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

Dennis Lim*

August 29, 2024

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 the bias and variance of estimates 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. In common 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. The synthetic control method (SCM) was proposed as a feasible solution, which works by estimating a weighted-average of control-units in the pre-treatment period and reconstructing the counterfactual outcome of the treated unit during the treatment period. Abadie, Diamond, and Hainmueller (2010) were amongst the first to provide theoretical justifications and conditions under which the SCM works. 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, commonly referred to as 'perfect pre-treatment fit' (PPTF). This assumption is has been maintained by Abadie and Gardeazabal (2003), Doudchenko and Imbens (2016) and Abadie, Diamond, and Hainmueller (2015), among many others. Under PPTF, they directly estimate the treatment effect during the

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

treated period without having to recover these weights by constructing an estimator, since weight recovery is generally infeasible as the vector of unknown factor-loadings (i.e. μ_j in (2.1)) are not observed.[1] This was pointed out by Abadie et al. (2010), who instead proposed an estimator that is approximately unbiased.[2] Ben-Michael, Feller, and Rothstein (2021)[Lemma 3] derived finite-sample bounds on the bias of the SCM and showed that the bounds they derive do not converge to zero when the number of control units are taken to be fixed and the number of pre-treatment period diverges to infinity, even under PPTF. In fact, Ferman and Pinto (2021) showed that it is impossible for this bound to converge to zero by showing that it is impossible for the SCM to recover the oracle weights. Among those trying to provide valid estimation for fixed control units, Amjad, Shah, and Shen (2018) proposed a two-step de-noising algorithm which requires a choice of hyperparameters, but requires data to be serially uncorrelated. In contrast, our method only requires the idiosyncratic-errors to be weakly-dependent in order to derive an unbiased estimate for the treatment effect. Our results also complement the works of Bai (2003), Abadie et al. (2010) and Ben-Michael et al. (2021) in recovering the pre-treatment weights under a fixed number of controls.

When the number of control units diverge instead, Arkhangelsky, Athey, Hirshberg, Imbens, and Wager (2021) provides an alternative method for inference on the treatment effect, but requires the errors to be identically and independently distributed Gaussian vectors. Xu (2017) relaxes this constraint and allows for weak serial dependence of the error terms, but requires error to be cross-sectionally independent and homoscedastic. Ferman (2021) provides conditions under which the demeaned SCM estimators is also asymptotically unbiased,

Our first contribution is the following:

- We propose a method such that under "perfect pre-treatment-fit", even when controls J are fixed, our SCM is valid using Chernozhukov, Wuthrich, and Zhu (2021)'s method, because our method consistently reproduces the pre-treatment weights. This complements Ben-Michael et al. (2021) in that our method does not require $J \to \infty$. It also complements Bai (2003)'s work in that he argued that estimating factor-loadings consistently when J is fixed under generally requires strong assumptions on the idiosyncratic shocks. In contrast, we can recover these factor-loadings without the need to make such strong assumptions. Abadie et al. (2010) also showed that we can recover factor-loadings close enough to the oracle/true values whenever the idiosyncratic variance is relatively small; our proposed method allows for significant idiosyncratic shocks. Ferman (2021) showed that these factor-loadings can be recovered whenever the weights are diluted among an increasing number of control-units, but assumes

[1] see σ_ε^2 in (2.3)

[2] Our paper differs from their paper in several facets: (A) They require the factor-model to be auto-regressive - see equation (5) of their paper; (B) They require $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\lambda_t\lambda_t'$ to be non-singular, where λ_t is the time-varying vector of common factors, explained in Assumption 1. We neither require (A) or (B).

that such pretreatment weights exists.

Under imperfect pre-treatment fit, Ferman and Pinto (2021)'s solution is to "skip" the step of estimating the pre-treatment weights but assume that $\mathbb{E}[\lambda_t] = 0$. The remaining nuisance term $(c_0 - c'W^{SCM})$ can be eliminated by demeaning. If $\mathbb{E}[\lambda_t] \neq 0$, their method does not work **Our second contribution is as follows:**

- Even when $\mathbb{E}[\lambda_t] \neq 0$, our method minimizes the variance of the estimated result, in the sense that the variance arising from $\lambda_t(\mu_0 - \mu W^{SCM})$ is the smallest as can be. Therefore our SCM should have better performance than the one produced by Ferman and Pinto (2021).

Under imperfect pre-treatment fit, if $\mathbb{E}[\lambda_t] = 0$, then Ferman and Pinto (2021) provide an estimator that is valid. We will need to show that under such condition,

- Our proposed SCM is also valid. Since this time pre-treatment weights do not exist, we have no way of recovering any sensible weights. However, we can demean as in Ferman and Pinto (2021) to remove nuisance term $(c_0 - c'W^{SCM})$, then apply Chernozhukov et al. (2021) as what they have done.

	$\widetilde{W}$	DID	SCM	Constrained-LASSO	FP
$\theta = \mathbf{0.05}$	0.061	0.716	0.522	0.112	0.547
$\theta = \mathbf{0.1}$	0.154	0.966	0.831	0.217	0.863

Table 1: size control for 1,000 replications using different weight estimator in the literature. θ is the nominal-size. Our proposed estimator is $\widetilde{W}$. See Section 4 for more details

An important contribution by Ferman and Pinto (2021) is the introduction of "imperfect pre-treatment fit", where such weights in general cannot be recovered.[3] 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 that even when the number of control-units is larger than the number of pre-treatment periods, well-known estimators are generally consistent. To understand 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 when the number of control-units are fixed. Our paper contributes to the

[3]See their Proposition 1

literature by providing a test that (1) is consistent when the number of control-units are fixed and (2) provides weights such that our estimator converges to a minimum variance (MV) estimate even when the time-varying common factor λ_t is non-stationary. We summarize this in Table 1.

Pre-Treatment fit	Ben-Michael et al. (2021)	Ferman and Pinto (2021)	Proposed SCM
Perfect, λ_t stationary			
Diverging control-units	Yes	Yes	Yes
Fixed control-units	No	Yes	Yes
Perfect, λ_t non-stationary			
Diverging control-units	Yes	Yes	Yes
Fixed control-units	No	No	Yes
Imperfect, λ_t stationary			
Diverging control-units	Yes	Yes	Yes
Fixed control-units	No	Yes	Yes
Imperfect, λ_t non-stationary			
Diverging control-units	Yes	Yes	Yes
Fixed control-units	No	No	No, but MV

Table 2: Summary of SCM in literature; "Yes" if unbiased estimator and "No" otherwise

Structure of Paper: Section 2 provides the model setup of our paper. Section 3 derives the main theoretical results. Section 4 provides the simulation results. The proofs of the result in the main text are contained in the Appendix.

2 Model

We assume that we observe a panel of $J+2$ individuals, in time $t \in \mathcal{T}_0 \cup \mathcal{T}_1$, where $\mathcal{T}_0$ is the period where no individuals are treated and $T_0 := card(\mathcal{T}_0)$ is the number of periods associated with $\mathcal{T}_0$. Furthermore, let $\mathcal{T}_1$ be the period that individual $j = 0$ is treated with the remaining individuals $j = 1, ..., J+1$ still untreated and $T_1 := card(\mathcal{T}_1)$ be the number of periods associated with $\mathcal{T}_1$; denote $T := T_0 + T_1$ to be the total number of periods under consideration. We consider the following factor model,[4] which has been extensively studied.[5]

Assumption 1 (Factor model potential outcome)**.** *The potential outcome for unit j at time t for the treated (y^I_{jt}) and non-treated (y^N_{jt}) are given by*

$$\begin{aligned} y^N_{jt} &= c_j + \delta_t + \lambda_t' \mu_j + \varepsilon_{jt} \\ y^I_{jt} &= \alpha_{jt} + y^N_{jt} \end{aligned} \tag{2.1}$$

[4]We assume without loss of generality that $c_{J+1} = 0$, since we can always replace δ_t by $\delta_t + c_{J+1}$.

[5]see Bai (2003) Abadie et al. (2010), Ferman and Pinto (2021) among many others

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

We are interested in testing

$$H_0 : \alpha_0 = \alpha \quad \text{versus} \quad H_1 : \alpha_0 \neq \alpha$$

where $\alpha_0 = \{\alpha_{0t}\}_{t \in \mathcal{T}_1}$ is some sequence of true parameter and $\alpha = \{\alpha_t\}_{t \in \mathcal{T}_1}$ is the sequence of parameters we hypothesize to be true. We assume throughout this paper that $T_0 \to \infty$. The case where $T_1 \to \infty$ has been studied extensively by Chernozhukov, Wuthrich, and Zhu (2022); in particular they introduce a t-test that provides valid inference for the average treatment effect (i.e. inference on $\frac{1}{T_1} \sum_{t \in \mathcal{T}_1} \alpha_{0t}$). Instead, our focus will be on the case where T_1 is fixed, so that the central-limit-theorem argument used to derive their t-test is no longer valid. We assume that our sampling procedure follows the given structure.

Assumption 2 (Sampling structure)**.** *We observe a realization of $\{y_{0t}, \dots, y_{Jt}\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$, 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 $j = 0, \dots, J$ as stochastic*

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

$$\widehat{W}^{SC} := \operatorname*{arg\,min}_{W : \sum_{i=1}^J W = 1} \frac{1}{T_0} \sum_{t \in \mathcal{T}_0} (y_{0t} - y_t' W)^2.$$

Fixing some weight W such that $\sum_{i=1}^J W_i = 1$, we can write

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

where $\mu := (\mu_1, ..., \mu_J)'$ and the last equality 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 shown that

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

The usual SCM estimator for α_{0t} is given as

$$\widehat{\alpha}_{0t}^{SCM}(W) := y_{0t} - y_t'W \xrightarrow{p} \alpha_{0t} + \lambda_t(\mu_0 - \mu'W) + (c_0 - c'W) + (\varepsilon_{0t} - \varepsilon_t'W) \tag{2.4}$$

Note that the asymptotic variance of the estimator $\widehat{\alpha}_{0t}(W)$ is given as $Q_0(W)$, so that the limit of $\widetilde{W}^{SC}$ is actually the minimizer argument for the asymptotic variance of the treatment effect estimator, i.e. $\overline{W} \in \arg\min_W avar(\widehat{\alpha}_t(W))$

We see that the σ_ε^2 given in (2.3) prevents us from recovering the pre-treatment weights when these weights exist, i.e. the variance of the error coming from $\varepsilon_{0t} - \varepsilon_t'W$ given in (2.2). Ferman and Pinto (2021) explain 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

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

(B) 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

$$\underset{W}{\arg\min}\, \widetilde{Q}_{T_0}(W) \xrightarrow{p} \underset{W}{\arg\min}\, Q_0(W) \tag{2.5}$$

for $Q_0(W) = (c_0 - c'W)^2 + (\mu_0 - \mu W)'\Omega_0(\mu_0 - \mu W)$, which generally holds whenever $Q_0(\cdot)$ has a unique solution. When this uniqueness does not hold, a strictly-convex penalty term can be introduced to induce a unique solution. Then intuitively, the block SCM approach can remove the need for the number of control units to increase. Furthermore, under PPTF, we are able to completely recover the unique weights. Our paper therefore contributes to the SCM literature in the following ways: (1) We can obtaining consistent confidence intervals around α_{0t} under the true null, while relaxing both assumptions (A) and (B), i.e. we allow the number of controls to be fixed;

(2) When the time-invariant fixed effect λ_t is non-stationary, even under imperfect pre-treatment fit, we minimize the weighted distance W between c_0 and $c'W$, as well as μ_0 and μW. Under perfect pre-treatment fit we show that we can fully recover the weights; (3) When λ_t is stationary we have exact size-control, as in the case of Ferman and Pinto (2021).

2.1 Test Statistic

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 $j \in \{0, 1, ..., J\}$, $s \in \{1, ..., r\}$ 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 $\overline{y}^s := \frac{1}{\Delta}(y_{s\Delta+1} + y_{s\Delta+2} + ... + y_{(s+1)\Delta})$. We define the demeaned-block-synthetic-control-method (DBSCM) weight as[6]

$$\widetilde{W}^{DBSCM}_T(f, \Lambda) := \underset{W \in \Delta^J_\eta}{\arg\min} \left\{ \frac{1}{r} \sum_{s=1}^{r} \left\{ \overline{y}^s_0 - (\overline{y}^s)'W - (\overline{y}_0 - \overline{y}'W) \right\}^2 + f(W, \Lambda) \right\} \tag{2.6}$$

where $\Lambda \geq 0$ is some given value, $\Delta^J_\eta := \{W \in \mathbb{R}^J : ||W||_2 \leq \eta\}$ is the set containing vectors of weight W whose Frobenius-norm is bounded by some fixed $\eta > 0$ and $f(W, \Lambda)$[7] is some non-negative penalty term with the property that it (is)

1. Strictly-convex in $W \in \mathbb{R}^J$
2. Equi-continuous in $(W, \Lambda) \in \Delta^J_\eta \times [0, 1]$ for $\Lambda > 0$
3. Converges to zero whenever $0 < \Lambda \downarrow 0$
4. Equals zero whenever Λ equals zero

If we define $f(W, \Lambda) := \Lambda ||W||_1$, then we have an ℓ_1-regularized penalty-term; if instead we take $f(W, \Lambda) := \Lambda ||W||^2_2$ then we have a ridge penalty-term. The main result of the paper is that under the correct null $\alpha_t = \alpha_{0t}$, for any $\theta \in (0, 1)$,

$$\mathbb{P}(\widehat{p}(\widetilde{W}^{DBSCM}_T(f, \Lambda_T)) \leq \theta) \to \theta$$

for some sequence of $0 < \Lambda_T \downarrow 0$ and $\widehat{p}(\cdot)$ is defined in section 3.3; this result is given in Corollary 3.4, which allows us to conduct inference on α_{0t} for any fixed $t \in \mathcal{T}_1$.

[6] Note that our simplex Δ^J_η allows for negative weights, which differs from Abadie et al. (2010), Abadie et al. (2015) in that weights are assumed to be non-negative

[7] In Theorem 3 we consider $f(\Lambda, W) := \Lambda \left(W'W + (\sum_{j=1}^{J} W_j - 1)^2 \right)$

3 Main Results

Throughout the rest of the paper, unless stated otherwise, we always take t to be a fixed value of $\mathcal{T}_1$. Define $Y_{jt} := y_{jt} - y_{J+1,t}$ for $j = 0, 1, ..., J$ so that we rewrite (2.1) as

$$\begin{aligned} Y_{jt}^N &= C_j + \lambda_t' M_j + u_{jt} \\ Y_{jt}^I &= \alpha_{jt} + Y_{jt}^N \end{aligned} \tag{3.1}$$

where $C_j := c_j - c_{J+1} = c_j$,[8] $M_j := \mu_j - \mu_{J+1}$ and $u_{jt} := \varepsilon_{jt} - \varepsilon_{J+1,t}$. We denote $C := (C_1, ..., C_J)'$ and $M := (M_1, ..., M_J)'$ as the $J \times 1$ vector and $J \times F$ matrix containing C_j and M_j respectively for $j = 1, ..., J$. Furthermore, define $\Delta_\eta^J := \{W \in \mathbb{R}^J : ||W|| \leq \eta\}$ for some given $\eta > 0$.

Assumption 3. *Suppose T_1 is fixed, $T_0 \to \infty$, $\mathbb{E}\varepsilon_{jt} = 0$ ε_t is stationary with holder-continuity of exponent $\beta > 1$ 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 $\frac{1}{T_0}\sum_{t \in \mathcal{T}_0} \varepsilon_t \varepsilon_t' \xrightarrow{p} \Sigma$, $\frac{1}{T_0}\sum_{t=1}^{T_0} \lambda_t \xrightarrow{p} 0$ and $\frac{1}{r}\sum_{s=1}^{r} \overline{\lambda}^s (\overline{\lambda}^s)' \xrightarrow{p} \Omega_0$, where Σ and Ω_0 are positive semi-definite matrices.*

Assumption 3 is similar to Ferman and Pinto (2021)[Assumption 4], except 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 $O_p(1)$, which is a weaker requirement. Note that the assumption $\frac{1}{T_0}\sum_{t=1}^{T_0} \lambda_t \xrightarrow{p} 0$ is without loss of generality.[9]

The SCM weights defined as Δ_η^J slightly differs from the usual case if two ways: (1) $||W|| = 1$ in the usual SCM weights and (2) $W_j \geq 0$ for $j = 1, ..., J$. The first is called the "adding-up" constraint, while the second is the "non-negativity" constraint. This is done so as to retain interpret-ability of the SCM weights (see Abadie et al. (2010)). However, if the unit of interest is an outlier relative to the control-units, then the adding-up constraint might fail. For instance, Abadie et al. (2010) studied the effect of the tobacco control program in California[10] by using per capita smoking as outcome. However, if the chosen treated unit is heavily dependent on tobacco, then this constraint might fail. Next, the non-negativity constraint ensures a unique solution as well as reduces the deviation of the estimated weights to the true weights by limiting the sum of squared weights which enters into the variance under estimation. This often ensures that the weights are non-zero only for a small subset of the control units which makes the weights easier to interpret. In many cases raw-correlations between the treated and control-units are positive; however, this does not mean

[8] Recall that we can assume $C_{J+1} = 0$ without loss of generality

[9] Suppose $\frac{1}{T_0}\sum_{t=1}^{T_0} \lambda_t \xrightarrow{p} w_0 \neq 0$. Then we can consider an observably equivalent model with $w_0 = 0$ by adjusting c_j for each $j = 1, ..., J+1$

[10] this program was called proposition 99

that the correlation between the treated and every control-unit must be non-negative. Allowing for negative weights can also improve out-of-sample predictions.

3.1 Perfect Pre-Treatment Fit

Definition 3.1. *We say that we have perfect pre-treatment fit whenever there exists some* $\widetilde{W}$ *such that* $\widetilde{W} \xrightarrow{P} \overline{W}$ *with* $\overline{W} \in \Phi := \{W \in \Delta_\eta^J : W'C = C_0 \text{ and } W'M = M_0\}$ *for some* $\eta > 0$

The term "perfect pre-treatment fit" (or PPTF as we have denoted it) usually refers to the case where some linear combination of controls units (in this case the $j = 1, ..., J+1$) can perfectly recover the time-invariant fixed effect of the treated individual c_0 and its unknown factor-loadings μ_0 in (2.1). Formally, the conventional definition of PPTF means that there exists some $W^* \in \widetilde{\Phi} := \{W \in \widetilde{\Delta}_1 : W'c = c_0 \ \ and \ \ W'\mu = \mu_0\}$, where $\widetilde{\Delta}_1 := \{W \in \mathbb{R}^{J+1} : \sum_{i=1}^{J+1} W_i = 1 \ \ and \ \ W_i \geq 0 \quad for \quad i = 1, ..., J+1.$[11] This is slightly different from definition 3.1. However, whenever we have PPTF in the usual sense, we will also have PPTF under definition 3.1. Formally we have the following:

Lemma 3.1. *Suppose there exists some* $W^* \in \widetilde{\Phi}$*. Then there exists some* $\eta > 0$ *and some* $\overline{W} \in \Phi$*. In particular, we can let* $\eta = 1$*.*

Therefore definition 3.1 can be seen as a more general version of the usual PPTF. We will work with this more general definition throughout this paper as it allows for extrapolation out of the usual simplex, possibly enabling better pre-treatment fit and ultimately out-of-sample prediction. We have an alternative characterization of PPTF: Define

$$G_0 := \begin{pmatrix} C_0 \\ M_0 \end{pmatrix} \quad and \quad G := \begin{pmatrix} C' \\ M \end{pmatrix}$$

Then we will have PPTF whenever there exists a $\overline{W} \in \Phi$ such that $GW = G_0$. If G has full column-rank, $W = (G'G)^{-1}G'(C_0, M_0')'$ is the solution to $GW = G_0$. In general, we can choose η to be large enough to ensure that at least one solution falls within the prescribed Δ_η^J. The only time when PPTF fails is when G_0 cannot be written as a linear combination of G, for instance, when the number of control-units J is relatively smaller than the $1 + F$ features coming from (C_j, M_j). In this case, our SCM weights $\widetilde{W}_T^{SC}(\Lambda)$ converges to weights W such that W minimizes the distance between GW and G_0 subject to some weighing matrix. This can be interpreted as our method trying to search for weights that best approximate G_0 by a linear combination of G as best as possible. We begin by defining the Block synthetic control method weight (BSCM) as

$$\widetilde{W}_T^{BSCM}(f, \Lambda) := \underset{W \in \Delta_\eta^J}{\arg\min} \left\{ \frac{1}{r} \sum_{s=1}^{r} \left\{ \overline{y}_0^s - (\overline{y}^s)'W \right\}^2 + f(W, \Lambda) \right\}$$

[11]Note that $\overline{W} \in \mathbb{R}^J$ while $W^* \in \mathbb{R}^{J+1}$

where f satisfies conditions 1–4. Then we have the following result.

Theorem 1 (Uniform approximation BSCM estimator)**.** *Suppose assumption 1, 2 and 3 holds. Then for any fixed $\gamma > 0$ and any f that satisfies conditions (1)-(4), we have,*

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

where $\overline{W}(\Lambda) := \arg\min_{W \in \Delta_\eta^J} \mathcal{A}(f,W,\Lambda)$ and $\mathcal{A}(f,W,\Lambda) := (C_0 - C'W)^2 + (M_0 - MW)'\Omega_0(M_0 - MW) + f(W,\Lambda)$; moreover, we can take γ to be zero whenever $(CC' + M'\Omega_0 M)$ is positive-definite. As $0 < \Lambda \downarrow 0$,

$$\overline{W}(\Lambda) \to \overline{W}(0)$$

where $\overline{W}(0) \in \min\arg\min_{W \in \Delta_\eta^J}(GW - G_0)'\widetilde{V}(GW - G_0)$, $\widetilde{V} := diag(1,\Omega_0)$ is a weighing matrix and $\min\arg\min(\cdot)$ is the minimum-norm vector in the space of $\arg\min(\cdot)$, where $\arg\min(\cdot)$ is the argument that minimizes $(\cdot)$.

Remark 1. *The reason we consider $\Lambda \downarrow 0$ instead of setting $\Lambda = 0$ and solving it directly is due to the fact that the objective function $\mathcal{A}(f,W,0)$ may not be strictly convex (by Ω_0 only being positive semi-definite), and therefore not unique. The solution set in this case could have cardinality greater than two, which renders the Newey and McFadden (1994) approach redundant. If Ω_0 is positive semi-definite and so $(GW - G_0)'\widetilde{V}(GW - G_0)$ does not have a unique solution, we could in principle set $\Lambda = 0$ to obtain some solution set for $\arg\min_{W \in \Delta_\eta^J} \frac{1}{r}\sum_{s=1}^r \{\overline{y}_0^s - (\overline{y}^s)'W)\}^2$, and then pick any of these solution, say $\check{W}$. However, this is not the most efficient choice: by recalling (2.3) and (2.4), the asymptotic variance of $\widehat{\alpha}_{0t}^{SCM}(\check{W})$ is at least as large as $\widehat{\alpha}_{0t}^{SCM}(\overline{W}(0))$, i.e.*

$$\begin{aligned} avar(\widehat{\alpha}_{0t}^{SCM}(\check{W})) &= Q_0(\check{W}) = \mathcal{A}(f,\check{W},0) + \sigma_\varepsilon^2(1 + \check{W}'\check{W}) \\ &\geq \mathcal{A}(f,\overline{W}(0),0) + \sigma_\varepsilon^2(1 + \overline{W}(0)'\overline{W}(0)) = Q_0(\overline{W}(0)) = avar(\widehat{\alpha}_{0t}^{SCM}(\overline{W}(0))) \end{aligned} \tag{3.2}$$

if we assume that errors are homoskedastic (i.e. $\Sigma = \sigma_\varepsilon^2 I_J$)[12]*, where we recall that $\check{W} \in \arg\min \mathcal{A}(f,W,0)$ and $\overline{W}(0) \in \min\arg\min \mathcal{A}(f,W,0)$.*

The difficulty in considering $\gamma = 0$ stems from the assumption that $\Omega_0 \equiv Plim_{T_0 \to \infty} \frac{1}{T_0}\sum_{t \in \mathcal{T}_0} \lambda_t \lambda_t'$ 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 that solves $\Omega_0(M_0 - MW) = 0$, due to this lack of unique identification. To be precise, note that for any $\Lambda > 0$, we can express

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

[12] The homoskedastic assumption is made for simplification of argument; it can be shown with some algebra that the inequality between the left and right-hand-side of (3.2) still holds under general heteroskedastic errors.

As $\Lambda \downarrow 0$, the term $(\Lambda I + cc' + \mu'\Omega_0\mu)^{-1}$ could potentially diverge 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. If we instead strengthen Ω_0 in assumption 3 to being positive-definite instead, then Theorem 1 implies that for **any** sequence of $0 \leq \Lambda_T \downarrow 0$,

$$\widetilde{W}_T^{BSCM}(f, \Lambda_T) \xrightarrow{p} \overline{W}(0)$$

An application of Theorem 4 below yields exact asymptotic size-control under the correct null.

Under PPTF, $\overline{W}(0)$ will be the pre-treatment weights that we seek. Note that $\overline{W}(0)$ is unique by Lemma A.1, even when Ω_0 is only positive semi-definite. In general we can ensure that $\overline{W}(0) \in \Phi$ whenever $\arg\min_{W\in\Delta_\eta^J}(GW - G_0)'\widetilde{V}(GW - G_0)$ has a single solution. A sufficient condition is for Ω_0 to be positive-definite. This usually occurs whenever $\{\lambda_t\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$ is a non-stationary process.

Example 1. *consider*

$$\{\lambda_1, \lambda_2, ..., \lambda_9\} = \{(1,0)', (1,0)', (0,1)', (0,1)', (1,1)', (1,1)', (1,1)', (1,1)', (1,1)'\}$$

and this time-dependent common factor repeats itself. Define

$$\omega_0 := \lim_{T_0\to\infty} \frac{1}{T_0} \sum_{t\in\mathcal{T}_0} \lambda_t = \frac{2}{9}(1,0)' + \frac{2}{9}(0,1)' + \frac{5}{9}(1,1)' = (0.777, 0.777)'$$

so that defining $\widetilde{\lambda}_t := \lambda_t - \omega_0$, we have

$$\lim_{T_0\to\infty} \frac{1}{T_0} \sum_{t\in\mathcal{T}_0} \widetilde{\lambda}_t = 0,$$

hence satisfying assumption 3. Consider a 3-fold cross fitting. Some algebraic manipulation yields

$$\overline{\lambda}^1 = (0.222, -0.777)', \qquad \overline{\lambda}^2 = (-0.777, 0.222)' \quad \textit{and} \quad \overline{\lambda}^3 = (0.222, 0.222)'$$

where $\overline{\lambda}_i := \frac{1}{\Delta_i}\sum_{t\in\Delta_i}\widetilde{\lambda}_t$ for $i = 1,2,3$. Then

$$\frac{1}{r}\sum_{s=1}^{r}(\overline{\lambda}^s)(\overline{\lambda}^s)' \to \begin{pmatrix} 0.567 & -0.432 \\ -0.432 & 0.567 \end{pmatrix},$$

which is positive-definite. □

A direct implication of Theorem 1 is the following:

Corollary 3.1. *Suppose assumption 1, 2 and 3 holds. Further assume that f satisfies conditions (1)-(4). Then for any $\xi > 0$, there exists a $\Lambda(\xi) > 0$ such that for any fixed $0 < \Lambda \leq \Lambda(\xi)$,*

$$|\widetilde{W}^{BSCM}(f, \Lambda) - \overline{W}(0)| \leq \xi + o_p(1)$$

Corollary 3.1 assures us that we can obtain as close an approximation to $\overline{W}(0)$. Therefore there exists some sequence of Λ_T such that $\widetilde{W}_T^{SC}(\Lambda_T)$ consistently estimates $\overline{W}(0)$. This is formalized below.

Corollary 3.2. *Suppose assumption 1, 2, 3 holds and f satisfies conditions (1)-(4). Then there exists a sequence $0 < \Lambda_T \downarrow 0$ such that*

$$\widetilde{W}_T^{BSCM}(f, \Lambda_T) = \overline{W}(0) + o_p(1)$$

If Ω_0 is positive-definite, then any sequence of $0 < \Lambda_T \downarrow 0$ will satisfy the preceding equation.

The block synthetic control-based weights $\widetilde{W}_T^{BSCM}(f, \Lambda_T)$ therefore provides a way to recover the pre-treatment weights whenever such weights exist within or even outside the simplex. In general, without making more assumptions it is impossible to obtain the sequence of penalty terms Λ_T given in corollary 3.2. Therefore, in applications, corollary 3.1 is more useful, and we simply choose an arbitrarily small $\Lambda > 0$.

3.2 Imperfect Pre-Treatment Fit

In the previous section, under perfect pre-treatment fit, we can "almost recover" the weights used to conduct unbiased inference. In this section we discuss the implications of our estimator under imperfect pre-treatment fit (IPTF), i.e. when $\nexists W^* \in \Phi$.

3.2.1 IPTF under Stationarity

We consider the case when $\{\lambda_t\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$ is a stationary process, so that Ω_0 will generally be a positive semi-definite matrix. In such a case we want our estimator to be at least efficient in the following sense: consider any fixed $t \in \mathcal{T}_1$ and let $\overline{W}(0)$ be the limit of $\widetilde{W}_T^{BSCM}(f, \Lambda_T)$ for some sequence of $0 < \Lambda_T \downarrow 0$, so that we recall from section 3.1 that

$$\widehat{\alpha}_{0t}^{SCM}(\widetilde{W}_T^{BSCM}(f, \Lambda_T)) \xrightarrow{p} \alpha_{0t} + (C_0 - C'\overline{W}(0)) + \lambda_t'(M_0 - M\overline{W}(0)) + (u_{0t} - u_t'\overline{W}(0)). \tag{3.3}$$

where $\widehat{\alpha}_{0t}^{SCM}(\cdot)$ was defined in (2.4). If Ω_0 is positive-definite, we will have by Theorem 1 that $\overline{W}(0) \in \Phi$, implying that the first term $C_0 - C'\overline{W}(0) = 0$; therefore $\widehat{\alpha}_{0t}^{SCM}(\widetilde{W}_T^{BSCM}(f, \Lambda_T))$ is an unbiased estimator for α_{0t} whenever $\mathbb{E}[\lambda_t] = 0$, as is the case under a stationary λ_t process. However, if instead Ω_0 is only positive-semi definite, the weights obtained from our BSCM procedure may not cancel out the first term (i.e. $C_0 \neq C'\overline{W}(0)$), making $\widehat{\alpha}_{0t}^{SCM}(\widetilde{W}_T^{BSCM}(f,\Lambda_T))$ a biased estimate. To overcome this, we can demean our BSCM and construct a "demeaned-version" of the BSCM defined as the DBSCM-based weights given in (2.6). Note that by Lemma A.1 this weight is uniquely defined. Then we have the following result:

Theorem 2 (Uniform approximation of DBSCM estimator)**.** *Suppose assumption 1, 2 and 3 holds. Then for any fixed $\gamma > 0$ and any f that satisfies conditions (1)-(4), we have*

$$\sup_{\Lambda\in[\gamma,1]}\left|\widetilde{W}_T^{DBSCM}(f,\Lambda) - \overline{W}^0(\Lambda)\right| = o_p(1)$$

where $\overline{W}^0(\Lambda) := \arg\min_{W\in\Delta_\eta^J}\mathcal{B}(f,W,\Lambda)$ and $\mathcal{B}(f,W,\Lambda) := (M_0 - MW)'\Omega_0(M_0 - MW) + f(W,\Lambda)$; moreover, we can take γ to be zero whenever Ω_0 is a positive-definite matrix. As $0 < \Lambda \downarrow 0$,

$$\overline{W}^0(\Lambda) \to \overline{W}^0(0)$$

where $\overline{W}^0(0) \in \min\arg\min_{W\in\Delta_\eta^J}(M_0 - MW)'\Omega_0(M_0 - MW)$

As a result of the preceding theorem, we can obtain a more efficient estimator of α_{0t} than using $\widehat{\alpha}_{0t}^{SCM}(\cdot)$ defined in (2.4). To this end, define the demeaned synthetic control-based estimator (which we denote as DSCM) for α_{0t} as

$$\widehat{\alpha}_{0t}^{DSCM}(W) := Y_{0t} - Y_t'W - \left(\overline{Y}_0 - \overline{Y}'W\right) \tag{3.4}$$

for any $t \in \mathcal{T}_0 \cup \mathcal{T}_1$. Then our demeaned block synthetic control method estimator (DBSCM) is given as

$$\widehat{\alpha}_{0t}^{DBSCM}(f,\Lambda) := \widehat{\alpha}_{0t}^{DSCM}\left(\widetilde{W}_T^{DBSCM}(f,\Lambda)\right) \xrightarrow{p} \alpha_{0t} + \lambda_t'(M_0 - M\overline{W}^0(\Lambda)) + (u_{0t} - u_t'\overline{W}^0(\Lambda)) \tag{3.5}$$

as $\Lambda_T \downarrow 0$. In this case, $\widehat{\alpha}_{0t}^{DBSCM}(f,\Lambda_T)$ is an unbiased estimator for α_{0t} whenever $\mathbb{E}[\lambda_t] = 0$, despite the existence of an imperfect pre-treatment fit. Furthermore, it is clear that for any $W \in \mathbb{R}^J$, $avar(\widehat{\alpha}_{0t}^{SCM}(W)) = avar(\widehat{\alpha}_{0t}^{DSCM}(W))$ by simply observing that $\widehat{\alpha}_{0t}^{SCM}(W)$ has an extra non-random term $(C_0 - C'W)$ in the limit (compare (3.3) and (3.5)). Therefore, by using $\widehat{\alpha}_{0t}^{DBSCM}(f,\Lambda_T)$ instead of $\widehat{\alpha}_{0t}^{SCM}(\widetilde{W}_T(f,\Lambda_T))$, we can retain efficiency while obtaining an unbiased estimator.

Remark 2. *As in Remark 1, it is more efficient to pick $\overline{W}^0(0)$ by letting $\overline{W}^0(\Lambda) \to \overline{W}^0(0)$ as $0 < \Lambda \downarrow 0$ instead of picking some arbitrary $\check{W} \in \arg\min_{W\in\Delta_\eta^J}\frac{1}{r}\sum_{s=1}^r\{\overline{y}_0^s - (\overline{y}^s)'W - (\overline{y}_0 - \overline{y}'W)\}^2$. To see this, simply observe that*

$$avar(\widehat{\alpha}_{0t}^{DSCM}(\check{W})) = \mathcal{B}(f,\check{W},0) + \sigma_\varepsilon^2(1 + \check{W}'\check{W})$$

$$\geq \mathcal{B}(f, \overline{W}^0(0), 0) + \sigma_\varepsilon^2(1 + \overline{W}^0(0)'\overline{W}^0(0)) = avar(\widehat{\alpha}_{0t}^{DBSCM}) \tag{3.6}$$

if we assume that errors are homoskedastic (i.e. $\Sigma = \sigma_\varepsilon^2 I_J$*)*[13]*, where we recall that* $\check{W} \in \arg\min \mathcal{B}(f, W, 0)$ *and* $\overline{W}^0(0) \in \min \arg\min \mathcal{B}(f, W, 0)$.

3.2.2 IPTF under Non-Stationarity and Efficiency of Existing Estimators

In this section we will discuss the implication of existing estimators under Imperfect Pre-Treatment Fit without the assumed stationarity of λ_t. We will also explore the efficiency and the bias of our proposed estimator together with existing estimators. To this end, note that the difference-in-difference (DID) estimator can be written as

$$\widehat{\alpha}_{0t}^{DID} \equiv \widehat{\alpha}_{0t}^{DSCM}\left(\frac{1}{J}\iota\right)$$

where $\iota = (1, ..., 1)' \in \mathbb{R}^J$ with $\widehat{\alpha}_{0t}^{DSCM}(\cdot)$ is defined in (3.4). In general, one can always consider the estimator $\widehat{\alpha}_{0t}^{DSCM}(W)$ defined in (2.4) for any given $W \in \mathbb{R}^J$ so that the DID-estimator is just a special case. Ferman and Pinto (2021)[Proposition 3] showed that their FP-based estimator $\widehat{\alpha}_{0t}^{FP}$ is more efficient (in the sense of weakly smaller asymptotic variance) than the $\widehat{\alpha}_{0t}^{DSCM}(\cdot)$ estimator, where

$$\widehat{\alpha}_{0t}^{FP} := \widehat{\alpha}_{0t}^{DSCM}\left(\widehat{W}^{FP}\right)$$

with

$$\widehat{W}^{FP} := \underset{W \in \widetilde{\Delta}_1}{\arg\min} \frac{1}{T_0} \sum_{t \in \mathcal{T}_0} \{y_{0t} - y_t'W - (\overline{y}_0 - \overline{y}'W)\}^2$$

We will begin by showing that the DBSCM estimator $\widehat{\alpha}_{0t}^{DBSCM}(f, \Lambda)$ defined in (3.5) is more efficient than the FP-based estimator for some penalty term $f(W, \Lambda)$ and parameter $\Lambda \geq 0$, and therefore also more efficient than the DID-based estimator under homoskedasticity or error terms and stationarity of λ_t. We begin with an assumption. which is consistent to Assumption 5 of Ferman and Pinto (2021), called the 'Stability in the pre- and post-treatment period' assumption.

Assumption 4. *For* $t \in \mathcal{T}_1$, $cov(\lambda_t, (\varepsilon_{0t}, \varepsilon_t) = 0$, $\mathbb{E}\lambda_t\lambda_t' = \Omega_0$ *and* $\mathbb{E}(\varepsilon_{0t}, \varepsilon_t')(\varepsilon_{0t}, \varepsilon_t')' = \sigma_\varepsilon^2 I_{J+1}$

We have the following result:

Theorem 3 (Efficiency of estimator)**.** *Under Assumptions 1–4 and* $\mathbb{E}[\lambda_t] = 0$*, if we define* $\overline{f}(\Lambda, W) := \Lambda(1 + W'W + (W'\iota - 1)^2)$ *and let* $\eta \geq 1$*, then the demeaned-block-synthetic-control method es-*

[13] The homoskedastic assumption is made for simplification of argument; it can be shown with some algebra that the inequality between the left-hand-side and right-hand-side in (3.6) still holds under general heteroskedastic errors.

timator $\widehat{\alpha}_{0t}^{DBSCM}(\overline{f}, \sigma_\varepsilon^2)$ *defined in* (3.5) *dominates both* $\widehat{\alpha}_{0t}^{FP}$ *and* $\widehat{\alpha}_{0t}^{DID}$ *in terms of asymptotic mean-squared-error (MSE).*

Note that the assumption of λ_t being is a stationary process automatically implies $\mathbb{E}[\lambda_t] = 0$ and $\mathbb{E}\lambda_t\lambda_t' = \Omega_0$ for every $t \in \mathcal{T}_0 \cup \mathcal{T}_1$. Theorem 3 tells us that under such stationarity, there is a special case for which our estimator has weakly smaller MSE. Furthermore, note that

$$\widehat{\alpha}_{0t}^{FP} \xrightarrow{p} \tau_{0t}\left(W_{-(J+1)}^{FP}\right) \qquad \text{and} \qquad \widehat{\alpha}_{0t}^{DBSCM}(f, \Lambda) \xrightarrow{p} \tau_{0t}\left(\overline{W}^0(\Lambda)\right)$$

where $W^{FP} = (W_{-(J+1)}^{FP}, W_{J+1}^{FP}) = Plim\widehat{W}^{FP}$ with $W_{-(J+1)}^{FP}$ denoting the first J terms (and omitting the $J+1$-th term) and

$$\tau_{0t}(W) := \alpha_{0t} + \lambda_t'(M_0 - MW) + \left(\varepsilon_{0t} + \sum_{j=1}^{J} W_j\varepsilon_{jt} - (1 - \sum_{j=1}^{J} W_j)\varepsilon_{J+1,t}\right) \tag{3.7}$$

When under non-stationarity of λ_t (so that $\mathbb{E}\lambda_t \neq 0$), then $\widehat{\alpha}_{0t}^{FP}$ and $\widehat{\alpha}_{0t}^{DBSCM}(f, \Lambda)$ is biased for any fixed $\Lambda > 0$ unless their respective weights cancel out the second term on the right-side of $\tau_{0t}(W)$ (i.e. unbiased only if $M_0 - MW = 0$). Note that the FP weights $\widehat{W}^{FP}$ come from

$$\underset{W \in \mathbb{R}^J : \Delta_1^J \cap W_i \geq 0}{\arg\min} \quad \mathcal{B}(\overline{f}, W, \sigma_\varepsilon^2) \tag{3.8}$$

where $\overline{f}(\cdot)$ is defined as in Theorem 3 and $\mathcal{B}(\cdot,\cdot,\cdot)$ is defined in Theorem 2, under homoskedastic error[14] so that in general the weights W^{FP} do not solve $MW = M_0$ even when such weights exist (i.e. $\exists W^* \in \Phi$), due to the fact that σ_ε^2 is present in the minimization problem of (3.8), leading $\widehat{\alpha}_{0t}^{FP}$ to be a biased estimator for α_{0t}; to see this, observe that WPA1,

$$\mathbb{E}[\widehat{\alpha}_{0t}^{FP}] = \mathbb{E}[\tau_{0t}(W_{-(J+1)}^{FP})] = \alpha_{0t} + \mathbb{E}[\lambda_t]'(M_0 - MW_{-(J+1)}^{FP}) \neq \alpha_{0t}$$

In spite of this, the non-stationarity of λ_t allows us to exploit the positive-definiteness of Ω_0, which by Theorem 2 allows us to recover $\overline{W}^0(0)$ whenever PPTF holds (i.e. $M_0 - M\overline{W}^0(0) = 0$). This implies that $\widehat{\alpha}_{0t}^{DBSCM}(f, \Lambda_T)$ is an unbiased estimator of α_{0t} under non-stationarity for any sequence of $\Lambda_T \downarrow 0$. It is also clear that even under stationarity, $\widehat{\alpha}_{0t}^{DBSCM}(f, \Lambda_T)$ remains an unbiased estimator (by $\mathbb{E}\lambda_t = 0$ and (3.5)). Fornally, we have the following.

Corollary 3.3. *Under assumptions 1–3 and assuming f satisfies 1–4, if either (i) λ_t is stationary or (ii) λ_t is non-stationary, Ω_0 is positive-definite and there exists some $\overline{W} \in \Phi$ (i.e. there exists some perfect pre-treatment fit), then $\widehat{\alpha}_{0t}^{DBSCM}(f, \Lambda_T)$ is an unbiased estimator of α_{0t} for any sequence of* $0 < \Lambda_T \downarrow 0$

[14] The assumption of homoskedastic error is made in order to simplify the arguments. Under heteroskedastic error the argument will still hold

So far we we have seen that our estimator $\widehat{\alpha}_{0t}^{DBSCM}(f, \Lambda_T)$ is an unbiased estimator for α_{0t} whenever we have either (i) Perfect Pre-Treatment Fit or (i) Imperfect Pre-Treatment Fit with $\{\lambda_t\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$ being a stationary process. However, when we have Imperfect Pre-Treatment Fit under non-stationary λ_t, then it is not possible to obtain an unbiased estimator. Intuitively, this is because unbiasedness requires the existence of some W^* such that the term $\mathbb{E}\left[\lambda_t'\right](M_0 - MW^*) = 0$ of (3.7); this is satisfied if either $\mathbb{E}[\lambda]_t = 0$ or $M_0 - MW^* = 0$. Under Imperfect Fit, $M_0 - MW^* \neq 0$ for any W^*, yet the lack of stationarity property from λ_t prevents us from estimating $\mathbb{E}[\lambda_t]$ for any fixed $t \in \mathcal{T}_1$ since we cannot "learn" from previous observations in $\mathcal{T}_0$. Despite this impossibility, we have shown that (A) existing estimators may be biased under the setting of (ii) in Corollary 3.3 and (B) our estimator is at least as efficient as existing estimators, meriting the use of the DBSCM estimator over other existing estimators.

Recall from Theorem 3 that the DBSCM estimator is efficient under some penalty term and fixed $\Lambda \equiv \sigma_\varepsilon^2$, whenever we have homoskedastic error and stationarity of λ_t. However, note that in general, for any f satisfying conditions 1–4,

$$\begin{aligned} avar\left(\widehat{\alpha}_{0t}^{DBSCM}(f, \Lambda_T))\right) &= \mathcal{B}(f, \overline{W}^0(0), 0) + \sigma_\varepsilon^2(1 + \overline{W}^0(0)'\overline{W}^0(0)) \\ &\geq \mathcal{B}(f, \overline{W}^0(\sigma_\varepsilon^2), 0) + \sigma_\varepsilon^2(1 + \overline{W}^0(\sigma_\varepsilon^2)'\overline{W}^0(\sigma_\varepsilon^2)) = avar\left(\widehat{\alpha}_{0t}^{DBSCM}(\overline{f}, \sigma_\varepsilon^2))\right) \end{aligned}$$

so that $\widehat{\alpha}_{0t}^{DBSCM}(f, \Lambda_T)$ may not necessarily be more efficient than either $\widehat{\alpha}_{0t}^{FP}$ or $\widehat{\alpha}_{0t}^{DID}$. There is therefore a bias-variance tradeoff in that the "cost" to obtain an unbiased estimator $\widehat{\alpha}_{0t}^{DBSCM}(f, \Lambda_T)$ is an increase in variance. We summarize this in Table 3.

	Asymptotic Mean	Asymptotic Variance
$\widehat{\alpha}^{DBSCM}(f, \Lambda_T)$		
λ_t stationary, PPTF	α_{0t}	$\sigma_\varepsilon^2(1+\overline{W}^0(0)'\overline{W}^0(0))$
λ_t stationary, IPTF	α_{0t}	$\sigma_\varepsilon^2(1+\overline{W}^0(0)'\overline{W}^0(0))+\mathcal{B}(\overline{W}^0(0))$
λ_t non-stationary, PPTF	α_{0t}	$\sigma_\varepsilon^2(1+\overline{W}^0(0)'\overline{W}^0(0))$
λ_t non-stationary, IPTF	$\alpha_{0t}+\mathbb{E}[\lambda_t]'\left(M_0-M\overline{W}^0(0)\right)$	$\sigma_\varepsilon^2(1+\overline{W}^0(0)'\overline{W}^0(0))+\mathcal{B}(\overline{W}^0(0))$
$\widehat{\alpha}^{DBSCM}(\overline{f}, \sigma_\varepsilon^2)$		
λ_t stationary, PPTF	α_{0t}	$\sigma_\varepsilon^2(1+\overline{W}^0(\sigma_\varepsilon^2)'\overline{W}^0(\sigma_\varepsilon^2))+\mathcal{B}(\overline{W}^0(\sigma_\varepsilon^2))$
λ_t stationary, IPTF	α_{0t}	$\sigma_\varepsilon^2(1+\overline{W}^0(\sigma_\varepsilon^2)'\overline{W}^0(\sigma_\varepsilon^2))+\mathcal{B}(\overline{W}^0(\sigma_\varepsilon^2))$
λ_t non-stationary, PPTF	$\alpha_{0t}+\mathbb{E}[\lambda_t]'\left(M_0-M\overline{W}^0(\sigma_\varepsilon^2)\right)$	$\sigma_\varepsilon^2(1+\overline{W}^0(\sigma_\varepsilon^2)'\overline{W}^0(\sigma_\varepsilon^2))+\mathcal{B}(\overline{W}^0(\sigma_\varepsilon^2))$
λ_t non-stationary, IPTF	$\alpha_{0t}+\mathbb{E}[\lambda_t]'\left(M_0-M\overline{W}^0(\sigma_\varepsilon^2)\right)$	$\sigma_\varepsilon^2(1+\overline{W}^0(\sigma_\varepsilon^2)'\overline{W}^0(\sigma_\varepsilon^2))+\mathcal{B}(\overline{W}^0(\sigma_\varepsilon^2))$
$\widehat{\alpha}_{0t}^{FP}$		
λ_t stationary, PPTF	α_{0t}	$\sigma_\varepsilon^2(1+(W^{FP})'W^{FP})+\mathcal{B}(W^{FP})$
λ_t stationary, IPTF	α_{0t}	$\sigma_\varepsilon^2(1+(W^{FP})'W^{FP})+\mathcal{B}(W^{FP})$
λ_t non-stationary, PPTF	$\alpha_{0t}+\mathbb{E}[\lambda_t]'\left(M_0-MW^{FP}(\sigma_\varepsilon^2)\right)$	$\sigma_\varepsilon^2(1+(W^{FP})'W^{FP})+\mathcal{B}(W^{FP})$
λ_t non-stationary, IPTF	$\alpha_{0t}+\mathbb{E}[\lambda_t]'\left(M_0-MW^{FP}(\sigma_\varepsilon^2)\right)$	$\sigma_\varepsilon^2(1+(W^{FP})'W^{FP})+\mathcal{B}(W^{FP})$
$\widehat{\alpha}_{0t}^{DID}$		
λ_t stationary, PPTF	α_{0t}	$\sigma_\varepsilon^2(1+(W^{DID})'W^{DID})+\mathcal{B}(W^{DID})$
λ_t stationary, IPTF	α_{0t}	$\sigma_\varepsilon^2(1+(W^{DID})'W^{DID})+\mathcal{B}(W^{DID})$
λ_t non-stationary, PPTF	$\alpha_{0t}+\mathbb{E}[\lambda_t]'\left(M_0-MW^{DID}(\sigma_\varepsilon^2)\right)$	$\sigma_\varepsilon^2(1+(W^{DID})'W^{DID})+\mathcal{B}(W^{DID})$
λ_t non-stationary, IPTF	$\alpha_{0t}+\mathbb{E}[\lambda_t]'\left(M_0-MW^{DID}(\sigma_\varepsilon^2)\right)$	$\sigma_\varepsilon^2(1+(W^{DID})'W^{DID})+\mathcal{B}(W^{DID})$

Table 3: Summary of different tests. We write $\mathcal{B}(\cdot) \equiv \mathcal{B}(f, \cdot, 0)$

3.3 Inference for Average Treatment of the Treated

In the previous sections we discussed how we can obtain an unbiased estimator even under non-stationarity, as long as PPTF exists. In this section we discuss how to make inference on $\alpha_0 = \{\alpha_{0t}\}_{t\in\mathcal{T}_1}$. Unfortunately, if $T_1 := card(\mathcal{T}_1)$ does not diverge to infinity, then the usual method of applying some normalization to the test statistic in order to obtain a central-limit-theorem to conduct inference would fail. To overcome this, we appeal to conformal-inference.

Suppose first that we have an estimator $\widetilde{W}$ of some sort, such that $\widetilde{W} \xrightarrow{P} \overline{W}$, with $\overline{W} \in \Phi$. Then recall from (3.4) and (3.7) that as $T_0 \to \infty$, given some estimator $\widetilde{W}$, for any fixed $t \in \mathcal{T}_1$,

$$\widehat{\alpha}_{0t}^{DSCM}(\widetilde{W}) \xrightarrow{p} \alpha_{0t} + \lambda_t\left(M_0 - M'\overline{W}\right) + \left(\varepsilon_{0t} + \sum_{j=1}^{J}\overline{W}_j\varepsilon_{jt} - (1-\sum_{j=1}^{J}\overline{W}_j)\varepsilon_{J+1,t}\right)$$

Following the notations of Chernozhukov et al. (2022), 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

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

For notational simplicity, define $\varepsilon_t := (\varepsilon_{1t}, ..., \varepsilon_{Jt})'$ and $y_t := (y_{1t}, ..., y_{Jt})'$. Then for any $t \in \mathcal{T}_0 \cup \mathcal{T}_1$, define

$$\begin{aligned}
&\widehat{P}_t^N(\widetilde{W}) := \widehat{\alpha}_{0t}^{DSCM}(\widetilde{W}) - \alpha_t =: -\widehat{v}_t(\widetilde{W}) \\
&P_t^N := Y_{0t}^N - \overline{W}' Y_t^N - \left(\overline{Y}_0^N - \overline{Y}^N \overline{W}\right) + \left(\overline{\lambda}'(M_0 - M\overline{W}) + (\overline{u}_0 - \overline{W}'\overline{u})\right) =: -v_t \\
&Y_{0t}^N = \begin{cases} Y_{0t} - \alpha_{0t} & \text{for } \; t \in \mathcal{T}_1 \\ Y_{0t} & \text{for } \; t \in \mathcal{T}_0 \end{cases}
\end{aligned}$$

where $\{\alpha_t\}_{t \in \mathcal{T}_1}$ is the hypothesized value of the treatment effect $\{\alpha_{0t}\}_{t \in \mathcal{T}_1}$ and $\widehat{\alpha}_{0t}^{DSCM}(\cdot)$ was defined in (3.4). By convention, we define $\alpha_t \equiv \alpha_{0t} \equiv 0$ for every $t \in \mathcal{T}_0$. Then define the p-value as

$$\widehat{p}(\widetilde{W}) := \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{v}_\pi(\widetilde{W})) \geq S(\widehat{v}(\widetilde{W}))\} \tag{3.9}$$

where

$$\begin{aligned}
S(\widehat{v}(\widetilde{W})) &:= \frac{|\sum_{t=T_0+1}^{T} \widehat{v}_t(\widetilde{W})|}{\sqrt{card(\mathcal{T}_1)}}, \\
\widehat{v}(\widetilde{W}) &= (\widehat{v}_1(\widetilde{W}), ..., \widehat{v}_T(\widetilde{W}))'
\end{aligned}$$

and $\widehat{v}_\pi(\widetilde{W}) = (\widehat{v}_{\pi(1)}(\widetilde{W}), ..., \widehat{v}_{\pi(T)}(\widetilde{W}))$ is the permutation of $\widehat{v}(\widetilde{W})$ by $\pi \in \Pi$. Then we have the following result.

Theorem 4. *Suppose $\widetilde{W} \xrightarrow{p} \overline{W}$ and assumptions 1–3 holds. If either*

1. *$\{\lambda_t\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$ is stationary*
2. *$\{\lambda_t\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$ is non-stationary and $\overline{W} \in \Phi$*

holds, then under the correct null of $\{\alpha_t\}_{t \in \mathcal{T}_1} = \{\alpha_{0t}\}_{t \in \mathcal{T}_1}$, for any $\theta \in (0, 1)$,

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

Theorem 2 and Corollary 3.3 gives us the immediate results, with states that the p-value based on the DBSCM-derived weights $\widetilde{W}_T^{DBSCM}(f, \Lambda_T)$ yield asymptotically valid results under the null.

Corollary 3.4. *Suppose assumptions 1–3 holds. If either*

1. $\{\lambda_t\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$ *is stationary*

2. $\{\lambda_t\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$ *is non-stationary,* Ω_0 *is positive-definite and* $\exists W^*\in\Phi$,

then under the correct null of $\{\alpha_t\}_{t\in\mathcal{T}_1}=\{\alpha_{0t}\}_{t\in\mathcal{T}_1}$ *and any* f *satisfying conditions 1–4, for large enough* $\eta>0$ *and any* $\theta\in(0,1)$,

$$\left|\mathbb{P}\left(\widehat{p}(\widetilde{W}_T^{DBSCM}(f,\Lambda_T)\right)\leq\theta)-\theta\right|=o_p(1)$$

for any sequence of $\Lambda_T\downarrow 0$.

Remark 3. *Under the conditions of Theorem 4.2, the DID and FP-based weights* $\widehat{W}^{FP}$ *and* $\widehat{W}^{DID}$ *defined in section 3.2.2 will generally not have size-control, i.e.* $\mathbb{P}(\widehat{p}(W)\leq\theta)\not\to\theta$ *for* $W\in\{\widehat{W}^{FP},\widehat{W}^{DID}\}$. *The reason is that they are biased estimators of* α_{0t} *in such a scenario (see Table 3).*

4 Simulation

We consider the factor model considered in Chernozhukov et al. (2021)[section G]. The model can be written as

$$\begin{aligned}y_{0t}&=y_t'W+u_t\quad\text{for } t\in\mathcal{T}_0\\ y_{0t}&=\alpha_{0t}+y_t'W+u_t\quad\text{for } t\in\mathcal{T}_1\end{aligned}$$

where $u_t=\rho_u u_{t-1}+v_t$, $v_t\overset{iid}{\sim}\mathcal{N}(0,1-\rho_u^2)$. The DGP of the y_t are given in (2.1), with $c_j=j/(J+1)=\mu_j$, $\delta_t\overset{iid}{\sim}\mathcal{N}(0,1)$, $\varepsilon_{jt}=\rho_\varepsilon\varepsilon_{j,t-1}+\xi_{jt}$, $\xi_{jt}\overset{iid}{\sim}\mathcal{N}(0,1-\rho_\varepsilon^2)$, $\lambda_t\overset{iid}{\sim}\mathcal{N}(t,1)$. We vary $\rho_u,\rho_\varepsilon,T_0$ and J, and set $|\mathcal{T}_1|=1$. Notice that $\mathbb{E}\lambda_t\neq 0$. The four DGPs for weight W that we consider are

	Weight specification	Correctly specified model
DGP1	$W=W_*$	DID, SC, Constrained Lasso
DGP2	$W=(\frac{1}{3},\frac{1}{3},\frac{1}{3},0,...,0)'$	SC, Constrained Lasso
DGP3	$W=-W_*$	Constrained Lasso
DGP4	$W=(1,-1,0,...,0)'$	own specification

where $W_*:=(\frac{1}{J},...,\frac{1}{J})'$. We choose $\Lambda=0.01$. We vary $J=20,50,100$, $T_0=20,50,100$, $\rho_u=\rho_\varepsilon=0,0.6$ for the $i.i.d.$ case or the dependent case respectively. We choose $r=2$ and obtain

DGP3	$\widetilde{W}_T^{DBSCM}(f, \Lambda_T)$	DID	SCM	Constrained-LASSO	FP
$\theta = \mathbf{0.05}$	0.061	0.716	0.522	0.112	0.547
$\theta = \mathbf{0.1}$	0.154	0.966	0.831	0.217	0.863

Table 4: size control for 1,000 replications: $T_0 = 50, J = 20, \Lambda = 0.01, \rho_u = \rho_\varepsilon = 0.6, r = 2$

where we let f be the ℓ_2-regularized penalty term and $\Lambda_T = 0.01$, i.e. a very small term. We run the test of

$$\mathbb{P}_n(\widehat{p}(W) \leq \theta)$$

for W equal to one of the weights in the top row of Table 4, where $\mathbb{P}_n$ is the empirical distribution over $1,000$ replications and $\widehat{p}(\cdot)$ is defined in (3.9). This shows the good property of our test.

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.

Abadie, A., A. Diamond, and J. Hainmueller (2015). Comparative politics and the synthetic control method. *American Journal of Political Science 59*(2), 495–510.

Abadie, A. and J. Gardeazabal (2003, March). The economic costs of conflict: A case study of the basque country. *American Economic Review 93*(1), 113–132.

Amjad, M., D. Shah, and D. Shen (2018, jan). Robust synthetic control. *Journal of Machine Learning. Research 19*(1), 802â€"852.

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2021, December). Synthetic difference-in-differences. *American Economic Review 111*(12), 4088â€"4118.

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

Ben-Michael, E., A. Feller, and J. Rothstein (2021). The augmented synthetic control method. *Journal of the American Statistical Association 116*(536), 1789–1803.

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.

Chernozhukov, V., K. Wuthrich, and Y. Zhu (2022). A t-test for synthetic controls.

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.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis 25*(1), 57–76.

Contents

A Proof of Lemmas

A.1 Auxiliary Lemma

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 except possibly only its diagonals, with these values denoted as $\sigma_1, ..., \sigma_r$. 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,

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

$\square$

A.2 Proof of Lemma 3.1

Define $W_{-J+1} := (W_1, ..., W_J)'$ and $W := (W_1, ..., W_J, W_{J+1})'$. Then by some algebraic manipulation we can obtain

$$M_0 - MW_{-(J+1)} = (\mu_0 - \mu W) - (1 - \sum_{i=1}^{J+1} W_i)\mu_{J+1} \tag{A.1}$$

and

$$C_0 - C'W_{-(J+1)} = (c_0 - c'W) + c_{J+1}(W_{J+1} - 1) = (c_0 - c'W) \tag{A.2}$$

If there exists some $W^* = (W_1^*, ..., W_J^*, W_{J+1}^*)' \in \widetilde{\Phi}$, then by recalling that $\sum_{i=1}^{J+1} W_i^* = 1$ and observing both (A.1) and (A.2), we have that $\overline{W} := (W_1^*, ..., W_J^*)' \in \Phi$.

B Proof of Theorems

B.1 Proof of Theorem 1

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

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

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

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

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

$$\begin{aligned}
A_1 &\le \frac{1}{T_0\Delta}\sum_{t=1}^{T_0}\varepsilon_{0t}^2+\frac{1}{\Delta}W'(\frac{1}{T_0}\sum_{t=1}^{T_0}\varepsilon_t\varepsilon_t')W\\
&=\frac{1}{\Delta}(\Sigma_{11}+o_p(1))+\frac{1}{\Delta}W'(\Sigma+o_p(1))W=o_p(1)
\end{aligned}$$

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

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

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

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

where (iv) follows from

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

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

step 2: We write $\widetilde{\mathcal{A}}_{T_0}(W,\Lambda) \equiv \widetilde{\mathcal{A}}_{T_0}(f,W,\Lambda)$ where f is a known fixed function, for notational simplicity. Then define

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

and

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

we want to show that

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

First we require a lemma:

Lemma B.1. *(Corollary 2.2 of Newey (1991)) Assume (1) $\Delta_\eta^J\times[\gamma,1]$ is compact, (2) $\widetilde{\mathcal{A}}_{T_0}(W,\Lambda) \xrightarrow{p} \mathcal{A}(W,\Lambda)$ for every $(W,\Lambda)\in\Delta_\eta^J\times[\gamma,1]$, (3) $\Delta_\eta^J\times[\gamma,1]$ 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) $\{\mathcal{A}(W,\Lambda)\}_{(W,\Lambda)\in\Delta_\eta^J\times[\gamma,1]}$ is equi-continuous. Then $\widetilde{\mathcal{A}}_{T_0}(W,\Lambda) \xrightarrow{p} \mathcal{A}(W,\Lambda)$ uniformly over $(W,\Lambda)\in\Delta_\eta^J\times[\gamma,1]$*

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

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

so that by

$$(a) \quad \frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\lambda}^s)' = \Omega_0 + o_p(1)$$

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

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

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

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

where the last equality in (d) follows from Markov-inequality, the simple inequality that $\mathbb{E}(\lambda_k^s\lambda_{j,K'}^s) \leq 2\mathbb{E}(\lambda_{j,k}^s)^2 + 2\mathbb{E}(\lambda_{k'}^s)^2 \leq 4C$ 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 that

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

It is clear that $\Delta_\eta^J \times [0,1]$ is compact, so condition (1) of Lemma B.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 (B.3). To show (B.2), it remains to prove that condition (4) of Lemma B.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

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

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

$$\begin{aligned}
B_{T_0} &:= \left|\left|\left(\frac{2}{r}\sum_{s=1}^{r}\{\widetilde{y}_0^s - (\widetilde{y}^s)'W_3\}(-\widetilde{y}^s - \overline{\delta}^s\iota) + ||W_3||^2 + 2\Lambda_3 W_3\right)\right|\right| \\
&\leq \left|\left|\left(\frac{2}{r}\sum_{s=1}^{r}\{\widetilde{y}_0^s - (\widetilde{y}^s)'W_3\}(\widetilde{y}^s)\right)\right|\right| + \left|\left|\left(\frac{2}{r}\sum_{s=1}^{r}\{\widetilde{y}_0^s - (\widetilde{y}^s)'W_3\}\overline{\delta}^s\right)\right|\right|\cdot||\iota|| + ||W_3||^2 + 2\Lambda_3||W_3|| \\
&\leq \left|\left|\frac{2}{r}\sum_{s=1}^{r}\widetilde{y}_0^s\widetilde{y}^s\right|\right| + \left|\left|\frac{2}{r}\sum_{s=1}^{r}\widetilde{y}^s(\widetilde{y}^s)'\right|\right|\times||W_3|| + \left|\left|\frac{2}{r}\sum_{s=1}^{r}\widetilde{y}_0^s\overline{\delta}^s\right|\right| + \left|\left|\frac{2}{r}\sum_{s=1}^{r}\widetilde{y}^s\overline{\delta}^s\right|\right|\times||W_3||\times||\iota|| + \eta^2 + 2\eta \\
&= ||2A_1|| + ||2A_2||\times\eta + ||2A_3|| + ||2A_4||\times\sqrt{J}\eta + \eta^2 + 2\eta
\end{aligned}$$

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

$$(a)\quad \frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^s(\overline{\lambda}^s)' = \Omega_0$$

$$(b)\quad \frac{1}{r}\sum_{s=1}^{r}\overline{\varepsilon}^s(\overline{\varepsilon}^s)' = O_p(1)$$

$$(c) \quad \frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^{s} = \frac{1}{T_0}\sum_{t=1}^{T_0}\lambda_t = o_p(1)$$

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

$$(e) \quad \frac{1}{r}\sum_{s=1}^{r}\overline{\lambda}^{s}(\overline{\varepsilon}^{s})' = O_p(1)$$

$$(f) \quad \frac{1}{r}\sum_{s=1}^{r}\overline{\varepsilon}_0^{s}(\overline{\varepsilon}^{s})' = O_p(1)$$

$$(g) \quad \frac{1}{r}\sum_{s=1}^{r}\overline{\delta}^{s} = \frac{1}{T_0}\sum_{t=1}^{T_0}\delta_t = O_p(1)$$

$$(h) \quad \frac{1}{r}\sum_{s=1}^{r}\overline{\delta}^{s}(\overline{\lambda}^{s})' = O_p(1)$$

$$(i) \quad \frac{1}{r}\sum_{s=1}^{r}\overline{\delta}^{s}(\overline{\varepsilon}^{s})' = O_p(1)$$

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

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

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

Step 3: For notational simplicity, we write $\widetilde{W}(\Lambda) \equiv \widetilde{W}^{BSCM}(f, \Lambda)$. We show that for any fixed $\gamma > 0$,

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

where $\overline{W}(\Lambda) = \arg\min_{W\in\Delta_\eta^J}\mathcal{A}(W, \Lambda)$.

Lemma B.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 $\widetilde{W}(\Lambda)$ for every $\Lambda \in [\gamma, 1]$. 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 (for $\Lambda > 0$), it has a uniquely-minimized solution. Clearly $\mathcal{A}(W, \Lambda)$ is continuous in Θ and condition (iv) of Lemma B.2 follows from equation (B.2), with $\widehat{Q}(\theta)$ as $\widehat{\mathcal{A}}_{T_0}(W, \Lambda)$. Therefore we have

$$\widetilde{W}(\Lambda) \xrightarrow{p} \overline{W}(\Lambda) \tag{B.5}$$

for every fixed $\Lambda \in [\gamma, 1]$. This satisfies condition (2) of Lemma B.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 $\Lambda_1, \Lambda_2 \in [\gamma, 1]$,

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

where (i) follows from V being orthogonal. Condition (5) of Lemma B.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

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

where

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

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 B.1 is satisfied, so that we obtain (B.4).

Step 4: In the case where $cc' + \mu'\Omega_0\mu$ is positive definite, then (B.5) holds for $\Lambda = 0$ by $\mathcal{A}(0, W)$ being strictly convex. We can repeat the proof as in **step 3** with $\gamma = 0$ and show that for any

$\Lambda_1, \Lambda_2 \in [0, 1]$,

$$\begin{aligned}||\overline{W}(\Lambda_1) - \overline{W}(\Lambda_2)|| &\leq ||\mu'\Omega_0\mu_0 + c_0 c||_1 \max_{i=1,\dots,J} \frac{|\Lambda_1 - \Lambda_2|}{(d_i + \Lambda_1)(d_i + \Lambda_2)} \\ &\leq ||\mu'\Omega_0\mu_0 + c_0 c||_1 \max\{d_1^{-1}, \dots, d_J^{-1}\}|\Lambda_1 - \Lambda_2|\end{aligned}$$

where we note that $\max\{d_1^{-1}, ..., d_J^{-1}\} < \infty$ by $\{d_i\}_{i=1}^J$ being the eigenvalues of $cc' + \mu'\Omega_0\mu$, a positive-definite matrix. In this case (B.4) holds with $\gamma = 0$.

Step 5: We show that $\overline{W}(\Lambda) \to W^* \equiv \overline{W}(0)$ as $0 < \Lambda \downarrow 0$

For any $\Lambda > 0$, $\mathcal{A}(W, \Lambda)$ is a strictly convex and continuous function, so that $\overline{W}(\Lambda) := \arg\min_{W \in \Delta_\eta^J} \mathcal{A}(W, \Lambda)$ is unique.[15] Recall that W^* is the unique solution that minimizes $\mathcal{H}(W) := (c_0 - c'W)^2 + (\mu_0 - \mu W)'\Omega_0(\mu_0 - \mu W)$ over $W \in \Delta_\eta^J$; this uniqueness follows from Lemma A.1. Note that for any other $W \in \Delta_\eta^J$ with the property that $\mathcal{H}(W) = \mathcal{H}(W^*)$, it must be that $||W^*|| < ||W||$ by the uniqueness property.

For any $W^\dagger \in \Delta_\eta^J$ with $||W^\dagger|| \geq ||W^*||$, we have $\mathcal{A}(W^*, \Lambda) < \mathcal{A}(W^\dagger, \Lambda)$ since $\mathcal{A}(W, \Lambda) \equiv \mathcal{H}(W) + \Lambda||W||^2$. Therefore $\overline{W}(\Lambda) \neq W^*$ whenever $||\overline{W}(\Lambda)|| \leq ||W^*||$.

Furthermore, we know that any $W \in \Delta_\eta^J \backslash \{W^*\}$ such that $||W|| \leq ||W^*||$ has the property that $\mathcal{H}(W) > 0$, since W^* is the unique minimum-norm solution. Define $\Delta(W) := W - W^*$ and consider any fixed $\delta > 0$. Then consider the open ball around W^*, defined as $B_\delta(W^*) \equiv \{W : ||\Delta(W)|| < \delta\}$. Since $\widetilde{\Delta} := \{W \in \Delta_\eta^J : ||W|| \leq ||W^*||\}$ is compact, then $\widetilde{\Delta} \cap B_\delta^c(W^*)$ is compact, where $B_\delta^c(W^*)$ is the complement set to $B_\delta(W^*)$, which is closed. By Weierstrass extreme-value-theorem, there exists a $W^\ddagger \in \widetilde{\Delta} \cap B_\delta^c(W^*)$ with $\mathcal{H}(W^\ddagger) \equiv \inf_{W \in \widetilde{\Delta} \cap B_\delta^c(W^*)} \mathcal{H}(W)$. Note that $0 < \mathcal{H}(W^\ddagger)$ by $W^\ddagger \in \widetilde{\Delta}$. Hence there must be a $\bar{c}(\delta) > 0$ such that whenever Λ is chosen such that $0 < \Lambda < \bar{c}(\delta)$, then $\Lambda\eta < \mathcal{H}(W^\ddagger) - \mathcal{H}(W^*)$.[16] We will show that

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

First note that any $W \in \Delta_\eta^J \backslash \{\widetilde{\Delta}\}$ cannot possibly minimize $\mathcal{A}(W, \Lambda)$, since W^* can always be chosen. Therefore $\overline{W}(\Lambda) \in \widetilde{\Delta}$. Second, by contradiction assume $\overline{W}(\Lambda) \in B_\delta^c(W^*)$. Then

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

which cannot be true since $\overline{W}(\Lambda)$ is the unique minimizer of $\mathcal{A}(W, \Lambda)$. Thus $\overline{W}(\Lambda) \in B_\delta(W^*)$, proving (B.6). Finally, we see that $\bar{c}(\delta) \downarrow 0$ as $\delta \downarrow 0$ because $\inf_{W \in \widetilde{\Delta} \cap B_\delta^c(W^*)} \mathcal{H}(W)$ is non-increasing with δ. This implies that as $0 < \Lambda \downarrow 0$, $\overline{W}(\Lambda) \to W^*$.

[15] The strict-convexity follows from $f(W, \Lambda)$ being strictly-convex for $\Lambda > 0$

[16] Note $\mathcal{H}(W^\ddagger) > \mathcal{H}(W^*)$, because $W^\ddagger \in \widetilde{\Delta}$.

B.2 Proof of Theorem 2

The proof follows in the exact same way as the proof of Theorem 1. For the sake of completeness, the only difference in proof comes from (B.3), i.e. we will instead show that

$$\widetilde{\mathcal{B}}_T(f, W, \Lambda) \xrightarrow{p} \mathcal{B}(f, W, \Lambda) \tag{B.7}$$

for any fixed $(W, \Lambda) \in \Delta_\eta^J \times [0, 1]$,

$$\widetilde{\mathcal{B}}_T(f, W, \Lambda) := \frac{1}{r} \sum_{s=1}^{r} \{\overline{y}_0^s - (\overline{y}^s)'W - (\overline{y}_0 - \overline{y}'W)\}^2 + f(W, \Lambda)$$

Using the notations of **Step 3** in Theorem 1, we can express

$$\begin{aligned}\widetilde{\mathcal{B}}_T(f, W, \Lambda) &= \widetilde{\mathcal{A}}_T(f, W, \Lambda) - (\overline{y}_0 - \overline{y}'W)^2 \overset{(i)}{=} \mathcal{A}(f, W, \Lambda) - (\overline{y}_0 - \overline{y}'W)^2 + o_p(1)\\ &\overset{(ii)}{=} \mathcal{A}(f, W, \Lambda) - (c_0 - c'W)^2 + o_p(1) = \mathcal{B}(f, W, \Lambda) + o_p(1),\end{aligned}$$

where (i) follows from (B.3) and (ii) follows from

$$(\overline{y}_0 - \overline{y}'W)^2 = \left((C_0 - C'W) + \overline{\lambda}(M_0 - MW) + (\overline{u}_0 - \overline{u})\right)^2 = (C_0 - C'W)^2 + o_p(1)$$

B.3 Proof of Theorem 3

The proof that $\widehat{\alpha}_{0t}^{FP}$ is more efficient than $\widehat{\alpha}_{0t}^{DID}$ is shown in Ferman and Pinto (2021)[Proposition 3]. It remains to show that $\widehat{\alpha}_{0t}^{DBSCM}$ is asymptotically more efficient than $\widehat{\alpha}_{0t}^{FP}$, which we do now. Observe

$$\begin{aligned}\widehat{\alpha}_{0t}^{FP} &\xrightarrow{p} \alpha_{0t} + \lambda_t'(\mu_0 - \mu W^{FP}) + (\varepsilon_{0t} - \varepsilon_t' W^{FP})\\ &= \alpha_{0t} + \lambda_t'(M_0 - MW_{-(J+1)}^{FP}) + (\varepsilon_{0t} - \varepsilon_t' W^{FP})\end{aligned}$$

where $W^{FP} = (W_{-(J+1)}^{FP}, W_{J+1}^{FP}) = Plim \widehat{W}^{FP}$ and the last equation follows from $\sum_{j=1}^{J+1} W_j^{FP} = 1$. The asymptotic variance is therefore

$$\begin{aligned}&avar(\widehat{\alpha}_{0t}^{FP}) = (M_0 - MW_{-(J+1)}^{FP})'\Omega_0(M_0 - MW_{-(J+1)}^{FP}) + \sigma_\varepsilon^2(1 + ||W^{FP}||_F^2)\\ &= (M_0 - MW_{-(J+1)}^{FP})'\Omega_0(M_0 - MW_{-(J+1)}^{FP}) + \sigma_\varepsilon^2 \left(1 + (W_{-(J+1)}^{FP})'W_{-(J+1)}^{FP} + (\sum_{j=1}^{J} W_j^{FP} - 1)^2\right)\\ &= \mathcal{B}(W_{-(J+1)}^{FP}, \sigma_\varepsilon^2)\end{aligned}$$

where

$$\mathcal{B}(W, \Lambda) := (M_0 - MW)'\Omega_0(M_0 - MW) + \Lambda\left(1 + W'W + (W'\iota - 1)^2\right).$$

Furthermore, by (3.5) and Assumption 4 we have that

$$\widehat{\alpha}_{0t}^{DBSCM}(\overline{f}, \sigma_\varepsilon^2) \xrightarrow{p} \alpha_{0t} + \lambda_t'(M_0 - M\overline{W}^0(0)) + (u_{0t} - u_t'\overline{W}^0(0))$$

$$= \alpha_{0t} + \lambda_t'(M_0 - M\overline{W}^0(0)) + (\varepsilon_{0t} - \varepsilon_t'\overline{W}^0(0) + (\sum_{j=1}^{J}\overline{W}_j^0(0) - 1)\varepsilon_{J+1,t})$$

so that

$$\begin{aligned}
&avar(\widehat{\alpha}_{0t}^{DBSCM}(\overline{f}, \sigma_\varepsilon^2)) \\
&= (M_0 - M\overline{W}^0(\sigma_\varepsilon^2))'\Omega_0(M_0 - M\overline{W}^0(\sigma_\varepsilon^2)) + \sigma_\varepsilon^2\left(1 + (\overline{W}^0(\sigma_\varepsilon^2))'\overline{W}^0(\sigma_\varepsilon^2) + (\sum_{j=1}^{J}\overline{W}_j^0(\sigma_\varepsilon^2) - 1)^2\right) \\
&\equiv \mathcal{B}(\overline{W}^0(\sigma_\varepsilon^2), \sigma_\varepsilon^2)
\end{aligned}$$

By the fact that $\overline{W}^0(\sigma_\varepsilon^2) \in \arg\min_{W\in\Delta_\eta^J}\mathcal{B}(W, \sigma_\varepsilon^2)$, we have that $\mathcal{B}(\overline{W}^0(\sigma_\varepsilon^2), \sigma_\varepsilon^2) \leq \mathcal{B}(W_{-(J+1)}^{FP}, \sigma_\varepsilon^2)$.

B.4 Proof of Theorem 4

We begin by fixing any $t \in \mathcal{T}_0 \cup \mathcal{T}_1$ and writing

$$\begin{aligned}
0 &= P_t^N + v_t \\
\alpha_t &= P_t^N + \alpha_{0t} + v_t
\end{aligned}$$

under the correct null, where $\alpha_{0t} \equiv 0$ for every $t \in \mathcal{T}_1$. Note that

$$\begin{aligned}
P_t^N \equiv -v_t &= \left\{(C_0 - C'\overline{W}) + \lambda_t'(M_0 - \overline{W}'M) + (u_{0t} - \overline{W}'u_t)\right\} \\
&\quad - \left\{(C_0 - C'\overline{W}) + \overline{\lambda}'(M_0 - \overline{W}'M) + (\overline{u}_0 - \overline{W}'\overline{u})\right\} + \left(\overline{\lambda}'(M_0 - M\overline{W}) + (\overline{u}_0 - \overline{W}'\overline{u})\right) \\
&= \lambda_t'(M_0 - \overline{W}'M) + (u_{0t} - \overline{W}'u_t)
\end{aligned}$$

Then under either case 4.1 or 4.2,

$$v_t \text{ is a mean-zero stationary process} \tag{B.8}$$

since $\{(u_{0t}, u_t)\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$ is stationary by assumption 3. Furthermore, under the correct null we have

$$P_t^N - \widehat{P}_t^N = -\lambda_t'(\overline{W} - \widetilde{W})'M) - (\overline{W} - \widetilde{W})'u_t + \left(\overline{\lambda}'(M_0 - M\widetilde{W}) + (\overline{u}_0 - \widetilde{W}'\overline{u})\right)$$

Therefore, writing $\widehat{P}^N := (\widehat{P}_1^N, ..., \widehat{P}_T^N)$ and $P^N := (P_1^N, ..., P_T^N)$, by noting the simple inequality of $(a + b + c + d)^2 \leq 8(a^2 + b^2 + c^2 + d^2)$,

$$\begin{aligned}
&||\widehat{P}^N - P^N||_2^2/T = \frac{1}{T}\sum_{t=1}^{T}(\widehat{P}_t^N - P_t^N)^2 \\
&\leq \frac{8}{T}\sum_{t=1}^{T}\left\{||M'(\overline{W} - \widetilde{W})||_2^2||\lambda_t||_2^2 + ||\overline{W} - \widetilde{W}||_2^2||u_t||_2^2 + ||\overline{\lambda}||_2^2||M_0 - M\widetilde{W}||_2^2 + ||\overline{u}_0 - \widetilde{W}'\overline{u}||_2^2\right\}
\end{aligned}$$

$$\stackrel{(i)}{=} o_p(1) \cdot \frac{8}{T} \sum_{t=1}^{T} \left\{ ||\lambda_t||_2^2 + ||u_t||_2^2 \right\} + o_p(1) + o_p(1) \stackrel{(ii)}{=} o_p(1) O_p(1) + o_p(1) + o_p(1) = o_p(1) \tag{B.9}$$

where (i) follows from $\widetilde{W} \xrightarrow{p} \overline{W}$ and assumption 3; (ii) follows from assumption 3. Finally, for any $t_1 \in \mathcal{T}_1$,

$$(\widehat{P}_{t_1}^N - P_{t_1}^N)^2 \leq 2||M'(\overline{W} - \widetilde{W})||_2^2 ||\lambda_{t_1}||_2^2 + 2||\overline{W} - \widetilde{W}||_2^2 ||u_t||_2^2 = o_p(1) O_p(1) = o_p(1). \tag{B.10}$$

Equations (B.8), (B.9) and (B.10) satisfy Assumptions 1,2 and 3 of Chernozhukov et al. (2021), so that an application of Chernozhukov et al. (2021)[Theorem 1] yields the result.

C Proof of Corollaries

C.1 proof of corollary 3.1

By Theorem 1, there exists a $\Lambda(\xi) > 0$ such that for any $0 < \Lambda \leq \Lambda(\xi)$, we have $|\overline{W}(\Lambda) - \overline{W}(0)| \leq \xi$. Define $\gamma := \Lambda$. Then by triangle inequality,

$$|\widetilde{W}^{BSCM}(f, \Lambda) - \overline{W}(0)| \leq |\widetilde{W}^{BSCM}(f, \Lambda) - \overline{W}(\Lambda)| + |\overline{W}(\Lambda) - \overline{W}(0)| \leq o_p(1) + \xi$$

so that the result is shown.

C.2 Proof of corollary 3.2

Consider any positive decreasing sequence $(\xi_m)_{m=1}^{\infty}$ that converges to 0. By corollary 3.1, for ξ_1, there is some $m_0(\xi_1) \in \mathbb{N}$ and $\Lambda(\xi_1) > 0$ such that for any $T \geq m_0(\xi_1)$, with probability at least $1 - \xi_1$ we have

$$|\widetilde{W}_T^{BSCM}(f, \Lambda(\xi_1)) - \overline{W}(0)| \leq \xi_1$$

Moving to ξ_2, there exists $m_0(\xi_2) > m_0(\xi_1)$ and $\Lambda(\xi_2) > 0$ such that for any $T \geq m_0(\xi_2)$, with probability at least $1 - \xi_2$,

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

We can express this recursively, 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 first part of the result holds.

For the second part, when Ω_0 is positive-definite, simply apply Theorem 1 and note that γ can be taken to be zero in this case, i.e. for any sequence of $\Lambda_T \downarrow 0$,

$$\begin{aligned}
\left|\widetilde{W}_T^{BSCM}(f, \Lambda_T) - \overline{W}(0)\right| &\leq \left|\widetilde{W}_T^{BSCM}(f, \Lambda_T) - \overline{W}(\Lambda_T)\right| + \left|\overline{W}(\Lambda_T) - \overline{W}(0)\right| \\
&\leq \sup_{\Lambda \in [0,1]} \left|\widetilde{W}_T^{BSCM}(f, \Lambda) - \overline{W}(\Lambda)\right| + \left|\overline{W}(\Lambda_T) - \overline{W}(0)\right| \\
&= o_p(1) + o(1) = o_p(1)
\end{aligned}$$

C.3 Proof of Corollary 3.3

(i) follows from (3.5), noting $\mathbb{E}[\lambda_t] = 0$ under the stationary case. (ii) follows from Theorem 2 and (3.5), which implies that $\overline{W}^0(0) \in \Phi$, i.e. $M_0 - M\overline{W}^0(0) = 0$.

C.4 Proof of Corollary 3.4

By Theorem 2 we know that $\widetilde{W}_T^{DBSCM}(f, \Lambda_T) \xrightarrow{p} \overline{W}^0(0)$, so that by Theorem 4 it suffices to show that $\overline{W}^0(0) \in \Phi$ whenever $\exists W^* \in \Phi$, but this follows again from Theorem 2.