (\Lambda I + cc' + \mu'\Omega_0\mu)^{-1}(\mu'\Omega_0\mu_0 + c_0 c)$$

We can take the spectral decomposition of $cc' + \mu'\Omega_0\mu = VDV'$, where $VV' = I = V'V$ and D is the diagonal matrix with non-negative eigenvalues $(d_1, ..., d_J)$ as its elements. Define $D^\Lambda = D + \Lambda I$. Then for any $\Lambda_1, \Lambda_2 \in [\gamma, 1]$,

$$\begin{aligned}
\left|\left|\overline{W}(\Lambda_1) - \overline{W}(\Lambda_2)\right|\right| &\leq ||(\Lambda_1 I + cc' + \mu'\Omega_0\mu)^{-1} - (\Lambda_2 I + cc' + \mu'\Omega_0\mu)^{-1}||_\infty \cdot ||\mu'\Omega_0\mu_0 + c_0 c||_1 \\
&= ||V(D^{\Lambda_1})^{-1} - (D^{\Lambda_1})^{-1})V'||_\infty \cdot ||\mu'\Omega\mu_0 + c_0 c||_1 \\
&\overset{(i)}{=} ||(D^{\Lambda_1})^{-1} - (D^{\Lambda_1})^{-1}||_\infty \cdot ||\mu'\Omega_0\mu_0 + c_0 c||_1 \\
&= ||\mu'\Omega_0\mu_0 + c_0 c||_1 \max_{i=1,\dots,J} \frac{|\Lambda_1 - \Lambda_2|}{(d_i + \Lambda_1)(d_i + \Lambda_2)} \\
&\leq \frac{||\mu'\Omega_0\mu_0 + c_0 c||_1}{\gamma^2} |\Lambda_1 - \Lambda_2|
\end{aligned}$$

where (i) follows from V being orthogonal. Condition (5) of Lemma C.1 is shown. For any $\Lambda_1, \Lambda_2 \in [\gamma, 1]$, we can diagonalize $\frac{1}{r}\sum_{s=1}^r \overline{y}^s(\overline{y}^s)' = V_T D_T V_T'$ and define $D_T^\Lambda := D_T + \Lambda I$, so that

$$\begin{aligned}
\left|\left|\widetilde{W}^{SC}(\Lambda_1) - \widetilde{W}^{SC}(\Lambda_2)\right|\right| &\leq ||\frac{1}{r}\sum_{s=1}^r \overline{y}_0^s \cdot \overline{y}^s||_1 \cdot \max_{i=1,\dots,J} \frac{|\Lambda_1 - \Lambda_2|}{(d_{i,T} + \Lambda_1)(d_{i,T} + \Lambda_2)} \\
&\leq \frac{||\frac{1}{r}\sum_{s=1}^r \overline{y}_0^s \cdot \overline{y}^s||_1}{\gamma^2} \cdot |\Lambda_1 - \Lambda_2| =: B_{T_0} \cdot |\Lambda_1 - \Lambda_2|
\end{aligned}$$

where

$$\gamma^2 \cdot \mathbb{E}(B_{T_0}) \leq \frac{1}{r}\sum_{j=1}^{J}\sum_{s=1}^{r}\mathbb{E}(|\overline{y}_0|\cdot|\overline{y}_j^s|)$$
$$\leq \frac{1}{r}\frac{1}{\Delta^2}\sum_{j=1}^{J}\sum_{s=1}^{r}\mathbb{E}\left(\sum_{\ell=1}^{\Delta}(|c_0|+|\delta_{s,\ell}|+|\lambda_{s,\ell}'\mu_0|+|\varepsilon_{0,s\Delta+\ell}|)\right)\cdot\left(\sum_{\ell=1}^{\Delta}(|c_j|+|\delta_{s,\ell}|+|\lambda_{s,\ell}'\mu_j|+|\varepsilon_{j,s\Delta+\ell}|)\right)$$
$$\leq \frac{1}{r}\sum_{j=1}^{J}\sum_{s=1}^{r}(1+|c_0|+|c_j|+||\mu_0||+||\mu_j||)\sigma^2 = J(1+|c_0|+|c_j|+||\mu_0||+||\mu_j||)\sigma^2 = O(1)$$

where the last inequality follows from the bounded second moments of δ_t, λ_t and ε_{jt} by assumption. By Markov-inequality, condition (4) of Lemma C.1 is satisfied, so that we obtain (C.4).

Step 4: We show that $\overline{W}(\Lambda) \to W^*$ as $0 < \Lambda \downarrow 0$

For any $\Lambda > 0$, $\mathcal{A}(W,\Lambda)$ is a strictly convex and continuous function, so $\overline{W}(\Lambda) := \arg\min \mathcal{A}(W,\Lambda)$ is unique. Define W^* as the minimum-norm least square solution that minimizes $\mathcal{H}(W) := (c_0 - c'W)^2 + (\mu_0 - \mu W)'\Omega_0(\mu_0 - \mu W)$ over $W \in \Delta_\eta^J$, i.e. $\mathcal{H}(W^*) = 0$ such that for any other $W \in \Delta_\eta^J$ with $\mathcal{H}(W) = 0$, $||W^*|| < ||W||$; this uniqueness follows from Lemma A.1. Note that $\mathcal{H}(W^*) = 0$ by assumption 4.

For any $W^\dagger \in \Delta_\eta^J$ with $||W^\dagger|| > ||W^*||$, we will have $\mathcal{A}(W^*,\Lambda) < \mathcal{A}(W^\dagger,\Lambda)$. Therefore we have that $W \neq W^*$, if $||\overline{W}(\Lambda)|| \leq ||W^*||$. Furthermore, we know that any $||W|| \leq ||W^*||$ has the property that $\mathcal{H}(W) > 0$, since W^* is the minimum-norm solution. Define $\Delta(W) := W - W^*$ and consider any fixed $\delta > 0$. Consider the open ball around W^*, defined as $B_\delta(W^*) \equiv \{W : ||\Delta(W)|| < \delta\}$. Since $\widetilde{\Delta} := \{W : ||W|| \leq ||W^*||\}$ is compact, then $\widetilde{\Delta} \cap B_\delta^c(W^*)$ is compact. By Weierstrass extreme-value-theorem, there exists a $W^\ddagger \in \widetilde{\Delta} \cap B_\delta^c(W^*)$ with $0 < \mathcal{H}(W^\ddagger) \equiv \inf_{W \in \widetilde{\Delta} \cap B_\delta^c(W^*)} \mathcal{H}(W)$. There must be a $\overline{c}(\delta) > 0$ such whenever $0 < \Lambda < \overline{c}(\delta)$, then $\Lambda\eta < \mathcal{H}(W^\ddagger)$. Then we can see that

$$\overline{W}^\Lambda \in \widetilde{\Delta} \cap B_\delta(W^*), \tag{C.6}$$

because of

$$\mathcal{A}(W^*,\Lambda) = \Lambda||W^*||^2 \leq \Lambda\eta < \mathcal{H}(W^\ddagger) \leq \mathcal{A}(W,\Lambda)$$

for any $W \in \widetilde{\Delta} \cap B_\delta^c(W^*)$, and the fact that any $W \in \Delta_\eta^J \backslash \{\widetilde{\Delta}\}$ cannot minimize $\mathcal{A}(W,\Lambda)$, i.e. $W \neq \overline{W}(\Lambda)$. We can expect that $\overline{c}(\delta) \downarrow 0$ since $\inf_{W \in \widetilde{\Delta} \cap B_\delta^c(W^*)} \mathcal{H}(W)$ is non-decreasing with $\delta \downarrow 0$. This implies that as $0 < \Lambda \downarrow 0$, $\overline{W}(\Lambda) \to W^*$.

D proof of corollary 3.2

For any given $\xi > 0$, we show that there exists a $\Lambda(\xi) > 0$ such that for any fixed $0 < \Lambda \leq \Lambda(\xi)$,

$$|\widetilde{W}^{SC}(\Lambda) - W^*| \leq \xi + o_p(1)$$

By Theorem 2, there exists a $\Lambda(\xi) > 0$ such that for any $0 < \Lambda \leq \Lambda(\xi)$, we have $|\overline{W}(\Lambda) - W^*| \leq \xi$. Define $\gamma := \Lambda$. Then by triangle inequality,

$$|\widetilde{W}^{SC}(\Lambda) - W^*| \leq |\widetilde{W}^{SC}(\Lambda) - \overline{W}(\Lambda)| + |\overline{W}(\Lambda) - W^*| \leq o_p(1) + \xi$$

so that the result is shown.

E Proof of corollary 3.3

Consider any positive decreasing sequence $(\xi_m)_{m=1}^{\infty}$ that converges to 0. By Theorem 2, for ξ_1, there is some $m_0(\xi_1) \in \mathbb{N}$ and $\Lambda(\xi_1) > 0$ such that for any $T \geq m_0(\xi_1)$,

$$|\widetilde{W}_T^{SC}(\Lambda(\xi_1)) - W^*| \leq |\widetilde{W}_T^{SC}(\Lambda(\xi_1)) - \overline{W}(\Lambda(\xi_1))| + |\overline{W}(\Lambda(\xi_1)) - W^*| \leq \xi_1 + \xi_1 = 2\xi_1$$

Moving to ξ_2, there exists $m_0(\xi_2) > m_0(\xi_1)$ and $\Lambda(\xi_2) > 0$ such that for any $T \geq m_0(\xi_2)$,

$$|\widetilde{W}_T^{SC}(\Lambda(\xi_1)) - W^*| \leq 2\xi_2$$

We can express this recursively, and by taking $\Lambda_T \equiv \Lambda(\xi_1)$ for $T = 1, ..., m_0(\xi_2)$, $\Lambda_T \equiv \Lambda(\xi_2)$ for $T = m_0(\xi_2) + 1, ..., m_0(\xi_3)$, so on and so forth. Then we see that the result holds.

F Proof of Theorem 3

The proof follows from an application of Chernozhukov et al. (2021)[Lemma H.1-H.5]. We include the proof for completeness. The first three lemmas are given to make the exposition self-contained. We write $n \equiv T!$ under moving permutation.

Lemma F.1 (Lemma H.5 Chernozhukov et al. (2021))**.** *Consider moving block permutations Π with T_1 fixed. Suppose that for some $Q > 0$, $|S(u) - S(v)| \leq Q||D_{T_1}(u - v)||$ for any $u, v \in \mathbb{R}^T$ and $D_{T_1} := Blockdiag(0_{T_0}, I_{T_1})$. If $||\widehat{P}^N(\Lambda) - P^N||/\sqrt{T} \leq \delta_T$ and $|\widehat{P}_t^N - P_t^N| \leq \delta_T$ for $t \in \mathcal{T}_0$ with probability at least $1 - \gamma_T$, $\gamma_T = o(1)$, then with probability at least $1 - \gamma_T$,*

$$(1) \quad \frac{1}{n}\sum_{\pi \in \Pi}[S(\widehat{u}_\pi(\Lambda)) - S(u_\pi)] \leq \delta_T^2$$

$$(2) \quad |S(\widehat{u}(\Lambda)) - S(u))| \leq \delta_T$$

Define $\widetilde{F}(x) := \frac{1}{n}\sum_{\pi \in \Pi} \mathbb{1}\{S(u_\pi < x)\}$ and $F(x) := \mathbb{P}(S(u) < x)$.

Lemma F.2 (Lemma H.2 Chernozhukov et al. (2021))**.** *Let Π be moving block permutations. Suppose that $\{u_t\}_{t=1}^T$ is a stationary and strong mixing. Assume the following conditions: (1) $\sum_{k=1}^{\infty} \alpha_{mixing}(k) < M$ for some M, (2) $T_0 \geq T_1 + 2$, and (3) $S(u)$ has bounded pdf. Then there exists a $M' > 0$ depending only on M such that for any $\delta_{1T} > 0$,*

$$\mathbb{P}(\sup_{x \in \mathbb{R}} |\widetilde{F}(x) - F(x)| \leq \delta_{1T}) \geq 1 - (M'\sqrt{\frac{T_1}{T_0}} logT_0 + \frac{T_1 + 1}{T_0 + T_1})/\delta_{1T}$$

Lemma F.3 (Lemma H.1 Chernozhukov et al. (2021)). *Suppose that with probability at least* $1-\gamma_{1T}$ *we have*

$$\sup_{x\in\mathbb{R}} |\widetilde{F}(x) - F(x)| \le \delta_{1T}$$

and with probability at least $1-\gamma_T$

$$\begin{aligned} &(1) \quad \frac{1}{n}\sum_{\pi\in\Pi} |S(\widehat{u}_\pi(\Lambda)) - S(u_\pi)| \le \delta_T^2, \\ &(2) \quad |S(\widehat{u}(\Lambda)) - S(u)| \le \delta_T \\ &(3) \quad \text{The pdf of } S(u) \text{ is bounded above by } D \end{aligned}$$

Then for any $\theta \in (0,1)$,

$$|\mathbb{P}(\widehat{p}(\Lambda) \le \theta) - \theta| \le 4\delta_{1T} + 4\delta_T + 2D(\delta_T + 2\sqrt{\delta_T}) + \gamma_{1T} + \gamma_T$$

We are now ready to prove Theorem 3. Observe that

$$\begin{aligned} \widehat{P}_t^N(\Lambda) - P_t^N &= -(y_{0t}^I - y_t'\widetilde{W}_T^{SC}(\Lambda) - \alpha_{0t}) - W^{*'} y_t^N + y_{0t}^N \\ &= -(y_{0t}^N - y_t'\widetilde{W}_T^{SC}(\Lambda)) - W^{*'} y_t^N + y_{0t}^N \\ &= (\widetilde{W}_T^{SC}(\Lambda) - W^*)' y_t^N \\ &= (\widetilde{W}_T^{SC}(\Lambda) - W^*)' c + (\widetilde{W}_T^{SC}(\Lambda) - W^*)'\mu\lambda_t + (\widehat{W}_T^{SC}(\Lambda) - W^*)'\varepsilon_t \end{aligned}$$

Therefore, by noting the simple inequality of $(a+b+c)^2 \le 4a^2 + 4b^2 + 4c^2$,

$$\begin{aligned} &||\widehat{P}^N(\Lambda) - P^N||_2^2/T = \frac{1}{T}\sum_{t=1}^T (\widehat{P}_t^N(\lambda) - P_t^N)^2 \\ &\le 4\left\{(\widetilde{W}_T^{SC}(\Lambda) - W^*)'c\right\}^2 + 4||(\widetilde{W}_T^{SC}(\lambda) - W^*)'\mu||^2 \left\{\frac{1}{T}\sum_{t=1}^T ||\lambda_t||^2\right\} \\ &+ 4||(\widetilde{W}_T^{SC}(\Lambda) - W^*)\mu||^2 \left\{\frac{1}{T}\sum_{t=1}^T ||\varepsilon_t||^2\right\} \\ &\le 4||c||^2 \cdot ||\overline{W}(\Lambda) - W^*||^2 + 8||\mu||^2 ||\overline{W}(\Lambda) - W^*||^2 \sigma^2 + o_p(1) \\ &= 4(||c||^2 + 2||\mu||^2\sigma^2)||\overline{W}(\Lambda) - W^*||^2 \end{aligned}$$

since $\mathbb{E}||\lambda_t||_2, \mathbb{E}||\varepsilon_t||_2 < \sigma^2 < \infty$ by assumption, Markov inequality, and $\widetilde{W}_T^{SC}(\Lambda) \xrightarrow{p} \overline{W}(\Lambda)$ by Theorem 2. Define $C_1 := 2\sqrt{(||c||^2 + 2||\mu||^2\sigma^2)}$ and $C_2 := ||c|| + \sigma^2||\mu|| + \sigma^2$ so that for some sequence of $\gamma_T = o(1)$, for probability at least $1-\gamma_T$, we have

$$||\widehat{P}^N(\Lambda) - P^N||_2/\sqrt{T} \le C_1 ||\overline{W}(\Lambda) - W^*||$$

and

$$|\widehat{P}_t^N(\Lambda) - P_t^N| \leq C_2||\overline{W}(\Lambda) - W^*||$$

Then define

$$\delta_T := \max(C_1, C_2)||\overline{W}(\Lambda) - W^*|| \tag{F.1}$$

Next, by setting $\delta_{1T} = (\frac{T_1}{T_0})^{1/4}$ and applying Lemma F.3 for (F.1), we obtain

$$\begin{aligned}
|\mathbb{P}(\widehat{p}(\Lambda) \leq \theta) - \theta| &\leq 4\delta_{1T} + 4\delta_T + 2D(\delta_T + 2\sqrt{\delta_T}) + \gamma_{1T} + \gamma_T \\
&\leq 4(T_1/T_0)^{1/4} + 4\delta_T + 2D(\delta_T + 2\sqrt{\delta_T}) + (M'\sqrt{\frac{T_1}{T_0}}logT_0 + \frac{T_1+1}{T_0+T_1})/(T_1/T_0)^{-1/4} + \gamma_T \\
&\leq \overline{C}\left\{(T_1/T_0)^{1/4}logT_0 + \delta_T + \sqrt{\delta_T}+\right\} + \gamma_T
\end{aligned}$$

G Regularity conditions for consistent estimator

Assumption 5. *For each $\Lambda \in \Theta$, where Θ is compact, $\widetilde{W}_T^{SC}(\Lambda) - \overline{W}(\Lambda) = o_p(1)$*

Assumption 6. *Suppose there is a $B_T = O_p(1)$ such that for all $\Lambda, \widetilde{\Lambda} \in \Theta$, we have $|\widetilde{W}_T^{SC}(\Lambda) - \widetilde{W}_T^{SC}(\widetilde{\Lambda})| \leq B_T h(d(\Lambda, \widetilde{\Lambda}))$, where $h(0) = 0$ and h is continuous at 0*

Lemma G.1. *Suppose assumptions 5 and 6 holds. Then*

$$\sup_{\Lambda\in\Theta}\left|\widetilde{W}_T^{SC}(\Lambda) - \overline{W}(\Lambda)\right| = o_p(1)$$

Proof of Lemma G.1
First we show that for every $\Lambda^\dagger \in \Theta$, there exists some $\delta > 0$ such that

$$\sup_{\Lambda\in B_\delta(\Lambda^\dagger)}\left|\widetilde{W}_T^{SC}(\Lambda) - \widetilde{W}_T^{SC}(\Lambda^\dagger)\right| \leq G_T(\varepsilon, \gamma) \tag{G.1}$$

where $\mathbb{P}(|G_T(\varepsilon, \gamma)| > \varepsilon) < \gamma$ for $T \geq T(\varepsilon, \gamma)$. Since $B_T = O_P(1)$, there exists an M such that for every T, $\mathbb{P}(B_T > \varepsilon M) < \gamma$. Define $G_T(\varepsilon, \gamma) := B_T/M$. Choose δ small enough so that $h(\ell) < 1/M$ for any $\ell < \delta$. Then

$$\sup_{\Lambda\in B_\delta(\Lambda^\dagger)}\left|\widetilde{W}_T^{SC}(\Lambda) - \widetilde{W}_T^{SC}(\Lambda^\dagger)\right| \leq B_T \sup_{\Lambda\in B_\delta(\Lambda^\dagger)} h(d(\Lambda, \Lambda^\dagger)) \leq B_T/M = G_T(\varepsilon, \gamma)$$

so that (G.1) is shown. Next we show the result. Define $R_T(\Lambda) := \widetilde{W}_T^{SC}(\Lambda) - \overline{W}(\Lambda)$. Then there are finite balls $B_\delta(\Lambda_i)$ for some $i = 1, .., m$ such that $\Theta \subset \cup_{i=1}^m B_\delta(\Lambda_i)$ and $\Lambda_i \in \Theta$, so

$$\sup_{\Lambda\in\Theta}|R_T(\Lambda)| \leq \max_{i=1,\dots,m}|R_T(\Lambda_i)| + \max_{i=1,\dots m}\sup_{\Lambda\in B_\delta(\Lambda_i)}|R_T(\Lambda) - R_T(\Lambda_i)| = o_p(1) + G_T(\varepsilon, \gamma)$$

where the equality follows from assumption 5. Therefore

$$\mathbb{P}(\sup_{\Lambda\in\Theta} > \varepsilon) \leq \mathbb{P}(o_p(1) > \varepsilon/2) + \mathbb{P}(G_T(\varepsilon,\gamma) > \varepsilon/2) < \eta$$

for large enough n, which follows from (G.1).

□

If assumption 6 holds, specifically that $B_T = O_p(1)$ over $\Lambda \in (0,1]$, then we have a more general result than (C.4), which allows us to show that for **any** sequence of $0 < \Lambda_T \downarrow 0$,

$$|\widetilde{W}^{SC}(\Lambda_T) - W^*| = o_p(1).$$

The implication is that we can then construct a feasible statistic $\widetilde{W}^{SC}(\Lambda_T)$ that is asymptotically recovers the treated individual's fixed effect c_0 and factor loading μ_0. However, as Λ converges to 0, $(\Lambda I + cc' + \mu'\Omega_0\mu)^{-1}$ will blow up if $cc' + \mu'\Omega_0\mu$ is positive semi-definite. This means that for any $\Lambda_1, \Lambda_2 \in (0,1]$,

$$\max_{i=1,\ldots,J} \frac{|\Lambda_1 - \Lambda_2|}{(d_i + \Lambda_1)(d_i + \Lambda_2)} = \frac{1}{\Lambda_1\Lambda_2}|\Lambda_1 - \Lambda_2|,$$

implying that the equi-continuity of $\overline{W}(\Lambda)_{\Lambda\in(0,1]}$ cannot hold, and this violates condition (5) of Lemma C.1. However, we can modify assumption 6 and obtain some analogous result to Lemma G.1. We need to assume some "smoothness" on the rate of convergence of $R_T(\Lambda)$.