

# Synthetic controls with imperfect pretreatment fit

BRUNO FERMAN

Sao Paulo School of Economics-FGV

CRISTINE PINTO

Sao Paulo School of Economics-FGV

We analyze the properties of the Synthetic Control (SC) and related estimators when the pre-treatment fit is imperfect. In this framework, we show that these estimators are generally biased if treatment assignment is correlated with unobserved confounders, even when the number of pre-treatment periods goes to infinity. Still, we show that a demeaned version of the SC method can improve in terms of bias and variance relative to the difference-in-difference estimator. We also derive a specification test for the demeaned SC estimator in this setting with imperfect pre-treatment fit. Given our theoretical results, we provide practical guidance for applied researchers on how to justify the use of such estimators in empirical applications.

**KEYWORDS.** Synthetic control, difference-in-differences, policy evaluation, linear factor model.

**JEL CLASSIFICATION.** C13, C21, C23.

## 1. INTRODUCTION

In a series of influential papers, Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010), and Abadie, Diamond, and Hainmueller (2015) proposed the Synthetic Control (SC) method as an alternative to estimate treatment effects in comparative case studies when there is only one treated unit. The main idea of the SC method is to use the pretreatment periods to estimate weights such that a weighted average of the outcomes of the control units reconstructs the pretreatment outcomes of the treated unit, and then use these weights to compute the counterfactual of the treated unit in case it were not treated. According to Athey and Imbens (2017), “*the simplicity of the*

---

Bruno Ferman: [bruno.ferman@fgv.br](mailto:bruno.ferman@fgv.br)

Cristine Pinto: [cristine.pinto@fgv.br](mailto:cristine.pinto@fgv.br)

We thank Alberto Abadie, Josh Angrist, Ivan Canay, Marcelo Fernandes, Jin Hahn, Guido Imbens, Aureo de Paula, Tong Li, Victor Filipe Martins-da-Rocha, Ricardo Masini, Pedro Sant’Anna, Rodrigo Soares, and conference and seminar participants at UC Berkeley, UC Santa Barbara, USC, UDELAR, UCL, University of Bristol, USP, George Washington University, University of Miami, PUC-Chile, IPEA, PUC-Rio, California Econometrics Conference 2016, Bristol Econometric Study Group 2016, IAAE conference 2017, EEA conference 2017, LACEA 2017, and the MIT Conference on Synthetic Controls and Related Methods for comments and suggestions. We also thank Deivis Angeli, Lucas Barros, and Luis Alvarez for excellent research assistance. Bruno Ferman gratefully acknowledges financial support from FAPESP. An earlier version of this paper circulated under the title “Revisiting the Synthetic Control Estimator.”

*idea, and the obvious improvement over the standard methods, have made this a widely used method in the short period of time since its inception,”* making it “*arguably the most important innovation in the policy evaluation literature in the last 15 years.*” As one of the main advantages that helped popularize the method, [Abadie, Diamond, and Hainmueller \(2010\)](#) derived conditions under which the SC estimator would allow confounding unobserved characteristics with time-varying effects, as long as a weighted average of the control units using the SC weights perfectly fits the outcomes of the treated unit for a long set of preintervention periods.

In this paper, we analyze the properties of the SC and related estimators when potential outcomes are determined by a linear factor model. More specifically, we consider that potential outcome of unit  $j$  at time  $t$ , in the absence of treatment, is given by

$$y_{jt}^N = c_j + \delta_t + \lambda_t \mu_j + \epsilon_{jt}, \quad (1)$$

where  $c_j$  and  $\delta_t$  are unit- and time-invariant fixed effects,  $\lambda_t$  is an  $1 \times F$  vector of unobserved common factors,  $\mu_j$  is an  $F \times 1$  vector of unknown factor loadings, and  $\epsilon_{jt}$  are unobserved idiosyncratic shocks. This is the structure considered by [Abadie, Diamond, and Hainmueller \(2010\)](#) and [Abadie \(2020\)](#) to derive the main theoretical justifications for the SC estimator.

Differently from [Abadie, Diamond, and Hainmueller \(2010\)](#), we consider the case in which the pretreatment fit is imperfect.<sup>1</sup> In a model with “nondiverging” common factors and a fixed number of control units ( $J$ ), we show that the estimated SC weights converge in probability to weights that do *not*, in general, reconstruct the factor loadings of the treated unit when the number of pretreatment periods ( $T_0$ ) goes to infinity.<sup>2</sup> This happens because, in this setting, the SC weights converge to weights that simultaneously attempt to match the factor loadings of the treated unit *and* to minimize the variance of a linear combination of the idiosyncratic shocks. Therefore, weights that reconstruct the factor loadings of the treated unit are not generally the solution to this problem, even if such weights exist. While in many applications  $T_0$  may not be large enough to justify large- $T_0$  asymptotics (e.g., [Doudchenko and Imbens \(2016\)](#)), our results can also be interpreted as the SC weights not converging to weights that reconstruct the factor loadings of the treated unit *even when  $T_0$  is large*.

As a consequence, the SC estimator is biased if treatment assignment is correlated with the factor structure ( $\lambda_t \mu_j$ ), even when the number of pretreatment periods goes to infinity. The intuition is the following: if treatment assignment is correlated with the factor structure in the post-treatment periods, then we would need a SC unit that is affected

<sup>1</sup>We refer to “imperfect pretreatment fit” as a setting in which it is not assumed existence of weights such that a weighted average of the outcomes of the control unit perfectly fits the outcome of the treated unit for all pretreatment periods. The perfect pretreatment fit condition is presented in equation (2) of [Abadie, Diamond, and Hainmueller \(2010\)](#).

<sup>2</sup>We refer to “nondiverging” common factors when the pretreatment average of of the first and second moments of the common factors converge in probability to a constant. We focus on the SC specification that uses the outcomes of all pretreatment periods as predictors. Specifications that use the average of the pretreatment periods outcomes and other covariates as predictors are also considered in Appendix A.5 in the Online Supplemental Material ([Ferman and Pinto \(2021\)](#)).

in exactly the same way by the factor structure as the treated unit, but did not receive the treatment, to obtain an unbiased estimator. However, this condition is not attained when the pretreatment fit is imperfect, even when  $T_0$  is large.<sup>3</sup> Our results are not as conflicting with the results from [Abadie, Diamond, and Hainmueller \(2010\)](#) as it might appear at first glance. The asymptotic bias of the SC estimator, in our framework, goes to zero when the variance of the idiosyncratic shocks is small. This is the case in which one should expect to have a close-to-perfect pretreatment fit when  $T_0$  is large, which is the setting the SC estimator was originally designed for. Our theory complements the theory developed by [Abadie, Diamond, and Hainmueller \(2010\)](#), by considering the properties of the SC estimator when the pretreatment fit is imperfect.

One important implication of the SC restriction to convex combinations of the control units is that the SC estimator may also be biased if the SC unit fails to reconstruct the time-invariant fixed effect of the treated unit. Therefore, the SC estimator may be biased in settings in which the difference-in-differences (DID) estimator would be unbiased. We consider a modified SC estimator, where we demean the data using information from the preintervention period, and then construct the SC estimator using the demeaned data.<sup>4</sup> An advantage of demeaning is that it is possible to, under some conditions, show that the SC estimator dominates the DID estimator in terms of variance and bias in this setting. Moreover, we provide a specification test for the validity of the demeaned SC estimator in this setting with an imperfect pretreatment fit. Finally, we also show that, in a setting with both nondiverging and diverging common factors, diverging common shocks would not generate asymptotic bias in the demeaned SC estimator, but we need that treatment assignment is uncorrelated with the nondiverging common factors to guarantee asymptotic unbiasedness.<sup>5</sup>

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.<sup>6</sup> However, with fixed  $J$ , it is only possible to estimate factor loadings consistently under strong assumptions on the idiosyncratic shocks (e.g., [Bai \(2003\)](#) and [Anderson \(1984\)](#)). Therefore, the asymptotic bias we find for the SC estimator is

<sup>3</sup>[Ben-Michael, Feller, and Rothstein \(2018\)](#) derived finite-sample bounds on the bias of the SC estimator, and show that the bounds they derive do not converge to zero when  $J$  is fixed and  $T_0 \rightarrow \infty$ . This is consistent with our results, but does not directly imply that the SC estimator is asymptotically biased when  $J$  is fixed and  $T_0 \rightarrow \infty$ . In contrast, our result on the asymptotic bias of the SC estimator imply that it would be impossible to derive bounds that converge to zero in this case. Moreover, we show the conditions under which the estimator is asymptotically biased.

<sup>4</sup>Demeaning the data before applying the SC estimator is equivalent to relaxing the nonintercept constraint, as suggested, in parallel to our paper, by [Doudchenko and Imbens \(2016\)](#). We formally analyze the implication of this modification to the bias of the SC estimator. The estimator proposed by [Hsiao, Ching, and Wan \(2012\)](#) relaxes not only the the nonintercept but also the adding-up and nonnegativity constraints. We consider the properties of the estimator proposed by [Hsiao, Ching, and Wan \(2012\)](#) in Remark 5.

<sup>5</sup>For this result, we need an assumption of existence of weights that reconstruct the factor loadings of the treated unit associated with the diverging common factors. This result holds for the demeaned SC estimator, but not for the original SC estimator.

<sup>6</sup>Assuming that it is possible to construct a linear combination of the factor loadings of the control units that reconstructs the factor loadings of the treated unit, then this linear combination of the control units' outcomes would provide an unbiased counterfactual for the treated unit.

consistent with the results from a large literature on factor models. We show that the asymptotic bias we derive for the SC estimator also applies to other related panel data approaches that have been studied in the context of an imperfect pre-treatment fit, such as Hsiao, Ching, and Wan (2012), Li and Bell (2017), Carvalho, Masini, and Medeiros (2018), Carvalho, Masini, and Cunha Medeiros (2016), and Masini and Medeiros (2019), when we consider settings with fixed  $J$ . We show that these papers rely on assumptions that implicitly imply no selection on unobservables, which clarifies why their consistency/unbiasedness results when  $J$  is fixed are not conflicting with our main results.

Also consistent with the literature on factor models, if we impose restrictions on the idiosyncratic shocks, then there are asymptotically unbiased alternatives. For example, Amjad, Shah, and Shen (2018) proposed a denoising algorithm, but it relies on idiosyncratic errors being serially uncorrelated.<sup>7</sup> However, this may not be an appealing assumption in common applications. To the best of our knowledge, there is no estimator that is asymptotically valid in settings with fixed  $J$  without assuming such kind of additional assumptions. Finally, Powell (2018) proposed a 2-step estimation in a setting with fixed  $J$  in which the SC unit is constructed based on the fitted values of the outcomes on unit-specific time trends. However, we show that the demeaned SC method is already very efficient in controlling for polynomial time trends

When both  $J$  and  $T_0$  diverge, Gobillon and Magnac (2016), Xu (2017), Athey et al. (2018), and Arkhangelsky et al. (2018) provided alternative estimation methods that are asymptotically valid when the number of both pretreatment periods and controls increase. This is also consistent with the literature on linear factor models, which shows that these models can be consistently estimated in large panels (e.g., Bai (2003), Bai and Ng (2002), Bai (2009), and Moon, Roger, and Weidner (2015)). Ferman (2019) provided conditions under which the original and the demeaned SC estimators are also asymptotically unbiased in this setting with large  $J$ /large  $T_0$ . The main requirement is that, as the number of control units increases, there are weights diluted among an increasing number of control units that recover the factor loadings of the treated unit. However, if  $J$  and  $T_0$  are not large, then we should expect from our results the SC estimator to be biased if treatment assignment is correlated with the factor structure, even if  $J$  and  $T_0$  are roughly of the same magnitudes. Moreover, even if  $J$  and  $T_0$  are large, we should also expect the bias we derive to be relevant if the condition on diluted weights that recover the factor loadings of the treated unit does not hold.<sup>8</sup>

The remainder of this paper proceeds as follows. In Section 2, we describe our setting and provide a brief review of the SC estimator. The main results are presented in

---

<sup>7</sup>This is also the case for an IV-like SC estimator we presented in an earlier version of this paper (Ferman and Pinto (2019)).

<sup>8</sup>In this case, there would not be a sequence of weights that recover the factor loadings of the treated unit such that the variance of a linear combination of the idiosyncratic shocks using those weights goes to zero when  $J$  and  $T_0$  diverge. Therefore, the competing goals of the SC weights that we describe in our paper—that they simultaneously attempt to match the factor loadings of the treated unit and to minimize the variance of a linear combination of the idiosyncratic shocks—would remain relevant even when  $J$  and  $T_0$  diverge. In contrast, if the condition stated by Ferman (2019) holds, then it is possible to match the factor loadings of the treated unit with weights such that the variance of the linear combination of the idiosyncratic shocks using those weights goes to zero. Therefore, minimizing the variance of this linear combination of the idiosyncratic shocks would become asymptotically irrelevant.

Section 3. We then present a Monte Carlo (MC) simulation in Section 4, and an empirical illustration in Section 5. In Section 6, we provide a guideline for applied researchers on how to justify the use of the SC method, based on our theoretical results. We conclude in Section 7.

## 2. BASE MODEL

Suppose we have a balanced panel of  $J + 1$  units indexed by  $j = 0, \dots, J$  observed on a total of  $T$  periods. We want to estimate the treatment effect of a policy change that affected only unit  $j = 0$ , and we have information before and after the policy change. Let  $\mathcal{T}_0$  ( $\mathcal{T}_1$ ) be the set of time indices in the pretreatment (post-treatment) periods. We assume that potential outcomes follow a linear factor model.

**ASSUMPTION 1** (Potential outcomes). *Potential outcomes when unit  $j$  at time  $t$  is treated ( $y_{jt}^I$ ) and nontreated ( $y_{jt}^N$ ) are given by*

$$\begin{cases} y_{jt}^N = c_j + \delta_t + \lambda_t \mu_j + \epsilon_{jt}, \\ y_{jt}^I = \alpha_{jt} + y_{jt}^N, \end{cases} \quad (2)$$

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

In principle, the terms  $\delta_t$  and  $c_j$  could be included in the linear factor structure  $\lambda_t \mu_j$ . We include these separately because we want to consider  $\lambda_t$  as a vector of common factors that do not have constant effects across units and that do not include a time-invariant fixed effect. Therefore, we can think of  $\lambda_t$  as time-varying unobservables that may affect different units differently. In order to simplify the exposition of our main results, we consider the model without observed covariates  $Z_j$ . In Appendix Section A.5.2, we consider the model with covariates.

The treatment effect on unit  $j$  at time  $t$  is given by  $\alpha_{jt}$ , and the main goal of the SC method is to estimate the effect of the treatment on unit 0 for each post-treatment  $t$ , that is  $\{\alpha_{0t}\}_{t \in \mathcal{T}_1}$ . However, we only observe  $y_{jt} = d_{jt} y_{jt}^I + (1 - d_{jt}) y_{jt}^N$ , where  $d_{jt} = 1$  if unit  $j$  is treated at time  $t$ .

We treat the vector of unknown factor loadings ( $\mu_j$ ), the unit fixed effects ( $c_j$ ), and the treatment assignment as fixed, while we consider the properties of the SC estimator under a repeated sampling framework over the distributions of common factors ( $\lambda_t$ ), time effects ( $\delta_t$ ), and idiosyncratic shocks ( $\epsilon_{jt}$ ). Alternatively, we can think that we have an underlying model where treatment assignment,  $\mu_j$ , and  $c_j$  are also stochastic, but we are conditioning on these variables. Assumption 2 defines the observed sample.

**ASSUMPTION 2** (Sampling). *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 equation (2). We treat  $\{c_j, \mu_j\}_{j=0}^J$  as fixed, and  $\{\lambda_t\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$ ,  $\{\delta_t\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$ , and  $\{\epsilon_{jt}\}_{t \in \mathcal{T}_0 \cup \mathcal{T}_1}$  for  $j = 0, \dots, J$  as stochastic.*

In the assumption below, we consider the identification assumption usually considered in the SC literature.

**ASSUMPTION 3 (Idiosyncratic shocks).**  $\mathbb{E}[\epsilon_{jt}] = 0$  for all  $j \in \{0, 1, \dots, J\}$  and  $t \in \mathcal{T}_1 \cup \mathcal{T}_0$ .

Assumption 3, combined with the fact that we consider treatment assignment and factor loadings as fixed, compose the main restrictions we impose on the treatment assignment mechanism. It is easier to think about the assignment mechanism if we consider an underlying model in which treatment assignment and factor loadings are stochastic, and the expectation in Assumption 3 is conditional on the realization of these variables. In this case, Assumption 3 implies that idiosyncratic shocks are mean-independent from the treatment assignment. However, it does not impose any restriction on the dependence between treatment assignment and the factor structure. In particular, Assumption 3 does not impose any restriction on the distribution of  $\lambda_t$  for  $t \in \mathcal{T}_1$ . We refer to that as “selection on unobservables,” meaning that treatment assignment may be correlated with the factor structure, but is uncorrelated with the idiosyncratic shocks.<sup>9</sup>

Let  $\boldsymbol{\mu} \equiv [\mu_1 \dots \mu_J]'$ ,  $\mathbf{c} \equiv [c_1 \dots c_J]'$ ,  $\mathbf{y}_t \equiv (y_{1t}, \dots, y_{Jt})$  and  $\boldsymbol{\epsilon}_t \equiv (\epsilon_{1t}, \dots, \epsilon_{Jt})$ . Following the original SC papers, we start restricting to convex combinations of the control units, so we consider weights in  $\Delta^{J-1} \equiv \{(w_1, \dots, w_J) \in \mathbb{R}^J | w_j \geq 0 \text{ and } \sum_{j=1}^J w_j = 1\}$ . We define  $\tilde{\Phi} = \{\mathbf{w} \in \Delta^{J-1} | \mu_0 = \boldsymbol{\mu}'\mathbf{w} \text{ and } c_0 = \mathbf{c}'\mathbf{w}\}$ . Therefore,  $\mathbf{w}$  in the set  $\tilde{\Phi}$  is such that a weighted average of the control units absorbs all factor structure associated to the treated unit,  $\lambda_t \mu_0$ , and also the time-invariant fixed effect of the treated unit ( $c_0$ ). Assuming  $\tilde{\Phi}$  is not empty, if we knew  $\mathbf{w}^*$  in  $\tilde{\Phi}$ , then we could consider an *infeasible* SC estimator using these weights,  $\hat{\alpha}_{0t}^* = y_{0t} - \mathbf{y}'_t \mathbf{w}^*$ . For a given  $t \in \mathcal{T}_1$ , we would have

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

Therefore, under Assumption 3,  $\mathbb{E}[\hat{\alpha}_{0t}^*] = \alpha_{0t}$ , which implies that this infeasible SC estimator is unbiased. Intuitively, the infeasible SC estimator constructs a SC unit for the counterfactual of  $y_{0t}$  that is affected in the same way as unit 0 by each of the common factors (i.e.,  $\mu_0 = \boldsymbol{\mu}'\mathbf{w}^*$ ) and has the same time-invariant fixed effect ( $c_0 = \mathbf{c}'\mathbf{w}^*$ ), but did not receive treatment. Therefore, the only difference between unit 0 and this SC unit, beyond the treatment effect, would be given by the idiosyncratic shocks, which are assumed to have mean zero (Assumption 3), implying that this infeasible SC estimator is unbiased.

It is important to note that [Abadie, Diamond, and Hainmueller \(2010\)](#) do not make any assumption on  $\tilde{\Phi}$  being not empty. Instead, they consider that there is a set of weights  $\tilde{\mathbf{w}}^* \in \Delta^{J-1}$  that satisfies  $y_{0t} = \mathbf{y}'_t \tilde{\mathbf{w}}^*$  for all  $t \in \mathcal{T}_0$ .<sup>10</sup> We call the existence of such

<sup>9</sup>This assumption is essentially the same as the ones considered by, for example, [Abadie, Diamond, and Hainmueller \(2010\)](#), [Gobillon and Magnac \(2016\)](#), and [Ben-Michael, Feller, and Rothstein \(2018\)](#) (in their Section 4.1), where they assume unconfoundness conditional on the unobserved factor loadings.

<sup>10</sup>[Abadie, Diamond, and Hainmueller \(2010\)](#) assumed that such weights also provided perfect balance in terms of observed covariates. [Botosaru and Ferman \(2019\)](#) analyzed the case in which the perfect balance on covariates assumption is dropped, but there is still perfect balance on pretreatment outcomes.

weights  $\tilde{\mathbf{w}}^*$  as a “perfect pretreatment fit” condition. While subtle, this reflects a crucial difference between our setting and the setting considered in the original SC papers. Abadie, Diamond, and Hainmueller (2010) and Abadie, Diamond, and Hainmueller (2015) considered the properties of the SC estimator conditional on having a perfect preintervention fit. As stated by Abadie, Diamond, and Hainmueller (2015), they “do not recommend using this method when the pretreatment fit is poor or the number of pretreatment periods is small.”

Abadie, Diamond, and Hainmueller (2010) provided conditions under which existence of  $\tilde{\mathbf{w}}^* \in \Delta^{J-1}$  such that  $y_{0t} = \mathbf{y}'_t \tilde{\mathbf{w}}^*$  for all  $t \in \mathcal{T}_0$  (for large  $T_0$ ) implies that those weights approximately reconstruct the factor loadings of the treated unit (i.e.,  $\mu_0 \approx \boldsymbol{\mu}' \tilde{\mathbf{w}}^*$ ). In this case, the bias of the SC estimator would be bounded by a function that goes to zero when  $T_0$  increases. We depart from the original SC setting in that we consider a setting with imperfect pretreatment fit, meaning that we do not assume existence of  $\tilde{\mathbf{w}}^* \in \Delta^{J-1}$  such that  $y_{0t} = \mathbf{y}'_t \tilde{\mathbf{w}}^*$  for all  $t \in \mathcal{T}_0$ . The motivation to analyze the SC method in our setting is that the SC estimator has been widely used even when the pretreatment fit is far from perfect. Therefore, it is important to understand the properties of the estimator in this setting. Moreover, we show that the estimator can provide important improvements relative to DID even when the fit is imperfect, although in this case we should be more careful about the conditions for unbiasedness.

In order to implement their method, Abadie, Diamond, and Hainmueller (2010) recommended a nested minimization problem using the preintervention data to estimate the SC weights. We focus on the case where one includes all preintervention outcome values as predictors. In this case, the nested optimization problem proposed by Abadie, Diamond, and Hainmueller (2010) simplifies to<sup>11</sup>

$$\hat{\mathbf{w}}^{\text{SC}} = \underset{\mathbf{w} \in \Delta^{J-1}}{\operatorname{argmin}} \frac{1}{T_0} \sum_{t \in \mathcal{T}_0} [y_{0t} - \mathbf{y}'_t \mathbf{w}]^2. \quad (4)$$

For a given  $t \in \mathcal{T}_1$ , the SC estimator is then defined by  $\hat{\alpha}_{0t} = y_{0t} - \mathbf{y}'_t \hat{\mathbf{w}}^{\text{SC}}$ . Ferman, Pinto, and Possebom (2020) provided conditions under which the SC estimator using all pretreatment outcomes as predictors will be asymptotically equivalent, when  $T_0 \rightarrow \infty$ , to any alternative SC estimator such that the number of pretreatment outcomes used as predictors goes to infinity with  $T_0$ , even for specifications that include other covariates. Therefore, our results are also valid for these SC specifications under these conditions. In Appendix A.5, we also consider SC estimators using (1) the average of the preintervention outcomes as predictor, and (2) other time-invariant covariates in addition to the average of the preintervention outcomes as predictors.

### 3. MAIN RESULTS

We consider the asymptotic properties of the SC and alternative estimators when  $T_0$  diverges and  $J$  is fixed. As we discuss in Remark 2, our results are also relevant for the

<sup>11</sup>See Kaul et al. (2015) and Doudchenko and Imbens (2016).

case in which  $T_0$  is small. We consider the properties of the original SC estimator in Section 3.1 in a setting in which common factors are “nondiverging,” in the sense that the second moments of the pretreatment averages of the common factors and of the idiosyncratic shocks converge in probability to nonstochastic constants. We propose and analyze a demeaned version of the SC estimator in Section 3.2 in this setting. Then we discuss in Section 3.3 a setting in which some common factors are “diverging.”

### 3.1 Asymptotic bias of the original SC estimator

We consider a settings in which the pretreatment averages of the first and second moments of the common factors and the idiosyncratic shocks converge in probability to nonstochastic constants. Importantly, note we do not require that the observed outcomes  $y_{jt}$  satisfy these conditions, because we do not impose any restriction on  $\delta_t$ . We discuss in Section 3.3 the case in which diverging common shocks may have heterogeneous effects across units. Let  $\varepsilon_t = (\varepsilon_{0t}, \dots, \varepsilon_{Jt})$ .

**ASSUMPTION 4** (Common and idiosyncratic shocks).  $\frac{1}{T_0} \sum_{t \in \mathcal{T}_0} \lambda_t \xrightarrow{P} 0$ ,  $\frac{1}{T_0} \sum_{t \in \mathcal{T}_0} \varepsilon_t \xrightarrow{P} 0$ ,  $\frac{1}{T_0} \sum_{t \in \mathcal{T}_0} \lambda_t' \lambda_t \xrightarrow{P} \Omega_0$  positive semidefinite,  $\frac{1}{T_0} \sum_{t \in \mathcal{T}_0} \varepsilon_t \varepsilon_t' \xrightarrow{P} \sigma_\varepsilon^2 I_{J+1}$ , and  $\frac{1}{T_0} \sum_{t \in \mathcal{T}_0} \varepsilon_t \lambda_t \xrightarrow{P} 0$  when  $T_0 \rightarrow \infty$ .

Assumption 4 allows for serial correlation for both idiosyncratic shocks and common factors. The only restriction on the serial correlation is that we can apply a law of large numbers so that these pretreatment averages converge in probability. We assume  $\frac{1}{T_0} \sum_{t \in \mathcal{T}_0} \varepsilon_t \varepsilon_t' \xrightarrow{P} \sigma_\varepsilon^2 I_{J+1}$  in order to simplify the exposition of our results. However, this can be easily replaced by  $\frac{1}{T_0} \sum_{t \in \mathcal{T}_0} \varepsilon_t \varepsilon_t' \xrightarrow{P} \Sigma$  for any symmetric positive definite  $(J+1) \times (J+1)$  matrix  $\Sigma$ , so that idiosyncratic shocks may be heteroskedastic and correlated across  $j$ . Assuming that  $\frac{1}{T_0} \sum_{t \in \mathcal{T}_0} \lambda_t \xrightarrow{P} \omega_0$ , setting  $\omega_0 = 0$  is without loss of generality.<sup>12</sup> Assumption 4 would be satisfied if, for example,  $(\varepsilon_t', \lambda_t)$  is  $\alpha$ -mixing with exponential speed, with uniformly bounded fourth moments in the pretreatment period, and  $\varepsilon_t$  and  $\lambda_t$  are independent. Note that this would allow the distribution of  $\lambda_t$  to be different when we consider pretreatment periods closer to the assignment of the treatment. In this case,  $\lambda_t$  would not be stationary, but Assumption 4 would still hold. Finally, note that we do not impose any restriction on  $\delta_t$ .

We consider in Proposition 1 the asymptotic distributions of the original SC in this setting.

**PROPOSITION 1.** Under Assumptions 1 to 4,  $\widehat{\mathbf{w}}^{\text{SC}} \xrightarrow{P} \bar{\mathbf{w}}^{\text{SC}} \equiv (\bar{w}_1^{\text{SC}}, \dots, \bar{w}_J^{\text{SC}})$  when  $T_0 \rightarrow \infty$ , where  $(c_0, \mu_0) \neq (\mathbf{c}' \bar{\mathbf{w}}^{\text{SC}}, \boldsymbol{\mu}' \bar{\mathbf{w}}^{\text{SC}})$ , unless  $\sigma_\varepsilon^2 = 0$  or  $\tilde{\Phi} \cap \text{argmin}_{\mathbf{w} \in \Delta^{J-1}} \{\mathbf{w}' \mathbf{w}\} \neq \emptyset$ . Moreover, for  $t \in \mathcal{T}_1$ ,

$$\begin{aligned} \hat{\alpha}_{0t} = y_{0t} - \mathbf{y}'_t \widehat{\mathbf{w}}^{\text{SC}} &\xrightarrow{P} \alpha_{0t} + \lambda_t (\mu_0 - \boldsymbol{\mu}' \bar{\mathbf{w}}^{\text{SC}}) + (c_0 - \mathbf{c}' \bar{\mathbf{w}}^{\text{SC}}) \\ &+ (\varepsilon_{0t} - \boldsymbol{\varepsilon}'_t \bar{\mathbf{w}}^{\text{SC}}) \quad \text{when } T_0 \rightarrow \infty. \end{aligned} \quad (5)$$

<sup>12</sup>If  $\omega_0 \neq 0$ , then we can consider an observably equivalent model with  $\omega_0 = 0$  by adjusting  $c_j$ .



Proposition 1 shows that the weights of the original SC estimators will generally not converge to weights that recover the factor loadings of the treated unit. The intuition is that  $\widehat{\mathbf{w}}^{\text{SC}}$  converges in probability to  $\mathbf{w} \in \Delta^{J-1}$  that minimizes the probability limit of equation (4), which is given by

$$Q_0(\mathbf{w}) = [(c_0 - \mathbf{c}'\mathbf{w})^2 + (\mu_0 - \boldsymbol{\mu}'\mathbf{w})' \Omega_0(\mu_0 - \boldsymbol{\mu}'\mathbf{w})] + \sigma_\epsilon^2(1 + \mathbf{w}'\mathbf{w}). \quad (6)$$

This objective function has two parts. The first one reflects the presence of common factors  $\lambda_t$  and differences in the fixed effects that remain after we choose the weights to construct the SC unit. If  $\tilde{\Phi}$  is not empty, then we can set this part equal to zero by choosing  $\mathbf{w}^*$  in the set  $\tilde{\Phi}$ . However, this objective function also depends on the variance of a weighted average of the idiosyncratic shocks  $\epsilon_{jt}$ , implying that choosing  $\mathbf{w}^*$  in the set  $\tilde{\Phi}$  will not generally be the solution to this problem. As a consequence, the SC weights will generally converge to weights that do not recover the factor loadings of the treated unit, even if  $\tilde{\Phi}$  is not empty. There are two conditions in which the SC weights would asymptotically recover the factor loadings of the treated unit. First, if  $\sigma_\epsilon^2$  equals zero then any weighted average of the idiosyncratic shocks would have variance equal to zero, so any  $\mathbf{w}$  in the set  $\tilde{\Phi}$  would minimize the objective function  $Q_0(\mathbf{w})$ . Given this rationale, the distortion on the SC weights will tend to be smaller when the common trends are much stronger than the idiosyncratic shocks (so that the second part of the objective function  $Q_0(\mathbf{w})$  becomes less relevant). Alternatively, if there are weights  $\mathbf{w}$  in the set  $\tilde{\Phi}$  that also minimize the variance of the weighted average of the idiosyncratic shocks, then such weights would also minimize the objective function  $Q_0(\mathbf{w})$ .<sup>13</sup> We present details of proof in Appendix A.1.1. Another intuition for this result is that the outcomes of the controls work as proxy variables for the factor loadings of the treated unit, but they are measured with error. We present this interpretation in more detail in Appendix A.2.

Proposition 1 also shows that the SC estimator converges in probability to the parameter we want to estimate ( $\alpha_{0t}$ ) plus linear combinations of contemporaneous idiosyncratic shocks and common factors.<sup>14</sup> By Assumption 3,  $\mathbb{E}[\epsilon_{jt}] = 0$ , so whether this estimator is asymptotically unbiased depends crucially on the differences in how the treated and the SC units are affected by the common shocks,  $\lambda_t(\mu_0 - \boldsymbol{\mu}'\bar{\mathbf{w}}^{\text{SC}})$ , and on whether the SC unit reconstructs  $c_0$ . We can guarantee asymptotic unbiasedness for the original SC estimator if we assume that, for  $t \in \mathcal{T}_1$ ,  $\mathbb{E}[\lambda_t^k] = 0$  for all common factors  $k$  such that the factor loadings of the treated unit associated with these common factors are not asymptotically reconstructed by the SC weights (i.e.,  $\mu_0^k \neq \sum_{j \neq 0} \bar{w}_j^{\text{SC}} \mu_j^k$ ), and also that the SC weights asymptotically reconstruct the fixed effect of treated unit (i.e.,  $c_0 = \mathbf{c}'\bar{\mathbf{w}}^{\text{SC}}$ ).<sup>15</sup> Therefore, once we relax the perfect fit condition, Assumption 3, which

<sup>13</sup>Note that, if we relax the assumption that the idiosyncratic errors are homoskedastic, then the weights that minimize the variance of the weighted average of the idiosyncratic shocks will not necessarily be  $1/J$  for all control units.

<sup>14</sup>For simplicity, we consider the case in which  $\alpha_{0t}$  is a fixed parameter. More generally, we could consider  $\alpha_{0t}$  stochastic, and redefine the parameter of interest as  $\mathbb{E}[\alpha_{0t}]$ . The intuition for all results would remain unchanged.

<sup>15</sup>There could also be linear combinations of biases arriving from different common factors that end up canceling out, but we see that as uninteresting “knife-edge” cases.

allows for selection on unobservables, is not sufficient to guarantee that the SC estimator is asymptotically unbiased. This means that, even if the true model for the potential outcomes follows a linear factor model as considered in Assumption 1, the SC estimator may be asymptotically biased. More specifically, Proposition 1 shows that, once we relax the perfect fit condition, the original SC estimator will generally be asymptotically biased when treatment assignment is correlated with time-varying unobservables, and when the SC weights fail to recover the levels of the treated unit. This second conclusion implies that the SC estimator may be biased even when a DID estimator would be unbiased.

REMARK 1. The discrepancy of our results with the results from Abadie, Diamond, and Hainmueller (2010) arises because we consider different frameworks. Abadie, Diamond, and Hainmueller (2010) considered the properties of the SC estimator conditional on having a perfect pretreatment fit. Our results are not as conflicting with the results from Abadie, Diamond, and Hainmueller (2010) as they may appear at first glance. In a model with nondiverging common factors, the probability that one has a data set at hand such that the SC weights provide a close-to-perfect preintervention fit with a moderate  $T_0$  is close to zero, unless the variance of the idiosyncratic shocks is small. Therefore, our results agree with the theoretical results from Abadie, Diamond, and Hainmueller (2010) in that the asymptotic bias of the SC estimator should be small in situations where one would expect to have a close-to-perfect fit for a large  $T_0$ .

REMARK 2. While many SC applications do not have a large number of pretreatment periods to justify large- $T_0$  asymptotics (see, e.g., Doudchenko and Imbens (2016)), our results can also be interpreted as the SC weights not converging to weights that reconstruct the factor loadings of the treated unit when  $J$  is fixed *even when  $T_0$  is large*. In Appendix A.2, we show that the problem we present remains if we consider a setting with finite  $T_0$ .

REMARK 3. Related to Remarks 1 and 2, if  $T_0$  is very small relative to  $J$ , then the objective function in equation (4) may be close to zero because the SC weights are chosen so that the idiosyncratic shocks compensate discrepancies between the factor loadings of the treated unit ( $\mu_0$ ) and the implied factor loadings of the SC unit ( $\mu' \hat{\mathbf{w}}^{\text{SC}}$ ). That is, a good pretreatment fit might be achieved due to overfitting. In this case, we should not expect that the SC weights approximately reconstruct  $\mu_0$ , and the bias we derive for the SC estimator when treatment assignment is correlated with time-varying unobservables remains relevant. Therefore, the bias we derive for the SC estimator in Proposition 1 does not come from the fact that it becomes harder to have a good pretreatment fit when  $T_0$  increases. On the contrary, this problem remains relevant even when  $T_0$  is small.<sup>16</sup>

<sup>16</sup>Consistent with this idea, the bounds on the bias of the SC estimator derived by Abadie, Diamond, and Hainmueller (2010) only goes to zero when  $T_0$  increases. Therefore, we have no guarantee that the bias of the SC estimator is small when  $T_0$  is not large, even when we consider a setting in which we have a perfect pretreatment fit.

3.2 Comparison with DID estimator & the demeaned SC estimator

In contrast to the SC estimator, the DID estimator for the treatment effect in a given post-intervention period  $t \in \mathcal{T}_1$  would be given by

$$\hat{\alpha}_{0t}^{DID} = y_{0t} - \frac{1}{J} \mathbf{y}'_t \mathbf{i} - \frac{1}{T_0} \sum_{\tau \in \mathcal{T}_0} \left[ y_{0\tau} - \frac{1}{J} \mathbf{y}'_{\tau} \mathbf{i} \right], \tag{7}$$

where  $\mathbf{i}$  is a  $J \times 1$  vector of ones.<sup>17</sup> Under Assumptions 1, 2, and 4, we have that

$$\hat{\alpha}_{0t}^{DID} \xrightarrow{P} \alpha_{0t} + \left( \epsilon_{0t} - \frac{1}{J} \boldsymbol{\epsilon}'_t \mathbf{i} \right) + \lambda_t \left( \mu_0 - \frac{1}{J} \boldsymbol{\mu}' \mathbf{i} \right) \text{ when } T_0 \rightarrow \infty. \tag{8}$$

Therefore, the DID estimator will be asymptotically unbiased in this setting if  $\mathbb{E}[\lambda_t] = 0$  for the factors such that the factor loadings of the treated unit are not reconstructed by a simple average of the control units (i.e.,  $\mu_0 \neq \frac{1}{J} \boldsymbol{\mu}' \mathbf{i}$ ). This would be the case if treatment assignment is uncorrelated with the time-varying common factors. Differently from the SC estimator, however, the DID estimator would *not* be biased if the average of the control units does not recover the fixed effect of the treated unit.

As an alternative to the standard SC estimator, we suggest a modification in which we calculate the pretreatment average for all units and demean the data. This is equivalent to a generalization of the SC method suggested, in parallel to our paper, by Doudchenko and Imbens (2016), which includes an intercept parameter in the minimization problem to estimate the SC weights and construct the counterfactual.<sup>18</sup> Here, we formally consider the implications of this alternative on the bias and variance of the SC estimator.

The demeaned SC estimator is given by  $\hat{\alpha}_{0t}^{SC'} = y_{0t} - \mathbf{y}'_t \hat{\mathbf{w}}^{SC'} - (\bar{y}_0 - \bar{\mathbf{y}}' \hat{\mathbf{w}}^{SC'})$ , where  $\bar{y}_0$  is the pretreatment average of unit 0, and  $\bar{\mathbf{y}}$  is an  $J \times 1$  vector with the pretreatment averages of the controls. We define  $\Phi = \{\mathbf{w} \in \Delta^{J-1} \mid \mu_0 = \boldsymbol{\mu}' \mathbf{w}\}$ . Therefore, any  $\mathbf{w}$  in the set  $\Phi$  is such that a weighted average of the control units absorbs all time correlated shocks of unit 0,  $\lambda_t \mu_0$ . However, such weights do not necessarily absorb the time-invariant fixed effects. In this case, the weights  $\hat{\mathbf{w}}^{SC'}$  are given by

$$\hat{\mathbf{w}}^{SC'} = \underset{\mathbf{w} \in \Delta^{J-1}}{\operatorname{argmin}} \frac{1}{T_0} \sum_{t \in \mathcal{T}_0} [y_{0t} - \mathbf{y}'_t \mathbf{w} - (\bar{y}_0 - \bar{\mathbf{y}}' \mathbf{w})]^2. \tag{9}$$

**PROPOSITION 2.** Under Assumptions 1, 2, 3, and 4,  $\hat{\mathbf{w}}^{SC'} \xrightarrow{P} \bar{\mathbf{w}}^{SC'} \equiv (\bar{w}_1^{SC'}, \dots, \bar{w}_J^{SC'})$  when  $T_0 \rightarrow \infty$ , where  $\mu_0 \neq \boldsymbol{\mu}' \bar{\mathbf{w}}^{SC'}$ , unless  $\sigma_{\epsilon}^2 = 0$  or  $\Phi \cap \operatorname{argmin}_{\mathbf{w} \in \Delta^{J-1}} \{\mathbf{w}' \mathbf{w}\} \neq \emptyset$ . Moreover, for  $t \in \mathcal{T}_1$ ,

$$\hat{\alpha}_{0t}^{SC'} \xrightarrow{P} \alpha_{0t} + (\epsilon_{0t} - \boldsymbol{\epsilon}'_t \bar{\mathbf{w}}^{SC'}) + \lambda_t (\mu_0 - \boldsymbol{\mu}' \bar{\mathbf{w}}^{SC'}) \text{ when } T_0 \rightarrow \infty. \tag{10}$$

<sup>17</sup>Note that the DID estimator in this case with one treated unit is numerically the same as the two-way fixed effects (TWFE) estimator using unit and time fixed effects. Since the goal in the SC literature is to estimate the effect of the treatment for unit 1 at a specific date  $t$ , this circumvents the problem of aggregating heterogeneous effects, as considered by de Chaisemartin and D’Haultfauille (2020), Callaway and Sant’Anna (2018), Athey and Imbens (2018), and Goodman-Bacon (2018) in the DID setting.

<sup>18</sup>Relaxing the nonintercept constraint was already a feature of Hsiao, Ching, and Wan (2012). The difference here is that we relax this constraint while maintaining the adding-up and nonnegativity constraints, which allows us to rank the demeaned SC with the DID estimator under some conditions.

Therefore, when potential outcomes follow a linear factor model, both the demeaned SC and the DID estimators are asymptotically unbiased when  $\mathbb{E}[\lambda_t] = 0$  for  $t \in \mathcal{T}_1$ .<sup>19</sup> This means that these estimators are asymptotically unbiased if treatment assignment is not correlated with time-varying unobservables. Importantly, differently from the original SC estimator, these estimators do not require that the weights recover the time-invariant fixed effect of the treated unit for unbiasedness. Therefore, Proposition 2 shows that the demeaned SC estimator is asymptotically unbiased under the usual identification assumptions considered when we rely on the DID estimator. The proof is essentially the same as the one for Proposition 1 (details in Appendix A.1.2).

With additional assumptions on  $(\epsilon_{0t}, \dots, \epsilon_{Jt}, \lambda_t')$  in the post-treatment periods, we can also assure that the demeaned SC estimator is asymptotically more efficient than DID.

**ASSUMPTION 5** (Stability in the pre and post-treatment periods). *For  $t \in \mathcal{T}_1$ ,  $\mathbb{E}[\lambda_t] = 0$ ,  $\mathbb{E}[\epsilon_t] = 0$ ,  $\mathbb{E}[\lambda_t' \lambda_t] = \Omega_0$ , and  $\mathbb{E}[\epsilon_t \epsilon_t'] = \sigma_\epsilon^2 I_{J+1}$ ,  $\text{cov}(\epsilon_t, \lambda_t) = 0$ .*

Assumptions 4 and 5 imply that idiosyncratic shocks and common factors have the same first and second moments in the pre and post-treatment periods. Again, the assumptions that idiosyncratic errors are homoskedastic is made just for simplification. What is crucial in this assumption is that the variance/covariance matrix of the idiosyncratic shocks in the post-treatment periods is the same as the long-run variance/covariance matrix of the idiosyncratic shocks in the pretreatment periods. From Proposition 2, Assumption 5 implies that the demeaned SC estimator is asymptotically unbiased. We now show that, when potential outcomes follow a linear factor model, this assumption also implies that the demeaned SC estimator has lower asymptotic MSE than the DID estimator.

**PROPOSITION 3.** *Under Assumptions 1 to 5, the demeaned SC estimator ( $\hat{\alpha}_{0t}^{\text{SC}'}$ ) dominates the DID estimator ( $\hat{\alpha}_{0t}^{\text{DID}}$ ) in terms of asymptotic MSE when  $T_0 \rightarrow \infty$ .*

The intuition of this result is that, under Assumption 5, the demeaned SC weights converge to weights that minimize a function  $\Gamma(\mathbf{w})$  such that  $\Gamma(\bar{\mathbf{w}}^{\text{SC}'}) = a \cdot \text{var}(\hat{\alpha}_{0t}^{\text{SC}'})$ , and  $\Gamma(\{\frac{1}{J}, \dots, \frac{1}{J}\}) = a \cdot \text{var}(\hat{\alpha}_{1t}^{\text{DID}})$ . Therefore, it must be that the asymptotic variance of  $\hat{\alpha}_{0t}^{\text{SC}'}$  is weakly lower than the variance of  $\hat{\alpha}_{1t}^{\text{DID}}$ . Moreover, these estimators are unbiased under these assumptions (details in Appendix A.1.3).

If treatment assignment is correlated with time-varying unobservables (i.e.,  $\mathbb{E}[\lambda_t] \neq 0$  for  $t \in \mathcal{T}_1$ ), then both the demeaned SC and the DID estimators would generally be asymptotically biased. In general, it is not possible to rank the demeaned SC and the DID estimators in terms of bias and MSE if treatment assignment is correlated with time-varying common factors. We provide in Appendix A.4 a specific example in which

<sup>19</sup>This is a sufficient condition. More generally, the demeaned SC estimator would be asymptotically unbiased if  $\mathbb{E}[\lambda_t^k] = 0$  for  $t \in \mathcal{T}_1$  for any common factor  $k$  such that  $\mu_0^k \neq \sum_{j \neq 0} \bar{w}_j^{\text{SC}'} \mu_j^k$ . However, as we show in Proposition 2, if  $\sigma_\epsilon^2 > 0$ , then we would only have  $\mu_0^k = \sum_{j \neq 0} \bar{w}_j^{\text{SC}'} \mu_j^k$  in knife-edge cases. Therefore, we focus on the sufficient condition  $\mathbb{E}[\lambda_t] = 0$  for  $t \in \mathcal{T}_1$ .

the DID can have a smaller bias relative to the demeaned SC estimator. This might happen when selection into treatment depends on common factors with low variance, and it happens that a simple average of the controls provides a good match for the factor loadings associated with these common factors. In general, however, we should expect a lower bias for the demeaned SC estimator, given that the demeaned SC weights are *partially* chosen to minimize the distance between  $\mu_0$  and  $\boldsymbol{\mu}'\widehat{\mathbf{w}}^{SC}$ , while the DID estimator uses weights that are not data driven.

Since the biases of these two estimators would generally differ when  $\mathbb{E}[\lambda_t] \neq 0$  for  $t \in \mathcal{T}_1$ , we can consider a specification test by contrasting these two estimators. More specifically, if the DID estimator is very different from the demeaned SC estimator, this would suggest that both estimators are biased (considering the setting in which the pretreatment fit is imperfect).

A potential problem in properly testing the equality of these two estimators is that they are generally not asymptotically normal. Still, if we consider a stronger assumption that  $\lambda_t$  and  $\epsilon_{jt}$  are stationary and weakly dependent for all periods (both pre and post-intervention)—which implies that  $\mathbb{E}[\lambda_t] = 0$  for  $t \in \mathcal{T}_1$ —then we can follow the idea from Chernozhukov, Wuthrich, and Zhu (2017) and test this condition using in-time placebos. More specifically, let  $\tilde{\mathbf{w}}$  be the demeaned SC weights using all periods to estimate the SC weights. We consider

$$\hat{u}_t = \left( \tilde{\mathbf{w}}' \mathbf{y}_t - \frac{1}{T_0 + T_1} \sum_{\tau \in \mathcal{T}_0 \cup \mathcal{T}_1} (\tilde{\mathbf{w}}' \mathbf{y}_\tau) \right) - \left( J^{-1} \mathbf{i}' \mathbf{y}_t - \frac{1}{T_0 + T_1} \sum_{\tau \in \mathcal{T}_0 \cup \mathcal{T}_1} (J^{-1} \mathbf{i}' \mathbf{y}_\tau) \right). \quad (11)$$

The idea is that  $\hat{u}_t$  contrasts the demeaned and the DID estimators. The outcomes for the treated unit do not appear directly in this expression because they cancel out when we contrast the two estimators, but they are used in the estimation of  $\tilde{\mathbf{w}}$ . Following Chernozhukov, Wuthrich, and Zhu (2017), we impose the null  $\mathbb{E}[\lambda_t] = 0$  for  $t \in \mathcal{T}_1$  and estimate the model using all periods of data to provide better finite sample properties.

The main idea is that, in this linear factor model setting,  $\lambda_t$  and  $\epsilon_{jt}$  stationary and weakly dependent, implies that  $\hat{u}_t$  will approximately be stationary and weakly dependent. Therefore, we can construct a test statistic  $S(\hat{\mathbf{u}}) = |\frac{1}{T_1} \sum_{t \in \mathcal{T}_1} \hat{u}_t|$ , and derive the distribution of the test statistic by considering the set of all moving block permutations of the time periods. Let  $\mathcal{T}_0 = \{1, \dots, T_0\}$  and  $\mathcal{T}_1 = \{T_0 + 1, \dots, T\}$ , and let  $\Pi$  be the set of permutations  $\pi_j$  indexed by  $j \in 0, \dots, T - 1$  such that

$$\pi_j(i) = \begin{cases} i + j & \text{if } i + j \leq T, \\ i + j - T & \text{otherwise.} \end{cases} \quad (12)$$

Then the  $p$ -value of the specification test is given by

$$\hat{p} = \frac{1}{T} \sum_{j=0}^{T-1} \mathbf{1}\{S(\hat{\mathbf{u}}) > S(\hat{\mathbf{u}}_{\pi_j})\}, \quad (13)$$

We formalize this idea in the following proposition.

PROPOSITION 4. Assume  $\lambda_t$  and  $\epsilon_{jt}$  are stationary and weakly dependent, with finite second moments, and that Assumptions 1 and 2 hold. Then, for any  $a \in (0, 1)$ ,  $\Pr(\hat{p} \leq a) \rightarrow a$  when  $T_0 \rightarrow \infty$  and  $T_1$  is fixed.

If we find a low  $\hat{p}$ , indicating that the DID and the demeaned SC estimators are significantly different, then this would be an indication that both estimators are biased (considering the case in which the pretreatment fit is imperfect). In contrast, a high  $\hat{p}$  would provide some evidence that the condition  $\mathbb{E}[\lambda_t] = 0$  for  $t \in \mathcal{T}_1$  is valid. If  $\lambda_t$  and  $\epsilon_{jt}$  are serially uncorrelated, then this test is exact. If there is serial correlation, though, then we may have distortions when  $T_0$  is finite, but the test is asymptotically valid when  $T_0 \rightarrow \infty$ .

Importantly, this test is completely uninformative about Assumption 3. Moreover, it relies on stationarity as an auxiliary assumption, which is not a necessary assumption for validity of the DID and the demeaned SC estimators in this setting. We also recommend applied researchers should plot the demeaned SC and the DID estimators to provide a visual inspection of the differences between these two estimators.

REMARK 4. In general, it is not possible to compare the original and the demeaned SC estimators in terms of bias and variance. For example, if units with similar factor loadings also have similar fixed effects, then matching also on the levels would help provide a better approximation to  $\mu_0$ . Moreover, the demeaning process may increase the variance of the estimator for a finite  $T_0$ . Finally, demeaning essentially implies extrapolation, while some may consider that one of the advantages of the original SC estimator is that it avoids extrapolation (e.g., Abadie, Diamond, and Hainmueller (2015)). Therefore, it is not clear whether demeaning is the best option in all applications, and the use of this estimator depends on the willingness of the researcher to allow for extrapolation.

REMARK 5. Our main result that the original and the demeaned SC estimators are generally asymptotically biased if there are unobserved time-varying confounders (Propositions 1 and 2) still applies if we also relax the nonnegative and the adding-up constraints, which essentially leads to the panel data approach suggested by Hsiao, Ching, and Wan (2012), and further explored by Li and Bell (2017). Our conditions for unbiasedness of the SC estimator also apply to the estimators proposed by Carvalho, Masini, and Medeiros (2018) and Carvalho, Masini, and Cunha Medeiros (2016) when  $J$  is fixed. In Appendix A.5.3, we show that these papers rely on assumptions that implicitly imply that there is no selection on time-varying unobservables. This clarifies what selection on unobservables means in this setting, and reconciles our findings with the asymptotic unbiasedness/consistency results in these papers.

### 3.3 Model with “diverging” common factors

While the assumptions considered in Sections 3.1 and 3.2 allow for outcomes with divergent pre-treatment averages (which would be the case when we consider, e.g., GDP or average wages), we restrict to settings in which such diverging common shocks affect all

units in the same way. In Appendix A.3, we modify Assumption 1 to consider the case in which we may have diverging common shocks with heterogeneous effects across unit.

Assuming that there exist weights that reconstruct the factor loadings of the treated unit associated to the diverging common shocks, we show that the asymptotic distribution of the *demeaned* SC estimator does not depend on such diverging common shocks when  $T_0$  diverges. The intuition is that, as  $T_0$  diverges, the variance of the weighted average of the idiosyncratic shocks becomes irrelevant relative to the cost of failing to recover the factor loadings associated with the diverging common shocks. This is consistent with the conclusion from Section 3.1 that the bias of the SC estimator should be less relevant when the common shocks are stronger relative to the idiosyncratic shocks. However, if we also have nondiverging common shocks, then the demeaned SC weights will generally not asymptotically recover the factor loadings of the treated unit associated with those nondiverging shocks. This implies that the demeaned SC estimator may be asymptotically biased if there is correlation between treatment assignment and these nondiverging shocks, for exactly the same reasons outlined in Section 3.1.

We also show that the conclusion that the demeaned SC estimator does not depend on the diverging common shocks is not valid for the original SC estimator. While the SC weights considering the original SC method converge in probability to weights that recover the factor loadings of the treated associated to the diverging common shocks, this convergence may not be fast enough to compensate that such common shocks are diverging. We present all details on this setting in Appendix A.3.

#### 4. MONTE CARLO SIMULATIONS

To illustrate our theoretical findings, we construct a MC simulation based on a real data set using the monthly employment data for 50 US states and the District of Columbia ( $J + 1 = 51$ ) from January 1982 to December 2019 ( $T = 456$ ). We construct these series by aggregating the Current Population Survey (CPS) microdata at the state  $\times$  month level.<sup>20</sup> We estimate a factor model that best approximates this data. In addition to the state and time fixed effects, we estimate four common factors.<sup>21</sup> We find evidence that these four factors and the 51 idiosyncratic shocks are stationary, suggesting that any nonstationary trends in the outcomes come from the time fixed effects,  $\delta_t$ .<sup>22</sup> This provides evidence that this data set is well approximated by the setting we consider in Sections 3.1 and 3.2. We consider state-specific Gaussian ARMA models for the distribution of each  $\epsilon_{jt}$  and also for the distribution of the four common factors  $\lambda_t$ .<sup>23</sup>

<sup>20</sup>We created our CPS extract using IPUMS (Ruggles et al. (2015)).

<sup>21</sup>We estimate the linear factor model using the iterated fixed effects method proposed by Bai (2009). The number of factors was selected using the  $IC_{p1}$  criterion in Bai and Ng (2002). We used the `interFE` function in the package `gsynth` (Xu (2017)).

<sup>22</sup>For each of these series, we consider the Augmented Dickey–Fuller test for a unit root, where the number of lags is chosen using the MAIC criterion of Ng and Perron (2001). We also test for the presence of deterministic trends using the test statistics proposed by Dickey and Fuller (1981).

<sup>23</sup>For each estimated factor and for each state time series of residuals, we use the `auto.arima` function in the R package `forecast` to perform grid search over the autoregressive and moving-average dimensions; and select the best model according to the BIC criterion (Hyndman and Khandakar (2008)).

For each simulation, we fix the estimated factor loadings  $\mu_j$  and consider random draws for  $\lambda_t$  and  $\epsilon_{jt}$ . We set  $T_1 = 12$  (1 year) and  $T_0 \in \{120, 240, 480, 1200\}$ , and we vary which state is considered as treated. We consider 5000 replications for each scenario. The case with  $T_0 = 480$  have approximately the same number of pretreatment periods as we have in our original data set, while the cases with smaller  $T_0$  reflect more common setting in which we have 10 or 20 years of pretreatment data. We include the case  $T_0 = 1200$  to approximate the asymptotic behavior of the estimators when  $T_0 \rightarrow \infty$ .

We consider in Panels A and B of Table 1 the case in which  $\mathbb{E}[\lambda_t]$  equals zero for  $t \in \mathcal{T}_1$ . In Panel A, we consider that the treated state is such that its time-invariant fixed effect ( $c_0$ ) is the second largest value in the distribution of  $c_j$ , while in Panel B the treated state has the second smallest value of  $c_0$ . We find that the original SC estimator is biased in both cases, even though it would be possible to have weights that reconstruct  $c_0$ . The problem is that such weights would be very concentrated on the state with largest (or smallest)  $c_j$ , so the SC weights would converge to weights that are more diluted, even if this means not recovering  $c_0$ . This is consistent with Proposition 1.

TABLE 1. Monte Carlo simulations.

	Bias			Standard error		
	SC (1)	Demeaned SC (2)	DID (3)	SC (4)	Demeaned SC (5)	DID (6)
Panel A: second largest $c_j$ , no break						
$T_0 = 120$	0.243	0.009	0.020	0.531	0.573	0.915
$T_0 = 240$	0.200	-0.002	0.006	0.505	0.532	0.960
$T_0 = 480$	0.196	0.006	0.007	0.501	0.514	1.007
$T_0 = 12,000$	0.167	-0.004	-0.010	0.503	0.500	1.005
Panel B: second smallest $c_j$ , no break						
$T_0 = 120$	-0.565	0.020	0.028	1.173	0.823	1.368
$T_0 = 240$	-0.601	-0.012	0.002	1.242	0.801	1.431
$T_0 = 480$	-0.596	-0.011	0.004	1.264	0.755	1.469
$T_0 = 12,000$	-0.581	-0.002	-0.003	1.242	0.714	1.442
Panel C: second largest $\mu_{1j}$ , break in factor 1						
$T_0 = 120$	1.142	1.043	2.143	0.772	0.819	1.247
$T_0 = 240$	0.934	0.797	2.129	0.728	0.764	1.308
$T_0 = 480$	0.808	0.675	2.110	0.713	0.728	1.337
$T_0 = 12,000$	0.718	0.598	2.112	0.688	0.691	1.310
Panel D: second smallest $\mu_{1j}$ , break in factor 1						
$T_0 = 120$	-1.821	-1.851	-3.176	1.200	1.167	1.777
$T_0 = 240$	-1.573	-1.532	-3.157	1.130	1.113	1.794
$T_0 = 480$	-1.415	-1.324	-3.110	1.082	1.058	1.816
$T_0 = 12,000$	-1.366	-1.230	-3.162	1.028	0.996	1.793

*Note:* This table presents the MC simulations discussed in Section 4. Panels A and B consider the case in which all common factors have mean zero in the post-treatment periods. In Panel A, the treated unit is the state with second largest fixed effect in the distribution of  $c_j$ , while in Panel B the treated unit is the state with the second smallest fixed effect. Panels C and D consider the case in which the first common factor has expected value equal to two times its standard deviation in the post-treatment periods. In Panel C, the treated unit is the state with second largest factor loadings associated to the first common factor in the distribution of  $\mu_{1j}$ , while in Panel D the treated unit is the state with the second smallest factor loading. In all simulations, the true treatment effect is equal to zero.



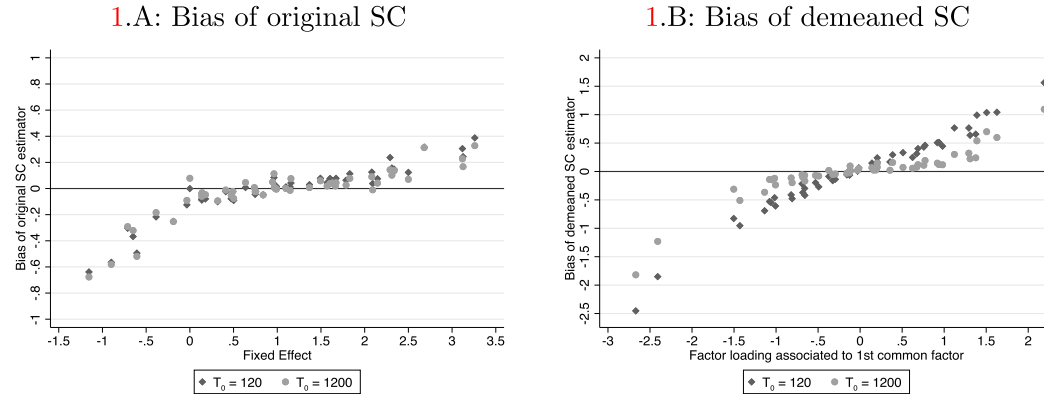


FIGURE 1. Monte Carlo simulations. Notes: Figure A presents the bias of the original SC estimator as a function of the time-invariant fixed effect of the treated state. We consider in this case a setting with no structural break for the common factors. Figure B presents the bias of the demeaned SC estimator as a function of the factor loadings associated to the first common factor of the treated state. We consider the case in which the first common factor has expected value equal to two times its standard deviation in the post-treatment periods. We present the settings with  $T_0 = 120$  and  $T_0 = 1200$ . In all simulations, the true treatment effect is equal to zero.

Figure 1.A shows the bias of the original SC estimator as a function of the time-invariant fixed effect of the treated state. We find relevant bias of the SC estimator when the time-invariant fixed effect of the treated state is in the extreme of the distribution of time-invariant fixed effects, while the bias is closer to zero when the treated state is more in the middle of this distribution. This happens because, as we consider a treated state with time-invariant fixed effect more toward the center of this distribution, we can have a weighted average of the control states with more diluted weights that reconstruct the time-invariant fixed effect of the treated. Therefore, the term in equation (6) related to the variance of the linear combination of idiosyncratic shocks becomes less relevant in the minimization problem. Indeed, if we consider a measure of concentration of weights given by  $\|\widehat{\mathbf{w}}^{\text{SC}}\|_2$ , then the correlation between the absolute value of the bias of the SC estimator and this measure ranges from 0.723 to 0.849 (depending on the value of  $T_0$ ).<sup>24</sup>

In contrast to the original SC estimator, both the demeaned SC and the DID estimators are unbiased in this setting, as presented in Panel A of Table 1. This is consistent with Proposition 2. Moreover, as expected from Proposition 3, the demeaned SC estimator is more efficient than the DID estimator, with roughly 40%–50% smaller standard errors.

In Panels C and D of Table 1, we consider a setting in which the first common factor has expected value equal to two times its standard deviation in the post-treatment periods. In Panel C, we consider the case in which the treated state is such that its factor loading associated to the first common factor ( $\mu_{10}$ ) is the second largest value in the distribution of  $\mu_{1j}$ , while in Panel D we consider the case in which it is the second smallest.

<sup>24</sup>For each  $T_0$  and a treated state, we calculate the average bias and the average  $\|\widehat{\mathbf{w}}^{\text{SC}}\|_2$ . Then we consider the correlation between these two variables for each  $T_0$ .

Therefore, again we are in a setting in which it would be possible to construct a SC state that is affected by  $\lambda_{1t}$  in the same way as the treated unit. Still, the results from columns 1 and 2 show that the original and the demeaned SC estimators are biased even when  $T_0$  is large. This happens because the SC weights fail to reconstruct  $\mu_{10}$ , despite the fact that there exist weights that would do so. This is consistent with the results from Propositions 1 and 2. Note also that the biases of the original and demeaned SC estimators are larger when  $T_0$  is smaller. See Ferman and Pinto (2019) for a more thorough discussion on that. Figure 1.B shows the bias of the demeaned SC estimator as a function of the factor loadings of the treated unit. Again, the bias is closer to zero when the treated unit is in the middle of the distribution of  $\mu_{1j}$ . If we consider a measure of concentration of weights given by  $\|\widehat{\mathbf{w}}^{SC}\|_2$ , then the correlation between the absolute value of the bias of the demeaned SC estimator and this measure ranges from 0.723 to 0.810 (depending on the value of  $T_0$ ).

While the biases of the original and the demeaned SC estimators do not converge to zero when there is selection on unobservables, these biases are substantially smaller than the bias of the DID estimator in this setting (column 3). Interestingly, the demeaned SC estimator attenuates the bias from the DID estimator, but does not eliminate it. The idea is that, when we move from the DID to the demeaned SC estimator, the weights move in the direction of weights that reconstruct the factor loadings. However, the SC weights do not generally go all the way through to recover  $\mu_0$  because of the idiosyncratic shocks. Moreover, the DID estimator presents a substantially larger standard error (columns 4 to 6). These results are consistent with the conclusions from Section 3.2, in that the demeaned SC estimator improves relative to DID in terms of bias and variance (under Assumption 5).<sup>25</sup>

Finally, we consider the size and power of the specification test proposed in Section 3.2 (see Appendix Table A.1). When there is no structural break (so both the demeaned SC and the DID estimators are asymptotically unbiased), the test presents relevant overrejection when  $T_0$  is small, but such distortions become less relevant when  $T_0$  increases. Such distortions with finite  $T_0$  arise because the common factors exhibit serial dependence. If we did not have dependence, then the test would be exact. Overall, the fact that the test has some overrejection when  $T_0$  is small is less worrisome than if we had underrejection, because this would lead researchers to be more cautious about the use of the demeaned SC estimator. When we consider the case in which there is a structural break, the test would have power to detect that the demeaned SC and the DID estimators are different, especially when the treated unit is on the extremes of the distribution of  $\mu_{1j}$ . When the treated unit is in the middle of this distribution, then the bias of both the demeaned SC and of the DID estimators become less relevant, so the probability of rejecting specification test becomes lower.

<sup>25</sup>In these simulations, the bias of the original SC estimator is only slightly larger than the bias of the demeaned SC estimator, suggesting that, in this setting, the SC state approximately reconstructs the fixed effect of the treated state. Note, however, that such comparison cannot be extrapolated to other settings, as discussed in Remark 4.

5. EMPIRICAL ILLUSTRATION

As an empirical illustration, we revisit the application presented by [Abadie and Gardeazabal \(2003\)](#). We present in Figure 2.A the per capita GDP time series for the Basque Country and for other Spanish regions, while in Figure 2.B we replicate their Figure 1, which displays the per capita GDP of the Basque Country contrasted with the per capita GDP of a SC unit constructed to provide a counterfactual for the Basque Country without terrorism. We construct three different SC units, with the original SC estimator using all pretreatment outcome lags as predictors, with the demeaned SC estimator using all pretreatment outcome lags as predictors, and with the specification considered by [Abadie](#)

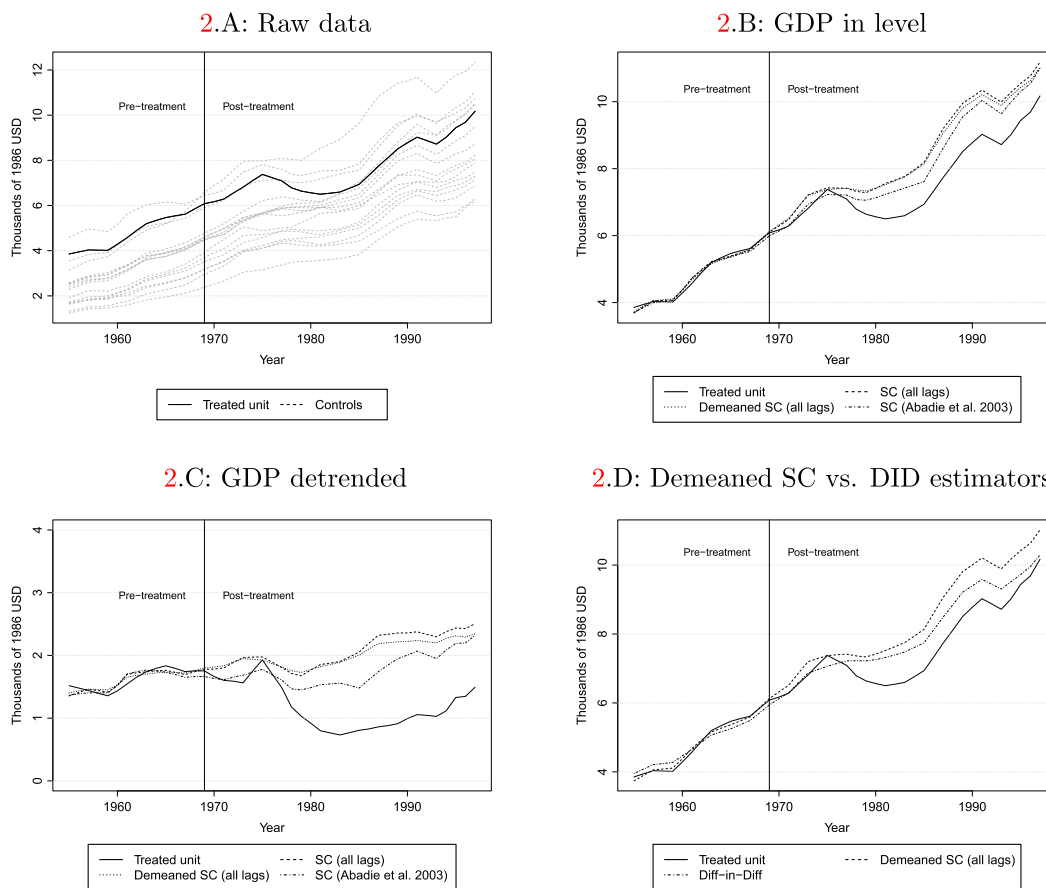


FIGURE 2. [Abadie and Gardeazabal \(2003\)](#) application. Notes: Figure A presents time series for the treated and for the control units used in the empirical application from [Abadie and Gardeazabal \(2003\)](#). In Figure B, we present the time series for the treated and for the SC units. We consider the SC unit estimated with the original SC estimator using all pretreatment periods lags, with the demeaned SC estimator using all pretreatment periods lags, and with the specification considered by [Abadie and Gardeazabal \(2003\)](#). In Figure C, we present the same information as in Figure B after subtracting the control groups' averages for each time period. In Figure D, we present the counterfactuals using the demeaned SC and the DID estimators.

and Gardeazabal (2003). All specifications point out to large negative treatment effects, although the estimated effects are slightly smaller for the original specifications considered by Abadie and Gardeazabal (2003).

Figure 2.B displays a remarkably good pretreatment fit, regardless of the specification. However, the per capita GDP series are clearly nonstationary, with all regions displaying similar trends before the intervention. Considering the results presented in Section 3, such nonstationarity may come either from time fixed effects  $\delta_t$ , or from nonstationary common shocks that may have heterogeneous effects across regions. If the nonstationarity comes from a common factor  $\delta_t$  that affects every unit in the same way, then the series  $\tilde{y}_{jt} = y_{jt} - \frac{1}{J} \sum_{j' \neq 0} y_{j't}$  would not display nonstationary trends. As shown in Figure 2.C, this appears to be the case in this application.<sup>26</sup>

In light of the results from Section 3, the distortions in the SC weights depend on the relative magnitudes of the variance of the *nondiverging* common factors relative to the variance of the idiosyncratic shocks. Therefore, Figure 2.C provides a better visual assessment of whether the pretreatment fit is good relative to Figure 2.B. While the pretreatment fit is still reasonably good after we discard the nonstationary part of the series, it is not as good as when we consider the series in levels.

We can consider whether we would be able to justify the use of the SC method in this setting without relying on theoretical results based on perfect pretreatment fit approximations. First, note that the estimated weights in this application are very concentrated among a few control regions, so we cannot rely on the theoretical results from Ferman (2019) to argue that the SC estimator is asymptotically unbiased in this setting (see Appendix Table A.2). Following the discussion in Section 3.2, we also contrast in Figure 2.D the DID and the demeaned SC estimators. If they were very similar, then we would have some support to rely on these estimators even if they failed to reconstruct the factor loadings of the treated region. However, we find the the estimated effect using the demeaned SC estimator is systematically larger. The  $p$ -value of the specification test proposed in Proposition 4 is 0.023. This suggests that we may have selection on time-varying unobservables, implying that both the demeaned and the DID estimators are asymptotically biased, although the bias of the demeaned SC estimator should be smaller.

Overall, since in this particular application the pretreatment fit is reasonably good even once we subtract the nonstationary trends, and the treatment effects are large relative to the pretreatment gaps, we should expect that any potential bias from the demeaned SC estimator does not explain a large proportion of the estimated effects. Moreover, given the discussions from Sections 3.2 and 4, we should expect the demeaned SC estimator to *partially* control for any bias that the DID estimator experience. Since the estimated effects with the demeaned SC estimator are stronger than the DID estimates, given this rationale, we should expect, if anything, that the demeaned SC estimator would provide a lower bound on the (absolute values of the) treatment effects. Therefore, a careful analysis of the potential problems of the SC method would not change the main conclusions from this empirical application. Still, in other settings in which the

<sup>26</sup>Given the adding-up constraint, note that the SC estimator is numerically the same if we estimated it using the original data or  $\tilde{y}_{jt}$ . If there were other sources of nonstationarity, then the series would remain nonstationary even after such transformation. In such cases, other strategies to detrend the series could be used, such as, for example, considering parametric trends.

pretreatment fit is worse, and in which moving from the DID to the demeaned SC estimator leads to weaker results, then it would not be possible to rely on the arguments used above, and the problems we highlight in this paper may undermine conclusions from the SC method.

## 6. RECOMMENDATIONS

Taken together, our results clarify the conditions in which the SC and related estimators can be reliably used, when we consider a setting in which potential outcomes are well approximated by a linear factor model. Based on these results, we provide guidance on how applied researchers could justify the use of these methods. First, a condition like the one we present in Assumption 3 is always necessary to justify the SC estimator. It states that treatment assignment is not related to shocks that are specific to the treated unit. It does allow, however, for unobserved confounders that may also affect other control units. Indeed, the main reason why a researcher should use these kind of methods is if he/she believes that there may be confounding factors that also affect the control units. In this case, information from the control units could be used to control for such confounders. Therefore, any applied paper relying on the SC method should discuss the possible unobserved confounders in the specific application, and argue that such confounders are not specific to the treated unit.

Importantly, even if it is plausible that idiosyncratic shocks are not correlated with the treatment assignment, whether the SC method is able to reliably control for the common shocks depends crucially on details of the empirical application. There are two settings that provide validity for the SC estimator even when there are time-varying unobserved confounders. First, [Abadie, Diamond, and Hainmueller \(2010\)](#) showed that the SC estimator is reliable if the pretreatment fit is good for a large number of pretreatment periods. This condition can be checked by contrasting the outcomes of the treated and of the SC units in the pretreatment periods. Based on our results, we recommend that applied researchers should also consider the pretreatment fit after discarding diverging trends, in order to provide a better understanding of the relative magnitude between the variances of the nondiverging common factors and of the idiosyncratic shocks.<sup>27</sup> Also, it is important that the number of control units in this case cannot be large in comparison to  $T_0$ , otherwise a good pretreatment fit might be a consequence of overfitting. In this case, the bias of the SC estimator we uncover in our paper may remain relevant even if we have a good pretreatment fit.

Second, when both  $J$  and  $T_0$  are large, [Ferman \(2019\)](#) showed that the SC estimator may be asymptotically unbiased even when the pretreatment fit is imperfect. This would be the case if the confounders affect a large number of control units, and in this case the SC weights would get diluted among an increasing number of control units when  $J \rightarrow \infty$ . Therefore, we recommend that applied researchers also report the  $L_2$  norm of the SC weights,  $\|\widehat{\mathbf{w}}^{\text{SC}}\|_2$ . If this is close to zero, then we would have evidence that we are

<sup>27</sup>Given the adding-up constraint, note that the SC estimator is numerically the same if we estimated it using the original data or if we detrend the data by subtracting a term  $a_t$  for all units in period  $t$ . This will be the case if we have a setting as the one considered in Section 5.

closer to the setting considered by [Ferman \(2019\)](#). In contrast, if the weights are concentrated, then we would have evidence that the bias we uncover in our paper is potentially relevant. As we show in our MC simulations in Section 4, the cases in which we find largest biases are exactly the ones in which the SC weights are more concentrated.

The results we derive in Section 3 are informative about the properties of the SC estimator when the conditions outlined by [Abadie, Diamond, and Hainmueller \(2010\)](#) and [Ferman \(2019\)](#) do not hold. This would be the case when (i) the pretreatment fit is imperfect and  $J$  is not large, (ii) the pretreatment fit is imperfect with large  $J$  and  $T_0$ , but the SC weights are not diluted among a large number of control units, or (iii) the pretreatment fit is good, but  $J$  is much larger than  $T_0$ , so such pretreatment fit is possibly good due to overfitting.

In these cases, we show that the SC estimator can still provide important gains relative to the DID estimator, but the applied researcher should be more careful in justifying the use of the method. If one considers the demeaned SC estimator, then the assumptions for unbiasedness would be the same as those for the DID estimator. That is, the researcher should argue that the relevant unobserved confounders are not time varying. The advantage of relying on the demeaned SC estimator relative to DID in this case is that it would be more efficient if common shocks are stable before and after the treatment, and that it should have lower bias in case there is correlation between treatment assignment and time-varying unobservables. We also show that contrasting the DID and the demeaned SC estimators is informative about whether these conditions for unbiasedness are valid, and propose a specification test based on that. If we find evidence that these two estimators are similar, then we should be more confident that the conditions for asymptotic unbiasedness of the demeaned SC estimator holds even when we are not in the settings considered by [Abadie, Diamond, and Hainmueller \(2010\)](#) or [Ferman \(2019\)](#).

Importantly, if the conditions considered by [Abadie, Diamond, and Hainmueller \(2010\)](#) or [Ferman \(2019\)](#) hold in a specific application, then the demeaned SC estimator would be asymptotically unbiased, while the DID estimator may be biased. In this case, any information from the specification test indicating that the demeaned SC and the DID estimators are different would not imply that the demeaned SC estimator is invalid. Therefore, it is crucial to understand the conditions under which each of these estimators are valid to interpret the conclusions from this specification test.

Finally, if one considers the original SC estimator, then one should have to inspect the pre-treatment fit. If the SC unit recovers the levels of the treated unit (even if the pretreatment fit is imperfect), then again the estimator would be reliable if there is no relevant time-varying unobserved confounders. If the SC unit does not recover the levels, then the original SC estimator should not be used. In such settings, the researcher should either use the demeaned SC estimator, or discard such application in case he/she does not want to rely on extrapolation.

## 7. CONCLUSION

We consider the properties of the SC and related estimators, in a linear factor model setting, when the pretreatment fit is imperfect. We show that, in this framework, the

SC estimator is generally biased if treatment assignment is correlated with the unobserved heterogeneity, and that such bias does not converge to zero even when the number of pretreatment periods is large. Still, we also show that a modified version of the SC method can improve relative to DID, even if the pretreatment fit is not close to perfect and if  $T_0$  is not large. Overall, we show that the SC method can provide substantial improvement relative to DID, even in settings where the method was not originally designed to work. However, researchers should be more careful in the evaluation of the identification assumptions in those cases. Importantly, our results clarify the conditions in which the SC and related estimators are reliable, and provide practical guidance on how applied researchers should justify the use of such estimators in empirical applications.

#### REFERENCES

- Abadie, A. (2020), “Using synthetic controls: Feasibility, data requirements, and methodological aspects.” *Journal of Economic Literature*. [1198]
- 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. [1197, 1198, 1199, 1202, 1203, 1206, 1217, 1218]
- Abadie, A., A. Diamond, and J. Hainmueller (2015), “Comparative politics and the synthetic control method.” *American Journal of Political Science*, 59 (2), 495–510. [1197, 1203, 1210]
- Abadie, A. and J. Gardeazabal (2003), “The economic costs of conflict: A case study of the Basque country.” *American Economic Review*, 93 (1), 113–132. [1197, 1215, 1216]
- Amjad, M., D. Shah, and D. Shen (2018), “Robust synthetic control.” *J. Mach. Learn. Res.*, 19 (1), 802–852. [1200]
- Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics—Applied Probability and Statistics Section Series. Wiley. [1199]
- Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2018), “Synthetic difference in differences.” E-prints, December 2018, [arXiv:1812.09970](https://arxiv.org/abs/1812.09970). [1200]
- Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2018), “Matrix completion methods for causal panel data models.” Working Paper 25132, National Bureau of Economic Research. [1200]
- Athey, S. and G. Imbens (2018), “Design-based analysis in difference-in-differences settings with staggered adoption.” [arXiv:1808.05293](https://arxiv.org/abs/1808.05293). Working Paper. [1207]
- Athey, S. and G. W. Imbens (2017), “The state of applied econometrics: Causality and policy evaluation.” *Journal of Economic Perspectives*, 31 (2), 3–32. [1197]

- Bai, J. (2003), “Inferential theory for factor models of large dimensions.” *Econometrica*, 71 (1), 135–171. [1199, 1200]
- Bai, J. (2009), “Panel data models with interactive fixed effects.” *Econometrica*, 77 (4), 1229–1279. [1200, 1211]
- Bai, J. and S. Ng (2002), “Determining the number of factors in approximate factor models.” *Econometrica*, 70 (1), 191–221. [1200, 1211]
- Ben-Michael, E., A. Feller, and J. Rothstein (2018), “The augmented synthetic control method.” arXiv:1811.04170. Working Paper. [1199, 1202]
- Botosaru, I. and B. Ferman (2019), “On the role of covariates in the synthetic control method.” *Econometrics Journal*, 1. [1202]
- Callaway, B. and P. H. C. Sant’Anna (2018), “Difference-in-differences with multiple time periods and an application on the minimum wage and employment.” arXiv:1803.09015. Working Paper. [1207]
- Carvalho, C. V., R. Masini, and M. C. Medeiros (2018), “ArCo: An artificial counterfactual approach for aggregate data.” *Journal of Econometrics*. (Forthcoming). [1200, 1210]
- Chernozhukov, V., K. Wuthrich, and Y. Zhu (2017), “An exact and robust conformal inference method for counterfactual and synthetic controls.” E-prints, December 2017, arXiv:1712.09089. [1209]
- Carvalho, C. V., R. Masini, and M. Cunha Medeiros (2016), “The perils of counterfactual analysis with integrated processes.” *Textos para discussão 654*, PUC-Rio (Brazil), Department of Economics. [1200, 1210]
- de Chaisemartin, C. and X. D’Haultfauille (2020), “Two-way fixed effects estimators with heterogeneous treatment effects.” *American Economic Review*, 110 (9), 2964–2996. [1207]
- Dickey, D. A. and W. A. Fuller (1981), “Likelihood ratio statistics for autoregressive time series with a unit root.” *Econometrica*, 49 (4), 1057–1072. [1211]
- Doudchenko, N. and G. W. Imbens (2016), “Balancing, regression, difference-in-differences and synthetic control methods: A synthesis.” Working Paper 22791, National Bureau of Economic Research. [1198, 1199, 1203, 1206, 1207]
- Ferman, B. (2019), “On the properties of the synthetic control estimator with many periods and many controls.” E-prints, June 2019, arXiv:1906.06665. [1200, 1216, 1217, 1218]
- Ferman, B. and C. Pinto (2019), “Synthetic controls with imperfect pre-treatment fit.” arXiv:1911.08521. version from November 19th, 2019. [1200, 1214]
- Ferman, B., C. Pinto (2021), “Supplement to ‘Synthetic controls with imperfect pretreatment fit.’” *Quantitative Economics Supplemental Material*, 12, <https://doi.org/10.3982/QE1596>. [1198]
- Ferman, B., C. Pinto, and V. Possebom (2020), “Cherry picking with synthetic controls.” *Journal of Policy Analysis and Management*, 39 (2), 510–532. [1203]



Gobillon, L. and T. Magnac (2016), “Regional policy evaluation: Interactive fixed effects and synthetic controls.” *The Review of Economics and Statistics*, 98 (3), 535–551. [1200, 1202]

Goodman-Bacon, A. (2018), “Difference-in-differences with variation in treatment timing.” Working Paper 25018, National Bureau of Economic Research. [1207]

Hsiao, C., H. S. Ching, and S. K. Wan (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 (5), 705–740. [1199, 1200, 1207, 1210]

Hyndman, R. J. and Y. Khandakar (2008), “Automatic time series forecasting: The forecast package for R.” *Journal of Statistical Software*, 26 (3), 1–22. [1211]

Kaul, A., S. Klöbner, G. Pfeifer, and M. Schieler (2015), “Synthetic control methods: Never use all pre-intervention outcomes as economic predictors.” Available at [www.oekonometrie.uni-saarland.de/papers/SCM\\_Predictors.pdf](http://www.oekonometrie.uni-saarland.de/papers/SCM_Predictors.pdf). [1203]

Li, K. T. and D. R. Bell (2017), “Estimation of average treatment effects with panel data: Asymptotic theory and implementation.” *Journal of Econometrics*, 197 (1), 65–75. [1200, 1210]

Masini, R. and M. C. Medeiros (2019), “Counterfactual analysis with artificial controls: Inference, high dimensions and nonstationarity.” Available at SSRN <https://ssrn.com/abstract=3303308>. [1200]

Moon, H. Roger, and M. Weidner (2015), “Linear regression for panel with unknown number of factors as interactive fixed effects.” *Econometrica*, 83 (4), 1543–1579. [1200]

Ng, S. and P. Perron (2001), “LAG length selection and the construction of unit root tests with good size and power.” *Econometrica*, 69 (6), 1519–1554. [1211]

Powell, D. (2018), “Imperfect synthetic controls: Did the Massachusetts health care reform save lives?” Available at SSRN <https://ssrn.com/abstract=3192710>. [1200]

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015), “Integrated public use microdata series: Version 6.0 [Machine-readable database].” [1211]

Xu, Y. (2017), “Generalized synthetic control method: Causal inference with interactive fixed effects models.” *Political Analysis*, 25 (2), 57–76. [1200, 1211]

---

Co-editor Andres Santos handled this manuscript.

Manuscript received 8 April, 2020; final version accepted 19 March, 2021; available online 31 March, 2021.