Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls

Bruno Ferman

To cite this article: Bruno Ferman (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, DOI: 10.1080/01621459.2021.1965613

To link to this article: https://doi.org/10.1080/01621459.2021.1965613

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 116, NO. 536, 1764–1772: Theory and Methods Special Section on Synthetic Control Methods
https://doi.org/10.1080/01621459.2021.1965613

On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls

Bruno Ferman

Sao Paulo School of Economics-FGV, Sao Paulo, Brazil

ABSTRACT

We consider the asymptotic properties of the synthetic control (SC) estimator when both the number of pretreatment periods and control units are large. If potential outcomes follow a linear factor model, we provide conditions under which the SC unit asymptotically recovers the factor structure of the treated unit, even when the pretreatment fit is imperfect. This happens when there are weights diluted among an increasing number of control units such that a weighted average of the factor structure of the control units asymptotically reconstructs the factor structure of the treated unit. In this case, the SC estimator is asymptotically unbiased even when treatment assignment is correlated with time-varying unobservables. Supplementary materials for this article, including a standardized description of the materials available for reproducing the work, are available as an online supplement.

ARTICLE HISTORY
Received October 2019
Accepted August 2021

KEYWORDS
Comparative studies; Counterfactual analysis; Factor models; Panel data; Policy evaluation

1. Introduction

The synthetic control (SC) estimator, proposed in a series of influential papers by Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010), and Abadie, Diamond, and Hainmueller (2015), quickly became one of the most popular methods for policy evaluation (Athey and Imbens 2017). An important advantage of the SC method is that it can potentially allow for correlation between treatment assignment and time-varying unobserved covariates. Assuming a perfect pretreatment fit condition, Abadie, Diamond, and Hainmueller (2010) showed that the bias of the SC estimator is bounded by a function that asymptotes to zero when the number of pretreatment periods increases and the number of control units is fixed. However, when the perfect pretreatment fit condition is relaxed and the number of control units is fixed, Ferman and Pinto (2021) showed that the SC estimator is generally biased when there are unobserved confounders. In settings where the number of control units and pretreatment periods are both large, there are alternative methods, many of them based on the original SC estimator, that allow for selection on time-varying unobservables (e.g., Arkhangelsky et al. 2021; Athey et al. 2021; Gobillon and Magnac 2016; Bai 2009; Xu 2017). However, the properties of the original SC estimator—which remains commonly used in empirical applications—when both the number of pretreatment periods and control units go to infinity received less attention.

In this article, we consider the asymptotic properties of the SC estimator when both the number of pretreatment periods and the number of control units increase, in a setting in which the pretreatment fit is imperfect. Considering a linear factor model structure for potential outcomes, we derive conditions under which the SC unit (which is a weighted average of the control units) asymptotically recovers the factor structure of the treated unit. We show that this will be the case when, as the number of control units goes to infinity, there are weights diluted among an increasing number of control units that (asymptotically) recover the factor structure of the treated unit. This holds even in settings in which the number of control units is at the same magnitude or even larger than the number of pretreatment periods, which is common in SC applications (e.g., Doudchenko and Imbens 2016).

The intuition is the following. Ferman and Pinto (2021) showed that, in a setting with a fixed number of control units and imperfect pretreatment fit, the SC weights, in general, converge to weights that do *not* recover the factor structure of the treated unit when the number of pretreatment periods increases. The reason is that the SC weights converge to weights that attempt to, at the same time, recover the factor structure of the treated unit *and* minimize the variance of a weighted average of the idiosyncratic shocks of the control units. However, 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 structure of the treated unit with weights that are diluted among an increasing number of control units. This implies that, asymptotically, the SC unit is affected by the common shocks in the same way as the treated unit. As a consequence, the SC estimator is asymptotically unbiased even when treatment assignment is correlated with time-varying unobservables.

While increasing the number of control units increases the number of parameters to be estimated, as shown by Chernozhukov, Wuthrich, and Zhu (2021), the nonnegativity and adding-up constraints work as a regularization method. This guarantees that the SC unit consistently estimates the factor structure of the treated unit even when the number of control units grows at a faster rate than the number of pretreatment periods. We also show that such regularization implies that, asymptotically, there is no over-fitting. Asymptotically, the SC

CONTACT Bruno Ferman bruno.ferman@fgv.br Sao Paulo School of Economics-FGV, Sao Paulo, 01332-000 Brazil
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

unit absorbs only the common factor structure, so that the pretreatment fit will *not* be perfect due to the idiosyncratic shocks, even when the number of control units increases. This highlights that the asymptotic unbiasedness of the SC estimator we derive does *not* come from improvements in the pretreatment fit due to an increased number of control units. Rather, it comes from the fact that, under the conditions we consider for the factor structure, it is possible to construct balancing weights such that the variance of a linear combination of the idiosyncratic shocks of the control units using those weights converges to zero.

Overall, these results extend the set of possible applications in which the SC estimator can be reliably used. While the original SC papers recommend that the method should only be used in applications that present a good pretreatment fit for a long series of pretreatment periods, we show that, under some conditions, it can still be reliable even when the pretreatment fit is imperfect. The conditions we derive for asymptotic unbiasedness provide a guideline on how applied researchers should justify the use of the method in empirical applications with imperfect pretreatment fit.

If we relax the nonnegativity constraint on the weights, then the estimator for the SC unit will still be an asymptotically unbiased estimator for the factor structure of the treated unit when both the number of pretreatment periods and the number of controls increase. However, due to the lack of regularization, this estimator may not be consistent. We provide a simple example showing that, while the bias of the estimator for the treatment effects when we relax these constraints converges to zero when the number of control units goes to infinity, the variance of its asymptotic distribution is increasing with the ratio between the number of control units and pretreatment periods. When this ratio becomes close to one, the variance of this asymptotic distribution diverges. This highlights the importance of using regularization methods when the number of pretreatment periods is not much larger than the number of control units.

We present a baseline SC setting in Section 2. In Section 3, we analyze the asymptotic properties of the original SC estimator when both the number of pretreatment periods and the number of control units go to infinity. In Section 4, we analyze the asymptotic properties of the SC estimator when we relax the non-negativity and adding-up constraints in this setting. We present a simple Monte Carlo exercise in Section 5 to illustrate the theoretical results presented in Sections 3 and 4. Section 6 concludes.

2. Setting

There are $i = 0, 1, \dots, J$ units, where unit 0 is treated and the other units are controls. Potential outcomes when unit i at time t is treated (y^I_{it}) and non-treated (y^N_{it}) are determined by a linear factor model,

$$\begin{cases} y^N_{it} = \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \epsilon_{it} \\ y^I_{it} = \alpha_{it} + y^N_{it}, \end{cases} \tag{1}$$

where $\boldsymbol{\lambda}_t = [\lambda_{1t} \ \dots \ \lambda_{Ft}]$ is an $1 \times F$ vector of unknown common factors, $\boldsymbol{\mu}_i$ is an $F \times 1$ vector of unknown factor loadings, and the error terms ϵ_{it} are unobserved idiosyncratic shocks.

We only observe $y_{it} = d_{it} y^I_{it} + (1 - d_{it}) y^N_{it}$, where $d_{it} = 1$ if unit i is treated at time t. We analyze the properties of the SC estimator considering a repeated sampling framework over the distribution of ϵ_{it}, conditional on a fixed sequence of $\boldsymbol{\lambda}_t$ and $\boldsymbol{\mu}_i$. We define $\mathbf{M}_J$ as the $J \times F$ matrix that collects the information on the factor loadings of the control units (that is, the ith row of $\mathbf{M}_J$ is equal to $\boldsymbol{\mu}_i'$). We observe $(y_{0t}, \dots, y_{Jt})$ for periods $t \in \{-T_0 + 1, \dots, -1, 0, 1, \dots, T_1\}$, where treatment is assigned to unit 0 after time 0. Therefore, we have T_0 pretreatment periods and T_1 posttreatment periods. Let $\mathcal{T}_0$ ($\mathcal{T}_1$) be the set of time indices in the pretreatment (post-treatment) periods. The main goal of the SC method is to estimate the effect of the treatment for unit 0 for each $t \in \mathcal{T}_1$, $\{\alpha_{01}, \dots, \alpha_{0T_1}\}$.

Assumption 2.1. (sampling) We observe $\{y_{0t}, y_{1t}, \dots, y_{Jt}\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$, where $y_{it} = y^I_{it}$ if $i = 0$ and $t \in \mathcal{T}_1$, and $y_{it} = y^N_{it}$ otherwise. Potential outcomes are defined by Equation (1). We treat $\{\boldsymbol{\mu}_i\}_{i=0}^J$, $\{\boldsymbol{\lambda}_t\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$, and $\{\alpha_{01}, \dots, \alpha_{0T_1}\}$ as fixed, and $\{\epsilon_{0t}, \epsilon_{1t}, \cdots, \epsilon_{Jt}\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$ as stochastic.

We define the operator $\mathbb{E}^*[]$ as the expectation with respect to the idiosyncratic shocks, holding treatment assignment, common factors, and factor loadings fixed. We impose the following assumption on the idiosyncratic errors.

Assumption 2.2. (selection on unobservables) $\mathbb{E}^*[\epsilon_{it}] = 0$ for all i and t.

Assumption 2.2, combined with the fact that we are considering treatment assignment and the factor structure as fixed (Assumption 2.1), compose the main restrictions we impose on the treatment assignment mechanism. In order to think about treatment assignment, it is easier to consider an underlying model in which treatment assignment, factor loadings, and common factors are stochastic. In this case, the expectation in Assumption 2.2 and all our analyses should be viewed as conditional on these variables. In such underlying model, we do not impose any restriction on the correlation between treatment assignment and the factor structure. The possibility that these variables are correlated is what we refer to as "selection on unobservables." In this underlying model, the main restriction on the treatment assignment mechanism that Assumption 2.2 imposes is that the idiosyncratic errors are mean independent from treatment assignment and factor structure. Note that this assumption would allow for the conditional variance of the idiosyncratic shocks to depend on treatment assignment and/or the factor structure.

If we have weights $\mathbf{w}^* \in \mathbb{R}^J$ such that $\boldsymbol{\mu}_0 = \mathbf{M}_J{}'\mathbf{w}^*$, it would be possible to consider an unfeasible SC unit for $t \in \mathcal{T}_1$ using those weights,

$$\mathbf{y}_t'\mathbf{w}^* = \boldsymbol{\lambda}_t \boldsymbol{\mu}_0 + \boldsymbol{\epsilon}_t'\mathbf{w}^*, \tag{2}$$

leading to an unfeasible SC estimator

$$\hat{\alpha}^*_{0t} = y_{0t} - \mathbf{y}_t'\mathbf{w}^* = \alpha_{0t} + \epsilon_{0t} - \boldsymbol{\epsilon}_t'\mathbf{w}^*. \tag{3}$$

Therefore, under Assumption 2.2, we have that $\mathbb{E}^*[\hat{\alpha}^*_{0t}] = \alpha_{0t}$, even when we have selection on unobservables. The main challenge, however, is that the factor loadings are unobserved, so we cannot choose $\mathbf{w}^*$ this way.

In a sequence of articles, Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010), and Abadie, Diamond, and Hainmueller (2015) proposed the SC method to estimate weights for the control units to construct a counterfactual for $\{y_{01}^N,\dots,y_{0T_1}^N\}$. In a version of the method where all pretreatment outcome lags are included as predictor variables, those weights are estimated by minimizing the pretreatment sum of squared residuals subject to the constraints that weights must be nonnegative and sum one. Abadie, Diamond, and Hainmueller (2010) showed that, if there are weights that provide a perfect pretreatment fit, then the bias of the SC estimator is bounded by a function that asymptotes to zero when T_0 increases, when J is fixed. By perfect pretreatment fit we mean that there is a $(w_1,\dots,w_J)\in\Delta^{J-1}$ such that $y_{0t}=\sum_{j=1}^J w_j y_{jt}$ for all $t\in\mathcal{T}_0$, where $\Delta^{J-1}\equiv\{(w_1,\dots,w_J)\in\mathbb{R}^J|w_j\geq 0 \text{ and } \sum_{j=1}^J w_j=1\}$. Botosaru and Ferman (2019) and Powell (2021) also considered the properties of the SC and related estimators under a perfect pretreatment fit condition. However, Ferman and Pinto (2021) showed that, if the pretreatment fit is imperfect, then the SC weights will generally not recover the factor structure of the treated unit, so the SC estimator will be biased if there is selection on unobservables. They show that this result is valid even when $T_0\to\infty$, as long as J is fixed. The main reason is that, for any $\mathbf{w}^*\in\mathbb{R}^J$ such that $\boldsymbol{\mu}_0=\mathbf{M}_J{}'\mathbf{w}^*$, it is possible to write

$$y_{0t}^N=\mathbf{y}_t'\mathbf{w}^*+\epsilon_{0t}-\boldsymbol{\epsilon}_t'\mathbf{w}^*, \tag{4}$$

where $\mathbf{y}_t=(y_{1t},\dots,y_{Jt})'$, and $\boldsymbol{\epsilon}_t=(\epsilon_{1t},\dots,\epsilon_{Jt})'$. Therefore, the outcomes of the control units serve as a proxy for the factor structure of the treated unit. However, the linear factor model structure inherently generates a correlation between $\mathbf{y}_t$ and the error in this model due to the idiosyncratic shocks $\boldsymbol{\epsilon}_t$. As a consequence, with J fixed, the SC weights will generally not converge in probability to a $\mathbf{w}^*$ such that $\boldsymbol{\mu}_0=\mathbf{M}_J{}'\mathbf{w}^*$, even when $T_0\to\infty$.

3. Asymptotic Behavior of the Original SC Estimator with Large T_0 and Large J

We analyze the properties of the SC estimator when both the number of control units (J) and the number of pretreatment periods (T_0) increase. This provides a better asymptotic approximation to settings in which the number of pretreatment periods and the number of control observations are roughly of the same size, as is common in SC applications (e.g., Doudchenko and Imbens 2016).

Considering a SC specification that includes all pretreatment outcome lags as predictors, the SC weights are given by

$$\widehat{\mathbf{w}}_{\text{SC}}\in\underset{\mathbf{w}\in\Delta^{J-1}}{\text{argmin}}\left\{\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(y_{0t}-\mathbf{w}'\mathbf{y}_t\right)^2\right\}. \tag{5}$$

The main challenge in analyzing the behavior of the SC estimator in a setting with large J and large T_0 is that, when $T_0\to\infty$, the dimension of $\widehat{\mathbf{w}}_{\text{SC}}$ increases. However, we are not inherently interested in $\widehat{\mathbf{w}}_{\text{SC}}$, but in the SC unit $\mathbf{y}_t'\widehat{\mathbf{w}}_{\text{SC}}$ for $t\in\mathcal{T}_1$, which is an estimator for the counterfactual y_{0t}^N. Note that $\mathbf{y}_t'\widehat{\mathbf{w}}_{\text{SC}}=\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{SC}}+\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}}$, where $\widehat{\boldsymbol{\mu}}_{\text{SC}}=\mathbf{M}_J{}'\widehat{\mathbf{w}}_{\text{SC}}$. We consider, therefore, the asymptotic behavior of $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{SC}}$ and $\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}}$. The solution to this minimization problem may not be unique. In this case, we can consider $\widehat{\mathbf{w}}_{\text{SC}}$ as being one of the solutions to this problem. As we show below, this does not impact the asymptotic distribution of the SC unit $\mathbf{y}_t'\widehat{\mathbf{w}}_{\text{SC}}$.

For a given $\mathbf{w}$, let $\boldsymbol{\mu}\equiv\mathbf{M}_J{}'\mathbf{w}$. From the objective function in Equation (5),

$$\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(y_{0t}-\mathbf{w}'\mathbf{y}_t\right)^2=\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\boldsymbol{\mu})+\epsilon_{0t}-\mathbf{w}'\boldsymbol{\epsilon}_t\right)^2. \tag{6}$$

Now define

$$\mathcal{H}_J(\boldsymbol{\mu})=\min_{\mathbf{w}\in\Delta^{J-1}:\,\mathbf{M}_J{}'\mathbf{w}=\boldsymbol{\mu}}\left\{\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\bar{\lambda}_t(\boldsymbol{\mu})-\mathbf{w}'\boldsymbol{\epsilon}_t)^2\right\}, \tag{7}$$

where $\bar{\lambda}_t(\boldsymbol{\mu})\equiv\boldsymbol{\lambda}_t(\boldsymbol{\mu}_0-\boldsymbol{\mu})+\epsilon_{0t}$. Then,

$$\begin{aligned}&\underset{\boldsymbol{\mu}\in\mathcal{M}_J}{\text{argmin}}\mathcal{H}_J(\boldsymbol{\mu})\\&=\left\{\boldsymbol{\mu}=\mathbf{M}_J{}'\mathbf{b}\mid\mathbf{b}\in\underset{\mathbf{w}\in\Delta^{J-1}}{\text{argmin}}\left\{T_0{}^{-1}\sum_{t\in\mathcal{T}_0}\left(y_{0t}-\mathbf{w}'\mathbf{y}_t\right)^2\right\}\right\},\end{aligned} \tag{8}$$

where $\mathcal{M}_J\equiv\{\boldsymbol{\mu}\in\mathbb{R}^F|\boldsymbol{\mu}=\mathbf{M}_J{}'\mathbf{w}\text{ for some }\mathbf{w}\in\Delta^{J-1}\}$ is the set of factor loadings that can be attained with weights $\mathbf{w}\in\Delta^{J-1}$ when there are J control units.

Then, for $t\in\mathcal{T}_1$, we consider conditions in which $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{SC}}\xrightarrow{p}\boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ and $\widehat{\mathbf{w}}_{\text{SC}}'\boldsymbol{\epsilon}_t\xrightarrow{p}0$ when T_0 and $J\to\infty$ for $t\in\mathcal{T}_1$. In this case, the SC unit asymptotically recovers the factor structure of the treated unit, and the SC estimator is asymptotically unbiased.

We consider the following assumptions on the idiosyncratic shocks, in addition to Assumption 2.2.

Assumption 3.1. (idiosyncratic shocks) (a) $\{\epsilon_{it}\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$ are independent across i; (b) $\{\epsilon_{0t},\dots,\epsilon_{Jt}\}_{t\in\mathcal{T}_0}$ is α-mixing; (c) ϵ_{it} have uniformly bounded fourth moments across i and t, and $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\mathbb{E}^*[\epsilon_{0t}^2]\to\sigma_0^2$; (d) $\exists\underline{\gamma}>0$ such that $\mathbb{E}^*[\epsilon_{it}^2]\geq\underline{\gamma}$ across i and t.

Assumption 3.1(a) implies that the idiosyncratic shocks are uncorrelated across units, so that all spatial correlation is captured by the factor structure. While we allow for serial correlation in ϵ_{it}, Assumption 3.1(b) restricts such dependence by assuming a mixing condition. Finally, while we do not require stationarity, Assumptions 3.1(c) and 3.1(d) impose some restrictions on the moments of ϵ_{it}.

We also consider the following assumptions on the sequence of factor loadings and common factors. Let $\|.\|_2$ be the Frobenius norm.

Assumption 3.2. (factor loadings) (a) As $T_0,J\to\infty$, there is a sequence $\mathbf{w}_J^*\in\Delta^{J-1}$ such that $\left\|\mathbf{M}_J{}'\mathbf{w}_J^*-\boldsymbol{\mu}_0\right\|_2\to 0$, and $\left\|\mathbf{w}_J^*\right\|_2\to 0$, and (b) the sequence $\boldsymbol{\mu}_i$ is uniformly bounded.

Assumption 3.2(a) implies that there is a sequence of weights $(\mathbf{w}_J^*)$ diluted among an increasing number of control units, and that are such that the implied factor loadings associated with those weights ($\boldsymbol{\mu}_J^*\equiv\mathbf{M}_J{}'\mathbf{w}_J^*$) reconstruct the factor loadings of the treated unit ($\boldsymbol{\mu}_0$) in the limit, implying that $\boldsymbol{\lambda}_t(\boldsymbol{\mu}_J^* - \boldsymbol{\mu}_0) \to 0$ for all t.

Recall that this analysis is conditional on a fixed sequence of factor loadings. If we assume, for example, that the underlying distribution of factor loadings $\{\boldsymbol{\mu}_i\}_{i=0,..,J}$ is iid with finite support $\{\mathbf{m}_1, \dots, \mathbf{m}_{\bar{q}}\}$, then the conditions imposed in Assumption 3.2 for the factor loadings would be satisfied with probability one (details in Appendix A.2.1). This assumption would also be satisfied with probability one even if we consider a case in which the distributions of $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_i$ for $i > 0$ are different, as long as every point in the support of the distribution of $\boldsymbol{\mu}_0$ is in the convex hull of $\{\mathbf{m}_1, \dots, \mathbf{m}_{\bar{q}}\}$. Assumption 3.2(b) guarantees that the parameter space $\mathcal{M} = \text{cl}\left(\cup_{J\in\mathbb{N}}\mathcal{M}_J\right)$, which is the closure of $\cup_{J\in\mathbb{N}}\mathcal{M}_J$, is compact.

Assumption 3.3. (common factors) $\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t \to \boldsymbol{\Omega}$ positive definite.

Assumption 3.3 implies that common factors generate enough independent variation so that we can identify the effects of each factor on the pretreatment outcomes. Abadie, Diamond, and Hainmueller (2010) considered a similar assumption. If we consider an underlying distribution for $\boldsymbol{\lambda}_t$ such that, for example, $\boldsymbol{\lambda}_t$ is α-mixing with uniformly bounded fourth moments, and that $T_0{}^{-1}\sum_{t\in\mathcal{T}_0}\mathbb{E}[\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t] \to \boldsymbol{\Omega}$, then $T_0{}^{-1}\sum_{t\in\mathcal{T}_0}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t \overset{a.s.}{\to} \boldsymbol{\Omega}$. In this case, Assumption 3.3 would be satisfied with probability one.

We also assume some technical conditions that are important to take into account that the number of control units goes to infinity with the number of pretreatment periods.

Assumption 3.4. (other assumptions) (a) $\max_{1\le j\le J}\left\{\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{0t}\epsilon_{jt}\right|\right\} = o_p(1)$ and, for all $f = 1, \dots, F$, $\max_{0\le j\le J}\left\{\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\lambda_{ft}\epsilon_{jt}\right|\right\} = o_p(1)$; (b) $\exists\ c > 0$ such that $\min_{1\le j\le J}\left\{\sum_{t\in\mathcal{T}_0}\left|\epsilon_{jt}^2\right|\right\} \ge cT_0$ with probability $1 - o(1)$, and $\max_{1\le i,j\le J, i\neq j}\left\{\left|\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\epsilon_{it}\epsilon_{jt}\right|\right\} = o_p(1)$.

These high-level conditions essentially determine the rate in which J can diverge when $T_0 \to \infty$. Whether these conditions are satisfied depend crucially on the rates in which J and T_0 diverge, on the dependence of ϵ_{it}, and on the number of uniformly bounded moments of ϵ_{it}. If we allow J to diverge at a faster rate than T_0, or we allow time-series dependence on ϵ_{it}, then we need a larger number of uniformly bounded moments of ϵ_{it}. See Appendix A.2.2 for some simple examples in which Assumption 3.4 is satisfied even when J diverges at a faster rate than T_0.

Given these conditions, we derive the following results.

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

The first result in Proposition 3.1 shows that, asymptotically, the treated and the SC units will be affected by the factor structure in the same way. The main idea of the proof is the following. We consider an extension of the function $\mathcal{H}_J(\boldsymbol{\mu})$ to the domain $\mathcal{M} = \text{cl}\left(\cup_{J\in\mathbb{N}}\mathcal{M}_J\right)$ such that $\underset{\boldsymbol{\mu}\in\mathcal{M}}{\text{argmin}}\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}) = \underset{\boldsymbol{\mu}\in\mathcal{M}_J}{\text{argmin}}\mathcal{H}_J(\boldsymbol{\mu})$. Then we show that $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu}_0) \overset{p}{\to} \sigma_0^2$, and that $\widetilde{\mathcal{H}}_J(\boldsymbol{\mu})$ is bounded from below by a function $\widetilde{\mathcal{H}}_{T_0}^{LB}(\boldsymbol{\mu})$ that converges uniformly in $\boldsymbol{\mu} \in \mathcal{M}$ to $\sigma_{\lambda}^2(\boldsymbol{\mu}) \equiv (\boldsymbol{\mu}_0 - \boldsymbol{\mu})'\boldsymbol{\Omega}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}) + \sigma_0^2$. Under Assumption 3.3, $(\boldsymbol{\mu}_0-\boldsymbol{\mu})'\boldsymbol{\Omega}(\boldsymbol{\mu}_0-\boldsymbol{\mu})+\sigma_0^2$ is uniquely minimized at $\boldsymbol{\mu}_0$, which implies that $\widehat{\boldsymbol{\mu}}_{\text{SC}} \overset{p}{\to} \boldsymbol{\mu}_0$. Therefore, for any $t \in \mathcal{T}_1$, we have that $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{SC}} \overset{p}{\to} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$. The details of the proof are presented in Appendix A.1.1.

Ferman and Pinto (2021) showed that, when the number of control units is fixed, the SC weights converge to weights that do not, in general, recover $\boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ for $t \in \mathcal{T}_1$. This happens because, in a setting with a fixed number of control units, the SC weights converge to weights that simultaneously attempt to minimize both the second moments of the remaining common structure, and the variance of a weighted average of the idiosyncratic shocks of the control units. Intuitively, this first result from Proposition 3.1 comes from the fact that, when both the number of pretreatment periods and the number of controls increase, the importance of this variance of a weighted average of the idiosyncratic shocks of the control units vanishes. As a consequence, the asymptotic bias of $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{SC}}$ disappears when both the number of pretreatment periods and the number of controls increase.

A crucial condition for this result is that, as the number of control units increases, it is possible to recover the factor structure of the treated with weights that are diluted among an increasing number of control units (Assumption 3.2). If we consider, for example, a setting such that there is only a fixed number of control units that can be used to recover $\boldsymbol{\lambda}_t\boldsymbol{\mu}_0$, and the additional control units are uncorrelated with y_{0t}, then the result from Ferman and Pinto (2021) would still apply, and $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{SC}}$ would generally not converge to $\boldsymbol{\lambda}_t\boldsymbol{\mu}_0$. Such setting would be inconsistent with Assumption 3.2.

Proposition 3.1 also shows that the SC weights will get diluted among an increasing number of control units, so that $\|\widehat{\mathbf{w}}_{\text{SC}}\|_2 \overset{p}{\to} 0$. An immediate consequence is that, if we assume that idiosyncratic shocks in the posttreatment periods are independent from the idiosyncratic shocks in the pretreatment periods, then, for any $t \in \mathcal{T}_1$, $\hat{\alpha}_{0t}^{\text{SC}} \equiv y_{0t} - \mathbf{y}_t'\widehat{\mathbf{w}}_{\text{SC}} \overset{p}{\to} \alpha_{0t} + \epsilon_{0t}$ when $T_0 \to \infty$.

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

This happens because not only $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{SC}} \overset{p}{\to} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$, but also $\widehat{\mathbf{w}}_{\text{SC}}$ is diluted among an increasing number of control units, implying that $\boldsymbol{\epsilon}_t'\widehat{\mathbf{w}}_{\text{SC}} \overset{p}{\to} 0$ (see details in Appendix A.1.2). Therefore, the SC unit converges in probability to the factor structure of the treated unit. As a consequence, under Assumption 2.2, the SC estimator is asymptotically unbiased for α_{0t} even when we have selection on unobservables. Moreover, the asymptotic distribution of the SC estimator depends only on the idiosyncratic shocks of the treated unit in period t. In Appendix A.2.3, we present Corollary A.1, in which we derive $\mathbf{y}_t'\widehat{\mathbf{w}}_{\text{SC}} \overset{p}{\to} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ and $\hat{\alpha}_{0t}^{\text{SC}} \overset{p}{\to} \alpha_{0t} + \epsilon_{0t}$ under a different set of assumptions, allowing for time dependence between idiosyncratic shocks in the pre- and posttreatment periods.

Our results are closely linked to Theorem 5 by Arkhangelsky et al. (2019), who considered a penalized version of the SC weights. These penalized SC weights solve the minimization problem presented in Equation (5) subject to the additional constraint that $||\mathbf{w}||_2 \leq a_w$. Since $||\mathbf{w}||_1 = 1 \Rightarrow ||\mathbf{w}||_2 \leq 1$, note that the original SC weights are equivalent to the penalized SC weights with $a_w = 1$. They showed that the approximation error for their low-rank matrix structure goes to zero if, among other conditions, $a_w \to 0$. In contrast, we show that, in our setting, the SC weights achieve such balancing even when the penalty term a_w equals one. In this case, the original SC method—which does not include an L_2 penalization term—is such that the SC unit asymptotically recovers $\boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ (since we have that, for $t \in \mathcal{T}_1$, both $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{SC}} \overset{p}{\to} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ and $\widehat{\mathbf{w}}_{\text{SC}}'\boldsymbol{\epsilon}_t \overset{p}{\to} 0$ when T_0 and $J \to \infty$).

Finally, under the assumptions considered in Proposition 3.1, $T_0^{-1}\sum_{t\in\mathcal{T}_0}\left(y_{0t} - \widehat{\mathbf{w}}_{\text{SC}}'\mathbf{y}_t\right)^2$ converges in probability to $\sigma_{\bar{\lambda}}^2(\boldsymbol{\mu}_0) = \sigma_0^2$, which is the asymptotic variance of ϵ_{0t}. Therefore, the SC unit will asymptotically absorb all variability of y_{0t} that is related to the factor structure, but will *not* overfit the idiosyncratic shocks of the treated unit. This happens because the nonnegativity and adding-up constraints on the weights work as a regularization method, as presented by Chernozhukov, Wuthrich, and Zhu (2021). This implies that we should *not* expect a perfect pretreatment fit in this setting, even when J grows at a faster rate than T_0. Therefore, we provide conditions in which the SC estimator can be reliably used even in a setting in which the original SC papers recommend that the method should not be used (e.g., Abadie, Diamond, and Hainmueller 2010, 2015). Moreover, this highlights that the asymptotic unbiasedness result from Proposition 3.1 does not come from a better pretreatment fit when we increase the number of control units. Rather, it comes from the fact that increasing the number of control units implies existence of balancing weights that are diluted among an increasing number of control units, implying that the problems highlighted by Ferman and Pinto (2021) become asymptotically irrelevant.

Remark 3.1. An important implication of Proposition 3.1 is that $\|\widehat{\mathbf{w}}_{\text{SC}}\|_2 \overset{p}{\to} 0$. Since $\|\widehat{\mathbf{w}}_{\text{SC}}\|_2$ is observed, this is an interesting measure that applied researchers can report to check whether the assumptions and the asymptotic approximations of Proposition 3.1 are reliable. While it would not be trivial to derive a general connection between $\|\widehat{\mathbf{w}}_{\text{SC}}\|_2$ and the possibility that the SC estimator is biased, a value of $\|\widehat{\mathbf{w}}_{\text{SC}}\|_2$ closer to zero should indicate that the conclusions from Proposition 3.1 are more reliable.

Remark 3.2. Proposition 3.1 remains valid if we consider a demeaned SC estimator, as proposed by Ferman and Pinto (2021), which is numerically the same as including a constant in the minimization problem (5), as proposed by Doudchenko and Imbens (2016). We show in Appendix A.2.4 that this would require only minor adjustments in the proof of Proposition 3.1.

Remark 3.3. While we focus on the SC specification that includes all pretreatment outcome lags as predictors, we consider a setting with covariates in Appendix A.2.5. We show that the conclusions from Proposition 3.1 remain valid for SC specifications that include time-invariant covariates as predictors, as long as the number of pretreatment outcomes lags used as predictors goes to infinity when $T_0 \to \infty$. This result is an extension of the conclusions from Ferman, Pinto, and Possebom (2020) for the case in which both J and T_0 diverge. We also present in Appendix A.2.5 Monte Carlo simulations considering a setting with covariates and different SC specifications. In this setting, both the SC weights estimated from Equation (5), and the SC weights using half of the pretreatment outcomes and covariates as predictors, approximately recover both the factor structure and the structure related to the time-invariant covariates of the treated unit when (T_0, J) are large. We also include in our simulations in Appendix A.2.5 a SC specification that does not satisfy the condition on the number of pretreatment outcomes used as predictors going to infinity. In this case, the SC weights failed to recover the factor structure of the treated unit even when (T_0, J) are large.

4. Relaxing the Nonnegativity Constraints

We consider now the importance of the regularization provided by the nonnegativity and adding-up constraints for the results presented in Section 3. We consider the case without both the adding-up and the nonnegativity constraints. The case with only the adding-up constraint is similar. We show that, if $T_0 > J$, then the estimator remains asymptotically unbiased when we relax these constraints, if we assume that there are weights diluted among an increasing number of control units that recover $\boldsymbol{\mu}_0$ (which is similar to Assumption 3.2). However, relaxing these constrains imply a larger asymptotic variance, except in the case in which T_0 increases at a much faster rate than J, so that $J/T_0 \to 0$.

In this case, the weights are estimated using the OLS regression

$$\widehat{\mathbf{b}}_{\text{OLS}} = \underset{\mathbf{b}\in\mathbb{R}^J}{\text{argmin}}\left\{\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left(y_{0t} - \mathbf{b}'\mathbf{y}_t\right)^2\right\}. \quad (9)$$

Following the same arguments presented in Section 3, we can define

$$\mathcal{H}_J^{\text{OLS}}(\boldsymbol{\mu}) = \min_{\mathbf{b}\in\mathbb{R}^J:\, \mathbf{M}_J'\mathbf{b}=\boldsymbol{\mu}}\left\{\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}(\bar{\lambda}_t(\boldsymbol{\mu}) - \mathbf{b}'\boldsymbol{\epsilon}_t)^2\right\}, \quad (10)$$

so that $\widehat{\boldsymbol{\mu}}_{\text{OLS}} \equiv \mathbf{M}_J'\widehat{\mathbf{b}}_{\text{OLS}}$ is the solution to $\underset{\boldsymbol{\mu}\in\mathcal{M}_J^{\text{OLS}}}{\text{argmin}}\,\mathcal{H}_{T_0}^{\text{OLS}}(\boldsymbol{\mu})$, where $\mathcal{M}_J^{\text{OLS}} \equiv \{\boldsymbol{\mu} \in \mathbb{R}^F | \boldsymbol{\mu} = \mathbf{M}_J'\mathbf{b} \text{ for some } \mathbf{b} \in \mathbb{R}^J\}$.

A crucial difference in this case is that, by not imposing any restriction on $\mathbf{b}$, this minimization problem is subject to over-fitting when J increases with T_0. As a consequence, the lower bound we derive in the proof of Proposition 3.1, which in this case would be given by $\min_{\mathbf{b}\in\mathbb{R}^J}\left\{T_0^{-1}\sum_{t\in\mathcal{T}_0}(\bar{\lambda}_t(\boldsymbol{\mu})-\mathbf{b}'\boldsymbol{\epsilon}_t)^2\right\}$, would not generally converge to $\sigma^2_{\bar{\lambda}}(\boldsymbol{\mu})$. In the extreme case in which $T_0 = J$, this lower bound would be equal to zero for all $\boldsymbol{\mu}$ with probability one. We can still show, however, that, when $J/T_0 \to c < 1, \boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{OLS}} \xrightarrow{p} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ for all $t \in \mathcal{T}_1$. Moreover, we can show that, under some conditions, the expected value of $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{OLS}}$ converges to the expected value of $\boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ for all $t \in \mathcal{T}_1$ even when $J/T_0 \to 1$.

Consider first the case in which $J/T_0 \to c < 1$. We continue to consider the properties of the estimator over the distribution of the idiosyncratic shocks, and conditional on fixed sequences of common factors and factor loadings. Regarding the idiosyncratic shocks, we continue to consider Assumptions 2.2 and 3.1. We impose the following assumptions on the sequence of factor loadings.

Assumption 4.1. (factor loadings) For some $\underline{a}, \bar{a} > 0$, let R be the number of disjoint groups of F control units we can arrange such that the $F \times F$ matrix with the factor loadings for each of those groups has its smallest eigenvalue greater than $\underline{a}$, and its largest eigenvalue smaller than $\bar{a}$. We assume that $R \to \infty$ when $J \to \infty$.

Differently from the setting considered in Section 3, note that there will always exist a $\boldsymbol{\beta} \in \mathbb{R}^J$ such that $\mathbf{M}_J{}'\boldsymbol{\beta} = \boldsymbol{\mu}_0$, as long as there is at least one group of F control units such that their factor loadings form a basis for $\mathbb{R}^F$. Assumption 4.1 guarantees that, as J increases, we can form many groups of F control units such that their factor loadings form a basis of $\mathbb{R}^F$. Moreover, the bounds on the eigenvalues of these matrices limit the norm of the linear combination of these factor loadings that we need to consider to recover $\boldsymbol{\mu}_0$. With this assumption, we can guarantee that we can find a sequence of weights $\boldsymbol{\beta}^*_J \in \mathbb{R}^J$ such that $\mathbf{M}_J{}'\boldsymbol{\beta}^*_J = \boldsymbol{\mu}_0$ and $\left\|\boldsymbol{\beta}^*_J\right\|_2 \to 0$, which is similar to the condition we considered in Assumption 3.2. If we consider an underlying distribution for the factor loadings as the ones discussed in Appendix A.2.1, then this assumption would be satisfied with probability one.

Let $\boldsymbol{\Lambda}$ be the $T_0 \times F$ matrix with rows equal to $\boldsymbol{\lambda}_t$ for $t \in \mathcal{T}_0$, and $\boldsymbol{\epsilon}_i$ be the $T_0 \times 1$ vector with ϵ_{it} for $t \in \mathcal{T}_0$. Also, let $\mathbf{Z}$ be a $T_0 \times (J-F)$ matrix with column $i \in \{1,\ldots,J-F\}$ given by $\boldsymbol{\epsilon}_{F+i}$ minus a linear combination of $\{\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_F\}$, and let $\mathbf{Q}$ be the residual maker matrix of an OLS regression on $\mathbf{Z}$. We assume that the sequence $\boldsymbol{\Lambda}$ satisfies the following conditions.

Assumption 4.2. (common factors) $\mathbf{Z}$ is full rank with probability one, and we have that $\frac{1}{T_0-J+F}\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\Lambda} = O_p(1)$, $\left(\frac{1}{T_0-J+F}\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\Lambda}\right)^{-1} = O_p(1)$, and $\frac{1}{T_0-J+F}\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\epsilon}_0 = o_p(1)$ when $T_0, J \to \infty$.

The condition that $\mathbf{Z}$ is full rank with probability one guarantees that the residual maker $\mathbf{Q}$ is a well-defined idempotent matrix with rank $T_0 - J + F$. If the underlying distribution for $\boldsymbol{\lambda}_t$ were iid normal with mean zero, and $\boldsymbol{\lambda}_t$ is independent from $\boldsymbol{\epsilon}_j$, then a realization of the sequence $\boldsymbol{\lambda}_t$ would satisfy Assumption 4.2 with probability one. In this particular case, we would have $(T_0-J+F)^{-1}\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\Lambda} = (T_0-J+F)^{-1}\sum_{q=1}^{T_0-J+F}\widetilde{\boldsymbol{\lambda}}_q'\widetilde{\boldsymbol{\lambda}}_q$, where $\widetilde{\boldsymbol{\lambda}}_q$ is iid and has the same distribution as $\boldsymbol{\lambda}_q$, which implies that $(T_0-J+F)^{-1}\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\Lambda} \xrightarrow{\text{a.s.}} \mathbb{E}[\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t]$. Likewise, if we also have ϵ_{0t} iid normal and independent from $\boldsymbol{\lambda}_t$, then $(T_0-J+F)^{-1}\boldsymbol{\Lambda}'\mathbf{Q}\boldsymbol{\epsilon}_0 = (T_0-J+F)^{-1}\sum_{q=1}^{T_0-J+F}\widetilde{\boldsymbol{\lambda}}_q\tilde{\epsilon}_{0t} \xrightarrow{a.s.} \mathbb{E}[\boldsymbol{\lambda}_q\epsilon_{0t}] = 0$.

Given these assumptions, we show that, when $J/T_0 \to c < 1$, $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{OLS}} \xrightarrow{p} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ for all $t \in \mathcal{T}_1$.

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

The intuition is the same as the intuition in Proposition 3.1. When the number of control units increases, we are able to have a diluted weighted average of the control units that recover the factor structure of the treated unit, $\boldsymbol{\lambda}_t\boldsymbol{\mu}_0$. This reduces the importance of the variance of the linear combination of the idiosyncratic shocks of the control units in the minimization problem (9) for the estimation of $\widehat{\mathbf{b}}_{\text{OLS}}$, making the problem raised by Ferman and Pinto (2021) less relevant. Since $J/T_0 \to c \in [0,1)$, $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{OLS}}$ for $t \in \mathcal{T}_1$ converges in probability even when $J \to \infty$. This is consistent with Theorem 1 from Cattaneo, Jansson, and Newey (2018), once we consider a change in variables so that we can divide the J control variables into a group of F variables such that their associated estimators give us $\widehat{\boldsymbol{\mu}}_{\text{OLS}}$, and a remaining group of $J-F$ variables that we are not inherently interested in. As in Cattaneo, Jansson, and Newey (2018), J can be a nonvanishing fraction of T_0, but we cannot have that $J/T_0 \to 1$. It is easy to show that the assumptions for Theorem 1 from Cattaneo, Jansson, and Newey (2018) hold if we assume that the data are iid normal. We consider here an alternative proof for Proposition 4.1 where we take advantage of the specific details of our application, so that we can consider a weaker set of assumptions. See details of the proof in Appendix A.1.3. Similar to Corollary 3.1, if, for all $t \in \mathcal{T}_1$, ϵ_{it} is independent from $\{\epsilon_{i\tau}\}_{\tau\in\mathcal{T}_0}$, then $\mathbb{E}^*\left[\hat{\alpha}_{0t}^{\text{OLS}} - \alpha_{0t}\right] \to 0$, where $\hat{\alpha}_{0t}^{\text{OLS}} = y_{0t} - \mathbf{y}_t'\widehat{\mathbf{b}}_{\text{OLS}}$. However, Proposition 4.1 does not guarantee $\left\|\widehat{\mathbf{b}}_{\text{OLS}}\right\|_2 \xrightarrow{p} 0$, so we may not have $\boldsymbol{\epsilon}_t'\widehat{\mathbf{b}}_{\text{OLS}} \xrightarrow{p} 0$.

When $J/T_0 \to 1$, we will not generally have $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{OLS}} \xrightarrow{p} \boldsymbol{\lambda}_t\boldsymbol{\mu}_0$. If we impose a stronger set assumptions, however, we can still show that the expected value of $\boldsymbol{\lambda}_t\widehat{\boldsymbol{\mu}}_{\text{OLS}}$ converges to the expected value of $\boldsymbol{\lambda}_t\boldsymbol{\mu}_0$ for all $t \in \mathcal{T}_1$, even when $J/T_0 \to 1$. The only restriction is that $T_0 \geq J$, so that the OLS estimator is well specified. In this case, we need to rely on stronger assumptions. Moreover, while we continue to condition on a sequence of $\boldsymbol{\mu}_i$, we now consider $\boldsymbol{\lambda}_t$ stochastic. Therefore, we replace Assumption 2.1 with the following assumption.

Assumption 4.3. (sampling) We observe $\{y_{0t}, y_{1t}, \cdots, y_{Jt}\}_{t\in\mathcal{T}_0\cup\mathcal{T}_1}$, where $y_{it} = y_{it}^I$ if $i = 0$ and $t \in \mathcal{T}_1$, and $y_{it} = y_{it}^N$ otherwise. Potential outcomes are defined by

Equation (1). We treat $\{\boldsymbol{\mu}_i\}_{i=0}^{J}$, and $\{\alpha_{01}, \ldots, \alpha_{0T_1}\}$ as fixed, and $\{\epsilon_{0t}, \epsilon_{1t}, \cdots, \epsilon_{Jt}\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$ and $\{\boldsymbol{\lambda}_t\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$ as stochastic.

To avoid confusion, we denote by $\mathbb{E}^{**}[]$ the expectation over idiosyncratic shocks and common factors, when treatment assignment and factor loadings are fixed. We consider the following assumptions on the idiosyncratic shocks and common factors.

Assumption 4.4. (distributions of ϵ_{it} and $\boldsymbol{\lambda}_t$) (i) $\epsilon_{it} \sim N(0, \sigma_i^2)$ iid across $t \in \mathcal{T}_0$ for all $i \in \mathbb{N} \cup \{0\}$, $\boldsymbol{\lambda}_t \overset{iid}{\sim} N(0, \boldsymbol{\Omega})$ across $t \in \mathcal{T}_0$, where $\boldsymbol{\Omega}$ is positive definite, and all these variables are mutually independent; (ii) For $t \in \mathcal{T}_1$, $\mathbb{E}^{**}[\epsilon_{it}|\mathbf{V}] = 0$ for all $i = 0, \ldots, J$ and $\boldsymbol{\lambda}_t \perp\!\!\!\perp \boldsymbol{V}$, where $\mathbf{V}$ is the set of all variables in all periods $\tau \in \mathcal{T}_0$. We also assume $\boldsymbol{\lambda}_t$ has finite first moment for $t \in \mathcal{T}_1$.

Note that Assumption 4.4 allows for $\mathbb{E}^{**}[\boldsymbol{\lambda}_t] \neq 0$ for $t \in \mathcal{T}_1$, which would imply that the estimator may be asymptotically biased if $\mathbb{E}^{**}[\widehat{\boldsymbol{\mu}}_{\text{OLS}} - \boldsymbol{\mu}_0]$ does not converge to zero.

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

The only assumption we make on the number of control units and pretreatment periods is that $T_0 \geq J$, so that the OLS estimator is well defined. Therefore, this conclusion is valid even when $T_0 - J$ does not go to infinity. See details of the proof of Proposition 4.2 in Appendix A.1.4. The assumption that $\mathbb{E}^{**}[\boldsymbol{\lambda}_t] = 0$ in the pretreatment periods can be relaxed if we consider the demeaned SC estimator.

To illustrate the trade-offs between using the constraints or not, we consider a very simple example in which we can derive the asymptotic distribution of $\widehat{\alpha}_{0t}^{\text{OLS}}$ when $T_0 \to \infty$, depending on the value of $c \in [0, 1)$ such that $J/T_0 \to c$. Consider a setting with $F = 1$, where $y_{it}^N = \lambda_t + \epsilon_{it}$ for all $i \in \mathbb{N} \cup \{0\}$, with $\epsilon_{it} \overset{\text{iid}}{\sim} N(0, \sigma^2)$ across all i and t, and $\lambda_t \overset{\text{iid}}{\sim} N(0, \sigma_\lambda^2)$ across t, and independent from all $\epsilon_{i\tau}$. From Proposition 3.1 and Corollary 3.1, we know that the SC estimator converges in distribution to a $N(\alpha_{0t}, \sigma^2)$ in this case. Moreover, from Proposition 4.1, we know that $\widehat{\mu}_{\text{OLS}} \overset{p}{\to} \mu_0 = 1$. In Appendix A.2.6, we show that $\widehat{\alpha}_{0t}^{\text{OLS}}$ converges in distribution to a $N\left(\alpha_{0t}, \sigma^2(1-c)^{-1}\right)$. Therefore, the asymptotic variance of $\widehat{\alpha}_{0t}^{\text{OLS}}$ equals the asymptotic variance of the SC estimator when $J/T_0 \to 0$. However, if $J/T_0 \to c > 0$, the asymptotic variance of $\widehat{\alpha}_{0t}^{\text{OLS}}$ is larger than the asymptotic variance of the SC estimator. Moreover, the asymptotic variance of $\widehat{\alpha}_{0t}^{\text{OLS}}$ diverges to infinity when $c \to 1$.

Overall, combining the results from Sections 3 and 4, we have that using an OLS regression to estimate the weights without any regularization method can be a reasonable idea when the number of control units is large, but the number of pretreatment periods is much larger than the number of controls units. An advantage relative to the original SC estimator is that Assumption 4.1 requires a sequence of factor loadings that reconstructs $\boldsymbol{\mu}_0$ without the constraints on the weights. However, an important disadvantage of using the OLS estimator without any regularization method is that the variance of the estimator may be larger. As we show in our simple example, this cost can be substantial when the number of pretreatment periods is not much larger than the number of control units. Including only the adding-up constraint (without the non-negativity constraint) only increases the number of degrees of freedom by one, so all results in this section remain valid in this case. When the number of pretreatment periods is not much larger than the number of control units, other regularization methods could be used, as considered by, for example, Doudchenko and Imbens (2016), Arkhangelsky et al. (2021), Carvalho, Masini, and Medeiros (2018), Chernozhukov, Wuthrich, and Zhu (2021), Hsiao, Ching, and Wan (2012), and Li and Bell (2017).

5. Monte Carlo Simulations

We present a simple Monte Carlo (MC) exercise to illustrate the main results presented in Sections 3 and 4. We consider a setting in which there are two common factors, λ_{1t} and λ_{2t}. Potential outcomes for the treated unit and for half of the control units depend on the first common factor, so $y_{jt} = \lambda_{1t} + \epsilon_{jt}$ for $j = 0, 1, \ldots, J/2$, while $y_{jt} = \lambda_{2t} + \epsilon_{jt}$ for $j = J/2 + 1, \ldots, J$. In this case, $\boldsymbol{\mu}_0 = (\mu_{1,0}, \mu_{2,0}) = (1, 0)$. Therefore, the goal of the SC method is to set positive weights only to units $j = 1, \ldots, J/2$, which would imply that the SC weights would recover the factor structure of the treated. The common factors are normally distributed with a serial correlation equal to 0.5 and variance equal to 1; λ_{1t} and λ_{2t} are independent. The idiosyncratic shocks ϵ_{jt} are iid normally distributed with variance equal to 1.

Columns 1 to 4 of Table 1 present results for the SC method. Panel A considers a setting with $T_0 = J + 5$, so the number of pretreatment periods and the number of control units are roughly of the same size. When the number of control units is small ($J = 4$ or $J = 10$), there is distortion in the proportion of weights allocated to the control units that follow the same common factor as the treated unit. For example, when there are 10 control units, around 82% of the weights are correctly allocated, while around 18% of the weights are misallocated. When J and T_0 increase, the proportion of misallocated weights goes to zero, which is consistent with Proposition 3.1. Interestingly, the standard error of $\widehat{\boldsymbol{\mu}}_{\text{SC}}$ goes to zero when J increases, even when J and T_0 remain roughly at the same size. Moreover, the standard error of the treatment effect one period ahead, $\hat{\alpha}_{01}^{\text{SC}}$, converges to the standard deviation of the idiosyncratic shocks, which is consistent with Corollary 3.1. We find similar results when $T_0 = 2 \times J$ (columns 1 to 4, Panel B).

Columns 5 to 8 of Table 1 present results using OLS to estimate the weights. In this case, $\mathbb{E}^{**}[\hat{\mu}_{1,0}^{\text{OLS}}] < 1$ when J is small, due to the endogeneity generated by the idiosyncratic shocks of the control units. When J increases, however, $\mathbb{E}^{**}[\hat{\mu}_{1,0}^{\text{OLS}}] \to 1$, which is consistent with Propositions 4.1 and 4.2. However, differently from the SC weights, the standard error of $\hat{\mu}_{1,0}^{\text{OLS}}$ does not go to zero, and remains roughly constant when J increases but J and T_0 remains roughly at the same size (Panel A). In contrast, when $J/T_0 \to c \in [0, 1)$ (Panel B), then the standard

Table 1. Monte Carlo Simulations

	SC estimator				Unrestricted OLS				OLS with adding-up constraint			
J	4 (1)	10 (2)	50 (3)	100 (4)	4 (5)	10 (6)	50 (7)	100 (8)	4 (9)	10 (10)	50 (11)	100 (12)
					Panel A: $T_0 = J + 5$							
$\mathbb{E}^{**}[\hat{\mu}_{01}]$	0.760	0.817	0.905	0.929	0.653	0.816	0.962	0.976	0.829	0.910	0.982	0.989
$se^{**}[\hat{\mu}_{01}]$	0.206	0.156	0.076	0.054	0.489	0.516	0.501	0.506	0.319	0.324	0.320	0.325
$\mathbb{E}^{**}[\hat{\mu}_{02}]$	0.240	0.183	0.095	0.071	−0.002	−0.002	−0.005	0.010	0.171	0.090	0.018	0.011
$se^{**}[\hat{\mu}_{02}]$	0.206	0.156	0.076	0.054	0.498	0.509	0.497	0.518	0.319	0.324	0.320	0.325
$se^{**}(\hat{\alpha})$	1.288	1.194	1.084	1.073	1.586	1.984	3.791	5.220	1.486	1.806	3.437	4.661
					Panel B: $T_0 = 2 \times J$							
$\mathbb{E}^{**}[\hat{\mu}_{01}]$	0.753	0.831	0.922	0.944	0.637	0.828	0.960	0.982	0.825	0.915	0.981	0.991
$se^{**}[\hat{\mu}_{01}]$	0.217	0.136	0.057	0.040	0.569	0.343	0.143	0.103	0.354	0.231	0.100	0.072
$\mathbb{E}^{**}[\hat{\mu}_{02}]$	0.247	0.169	0.078	0.056	0.001	0.003	0.000	0.001	0.175	0.085	0.019	0.009
$se^{**}[\hat{\mu}_{02}]$	0.217	0.136	0.057	0.040	0.582	0.335	0.143	0.102	0.354	0.231	0.100	0.072
$se^{**}(\hat{\alpha})$	1.297	1.186	1.050	1.047	1.798	1.586	1.420	1.444	1.571	1.519	1.411	1.437

Notes: this table presents MC simulation results for the expected value and the standard error of the estimators for $\boldsymbol{\mu}_0 = (\mu_{01}, \mu_{02})$. These expectations consider both the idiosyncratic shocks and the common factor as stochastic. It also presents the standard error of $\hat{\alpha}$. Since $\mathbb{E}^{**}[\lambda_t] = 0$, $\mathbb{E}^{**}[\hat{\alpha}_{01}] = 0$, which is the true treatment effect. Panel A presents results with $T_0 = J + 5$, while Panel B presents results with $T_0 = 2 \times J$. Columns 1 to 4 present the results using the SC estimator to estimate the weights. Columns 5 to 8 present results using OLS estimator with no constraint. Columns 9 to 12 present results using OLS estimator with adding-up constraint. Results based on 5000 simulations. The DGP is described in detail in Section 5.

error of $\hat{\mu}_{1,0}^{\text{OLS}}$ goes to zero. The standard error of $\hat{\alpha}_{01}^{\text{OLS}}$ diverges with J when $T_0 = J + 5$. In contrast, it is decreasing with J when $T_0 = 2 \times J$, although it never reaches the standard error of $\hat{\alpha}_{01}^{\text{SC}}$. These results are consistent with the simple example presented in Section 4.

When weights are estimated with OLS using only the adding-up constraint, results are similar to the unrestricted OLS. The only difference is that $\mathbb{E}^{**}[\hat{\mu}_{2,0}^{\text{OLS}}] = 0$ regardless of J when we consider the unrestricted OLS. This happens because $\mu_{2,0} = 0$, so there is no endogeneity problem for this parameter when we consider the unrestricted OLS. In contrast, there is distortion in $\hat{\mu}_{2,0}$ when we include the restriction that weights should sum one (see columns 9 to 12 of Table 1).

In Appendix A.2.5, we present Monte Carlo simulations considering a setting with covariates and different SC specifications.

6. Conclusion

We provide conditions under which the SC estimator is asymptotically unbiased when both the number of pretreatment periods and the number of control units increase. This will be the case when, as the number of control units goes to infinity, there are weights diluted among an increasing number of control units that asymptotically recover the factor structure of the treated unit. Under this condition, the SC estimator can be asymptotically unbiased even when treatment assignment is correlated with time-varying unobserved confounders.

We show that the nonnegative and adding-up constraints are crucial for this result, as they provide regularization for cases in which the number of parameters to be estimated is larger than the number of pretreatment periods. Without these constraints, the estimator for the treatment effect remains asymptotically unbiased, but it will generally have a larger variance, unless the number of pretreatment periods is much larger than the number of control units.

Overall, our results extend the set of possible applications in which the SC estimator can be reliably used. While the original SC articles recommend that the method should only be used in applications that present a good pretreatment fit for a long series of pretreatment periods, we show that, under some conditions, it can still allow for time-varying unobserved confounders even when the pretreatment fit is imperfect. In this case, however, researchers would have to evaluate the plausibility of the conditions we present in this article. Observing that the SC weights are diluted among a large number of control units in a given application provides supportive evidence that these conditions hold, although it would not be a sufficient condition for the validity of the method. In this case, we need that possible time-varying unobserved confounders that may be correlated with treatment assignment also affect a large number of control units, so that a weighted average of the control units with diluted weights could absorb such effects. Under these conditions, the SC estimator would be asymptotically unbiased when both the number of pretreatment periods and the number of control units increase, even in settings where we should not expect to have a good pretreatment fit.

Supplementary Materials

Online supplemental appendix: Appendix A.1 presents the proofs of the main results, while Appendix A.2 presents extensions and additional theoretical results.

Replication package: R codes to replicate Tables 1 and A.1.

Acknowledgments

I thank to Alberto Abadie, Matias Cattaneo, the associate editor, and three anonymous referees for extremely valuable feedback that allowed me to substantially improve this article. I also thank to Aureo de Paula, Marcelo Fernandes, Antonio Galvao, Victor Filipe Martins-da-Rocha, Ricardo Masini, Eduardo Mendes, Whitney Newey, Vitor Possebom, Pedro

Sant'Annam and Kaspar Wuthrich for comments and suggestions. I also thank seminar and conference participants at the Chamberlain seminar, Warwick University, University of Cambridge, Queen Mary University, and of the Simonsen Lecture of the 2019 LACEA/LAMES conference. Luis Alvarez and Lucas Barros provided truly exceptional research assistance, being fundamental in both the derivation of the theoretical results and in the simulations. I also gratefully acknowledge financial support from FAPESP.

ORCID

Bruno Ferman http://orcid.org/0000-0001-7286-5677

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statiscal Association*, 105, 493–505. [1764,1765,1766,1767,1768]

Abadie, A., Diamond, A., and Hainmueller, J. (2015), "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 59, 495–510. [1764,1765,1768]

Abadie, A., and Gardeazabal, J. (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 93, 113–132. [1764,1765]

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2019), "Synthetic Difference in Differences," arXiv:1812.09970. [1768]

——— (2021), "Synthetic Difference in Differences." arXiv:1812.09970. [1764,1770]

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). "Matrix Completion Methods for Causal Panel Data Models," *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2021.1891924. [1764]

Athey, S., and Imbens, G. W. (2017), "The State of Applied Econometrics: Causality and Policy Evaluation," *Journal of Economic Perspectives*, 31, 3–32. [1764]

Bai, J. (2009), "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279. [1764]

Botosaru, I., and Ferman, B. (2019), "On the Role of Covariates in the Synthetic Control Method," *The Econometrics Journal*, 22, 117–130. [1766]

Carvalho, C., Masini, R., and Medeiros, M. C. (2018), "Arco: An Artificial Counterfactual Approach for High-Dimensional Panel Time-Series Data," *Journal of Econometrics*, 207, 352–380. [1770]

Cattaneo, M. D., Jansson, M., and Newey, W. K. (2018), "Inference in Linear Regression Models with many Covariates and Heteroscedasticity," *Journal of the American Statistical Association*, 113, 1350–1361. [1769]

Chernozhukov, V., Wuthrich, K., and Zhu, Y. (2021), "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls," *Journal of the American Statistical Association*, 1–16. [1764,1768,1770]

Doudchenko, N., and Imbens, G. (2016), "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis," available at *https://www.nber.org/papers/w22791*. [1764,1766,1768,1770]

Ferman, B., and Pinto, C. (2021), "Synthetic Controls with Imperfect Pre-Treatment Fit," *Quantitative Economics*, forthcoming. [1764,1766,1767,1768,1769]

Ferman, B., Pinto, C., and Possebom, V. (2020), "Cherry Picking With Synthetic Controls," *Journal of Policy Analysis and Management*, 39, 510–532. [1768]

Gobillon, L., and Magnac, T. (2016), "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *The Review of Economics and Statistics*, 98, 535–551. [1764]

Hsiao, C., Steve Ching, H., and Ki Wan, S. (2012), "A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China," *Journal of Applied Econometrics*, 27, 705–740. [1770]

Li, K. T., and Bell, D. R. (2017), "Estimation of Average Treatment Effects with Panel Data: Asymptotic Theory and Implementation," *Journal of Econometrics*, 197, 65–75. [1770]

Powell, D. (2021), "Synthetic Control Estimation Beyond Comparative Case Studies: Does the Minimum Wage Reduce Employment?" *Journal of Business & Economic Statistics*, 1–13. [1766]

Xu, Y. (2017), "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models," *Political Analysis*, 25, 57–76. [1764]